Simulation-based estimation with many auxiliary statistics applied to long-run dynamic analysis*

Bertille Antoine† and Wenqian Sun
Simon Fraser University

February 28, 2024

Abstract

The existing asymptotic theory for estimators obtained by simulated minimum distance does not cover situations in which the number of components of the auxiliary statistics (or number of matched "moments") is large - typically larger than the sample size. We establish the consistency of the simulated minimum distance estimator in this situation and derive its asymptotic distribution.

Our 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. As a result, we are able to exploit, among other things, long-run information. We illustrate the implementation of the proposed method through Monte-Carlo simulation experiments based on small- and medium-scale New Keynesian models. These examples highlight how to conveniently exploit valuable information from matching a large number of impulse responses including at long-run horizons.

Keywords: many moments, regularization, bootstrap, impulse response matching.

JEL Classification: C32, C52, E30, E50.

*We would like to thank M. Carrasco, S. Gonçalves, A. Guay, L. Khalaf, D. Kim, E. Renault, the Editor (S. Ng), an Associate Editor, two Referees, and seminar and conference participants for helpful discussions. All errors remain our own.

†Corresponding author: bertille_antoine@sfu.ca, Dept of Economics, 8888 Univ. Dr., V5A 1S6 Burnaby, BC, Canada. Antoine acknowledges financial support from Social Sciences and Humanities Research Council of Canada (SSHRC).

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 paper is to extend the above-mentioned SAS or I-I estimation procedures when a large number of auxiliary statistics is available to estimate the finite dimensional parameter of interest, so-called simulation with many auxiliary statistics or SMAS. We 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. Even if, in practice, the number of auxiliary statistics is finite, our theory, developed for an infinite number of auxiliary statistics, remains relevant. Indeed, in many applications[1], the number of auxiliary statistics one would like to include is *large* - sometimes larger than the sample size - leading to invertibility and computational issues that can be dealt with by the above-mentioned regularization approach.

Our paper contributes to the literature on minimum distance estimation of a finite-dimensional vector of parameters when a large number of moments is available. We build on Carrasco and Florens (2000) who extend the generalized method of moments procedure to the case of a continuum of moment conditions.[2] 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 bootstrap procedure 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.

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. 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 long ones) after a monetary shock: these impulse re-

[1] In the context of impulse response matching - as considered in our numerical illustrations - Fève et al. (2009) highlight that e.g. the number of included impulse responses is much larger than the number of estimated parameters, leading to collinearity and invertibility issues.

[2] See also Guay and Pelgrin (2023) for results in the frequency domain.

sponses 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), Guerron-Quintana et al. (2017), or Sokullu (2020). More specifically, Guerron-Quintana et al. (2017) consider VAR-based impulse response matching estimation of the parameters of dynamic stochastic equilibrium (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. Sokullu (2020) considers the problem of impulse response (IR) matching with many statistics in DSGE models when these IRs are available in closed-form, and also relies on Tikhonov regularization to obtain the operator. Our framework extends hers to the case where underlying macro models are too complex to obtain closed-form solutions. In addition, we propose a bootstrap procedure to compute the optimal covariance operator.

Our paper is organized as follows. In section 2, we introduce our framework and estimator. In sections 3 and 4, we derive its asymptotic properties and design a simulation-based approach to compute the optimal operator in practice. In section 5, we illustrate its implementation through Monte-Carlo simulation experiments based on New Keynesian models where we match a large number of impulse responses, including long-run ones. Section 6 concludes. Proofs, tables of results and graphs are collected in the Appendix, as well as in a Supplementary Appendix.

2 Framework and notations

We first introduce 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[3]:

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

for some (H, H) weighting matrix W_T 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[4] $\mathbb{N}$ 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 square-integrable real-valued functions through a linear operator denoted B. Our estimator based on simulation with many auxiliary statistics (or SMAS) is defined as

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

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

[3]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).

[4]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).

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 consider instead infinitely many impulse responses 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 5, we show that long-run information is valuable and how it can easily be used to estimate the structural parameters of interest.

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.

3.1 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 independently from $\mathcal{X}_T$.

Several comments are worth mentioning.

(i) Strictly speaking, our simulator is not assumed to coincide exactly with the true model, and we rather have in mind a parametric model that is not "too far" from

the true DGP: for example, our simulator could be obtained as a reduced form model of the structural model of interest. Our framework does allow for "small" (local) deviations between the simulator and the true DGP as formalized by our regularity assumptions: see e.g. Assumption 7 and associated discussions.

(ii) To simplify our exposition, we maintain throughout that $\mathcal{X}_T^s(\theta)$ are drawn independently from the observations $\mathcal{X}_T$. This assumption is satisfied in the models considered in our simulation study in section 5. It would, however, be violated if the simulated data were drawn conditional on some observed variables: for example, when the observed data contain some endogenous variables y_t and some exogenous variables w_t, so that $x_t = (y_t, w_t)$, the simulator usually delivers $(y_1^s, \cdots, y_T^s)$ for given θ, $(w_1, \cdots, w_T)$ and some initial conditions $\underline{y}_0$. See e.g. Gouriéroux and Monfort (1996) for extensive discussions.

(iii) 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 a nuisance parameter. 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 deliver identification-robust SAS or I-I based inference on θ and allow θ_2 to be *approximately* calibrated in the sense that it may not be correctly calibrated.

Our estimation strategy consists in *matching* the chosen auxiliary statistic computed on observed data with that computed on simulated 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*

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

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

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 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 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, $z(h, \theta) = 0$, $\forall h \in \mathbb{N}$, 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 H). Assumption 2(ii) is an identification assumption[7]: 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)*

[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. In our applications of interest, $z(.)$ is indeed square-integrable (see e.g. Appendix B.1.1), and we proceed without introducing such a probability measure.

[7] See Antoine et al. (2023) for identification-robust simulation-based matching inference with a finite number of auxiliary statistics.

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

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

(iii) *$Bz(.,\theta) = 0 \Leftrightarrow \theta = \theta_0$.*

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. Notice that B is not assumed to be non-singular, and, indeed, we allow its null space $N(B)$ to contain (other) non-zero elements, as $N(B) = \{0\}$ is rarely satisfied; see e.g. Remark 1 in Carrasco and Florens (2000) for additional discussions and examples. When B is assumed non-singular, its null space is equal to $\{0\}$, and Assumption 3 simplifies as (iii) becomes redundant: B non-singular, together with Assumption 2(ii), ensures that the equation $Bz(,\theta) = 0$ has a unique solution. In the finite dimensional case, such a condition reduces to a full rank assumption on the weighting matrix $B'B$.

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

(i) *Let B_T be a sequence of random bounded linear operators, $B_T : \mathcal{D}(B_T) \subset H \to H$. 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 $Q_T(\theta) \equiv \left\|B_T z_T(.,\theta)\right\|$ is a continuous function of θ.*

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

Assumption 4(i) guarantees the continuity of the objective function in θ whereas (ii) implies that the empirical weighted distance converges to its population value.

