Essays on Simulation-based Methodology with Many Auxiliary Statistics

by

Wenqian Sun

M.A., Simon Fraser University, 2016
M.Sc., Columbia University, 2015
B.Econ., Central University of Finance and Economics, 2013

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Economics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Summer 2023

Declaration of Committee

Name:	**Wenqian Sun**
Degree:	**Doctor of Philosophy**
Thesis title:	**Essays on Simulation-based Methodology with Many Auxiliary Statistics**
Committee:	**Chair:** Simon Woodcock Associate Professor, Economics
	Bertille Antoine Supervisor Professor, Economics
	Dongwoo Kim Committee Member Assistant Professor, Economics
	Brian Krauth Examiner Associate Professor, Economics
	Marine Carrasco External Examiner Professor, Economics University of Montreal

Abstract

The thesis focuses on simulation-based methods for estimation and inferences when many auxiliary statistics are available. In the first chapter, we establish the consistency of the simulated minimum distance estimator in such a situation and derive its asymptotic distribution. Our estimator contributes to the asymptotic theory for estimators obtained by simulated minimum distance in situations where the number of auxiliary statistics (or the number of matched moments) is large, which has not been covered in the existing literature. The estimator is easy to implement and allows us to exploit all the informational content of a large number of auxiliary statistics without having to, (i) know these functions explicitly, or (ii) choose *a priori* which functions are the most informative. This allows us to exploit, among other things, long-run information.

In the second chapter, we illustrate the implementation of the proposed method through Monte-Carlo simulation experiments based on small- and medium-scale New Keynesian models. These examples illustrate how to exploit information from matching a large number of impulse responses including at long-run horizons. It is revealed that the utilization of many auxiliary statistics and data-driven regularization effectively improves estimation in terms of precision and coverage rate.

In the third chapter, we propose tests of the null hypothesis of autoregressive models against ones with Markov-switching autoregressive components. The empirical simulation-based tests allow for unknown distributions and use Monte-Carlo test techniques. The approach is flexible and computationally simple. The designed test statistic allows for a large number of empirical moments and relies on simulations under the null hypothesis which permits the use of higher-order moments. Our simulation experiments demonstrate that more information can be harvested with more moments matched, with evidence of increased empirical power. The Monte-Carlo testing methodology is illustrated with a mean-variance switching model, an autoregressive coefficient switching model, and an application to US output growth modeling.

Keywords: Many moments; Bootstrap; Impulse response matching; Monte-Carlo tests; Markov-switching

Dedication

For my parents and my husband, who raise me up to more than I can be.
感谢我的父亲孙建平、母亲闻豫给予我最大的支持和包容。
感谢我的爱人张昊玮十年如一日的鼓励和陪伴。

For my friends Tianyi Chen, Mo Li, Yun Shen, Zhe Zhang, and Viona Zhuang, who volunteer to enter the library database of SFU with my thesis as digital lives.

Acknowledgements

I am deeply grateful to my supervisor Bertille Antoine for her exceptional guidance, encouragement, caring, and wisdom. Her continuous support and brilliant advice have guided me through my Ph.D. years, and I feel fortunate to have had the opportunity to work with her.

I am also grateful to the anonymous referees for their insightful comments, which have helped to improve the quality of our work.

I want to express my appreciation to my Ph.D. cohort, including Shirleen Manzur, Xiaolin Sun, and Boxi Yang, for their unwavering support and friendship.

I am thankful to the professors and staff in the department for their valuable advice and assistance.

Finally, I want to convey my most sincere thanks to Meiyu Li, without whom I would never have become an econometrician. I miss her dearly.

Table of Contents

Chapter 1

Simulation-based Estimation with Many Auxiliary Statistics: Theory

1.1 Introduction

Estimation methods based on simulation with auxiliary statistics (or SAS) have become very popular to estimate the underlying parameters of complex structural models, and include such estimators as Indirect Inference (I-I, Gouriéroux et al. (1993)), Simulated Method of Moments (Duffie and Singleton (1993), Smith (1993)), or Efficient Methods of Moments (Gallant and Tauchen (1996)); see Forneron and Ng (2018) and associated references for a recent review. These estimation procedures have the advantage of bypassing the characterization of a likelihood function - often difficult to obtain for complex models - by focusing instead on "matching" auxiliary statistics chosen to summarize key features of the data generating process of interest. More specifically, these estimators are obtained by minimizing the distance between the auxiliary statistic computed with observed data and an average of the auxiliary statistics computed with simulated data for a given parameter value.

The main objective of this chapter is to consider the extension of the above-mentioned SAS or I-I estimation procedures to the case where a large number of auxiliary statistics is chosen to estimate the finite-dimensional parameter of interest, so-called simulation with many auxiliary statistics or SMAS. We will consider cases where the number of components of the vector of auxiliary statistics is large and typically larger than the sample size. Our framework offers two main advantages. First, the practitioner does not need to select *a priori* a small number of auxiliary statistics. In general, it is actually difficult to determine which statistics are most informative. Second, long-run information can easily be incorporated: for example, the long-run dynamic responses of macro variables to unitary shocks contain information that can be harvested by including their impulse responses at large horizons in the auxiliary statistics.

With a finite-dimensional vector of auxiliary statistics denoted $\hat{\psi}$, a SAS or I-I based estimator minimizes the L_2 distance between $\hat{\psi}$ and an average of the auxiliary statistics computed with simulated data for given parameter value θ, say $\hat{\psi}^s(\theta)$. When the number of matched auxiliary statistics becomes large, the norm is rather determined in a Hilbert space and requires the introduction of an operator. We establish the consistency of the associated SMAS-based estimator in this situation and derive its asymptotic distribution. We also derive the optimal (covariance) operator that delivers asymptotic efficiency and design a bootstrap-based procedure to estimate it. To implement our efficient estimator, it is necessary to invert the optimal operator which is highly unstable due to the large number of underlying auxiliary statistics. We rely on Tikhonov regularization to stabilize its inverse, and design a cross-validation procedure to choose the associated tuning parameter.

Our work contributes to the literature on minimum distance estimation of a finite-dimensional vector of parameters when a large number of moments is available. More specifically, we build on Carrasco and Florens (2000) who extend the generalized method of moments procedure to the case of a continuum of moment conditions. We consider instead auxiliary statistics that are not always moments, and that are not necessarily known analytically but rather simulation-based. In that sense, our work is also related to section 5 in Carrasco et al. (2007) where the authors explain how to handle characteristic-function based estimation when the characteristic function is not available in closed form (e.g. because the model involves latent variables). In such a case, ML efficiency is still achievable and the associated optimal operator is obtained through the same kernel-based estimation as with a tractable characteristic function. We instead need to design a residual-based bootstrap to estimate the optimal operator. Further, it is important to mention that our interest in considering a large number of auxiliary statistics is not directly related to efficiency in the sense that we have no hope of achieving the Cramer-Rao efficiency bound in the *complex structural models* we have in mind - even with such a large number of matched auxiliary statistics. Our motivation to consider a large number of statistics stems from two main practical reasons: (i) to avoid the *ad hoc* pre-selection of a small number of statistics; (ii) to incorporate information from the DGP that can only be harvested - as far as we know - from allowing a large number of auxiliary statistics: e.g. by letting the horizon of matched impulse responses grow to infinity to incorporate long-run dynamic responses of key macro variables.

The main theoretical results of the SMAS estimator are presented in this chapter. In Section 1.2, the estimation methodology and general algorithm are introduced. Section 1.3 demonstrates the asymptotic normality of the SMAS estimator and its optimal asymptotic covariance. The implementation of the estimator in practice with regularized inversion is explained in Section 1.4. Section 1.5 concludes. The proofs of our theoretical results are presented in Appendix A.

1.2 Framework and notations

We start by introducing our framework and notations through a brief review of the traditional SAS or I-I based estimation procedure. It can be understood as an extension of classical minimum distance estimation (such as GMM) when the underlying "moments" are not analytically tractable, but can easily be evaluated on simulated data.

Consider the sample of observed data of length T denoted $\mathcal{X}_T = (x_1, \cdots, x_T)'$. We assume throughout that $\mathcal{X}_T$ are stationary and can be represented by a parametric model with probability measure P_{θ^0} where $\theta^0 \in \Theta \subset \mathbb{R}^p$. In that sense, our interest lies in estimating θ^0 the "true" unknown value of the parameter θ that has generated the data. SAS or I-I based estimation traditionally relies on matching a vector of $H \geq p$ auxiliary statistics $\hat{\psi}_T \equiv \hat{\psi}(\mathcal{X}_T)$ evaluated at the observed data $\mathcal{X}_T$ with its counterpart evaluated on simulated data $\hat{\psi}_T^s(\theta) \equiv \hat{\psi}_s(\mathcal{X}_T^s(\theta))$ where $\mathcal{X}_T^s(\theta) \equiv \mathcal{X}_T^s(\epsilon^s, \theta)$ represents a sample of T simulated data for given θ with errors ϵ^s drawn from an assumed distribution F_ϵ. The SAS or I-I estimator is then defined by minimizing the L_2 norm between $\hat{\psi}_T$ and $\hat{\psi}_T^s(\theta)$, specifically[1]:

$$\arg\min_{\theta\in\Theta} \left[z_T'(\theta) W_T z_T(\theta) \right] \qquad \text{with} \qquad z_T(\theta) = \hat{\psi}_T - \frac{1}{S}\sum_{s=1}^{S} \hat{\psi}_T^s(\mathcal{X}_T^s(\theta)) \,,$$

for some weighting matrix W_T of size (H, H) that converges to a positive-definite matrix W.

We now propose to generalize SAS or I-I based estimation to allow for a large (possibly infinite) number of auxiliary statistics to be matched to estimate θ. Accordingly, we introduce the "distance" function, $z_T(., \theta)$, the real-valued function defined over the set of integers[2] $\mathbb{N}$ which corresponds to the difference between the auxiliary statistic computed on observed and simulated data. Intuitively, we are looking for the value of θ that will make $z_T(., \theta)$ as close as possible to 0. Following Carrasco and Florens (2000) - and forgetting for the time being that $z_T(.)$ is not analytically tractable - the appropriate norm is defined in the Hilbert space of squared-integrable real-valued functions through a linear operator denoted B and our estimator based on simulation with many auxiliary statistics (or SMAS) is defined as

$$\hat{\theta}_{SMAS} \equiv \arg\min_{\theta\in\Theta} \|B_T z_T(., \theta)\| \quad \text{with} \quad z_T(h, \theta) = \hat{\psi}_T(h) - \frac{1}{S}\sum_{s=1}^{S} \hat{\psi}_T^s(h, \mathcal{X}_T^s(\theta)) \quad \forall h \in \mathbb{N} \,,$$

[1] It is common practice to consider the average of $\hat{\psi}^s(\theta)$ obtained with S simulated paths $\mathcal{X}_T^s(\theta)$ ($s = 1, \cdots, S$); however S can be as small as 1 as discussed in Gouriéroux et al. (1993).

[2] For ease of exposition, we present our results with a function $z(., \theta)$ defined over the set of integers. However, our results are not restricted to this particular indexation and can be extended to e.g. $h \in [0, 1]$; see related discussions in Carrasco and Florens (2000).

where B_T converges to B.

The introduction of the Hilbert space (and associated operator B) provides a convenient framework that allows the information contained in the entire function $z(.,\theta)$ to be harvested - rather than evaluating it at a small number of chosen points, say $[z_T(h_1,\theta), z_T(h_2,\theta), \cdots, z_T(h_J,\theta)]$. Compared to the literature that relies on "many instruments", when $J \to \infty$, our framework also avoids having to specify the growth rate of J in relation to the sample size T.

For example, the impulse response matching estimator of Guerron-Quintana et al. (2017) relies on a finite vector of auxiliary statistics chosen as the first J impulse responses of key macro variables: hence, $z_T(\theta)$ is a vector of size J with components $z_T(h,\theta)$ with $h = 1, \cdots, J$. Such an estimator focuses on the short-run dynamic behavior of these macro variables. We propose instead to consider infinitely many impulse responses in order to incorporate the long-run dynamic behavior of these variables through $z(h,\theta)$ for any $h \in \mathbb{N}$. In our simulation study in section 2.1, we show that long-run information can easily be used to estimate the structural parameters of interest.

We conclude this section by providing the algorithm that describes the key steps of our simulation-based approach with many auxiliary statistics taken as impulse responses.

Algorithm 1. *(Practical implementation)*

1. *Using the sample of T observations, compute the chosen impulse responses $\hat{\psi}_T(\mathcal{X}_T)$ as well as the transition matrix and the residuals $\hat{\epsilon}_T$.* [3]
2. *For given $\theta \in \Theta$, use the simulator to generate S independent samples of T observations; compute the associated (chosen) impulse responses, $\hat{\psi}_T^s(\theta)$ with $s = 1, \cdots, S$, as well as $z_T(\theta) = \hat{\psi}_T(\mathcal{X}_T) - \sum_s \hat{\psi}_T^s(\theta)/S$.*
3. *Estimate the optimal operator $\hat{K}_{T,a}^{-1/2*}$ (a the regularizing parameter).*
   (a) *Re-sample with replacement from the residuals $\hat{\epsilon}_T$ to get $\epsilon_{T,n}^*$.*
   (b) *Use the transition matrix and $\epsilon_{T,n}^*$ to generate $\mathcal{X}_{T,n}^*$ and compute $\hat{\psi}_{T,n}^*$.*[4]
   (c) *Repeat independently N times to get $\hat{\psi}_{T,n}^*$ with $n = 1, \cdots, N$.*
   (d) *Compute $\hat{K}_T^{-1/2*}$ using $\hat{\psi}_{T,n}^*$ by following the procedure described in section 1.4.*
4. *Obtain $\hat{\theta}_{SMAS}$ as the minimizer over θ of $\|\hat{K}_{T,a}^{-1/2*} z_T(\theta)\|$.*

1.3 Asymptotic properties of the SMAS estimator

In this section, we present our main theoretical results, namely consistency and asymptotic normality of our SMAS estimator.

[3] Computation of structural impulse response functions with a VAR model is explained in Appendix B.1.

[4] Alternatively, compute $\hat{\psi}_{T,n}^*$ with the simulator, selected θ, and error term ϵ_T (the *double-bootstrap* approach). See section 1.4 for further explanation and Algorithm 4 in Appendix A.2.

1.3.1 Consistency of the SMAS estimator

Let X be a random element defined on a complete probability space $(\Omega, \mathcal{F}, P_0)$ which can be represented by a parametric model with probability measure $P_0 \equiv P_{\theta^0}$ where $\theta^0 \in \Theta \subset \mathbb{R}^p$. X takes its values in $(S, \mathcal{S})$.

Assumption 1. *(Data Generating Process)*
The observed sequence $\mathcal{X}_T = (x_1, \ldots, x_T)$ is a stationary realization of the stochastic process X.

To formally characterize the simulated data $\mathcal{X}_T^s(\theta)$, let $(\Omega^s, \mathcal{F}^s, P^s)$ denote the associated probability space, and let X^s be a random element defined on the product probability space $(\Omega, \mathcal{F}, P_0) \times (\Omega^s, \mathcal{F}^s, P^s)$ that takes its values in $(S^s, \mathcal{S}^s)$. The joint probability measure is denoted $\mathbb{P} \equiv P_0 \times P^s$.

Definition 1. *(Simulator)*
For a fixed vector θ of size p, $\mathcal{X}_T^s(\theta) \equiv \mathcal{X}_T^s(\epsilon_s, \theta)$ denotes a sample of size T of data simulated under θ with errors ϵ_s drawn from an assumed distribution F_ϵ.

Several comments are worth mentioning.

(i) In many cases, the observed data contains some endogenous variables y_t and some exogenous variables w_t, so that $x_t = (y_t, w_t)$. In such cases, the simulator will deliver $(y_1^s, \cdots, y_T^s)$ for given θ, $(w_1, \cdots, w_T)$ and some initial conditions $\underline{y}_0$. To simplify our notations, we keep referring to the simulated data as $\mathcal{X}_T^s(\theta)$; see e.g. Gouriéroux and Monfort (1996) for extensive discussions.

(ii) The parameter of interest (say θ_1) is typically a subset of the full set of parameters θ (with $\theta = (\theta_1, \theta_2)$) needed to simulate the chosen auxiliary statistics, while the remaining parameter θ_2 can be seen as nuisance parameters: e.g. the underlying parameters of the distribution F_ϵ. To simplify our presentation, we work on the full vector of parameters θ. For alternatives, see Dridi et al. (2007) who introduce the Partial Indirect Inference where θ_2 is estimated, Khalaf et al. (2019) who conduct fully parametric inference in a DSGE framework where θ_2 is known under the null, Khalaf and Saunders (2019) who derive statistics invariant[5] to θ_2 for inference in autoregressive panels, and Antoine et al. (2023) who extend SAS or I-I based inference to allow for θ to be weakly identified and for θ_2 to be *approximately* calibrated in the sense that it may not be correctly calibrated.

As explained in the previous section, our estimation strategy consists in *matching* the chosen auxiliary statistic computed on observed data with that computed on simulated

[5] As models' complexity increases, invariant statistics are typically hard to come by: for example in models such as DSGE.

data. To formalize our analysis, we introduce the real-valued distance function $z(.,\theta)$ as the difference between the (population) function of interest denoted $\psi(.)$ and its simulation-based counterpart denoted $\psi^s(.)$. The function $\psi(.)$ implicitly depends on the DGP X, while the function $\psi^s(.)$ depends on the simulator X^s and the associated vector of parameters θ.

Definition 2. *(Distance function)*
The distance function $z(.)$ is defined on $(\mathbb{N} \times S \times S^s \times \Theta)$ as

$$z(h, X, X^s, \theta) \equiv \psi(h, X) - \psi^s(h, X^s, \theta)\,.$$

It is important to mention that $\psi(.)$ is not itself random, as it rather corresponds to the *population* function of interest. To fix ideas, consider the following two examples:

- when matching moments of X, $\psi(h, X)$ corresponds to the moment of order h of X computed with respect to P_0;

- when matching structural impulse responses - as in our simulation study - $\psi(h, X)$ corresponds to the dynamic response of interest at horizon h which can be expressed implicitly as a function of the first and second moments of X; see e.g. discussions in Guerron-Quintana et al. (2017) p146.

Our analysis requires the introduction of $\mathbb{H}$, the Hilbert space of square integrable real-valued functions defined over the set of integers with the inner product $(.,.)$ and associated norm $\|.\|$. Specifically, the inner product is defined[6] as

$$(f, g) = \sum_{j \in \mathbb{N}} f_j g_j\,.$$

Assumption 2. *(Regularity of the auxiliary statistics)*

(i) As a function of $h \in \mathbb{N}$, the distance function $z(.)$ (see Definition 2) belongs to the Hilbert space $\mathbb{H}$. It is also a measurable function of (h, X, X^s) for any θ, and it is continuous in θ, $\forall \theta \in \Theta \subset \mathbb{R}^p$ with Θ compact. When there is no confusion, we simply write $z(.,\theta)$ or $z(\theta)$.

(ii) The equation, $z(h, \theta) = 0$ for all $h \in \mathbb{N}$, almost surely, has a unique solution θ_0 in the interior of Θ.

Assumption 2(i) requires $z(.,\theta)$ to be square integrable (as an element of the Hilbert space $\mathbb{H}$). When matching impulse responses, this follows from Assumption 1 maintained

[6]Our setup implicitly rules out applications where $z(.)$ is not square-integrable. Alternatively, we could consider square-integrability with respect to a given probability measure, which would alter the definition of the inner product. As explained after Assumption 2, in our applications of interest $z(.)$ is indeed square-integrable, and we proceed without introducing such a probability measure.

on the underlying DGP: in general, it is known that weakly dependent sequences are characterized by absolutely summable impulse response coefficients of their Wold decomposition; see e.g. Hassler and Kokoszka (2010). Assumption 2(ii) is an identification assumption: intuitively, θ_0 is the only value of the (unknown) parameter θ that allows for a "perfect match" between $\psi(.)$ and $\psi^s(.)$ for all possible values of h.

Assumption 3. *(Operator)*

(i) *B is a nonrandom, nonsingular, bounded linear operator defined on $\mathcal{D}(B) \subset \mathbb{H}$ valued in $\mathbb{H}$. The operator does not depend on θ but may depend on θ_0.*

(ii) *$z(.,\theta) \in \mathcal{D}(B)\ \forall \theta \in \Theta$.*

Assumption 3 maintains regularity assumptions on the operator B to ensure that the population objective function (or the norm of the operator applied to z) is well-defined and uniquely minimized at θ_0. Since B is assumed non-singular, its null space is equal to $\{0\}$ which ensures that the equation $Bz(.,\theta) = 0$ has a unique solution.

Assumption 4. *(Sample counterparts of the operator and objective function)*

(i) *Let B_T be a sequence of nonsingular random bounded linear operators such that $B_T : \mathcal{D}(B_T) \subset \mathbb{H} \to \mathbb{H}$. Let $z_T(.,\theta)$ denote the sample counterpart of $z(.,\theta)$, that is the difference between the estimated auxiliary statistics obtained with observed and simulated data. We assume that $z_T(.,\theta) \equiv z_T(.,\mathcal{X}_T,\mathcal{X}_T^s,\theta) \in \mathcal{D}(B_T), \forall\theta$ and that $Q_T(\theta) \equiv \|B_T z_T(.,\theta)\|$ is a continuous function of θ.*

(ii) *$Q_T(\theta) \xrightarrow{\mathbb{P}} Q_0(\theta) \equiv \|Bz(.,\theta)\|$ uniformly in $\theta \in \Theta$.*

For details on how to compute impulse response functions, see e.g. Hamilton (1994).

Theorem 1. *(Consistency of the SMAS estimator)*
Under Assumptions 1 to 4, the SMAS estimator defined as,

$$\hat{\theta}_{SMAS} \equiv \arg\min_{\theta\in\Theta} Q_T(\theta)\,, \tag{1.1}$$

is consistent for θ_0, that is $\hat{\theta}_{SMAS} \xrightarrow{\mathbb{P}} \theta_0$.

1.3.2 Asymptotic distribution

In order to derive the asymptotic distribution of our SMAS estimator, additional regularity conditions are needed.

Assumption 5. *(Differentiability)*

(i) *The function $\theta \to z(h,\theta)$ is differentiable with respect to θ with $G_j(.,\theta) \equiv \partial z(.,\theta)/\partial\theta_j \in \mathcal{D}(B)$ for $j = 1,\cdots,p$.*

(ii) The (p,p)-matrix $\Big(BG(.,\theta), BG(.,\theta)\Big) = \|BG(.,\theta)\|^2$ with element (i,j) defined as $(BG_i(.,\theta), BG_j(.,\theta))$ (for $i,j = 1,\cdots,p$), is positive definite and symmetric.

Assumption 6. *(Commutativity)*

(i) For any functions $u(.,\theta)$ and $v(.,\theta)$ in $\mathbb{H}$, we have:

$$\frac{\partial}{\partial\theta'}(u(.,\theta), v(.,\theta)) = \Big(\frac{\partial}{\partial\theta'}u(.,\theta), v(.,\theta)\Big) + \Big(u(.,\theta), \frac{\partial}{\partial\theta'}v(.,\theta)\Big)\,.$$

(ii) B and B_T commute with the differential operator, that is $\partial[Bu(.,\theta)]/\partial\theta' = B[\partial u(.,\theta)/\partial\theta']$ for any function $u(.,\theta) \in \mathcal{D}(B)$.

Assumption 7. *(Convergence in norm and Functional convergence)*

(i) Define $\|B\| = \sup_{\|f\|\leq 1} \|Bf\|$. We have: $\|B_T - B\| \to 0$ in probability.

(ii) $\sqrt{T}z_T(.,\theta_0) \xrightarrow{d} Z \sim \mathcal{N}(0,K)$ on $\mathbb{H}$ as T goes to infinity. Z is the Gaussian random element of H that has mean zero and covariance operator K. In addition, $Z \in \mathcal{D}(B)$ with probability 1.

Assumption 7 is key to ensure that $B_T(\sqrt{T}z_T(.,\theta_0))$ is well-behaved asymptotically. Sufficient conditions to ensure the convergence of the $\mathbb{H}$-valued random element $z_T(.,\theta_0)$ stated in Assumption 7(ii) include e.g. mixing conditions: see Chen and White (1998). Asymptotic properties of estimated impulse response functions of weakly dependent processes are discussed, e.g., in Lütkepohl (1990).

We are now ready to state our main result.

Theorem 2. *(Asymptotic distribution of SMAS)*
Under Assumptions 1 to 7,

$$\sqrt{T}(\hat{\theta}_{SMAS} - \theta_0) \xrightarrow{d} \mathcal{N}(0,V)\,,$$

$$\textit{where} \qquad V = \|BG(\theta_0)\|^{-2}\Big(BG(\theta_0), (BKB')BG(\theta_0)\Big)\|BG(\theta_0)\|^{-2}\,,$$

with B' the adjoint operator of B.

The asymptotic covariance V displays the typical "sandwich form", which should yield the optimal choice of the operator B as the one such that BKB' equals the identity operator with associated $V = \|BG(\theta_0)\|^{-2}$. Unfortunately, one cannot directly choose B as $K^{-1/2}$ does not satisfy Assumption 3 and $B_T(\sqrt{T}z_T(.,\theta_0))$ would not be well-defined asymptotically in the Hilbert space[7]. The choice of the (optimal) operator is postponed until the end of the next section, after we explain how to estimate the covariance operator K.