Theorem 1. *(Consistency of the SMAS estimator)*
Under Assumptions 1 to 4, the SMAS estimator $\hat{\theta}_{SMAS}$ is consistent for θ_0, that is,

$$\hat{\theta}_{SMAS} \xrightarrow{\mathbb{P}} \theta_0 \qquad with \qquad \hat{\theta}_{SMAS} \equiv \arg\min_{\theta\in\Theta} Q_T(\theta)\,. \tag{1}$$

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 $\big(BG(.,\theta), BG(.,\theta)\big) = \big\|BG(.,\theta)\big\|^2$ 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 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 5 guarantees first-order identification (see e.g. Sargan (1983)), while Assumption 6 imposes further regularity conditions needed to derive the asymptotic properties of our estimator.

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 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 functional convergence of the H-valued random element $z_T(.,\theta_0)$ maintained in Assumption 7(ii) include e.g. mixing conditions: see Chen and White (1998). As previously mentioned, our framework allows for "small" deviations between the simulator and the true DGP which can be understood as a certain form of local misspecification. For example, the estimated auxiliary statistics obtained with observed and simulated data do not have to converge to the (same) limit, say ψ_0, as long as their difference z_T converges to zero sufficiently fast to ensure Assumption 7(ii) holds. These local deviations do not affect the computation of our estimator or the estimation of K as explained in Section 4.

Theorem 2. *(Asymptotic distribution of SMAS)*
Under Assumptions 1 to 7, $\sqrt{T}(\hat{\theta}_{SMAS} - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$*, where*
$V = \left\|BG(\theta_0)\right\|^{-2} \left(BG(\theta_0), (BKB')BG(\theta_0)\right) \left\|BG(\theta_0)\right\|^{-2}$ *with* B' *the adjoint of* B*.*

The asymptotic covariance V depends on the number of simulated paths[8] through the covariance operator K. In addition, 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 $V = \left\|BG(\theta_0)\right\|^{-2}$. Unfortunately, one cannot directly choose B as the inverse of $K^{1/2}$, since it does not satisfy Assumption 3 and $B_T(\sqrt{T} z_T(., \theta_0))$ would not be well-defined asymptotically in the Hilbert space[9].

Next, we explain how to obtain valid inference on $\hat{\theta}_{SMAS}$ using a block-bootstrap resampling scheme, before returning to the choice of the (optimal) operator B and the computation of the associated optimal SMAS estimator.

4 Practical implementation of the SMAS estimator

In this section, we first explain how to obtain valid inference on $\hat{\theta}_{SMAS}$ using a block-bootstrap resampling scheme. Then, we explain how to compute the (optimal) SMAS estimator which requires: (i) estimating the kernel operator K and (ii) computing its inverse which is ill-behaved and needs to be regularized. Finally, we derive its asymptotic properties and propose a data-driven procedure to select the associated regularization parameter.

4.1 Bootstrap-based inference for $\hat{\theta}_{SMAS}$

In practice, we do not recommend direct estimation of the asymptotic variance of $\hat{\theta}_{SMAS}$ derived in Theorem 2 for two main reasons. First, it depends on the derivative of a function that is not known in closed-form. This is actually related to our framework of interest where auxiliary statistics are not known in closed-form. Second, even if these were available, it is well-documented that the asymptotic distribution of standard minimum distance estimators such as GMM or Indirect Inference may not accurately reflect finite-sample performance. To this end, we propose a bootstrap approach.

[8] The multiplying factor $(1 + 1/S)$ - well-known for indirect inference estimation - can be extracted under the maintained assumptions that the S paths $\mathcal{X}_T^s$ $(s = 1, \cdots, S)$ are generated independently of $\mathcal{X}_T$ and of each other, and that ψ_T and ψ_T^s share the same distribution.

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

The block bootstrap methodology is a general resampling scheme applicable to time series data obeying a weak dependence structure; see e.g. Lahiri (2003) for an overview. Gonçalves and White (2004) provide a unified framework for analyzing bootstrapped extremum estimators of nonlinear dynamic models for heterogeneous dependent stochastic processes. They prove the first-order asymptotic validity of the bootstrap approximation to the true distribution of quasi-maximum likelihood estimators (QMLE) for a broad class of models and data generating processes. Even though Gonçalves and White (2004) focus on QMLE, they explain that their results can be applied to prove the validity of bootstrap methods for other extremum estimators - such as ours[10]. Indeed, the key lemmas which are used to prove their two main theorems are written for a general objective function. We verify that, under our maintained assumptions, these lemmas can be applied to our objective function.

To formalize our analysis, we follow Gonçalves and White (2004) (see their Appendix A). Given our (product) probability space $(\Omega, \mathcal{F}, P_0) \times (\Omega^s, \mathcal{F}^s, P^s)$ introduced in section 3 and our observed sample $\mathcal{X}_T$ of size T, $\mathcal{X}_T^*$ is 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 of randomness, one related to the observed data and the other related to the bootstrap mechanism. When the joint randomness is of interest, the bootstrap statistic can be viewed as being defined on the product probability space $\big[(\Omega, \mathcal{F}, P_0) \times (\Omega^s, \mathcal{F}^s, P^s)\big] \times (\Omega^*, \mathcal{F}^*, P^*)$. Recall also that $\mathbb{P} \equiv P_0 \times P^s$. Given any bootstrap statistic X_T^*, we follow Gonçalves and White (2004) and define:

- $X_T^* \xrightarrow{P^*} 0$ in prob-$\mathbb{P}$ if for any $\epsilon, \delta > 0$, $\mathbb{P}\big(P^*\big(\big|X_T^*\big| > \epsilon\big) > \delta\big) \to 0$ 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 $\mathbb{P}\big(P^*\big(\big|X_T^*\big| \geq M\big) > \delta\big) \to 0$ 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.

Given the original sample $\mathcal{X}_T$, let $\hat{\theta}_T^*$ be a bootstrap version of $\hat{\theta}_{SMAS}$ defined in Theorem 1 solving:

$$\hat{\theta}_T^* = \arg\min_{\theta \in \Theta} Q_T^*(\theta) \quad \text{where} \quad Q_T^*(\theta) = \|B_T z_T^*(., \theta)\| = \|B_T z_T(., \mathcal{X}_T^*, \mathcal{X}_T^s, \theta)\|,$$

[10]We thank Sílvia Gonçalves for helpful discussions.

with $\mathcal{X}_T^*$ bootstrap data generated by block bootstrap (with blocks of length ℓ) as explained in Gonçalves and White (2004); see also Appendix A.3 for details on the block bootstrap methods implemented in Section 5.

Theorem 3. *Under Assumptions 1 to 7,*
(i) if $\ell \to \infty$ *and* $\ell = o(T)$*, then* $\hat{\theta}_T^* - \hat{\theta}_{SMAS} \xrightarrow{P^*} 0$ *in prob-*$\mathbb{P}$*.*
(ii) if $\ell \to \infty$ *and* $\ell = o(\sqrt{T})$*, then for any* $\epsilon > 0$*,*

$$\mathbb{P}\left[\sup_{x\in\mathbb{R}^p} |P^*\left[\sqrt{T}(\hat{\theta}_T^* - \hat{\theta}_{SMAS}) \leq x\right] - \mathbb{P}\left[\sqrt{T}(\hat{\theta}_{SMAS} - \theta^0) \leq x\right]| > \epsilon\right] \to 0\,.$$

Theorem 3 justifies using order statistics of the bootstrap distribution to form percentile confidence intervals for θ_0 with asymptotically correct coverage probabilities.