[7]This issue is not specific to our framework; see e.g. Remark 4 in Carrasco and Florens (2000).

1.4 Practical implementation of the SMAS estimator

In this section, we explain how to compute the (optimal) SMAS estimator in practice. We first explain how to estimate the kernel operator K by bootstrap. Then, we explain how to compute its inverse which is ill-behaved and needs to be regularized in practice. Finally, we propose a data-driven procedure to select the regularization parameter.

1.4.1 Estimation of the covariance operator K

The estimation of the kernel operator K amounts to estimating the covariance between $z(h, \theta_0)$ and $z(s, \theta_0)$ for all possible pairs (h, s). In practice, we may consider relying on the sample counterparts $\hat{z}_T(h, \theta_0)$ and $\hat{z}_T(s, \theta_0)$. However, it is important to realize that in our general framework (e.g. when matching impulse responses), $\hat{z}_T(h, \theta_0)$ actually depends on the entire sample of observations $\mathcal{X}_T$ - as well as the entire simulated sample $\mathcal{X}_T^s$. As a result, direct estimation of the kernel operator is not feasible, and we are going to bootstrap it instead.

We consider two approaches to generate bootstrap samples. First, we assume that a bootstrap sample $\mathcal{X}_T^*$ is obtained by resampling with replacement from $\mathcal{X}_T$. As a result, the sample counterpart of the operator does not require a preliminary (first-step) estimator of the vector of parameters θ. A possible drawback of such a simple approach is associated with the fact that the auxiliary model is not assumed to be the true DGP, but only an approximation. This may put in jeopardy the validity of the above-mentioned resampling scheme[8]: e.g. if the errors inherit some of the serial dependence of the underlying DGP. As an alternative, we propose a *double-bootstrap* approach in which the simulator is used to generate bootstrap paths (for a given parameter value, say $\bar{\theta}$): the second layer conditions on the first layer to "recenter" the criterion function at the appropriate "true" value. This approach was recently formalized in a much more general framework (e.g. without assuming identification of the true unknown parameter value) by Antoine et al. (2022). Compared to the resampling scheme, it is computationally more demanding as it is *conditional* on the parameter value. Both procedures will be considered in our Monte-Carlo study.

To formalize the bootstrap, we follow Dovonon and Gonçalves (2017) (see their section B.1). Given our underlying (product) probability space $(\Omega, \mathcal{F}, P_0) \times (\Omega^s, \mathcal{F}^s, P^s)$ introduced in section 1.3 and our observed sample of size T $\mathcal{X}_T$, we assume that a bootstrap sample $\mathcal{X}_T^*$ is obtained by resampling[9] from $\mathcal{X}_T$: $\mathcal{X}_T^*$ is then viewed as a realization of a stochastic process defined on another probability space $(\Omega^*, \mathcal{F}^*, P^*)$. $\mathcal{X}_T^*$ actually depends on two sources

[8] We thank a referee for bringing this concern to our attention.

[9] As a result, the sample counterpart of the operator does not require a preliminary (first-step) estimator of the vector of parameters θ. Alternatively, we could also rely on the simulator to generate bootstrap paths, which would require knowing θ. Both procedures will be considered in our Monte-Carlo study.

of randomness, one related to the observed data and the other related to the resampling mechanism. When the joint randomness is of interest, the bootstrap statistic can be viewed as being defined on the product probability space $\left[(\Omega, \mathcal{F}, P_0) \times (\Omega^s, \mathcal{F}^s, P^s)\right] \times (\Omega^*, \mathcal{F}^*, P^*)$. Recall also that $\mathbb{P} \equiv P_0 \times P^s$. Given any bootstrap statistic X_T^*, we follow Dovonon and Gonçalves (2017) and define[10]:

- $X_T^* \xrightarrow{P^*} 0$ in prob-$\mathbb{P}$ if for any $\epsilon, \delta > 0$, $\mathbb{P}\Big(P^*(|X_T^*| > \epsilon) > \delta\Big) \to 0$ as $T \to \infty$.
- $X_T^* = \mathcal{O}_{P^*}(1)$ in prob-$\mathbb{P}$ if for any $\delta > 0$, there exists $0 < M < \infty$ such that $\mathbb{P}\Big(P^*(|X_T^*| \geq M) > \delta\Big) \to 0$ as $T \to \infty$.
- $X_T^* \xrightarrow{d^*} X^*$ in prob-$\mathbb{P}$ if $E^* g(X_T^*) \to E_{\mathbb{P}} g(X^*)$ in prob-$\mathbb{P}$ for every continuous and bounded function g, where $E^*(\cdot)$ is the expectation operator with respect to the bootstrap probability measure conditional on the data.

Assumption 8. *The sample distance function z_T and its bootstrap counterpart z_T^* are such that $\sqrt{T}\Big(z_T(., \theta_0) - z_T^*(., \theta_0)\Big) = o_{P^*}(1)$ in prob-$\mathbb{P}$.*

Assumption 8 is essential to ensure that the population covariance operator K can be estimated from the bootstrap sample. More specifically, we will estimate the covariance operator associated with the following kernel k^* which corresponds to the long-run covariance of $z^*(., \theta_0)$ defined as the limit of $z_T^*(., \theta_0)$.

Assumption 9. *We assume that there exists $z^*(., \theta_0)$ such that*

$$z_T^*(., \theta^0) \xrightarrow{d^*} z^*(., \theta^0) \quad \text{in prob-}\mathbb{P}.$$

Then, the bootstrap covariance kernel $k^*(h, s)$ is the L^2-kernel defined as

$$\begin{aligned} k^*(h,s) \quad \equiv \quad & \lim_{\tau \to \infty} \sum_{-\tau}^{\tau} E_{\mathbb{P}} \left[(z^*(h, X_t^*, \theta_0) - E_{\mathbb{P}} z^*(h, X_t^*, \theta_0)) \right. \\ & \left. \times (z^*(s, X_{t+\tau}^*, \theta_0) - E_{\mathbb{P}} z^*(s, X_{t+\tau}^*, \theta_0)) \right], \end{aligned}$$

where, with a slight abuse of notations, $X_{t'}^*$ corresponds to the bootstrap path starting at observation t'.

The maintained assumption that $k^*(.)$ is an L^2-kernel requires that,

$$\sum_h \sum_s k^*(h, s)^2 < \infty \,.$$

Since $k^*(.)$ corresponds to the long-run covariance of $z^*(., \theta_0)$, sufficient conditions are well-known, including e.g. boundedness and mixing conditions. Related conditions are

[10] See also Guerron-Quintana et al. (2017)

maintained on $z_T(.,\theta_0)$ in Assumption 7; see also Assumption 8.
Assumption 9 implies that the underlying covariance operator K is an Hilbert-Schmidt operator. In our proofs, we rely on the fact that: (i) the associated (Hilbert-Schmidt) norm is finite; and (ii) K can be approached by a sequence of bounded operators.

The sample counterpart of k^*, denoted k_T^*, is obtained using $N(T)$ (independent) bootstrap paths of length T denoted $\mathcal{X}_T^n$ (with $n = 1, \cdots, N(T)$), and associated distance functions denoted $z_T^{*(n)}(.,\theta_0)$. Thus, we have:

$$k_T^*(h,s) \equiv \sum_{m=-T+1}^{T-1} \omega\left(\frac{m}{M_T}\right) \hat{\Gamma}_T(|m|),$$

where $\omega(.)$ is a kernel, M_T is a bandwidth that diverges with T, and, for $m \geq 0$,

$$\begin{aligned}\hat{\Gamma}_T(m) &= \frac{1}{N(T)} \sum_{n=1}^{N(T)} \left[z_T^{*(n)}(h, X_1^{(n)}(T-m), \theta_0) - \overline{z}_T^*(h,\theta_0) \right] \\ &\quad \times \left[z_T^{*(n)}(s, X_{1+m}^{(n)}(T-m), \theta_0) - \overline{z}_T^*(s,\theta_0) \right]. \\ \text{with} \qquad \overline{z}_T^*(h,\theta_0) &\equiv \frac{1}{N(T)} \sum_{n=1}^{N(T)} z_T^{*(n)}(h, X_1^{(n)}(T), \theta_0),\end{aligned}$$

and $X_t^{(n)}(L)$ the n-th bootstrap path of length L starting at observation t. The chosen kernel and bandwidth will affect the asymptotic properties of the sample operator K_T associated with k_T^*. To simplify our exposition, we consider throughout the Bartlett kernel,

$$\omega(x) = \begin{cases} 1 - |x| & \text{if } x \leq 1 \\ 0 & \text{otherwise} \end{cases},$$

and the bandwidth is set according to the rule-of-thumb commonly used in practice, that is $M_T = \lfloor 0.75 \times T^{1/3} \rfloor$. For a thorough treatment of the links between the chosen kernel and bandwidth and the asymptotic properties of the associated operator, see e.g. Carrasco et al. (2007). Notice that in an i.i.d. framework where the autocorrelations can be ignored, the bootstrap covariance kernel's expression simplifies to,

$$E_{\mathbb{P}}\left[(z^*(h, X_t^*, \theta_0) - E_{\mathbb{P}} z^*(h, X_t^*, \theta_0))(z^*(s, X_t^*, \theta_0) - E_{\mathbb{P}} z^*(s, X_t^*, \theta_0)) \right],$$

and its sample counterpart does not involve any kernel,

$$\frac{1}{N(T)} \sum_{n=1}^{N(T)} \left[z_T^{*(n)}(h, X_1^{(n)}(T), \theta_0) - \overline{z}_T^*(h,\theta_0) \right] \times \left[z_T^{*(n)}(s, X_1^{(n)}(T), \theta_0) - \overline{z}_T^*(s,\theta_0) \right].$$

The sample operator K_T associated with k_T^* has the following representation, for any function $g \in \mathcal{D}(K_T)$,

$$(K_T g)(h) \quad = \quad \sum_{s=1}^{\infty} k_T^*(h,s) g(s) \, .$$

The operator K_T has a degenerate kernel and therefore, contrary to K, has a finite-dimensional closed range. As a result, the number of its eigenvalues and eigenfunctions is finite (equal to $N(T)$): as we explain next, these can be computed by solving a linear system of $N(T)$ equations. This extends section 3 in Carrasco and Florens (2000) to a time series framework[11].

Lemma 3. *(Computation of the eigenvalues and eigenvectors of K_T)*
Let $\phi_j^{(T)}$ denote the j-th eigenfunction and $\lambda_j^{(T)}$ the associated eigenvalue of K_T with $j = 1, \cdots, N(T)$. These eigenvalues and eigenfunctions are obtained as follows:

1. *Find the eigenvalues $\mu_j^{(T)}$ and the associated eigenvectors $\mathcal{B}_j = [\beta_j^1 \ \beta_j^2 \ ... \ \beta_j^{N(T)}]'$ $(j = 1, \ldots, N(T))$ of the $(N(T), N(T))$-matrix C with (n, n')-element*

$$\begin{aligned} c_{n,n'} \quad = \quad & \frac{1}{N(T)} \sum_{s=1}^{\infty} \sum_{m=-T+1}^{T-1} \omega\left(\frac{m}{M_T}\right) \left(z_T^{*(n)}(s, X_{1+m}^{(n)}(T-m), \theta_0) - \bar{z}_T^*(s, \theta_0) \right) \\ & \times \left(z_T^{*(n')}(s, X_1^{(n')}(T-m), \theta_0) - \bar{z}_T^*(s, \theta_0) \right) . \end{aligned}$$

2. *The eigenvalues of K_T are such that: $\lambda_j^{(T)} = \mu_j^{(T)}$ for $j = 1, \cdots, N(T)$.*

3. *The eigenfunctions of K_T are such that, for $j = 1, \ldots, N(T)$,*

$$\phi_j^{(T)}(h) = \frac{1}{N(T)} \left(z_h^{(T)} \right)' \mathcal{B}_j$$

$$\text{with} \qquad z_h^{(T)} \equiv \begin{bmatrix} \sum_{m=-T+1}^{T-1} \omega\left(\frac{m}{M_T}\right) \left(z_T^{*(1)}(h, X_1^{(1)}(T-m), \theta_0) - \bar{z}_T^*(h, \theta_0) \right) \\ \sum_{m=-T+1}^{T-1} \omega\left(\frac{m}{M_T}\right) \left(z_T^{*(2)}(h, X_1^{(2)}(T-m), \theta_0) - \bar{z}_T^*(h, \theta_0) \right) \\ \vdots \\ \sum_{m=-T+1}^{T-1} \omega\left(\frac{m}{M_T}\right) \left(z_T^{*(N(T))}(h, X_1^{(N(T))}(T-m), \theta_0) - \bar{z}_T^*(h, \theta_0) \right) \end{bmatrix}$$

From now on, we assume that the eigenvalues $\lambda_j^{(T)}$ are ranked in descending order, and that the eigenfunctions $\phi_j^{(T)}(h)$ have been orthonormalized. Our next result, Theorem 4,

[11] An alternative approach that does not involve the computation of eigenvalues and eigenfunctions is developed in section 3.3 in Carrasco et al. (2007): these authors argue that it may have some computational advantages, particularly in large samples. The sample sizes we consider are small to moderately large in accordance with our macro applications.

guarantees that the sample covariance operator and the associated eigenvalues are well-behaved asymptotically.

Assumption 10. $E_{\mathbb{P}}\|z^*\|^4 < \infty$.

Theorem 4. *(Asymptotic behavior of the covariance operator)*
Under Assumptions 1 to 10, when $T/N(T) \to \zeta$ as $T \to \infty$ with $0 < \zeta < \infty$, we have:

$$\|K_T - K\| \quad = \quad \mathcal{O}_{P^*}\left(\frac{1}{T^\nu}\right) \quad \textit{in prob-}\mathbb{P}\,,$$

with $\nu = 1/3$ in the general case with serial dependence, and $\nu = 1/2$ in the i.i.d. case.

1.4.2 Estimation of the optimal operator $K^{-\frac{1}{2}}$

To get the optimal SMAS estimator, the control operator B should be set equal to $K^{-\frac{1}{2}}$. Finding the inverse of the covariance operator K amounts to solving a Fredholm equation of the first kind in Φ, $K\Phi = g$, for some known $g \in L^2$, which is, in general, an ill-posed problem: that is, the solution Φ is unstable for small variations of g; see e.g. Wahba (1973), Groetsch (1993), Carrasco and Florens (2000), Carrasco et al. (2007), and Amengual et al. (2020). In order to stabilize the above solution (and the associated inverse of the covariance operator), we rely on Tikhonov regularization[12] and replace K^{-1} by $K_a^{-1} \equiv (K^2 + aI)^{-1}K$ for some positive a that converges to 0.

The optimal SMAS estimator is obtained as:

$$\hat{\theta}^{opt}_{SMAS} \equiv \arg\min_\theta \left\| K_{T,a}^{-1/2} z_T(\theta) \right\| = \arg\min_\theta \sum_{j=1}^{\infty} \frac{\lambda_j^{(T)}}{(\lambda_j^{(T)})^2 + a} |(z_T(\theta), \phi_j^{(T)})|^2 \,. \tag{1.2}$$

In order to derive its asymptotic properties, Assumptions 3 and 5 need to be updated: this is because the regularity properties maintained on the operator B (e.g. boundedness) are not satisfied by $K^{-1/2}$. As in Nashed and Wahba (1974), we let $\mathcal{H}(K)$ denote the domain of the operator $K^{-1/2}$, that is, the reproducing kernel Hilbert space of K.

Assumption 11. *(i) $z(.,\theta) \in \mathcal{H}(K) + \mathcal{H}(K)^\perp \; \forall \theta \in \Theta$.*

(ii) The function $\theta \to z(h,\theta)$ is differentiable with respect to θ with $G_j(.,\theta) \equiv \partial z(.,\theta)/\partial\theta_j \in \mathcal{D}(K^{-1/2})$ for $j = 1, \cdots, p$.

(iii) The (p,p)-matrix $\left(K^{-1/2}G(.,\theta), K^{-1/2}G(.,\theta)\right) = \left\|K^{-1/2}G(.,\theta)\right\|^2$ with element (i,j) defined as $(K^{-1/2}G_i(.,\theta), K^{-1/2}G_j(.,\theta))$ (for $i,j = 1, \cdots, p$), is positive definite and symmetric.

[12] Other regularization schemes have been used in practice: see e.g. Carrasco (2012).

We also need to introduce an additional regularity assumption.

Assumption 12. *(i)* $\left\|z_T(.,\theta) - z(.,\theta)\right\| = \mathcal{O}_{\mathbb{P}}(\frac{1}{\sqrt{T}})$ *uniformly in* $\theta \in \Theta$*;*

(ii) $\left\|z_T^*(.,\theta) - z(.,\theta)\right\| = \mathcal{O}_{P^*}(\frac{1}{\sqrt{T}})$ *in prob-*$\mathbb{P}$ *uniformly in* $\theta \in \Theta$*;*

(iii) $\left\|\frac{\partial z_T(.,\theta)}{\partial \theta} - \frac{\partial z(.,\theta)}{\partial \theta}\right\| = \mathcal{O}_{\mathbb{P}}(\frac{1}{\sqrt{T}})$ *uniformly in* $\theta \in \Theta$*;*

(iv) $\left\|\frac{\partial z_T^*(.,\theta)}{\partial \theta} - \frac{\partial z(.,\theta)}{\partial \theta}\right\| = \mathcal{O}_{P^*}(\frac{1}{\sqrt{T}})$ *in prob-*$\mathbb{P}$ *uniformly in* $\theta \in \Theta$*.*

The next theorem presents the asymptotic properties of the optimal SMAS estimator.

Theorem 5. *(Asymptotic behavior of the optimal SMAS estimator)*
Under Assumptions 1, 2, 4, and 6 to 12, the optimal SMAS estimator defined in (1.2) is $\sqrt{T}$*-consistent and asymptotically normally distributed with mean zero and variance* $\left\|K^{-1/2}G(\theta_0)]\right\|^{-2}$*, as* $T \to \infty$*,* $T^\nu a^{3/2} \to \infty$*,* $a \to 0$*, and* ν *as in Theorem 4.*

In practice, the regularization parameter a is selected by cross-validation, and we explain how to implement such a cross-validation procedure next.

1.4.3 Selection of the regularization parameter

In this subsection, we introduce our data-driven procedure to select the regularization a_T. In practice, we explain how to choose the parameter c by cross-validation where $a_T \equiv c/T^\ell$ with given $0 < \ell < 2\nu/3$. Notice that the rate of decay of a_T to 0 is motivated by our theoretical (asymptotic) results for the optimal SMAS in Theorem 5.

We first sketch how cross-validation works in practice. We start by splitting the sample of T observations into two subsamples: the training subsample - labelled "tr" - which corresponds to the first $\tilde{T}$ observations, and the testing subsample - labelled "test" - which corresponds to the remaining observations. For each candidate value for the parameter c, say c_j, we compute the corresponding optimal SMAS estimator over the training sample, say $\hat{\theta}^{opt}_{SMAS}(c_j)$: e.g. using the regularized optimal operator with regularization parameter $a_j = c_j/\tilde{T}^\nu$. This SMAS estimator is then used to simulate a pseudo-testing-sample and to compute the corresponding (optimal) SMAS objective function over the testing sample. The regularization parameter c_j is then chosen as the one that minimizes the objective function over the testing sample.

Let $\mathcal{C}$ denote the grid of candidate values for the parameter c.

Algorithm 2. *(Cross-validation to select the regularization parameter)*

1. *Split the sample of* T *observations into the training subsample "tr" which collects the first* $\tilde{T} \equiv \lfloor 2T/3 \rfloor$ *observations, and the testing subsample "test" with the remaining* $(T - \tilde{T})$ *observations.*
2. *Given* $c \in \mathcal{C}$*:*

(a) *Using the training sample, follow Algorithm 1 to compute the regularized optimal SMAS estimator (as in (1.2)) obtained with the regularized optimal operator* $\hat{K}^{-1/2}_{\tilde{T},a_{\tilde{T}}}$ *with* $a_{\tilde{T}} = c/\tilde{T}^{\nu}$*, that is:*

$$\hat{\theta}^{opt}_{SMAS}(c) = \arg\min_{\theta\in\Theta} \|K^{-1/2}_{\tilde{T},a_{\tilde{T}}} z_{tr}(\theta)\| .$$

(b) *Use the simulator with* $\hat{\theta}^{opt}_{SMAS}(c)$ *to generate* S *independent sample of* $(T - \tilde{T})$ *observations; compute the associated auxiliary statistics and match them to the auxiliary statistics computed over the testing sample to get* $z_{test}(\hat{\theta}^{opt}_{SMAS}(c))$*. Evaluate the associated SMAS objective function.*

3. *The regularization parameter (for the whole sample of size* T*) is* $a^*_T = c^*/T^{\nu}$ *where* c^* *is obtained by minimizing the SMAS objective function over the testing sample with respect to* c*.*

1.5 Conclusion

In the first chapter, we propose a new simulation-based estimator which handles a large number of auxiliary statistics simultaneously. It extends the SAS or I-I estimation as we focus on "matching" auxiliary statistics chosen to summarize key features of the data generating process of interest. Unlike classical minimum distance estimators, we do not rely on the characterization of a likelihood function that leads to an analytical formula of the estimator. We demonstrate the consistency and asymptotic normality of the SMAS estimator. Its asymptotic distribution displays the typical "sandwich form", which implies an optimal choice of the operator that governs the matched statistics. It requires the estimation of the covariance operator, which is implemented through a bootstrap-based procedure. Due to its unstable nature with a large number of underlying auxiliary statistics, Tikhonov regularization is introduced to stabilize its inverse. We also present a cross-validation procedure for selecting the associated tuning parameter.

Chapter 2

Simulation-based Estimation with Many Auxiliary Statistics: Applications to Long-run Dynamic Analysis

2.1 Introduction

We illustrate the implementation and performance of our proposed estimator through Monte-Carlo simulation experiments based on two well-known small- and medium-scale New Keynesian models. The New Keynesian approach is built on the DSGE framework originally developed for Real Business Cycle models, wherein agents optimize their utility functions while nominal wages and/or prices exhibit rigidity. These models are composed of equations that describe the micro-founded movements in macroeconomics; see Schorfheide (2008) for a survey on DSGE model-based estimation of the New Keynesian model. Compared to full information maximum likelihood estimators of DSGE models, impulse response matching estimation is preferred by some researchers as it allows for a focus on the aspects of the model that are most important for macroeconomics (Dridi et al. (2007)).

Our large number of auxiliary statistics corresponds to the dynamic responses of key macro indices (such as inflation and interest rate) at various horizons (including large ones) after a monetary shock: these impulse responses are not known in closed form and are rather obtained by simulation. Our examples show how to incorporate long-run information that can be used to improve the precision of estimates of structural parameters of interest. Our paper then contributes to the literature on impulse response matching estimation as done in Christiano et al. (2005), Inoue and Kilian (2013), or Guerron-Quintana et al. (2017). More specifically, Guerron-Quintana et al. (2017) consider VAR-based impulse response matching estimation of the parameters of DSGE models

when the number of impulse responses exceeds the number of VAR model parameters, but remains fixed. We extend Guerron-Quintana et al. (2017) to allow the number of components of the chosen auxiliary statistic to be infinitely large. In our Monte-Carlo experiments, we illustrate how the performance of their procedure deteriorates when a large number of "moments" is matched. And, more importantly, how long-run information can easily be incorporated to estimate the structural parameters with our proposed SMAS in the previous chapter. Sokullu (2020) proposes an alternative approach to constructing an I-I estimator using IRF matching, also employing Tikhonov regularization for the optimal weighting matrix. Their minimum distance estimator minimizes the difference between the empirical IRFs estimated from a VAR model and the theoretical IRFs implied by the structural model. In contrast, our SMAS estimator does not require an analytical expression for the IRFs.

The rest of this chapter unfolds as follows. In Section 2.2, we illustrate the small sample properties of the SMAS estimator by revisiting the small-scale New Keynesian model in Guerron-Quintana et al. (2017) with two macroeconomic indices. We compare the performance of six estimators: regularized SMAS with the diagonal operator, SMAS with the diagonal operator, regularized SMAS with the optimal operator, SMAS with the optimal operator, and two more estimators originally proposed in Guerron-Quintana et al. (2017), under a variety of simulation conditions. We also demonstrate the robustness of the covariance operator by an alternative constructed conditional on the values of the parameters of interest, instead of the one constructed by bootstrapping, using the SMAS estimator with the diagonal operator. In Section 2.3, we illustrate the performance of the SMAS estimator with a medium-scale New Keynesian model proposed by Smets and Wouters (2007), by matching impulse responses of all seven indices and of a subset of four indices, when estimating the degree of price indexation, the Calvo parameter, and both of them together. In Section 2.2.3, we consider the baseline stylized DSGE model from Fernández-Villaverde et al. (2016) to quantify the loss in estimation efficacy due to the infeasibility of the analytical IRF.

2.2 Small-scale New Keynesian model with 2 indices

2.2.1 Model and estimator

In our first set of experiments, we revisit the small-scale New Keynesian model of Guerron-Quintana et al. (2017) and focus on the estimation of price stickiness. In the model, π, R,

and x denote respectively, the inflation rate, interest rate, and output gap:

$$\begin{aligned}
\pi_t &= \kappa x_t + \beta \mathbb{E}_t(\pi_{t+1}) && \text{(Phillips Curve)} \\
R_t &= \rho_r R_{t-1} + (1-\rho_r)\phi_\pi \pi_t + (1-\rho_r)\phi_x x_t + \xi_t && \text{(Taylor rule)} \\
x_t &= \mathbb{E}_t(x_{t+1}) - \sigma[\mathbb{E}_t(R_t) - \mathbb{E}_t(\pi_{t+1}) - z_t] && \text{(Investment-Savings)} \\
z_t &= \rho_z z_{t-1} + \sigma^z \epsilon_t^z && \text{(real output shock)} \\
\xi_t &= \sigma^r \epsilon_t^r && \text{(monetary policy shock)}
\end{aligned}$$

where $\kappa = \dfrac{(1-\alpha)(1-\alpha\beta)}{\alpha} \dfrac{\omega+\sigma}{\sigma(\omega+\theta)}$, with α the probability to fix price, β the discounting factor, ω the disutility to work, and σ and θ the elasticities of substitution across time and across various commodities. ϕ_π and ϕ_x measure how responsive the central bank is to changes in inflation and the output gap. ϵ^z and ϵ^r in the shocks' dynamics are white noise processes.

Expectations are history based, and we generate our sample of observations on inflation and interest rate using `Dynare`[1]. Because inflation and output gap do not react to concurrent monetary shock, but react to the concurrent real shock, the structural shocks in the VAR model can be identified. The true values of structural parameters are provided as follows, $\alpha = 0.75, \beta = 0.99, \omega = 1, \sigma = 1, \theta = 6, \rho_r = 0.75, \rho_z = 0.90, \phi_\pi = 1.5, \phi_x = 0.125, \sigma^z = 0.30, \sigma^r = 0.20$. In this experiment, we focus on estimating α, and calibrate all the other parameters to their true value[2].

Given the sample of T observations on inflation and interest rate, a VAR(p) model is estimated; see section B.1 in the Appendix. In what follows, we refer to the VAR model as the *reduced-form model*, and to the DGP as the *structural model*. Based on the estimated VAR(p) model, associated transition matrix, residuals and impulse responses[3] of chosen horizons (e.g. from 1 to H) are obtained. Since in each period two shocks cast influence on two indices, a total of $4 \times H$ impulse responses, denoted $\hat{\psi}_T$, are obtained and matched with the corresponding impulse responses obtained from *simulated data* generated for a given α value, denoted $\hat{\psi}_T^s(\alpha)$. In addition, N bootstrap datasets of length T are generated using the reduced-form model with the estimated transition matrix and associated residuals, from which $z_{T,h}^{*(n)}$ and $\bar{z}_{T,h}^*$ can be computed, along with the eigenvalues $\lambda_j^{(T)}$ and eigenfunctions

[1]Dynare is a software platform for handling a wide class of economic models, in particular dynamic stochastic general equilibrium (DSGE). It is used to generate our sample of observations according to the structural model given the true model parameters.

[2]For results that allow parameters to be incorrectly calibrated, see the recent work of Antoine et al. (2022)

[3]We focus here on matching (so-called) structural impulse responses (see details in Appendix B.1). However, our procedure can easily accommodate other (dynamic) responses such as local projections (see e.g. Jordà (2005) and more recently Plagborg-Møller and Wolf (2021)), or nonlinear impulse responses (see Goncalves et al. (2021)).

$\phi_j^{(T)}$ of the covariance operator as explained in section 1.4.1. The (optimal) SMAS estimator of α is then

$$\hat{\alpha}_T = \arg\min_{\alpha} \sum_{j=1}^{\infty} \frac{\lambda_j^{(T)}}{(\lambda_j^{(T)})^2 + a_T} | < z_T(\alpha), \phi_j^{(T)} > |^2 \qquad \text{with } z_T(\alpha) = \hat{\psi}_T - \frac{1}{S}\sum_{s=1}^{S} \hat{\psi}_T^s(\alpha)\,,$$

and a_T chosen by cross-validation as explained in section 1.4.3. In practice, the optimization problem is solved by conducting a grid search over potential values of α. The probability of the price stickiness α is naturally bounded between 0 and 1, and the grid of candidate values is set at [0.005, 0.995] with a step of 0.005.

2.2.2 Monte-Carlo experiments

In our baseline experiment, a VAR(2) model is fitted to our sample of $T = 232$ observations (which corresponds to 58 years of quarterly observations) and impulse responses are matched up to horizon H - chosen between 20 and 80 (that is, between 5 and 20 years with quarterly data) - for a total of $4H$ matched impulse responses: in other words, we are matching between 80 and 320 impulse responses, therefore considering cases where it exceeds our sample size. For the estimation of the optimal operator and its inverse, we consider $N = 199$ bootstrapped samples and regularization parameter a_T selected by cross-validation (as explained in section 1.4.3). The sample impulse response values are obtained by averaging the impulse response vectors from $S = 10$ iterations.

Implementation details

We detail the implementation of optimal SMAS estimation of the aforementioned model with the resampling method[4], assuming that the auxiliary VAR model is deemed suitable. Consider a sample of $T = 232$ observations of interest rate R_T and inflation π_T, the fitted model VAR(2), and a matching horizon of $H = 80$.

- Estimate $\hat{\alpha}_{SMAS}^{opt}$ from the full sample of observations.

1. Fit a VAR(2) model with the observations and get the transition matrix A_f, the model residuals $\hat{\epsilon}_T$, and the covariance σ_f^2. Compute the structural impulse responses $\hat{\psi}_T$ with A_f and σ_f^2 as introduced in Appendix B.1.

2. Using $A_f, \hat{\epsilon}_T, \sigma_f^2$, and randomly selected initial values from true observations, get $N = 199$ bootstrapped samples of impulse responses to construct the optimal operator $\hat{K}_{T,a_T}^{-1/2}, a_T = c^*/T^{1/3}$ (c^* the selected tuning parameter).

   (a) Re-sample with replacement from the residuals $\hat{\epsilon}_T$ to get $\epsilon_{T,n}^*$.

[4]Details of the double-bootstrap method are presented in Appendix B.2

(b) Generate $\mathcal{X}^*_{T,n}$ and compute $\hat{\psi}^*_{T,n}$.

(c) Repeat independently N times to get $\hat{\psi}^*_{T,n}$ with $n = 1, \cdots, N$.

(d) Compute $\hat{K}_T^{-1/2*}$ using $\hat{\psi}^*_{T,n}$.

3. Iterate over the grid of α and simulate S samples of size $T = 232$, then generate impulse responses $\hat{\psi}^s_T(\mathcal{X}^s_T(\alpha))$. Obtain the optimal estimator as the minimizer over the search grid of α,

$$\hat{\alpha}^{opt}_{SMAS} = \arg\min_{\alpha} \left\| \hat{K}^{-1/2}_{T,a_T} z_T(\alpha) \right\|, \quad z_T(\alpha) = \hat{\psi}_T - \frac{1}{S}\sum_{s=1}^{S} \hat{\psi}^s_T(\mathcal{X}^s_T(\alpha))$$

• Use cross-validation to determine tuning parameter c^* from the search grid C following Algorithm 2.

1. Split the sample of size $T = 232$ into two parts, a training set of the first $\tilde{T} = 155$ observations and a testing set of the rest $T - \tilde{T} = 77$ observations.

2. Given $c \in C$, traverse the grid of α and compute the regularized optimal SMAS estimator obtained with the regularized optimal operator $\hat{K}^{-1/2}_{\tilde{T},a_{\tilde{T}}}, a_{\tilde{T}} = c/\tilde{T}^{1/3}$, using the training sample.

$$\hat{\alpha}^{opt}_{SMAS}(c) = \arg\min_{\alpha} \left\| K^{-1/2}_{\tilde{T},a_{\tilde{T}}} z_{tr}(\alpha) \right\|.$$

$z_{tr}(\alpha)$ denotes the difference between the chosen impulse responses (up to horizon $H = 80$) computed from the training set and that computed from simulated samples of size $\tilde{T} = 155$ generated with α. $K^{-1/2}_{\tilde{T},a_{\tilde{T}}}$ is constructed the same way as in the estimation step.

3. Use the simulator with $\hat{\alpha}^{opt}_{SMAS}(c)$ to generate S independent sample of $(T - \tilde{T}) = 77$ observations; compute the impulse responses (up to horizon $H = 80$) and match them to those computed over the testing sample to get $z_{test}(\hat{\alpha}^{opt}_{SMAS}(c))$. Evaluate the associated SMAS objective function.

4. The regularization parameter (for the whole sample of size T) is $a^*_T = c^*/T^{1/3}$ where c^* is obtained by minimizing the SMAS objective function over the testing sample with respect to c.

We compare the performance of the (efficient) SMAS estimator, which uses either the diagonal or the optimal operator, with and without regularization, to the two estimators developed in Guerron-Quintana et al. (2017), respectively with a diagonal weighting matrix and the optimal one. Estimation suggested in Guerron-Quintana et al. (2017) resembles the SMAS estimator in many ways. However, two major differences between the

GIK estimator and the SMAS estimator are captured: (i)GIK estimator minimizes the difference between the impulse response function of a VAR model based on observations and their *theoretical expression* based on the same VAR whereas SMAS estimator minimizes the difference between the impulse response function computed on the observations and their simulated counterpart without relying on their theoretical expressions; (ii) GIK estimator does not incorporate any form of regularization whereas SMAS estimator utilizes Tikhonov regularization to stabilize the behavior of the optimal operator.

Tables B.1, B.2 and B.3 summarize the performance of these six estimators in terms of Monte-Carlo average, standard deviation, Root Mean Squared Error (RMSE), and Mean Absolute Deviation (MAD), as well as effective coverage rates of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications. In addition, Figures B.2 and B.3 display the Monte-Carlo histograms of these six estimators. We match either: (i) a medium to large number of IR (from 5 to 20 years with horizons $H = 20$ to 80) in Table B.1, Table B.2, and Figure B.2; or (ii) a small number of IR (from 0.5 to 2 years with horizons $H = 2$ to 8) in Table B.3 and Figure B.3. We also compare a smaller sample size of $T = 100$ observations (or 25 years of quarterly observations) when matching a medium to large number of IR in Table B.4. Moreover, we present a robustness check with the operator obtained by double-bootstrap mentioned in Section 1.4.1 which depends on the parameter value, in comparison to that generated from resampling.

Experiment results

• Regularized SMAS vs non-regularized SMAS:
Overall, the regularized estimator behaves much better than the non-regularized one according to all reported measures of performance (including RMSE and MAD) when matching a medium to large number of IR. The Monte-Carlo distribution of the regularized SMAS is accordingly much better behaved than that of the non-regularized one: specifically, the distribution is closer to being symmetric around the true unknown parameter value, more concentrated around it, and closer to being bell-shaped. Without regularization, the empirical coverage rates often fall below the nominal level (95% and 90%).
In addition, when only matching a small number of IR, the two estimators are very close to each other, with the regularized one slightly better in terms of SD and RMSE, and often equally as good in terms of MAD and Bias.

• SMAS with diagonal operator vs SMAS with optimal operator:
Overall, the SMAS estimator with the optimal operator outperforms the one with the diagonal operator. Specifically, the SMAS estimator with the optimal operator presents a smaller RMSE even when the MAD is slightly larger, showing it is more concentrated

around the true unknown parameter value. It is true with short and long horizons matched. Effective coverage rates are often equally good for both estimators.

• SMAS vs benchmark estimation in Guerron-Quintana et al. (2017) (GIK estimators):

Since the impulse response functions in Guerron-Quintana et al. (2017) are generated by the fitted structural VAR model, theoretically, if the VAR model approximates well the DGP, then the GIK estimators would excel. This advantage is demonstrated by the comparisons between the SMAS estimator with the diagonal (optimal) operator but not regularized and the GIK estimator with the diagonal (optimal) operator where the GIK estimator usually dominates.

Overall, the regularized SMAS estimator behaves much better than both GIK estimators. For all the horizons we consider - even the shorter ones, it always outperforms them according to all measures of performance, and displays the smallest bias, MAD, SD, and RMSE throughout - even when considering smaller sample sizes. In addition, effective coverage rates are much closer to their nominal levels for the regularized SMAS estimator than for the GIK ones.

• Performance as a function of the horizon H and the order p of the fitted VAR:

Overall, the performance of the regularized SMAS estimator is remarkably stable as a function of H - including when H is large and/or when the number of matched conditions exceeds the sample size. Finally, the choice of the order p of the fitted VAR model does not seem to affect the performance of the regularized SMAS estimator much.

2.2.3 Robustness check

As a robustness check, we consider the *double-bootstrap* approach proposed in Section 1.4.1. Specifically, the covariance operator is computed using IRs generated from simulations conditional on the parameter value.

We implement estimation with a VAR(2) model fitted to our sample of $T = 232$, and a matching horizon of $H = 80$. As shown in Table B.6, the estimator with the diagonal operator or optimal operator conditional on the parameter value is nearly as effective as the general SMAS estimator, with only a slight difference in the chosen measures of performance, with regularization in particular. When employing the double-bootstrap approach with regularization, the estimator with the optimal operator performs slightly weaker than that with the diagonal operator in the chosen measure. The outcome may be attributed to the compounded complexity arising from both correlation and varying parameter values. The findings indicate that our proposed estimator is resilient in its ability to handle variations in the structural model specification.

2.3 Medium-scale New Keynesian model with 7 indices

In our second set of experiments, we consider a medium-scale New Keynesian model in the style of Smets and Wouters (2007), and focus on the estimation of the price stickiness (Calvo parameter) and the degree of price indexation. The model corresponds to a 30-equation dynamic linear system, which involves output, consumption, investment, wage, working hours, inflation, and interest rate; see e.g. section 1 in Smets and Wouters (2007) p588. Our sample of observations on the seven above-mentioned indices is generated using `Dynare`[5] with parameter values set as in Table B.7: in particular, the Calvo parameter ξ_p is set at 0.908 and the degree of price indexation γ_p is set at 0.469.

Our estimation procedure closely follows what was done in the previous section: first, a VAR(p) model is applied to our sample of T observations, and transition matrix, residuals and chosen impulse responses are obtained accordingly; second, these impulse responses are matched to corresponding ones obtained from simulated data generated for a given value of the unknown parameter; third, the (optimal) covariance operator - and corresponding eigenvalues and eigenfunctions - is obtained from N additional bootstrap samples. The regularization parameter is chosen once again by cross-validation. Both parameters of interest are naturally bounded between 0 and 1. The grid of candidate values for γ_p is set at [0.05, 0.95] with a step of 0.05. The search step of candidate values for ξ_p is set at 0.05 in the interval [0.1, 0.6], and at 0.005 in the interval [0.61, 1], respectively. See also the previous subsection for implementation details.

We focus here on the impact of a monetary policy shock, and consider dynamic responses associated with a unit shock in interest rate. In our baseline experiment, a VAR(4) model is fitted to our sample of $T = 236$ observations (59 years of quarterly observations), and impulse responses are matched up to horizon H - chosen between 20 and 80 (that is, between 5 and 20 years after the shock with quarterly data). We either consider the dynamic responses in all seven indices, or in a subset of 4 indices (consumption, labor force, inflation, and interest rate). For the estimation of the optimal operator and its inverse, we consider $N = 100$ bootstrapped samples.

(1) Estimation of the degree of price indexation alone.

Table B.8 summarizes the performance of the (efficient) SMAS estimator (with and without regularization) in terms of Monte-Carlo average, standard deviation, Root Mean Squared Error (RMSE), and Mean Absolute Deviation (MAD), as well as effective coverage rates of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

[5]We thank Nicola Viegi (`http://www.nviegi.net/teaching/master/monmas.htm`) and Johannes Pfeifer (`https://github.com/JohannesPfeifer/DSGE_mod/tree/master/Smets_Wouters_2007`) for sharing their `Dynare` code.

Overall, the performance of the regularized SMAS estimator dominates that of the non-regularized one, and we focus on the regularized SMAS estimator in our discussion of the results below.

• Matching IR up to medium horizons vs up to long horizons:
The performance of the regularized SMAS estimator improves when adding IR at long horizons. When comparing the first two columns in Table B.8, all measures of performance indicate that there is useful information contained in long-run IR. This holds whether one is matching IR on all seven indices (Panel A), or on a subset of 4 indices (Panel B) - though the improvements from column 1 to column 2 are more modest in Panel B. Also, such long-run information does not improve the coverage rates much, and inference remains conservative with effective coverage rates above their nominal levels.

• Matching IR up to long horizons vs from medium to long horizons:
The performance of the regularized SMAS estimator improves when adding IR at short horizons. When comparing the last two columns in Table B.8, all measures of performance indicate that there is useful information contained in short-run IR. This holds whether one is matching IR on all seven indices (Panel A), or on a subset of 4 indices (Panel B) - though the improvements from column 3 to column 2 are quite modest in Panel A. Also, such short-run information does not improve the coverage rates much, and inference remains conservative with effective coverage rates above their nominal levels.

• Robustness check:
As a robustness check, we report in Table B.9 results obtained when using alternate VAR models. Overall, the choice of the order of the fitted VAR model does not seem to affect much the performance of the regularized SMAS estimator.

(2) Estimation of the Calvo parameter alone.
By design, the estimation of the Calvo parameter is more challenging as its true unknown parameter value is much closer to the upper bound 1. Table B.10 summarizes the performance of the (efficient) SMAS estimator (with and without regularization) in terms of Monte-Carlo average, standard deviation, Root Mean Squared Error (RMSE), and Mean Absolute Deviation (MAD), as well as effective coverage rates of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

Overall, the performance of the regularized SMAS estimator is excellent and dominates that of the non-regularized one which suffers from very large size distortions, even when large sample sizes are considered. Similarly to what was reported for the estimation of the degree of price indexation, there is useful information contained in long-run IR, and both SD and RMSE of regularized SMAS do improve when matching up to horizon 80 - even with a relatively small sample size.

(3) Joint estimation of the degree of price indexation and the Calvo parameter.
When it comes to the joint estimation of the degree of price indexation and the Calvo

parameter, Table B.11 summarizes the performance of the (efficient) SMAS estimator (with and without regularization) in terms of Monte-Carlo average, standard deviation, Root Mean Squared Error (RMSE), and Mean Absolute Deviation (MAD), as well as effective coverage rates of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

Overall, the performance of the regularized SMAS estimator is excellent and dominates that of the non-regularized one which still suffers from large size distortions. Similarly to what was reported when estimating each parameter individually, there is useful information contained in long-run IR, and both SD and RMSE of regularized SMAS do improve when matching up to horizon 80.

2.4 Baseline stylized DSGE model

In our last set of experiments, we consider the baseline stylized DSGE model from Fernandez-Villaverde et al. (2016) as adapted from DelNegro and Schorfheide (2008) where the IRs are known analytically. This allows us to compare the performance of two versions of our regularized SMAS estimator: the feasible one - as previously described - as well as the *infeasible* one which relies on the analytical IRs instead of computing them by simulation[6].

The stylized DSGE model consists of several sectors including households, intermediate and final goods producers, and a monetary authority. A Calvo assumption introduces nominal rigidity in prices, and firms that cannot reoptimize their prices at a given time adjust these by the steady-state inflation rate. This baseline model is designed to have a state-space representation which is used to obtain associated IRs analytically. Details can be found in Appendix B.3.3.

Once again, we focus on delivering inference on only one parameter of the model, namely the Calvo parameter, while the remaining structural parameters are calibrated to values suggested in the literature. In our experiment, a VAR(4) model is fitted to our sample of $T = 200$ (or 400) observations and impulse responses to a monetary policy shock are matched up to horizon $H = 20$ or 80. The results are reported in Table B.13.

Overall, the infeasible SMAS estimator performs better than the SMAS, both in terms of bias and standard deviation. However, the differences remain small. It is interesting to note that for the smaller sample sizes, differences between the two versions of SMAS are mainly driven by biases, whereas for larger sample sizes they are mainly driven by standard

[6]The second SMAS is labelled "infeasible" because, in general, impulse responses are not known analytically. In our previous two experiments, IRs are not known analytically and the infeasible SMAS is not available.

deviations. There are little to no size distortions, and the coverage of associated confidence intervals remains very close to the nominal sizes for the feasible SMAS estimator.

2.5 Conclusion

The simulation experiments conducted on three models reveal that the incorporation of many auxiliary statistics and data-driven regularization significantly enhance the precision and coverage rate of the estimation. Our proposed SMAS estimator provides a straightforward solution for estimating complex models such as the New Keynesian equation system, without relying on the often unattainable likelihood function or analytical expressions of the objective function. We also demonstrate the robustness of our estimator to variations in the simulated operator approximation. Moreover, the efficiency loss due to the constraint of feasibility of the SMAS estimator is negligible and can be compensated with additional observations.

Chapter 3

Simulation-based Inference with Many Auxiliary Statistics: Hypothesis Testing with Many Moments

3.1 Introduction

The autoregressive model is a prevalent method for analyzing the temporal progression of economic phenomena. This concise yet robust tool enables researchers to simulate business cycles for significant economic indicators, model stock market volatility, evaluate monetary policy, exchange rates, and more. Hamilton (1989) adds a two-regime switching component, which follows a Markov process, to the linear autoregressive specification, enabling the economic trajectory to be influenced by either a positive growth state or a negative growth state. Kim and Nelson (1999) in their paper on the business cycles in the US advocate the Markov-switching model to characterize the structural break in the US real GDP growth towards more stabilization. Outside economics, the hidden Markov model (HMM) is a popular statistical model for many real-world applications like speech recognition, facial expression recognition, musical composition, and bio-informatics (see Mor et al. (2021)).

Typically, estimating statistical models with unobserved Markov processes requires creating and analyzing likelihood functions, which can be computationally intensive. Similarly, performing tests to compare models with one state to those with two or more states using likelihood-based methods can also pose computational challenges and potential issues. Davies (1977, 1987, 2002) address the issue that the likelihood ratio (LR) does not possess the standard χ^2 distribution when a hidden nuisance parameter exists under different scenarios. A second problem arises when the regularity conditions are violated, for example, with identically zero scores and the scores not having positive

variances. It also leads to LR test statistic with a non-standard distribution. Hansen (1992, 1996) and Garcia (1998) design LR tests specifically for addressing these types of violations. Though treating the nuisance parameters vector differently, both of their methods involve exploring the intervening nuisance parameter space and assessing the Markov-switching likelihood function at each grid point, which makes them computationally demanding.

Carrasco et al. (2014) reduces the computational cost by suggesting a parametric bootstrap approach that needs model estimation under the null hypothesis only to determine critical values. Their proof of optimality involves showing that, for fixed values of the nuisance parameters, the proposed test is asymptotically locally equivalent to the LR test. On the other hand, Dufour and Luger (2017) addresses the statistical challenge of establishing the LR function by presenting test statistics based on the moments of the least-squares residuals. Their pivotal compound statistics incorporate the mean, variance, skewness, and kurtosis of the residuals of the model.

We propose a novel approach that also employs the Monte-Carlo framework to overcome the challenges in deriving and estimating the LR functions. Our method adds to the regime-switch testing literature by allowing the incorporation of higher-order moments in the design of test statistics, without being limited by the availability of analytical expressions of moments. By introducing a norm that is controlled by a specialized operator in a Hilbert space of real-valued functions that are squared-integrable, we can combine infinitely many discrete empirical moments into a single squared norm. This method avoids the need to find the moment generating functions and the operator presented in Section 3.2.2 aids in the practical integration of a large number of moments.

In a preceding paper, Antoine and Sun (2022) introduced a simulation-based estimation method, which matches a large number of auxiliary statistics; see also Chapter 1. This method was inspired by the generalized method of moments procedure, extended to a continuum of moment conditions with regularization (Carrasco and Florens (2000)). Unlike that paper, our current work does not aim to estimate the underlying parameters, nor derive the asymptotic distribution of the associated estimator, rather it designs a specification test to decide whether a (simple) stable model is appropriate. Our motivation for considering a large number of moments is not related to optimality (e.g. to get "close" to the LR test), but rather to extract information from the higher-order moments of the data. In addition, it is worth mentioning that our procedure is easy to implement since it relies on simulation rather than finding a pivotal statistic that may not be available, or convenient to compute.

The performance of our proposed test is demonstrated through a series of Monte-Carlo experiments. First, we examine the mean-variance switching model, with weak and strong inter-temporal correlations, and demonstrate that our test gains power with higher-order moments matched, in comparison to examples provided by Dufour and Luger (2017).

Second, we test the changing slope coefficients and demonstrate that our test can detect significant differences against the generic AR model under the null. Finally, we revisit Hamilton's Markov-switching model of postwar US output growth and show that the linear AR(4) model fails to accurately capture the dynamics between 1952 and 2010. Overall, we conclude that a higher number of matched moments leads to the acquisition of more information, with evidence of increased empirical power.

The paper proceeds as follows. In Section 3.2, we introduce the framework of testing a Markov-switching model, leading with a simple example of the moment-matching test by bootstrap, and formulate the algorithm. In Section 3.3, we illustrate the performance of our proposed test through Monte-Carlo experiments with autoregressive models. In Section 3.4, we apply our testing procedure to revisit Hamilton's Markov-switching model of the US output growth, with seasonal observations of 32 years and 58 years, respectively. Section 3.5 concludes.

3.2 Test Design

Hypothesis testing of the discrete-state switching means and variances suggested in Dufour and Luger (2017) exploits four statistics based on mean, variance, skewness, and kurtosis of the samples. In the same spirit of moment matching, our method extends the number and choice of moments - allowing for many, possibly higher-order, moments - without requiring their analytical expressions. In this section, we exemplify the simulation-based hypothesis test design using an AR(1) model with a two-state Markov-switching component.