4.2 Optimal SMAS estimator

To obtain the optimal SMAS estimator, the control operator B should be set equal to the inverse of $K^{1/2}$ which presents a number of challenges. First, 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). To stabilize the above solution (and the associated inverse of the covariance operator), we rely on Tikhonov regularization[11] and replace the inverse of K by $K_a^{-1} \equiv (K^2 + aI)^{-1}K$ for some positive a that converges to 0. Second, the estimation of the covariance operator K amounts to estimating the covariance between $z(h_1, \theta_0)$ and $z(h_2, \theta_0)$ for all possible pairs (h_1, h_2). In practice, we cannot rely on associated sample counterparts $z_T(h_1, \theta_0)$ and $z_T(h_2, \theta_0)$ since, in general, they are not known in closed-form (e.g. when matching impulse responses), and they may actually depend 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 K is not feasible. We rely instead on a block bootstrap approach to obtain K_T and denote its regularized inverse $K_{T,a}^{-1}$. Accordingly, the optimal SMAS estimator is obtained as:

$$\hat{\theta}_{SMAS}^{opt} \equiv \arg\min_{\theta} \left\| K_{T,a}^{-1/2} z_T(\theta) \right\| . \tag{2}$$

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

Recall that, by assumption, simulated data $\mathcal{X}_T^s$ are drawn independently from observations $\mathcal{X}_T$, and that $\hat{\psi}_T(.)$ and $\hat{\psi}_T^s(.)$ are close to each other asymptotically. Consequently, K should be well approximated by $K_T \equiv (1+1/S)K_T^\psi$ with K_T^ψ the integral operator with kernel $k_T^*(h_1,h_2)$ that estimates $cov(\hat{\psi}_T(h_1),\hat{\psi}_T(h_2))$. Specifically, K_T^ψ is obtained using $N(T)$ bootstrap data of length T (generated independently by block bootstrap) denoted $\mathcal{X}_T^n$ (with $n=1,\cdots,N(T)$) such that

$$
\begin{aligned}
(K_T^\psi g)(h_1) &= \sum_{h_2} k_T^*(h_1,h_2)g(h_2) \quad \text{for any function } g \in \mathcal{D}(K_T) \\
\text{where} \qquad k_T^*(h_1,h_2) &\equiv \frac{1}{N(T)}\sum_{n=1}^{N(T)} \left[\hat{\psi}_T(h_1,\mathcal{X}_T^n) - \overline{\psi}_T^*(h_1)\right] \times \left[\hat{\psi}_T(h_2,\mathcal{X}_T^n) - \overline{\psi}_T^*(h_2)\right] \\
\text{with} \qquad \overline{\psi}_T^*(h) &\equiv \frac{1}{N(T)}\sum_{n=1}^{N(T)} \hat{\psi}_T(h,\mathcal{X}_T^n)\,.
\end{aligned}
$$

Notice that K_T (or $K_{T,a}$) does not directly depend on θ or $\hat{\psi}_T^s$. Further, the computation of the optimal objective function greatly simplifies in practice when the number of bootstrap samples $N(T)$ can be chosen larger than the effective number of values of h, say $\tilde{H}$, selected to compute the SMAS estimator - and that, even when $\tilde{H}$ is larger than T, the sample size. In Section 5, the number of bootstrap samples is always larger than the number of matched impulse responses: e.g. in our baseline experiment with the small scale New Keynesian model $N(T)=500$, $\tilde{H}=320$ and $T=232$. The optimal objective function can then be rewritten as follows:

$$
\hat{\theta}_{SMAS}^{opt} = \arg\min_\theta \left[Z_T(\theta)'(K_T+aI)^{-1}K_T Z_T(\theta)\right]\,,
$$

with $Z_T(\theta)$ the $\tilde{H}$ vector with j-th component $z_T(h_j,\theta)$ and K_T the square matrix of size $\tilde{H}$ with (i,j) component $k_T^*(h_i,h_j)$. When it is not possible to choose $N(T)$ larger than $\tilde{H}$, the objective function is computed by spectral decomposition: we discuss this extension in Supplementary Appendix S.2.

To derive the asymptotic properties of $\hat{\theta}_{SMAS}^{opt}$, Assumptions 3 and 5 need to be updated, because the regularity properties originally maintained on the operator B (e.g. boundedness) are not satisfied by the optimal operator. 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 8. *(i) The kernel k associated with the covariance operator K is L^2.*

(ii) $z(.,\theta) \in \mathcal{H}(K) + \mathcal{H}(K)^{\perp}\ \forall \theta \in \Theta$.

(iii) *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* $j = 1, \cdots, p$.

(iv) *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.*

We maintain that $k(.)$ is an L^2-kernel, which requires that,

$$\sum_{h_1}\sum_{h_2} k(h_1,h_2)^2 < \infty\,.$$

Since $k(.)$ corresponds to the (asymptotic) covariance of $z_T(.,\theta_0)$, sufficient conditions are well-known, including e.g. boundedness and mixing conditions maintained on $z_T(.,\theta_0)$ in Assumption 7. It implies that the underlying covariance operator K is a 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. Next, we introduce additional (high-level) regularity assumptions to ensure that population quantities (such as the covariance operator K) can be appropriately estimated by bootstrap.

Assumption 9. *(i)* $\sqrt{T}\big(z_T(.,\theta_0) - z_T^*(.,\theta_0)\big) = o_{P^*}(1)$ *in prob-*$\mathbb{P}$.

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

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

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

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

Next, we present the asymptotic properties of the optimal SMAS estimator.

Theorem 4. *(Asymptotic behavior of the optimal SMAS estimator)*
Under Assumptions 1, 2, 4, and 6 to 9, the optimal SMAS estimator defined in (2) is $\sqrt{T}$*-consistent and asymptotically normally distributed with mean zero and variance* $\left\|K^{-1/2}G(\theta_0)]\right\|^{-2}$*, as* $T \to \infty$*,* $\sqrt{T}a^{3/2} \to \infty$*,* $a \to 0$.

To select the regularization parameter a, we develop a data-driven selection procedure. In Appendix A.2, we explain how to choose the parameter c by cross-validation where $a \equiv c/T^{\nu}$ with given $0 < \nu < 1/3$ and the rate of decay of a to 0 is motivated by our theoretical (asymptotic) results for the optimal SMAS in Theorem 4.

Algorithm 1 in Appendix A explains how to compute our optimal estimator $\hat{\theta}^{opt}_{SMAS}$ in practice. To obtain valid inference on $\hat{\theta}^{opt}_{SMAS}$, we recommend using the bootstrap approach highlighted in section 4.1 after replacing B_T by $K^{-1/2}_{T,a}$.

5 Simulation study

We illustrate the small sample properties of the SMAS estimator by revisiting two well-known small- and medium-scale New Keynesian models. As a robustness check, we also consider a baseline stylized DSGE model where the IRs are known analytically.

5.1 Small-scale New Keynesian model with 2 indices

We first revisit the small-scale New Keynesian model of Guerron-Quintana et al. (2017) and focus on the estimation of the price stickiness. In the model, π, R, and x denote respectively, inflation rate, interest rate, and output gap:

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

with α the probability to fix price, β the discounting factor, ω the disutility to work, and σ and τ the elasticities of substitution across time and various commodities. ϕ_π and ϕ_x measure how responsive the central bank is to changes in inflation and output gap. ϵ^z and ϵ^r are white noise processes, independently distributed as standard normals. The vector of unknown parameters and its true value are, respectively,

$$\theta = (\alpha, \beta, \omega, \sigma, \tau, \rho_r, \rho_z, \phi_\pi, \phi_x, \sigma^z, \sigma^r)', \quad \theta_0 = (0.75, 0.99, 1, 1, 6, 0.75, 0.9, 1.5, 0.125, 0.3, 0.2)'.$$

Here, we estimate α, and calibrate all the other parameters to their true value[12]. Expectations are history based, and we generate our sample of observations on inflation and interest rate using `Dynare`[13]. The regularity assumptions maintained in Section 3 are discussed in the context of this model in Appendix B.1.1.

Given the sample of T observations on inflation and interest rate, a VAR(p) model is estimated (see Appendix C). Based on the estimated VAR(p) model, associated transition matrix, residuals and impulse responses[14] 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 $\hat{\psi}^s_T(\alpha)$. As explained in section 4.2, we generate (independently) N bootstrap data of length T (with $N > 4H$) to compute the covariance operator K_T (which is a square matrix of size $4H$). The (optimal) SMAS estimator of α is then

$$\hat{\alpha}_T = \arg\min_{\alpha} \left[z_T(\alpha)'(K_T + a_T I)^{-1} K_T z_T(\alpha) \right] \quad \text{with } z_T(\alpha) = \hat{\psi}_T - \frac{1}{S}\sum_{s}^{S} \hat{\psi}^s_T(\alpha)\,,$$

with a_T chosen by cross-validation[15]. In practice, the optimization problem is solved by conducting a grid search over potential values of α. Since the probability of the price stickiness α is naturally bounded between 0 and 1, the grid of candidate values is set at [0.005, 0.995] with a step of 0.005. See Appendix A for implementation details.

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 generate $N = 500$ bootstrap samples.

We consider four SMAS estimators, computed with the optimal operator and the

[12]Recent work by Antoine et al. (2023) allows parameters to be incorrectly calibrated

[13]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.

[14]We focus here on matching (so-called) structural impulse responses (see details in Appendix C). 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)).

[15]As explained in Appendix A.2, our cross-validation is implemented using the following grid of candidates $\mathcal{C} = [2, 1, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0001]$.

diagonal one[16] - with and without regularization - and the two estimators proposed by Guerron-Quintana et al. (2017) (hereafter GIK), respectively with the optimal weighting matrix and the diagonal one. Tables 1 to 4 summarize the performance of these estimators in terms of Monte-Carlo average (mean), standard deviation (SD), Root Mean Squared Error (RMSE), Mean Absolute Deviation (MAD), and effective coverage rates of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications. Figures 2 and 3 display the Monte-Carlo histograms of these estimators. We match: (i) a medium number of IR (up to $H = 20$) in Table 1; (ii) a large number of IR (up to $H = 80$) in Table 2 and Figure 2; (iii) a small number of IR (up to $H = 2$ and 8) in Table 3 and Figure 3. In Table 4, we report results obtained with a sample of only 100 observations (or 25 years of quarterly data) when matching up to $H = 20$ and 80. The bootstrap samples are obtained by block-bootstrap resampling as explained in Appendix A.3.

• 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.

When matching a small number of IR (e.g. 2), the SMAS estimators with and without regularization are very close to each other, with the non-regularized one only slightly better in terms of bias, but not in terms of SD or RMSE. This suggests that there is little to no drawback in applying regularization, even when it is not needed. Since, in practice, it is not always clear how to determine whether regularization is needed or not, we recommend to always implement SMAS estimation with regularization.

• Optimal vs Diagonal operator:

Overall - and as predicted by theory - the performance of SMAS computed with the optimal operator (hereafter optimal SMAS) is better than that of SMAS computed with the diagonal operator as it delivers smaller SD and RMSE. However, the differences remain small, particularly when H is small. Still, when matching up to horizon 80 in the benchmark model, SD of optimal SMAS is approximately 6.5% smaller.

Notice that this is not the case for GIK: e.g. optimal GIK performs worst than diagonal GIK (with larger SD, MAD, and RMSE), even with H as small as 8.

[16]The diagonal operator imposes zeros off-diagonal on K_T; similar labels are used to distinguish the two estimators from Guerron-Quintana et al. (2017).

• SMAS vs GIK:

Overall, regularized SMAS estimators behave much better than the GIK estimators, for all the horizons we consider - even the shorter ones: SMAS outperforms GIK according to all measures of performance, and displays the smallest bias, MAD, SD and RMSE throughout - even when considering smaller sample sizes, or $H = 2$. In addition, effective coverage rates are much closer to their nominal levels for SMAS than for GIK. For example, when matching up to $H = 80$, the RMSE of optimal SMAS is nearly half of that of optimal GIK (0.067 vs 0.121) with a significantly smaller SD (0.066 vs 0.104); yet, optimal GIK displays size distortions, whereas SMAS does not: the corresponding Monte-Carlo empirical coverage rates of 95% confidence intervals are respectively 95.1% for SMAS and 90.5% for GIK.

• Robustness checks: horizon H, order p, and distribution of the errors

Overall, the performance of regularized SMAS is remarkably stable as a function of the order p of the fitted VAR model and of H - including when H is large and/or when the number of IR exceeds the sample size. This should alleviate potential concerns about the robustness of SMAS when the number of auxiliary statistics is large.

Finally, we consider deviations from the model assumptions by redoing our simulation study when the error terms in the real DGP are not normally distributed, but rather t-distributed[17] - with either 4 or 20 degrees of freedom. Everything else remains exactly as before. Table 5 collects our results for the baseline experiment where we fit a VAR(2) model to a sample of $T = 232$ observations and match IR up to $H = 20$ or 80. Similar to the results obtained previously (see e.g. Tables 1 and 2), our optimal SMAS estimator performs best throughout, displaying smaller RMSEs with effective coverage rates closer to their nominal levels. As expected, when deviations from normality are more severe (e.g. with t(4)), its performance deteriorates slightly with increased MAD and SD; nonetheless, it still outperforms other estimators. All in all, our results suggest that SMAS can indeed withstand (small) deviations from the model assumptions, including deviations in the distributions of the error terms.

5.2 Medium-scale New Keynesian model with 7 indices

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

[17] We thank Willi Mutschler (`https://mutschler.eu/`) for sharing his `Dynare` code; see also Mutschler (2018).

system which involves output, consumption, investment, wage, working hours, inflation, and interest rate; see section 1 in Smets and Wouters (2007) p588. Our sample of observations on the seven above-mentioned indices is generated using `Dynare`[18] with parameter values as in Table 6: in particular, the Calvo parameter ξ_p is set at 0.908 and the degree of price indexation γ_p at 0.469.