3.2.1 Moment-matching test by bootstrap

We start by introducing moment-matching testing with an example of the test for the difference between two means. With observations of sample A, $X^A = \{x_1^A, x_2^A, \ldots, x_{L_1}^A\}$, and sample B, $X^B = \{x_1^B, x_2^B, \ldots, x_{L_2}^B\}$[1], consider testing whether the two (unknown) population means, respectively μ_A and μ_B, are equal:

$$H_0 : \mu_A - \mu_B = 0 \text{ vs. } H_1 : \mu_A - \mu_B \neq 0 \tag{3.1}$$

Under standard regularity conditions, according to the central limit theorem, we can calculate the sample mean difference and derive its asymptotic distribution, then use it as a test statistic for the above null. The Monte-Carlo testing design does not require knowledge of the (asymptotic) distribution of the test statistics. Instead, it is replaced with an approximation by bootstrap.

[1]Two integers L_1, L_2 are not assumed to be the same.

The hypothesis can be written as $\mathcal{H}_0 : (\mu_A - \mu_B)^2 = 0$ vs. $\mathcal{H}_1 : (\mu_A - \mu_B)^2 > 0$. The sample test statistic $z_{L_1,L_2} = (\bar{X}_A - \bar{X}_B)^2$ is the squared norm of the sample mean difference. The distribution of the test statistic under the null is established by bootstrap. A bootstrap sample is generated by random selection with replacement[2], of size L_1 for sample A and size L_2 for sample B, and the bootstrapped sample mean is denoted $\bar{X}^*_{n,A}$ or $\bar{X}^*_{n,B}$. With $N-1$ pairs of such samples, we compute the squared norm of the difference between the sample average $z^*_n = (\bar{X}^*_{n,A} - \bar{X}^*_{n,B})^2,\ n = 1, 2, \ldots, N-1$. The MC p-value based on the realizations of test statistics z^*_n's is then defined as

$$G_z[z_{L_1,L_2}; N] = \frac{N + 1 - R_z[z_{L_1,L_2}; N]}{N} \tag{3.2}$$

where $R_z[z_{L_1,L_2}; N]$ is the rank of z_{L_1,L_2} when $z_{L_1,L_2}, z^*_1, z^*_2, \ldots, z^*_{N-1}$ are placed in an ascending order. The number of iterations N is preset and can be as small as 20 to obtain a test with an exact level of 5%. A larger N reduces randomness within the simulations and, in general, leads to the power increase; however, the gain is limited for N greater than 100 (Dufour et al. (2004)). For a more detailed discussion on the critical region and empirical p-value of the Monte-Carlo test method, see Dufour and Khalaf (2001).

This illustrates the bootstrap-based test for the equality of the means of two populations. Dufour and Luger (2017) create the pivotal test statistic using sample moments with an analytical function, and build the distribution of the composite statistic under the null by the Monte-Carlo method. Under a similar framework, our method puts no limit on the number of moments, since we do not need their closed-form expression.

3.2.2 Framework and motivation

Consider a sequence of $X_T = \{x_1, x_2, \ldots, x_T\}$, generated from an autoregressive process containing a regime-switching component following a first-order Markov chain process. Let S_t denote the random variable of the state.

$$x_t = \mu_{s_t} + \rho(x_{t-1} - \mu_{s_{t-1}}) + \sigma_{s_t}\epsilon_t, \quad t = 2, \ldots, T \tag{3.3}$$

The latent random variable S_t takes the integer value 1 or 2 in each and every period with probability

$$\Pr(S_t = j) = \sum_{i=1}^{2} p_{ij} \Pr(S_{t-1} = i), p_{ij} = \Pr(S_t = j | S_{t-1} = i),\ j = 1, 2$$

[2]In our MC test, we go beyond this simple example and use a simulator function to generate the "bootstrap samples".

The transition matrix is

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$$

where $p_{i1} + p_{i2} = 1$, and $\epsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. S_t and ϵ_t are independent.

Let θ denote the set of all model parameters, both constant and varying included. As in model (3.3) above, $\theta_{s_t} = \{\rho, \mu_{s_t}, \sigma_{s_t}\}$. The proposed hypothesis is $\mathcal{H}_0$: all elements of θ_{s_t} are stable, $\theta_1 = \theta_2 = \theta_0$. Alternative expressions of the null hypothesis are $p_{11} = p_{21} = 1$ or $p_{12} = p_{22} = 1$. Though this method tests the closeness between the observations and the proposed distribution in the hypothesis by moment-matching, it requires no analytical formula of moments or likelihood function. Instead, we rely on simulations to build the test statistics and rely on the Monte-Carlo method to find the empirical p-values. In practice, our method is less demanding in terms of the maintained assumptions of the distribution functions, and the computational complexity only depends on the choice of simulator under the null hypothesis.

The data generating process (DGP) under the null hypothesis is denoted $g(\cdot, \theta)$. Let $\hat{\theta}_T^0$ be the restricted consistent estimator of θ and $g(\cdot, \hat{\theta}_T^0)$ which will be used as the simulator for later bootstrapping. The generated data is denoted $\widetilde{X}_T(\hat{\theta}_T^0)$. To execute the MC test, we only assume that the moments of data can be simulated under the null hypothesis and they take continuous values.

For the case of model (3.3), the data is tested against the hypothesis of an AR(1) process with no regime change. The observations of size T, $\{x_1, x_2, \ldots, x_T\}$, are fitted to an AR(1) model with maximum likelihood estimation[3], $x_t = \rho_0 + \rho_1 x_{t-1} + \epsilon_t$, $t = 2, \ldots, T$. Let $\hat{\rho}_0$ and $\hat{\rho}_1$ be the consistent estimates. The h-th order empirical moment $\xi_{t,h}$ can be computed for the true observations and bootstrapped data at time t. Type A moments are moments of demeaned observations, $\xi_{t,h}^A = (x_t - \bar{x}_T)^h$, where $\bar{x}_T$ is the sample average. Type B moments are moments of demeaned sample residuals, $\xi_{t,h}^B = (z_t - \bar{z}_T)^h$ where $z_t = x_t - \hat{\rho}_0 - \hat{\rho}_1 x_{t-1}$, and $\bar{z}_T$ being the average of residuals. For the selected h, the stacked vector of moments is denoted $\xi_h = [\xi_{1,h}\ \xi_{2,h}\ \ldots\ \xi_{T,h}]'$, and $\bar{\xi}_h = \frac{1}{T}\sum_{t=1}^{T} \xi_{t,h}$. All of the computed empirical moments are collected in the stacked vector of H moments $\underline{\xi}_T = [\bar{\xi}_1, \bar{\xi}_2, \ldots, \bar{\xi}_H]$.

In practice, we define a norm governed by special operator $\mathcal{B}$ in a Hilbert space of squared-integrable real-valued functions to harvest the information in the higher-order moments. We also define the associated inner product $(\cdot, \cdot)$ and the norm $\|.\|$ in the Hilbert space. See Carrasco and Florens (2000) and Antoine and Sun (2022) for more discussion about the operator, the inner product, and the norm.

Next, we introduce the test statistic with $\mathcal{B}$. Let $\underline{\widetilde{\xi}}_{null}$ be the values of H moments under the null hypothesis, which is computed from the simulated sample $\widetilde{X}_T(\hat{\theta}_T^0)$.

[3]According to the specific assumptions held, consistent $\hat{\theta}_T^0$ might be estimated in a different way.

$$q = \left\| \mathcal{B}_T(\underline{\xi}_T - \widetilde{\underline{\xi}}_{null}) \right\|^2 = \sum_{j=1}^{\infty} \frac{\lambda_j^{(T)}}{(\lambda_j^{(T)})^2 + a} |(\underline{\xi}_T - \widetilde{\underline{\xi}}_{null}, \phi_j^{(T)})|^2 \tag{3.4}$$

Let K be the covariance operator of the distance $\underline{\xi}_T - \widetilde{\underline{\xi}}_{null}$, and K_T be its sample counterpart with a degenerate kernel. $\phi_j^{(T)}$ is defined as the j-th eigenfunction of K_T and $\lambda_f^{(T)}$ the associated eigenvalue. a is a preset regularization parameter for covariance operator inversion.

We now formulate the MC test as stated in this section.

Algorithm 3. *(Monte-Carlo test)*

1. *Fit model $x_t = \hat{\rho}_0 + \hat{\rho}_1 x_{t-1} + e_t$ under the null hypothesis, as the simulator;*
2. *Compute sample moments of true observations $\underline{\xi}_T$;*
3. *Simulate data under the null hypothesis using the linear model $g(\cdot, \hat{\theta}_T^0)$ and error term as assumed, and compute their moments $\widetilde{\underline{\xi}}_{null}$* [4]*;*
4. *Construct the test statistic $q = \left\| \mathcal{B}_T(\underline{\xi}_T - \widetilde{\underline{\xi}}_{null}) \right\|^2$;*
5. *Build the bootstrapped distribution of the test statistic under the null hypothesis;*
   - *Simulate $N-1$ independent samples under the null hypothesis (the same way as described in step 3) and compute their moments $\widetilde{\underline{\xi}}_{null}^n$, $n = 1, 2, \ldots, N-1$;*
   - *Calculate $q^n = \left\| \mathcal{B}_T(\widetilde{\underline{\xi}}_{null}^n - \underline{\xi}_{null}) \right\|^2$ for each n;*
6. *Rank $q, q^1, q^2, \ldots, q^{N-1}$ in ascending order, then calculate the empirical p-value defined in (3.2), $G_q[q; N] = \dfrac{N + 1 - R_q[q; N]}{N}$.*

3.3 Simulation Studies

In this section, we illustrate the proposed Monte-Carlo hypothesis testing method with two autoregressive models. The first one is a mean-variance switch model from Dufour and Luger (2017), and the second one is a Hamilton-type regression model from Ghysels and Marcellino (2018). Here we provide the Monte-Carlo experiments results of the models. For both of them, we present the empirical size and empirical power of the tests, with two families of moments introduced in Section 3.2.2, the moments of demeaned observations (Type A) and moments of demeaned residuals (Type B).

In this section, all simulation-based tests use a total of $H = 4$, 12, or 20 moments. The empirical size is computed from rejections out of 5000 independent replications. For

[4] In practice, $\widetilde{\underline{\xi}}_{null}$ is imputed by the sample average from N_{null} independent samples

each replication, $N-1=499$ simulation samples are generated to construct the empirical distribution of the test statistic under the null hypothesis. We consider $N_{null}=20$ samples generated under the null and compute the value of moments whose sample mean is used as the moments under the null. For the estimation of the empirical operator $\mathcal{B}_T$ and its inverse, we consider $N_{boot}=20$ bootstrapped samples and arbitrary regularization parameter $a=0.0001$.

3.3.1 Mean-variance Markov-switching autoregression model

We revisit the model considered by Dufour and Luger (2017):

$$x_t=\mu_{s_t}+\sum_{k=1}^{r}\rho_k(x_{t-k}-\mu_{s_{t-k}})+\sigma_{s_t}\varepsilon_t \tag{3.5}$$

The state variable $\{S_t\}$ can take two values, 1 and 2. It follows the first-order Markov chain process with transition matrix

$$\mathbf{P}=\begin{bmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{bmatrix}$$

where $\sum_{j=1}^{2}p_{ij}=1$, for $i=1,2$.

Without loss of generality, consider the AR(1) case of model (3.5),

$$x_t=\mu_{s_t}+\rho(x_{t-1}-\mu_{s_{t-1}})+\sigma_{s_t}\epsilon_t \tag{3.6}$$

where $\mu_{s_t}=\mu_1\mathbb{I}[S_t=1]+\mu_2\mathbb{I}[S_t=2]$, $\sigma_{s_t}=\sigma_1\mathbb{I}[S_t=1]+\sigma_2\mathbb{I}[S_t=2]$, $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ are constants. ϵ_t is i.i.d. white noise.

Rewrite model (3.6) as

$$x_t=\sum_{i=1}^{2}\mu_i\mathbb{I}[S_t=i]+\rho(x_{t-1}-\sum_{i=1}^{2}\mu_i\mathbb{I}[S_{t-1}=i])+\sum_{i=1}^{2}\sigma_i\mathbb{I}[S_t=i]\varepsilon_t \tag{3.7}$$

Define $z_t(\rho)=x_t-\rho x_{t-1},\ t=2,\dots,T$, then

$$z_t(\rho)=\mu_1^*\mathbb{I}[S_t^*=1]+\mu_2^*\mathbb{I}[S_t^*=2]+\mu_3^*\mathbb{I}[S_t^*=3]+\mu_4^*\mathbb{I}[S_t^*=4]+(\sigma_1\mathbb{I}[S_t=1]+\sigma_2\mathbb{I}[S_t=2])\varepsilon_t$$

where

$$\mu_1^*=\mu_1(1-\rho),\ \mu_2^*=\mu_2-\rho\mu_1,\ \mu_3^*=\mu_1-\rho\mu_2,\ \mu_4^*=\mu_2(1-\rho)$$

S_t^* denotes the four possible cases depending on the constant values. Only when $\mu_1=\mu_2$ and $\sigma_1=\sigma_2$, z_t has constant mean and variance over time. Depending on the true value of ρ, when $\mu_1\neq\mu_2$, z_t reveals to be a mixture of two (when $\rho=0$), three (when $\rho=1$ or -1), or four normals (with other values of ρ). Nonetheless, under the null hypothesis $\mu_1=\mu_2$

and $\sigma_1 = \sigma_2$, the moments of z_t are non-varying, the moment-based test thus identifies the discrepancies in some or all of the moments when the switch is present.

In the simulation study, observations are tested against a linear AR(1) specification. Model parameters are unknown. The maximum likelihood estimation of the AR(1) model's constant is denoted $\hat{\rho}_0$, and that of the slope coefficient is denoted $\hat{\rho}$. In example model (3.6), we obtain the demeaned model residuals $\hat{z}_t = x_t - \hat{\rho}_0 - \hat{\rho} x_{t-1},\ t = 2, 3, \ldots, T$ with consistent AR coefficient $\hat{\rho}$, whose moments are computed for the test purpose. As a comparison, we also obtain (demeaned) moments of X_T and execute the moment-based test.

The Monte-Carlo experiments feature the same setup as Table 2-4 in Dufour and Luger (2017). Overall, our results are in line with theirs and with less computational complexity from deriving the pivotal statistics.

Empirical size of tests. Time series samples of T observations are generated by model (3.6) under the hypothesis that no latent state switch in model parameters exists. The DGP is AR(1) with no constant and coefficient of the first lag ρ. The error term is characterized by the independent standard normal distribution.

Table C.1 reports the empirical size of the test at a nominal significance level of 5%. In practice, ρ is set at 0.1 or 0.9, to show low and high intertemporal correlation. Tests are executed with sample sizes of $T = 100$ or 200. For both values of ρ, reported empirical sizes are significantly lower than the nominal level 5% with only the first 4 and 12 moments of observations. With the first 20 moments incorporated and a larger sample size, the reported empirical sizes are close to 5% for the Type B test.

Empirical power of tests. Samples of T observations are generated by model (3.7), $T = 100, 200$. The DGP is AR(1) with two latent states which follow a first-order Markov process. Three transition matrices are considered: $(p_{11}, p_{22}) = (0.9, 0.9), (0.9, 0.5), (0.9, 0.1)$. The model parameters are summarized as,

	State 1	State 2
Case 1	$\mu_1 = 0, \sigma_1 = 1$	$\mu_2 = 2, \sigma_2 = 1$
Case 2	$\mu_1 = 2, \sigma_1 = 1$	$\mu_2 = 2, \sigma_2 = 2$
Case 3	$\mu_1 = 0, \sigma_1 = 1$	$\mu_2 = 2, \sigma_2 = 3$

Table C.2 and C.3 reports the empirical power of tests with $\rho = 0.1$ and $\rho = 0.9$, respectively. The nominal level of significance is set at 5%. The power of the proposed test increases with the sample size, and disparity in μ and σ. The test power is the lowest with inactive regime switch $(p_{11}, p_{22}) = (0.9, 0.9)$, and increases with the frequency of the switch. With $(\Delta\mu, \Delta\sigma) = (2, 2)$, the empirical power exceeds 96% when $\rho = 0.1$ for both types of test, and it exceeds 92% when $\rho = 0.9$ for Type B test.

3.3.2 Markov-switching (auto)regression model with switching coefficients

We present the Markov-switching regression model used in Ghysels and Marcellino (2018). Unlike the example in the previous section, the slope coefficient is dynamic and determined by the current state. $X_T = \{x_1, x_2, \ldots, x_T\}$ is generated from an autoregressive process with state-dependent AR coefficients and constant,

$$x_t = \rho_{0,s_t} + \sum_{k=1}^{r} \rho_{k,s_t} x_{t-k} + \epsilon_t \tag{3.8}$$

where $\{S_t\}$ takes values $\{1, 2\}$ and follows a first-order Markov process, but states are not observed. ϵ_t is i.i.d. white noise.

Even with the simple illustration of AR(1) model $x_t = \rho_{0,s_t} + \rho_{1,s_t} x_{t-1} + \epsilon_t$, the distribution of x_t is not as straightforward as a mixture of normal distributions when the parameters are state-sensitive. $\{x_t - \rho_{0,s_t} - \rho_{1,s_t} x_{t-1}\}$ would be a standard normal process with static ρ_0 and ρ_1.

Consider the two-state case where $\rho_{k,1} \neq \rho_{k,2}$, $k = 0, 1$,

$$\begin{aligned} x_t &= \rho_{0,s_t} + \rho_{1,s_t} x_{t-1} + \epsilon_t \\ &= (\rho_{0,1} + \rho_{1,1} x_{t-1}) \mathbb{I}[S_t = 1] + (\rho_{0,2} + \rho_{1,2} x_{t-1}) \mathbb{I}[S_t = 2] + \epsilon_t \end{aligned} \tag{3.9}$$

Let $\hat{\rho}_0$ and $\hat{\rho}_1$ be the maximum likelihood estimation of the AR(1) model parameters with no switch. $\mathbb{I}(\cdot)$ is the indicator function.

$$\begin{aligned} \hat{z}_t &= x_t - \hat{\rho}_0 - \hat{\rho}_1 x_{t-1} \\ &= (\rho_{0,1} - \hat{\rho}_0) \mathbb{I}[S_t = 1] + (\rho_{0,2} - \hat{\rho}_0) \mathbb{I}[S_t = 2] \\ &+ (\rho_{1,1} - \hat{\rho}_1) x_{t-1} \mathbb{I}[S_t = 1] + (\rho_{1,2} - \hat{\rho}_1) x_{t-1} \mathbb{I}[S_t = 2] + \epsilon_t \end{aligned} \tag{3.10}$$

The model residuals fluctuate with the changing regimes thus the sequence is not stationary. Since the slope coefficients are state-dependent, it is difficult to transform $\hat{z}_t$ into a mixture of multiple normal processes like in the previous example. The model residuals should follow a normal distribution without a state switch.

Model illustrated in the experiment is an AR(2) model with a constant. The coefficient on the first lag is dynamic, and the coefficient on the second lag is static.

$$x_t = \rho_{0,s_t} + \rho_{1,s_t} x_{t-1} + \rho_2 x_{t-2} + \epsilon_t \tag{3.11}$$

which can be rewritten as

$$x_t = \sum_{i=1}^{2} \rho_{0,i} \mathbb{I}[S_t = i] + \sum_{i=1}^{2} \rho_{1,i} x_{t-1} \mathbb{I}[S_t = i] + \rho_2 x_{t-2} + \epsilon_t \tag{3.12}$$

Ghysels and Marcellino (2018) compares the one-step ahead recursive forecast performance of an AR(2) model and a Markov-switching AR(2) model estimated with data from a Markov-switching AR(2) DGP. They conclude that both models have similar forecasting accuracy. We adopt the same DGP, however, our test rejects the null hypothesis of no switching regimes.

Empirical size of tests Samples are generated from model (3.11) with constant $\rho_0 = -0.2$, $\rho_1 = 1.2$, and $\rho_3 = -0.3$. Sample sizes are 500 or 1000.

The upper panel of Table C.4 shows the test results. With only the first four moments, almost no rejections are reported out of the 5000 random samples. The empirical size grows closer to 5% as more moments are incorporated, and when the sample size doubles, as expected.

Empirical power of the tests Samples of T observations are generated from model (3.12), $T = 500, 1000$. When $S_t = 1$, the model parameters are the same as in the experiment for empirical size. When $S_t = 2$, the parameters change to $\rho_{0,2} = 0.2$ and $\rho_{1,2} = 0.4$. ρ_2 remains constant. The transition matrix is $(p_{11}, p_{22}) = (0.83, 0.75)$.

The lower panel of Table C.4 shows the empirical power of the tests. With the suggested model, when higher order of moments is applied, Type B test outperforms Type A test, and the rejection rate ranges between 50% to 92%. It demonstrates that the moment-based test captures the significant difference between the DGP and the generic AR model under the null hypothesis.

3.4 US Real GNP Growth

Hamilton (1989) suggests modeling the US real GNP growth from the year 1952 to 1984 with a Markov-switching specification as in model (3.5) with $r = 4$, MSAR(4), where only the mean is subject to regime changes. The author argues that the business cycles of expansion and recession are reflected by the switch between high and low growth rates in the economy, thus the MSAR(4) model outperforms the linear model by featuring the changes. In this section, we apply the proposed simulation-based MC procedures to test against the linear AR(4) specification by examining the original Hamilton's model of observations from 1952Q2 to 1984Q4, and an extended sample from 1952Q2 to 2010Q4 as in Dufour and Luger (2017).

The quarterly US GNP data is acquired from the Federal Reserve Economic Database[5]. The quarterly growth rate at period t is computed as

$$r_t = (\ln GNP_t - \ln GNP_{t-1}) \times 100$$

[5]This data is available from U.S. Bureau of Economic Analysis, Gross National Product [GNP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GNP

Since we consider the AR(4) model specification, four more observations are added prior to the start of the data. The sample size equals 135 for Hamilton's original data set, and it equals 239 for the extended sample.

We estimate the model by maximum likelihood estimation with observations of r_t's,

$$r_t = \rho_0 + \rho_1 r_{t-1} + \rho_2 r_{t-2} + \rho_3 r_{t-3} + \rho_4 r_{t-4} + \epsilon_t \tag{3.13}$$

and obtain the constant and slope coefficients estimates $\hat{\rho}_0, \hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4$. The fitted model is then used as the simulator in the test. The Type B test is based on the model residuals:

$$\hat{z}_t = r_t - \hat{\rho}_0 - \hat{\rho}_1 r_{t-1} - \hat{\rho}_2 r_{t-2} - \hat{\rho}_3 r_{t-3} - \hat{\rho}_4 r_{t-4} \tag{3.14}$$

Table C.5 summarizes the empirical p-values out of 1000 repeated tests when the order of moment matched $H = 4$, 12, or 20. For the estimation of the optimal operator and its inverse, we consider $N = 20$ bootstrapped samples and an arbitrary tuning parameter $a = 0.0001$.

In Table C.5, results displayed in Panel A are based on observations between 1952Q2 and 1984Q4. Both types of tests with smaller and higher orders of moments show that the null hypothesis of a linear AR(4) model cannot be rejected at the usual levels of significance. It demonstrates that Hamilton's selected data is compatible with a linear and stationary autoregressive presentation. Panel B presents test results using the extended observations from 1952Q2 to 2010Q4. For Type A tests using moments of demeaned observations, with smaller numbers of moments $H = 4$ and 12, the null hypothesis is rejected at a significance level of 10% but not rejected at a significance level of 5% or lower. When the order of moments increases to 20, more information can be harvested so that the null hypothesis can be rejected at a significance level of 5% or 2.5%. The result is in line with the findings of Dufour and Luger (2017) that echo the detection of the structural break in the US economic growth by Kim and Nelson (1999) and McConnell and Perez-Quiros (2000). It is worth noting that in this case, the Type B test gives a different conclusion. A possible explanation can be that a sample size of 239 is still not large enough and part of the residual information is lost when the data is filtered by the AR(4) specification.

3.5 Conclusion

The present study proposes a novel Monte-Carlo test for comparing a Markov-switching autoregressive model against a linear specification. Our approach allows for the incorporation of a large number of empirical moments without requiring analytical expressions. The test's size is controlled by the Monte-Carlo method, and our simulation experiments demonstrate that more information can be obtained with a higher number of matched moments, with evidence of increased empirical power. We apply our method to

test the specification of a Markov-switching AR(4) model for the US output growth rate and find that the null hypothesis of linearity is not rejected using Hamilton's original dataset, but it is rejected when using an extended sample that includes more recent observations. This result aligns with previous research on US business cycles. Potential avenues for future research include developing a data-driven method to determine the regularization parameter and exploring alternative testing statistics beyond empirical moments.

Bibliography

Amengual, D., M. Carrasco, and E. Sentana (2020). Testing distributional assumptions using a continuum of moments. *Journal of Econometrics 218*, 655–689.

Andrews, D. (1991). Heteroskedasticity and autocorreltaion consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Antoine, B., L. Khalaf, M. Kichian, and Z. Lin (2022). Simulation-based matching inference with applications to DSGE models. *forthcoming Journal of Business & Economic Statistics*.

Antoine, B., L. Khalaf, M. Kichian, and Z. Lin (2023). Identification-robust inference with simulation-based pseudo-matching. *Journal of Business & Economic Statistics 41*(2), 321–338.