Our estimation procedure closely follows subsection 5.1: first, a VAR(p) model is applied to our sample of 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[19]. Both parameters of interest are naturally bounded between 0 and 1, and the grid of candidate values is set at [0.05, 0.95] for γ_p with a step of 0.05, and at [0.1,0.995] for ξ_p with a step of 0.05 up to 0.6 and 0.005 afterwards.

We focus 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 (in all seven indices) are matched up to horizon H - chosen between 20 and 80 (that is, between 5 and 20 years after the shock with quarterly data). For the estimation of the optimal operator and its inverse, we generate $N = 100$ samples by block-bootstrap resampling.

We report the performance of four SMAS estimators, respectively with the optimal operator and the diagonal one, with and without regularization in terms of Monte-Carlo average (mean), standard deviation (SD), Root Mean Squared Error (RMSE), Mean Absolute Deviation (MAD), and effective coverage rates of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

(1) *Estimation of the Calvo parameter alone.*
By design, the estimation of the Calvo parameter is challenging as its true unknown parameter value is much closer to the upper bound 1. Estimation results are reported in Table 7. Overall, the performance of the regularized SMAS estimator is very good and dominates that of the non-regularized one according to all reported measures of

[18]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.

[19]The grid of candidates is $\mathcal{C} = [0.1, 0.05, 0.01, 0.005, 0.001, 0.0001, 10^{-5}, 10^{-6}]$.

performance. We do notice some small size distortions: those tend to disappear for regularized SMAS when the sample size increases, but not for non-regularized SMAS. As predicted by theory, the performance of optimal SMAS is better than that of diagonal SMAS as it delivers smaller SD and RMSE, and improved coverage rates - especially with $T = 236$: SD is approximately 10% smaller with $T = 236$ and $H = 80$. We focus on that SMAS estimator in our discussions below. The performance of SMAS does improve when adding IR at long horizons according to all measures of performance. We specifically notice the reduction in MAD and SD: by over 5% and 7%, respectively, when $T = 236$. Such results clearly indicate that there is useful information in long-run IR and that our estimation procedure is able to harvest it.

(2) *Joint estimation of the degree of price indexation and the Calvo parameter.* Results for the joint estimation of the degree of price indexation and the Calvo parameter are reported in Table 8. Overall, the performance of the regularized optimal SMAS estimator is very good and in line with the results obtained when estimating the Calvo parameter alone. Particularly, we find that there is useful information contained in long-run IR, and that both parameters are more precisely estimated when matching impulse responses up to long horizons.

5.3 Baseline stylized DSGE model

We now 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[20].

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[21]. 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.

We focus on delivering inference on only one parameter of the model, namely the

[20]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.

[21]Our SMAS estimator is implemented using candidates for the Calvo parameter set at [0.005,0.995] with step 0.005; for the cross-validation, the grid is $\mathcal{C} = [0.1, 0.05, 0.01, 0.005, 0.001, 0.0001, 10^{-5}, 10^{-6}]$.

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. Results are reported in Table 10.

Overall, the infeasible SMAS estimator performs better, both in terms of bias and standard deviation. However, the differences remain small, and tend to decrease when the sample size increases.

6 Conclusion

We establish the consistency and derive the asymptotic distribution of the simulated minimum distance estimator when the number of components of the auxiliary statistics is large - typically larger than the sample size. Our SMAS estimator is easy to implement and exploits 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. We design a bootstrap-based procedure to estimate the optimal (covariance) operator associated with our efficient estimator, and rely on Tikhonov regularization to stabilize its inverse. We also propose a cross-validation procedure to select the associated tuning parameter.

We illustrate the performance of our proposed estimator through a Monte-Carlo study based on small- and medium-scale New Keynesian models when matching a large number of impulse responses including at long-run horizons. Overall, our simulation experiments reveal that: (i) our estimator works well in practice; (ii) long-run information is valuable, and we show how it improves the quality of estimates of structural parameters of interest - including their precision. These results are encouraging, and our estimator promises to be useful in a wide range of applications in applied macro and beyond. We leave these investigations for future work.

References

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

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

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.

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.

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.

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

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

Fève, P., J. Matheron, and J.-G. Sahuc (2009). Minimum distance estimation and testing of dsge models from structural vars. *Oxford Bulletin of Economics and Statistics 71*, 883–894.

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

Gonçalves, S. and H. White (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics 119*, 199–220.

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.

Guay, A. and F. Pelgrin (2023). Structural VAR models in the Frequency Domain. *Forthcoming, Journal of Econometrics*.

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.

Hassler, U. and P. Kokoszka (2010). Impulse Responses of Fractionally 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*.

Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Lahiri, S. (2003). *Resampling Methods for Dependent Data*. Springer Series in Statistics. NY: Springer.

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.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer Berlin, Heidelberg.

Mutschler, W. (2018). Higher-order statistics for DSGE models. *Econometrics and Statistics 6*, 44–56.

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.

Politis, D. and H. White (2004). Automatic Block-Length Selection for the Dependent Bootstrap. *Econometrics Reviews 23*, 53–70.

Sargan, J. (1983). Identification and lack of identification. *Econometrica 51*, 1605–1633.

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*, 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

In Appendix A, we detail our implementation procedure including the selection of the regularization parameter by cross-validation and the block-bootstrap resampling. In Appendix B, we first review the maintained regularity assumptions in the context of the small-scale model of Section 5.1. Then, we display the tables of results of our Monte-Carlo study. In Appendix C, we explain how to compute the impulse responses. In the Supplementary Appendix, we present: (i) the proofs of our theoretical results; (ii) the computation of the optimal SMAS estimator by spectral decomposition; (iii) additional Monte-Carlo results in the context of the medium-scale model of Section 5.2; (iv) results on the direct estimation of the asymptotic variance of SMAS in the context of the small-scale model of Section 5.1.

A Implementation details

A.1 General implementation

The first algorithm 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 $\hat{\epsilon}_T$ as explained in Appendix C.*

2. *For given $\theta \in \Theta$:*

   (a) *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 $z_T(\theta) = \hat{\psi}_T(\mathcal{X}_T) - \sum_s \hat{\psi}_T^s(\theta)/S$.*

   (b) *Generate independently N bootstrap data $\mathcal{X}_T^n(\theta)$ of length T; compute the associated (chosen) impulse responses $\hat{\psi}_T^{*(n)}(\mathcal{X}_T^n)$ with $n = 1, \cdots, N$; compute the estimator $\hat{K}_T$ of the operator K as explained in section 4.*

   (c) *Select the regularization parameter a_T by following Algorithm 2.*

3. *Obtain $\hat{\theta}_{SMAS}$ as the minimizer over θ of $\|\hat{K}_{T,a_T}^{-1/2} z_T(\theta)\|$.*

A few comments are warranted.

(i) As detailed in section 4, the optimal operator only depends on the demeaned distance $z_T^{*(n)}$ - and the demeaning is across the bootstrap samples. Since the auxiliary

statistics computed on simulated data are fixed across n, the computation of the optimal operator only involves $\hat{\psi}_T^{*(n)}$ and not z_T^{*n}.
(ii) When the bootstrap data are generated by resampling, they do not actually depend on θ; they can then be denoted more simply as $\mathcal{X}_T^n$. This observation, together with the previous comment, means that, in this case, the computation of the optimal operator does not actually depend on θ.

A.2 Selection of the regularization parameter

We now 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 $a_T \equiv c/T^\nu$ with given $0 < \nu < 1/3$. Notice that the rate of decay of a_T to 0 is motivated by our theoretical (asymptotic) results for the optimal SMAS in Theorem 4.

We first sketch how the cross-validation works in practice. We start by splitting the sample of T observations into two subamples: the training subsample, which corresponds to the first $\tilde{T}$ observations, and the testing subsample, 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.

Next, Algorithm 2 details the selection of the regularization parameter by cross-validation. 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 (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 (non-optimal) SMAS objective function with identity matrix.*

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

A.3 Residual-based block bootstrap resampling

The block bootstrap methodology is a general resampling scheme applicable to time series data obeying a weak dependence structure; see e.g. Lahiri (2003) for an overview and chapter 12 in Kilian and Lütkepohl (2017) for applications to SVAR models. Here, we rely on residual-based block bootstrap methods, which can be summarized as follows. Consider the VAR(p) approximation of the vector of observables:

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \,.$$

For a given (fixed) p, the above VAR model can only approximately capture the covariance properties of the data, and the associated u_t remain e.g. serially dependent. In that sense, the VAR model can be intuitively understood as a particular linear projection technique. Having fitted the VAR(p) model to the data and having recovered estimates of the model parameters and the model residuals

$$\hat{u}_t = y_t - \hat{\nu} + \hat{A}_1 y_{t-1} + \cdots + \hat{A}_p y_{t-p} \,,$$

for $t = 1, \cdots, T$, we arrange the residuals in the form of a matrix

$$\begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \cdots & \hat{u}_l \\ \hat{u}_2 & \hat{u}_3 & \cdots & \hat{u}_{l+1} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{u}_{T-l+1} & \hat{u}_{T-l+2} & \cdots & \hat{u}_T \end{bmatrix} ,$$

where each row denotes a block of l consecutive residuals and the blocks are overlapping. The number of blocks is $s = T - l + 1$. The bootstrap innovations are obtained by drawing at random with replacement from the rows of this matrix, laying these blocks of residuals end-to-end, and retaining the first T innovations denoted $[\tilde{u}^*_1, \cdots, \tilde{u}^*_T]$.

The corresponding bootstrap innovations denoted $[u_1^*, \cdots, u_T^*]$ are then obtained after recentering, that is

$$u_{jl+i}^* = \tilde{u}_{jl+i}^* - \frac{1}{T-l+1} \sum_{r=0}^{T-l} \hat{u}_{i+r} \,,$$

for $i = 1, 2, \cdots, l$ and $j = 0, 1, \cdots, s-1$.
The corresponding bootstrap data, $[y_{-p+1}^*, \cdots, y_T^*]$, are generated recursively using the estimated VAR coefficients and initial historical conditions

$$y_t^* = \hat{\nu} + \hat{A}_1 y_{t-1}^* + \cdots + \hat{A}_p y_{t-p}^* + u_t^* \,.$$

Bootstrap realizations of the estimators of interest (e.g. impulse response functions) are obtained by fitting a VAR(p) model to this sequence of bootstrap data, and proceeding exactly as done on the sequence of observed data.

To select the block length, we suggest Politis and White (2004) automatic block-length selection procedure[22].

B Monte-Carlo simulation study

B.1 Small-scale model

B.1.1 Regularity assumptions

In our small-scale model, inflation and real output do not react contemporaneously to the monetary policy shock ξ_t, but they do respond contemporaneously to the shock to the investment-savings relationship, z_t. These restrictions ensure that the structural shocks of interest can be identified in the VAR model. For additional discussions, see e.g. Guerron-Quintana et al. (2017). We now review the maintained regularity assumptions of Section 3 in the context of the above model.

• Assumption 1 (DGP):
In our simulation study, our DGP is stationary. In practice, it is often assumed that the two series of inflation and interest rate are stationary - especially over a well-chosen (small) sample. For additional discussions, see e.g. Christiano et al. (2005).

• Assumption 2 (Auxiliary statistic):
Assumption 2(i) follows from Assumption 1 maintained on the underlying DGP: in general, it is known that weakly dependent sequences are characterized by abso-

[22]The matlab code developed by Andrew Patton is available here.

lutely summable impulse response coefficients of their Wold decomposition; see e.g. Hassler and Kokoszka (2010). Assumption 2(ii) is an identification assumption which ensures that θ_0 is the only value that ensures a perfect match between the IR function computed from the DGP and its simulated counterpart.

• Assumption 3 (Operator):

Assumption 3 maintains regularity assumptions on the operator B to ensure that the population objective function is well-defined and uniquely minimized at θ_0. When matching impulse responses, Sokullu (2020) argues that it is useful to relax the non-singularity of B to allow for the singularity of the covariance operator[23] of impulse-response functions K. See related discussions in Guerron-Quintana et al. (2017) when the number of IR parameters exceeds the number of VAR model parameters.

• Assumptions 4, 5, and 6 (Sample counterparts, differentiability, commutativity):

Computation of impulse response functions is detailed in Section C; see also e.g. Chapter 11 in Hamilton (1994). Under our maintained assumptions, sample counterparts will satisfy the maintained regularity conditions: e.g. continuity, differentiability, and convergence to corresponding population values.

• Assumption 7 (Functional convergence):

Asymptotic properties of estimated impulse response functions are derived in e.g. Lütkepohl (1990) when the VAR corresponds to the DGP; for more general results, see e.g. chapter 15 in Lütkepohl (2005) or chapter 12 in Kilian and Lütkepohl (2017). Under our maintained assumptions (e.g. weak dependence of underlying processes), the IR function will satisfy the maintained functional convergence.

[23]Recall, however, that B cannot be directly chosen optimally as the inverse of $K^{1/2}$.

B.1.2 Results

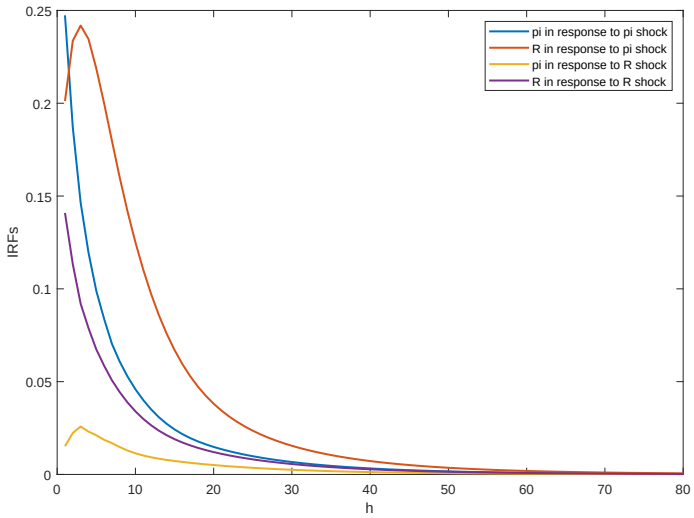

Figure 1: Impulse responses of the small-scale DSGE model as a function of the horizon.