Antoine, B. and W. Sun (2022). Simulation-based estimation with many auxiliary statistics applied to long-run dynamic analysis. Working paper, Simon Fraser University.

Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics 170*(2), 383–398.

Carrasco, M., M. Chernov, J.-P. Florens, and E. Ghysels (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics 140*(2), 529 – 573.

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

Carrasco, M., L. Hu, and W. Ploberger (2014). Optimal test for markov switching parameters. *Econometrica 82*(2), 765–784.

Chen, X. and H. White (1998). Central limit and functional central limit theorems for hilbert-valued dependent heterogeneous arrays with applications. *Econometric Theory 14*(2), 260–284.

Christiano, L., M. Eichenbaum, and C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy 113*, 1–45.

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 64*(2), 247–254.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika 74*(1), 33–43.

Davies, R. B. (2002). Hypothesis testing when a nuisance parameter is present only under the alternative: linear model case. *Biometrika*, 484–489.

DelNegro, M. and F. Schorfheide (2008). Forming priors for dsge models (and how it affects the assessment of nominal rigidities). *Journal of Monetary Economics 55*, 1191–1208.

Dovonon, P. and S. Gonçalves (2017). Bootstrapping the gmm overidentification test under first-order underidentification. *Journal of Econometrics 201*(1), 43 – 71.

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

Duffie, D. and K. Singleton (1993). Simulated moments estimation of markov models of asset prices. *Econometrica 61*, 929–952.

Dufour, J.-M. and L. Khalaf (2001). Monte carlo test methods in econometrics. *Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, UK*, 494–519.

Dufour, J.-M., L. Khalaf, J.-T. Bernard, and I. Genest (2004). Simulation-based finite-sample tests for heteroskedasticity and arch effects. *Journal of Econometrics 122*(2), 317–347.

Dufour, J.-M. and R. Luger (2017). Identification-robust moment-based tests for markov switching in autoregressiv models. *Econometric Reviews 36*(6-9), 713–727.

Fernandez-Villaverde, J., J. Rubio-Ramirez, and F. Schorfheide (2016). Solution and estimation methods for dsge models. *Handbook of Macroeconomics 2*, 527–724.

Fernández-Villaverde, J., J. Rubio-Ramírez, and F. Schorfheide (2016). Chapter 9 - solution and estimation methods for dsge models. Volume 2 of *Handbook of Macroeconomics*, pp. 527–724. Elsevier.

Forneron, J.-J. and S. Ng (2018). The abc of simulation estimation with auxiliary statistics. *Journal of econometrics 205*, 112–139.

Gallant, R. and G. Tauchen (1996). Which moments to match? *Econometric Theory 12*, 657–681.

Garcia, R. (1998). Asymptotic null distribution of the likelihood ratio test in markov switching models. *International Economic Review*, 763–788.

Ghysels, E. and M. Marcellino (2018). *Applied economic forecasting using time series methods*. Oxford University Press.

Goncalves, S., A. Herrera, L.Kilian, and E. Pesavento (2021). Impulse response analysis for structural dynamic models with nonlinear regressors. *Journal of Econometrics*.

Gouriéroux, C. and A. Monfort (1996). *Simulation-Based Econometric Methods*. Oxford: Oxford University Press.

Gouriéroux, C. S., A. Monfort, and E. M. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*(S1), S85 – S118.

Groetsch, C. W. (1993). *Inverse problems in the mathematical sciences*, Volume 52. Springer.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2017). Impulse response matching estimators for dsge models. *Journal of Econometrics 196*(1), 144 – 155.

Hamilton, J. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, 357–384.

Hansen, B. E. (1992). The likelihood ratio test under nonstandard conditions: testing the markov switching model of gnp. *Journal of applied Econometrics 7*(S1), S61–S82.

Hansen, B. E. (1996). Erratum: The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp. *Journal of Applied econometrics*, 195–198.

Hassler, U. and P. Kokoszka (2010). Impulse Responses of Factionally Integrated Processes with Long Memory. *Econometric Theory 26*, 1855–1861.

Inoue, A. and L. Kilian (2013). Inference on impulse response functions in structural VAR models. *Journal of Econometrics 177*, 1–13.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review 95*, 161–182.

Khalaf, L., Z. Lin, and A. Reza (2019). Beyond the Linearized Straight Jacket: Finite Sample Inference for - Possibly Singular - DSGE Models. *Working paper*.

Khalaf, L. and C. Saunders (2019). Monte-Carlo two-stage indirect inference (2SIF) for autoregressive panels. *Journal of Econometrics forthcoming*.

Kim, C.-J. and C. R. Nelson (1999). Has the us economy become more stable? a bayesian approach based on a markov-switching model of the business cycle. *Review of Economics and Statistics 81*(4), 608–616.

Lütkepohl, H. (1990). Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. *The Review of Economics and Statistics 71*, 116–125.

McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review 90*(5), 1464–1476.

Mor, B., S. Garhwal, and A. Kumar (2021). A systematic review of hidden markov models and their applications. *Archives of computational methods in engineering 28*, 1429–1448.

Nashed, M. Z. and G. Wahba (1974). Regularization and approximation of linear operator equations in reproducing kernel spaces. *Bull. Amer. Math. Soc. 80*, 1213–1218.

Plagborg-Møller, M. and C. Wolf (2021). Local projections and vars estimate the same impulse responses. *Econometrica 89*(2), 955–980.

Schorfheide, F. (2008). Dsge model-based estimation of the new keynesian phillips curve. *FRB Richmond Economic Quarterly 94*(4), 397–433.

Smets, F. and F. Wouters (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review 97*(3), 586–606.

Smith, A. (1993). Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*, S63–S84.

Sokullu, S. (2020). A regularization approach to the minimum distance estimation: application to structural macroeconomic estimation using irfs. *Oxford Economic Papers 72*(2), 546–565.

Wahba, G. (1973). Convergence rates of certain approximate solutions to fredholm integral equations of the first kind. *Journal of Approximation Theory 7*(2), 167 – 185.

Appendix A

Simulation-based Methodology with Many Auxiliary Statistics: An Estimator

A.1 Proofs

• Proof of Theorem 1:
Under our regularity assumptions, the consistency of extremum estimators follows, and the proof is rather standard. It requires showing that

$$\sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| = o_{\mathbb{P}}(1) .$$

This, together with the fact that θ_0 is the unique solution of $z(\theta_0) = 0$ - and unique minimizer of $Q(\theta)$ over Θ - delivers the result. ■

• Proof of Theorem 2:
We start our proof by showing a preliminary result.

Lemma A.1.1. *Assumption 7 implies that* $B_T(\sqrt{T} z_T(., \theta_0)) \xrightarrow{d} BZ \sim \mathcal{N}(0, BKB')$ *with* B' *the adjoint operator of* B.

Proof of Lemma A.1.1:
Throughout, we write $z_T(\theta)$ for $z_T(., \theta)$. By assumption 7(ii), the random element $\sqrt{T} z_T(\theta_0)$ is bounded for T large enough, and it converges to Z in distribution as $T \to \infty$. By definition, the covariance of Z is $E\Big[(Z - EZ)(Z - EZ)\Big]$. Then, for any f well-defined in the Hilbert

space H, the covariance of the inner product (Z, f) is:

$$\begin{aligned} E\Big([(Z,f)-E(Z,f)][(Z,f)-E(Z,f)]\Big) &= E\Big([(Z,f)-(EZ,f)][(Z,f)-(EZ,f)]\Big) \\ &= E\Big[(Z-EZ,f)(Z-EZ,f)\Big] \\ &= E\Big((Z-EZ,f)(Z-EZ),f\Big), \end{aligned}$$

where $E\Big[(Z-EZ,f)(Z-EZ)\Big] \equiv Kf$ defines the covariance operator K.

Then, we can show that $B_T\sqrt{T}z_T(\theta_0) \xrightarrow{d} BZ$:

$$\begin{aligned} \Big\|BZ - B_T\sqrt{T}z_T(\theta_0)\Big\| &= \Big\|BZ - B_TZ + B_TZ - B_T\sqrt{T}z_T(\theta_0)\Big\| \\ &\leq \|BZ - B_TZ\| + \Big\|B_TZ - B_T\sqrt{T}z_T(\theta_0)\Big\| \\ &\leq \|B - B_T\|\|Z\| + \|B_T\|\Big\|Z - \sqrt{T}z_T(\theta_0)\Big\| \\ &\xrightarrow{\mathbb{P}} 0 \end{aligned}$$

which follows from Assumption 7(i) and (ii) which ensure that each term is either bounded or converging to 0 appropriately.
Similarly, the covariance of the inner product (BZ, f) is

$$\begin{aligned} & E\Big([(BZ,f)-E(BZ,f)][(BZ,f)-E(BZ,f)]\Big) \\ = \; & E\Big([(BZ,f)-(EBZ,f)][(BZ,f)-(EBZ,f)]\Big) \\ = \; & E\Big[(BZ-B(EZ),f)(BZ-B(EZ),f)\Big] \\ = \; & E\Big[(B(Z-EZ),f)(B(Z-EZ),f)\Big] \\ = \; & E\Big((B[Z-EZ],f)B[Z-EZ],f\Big). \end{aligned}$$

Define the operator BKB' such that

$$(BKB')(f) \equiv E\Big[(B[Z-EZ],f)B[Z-EZ]\Big].$$

The covariance of (BZ, f) is then $\Big(f, (BKB')(f)\Big)$. □

We now return to the proof of Theorem 2. From the definition of the SMAS estimator, we have:

$$\hat{\theta}_{SMAS} \equiv \arg\min_{\theta\in\Theta}\|B_Tz_T(.,\theta)\| = \arg\min_{\theta\in\Theta}\Big(B_Tz_T(.,\theta), B_Tz_T(.,\theta)\Big).$$

By Assumption 6 and from the symmetry of the inner product, the first order conditions (FOC) are:

$$\left(B_T \frac{\partial}{\partial \theta'} z_T(\hat{\theta}_{SMAS}), B_T z_T(\hat{\theta}_{SMAS}) \right) \quad = \quad 0 \,.$$

From Theorem 1, we know $\hat{\theta}_{SMAS} \xrightarrow{a.s.} \theta_0$; then, and a mean value expansion of $z_T(\hat{\theta}_{SMAS})$ around θ_0 yields:

$$z_T(\hat{\theta}_{SMAS}) = \Big(z_T(\theta_0) + \frac{\partial}{\partial \theta'} z_T(\bar{\theta})(\hat{\theta}_{SMAS} - \theta_0) \Big) = \Big(z_T(\theta_0) + G_T(\bar{\theta})(\hat{\theta}_{SMAS} - \theta_0) \Big) \,,$$

where $\bar{\theta}$ lies between θ_0 and $\hat{\theta}_{SMAS}$ component by component, and $G_T(\theta) \equiv \partial z_T(\theta)/\partial \theta'$. Substitute back into the FOC to get:

$$\begin{aligned}
& \left(B_T G_T(\hat{\theta}_{SMAS}), B_T \Big(z_T(\theta_0) + G_T(\bar{\theta})(\hat{\theta}_{SMAS} - \theta_0) \Big) \right) = 0 \\
\Leftrightarrow \quad & \left(B_T G_T(\hat{\theta}_{SMAS}), B_T z_T(\theta_0) + B_T G_T(\bar{\theta})(\hat{\theta}_{SMAS} - \theta_0) \right) = 0 \\
\Leftrightarrow \quad & \left(B_T G_T(\hat{\theta}_{SMAS}), B_T z_T(\theta_0) \right) + \left(B_T G_T(\hat{\theta}_{SMAS}), B_T G_T(\bar{\theta})(\hat{\theta}_{SMAS} - \theta_0) \right) = 0 \,.
\end{aligned}$$

Under our regularity assumptions, combined with Lemma A.1.1, we can write, for T large enough:

$$\sqrt{T}(\hat{\theta}_{SMAS} - \theta_0) = \Big(BG(\theta_0), BG(\theta_0) \Big)^{-1} \Big(BG(\theta_0), BZ \Big) + o_{\mathbb{P}}(1) \,.$$

Since $BZ \sim \mathcal{N}(0, BKB')$ and $BG(\theta_0) \in H$, we have, by definition

$$(BG(\theta_0), BZ) \sim \mathcal{N}(0, \Big(BG(\theta_0), (BKB')BG(\theta_0) \Big)) \,.$$

Therefore, as $T \to \infty$, we have:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, V) \,,$$

$$\text{with} \quad V = \|BG(\theta_0)\|^{-2} \Big(BG(\theta_0), (BKB')BG(\theta_0) \Big) \|BG(\theta_0)\|^{-2} \,.$$

■

• Proof of Lemma 3:
In this proof, we keep the length of the bootstrap path used to compute $z_T^{*(n)}(h, .)$ equal to $(T - Q)$ (where Q corresponds to the cutoff determined by the rule-of-thumb used to choose the bandwidth associated with the Bartlett kernel). In addition, when there is no confusion, we simplify our notations as follows:

$$\begin{aligned}
z_{T,h,1}^{*(n)} &\equiv z_T^{*(n)}(h, X_1^{(n)}(T - Q), \theta_0) \\
z_{T,h,1+|m|}^{*(n)} &\equiv z_T^{*(n)}(h, X_{1+|m|}^{(n)}(T - Q), \theta_0) \\
\bar{z}_{T,h}^{*} &\equiv \bar{z}_T^{*}(h, \theta_0)
\end{aligned}$$

for any integer h and any $m = -Q, \cdots, Q$.

By definition, any eigenfunction $\phi^{(T)}$ and corresponding eigenvalue $\lambda^{(T)}$ of the operator K_T are such that:

$$\begin{aligned} & (K_T\phi^{(T)})(h) = \lambda^{(T)}\phi^{(T)}(h) \\ \Leftrightarrow \quad & \frac{1}{N(T)}\sum_{n=1}^{N(T)}\sum_{m=-Q}^{Q}\omega\left(\frac{m}{M_T}\right)\left(z_{T,h,1}^{*(n)} - \bar{z}_{T,h}^{*}\right)\sum_{s=1}^{\infty}\left(z_{T,s,1+|m|}^{*(n)} - \bar{z}_{T,s}^{*}\right)\phi^{(T)}(s) \\ & = \lambda^{(T)}\phi^{(T)}(h)\,, \end{aligned}$$

from the representation of K_T. Since $\lambda^{(T)}$ is a scalar, and since

$$\sum_{s=1}^{\infty}\left(z_{T,s,1+|m|}^{*(n)} - \bar{z}_{T,s}^{*}\right)\phi^{(T)}(s)$$

does not depend on s, there must exist some β^n such that:

$$\phi^{(T)}(h) = \frac{1}{N(T)}\sum_{n=1}^{N(T)}\left(z_{T,h,1}^{*(n)} - \bar{z}_{T,h}^{*}\right)\beta^n\,.$$

Overall, we can write:

$$\begin{aligned} & (K_T\phi^{(T)})(h) = \lambda^{(T)}\phi^{(T)}(h) \\ \Leftrightarrow \quad & \frac{1}{N(T)}\sum_{n=1}^{N(T)}\left(z_{T,h,1}^{*(n)} - \bar{z}_{T,h}^{*}\right)\sum_{s=1}^{\infty}\sum_{m=-Q}^{Q}\omega\left(\frac{m}{M_T}\right)\left(z_{T,s,1+|m|}^{*(n)} - \bar{z}_{T,s}^{*}\right) \\ & \times\frac{1}{N(T)}\sum_{n'=1}^{N(T)}\left(z_{T,s,1}^{*(n')} - \bar{z}_{T,s}^{*}\right)\beta^{n'} \\ & = \lambda^{(T)}\frac{1}{N(T)}\sum_{n=1}^{N(T)}\left(z_{T,h,1}^{*(n)} - \bar{z}_{T,h}^{*}\right)\beta^n \end{aligned}$$

To solve for $\lambda^{(T)}$ and $\mathcal{B} = [\beta^1\ \beta^2\ ...\ \beta^{N(T)}]'$ in the previous equation, it is equivalent to solve the following system of $N(T)$ equations, for $n = 1, \ldots, N(T)$:

$$\frac{1}{N(T)}\sum_{n'=1}^{N(T)}\beta^{n'}\sum_{s=1}^{\infty}\left(z_{T,s,1}^{*(n')} - \bar{z}_{T,s}^{*}\right)\sum_{m=-Q}^{Q}\omega\left(\frac{m}{M_T}\right)\left(z_{T,s,1+|m|}^{*(n)} - \bar{z}_{T,s}^{*}\right) \quad = \quad \lambda^{(T)}\beta^n\,.$$

And, with the $(N(T), N(T))$-matrix C as defined in Lemma 3, the above system of linear equations can be rewritten as:

$$C\mathcal{B} = \lambda^{(T)}\mathcal{B}\,.$$

It is then easy to see that the eigenvalues of K_T, namely $\lambda_j^{(T)}$, are also the eigenvalues of the matrix C with associated eigenvectors $\mathcal{B}_j = [\beta_j^1\ \beta_j^2\ ...\ \beta_j^{N(T)}]'$, $j = 1, \ldots, N(T)$. Further,

the eigenfunctions of K_T are such that, for $j = 1, \ldots, N(T)$,

$$\phi_j^{(T)}(h) = \frac{1}{N(T)}\left(z_h^{(T)}\right)' \mathcal{B}_j \quad \text{with} \qquad z_h^{(T)} \equiv \begin{bmatrix} \sum_{m=-Q}^{Q} \omega\left(\frac{m}{M_T}\right)\left(z_{T,h,1+|m|}^{*(1)} - \bar{z}_{T,h}^{*}\right) \\ \sum_{m=-Q}^{Q} \omega\left(\frac{m}{M_T}\right)\left(z_{T,h,1+|m|}^{*(2)} - \bar{z}_{T,h}^{*}\right) \\ \vdots \\ \sum_{m=-Q}^{Q} \omega\left(\frac{m}{M_T}\right)\left(z_{T,h,1+|m|}^{*(N(T))} - \bar{z}_{T,h}^{*}\right) \end{bmatrix}.$$

■

• **Proof of Theorem 4**:
We start our proof by showing a preliminary result about the asymptotic behavior of the eigenvalues of the operator.

Theorem A.1.1. *Under Assumptions 1 to 10, when $T/N(T) \to \zeta$ as $T \to \infty$ with $0 < \zeta < \infty$, we have:*

$$(\lambda_j^{(T)} - \lambda_j) \quad = \quad \begin{cases} \mathcal{O}_{P^*}(\frac{1}{\sqrt{T}}) \text{ in prob-}\mathbb{P} & \text{when the autocorrelations can be ignored} \\ \mathcal{O}_{P^*}(\frac{1}{T^{1/3}}) \text{ in prob-}\mathbb{P} & \text{otherwise} \end{cases}$$

Proof of Theorem A.1.1:
Our proof builds on the proof of Theorem 3 in Carrasco and Florens (2000). We consider λ_j as a function of F, the c.d.f of the joint probability measure $\mathbb{P}$, that is $\lambda_j = \Lambda(F)$. The bootstrap counterpart of F is denoted F_T, and, accordingly, we have $\lambda_j^{(T)} = \Lambda(F_T)$. We define $D\Lambda_F$ as the Fréchet derivative of the operator Λ_j with respect to F. A first-order Taylor expansion with Fréchet derivative gives

$$\lambda_j^{(T)} - \lambda_j = D\Lambda_F(F_T - F) + \epsilon(F_T - F)\|F_T - F\| \,.$$

where norm is the sup-norm. Also, under Assumptions 7 and 8, the term $\epsilon(F_T - F)$ converges to zero, and $\sqrt{T}\|F_T - F\|$ is bounded in the sense that

$$\epsilon(F_T - F) \xrightarrow{P^*} 0 \quad \text{in prob-}\mathbb{P} \quad \text{and} \quad \sqrt{T}\|F_T - F\| = \mathcal{O}_{P^*}(1) \quad \text{in prob-}\mathbb{P}.$$

As a result, we have:

$$\sqrt{T}(\lambda_j^{(T)} - \lambda_j) = \sqrt{T} D\Lambda_F(F_T - F) + o_{P^*}(1) \text{ in prob} - \mathbb{P}\,. \tag{A.1}$$

Rewrite the following equation $(K\phi_j)(h) = \lambda_j \phi_j(h)$ as

$$\begin{aligned} \sum_s k^*(h,s)\phi_j(s) = \lambda_j \phi_j(h) \quad &\Leftrightarrow \quad \sum_s E_{\mathbb{P}}(z^*(h,s))\phi_j(s) = \lambda_j \phi_j(h) \\ &\Leftrightarrow \quad \sum_s E_F(z^*(h,s))\phi_j(s) = \lambda_j \phi_j(h)\,, \end{aligned} \tag{A.2}$$

with

$$\begin{aligned} & z^*(h,s) \\ \equiv & \lim_m \sum_{-m}^{m} (z^*(h, X_t^*, \theta_0) - E_{\mathbb{P}}(z^*(h, X_t^*, \theta_0)))(z^*(s, X_{t+m}^*, \theta_0) - E_{\mathbb{P}}(z^*(s, X_{t+m}^*, \theta_0))) \end{aligned}$$

Differentiate equation (A.2) with respect to F to get:

$$\sum_s E_{\widetilde{F}}(z^*(h,s))\phi_j(s) + \sum_s E_F(z^*(h,s))\widetilde{\phi}_j(s) = \lambda_j \widetilde{\phi}_j(h) + \widetilde{\lambda}_j \phi_j(h)\,.$$

where $E_{\widetilde{F}}$, $\widetilde{\phi}$, and $\widetilde{\lambda}$ denote the differential elements respectively[1].

Multiply by $\phi_j(h)$ on both sides, and integrate with respect to h to get:

$$\begin{aligned} & \sum_h \sum_s E_{\widetilde{F}}(z^*(h,s))\phi_j(h)\phi_j(s) + \sum_h \sum_s E_F(z^*(h,s))\phi_j(h)\widetilde{\phi}_j(s) \\ = & \sum_h \lambda_j \widetilde{\phi}_j(h)\phi_j(h) + \sum_h \widetilde{\lambda}_j \phi_j^2(h)\,. \end{aligned}$$

Assume now that the eigenvalues λ_j (and $\lambda_j^{(T)}$) are ranked in descending order, and that the eigenfunctions $\phi_j(.)$ (and $\phi_j^{(T)}(.)$) are orthonormalized (e.g. $\sum_h \phi_j^2(h) = 1$). The previous equation simplifies to:

$$\begin{aligned} & \sum_h \sum_s E_{\widetilde{F}}(z^*(h,s))\phi_j(h)\phi_j(s) + \sum_h \sum_s E_F(z^*(h,s))\phi_j(h)\widetilde{\phi}_j(s) \\ = & \sum_h \lambda_j \widetilde{\phi}_j(h)\phi_j(h) + \widetilde{\lambda}_j\,. \end{aligned}$$

Using (A.2), the second term on the left hand side can be rewritten as:

$$\begin{aligned} \sum_h \sum_s E_F(z^*(h,s))\phi_j(h)\widetilde{\phi}_j(s) &= \sum_s \widetilde{\phi}_j(s) \sum_h E_F(z^*(h,s))\phi_j(h) \\ &= \sum_s \widetilde{\phi}_j(s)\lambda_j\phi_j(s) \\ &= \sum_h \lambda_j \widetilde{\phi}_j(h)\phi_j(h) \end{aligned}$$

Therefore, we get:

$$\begin{aligned} \widetilde{\lambda}_j &= \sum_h \sum_s E_{\widetilde{F}}(z^*(h,s))\phi_j(h)\phi_j(s) \\ &= \sum_h \sum_s E_{F_T}(z_T^*(h,s))\phi_j(h)\phi_j(s) - \sum_h \sum_s E_F(z^*(h,s))\phi_j(h)\phi_j(s) \\ &\quad + \epsilon'(F_T - F)\|F_T - F\| \\ &= \sum_h \sum_s E_{F_T}(z_T^*(h,s))\phi_j(h)\phi_j(s) - \lambda_j + \epsilon'(F_T - F)\|F_T - F\|\,. \end{aligned}$$

[1]For example, $\widetilde{\lambda} = D\Lambda(\Delta F)$

where $\epsilon'(F_T - F) \xrightarrow{P^*} 0$ in prob-$\mathbb{P}$ and $\sqrt{T}\|F_T - F\| = \mathcal{O}_{P^*}(1)$ in prob-$\mathbb{P}$.

After substituting into (A.1), we have:

$$\begin{aligned}
& \sqrt{T}(\lambda_j^{(T)} - \lambda_j) \\
= \; & \sqrt{T}\left[\sum_h \sum_s E_{F_T}(z_T^*(h,s))\phi_j(h)\phi_j(s) - \lambda_j\right] + o_{P^*}(1) \\
= \; & \frac{\sqrt{T}}{N(T)} \sum_{n=1}^{N(T)} \left[\sum_h \sum_s \sum_{m=-N(T)+1}^{N(T)-1} \omega\left(\frac{m}{M_T}\right)\left(z_{T,h,1}^{*(n)} - \bar{z}_{T,h}^*\right)\left(z_{T,s,1+|m|}^{*(n)} - \bar{z}_{T,s}^*\right) - \lambda_j\right] \\
& + o_{P^*}(1) \\
= \; & \frac{\sqrt{T}}{N(T)^{1/3}} \times \frac{N(T)^{1/3}}{N(T)} \sum_{n=1}^{N(T)} \left[\sum_h \sum_s \sum_{m=-N(T)+1}^{N(T)-1} \omega\left(\frac{m}{M_T}\right)\right. \\
& \left. \times\left(z_{T,h,1}^{*(n)} - \bar{z}_{T,h}^*\right)\left(z_{T,s,1+|m|}^{*(n)} - \bar{z}_{T,s}^*\right) - \lambda_j\right] + o_{P^*}(1)
\end{aligned}$$

Under Assumption 10, and the maintained regularity conditions on the chosen Bartlett kernel and associated bandwidth, a bootstrap CLT applies to

$$\left[\sum_h \sum_s \sum_{m=-N(T)+1}^{N(T)-1} \omega\left(\frac{m}{M_T}\right)\left(z_{T,h,1}^{*(n)} - \bar{z}_{T,h}^*\right)\left(z_{T,s,1+|m|}^{*(n)} - \bar{z}_{T,s}^*\right) - \lambda_j\right],$$

with the fastest available rate of $N(T)^{1/3}$ as shown in Andrews (1991).
Since $T/N(T) \to \zeta$ some positive constant, the result follows.