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 1: Estimation of the price stickiness $\alpha = 0.75$ in the small-scale DSGE model. Performance of the SMAS estimators (with and without regularization) and the 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 2: Estimation of the price stickiness $\alpha = 0.75$ in the small-scale DSGE model. Performance of the SMAS estimators (with and without regularization) and the estimators of Guerron-Quintana et al. (2017) for different simulations designs when matching IR over 80 periods 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					
	SMAS				GIK	
	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
MAD	0.0048	10^{-4}	0.0052	0.0003	0.0083	0.0203
SD	0.0536	0.0578	0.0536	0.0577	0.0666	0.0584
RMSE	0.0538	0.0578	0.0539	0.0577	0.0671	0.0618
Coverage						
95%	95.3	96.2	95.1	96.0	94.7	92.8
90%	90.9	91.4	90.7	90.5	91.9	88.2
	Panel B: Matching at horizons 1 to 8					
	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%	94.5	96.1	94.8	96.3	94.1	92.2
90%	91.7	93.3	91.7	93.9	90.9	87.2

Table 3: Estimation of the price stickiness $\alpha = 0.75$ in the small-scale DSGE model. 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					
	SMAS				GIK	
	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
MAD	0.0175	0.1465	0.0207	0.1437	0.0411	0.0459
SD	0.1060	0.2491	0.1040	0.2485	0.1346	0.1086
RMSE	0.1074	0.2890	0.1060	0.2871	0.1407	0.1179
Coverage						
95%	94.6	87.2	94.3	86.8	92.4	90.8
90%	91.4	82.5	91.2	82.9	90.3	87.5
	Panel B: Matching at horizons 1 to 80					
	SMAS				GIK	
	diagonal reg.	diagonal non-reg.	optimal reg.	optimal non-reg.	diagonal	optimal
mean	0.7323	0.4753	0.7263	0.4692	0.6853	0.6493
MAD	0.0177	0.2747	0.0237	0.2808	0.0647	0.1007
SD	0.1051	0.2522	0.1056	0.2539	0.1628	0.1425
RMSE	0.1065	0.3729	0.1082	0.3785	0.1752	0.1745
Coverage						
95%	93.8	75.1	93.7	75.1	92.5	88.5
90%	91.6	67.9	91.8	65.8	88.7	83.7

Table 4: Estimation of the price stickiness $\alpha = 0.75$ in the small-scale DSGE model. 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.

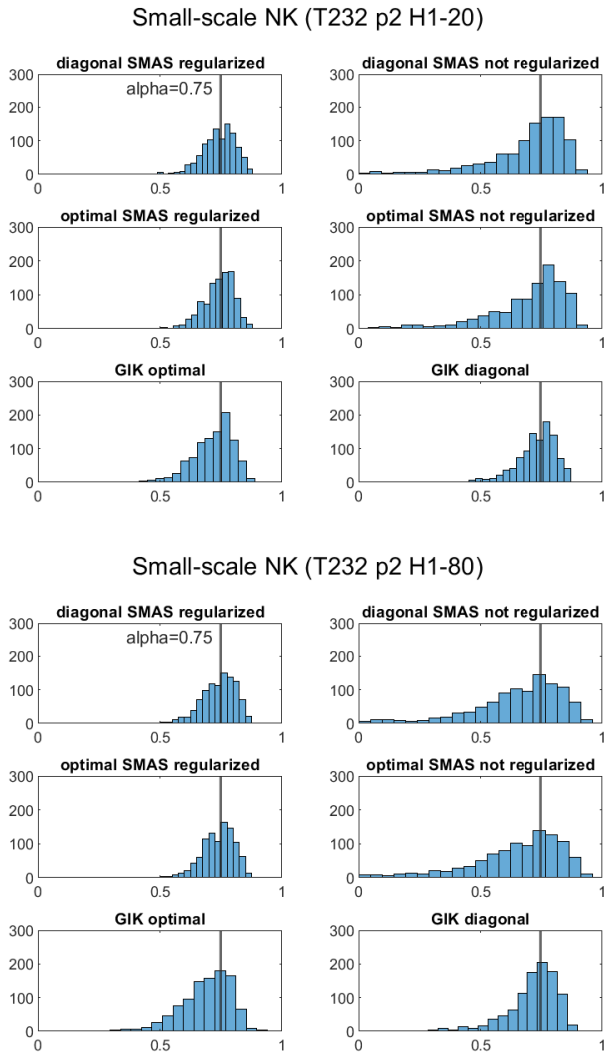

Figure 2: Estimation of the price stickiness $\alpha = 0.75$ in the small-scale DSGE model. Monte-Carlo distribution of the following six estimates obtained when matching IR up to $H = 20$ (top three rows) and up to $H = 80$ (bottom three rows) over 1,000 replications: regularized SMAS with diagonal 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 red vertical line represents the true value of the parameter.

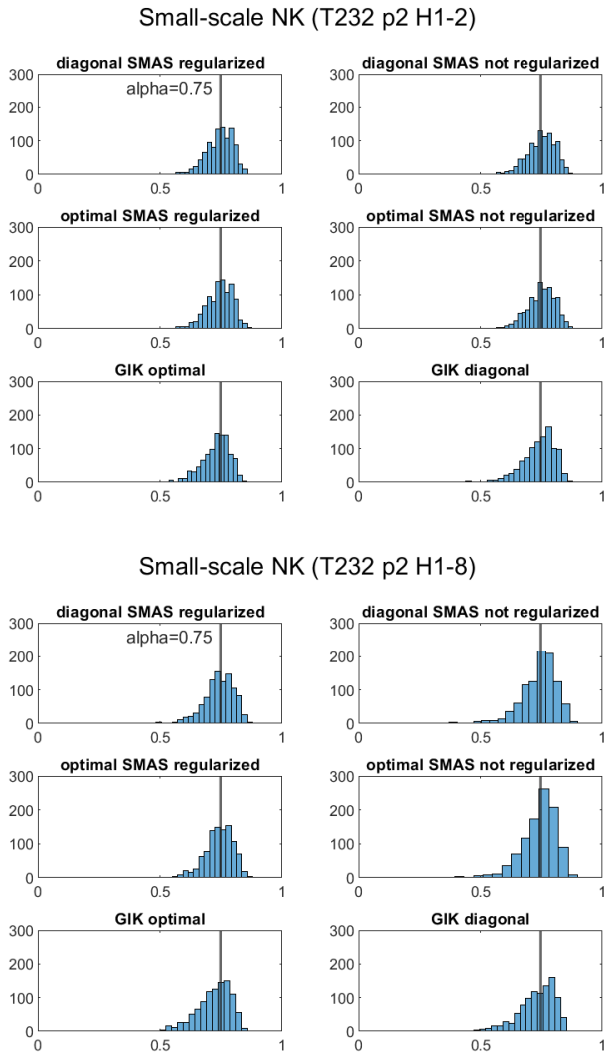

Figure 3: Estimation of the price stickiness $\alpha = 0.75$ in the small-scale DSGE model. Monte-Carlo distribution of the following six estimates obtained when matching IR up to $H = 2$ (top three rows) and up to $H = 8$ (bottom three rows) over 1,000 replications: regularized SMAS with diagonal operator (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 red vertical line represents the true value of the parameter.