When the autocorrelations can be ignored, the expression for $z_T^*(h,s)$ simplifies as it does not involve a kernel and, in the summation over m only one term remains ($m = 0$). In that case, a bootstrap CLT applies instead to

$$\left[\sum_h \sum_s \left(z_{T,h,1}^{*(n)} - \bar{z}_{T,h}^*\right)\left(z_{T,s,1}^{*(n)} - \bar{z}_{T,s}^*\right) - \lambda_j\right],$$

with rate $\sqrt{N(T)}$, and the result follows. ■

We now return to the proof of Theorem 4. Our proof builds on the proof of Theorem 4 in Carrasco and Florens (2000) and the proof of Theorem 3.3 in Carrasco et al. (2007). By Assumption 9, the kernel $k^*(h,s)$ satisfies

$$\sum_h \sum_s k^*(h,s)^2 = \sum_{j=1}^{\infty} \lambda_j^2 < \infty,$$

and the Hilbert-Schmidt norm of K is defined as:

$$\|K\|_{HS} = \left(\sum_{j=1}^{\infty} \lambda_j^2\right)^{\frac{1}{2}}.$$

Since $\|K_T - K\| \leq \|K_T - K\|_{HS}$, we have:

$$\begin{aligned}
\|K_T - K\|^2 &\leq \sum_h \sum_s \left[k_T^*(h,s) - k^*(h,s)\right]^2 \\
&= \sum_h \sum_s \left[\frac{1}{N(T)} \sum_{n=1}^{N(T)} k_T^{*(n)}(h,s) - k^*(h,s)\right]^2 \\
&= \frac{1}{N(T)^2} \sum_{n,n'=1}^{N(T)} \sum_{h,s} \left[k_T^{*(n)}(h,s) - k^*(h,s)\right]\left[k_T^{*(n')}(h,s) - k^*(h,s)\right]
\end{aligned}$$

where, with the notations introduced in the proof of Lemma 3,

$$k_T^{*(n)}(h,s) \equiv \sum_{m=-N(T)+1}^{N(T)} \omega\left(\frac{m}{M_T}\right)\left(z_{T,h,1}^{*(n)} - \bar{z}_{T,h}^*\right)\left(z_{T,s,1+|m|}^{*(n)} - \bar{z}_{T,s}^*\right)$$

Because the n-th and n'-th simulation paths are independent, we have

$$E_{P^*}\Big\{ \sum_{h,s} \left[k_T^{*(n)}(h,s) - k^*(h,s)\right]\left[k_T^{*(n')}(h,s) - k^*(h,s)\right] | \mathcal{X}_T^{n'} \Big\} = 0$$

Under Assumption 10, and the maintained regularity conditions on the chosen Bartlett kernel and associated bandwidth, we get, using Theorem A.1.1 and results established in its proof:

$$\|K_T - K\|^2 = \begin{cases} \mathcal{O}_{P^*}(1/T) \text{ in prob-}\mathbb{P} & \text{when the autocorrelations can be ignored} \\ \mathcal{O}_{P^*}(1/T^{2/3}) \text{ in prob-}\mathbb{P} & \text{otherwise} \end{cases}$$

And the expected results follow. ■

• Proof of Theorem 5:
We start our proof by showing a preliminary result to ensure that our simulation-based objective function converges to its population counterpart asymptotically.

Theorem A.1.2. *For any g, g_T such that $\|g_T - g\| = \mathcal{O}_{P^*}(\frac{1}{\sqrt{T}})$ in prob-$\mathbb{P}$, and assuming $\|K_T - K\| = \mathcal{O}_{P^*}(1/T^\nu)$ in prob-$\mathbb{P}$ for some $\nu > 0$, we have:*
(i) $\left\|K_{T,a}^{-\frac{1}{2}} g_T - K^{-\frac{1}{2}} g\right\| \xrightarrow{P^*} 0$ *in prob-$\mathbb{P}$, when $g \in \mathcal{H}(K) + \mathcal{H}(K)^\perp$, as $a \to 0$ and $T^\nu a^{\frac{3}{4}} \to \infty$;*
(ii) $\left\|K_{T,a}^{-1} g_T - K^{-1} g\right\| \xrightarrow{P^*} 0$ *in prob-$\mathbb{P}$, when $g \in \mathcal{D}(K^{-1})$, as $a \to 0$ and $T^\nu a^{3/2} \to \infty$.*

Proof of Theorem A.1.2:
Our proof builds on the proofs of Theorem 7 in Carrasco and Florens (2000) and Lemma B.2 in Carrasco et al. (2007). First, notice that:

$$\left\|K_{T,a}^{-\frac{1}{2}} g_T - K^{-\frac{1}{2}} g\right\| \leq \left\|K_{T,a}^{-\frac{1}{2}} g_T - K_{T,a}^{-\frac{1}{2}} g\right\| + \left\|K_{T,a}^{-\frac{1}{2}} g - K_a^{-\frac{1}{2}} g\right\| + \left\|K_a^{-\frac{1}{2}} g - K^{-\frac{1}{2}} g\right\|$$

We study each of the 3 terms on the RHS separately to show that:

$$\text{(A)} \quad \left\| K_{T,a}^{-\frac{1}{2}} g_T - K_{T,a}^{-\frac{1}{2}} g \right\| \xrightarrow{P^*} 0 \text{ in prob-}\mathbb{P} \text{ as } T\sqrt{a} \to \infty \text{ and } a \to 0$$

$$\text{(B)} \quad \left\| K_{T,a}^{-\frac{1}{2}} g - K_a^{-\frac{1}{2}} g \right\| \xrightarrow{\mathbb{P}} 0 \text{ as } T^\nu a^{3/4} \to \infty \text{ and } a \to 0$$

$$\text{(C)} \quad \left\| K_a^{-\frac{1}{2}} g - K^{-\frac{1}{2}} g \right\| \to 0 \text{ as } a \to 0$$

Then, the expected result follows by applying, e.g. Lemma B.2 from Dovonon and Gonçalves (2017).

• Part (A):

$$\begin{aligned} \left\| K_{T,a}^{-\frac{1}{2}} g_T - K_{T,a}^{-\frac{1}{2}} g \right\| &\leq \left\| K_{T,a}^{-\frac{1}{2}} \right\| \|g_T - g\| \\ \text{with} \quad \left\| K_{T,a}^{-\frac{1}{2}} \right\|^2 &= \left((K_T^2 + aI)^{-\frac{1}{2}} K_T^{\frac{1}{2}}, (K_T^2 + aI)^{-\frac{1}{2}} K_T^{\frac{1}{2}} \right) \\ &= \left((K_T^2 + aI)^{-\frac{1}{2}}, (K_T^2 + aI)^{-\frac{1}{2}} K_T \right) \\ &\leq \left\| (K_T^2 + aI)^{-\frac{1}{2}} \right\| \left\| (K_T^2 + aI)^{-\frac{1}{2}} K_T \right\| . \end{aligned}$$

The second term is bounded by 1, while the first term is bounded by $1/\sqrt{a}$ for T large enough. As a result, since $\|g_T - g\| = \mathcal{O}_{P^*}(\frac{1}{\sqrt{T}})$ by assumption, the result follows as long as $\sqrt{T} a^{1/4} \to \infty$.

• Part (B):

$$\begin{aligned} \left\| K_{T,a}^{-\frac{1}{2}} g - K_a^{-\frac{1}{2}} g \right\| \leq & \left\| (K_T^2 + aI)^{-\frac{1}{2}} K_T^{\frac{1}{2}} g - (K_T^2 + aI)^{-\frac{1}{2}} K^{\frac{1}{2}} g \right\| & \text{(B1)} \\ & + \left\| (K_T^2 + aI)^{-\frac{1}{2}} K^{\frac{1}{2}} g - (K^2 + aI)^{-\frac{1}{2}} K^{\frac{1}{2}} g \right\| & \text{(B2)} \end{aligned}$$

- Part (B1)

$$\begin{aligned} \left\| (K_T^2 + aI)^{-\frac{1}{2}} K_T^{\frac{1}{2}} g - (K_T^2 + aI)^{-\frac{1}{2}} K^{\frac{1}{2}} g \right\| &= \left\| (K_T^2 + aI)^{-\frac{1}{2}} (K_T^{\frac{1}{2}} - K^{\frac{1}{2}}) g \right\| \\ &\leq \left\| (K_T^2 + aI)^{-\frac{1}{2}} \right\| \left\| K_T^{\frac{1}{2}} - K^{\frac{1}{2}} \right\| \|g\| \end{aligned}$$

The first term is bounded by $1/\sqrt{a}$ as discussed in Part (A). The second term is such that: $\left\| K_T^{\frac{1}{2}} - K^{\frac{1}{2}} \right\| = \mathcal{O}_{P^*}(\frac{1}{T^\nu})$ in prob-$\mathbb{P}$ which follows from Theorem 4 and the continuity of the square-root transformation. Hence, overall, (B1) goes to zero as $T^\nu \sqrt{a} \to \infty$.

- Part (B2)

$$
\begin{aligned}
&\left\|(K_T^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g-(K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g\right\| \\
\leq&\left\|(K_T^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g-(K_T^2+aI)^{-\frac{1}{2}}K(K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g\right\| &\text{(B2.1)}\\
+&\left\|(K_T^2+aI)^{-\frac{1}{2}}K(K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g-(K_T^2+aI)^{-\frac{1}{2}}K_T(K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g\right\| &\text{(B2.2)}\\
+&\left\|(K_T^2+aI)^{-\frac{1}{2}}K_T(K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g-(K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g\right\| &\text{(B2.3)}
\end{aligned}
$$

We study each term on the right-hand side separately:

$$\text{(B2.1)} \leq \left\|(K_T^2+aI)^{-\frac{1}{2}}K\right\|\left\|(K^{-\frac{1}{2}}-K_a^{-\frac{1}{2}})g\right\|$$

The first term goes to one as $a \to 0$ and $T \to \infty$, while the second one goes to zero as $a \to 0$ as shown in Part (C).

$$\text{(B2.2)} \leq \left\|(K_T^2+aI)^{-\frac{1}{2}}\right\|\|K_T-K\|\left\|K_a^{-\frac{1}{2}}g\right\|$$

The first term is bounded by $1/\sqrt{a}$ as discussed in Part (A). From Theorem 4, the second term is such that: $\|K_T - K\| = \mathcal{O}_{P^*}(\frac{1}{T^\nu})$ in prob-$\mathbb{P}$. Finally, the third term is bounded by $a^{-1/4}\|g\|$ as shown in Part (A).

$$
\begin{aligned}
\text{(B2.3)} &= \left\|((K_T^2+aI)^{-\frac{1}{2}}K_T^{\frac{1}{2}}-K_T^{-\frac{1}{2}})K_T^{\frac{1}{2}}(K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g\right\| \\
&= \left\|(K_{T,a}^{-\frac{1}{2}}-K_T^{-\frac{1}{2}})K_T^{\frac{1}{2}}(K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g\right\| \\
&\leq \left\|K_T^{\frac{1}{2}}(K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}\right\|\left\|(K_{T,a}^{-\frac{1}{2}}-K_T^{-\frac{1}{2}})g\right\|
\end{aligned}
$$

Similar to (B2.1), the first term converges to one as $a \to 0$ and $T \to \infty$. The second term converges to 0 when $a \to 0$ for T large enough. Hence, overall, (B2) goes to 0 as $T^\nu a^{3/4} \to \infty$.

When we combine the properties of (B1) and (B2), we obtain the expected result.

• Part (C):
By definition, with λ_j and ϕ_j the eigenvalues and eigenfunctions of K, we have:

$$K^{-\frac{1}{2}}g = \sum_{j=1}^{\infty}\frac{1}{\sqrt{\lambda_j}}(g,\phi_j)\phi_j\,.$$

We also have:

$$K_a^{-\frac{1}{2}}g = \sum_{j=1}^{\infty}\frac{\sqrt{\lambda_j}}{\sqrt{\lambda_j^2+a}}(g,\phi_j)\phi_j\,,$$

since

$$K_a^{-\frac{1}{2}}g = (K^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g = [(K^2+aI)K^{-1}]^{-\frac{1}{2}}g\,,$$

and we can easily show that $(K^2 + aI)K^{-1}$ has eigenvalues $\frac{\lambda_j^2+a}{\lambda_j}$ and eigenfunctions ϕ_j. Indeed, we have:

$$\begin{aligned}
& K\phi_j = \lambda_j \phi_j \\
\Rightarrow \quad & K^2\phi_j = \lambda_j^2 \phi_j \quad \text{and} \quad K^{-1}\phi_j = (1/\lambda_j)\phi_j \\
\Rightarrow \quad & (K^2 + aI)\phi_j = (\lambda_j^2 + a)\phi_j \quad \text{and} \quad K^{-1}\phi_j = (1/\lambda_j)\phi_j \\
\Rightarrow \quad & (K^2 + aI)K^{-1}\phi_j = \frac{\lambda_j^2 + a}{\lambda_j}\phi_j
\end{aligned}$$

Thus, we have:

$$\begin{aligned}
K_a^{-\frac{1}{2}} g - K^{-\frac{1}{2}} g &= \sum_{j=1}^{\infty} \left(\frac{\sqrt{\lambda_j}}{\sqrt{\lambda_j^2 + a}} - \frac{1}{\sqrt{\lambda_j}} \right) (g, \phi_j)\phi_j \\
\Rightarrow \left\| K_a^{-\frac{1}{2}} g - K^{-\frac{1}{2}} g \right\|^2 &= \sum_{j=1}^{\infty} \left(\frac{\sqrt{\lambda_j}}{\sqrt{\lambda_j^2 + a}} - \frac{1}{\sqrt{\lambda_j}} \right)^2 (g, \phi_j)^2 \phi_j^2 \\
&= \sum_{j=1}^{\infty} \left(\frac{\sqrt{\lambda_j}}{\sqrt{\lambda_j^2 + a}} - \frac{1}{\sqrt{\lambda_j}} \right)^2 (g, \phi_j)^2 \\
&\leq \sum_{j=1}^{\infty} \frac{1}{\lambda_j} (g, \phi_j)^2 < \infty
\end{aligned}$$

since it is easy to show that

$$\left(\frac{\sqrt{\lambda_j}}{\sqrt{\lambda_j^2 + a}} - \frac{1}{\sqrt{\lambda_j}} \right)^2 \leq \frac{1}{\lambda_j} \quad \forall j \,.$$

To compute the limit of the LHS when $a \to 0$, we switch the summation and the limit and conclude that it converges to 0.

Overall, $\left\| K_{T,a}^{-\frac{1}{2}} g_T - K^{-\frac{1}{2}} g \right\| \xrightarrow{P^*} 0$ in prob-$\mathbb{P}$ when $a \to 0$ and $T^\nu a^{\frac{3}{4}} \to \infty$. The proof of part (ii) is similar to that of part (i). $\square$

We now return to the proof of Theorem 5.
The consistency of the estimator (as $T \to \infty$, $a \to 0$, and $T^\nu a^{3/4} \to \infty$) directly follows from Theorem A.1.2 applied to z_T^* and z under Assumption 12:

$$\begin{aligned}
& \left\| K_{T,a}^{-\frac{1}{2}} z_T^*(., \theta) - K^{-\frac{1}{2}} z(., \theta) \right\| \xrightarrow{P^*} 0 \quad \text{in prob-}\mathbb{P} \\
& \left\| K_{T,a}^{-1} \partial z_T^*(., \theta)/\partial\theta - K^{-1} \partial z(., \theta)/\partial\theta \right\| \xrightarrow{P^*} 0 \quad \text{in prob-}\mathbb{P} \,,
\end{aligned}$$

Following similar steps as those taken in the proof of Theorem 2, we can show that:

$$\begin{aligned} & (K_{T,a}^{-1/2} G_T(\hat{\theta}_{SMAS}^{opt}), K_{T,a}^{-1/2} z_T(\theta_0)) + (K_{T,a}^{-1/2} G_T(\hat{\theta}_{SMAS}^{opt}), K_{T,a}^{-1/2} G_T(\overline{\theta})(\hat{\theta}_{SMAS}^{opt} - \theta_0)) = 0 \\ \Leftrightarrow \quad & (K_{T,a}^{-1/2} G_T(\hat{\theta}_{SMAS}^{opt}), K_{T,a}^{-1/2} G_T(\overline{\theta}))\sqrt{T}(\hat{\theta}_{SMAS}^{opt} - \theta_0) = -(K_{T,a}^{-1} G_T(\hat{\theta}_{SMAS}^{opt}), \sqrt{T} z_T(\theta_0)) \end{aligned}$$

where $\overline{\theta}$ lies between θ_0 and $\hat{\theta}_{SMAS}^{opt}$ component by component.
We focus on the RHS term:

$$\begin{aligned} & (K_{T,a}^{-1} G_T(\hat{\theta}_{SMAS}^{opt}), \sqrt{T} z_T(\theta_0)) \\ = \quad & (K_{T,a}^{-1} G_T(\hat{\theta}_{SMAS}^{opt}) - K^{-1} G(\theta_0), \sqrt{T} z_T(\theta_0)) + (K^{-1} G(\theta_0), \sqrt{T} z_T(\theta_0)) \end{aligned}$$

Since $\sqrt{T} z_T(\theta_0) \xrightarrow{d} Z \sim \mathcal{N}(0, K)$, we have:

$$(K^{-1} G(\theta_0), \sqrt{T} z_T(\theta_0)) \xrightarrow{d} \mathcal{N}(0, (K^{-1} G(\theta_0), K^{-1} G(\theta_0)) .$$

In addition, we have:

$$\begin{aligned} & (K_{T,a}^{-1} G(\hat{\theta}_{SMAS}^{opt}) - K^{-1} G(\theta_0), \sqrt{T} z_T(\theta_0)) \\ \leq \quad & \| K_{T,a}^{-1} G_T(\hat{\theta}_{SMAS}^{opt}) - K^{-1} G(\theta_0) \| \times \| \sqrt{T} z_T(\theta_0) \| \\ = \quad & o_{P^*}(1) \quad \text{in prob-}\mathbb{P} \end{aligned}$$

since $\| \sqrt{T} z_T(\theta_0) \| = \mathcal{O}_{\mathbb{P}}(1)$ and

$$\begin{aligned} & \| K_{T,a}^{-1} G_T(\hat{\theta}_{SMAS}^{opt}) - K^{-1} G(\theta_0) \| \\ \leq \quad & \| K_{T,a}^{-1} \| \| G_T(\hat{\theta}_{SMAS}^{opt}) - G_T^*(\hat{\theta}_{SMAS}^{opt}) \| + \| K_{T,a}^{-1} G_T^*(\hat{\theta}_{SMAS}^{opt}) - K^{-1} G(\theta_0) \| \\ = \quad & o_{P^*}(1) \quad \text{in prob-}\mathbb{P} \end{aligned}$$

where the last equality follows from Theorem A.1.2 - and intermediate results in its proof such as $\| K_{T,a}^{-1} \| \leq 1/\sqrt{a}$ for T large enough - as well as

$$\begin{aligned} & \| G_T(\hat{\theta}_{SMAS}^{opt}) - G_T^*(\hat{\theta}_{SMAS}^{opt}) \| \\ \leq \quad & \| G_T(\hat{\theta}_{SMAS}^{opt}) - G(\hat{\theta}_{SMAS}^{opt}) \| + \| G(\hat{\theta}_{SMAS}^{opt}) - G_T^*(\hat{\theta}_{SMAS}^{opt}) \| \\ = \quad & \mathcal{O}_{P^*}(1/\sqrt{T}) \quad \text{in prob-}\mathbb{P} \end{aligned}$$

which holds under the uniform convergence results maintained in Assumption 12 combined with Lemma B.2 in Dovonon and Gonçalves (2017).

To complete the proof, notice that we have, for T large enough with $a \to 0$ and $T^{\nu} a^{3/2} \to \infty$:

$$\begin{aligned} \sqrt{T}(\hat{\theta}_{SMAS}^{opt} - \theta_0) \quad = \quad & -(K^{-1/2} G(\theta_0), K^{-1/2} G(\theta_0))(K^{-1} G(\theta_0), Z) + o_{\mathbb{P} \times P^*}(1) \\ \text{with} \quad & (K^{-1/2} G(\theta_0), K^{-1/2} G(\theta_0))(K^{-1} G(\theta_0), Z) \sim \mathcal{N}(0, \| K^{-1/2} G(\theta_0) \|^{-2}) . \end{aligned}$$

Finally, the optimality of $\hat{\theta}_{SMAS}^{opt}$ amounts to showing that $(V - \| K^{-1} G(\theta_0) \|^{-2})$ is positive definite - with V the asymptotic variance derived in Theorem 2. A similar result has already been shown in Carrasco and Florens (2000) at the end of the proof of Theorem 8 . ■

A.2 Algorithm 1 supplement: double-bootstrap

The double-bootstrap approach features a different way to construct the controlling operator whose unit sample paths are generated by the simulator and depend on the values of parameters. See Section 1.4.

Algorithm 4. *(double-bootstrap implementation)*

1. *Using the sample of T observations, compute the chosen impulse responses $\hat{\psi}_T(\mathcal{X}_T)$ as well as the transition matrix and the residuals $\hat{\epsilon}_T$ as explained in Appendix B.1.*
2. *For given $\theta \in \Theta$, use the simulator to generate S independent samples of T observations; compute the associated (chosen) impulse responses, $\hat{\psi}_T^s(\theta)$ with $s = 1, \cdots, S$, as well as $z_T(\theta) = \hat{\psi}_T(\mathcal{X}_T) - \sum_s \hat{\psi}_T^s(\theta)/S$.*
3. *With the same θ in Step 2, estimate the optimal operator $\hat{K}_{T,a}^{-1/2*}(\theta)$ with the double-bootstrap method (a the regularizing parameter).*
   (a) *Use the simulator (same as in Step 2) to generate $\mathcal{X}_{T,n}^*(\theta)$ and compute $\hat{\psi}_{T,n}^*(\theta)$.*
   (b) *Repeat independently N times to get $\hat{\psi}_{T,n}^*(\theta)$ with $n = 1, \cdots, N$.*
   (c) *Compute $\hat{K}_T^{-1/2*}(\theta)$ using $\hat{\psi}_{T,n}^*(\theta)$ by following the procedure described in section 1.4.*
4. *Obtain $\hat{\theta}_{SMAS}$ as the minimizer over θ of $\|\hat{K}_{T,a}^{-1/2*}(\theta) z_T(\theta)\|$.*

Appendix B

Simulation-based Methodology with Many Auxiliary Statistics: Applications to Long-run Dynamic Analysis

B.1 Computation of the (structural) impulse responses

We start by postulating a reduced-form VAR model of order p to represent the dynamics of the vector of observables $\mathcal{X}_T$ on inflation and interest rate:

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \ldots + \Phi_p x_{t-p} + \Phi_0 + u_t, \qquad u_t \sim i.i.d.(0, \Sigma)$$

Assuming that the reduced-form errors u_t are linked to the structural model innovations ϵ_t via the equation $Pu_t = \epsilon_t$ with $P\Sigma P' = I$, a Choleski decomposition can be applied to the variance-covariance matrix Σ. The impulse response of the structural shock $\epsilon_{j,t}$ on the variable $x_{i,t}$ at horizon h is defined as

$$IRF(i, j, h) = \partial x_{i,t+h} / \partial \epsilon_{j,t}$$

and given by the appropriate coefficient in the following model,

$$x_t = \Theta(L) P^{-1} P u_t \equiv \Psi(L)\epsilon_t, \qquad \epsilon_t \sim i.i.d.(0, I)$$

After estimating the above model, we obtain the impulse responses at chosen horizon h, as well as the residuals $\hat{\epsilon}_t$ and the transition matrix $\hat{\Psi}(L)$.

B.2 Implementation details of *double-bootstrap* method

Consider a sample of $T = 232$ observations of interest rate R_T and inflation π_T, the fitted model VAR(2), and a matching horizon of $H = 80$.

• Estimate $\hat{\alpha}^{opt}_{SMAS}$ from the full sample of observations.

The subsequent discussion centers on the *double-bootstrap* method, which formalizes a more general framework to establish the governing operator with the simulator instead of the reduced-form VAR model.

1. Fit a VAR(2) model with the observations and get the transition matrix A_f, the model residuals $\hat{\epsilon}_T$, and the covariance σ_f^2. Compute the structural impulse responses $\hat{\psi}_T$ with A_f and σ_f^2 as introduced in Appendix B.1.

2. For given α, simulate $N = 199$ independent samples with the simulator and compute impulse responses for each of them. Construct the optimal operator with N IRFs $\hat{K}^{-1/2}_{T,a_T}(\alpha), a_T = c^*/T^{1/3}$ (c^* the selected tuning parameter).

   (a) Generate $\mathcal{X}^*_{T,n}(\alpha)$ with the simulator and compute $\hat{\psi}^*_{T,n}(\alpha)$.

   (b) Repeat independently N times to get $\hat{\psi}^*_{T,n}(\alpha)$ with $n = 1, \cdots, N$.

   (c) Compute $\hat{K}^{-1/2*}_T(\alpha)$ using $\hat{\psi}^*_{T,n}(\alpha)$.

3. With the same α in Step 2, simulate S samples of size $T = 232$, then generate impulse responses $\hat{\psi}^s_T(\mathcal{X}^s_T(\alpha))$. Obtain the optimal estimator as the minimizer over the search grid of α,

$$\hat{\alpha}^{opt}_{SMAS} = \arg\min_{\alpha} \left\| \hat{K}^{-1/2}_{T,a_T}(\alpha) z_T(\alpha) \right\|, \quad z_T(\alpha) = \hat{\psi}_T - \frac{1}{S}\sum_{s=1}^{S} \hat{\psi}^s_T(\mathcal{X}^s_T(\alpha))$$

• Use cross-validation to determine tuning parameter c^* from the search grid C following Algorithm 2.

1. Split the sample of size $T = 232$ into two parts, a training set of the first $\tilde{T} = 155$ observations and a testing set of the rest $T - \tilde{T} = 77$ observations.

2. Given $c \in C$, traverse the grid of α and compute the regularized optimal SMAS estimator obtained with the regularized optimal operator $\hat{K}^{-1/2}_{\tilde{T},a_{\tilde{T}}}(\alpha), a_{\tilde{T}} = c/\tilde{T}^{1/3}$, using the training sample.

$$\hat{\alpha}^{opt}_{SMAS}(c) = \arg\min_{\alpha} \left\| K^{-1/2}_{\tilde{T},a_{\tilde{T}}}(\alpha) z_{tr}(\alpha) \right\|.$$

   $z_{tr}(\alpha)$ denotes the difference between the chosen impulse responses (up to horizon $H = 80$) computed from the training set and that computed from simulated samples of size $\tilde{T} = 155$ generated with α. $K^{-1/2}_{\tilde{T},a_{\tilde{T}}}(\alpha)$ is constructed the same way as in the estimation step.

3. Use the simulator with $\hat{\alpha}^{opt}_{SMAS}(c)$ to generate S independent sample of $(T - \tilde{T}) = 77$ observations; compute the impulse responses (up to horizon $H = 80$) and match them to those computed over the testing sample to get $z_{test}(\hat{\alpha}^{opt}_{SMAS}(c))$. Evaluate the associated SMAS objective function.

4. The regularization parameter (for the whole sample of size T) is $a^*_T = c^*/T^{1/3}$ where c^* is obtained by minimizing the SMAS objective function over the testing sample with respect to c.

B.3 Results of the Monte-Carlo simulation study

B.3.1 Small-scale model

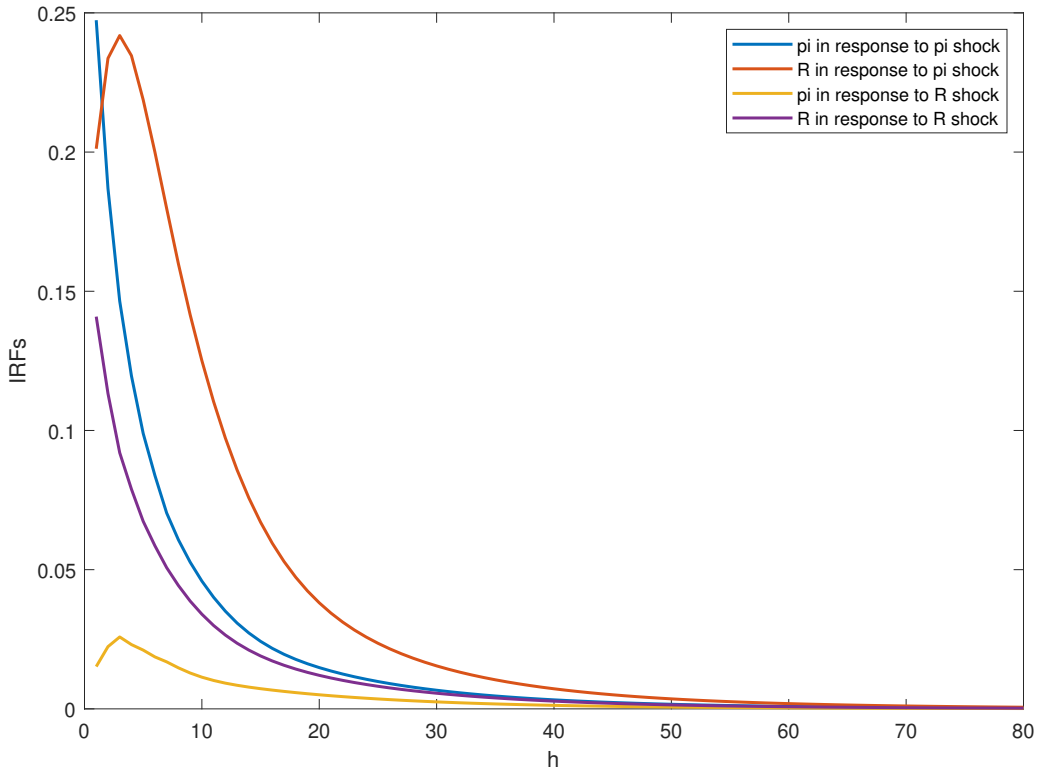

Figure B.1: Impulse responses of the small-scale DSGE model as a function of the horizon.

Matching IR at horizons 1 to 20, sample size $T = 232$						
	VAR(2)					
	SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.7412	0.6935	0.7396	0.6944	0.7311	0.7177
MAD	0.0088	0.0565	0.0104	0.0556	0.0189	0.0323
SD	0.0673	0.1666	0.0653	0.1630	0.0813	0.0829
RMSE	0.0679	0.1760	0.0661	0.1722	0.0835	0.0890
Coverage						
95%	95.8	92.0	95.3	92.3	94.2	93.1
90%	91.4	89.1	91.1	89.4	91.4	88.2
	VAR(4)					
	SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.7412	0.6920	0.7380	0.6940	0.7327	0.7084
MAD	0.0088	0.0580	0.0120	0.0560	0.0173	0.0416
SD	0.0761	0.1808	0.0716	0.1818	0.0858	0.0880
RMSE	0.0766	0.1899	0.0726	0.1902	0.0875	0.0973
Coverage						
95%	94.6	91.9	93.9	92.2	94.3	91.7
90%	91.2	89.5	91.0	90.0	92.8	88.0
	VAR(6)					
	SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.7406	0.6897	0.7418	0.6927	0.7393	0.7088
MAD	0.0094	0.0603	0.0082	0.0573	0.0107	0.0412
SD	0.0798	0.1922	0.0749	0.1860	0.0845	0.0817
RMSE	0.0804	0.2014	0.0753	0.1946	0.0852	0.0915
Coverage						
95%	94.8	91.6	94.4	91.8	93.9	90.7
90%	92.5	89.7	91.1	89.5	91.1	87.2

Table B.1: Small-scale model estimation using medium-term IRs

Note: Performance of the SMAS estimators (with and without regularization) and two estimators of Guerron-Quintana et al. (2017) for different simulations designs when matching IR over 20 periods (5 years) with a sample size $T = 232$ and VAR of orders 2, 4 and 6. We report the Monte-Carlo Mean, Mean Absolute Deviation (MAD), Standard deviation (SD), RMSE, and effective coverage probabilities of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

Matching IR at horizons 1 to 80, sample size $T = 232$						
	VAR(2)					
	SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.7401	0.6532	0.7385	0.6485	0.7163	0.6887
MAD	0.0099	0.0968	0.0115	0.1005	0.0337	0.0613
SD	0.0706	0.1860	0.0663	0.1879	0.1066	0.1041
RMSE	0.0713	0.2096	0.0673	0.2131	0.1118	0.1208
Coverage						
95%	94.8	91.1	95.1	90.5	93.4	90.5
90%	92.7	87.2	91.6	86.7	90.3	86.3
	VAR(4)					
	SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.7409	0.6450	0.7399	0.6415	0.7106	0.6565
MAD	0.0091	0.1050	0.0101	0.1085	0.0394	0.0935
SD	0.0802	0.2126	0.0721	0.2107	0.1155	0.1326
RMSE	0.0807	0.2371	0.0728	0.2370	0.1220	0.1623
Coverage						
95%	95.7	89.4	94.8	89.5	93.7	89.6
90%	92.4	86.4	91.7	85.0	91.1	83.8
	VAR(6)					
	SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.7437	0.6338	0.7422	0.6394	0.7170	0.6495
MAD	0.0063	0.1162	0.0078	0.1106	0.0330	0.1005
SD	0.0815	0.2268	0.0757	0.2226	0.1095	0.1405
RMSE	0.0817	0.2548	0.0761	0.2486	0.1144	0.1727
Coverage						
95%	95.7	88.6	95.7	88.6	93.0	87.7
90%	91.8	84.7	91.7	85.8	90.3	82.8

Table B.2: Small-scale model estimation using long-term IRs

Note: Performance of the SMAS estimators (with and without regularization) and two estimators of Guerron-Quintana et al. (2017) for different simulations designs when matching IR over 80 periods (20 years) with a sample size $T = 232$ and VAR of orders 2, 4 and 6. We report the Monte-Carlo Mean, Mean Absolute Deviation (MAD), Standard deviation (SD), RMSE, and effective coverage probabilities of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

Matching IR at short-term horizons with sample size $T = 232$												
	Panel A: Matching at horizons 1 to 2						*Panel B: Matching at horizons 1 to 8*					
	SMAS				GIK		SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.7452	0.7501	0.7448	0.7497	0.7417	0.7297	0.7423	0.7397	0.7418	0.7395	0.7216	0.7342
MAD	0.0048	10^{-4}	0.0052	0.0003	0.0083	0.0203	0.0077	0.0103	0.0082	0.0105	0.0284	0.0158
SD	0.0536	0.0578	0.0536	0.0577	0.0666	0.0584	0.0614	0.0771	0.0589	0.0786	0.0738	0.0707
RMSE	0.0538	0.0578	0.0539	0.0577	0.0671	0.0618	0.0619	0.0778	0.0595	0.0793	0.0791	0.0724
Coverage												
95%	95.3	96.2	95.1	96.0	94.7	92.8	94.5	96.1	94.8	96.3	94.1	92.2
90%	90.9	91.4	90.7	90.5	91.9	88.2	91.7	93.3	91.7	93.9	90.9	87.2

Table B.3: Small-scale model estimation using short-term IRs

Note: Performance of the SMAS estimators (with and without regularization) and two estimators of Guerron-Quintana et al. (2017) for different simulations designs when matching IR over short term horizons with sample size $T = 232$. We report the Monte-Carlo Mean, Mean Absolute Deviation (MAD), Standard deviation (SD), RMSE, and effective coverage probabilities of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

Matching IR over medium and long horizons with sample size $T = 100$												
	Panel A: Matching at horizons 1 to 20						*Panel B: Matching at horizons 1 to 80*					
	SMAS				GIK		SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.7325	0.6035	0.7293	0.6063	0.7089	0.7041	0.7323	0.4753	0.7263	0.4692	0.6853	0.6493
MAD	0.0175	0.1465	0.0207	0.1437	0.0411	0.0459	0.0177	0.2747	0.0237	0.2808	0.0647	0.1007
SD	0.1060	0.2491	0.1040	0.2485	0.1346	0.1086	0.1051	0.2522	0.1056	0.2539	0.1628	0.1425
RMSE	0.1074	0.2890	0.1060	0.2871	0.1407	0.1179	0.1065	0.3729	0.1082	0.3785	0.1752	0.1745
Coverage												
95%	94.6	87.2	94.3	86.8	92.4	90.8	93.8	75.1	93.7	75.1	92.5	88.5
90%	91.4	82.5	91.2	82.9	90.3	87.5	91.6	67.9	91.8	65.8	88.7	83.7

Table B.4: Small-scale model estimation using medium- and long-term IRs, small samples

Note: Performance of the SMAS estimators (with and without regularization) and two estimators of Guerron-Quintana et al. (2017) for different simulations designs when matching IR over medium to long horizons with a sample size $T = 100$ and VAR of order 2. We report the Monte-Carlo Mean, Mean Absolute Deviation (MAD), Standard deviation (SD), RMSE, and effective coverage probabilities of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

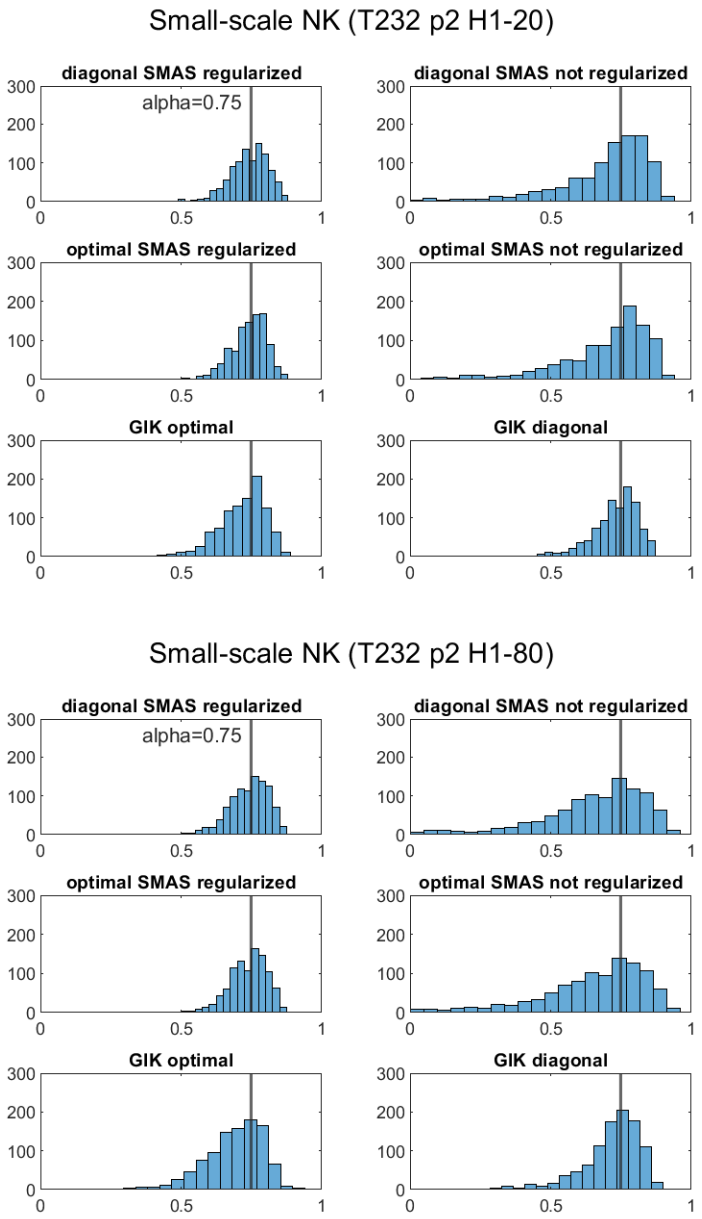

Figure B.2: Monte-Carlo distribution: estimation of Small-scale model, medium- and long-term IR

Note: Monte-Carlo distribution of the following four estimates obtained when matching IR up to $H = 20$ (top three rows) and up to $H = 80$ (bottom three rows) over 1,000 replications: regularized SMAS with diagonal operator(top left), SMAS with diagonal operator(top right), regularized SMAS with optimal operator(middle left), SMAS with optimal operator(middle right), Guerron-Quintana et al. (2017) with optimal weighting matrix (bottom left), and Guerron-Quintana et al. (2017) with diagonal weighting matrix (bottom right). The vertical line represents the true value of the parameter.

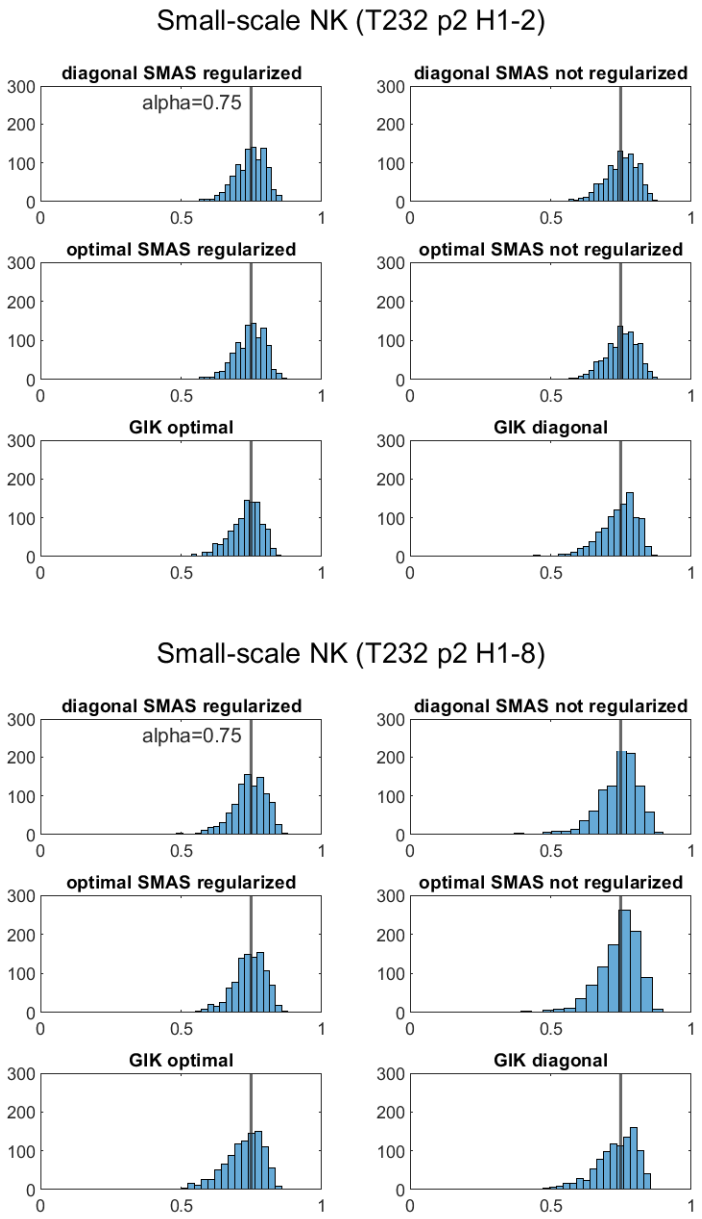

Figure B.3: Monte-Carlo distribution: estimation of Small-scale model, short-term IR

Note: Monte-Carlo distribution of the following six estimates obtained when matching IR up to $H = 2$ (top three rows) and up to $H = 8$ (bottom three rows) over 1,000 replications: regularized SMAS with diagonal operator(top left), SMAS with diagonal operator(top right), regularized SMAS with optimal operator(middle left), SMAS with optimal operator(middle right), Guerron-Quintana et al. (2017) with optimal weighting matrix (bottom left), and Guerron-Quintana et al. (2017) with diagonal weighting matrix (bottom right). The vertical line represents the true value of the parameter.

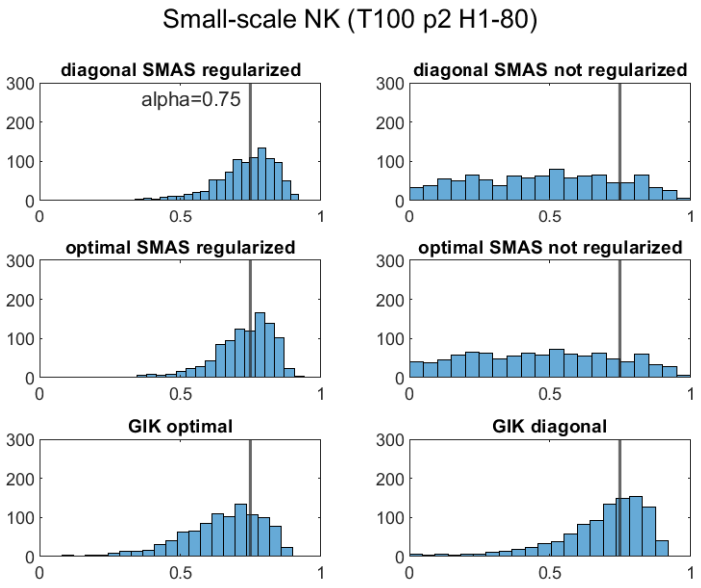

Figure B.4: Monte-Carlo distribution: estimation of Small-scale model, sample size $T = 100$

Note: Monte-Carlo distribution of the following six estimates obtained when matching IR up to $H = 80$ with a smaller sample size $T = 100$ over 1,000 replications: regularized SMAS with diagonal operator(top left), SMAS with diagonal operator(top right), regularized SMAS with optimal operator(middle left), SMAS with optimal operator(middle right), Guerron-Quintana et al. (2017) with optimal weighting matrix (bottom left), and Guerron-Quintana et al. (2017) with diagonal weighting matrix (bottom right). The vertical line represents the true value of the parameter.

Matching IR at medium to long-term horizons with sample size $T = 232$						
	SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.6139	0.4593	0.6227	0.4647	0.6422	0.6673
MAD	0.1361	0.2907	0.1273	0.2853	0.1078	0.0827
SD	0.2387	0.2904	0.2378	0.2901	0.2040	0.1264
RMSE	0.2748	0.4109	0.2697	0.4069	0.2307	0.1511
Coverage						
95%	86.3	75.6	86.7	76.3	88.6	89.6
90%	82.9	66.4	83.6	67.3	85.1	86.7

Table B.5: Small-scale model estimation, matching IR at horizons 21 to 40

Note: Performance of the SMAS estimators (with and without regularization) and two estimators of Guerron-Quintana et al. (2017) for different simulations designs when matching IR in the long-term but ignoring the short-term, $H = 21 - 40$, with sample size $T = 232$. We report the Monte-Carlo Mean, Mean Absolute Deviation (MAD), Standard deviation (SD), RMSE, and effective coverage probabilities of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

Matching IR with diagonal operator conditional on the parameter value				
	Operator obtained by double-bootstrap		Operator obtained by resampling	
	diagonal reg.	diagonal non-reg.	diagonal reg.	diagonal non-reg.
Mean	0.7338	0.6722	0.7401	0.6532
MAD	0.0162	0.0778	0.0099	0.0968
SD	0.0719	0.1690	0.0706	0.1860
RMSE	0.0737	0.1861	0.0713	0.2096
Coverage				
95%	94.5	90.6	94.8	91.1
90%	91.9	87.7	92.7	87.2
Matching IR with optimal operator conditional on the parameter value				
	Operator obtained by double-bootstrap		Operator obtained by resampling	
	diagonal reg.	diagonal non-reg.	diagonal reg.	diagonal non-reg.
Mean	0.7292	0.6303	0.7385	0.6485
MAD	0.0208	0.1198	0.0115	0.1005
SD	0.0768	0.2155	0.0663	0.1879
RMSE	0.0796	0.2466	0.0673	0.2131
Coverage				
95%	93.8	88.0	95.1	90.5
90%	90.0	85.5	91.6	86.7

Table B.6: Small-scale model estimation with operators conditional on the parameter value

Note: Performance of the SMAS estimators (with and without regularization) when matching IR in the long-term $H = 80$, with sample size $T = 232$. The sequence of data is fitted to a VAR(2) model. The diagonal operator is generated conditional on the value of the parameter of interest.

B.3.2 Medium-scale model

model variable		
W	real wage	
Y	output	
I	investment	
C	consumption	
L	hours worked	
π	inflation	
R	nominal interest rate	
parameter		parameter value
α	capital share	0.30
β	time discount factor	0.99
τ	capital accumulation	0.025
c_y	consumption-output ratio	0.6
i_y	investment-output ratio	0.22
λ_w	wage markup	0.5
ϕ_i	investment adjustment cost	6.771
σ_c	risk aversion	1.353
h	external habit formation	0.573
ξ_w	Calvo parameter wage	0.737
σ_L	inverse elasticity of labor supply	2.400
ξ_p	Calvo parameter price	0.908
ξ_e	fraction of firms able to adjust employment	0.599
γ_w	degree of wage indexation	0.763
γ_p	degree of price indexation	0.469
ψ	capital utilization cost	0.169
ϕ_y	one plus share of the fixed cost in production	1.408
r_π	Taylor rule inflation feedback	1.684
$r_{\Delta\pi}$	Taylor rule inflation change feedback	0.14
ρ	degree of interest rate smoothing	0.961
r_y	Taylor rule output level feedback	0.099
$r_{\Delta y}$	Taylor rule output growth feedback	0.159
ρ_a	persistence productivity shock	0.823
ρ_b	persistence risk premium shock	0.855
ρ_g	persistence spending shock	0.949
ρ_l	persistence labor shock	0.889
ρ_i	persistence investment shock	0.927
ρ_π	persistence price markup shock	0.924

Table B.7: Parameter values in the medium-scale model.

Panel A: Matching IR over medium to long horizons on all 7 indices						
	Horizons 1 to 20		Horizons 1 to 80		Horizons 21 to 80	
	regularized	non-reg.	regularized	non-reg.	regularized	non-reg.
mean	0.4862	0.5133	0.4740	0.5327	0.4720	0.5606
MAD	0.0172	0.0443	0.0050	0.0637	0.0030	0.0916
SD	0.2714	0.2794	0.2608	0.2850	0.2638	0.2889
RMSE	0.2719	0.2829	0.2608	0.2920	0.2638	0.3031
Coverage						
95%	100.0	100.0	100.0	100.0	100.0	100.0
90%	95.8	92.6	95.7	91.4	97.2	94.6
Panel B: Matching IR over medium to long horizons on 4 indices						
	Horizons 1 to 20		Horizons 1 to 80		Horizons 21 to 80	
	regularized	non-reg.	regularized	non-reg.	regularized	non-reg.
mean	0.4845	0.4915	0.4802	0.5514	0.4964	0.4330
MAD	0.0155	0.0225	0.0112	0.0824	0.0274	0.0360
SD	0.2718	0.2873	0.2674	0.2929	0.2730	0.2575
RMSE	0.2722	0.2882	0.2676	0.3043	0.2744	0.2600
Coverage						
95%	100.0	100.0	100.0	100.0	100.0	100.0
90%	96.6	94.2	94.9	95.4	95.6	94.6

Table B.8: Estimation of the degree of price indexation, medium- to long-run

Note: Estimation of the degree of price indexation $\gamma_p = 0.469$. Performance of SMAS estimators (with and without regularization) when matching impulse responses over medium to long horizons in all indices (Panel A), or in a subset of 4 indices, consumption, labor force, inflation, interest rate (Panel B). We consider dynamic responses obtained up to five years after the shock (20 periods), up to twenty years (80 periods), or between five and twenty years (60 periods). The sample size is $T = 236$; we consider VAR(4), $N = 1,000$ replications.

Panel A: Matching IR over long horizons on all 7 indices						
	VAR(2) model		VAR(4) model		VAR(8) model	
	regularized	non-reg.	regularized	non-reg.	regularized	non-reg.
mean	0.4796	0.4901	0.4740	0.5327	0.4613	0.5208
MAD	0.0106	0.0211	0.0050	0.0637	0.0077	0.0518
SD	0.2618	0.2929	0.2608	0.2850	0.2643	0.2799
RMSE	0.2620	0.2937	0.2608	0.2920	0.2644	0.2847
Coverage						
95%	100.0	100.0	100.0	100.0	100.0	100.0
90%	93.7	92.0	95.7	91.4	94.4	94.1
Panel B: Matching IR over long horizons on 4 indices						
	VAR(2) model		VAR(4) model		VAR(8) model	
	regularized	non-reg.	regularized	non-reg.	regularized	non-reg.
mean	0.4773	0.4458	0.4802	0.5514	0.4847	0.5358
MAD	0.0083	0.0232	0.0112	0.0824	0.0157	0.0668
SD	0.2595	0.2298	0.2674	0.2929	0.2651	0.2807
RMSE	0.2596	0.2310	0.2676	0.3043	0.2656	0.2885
Coverage						
95%	100.0	99.0	100.0	100.0	100.0	100.0
90%	95.9	91.6	94.9	95.4	93.1	92.0

Table B.9: Estimation of the degree of price indexation, VAR order $p = 2, 4, 8$

Note: Estimation of the degree of price indexation $\gamma_p = 0.469$. Performance of SMAS estimators (with and without regularization) when matching impulse responses up to twenty years after the shock (80 periods) in all indices (Panel A), or in a subset of 4 indices, consumption, labor force, inflation, interest rate (Panel B). We consider different reduced-form models, VAR(2), VAR(4), and VAR(8). The sample size is $T = 236$ with $N = 1,000$ replications.

Matching IR over medium to long horizons on all 7 indices								
	Horizons 1 to 20				Horizons 1 to 80			
	$T = 236$		$T = 944$		$T = 236$		$T = 944$	
	regularized	non-reg.	regularized	non-reg.	regularized	non-reg.	regularized	non-reg.
mean	0.7954	0.5317	0.8399	0.4728	0.7884	0.5119	0.8445	0.5179
MAD	0.1126	0.3763	0.0681	0.4352	0.1196	0.3961	0.0635	0.3901
SD	0.2103	0.3046	0.1901	0.3359	0.1922	0.3325	0.1781	0.3354
RMSE	0.2385	0.4841	0.2019	0.5498	0.2264	0.5172	0.1891	0.5145
Coverage								
95%	91.3	60.3	92.9	56.3	92.2	59.3	93.6	63.9
90%	89.9	55.7	92.6	49.7	91.4	55.9	92.9	58.4

Table B.10: Estimation of the Calvo parameter.

Note: Estimation of the Calvo parameter $\zeta_p = 0.908$. Performance of SMAS estimators (with and without regularization) when matching impulse responses over medium to long horizons in all indices. We consider dynamic responses obtained up to five years after the shock (20 periods), or up to twenty years (80 periods) with sample sizes $T = 236$, or $T = 944$. We fit VAR(4) and consider $N = 1,000$ replications.

Matching IR over medium to long horizons on all 7 indices with $T = 232$								
	Horizons 1 to 20				Horizons 1 to 80			
	Regularized		Non-regularized		Regularized		Non-regularized	
	γ_p	ζ_p	γ_p	ζ_p	γ_p	ζ_p	γ_p	ζ_p
mean	0.4674	0.7888	0.4130	0.4334	0.4722	0.7915	0.4400	0.5670
MAD	0.0016	0.1192	0.0559	0.4746	0.0032	0.1165	0.0290	0.3410
SD	0.2779	0.2279	0.2608	0.3551	0.2635	0.1844	0.2752	0.3308
RMSE	0.2779	0.2571	0.2668	0.5927	0.2635	0.2181	0.2767	0.4751
Coverage								
95%	100	90.0	100	48.8	100	92.7	100	68.5
90%	95.8	89.4	94.9	46.8	95.9	91.7	95.2	67.0
Joint Coverage								
95%	90.5		84.5		93.2		75.8	
90%	89.6		49.1		91.5		67.4	

Table B.11: Joint estimation of the degree of price indexation and the Calvo parameter.

Note: Joint estimation of the degree of price indexation $\gamma_p = 0.469$ and the Calvo parameter $\zeta_p = 0.908$. Performance of SMAS estimators (with and without regularization) when matching impulse responses over medium to long horizons in all indices. We consider dynamic responses obtained up to five years after the shock (20 periods), or up to twenty years (80 periods) with sample sizes $T = 236$. We fit VAR(4) and consider $N = 1,000$ replications.

B.3.3 Baseline stylized DSGE model

We consider the baseline stylized DSGE model from Fernandez-Villaverde et al. (2016) as adapted from DelNegro and Schorfheide (2008). The log-linearized equilibrium conditions of the model for output, X_t, labor share, lsh_t, inflation, π_t and interest rate, R_t, are given by:

$$\hat{x}_t = E_t[\hat{x}_{t+1}] - (\hat{R}_t - E_t[\hat{\pi}_{t+1}]) + E_t[z_{t+1}] \qquad \widehat{lsh}_t = \hat{x}_t + \phi_t \,,$$
$$\hat{\pi}_t = \beta E_t[\hat{\pi}_{t+1}] + \frac{(1-\zeta_p\beta)(1-\zeta_p)}{\zeta_p}(\widehat{lsh}_t + \lambda_t) \qquad \hat{R}_t = \frac{1}{\beta}\hat{\pi}_t + \sigma_R \epsilon_{R,t}.$$

where the log deviation of a variable w_t from its steady-state is denoted by $\hat{w}_t$; β is the stochastic discount rate and ζ_p is the Calvo parameter (or probability with which a given firm is unable to re-optimize its price). Four exogenous shocks influence the dynamics of the variables: a technology shock, z_t, a price markup shock, λ_t, a shock that affects the preference for leisure, ϕ_t, and a monetary policy shock, $\epsilon_{R,t}$. Except for the monetary policy shock, which is assumed to be independently and identically normally distributed with mean zero and variance one, the remaining shocks are assumed to follow autoregressive processes. Thus, for each shock $i = z, \lambda, \phi$, the autoregression coefficient is ρ_i and the standard deviation is σ_i. Overall, the unknown structural parameters of the model are $[\zeta_p, \beta, \gamma, \lambda, \pi^*, \rho_\phi, \rho_\lambda, \rho_z, \sigma_\phi, \sigma_\lambda, \sigma_z, \sigma_R]'$, where γ is the growth rate of technology, λ is the steady-state markup charged by the intermediate goods producers, and π^* is the steady-state inflation rate. The steady-states for the interest rate and for the labor share can be obtained from the expressions $\bar{R} = \pi^*\gamma/\beta$, and, $\bar{lsh} = 1/(1+\lambda)$, respectively.

This baseline model is designed to have a state-space representation which is used to obtain the associated IRs analytically. Let γ_t and s_t denote the vector of observables and state variables, respectively, with $\gamma_t = M_\gamma'[\log(X_t/X_{t-1}), \log lsh_t, \log \pi_t, \log R_t]'$ - with M_γ' a selection matrix - and $s_t = [\phi_t, \lambda_t, z_t, \epsilon_{R,t}, \hat{x}_{t-1}]'$. Then, we have:

$$\begin{aligned} \gamma_t &= \Psi_0(\theta) + \Psi_1(\theta) s_t \\ s_t &= \Phi_1(\theta) s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t \,, \end{aligned}$$

with

$$\Psi_0(\theta) = M_\gamma' \begin{bmatrix} \log\gamma \\ \log(lsh) \\ \log\pi^* \\ \log(\pi^*\gamma/\beta) \end{bmatrix}, x_\phi = \frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\phi}, x_\lambda = \frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\lambda}, x_z = \frac{\rho_z\psi_p}{1-\psi_p\rho_z}, x_{\epsilon_R} = -\psi_p\sigma_R$$

$$\Psi_1(\theta) = M_\gamma' \begin{bmatrix} x_\phi & x_\lambda & x_z+1 & x_{\epsilon_R} & -1 \\ 1+(1+v)x_\phi & (1+v)x_\lambda & (1+v)x_z & (1+v)x_{\epsilon_R} & 0 \\ \frac{\kappa_p}{1-\beta\rho_\phi}(1+(1+v)x_\phi) & \frac{\kappa_p}{1-\beta\rho_\lambda}(1+(1+v)x_\lambda) & \frac{\kappa_p}{1-\beta\rho_z}(1+v)x_z & \kappa_p(1+v)x_{\epsilon_R} & 0 \\ \frac{\kappa_p/\beta}{1-\beta\rho_\phi}(1+(1+v)x_\phi) & \frac{\kappa_p/\beta}{1-\beta\rho_\lambda}(1+(1+v)x_\lambda) & \frac{\kappa_p/\beta}{1-\beta\rho_z}(1+v)x_z & \kappa_p(1+v)x_{\epsilon_R}/\beta+\sigma_R & 0 \end{bmatrix}$$

$$\Phi_1(\theta) = \begin{bmatrix} \rho_\phi & 0 & 0 & 0 & 0 \\ 0 & \rho_\lambda & 0 & 0 & 0 \\ 0 & 0 & \rho_z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ x_\phi & x_\lambda & x_z & x_{\epsilon_R} & 0 \end{bmatrix}, \Phi_\epsilon(\theta) = \begin{bmatrix} \sigma_\phi & 0 & 0 & 0 \\ 0 & \sigma_\lambda & 0 & 0 \\ 0 & 0 & \sigma_z & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

From the above state-space representation, the IR function to shock j at horizon h can be written as:

$$\psi_0(.,j,h) = \Psi_1 \Phi_1^h [\Phi_\epsilon]_{.j}$$

where $[A]_{.j}$ is the $j-$th column of a matrix A.

	Parameter	**Value**
β	stochastic discount rate	0.98
γ	growth rate of technology	1.005
λ	steady-state intermediate goods markup	0.15
π^*	steady-state inflation rate	1.005
ρ_z	autoregression parameter of the technology shock	0.13
ρ_λ	autoregression parameter of the price markup shock	0.88
ρ_ϕ	autoregression parameter of the shock that affects the preference for leisure	0.30
σ_z	standard deviation of the technology shock	1.50
σ_λ	standard deviation of the price markup shock	0.50
σ_ϕ	standard deviation of the shock that affects the preference for leisure	3.00
σ_R	standard deviation of the monetary policy shock	1.00

Table B.12: Parameter values in the baseline stylized model.

	Matching horizons H = 1-20				*Matching horizons H = 1-80*			
	T=200		T=400		T=200		T=400	
	SMAS	infeasible SMAS	SMAS	infeasible SMAS	SMAS	infeasible SMAS	SMAS	infeasible SMAS
Mean	0.6132	0.6317	0.6319	0.6378	0.6080	0.6212	0.6333	0.6353
MAD	0.0368	0.0183	0.0181	0.0122	0.0420	0.0288	0.0167	0.0147
SD	0.1854	0.1725	0.1740	0.1307	0.1797	0.1653	0.1768	0.1268
RMSE	0.0357	0.0301	0.0306	0.0172	0.0340	0.0282	0.0315	0.0163
Coverage								
95%	93.5	94.1	95.0	93.2	93.2	90.2	95.5	93.4
90%	90.8	86.5	92.4	89.9	89.6	87.2	91.8	89.8

Table B.13: Estimation of the Calvo parameter in the baseline stylized model.

Note: Estimation of the Calvo parameter $\zeta_p = 0.65$ in the baseline stylized model. Performance of the feasible and infeasible SMAS estimators with regularization when matching impulse responses over medium to long horizons. We consider dynamic responses obtained up to five years after the shock (20 periods), or up to twenty years (80 periods) with sample sizes $T = 200$ or 400. We fit VAR(4) and consider $N = 1,000$ replications.

Appendix C

Simulation-based Methodology with Many Auxiliary Statistics: Hypothesis Testing with Many Moments

C.1 Tables

Standard AR(1)	$\rho = 0.1$				$\rho = 0.9$			
	T=100		T=200		T=100		T=200	
H	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B
4	0.32	0.18	0.24	0.16	1.14	0.18	1.62	0.26
12	1.30	1.46	0.88	0.94	2.28	1.12	2.40	1.12
20	3.52	3.68	4.14	5.06	2.68	3.96	2.60	4.82

Table C.1: Empirical size of tests for AR(1) model with no switch

Note: We report the empirical rejection rates under the null in Section 3.3.1. The DGP is AR(1) and has no Markov-switching component. The nominal level of significance is 5%. Type A test uses the moments of y's and Type B test uses the moments of residuals.

Panel A: Mean-variance switching autoregressive model $\rho = 0.1, \Delta\mu = 2, \Delta\sigma = 0$												
	$(p_{11}, p_{22}) = (0.9, 0.9)$				$(p_{11}, p_{22}) = (0.9, 0.5)$				$(p_{11}, p_{22}) = (0.9, 0.1)$			
	T=100		T=200		T=100		T=200		T=100		T=200	
Number of moments	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B
4	1.10	0.32	1.78	0.58	4.84	2.46	20.34	11.32	3.60	3.36	13.54	13.52
12	0.34	1.70	0.22	1.40	1.10	1.96	1.56	2.38	2.32	2.64	2.12	1.90
20	1.08	3.42	1.44	3.24	5.64	5.96	5.94	6.48	9.44	8.58	11.54	11.04
Panel B: Mean-variance switching autoregressive model $\rho = 0.1, \Delta\mu = 0, \Delta\sigma = 1$												
	$(p_{11}, p_{22}) = (0.9, 0.9)$				$(p_{11}, p_{22}) = (0.9, 0.5)$				$(p_{11}, p_{22}) = (0.9, 0.1)$			
	T=100		T=200		T=100		T=200		T=100		T=200	
Number of moments	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B
4	9.02	8.28	18.30	16.60	8.96	7.40	17.34	15.76	5.30	4.74	10.38	10.24
12	11.28	10.08	15.78	14.26	9.64	7.34	13.32	8.56	5.74	4.60	7.90	5.64
20	22.08	21.54	30.86	30.72	38.42	37.64	56.34	55.88	34.74	34.24	51.18	50.84
Panel C: Mean-variance switching autoregressive model $\rho = 0.1, \Delta\mu = 2, \Delta\sigma = 2$												
	$(p_{11}, p_{22}) = (0.9, 0.9)$				$(p_{11}, p_{22}) = (0.9, 0.5)$				$(p_{11}, p_{22}) = (0.9, 0.1)$			
	T=100		T=200		T=100		T=200		T=100		T=200	
Number of moments	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B
4	45.90	38.44	80.98	74.74	73.56	67.06	97.04	95.90	57.94	56.56	88.20	88.08
12	29.42	29.58	38.80	39.96	60.86	55.22	75.12	70.58	52.46	50.76	66.02	62.16
20	31.32	32.74	37.66	41.34	83.76	81.38	96.48	96.40	87.56	86.98	98.12	98.14

Table C.2: Empirical power of tests for Markov-Switching AR(1) model with $\rho = 0.1$

Note: We report the empirical power for the Monte Carlo tests in Section 3.3.1. Results from tests using the first 4, 12, and 20 moments of observation are presented. The DGP is AR(1). The switch in mean and variance follows a first-order Markov process. Autoregressive coefficient is 0.1. The nominal level of significance is 5%. Type A test uses the moments of y's and Type B test uses the moments of residuals.

Panel A: Mean-variance switching autoregressive model $\rho = 0.9, \Delta\mu = 2, \Delta\sigma = 0$												
	$(p_{11}, p_{22}) = (0.9, 0.9)$				$(p_{11}, p_{22}) = (0.9, 0.5)$				$(p_{11}, p_{22}) = (0.9, 0.1)$			
	T=100		T=200		T=100		T=200		T=100		T=200	
Number of moments	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B
$\Delta\mu = 2, \Delta\sigma = 0$												
4	1.60	0.62	1.58	1.48	1.40	0.78	1.96	1.40	1.16	0.94	1.82	2.28
12	1.92	1.66	2.00	0.96	1.98	1.14	2.90	0.94	1.78	1.38	2.82	1.38
20	2.20	10.76	1.92	14.18	2.30	7.44	2.86	9.02	2.56	8.10	3.32	7.86
Panel B: Mean-variance switching autoregressive model $\rho = 0.9, \Delta\mu = 0, \Delta\sigma = 1$												
	$(p_{11}, p_{22}) = (0.9, 0.9)$				$(p_{11}, p_{22}) = (0.9, 0.5)$				$(p_{11}, p_{22}) = (0.9, 0.1)$			
	T=100		T=200		T=100		T=200		T=100		T=200	
Number of moments	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B
4	4.16	7.96	5.44	17.28	3.30	8.00	3.26	17.94	2.02	4.42	2.22	10.22
12	6.90	10.78	9.44	14.68	5.56	6.56	7.38	9.22	3.92	4.82	4.80	6.84
20	7.58	23.28	9.72	30.38	6.14	39.04	8.40	57.30	3.84	34.04	5.32	51.98
Panel C: Mean-variance switching autoregressive model $\rho = 0.9, \Delta\mu = 2, \Delta\sigma = 2$												
	$(p_{11}, p_{22}) = (0.9, 0.9)$				$(p_{11}, p_{22}) = (0.9, 0.5)$				$(p_{11}, p_{22}) = (0.9, 0.1)$			
	T=100		T=200		T=100		T=200		T=100		T=200	
Number of moments	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B	Type A	Type B
4	6.50	30.06	10.56	61.50	5.64	53.50	9.18	85.20	3.80	43.20	4.86	75.12
12	11.98	32.26	16.12	43.14	17.00	55.92	24.10	70.04	13.24	49.76	17.86	62.56
20	12.66	35.70	16.22	45.98	18.52	77.42	24.92	92.22	14.52	80.84	19.16	94.98

Table C.3: Empirical power of tests for Markov-Switching AR(1) model with $\rho = 0.9$

Note: We report the empirical power for the Monte Carlo tests in Section 3.3.1. Results from tests using the first 4, 12, and 20 moments of observation are presented. The DGP is AR(1). The switch in mean and variance follows a first-order Markov process. Autoregressive coefficient is 0.9. The nominal level of significance is 5%. Type A test uses the moments of y's and Type B test uses the moments of residuals.

Model 1 (no switch) $y_t = -0.2 + 1.2y_{t-1} - 0.3y_{t-2} + \epsilon_t$, $\epsilon_t \sim \mathcal{N}(0,1)$				
	T=500		T=1000	
Number of moments	Type A	Type B	Type A	Type B
4	0.02	0	0	0
12	2.96	3.92	3.8	4.56
20	3.24	4.34	3.98	4.76
Model 2 (2-state AR coefficient switch) State 1: $y_t = -0.2 + 1.2y_{t-1} - 0.3y_{t-2} + \epsilon_t$ State 2: $y_t = 0.2 + 0.4y_{t-1} - 0.3y_{t-2} + \epsilon_t$ $\epsilon_t \sim \mathcal{N}(0,1)$ transition matrix: $(p_{11}, p_{22}) = (0.83, 0.75)$				
	T=500		T=1000	
Number of moments	Type A	Type B	Type A	Type B
4	8.12	0.18	50.66	6.74
12	14.46	54.8	28.78	91.22
20	9.36	50.86	19.72	86.42

Table C.4: Tests of Markov-switching model with changing slope coefficients

Note: We report the Monte Carlo test results for the model with a switching component in level and the first order autoregressive coefficient in Section 3.3.2. Results from tests using the first 4, 12, and 20 moments of observation are presented. The DGP is AR(2) and the sample sizes are $T = 500$ or $T = 1000$ respectively. The upper panel shows the empirical size of the tests. The lower panel shows the empirical power of the tests. The switch in mean and variance follows a first-order Markov process. The nominal level of significance is 5%. Type A test uses the moments of y's and Type B test uses the moments of residuals.

Panel A: USA real GNP growth 1952Q2 - 1984Q4		
Number of moments	Type A	Type B
4	0.5119	0.3534
12	0.2149	0.4200
20	0.6781	0.8145
Panel B: USA real GNP growth 1952Q2 - 2010Q4		
Number of moments	Type A	Type B
4	0.0985	0.1850
12	0.0971	0.2296
20	0.0245	0.4078

Table C.5: Test results: US GNP growth

Note: We report the empirical p-value when the output growth rate data is tested against a stationary linear AR(4) model with no regime switch. Panel A presents the p-values of tests of the US GNP data from 1952 to 1984 using different numbers of moments, and Panel B features an extended sample period from 1952 to 2010. The empirical p-value is calculated as the average p-value out of 1000 repeated tests. For the output growth rate between 1952Q2 and 2010Q4, the hypothesis that the observations follow a linear AR(4) model is rejected at $\alpha = 0.05$ with our simulation-based MC test using $H = 20$.

C.2 eigenvalues and eigenvectors of K_T

K is defined as the asymptotic covariance operator of the stacked moments. Let K_T denote its sample counterpart with a degenerate kernel. For assumptions needed about the operator $\mathcal{B}$ and $\mathcal{B}_T$, and their properties, see Antoine and Sun(2022).

To compute the eigenfunction $\phi_j^{(T)}$ and the associated eigenvalue $\lambda_j^{(T)}$, we simulate another set of S samples of size T. This batch of simulations is not generated under the null hypothesis as for data used to compute $\widetilde{\underline{\xi}}_{null}$ and $\widetilde{\underline{\xi}}^n_{null}$. Instead, the S samples $X_T^{*,s}(\hat{\theta}_T^0)$, $s = 1, \ldots, S$ are generated by residual bootstrapping with simulator $g(\cdot, \hat{\theta}_T^0, \hat{e}_T)$, where $\hat{e}_T$ is the vector of sample residuals.

Therefore, to compute $\phi_j^{(T)}$ and $\lambda_j^{(T)}$, $j = 1, \ldots, S$, we need to

1. Get S samples $X_T^{*,s}(\hat{\theta}_T^0)$, $s = 1, \ldots, S$;
2. Compute the values of H moments $\underline{\xi}_{T,h}^{*,s}$, $h = 1, \ldots, H$ and their average $\bar{\underline{\xi}}_T^*$
3. Find the eigenvalues $\mu_j^{(T)}$ and the associated eigenvectors $\underline{\beta}_j = [\beta_j^1 \ \beta_j^2 \ \ldots \ \beta_j^S]'$ ($j = 1, \ldots, S$) of the matrix (S, S)-matrix C with (s, s') element;

$$c_{ss'} = \frac{1}{S} \sum_{h=1}^{H} \left(\underline{\xi}_{T,h}^{*,s} - \bar{\underline{\xi}}_{T,h}^* \right) \left(\underline{\xi}_{T,h}^{*,s'} - \bar{\underline{\xi}}_{T,h}^* \right);$$

4. Get $\lambda_j^{(T)} = \mu_j^{(T)}$ for $j = 1, \cdots, S$;
5. Compute

$$\phi_j^{(T)}(h) = \frac{1}{S} \left(\underline{\xi}_h^{*(T)} \right)' \underline{\beta}_j \qquad \text{with} \qquad \underline{\xi}_h^{*(T)} \equiv \begin{bmatrix} \underline{\xi}_{T,h}^{*,1} - \bar{\underline{\xi}}_{T,h}^* \\ \underline{\xi}_{T,h}^{*,2} - \bar{\underline{\xi}}_{T,h}^* \\ \vdots \\ \underline{\xi}_{T,h}^{*,S} - \bar{\underline{\xi}}_{T,h}^* \end{bmatrix}$$

, for $j = 1, \ldots, S$