	Error terms are t(20)						**Error terms are t(4)**					
	Panel A.1: Matching at horizons 1 to 20						*Panel B.1: Matching at horizons 1 to 20*					
	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	0.7433	0.6928	0.7412	0.6907	0.7287	0.7204	0.7385	0.6782	0.7371	0.6817	0.7325	0.7189
MAD	0.0067	0.0573	0.0088	0.0593	0.0213	0.0296	0.0115	0.0718	0.0129	0.0683	0.0175	0.0311
SD	0.0667	0.1587	0.0633	0.1625	0.0799	0.0841	0.0663	0.1829	0.0638	0.1773	0.0808	0.0834
RMSE	0.0671	0.1687	0.0639	0.1730	0.0827	0.0891	0.0673	0.1965	0.0651	0.1900	0.0826	0.0890
Coverage												
95%	94.7	92.2	94.9	92.4	94.3	93.0	94.5	91.7	93.6	91.2	94.3	92.3
90%	91.6	88.8	91.9	89.6	90.7	89.5	89.9	89.7	90.2	88.9	92.1	89.3
	Panel A.2: Matching at horizons 1 to 80						*Panel B.2: 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	0.7443	0.6543	0.7426	0.6521	0.7081	0.6916	0.7382	0.6479	0.7376	0.6446	0.7157	0.6936
MAD	0.0057	0.0957	0.0074	0.0979	0.0419	0.0584	0.0118	0.1021	0.0124	0.1054	0.0343	0.0564
SD	0.0686	0.1867	0.0640	0.1892	0.1188	0.0998	0.0714	0.1859	0.0654	0.1936	0.1105	0.1084
RMSE	0.0689	0.2098	0.0644	0.2131	0.1260	0.1156	0.0723	0.2121	0.0665	0.2205	0.1157	0.1222
Coverage												
95%	94.7	89.7	94.2	89.9	94.3	90.1	94.4	90.1	94.4	90.1	93.6	92.5
90%	90.5	85.8	91.5	86.4	92.5	85.3	91.6	86.2	91.5	86.9	91.4	88.2

Table 5: Estimation of the price stickiness $\alpha = 0.75$ in the small-scale DSGE model when the error terms are t-distributed, either t(20) in Panels A.1 and A.2, or t(4) in Panels B.1 and B.2. 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 $T = 232$ 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.

B.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 6: Parameter values in the medium-scale model.

Panel A: Matching IR over medium horizons 1-20 on all 7 indices								
	$T = 236$				$T = 944$			
	Diagonal		Optimal		Diagonal		Optimal	
	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.
mean	0.6974	0.6467	0.7561	0.7253	0.8307	0.7877	0.8461	0.7908
MAD	0.2106	0.2613	0.1519	0.1827	0.0773	0.1203	0.0619	0.1172
SD	0.2376	0.2474	0.2175	0.2385	0.1627	0.2036	0.1552	0.2048
RMSE	0.3175	0.3598	0.2653	0.3005	0.1801	0.2365	0.1671	0.2360
Coverage								
95%	84.3	79.8	88.2	85.2	95.1	91.7	95.5	90.9
90%	82.0	76.8	86.8	84.5	91.2	90.3	91.4	90.2
Panel B: Matching IR over long horizons 1-80 on all 7 indices								
	$T = 236$				$T = 944$			
	Diagonal		Optimal		Diagonal		Optimal	
	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.
mean	0.6902	0.6487	0.7639	0.7273	0.8194	0.7723	0.8530	0.7852
MAD	0.2178	0.2593	0.1441	0.1807	0.0886	0.1357	0.0550	0.1228
SD	0.2225	0.2444	0.2031	0.2305	0.1648	0.2006	0.1300	0.2008
RMSE	0.3114	0.3563	0.2491	0.2929	0.1871	0.2422	0.1412	0.2354
Coverage								
95%	84.1	80.6	90.1	86.5	93.8	90.4	92.2	90.3
90%	82.9	77.9	88.3	85.1	89.7	89.5	87.9	89.7

Table 7: Estimation of the Calvo parameter $\zeta_p = 0.908$ in the medium-scale model. Performance of SMAS estimators (with and without regularization) when matching impulse responses over medium (Panel A) and long horizons (Panel B) in all indices with sample size $T = 236$ and 944. We fit VAR(4) and consider $N = 1,000$ replications.

Matching IR over medium and long horizons in all 7 indices								
	$H = 20$				$H = 80$			
	Regularized		Non-regularized		Regularized		Non-regularized	
	γ_p	ζ_p	γ_p	ζ_p	γ_p	ζ_p	γ_p	ζ_p
mean	0.4945	0.7638	0.5037	0.6987	0.4872	0.7498	0.5069	0.7002
MAD	0.0255	0.1442	0.0347	0.2093	0.0182	0.1852	0.0378	0.2078
SD	0.2832	0.2356	0.2725	0.2641	0.2722	0.2333	0.2711	0.2567
RMSE	0.2843	0.2762	0.2747	0.3369	0.2728	0.2819	0.2738	0.3303
Coverage								
95%	100	88.4	100	83.6	100	87.9	100	83.9
90%	94.3	87.5	95.2	81.8	95.2	86.7	94.8	81.9
Joint Coverage								
95%	89.2		84.9		88.6		85.9	
90%	88.0		83.2		86.9		83.1	

Table 8: Joint estimation of the degree of price indexation $\gamma_p = 0.469$ and the Calvo parameter $\zeta_p = 0.908$. Performance of optimal SMAS estimator (with and without regularization) when matching impulse responses over medium horizons ($H = 20$) and long horizons ($H = 80$) in all indices. The sample size is $T = 236$; we fit VAR(4), and consider $N = 1,000$ replications.

B.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 9: Parameter values in the baseline stylized model.

Panel A: Matching IR over medium horizons $H = 20$								
	$T = 200$				$T = 400$			
	Infeasible		SMAS		Infeasible		SMAS	
	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.
mean	0.6187	0.7453	0.5905	0.7490	0.6435	0.7423	0.6273	0.7357
MAD	0.0313	0.0953	0.0595	0.0990	0.0065	0.0923	0.0227	0.0857
SD	0.1637	0.3557	0.1939	0.3540	0.1762	0.3458	0.1796	0.3517
RMSE	0.1667	0.3682	0.2028	0.3675	0.1763	0.3580	0.1811	0.3620
Coverage								
95%	94.0	100.0	92.4	100.0	95.8	100.0	96.1	100.0
90%	91.1	93.9	89.3	93.6	90.5	93.0	88.3	92.4
Panel B: Matching IR over long horizons $H = 80$								
	$T = 200$				$T = 400$			
	Infeasible		SMAS		Infeasible		SMAS	
	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.
mean	0.6201	0.7397	0.6054	0.7490	0.6512	0.7692	0.6352	0.7682
MAD	0.0299	0.0897	0.0446	0.0990	0.0012	0.1192	0.0148	0.1182
SD	0.1780	0.2900	0.2058	0.2775	0.1809	0.2649	0.1853	0.2646
RMSE	0.1805	0.3036	0.2106	0.2946	0.1809	0.2905	0.1859	0.2898
Coverage								
95%	95.2	95.7	93.3	96.6	97.5	96.5	95.6	97.4
90%	88.6	93.7	88.8	94.3	89.5	95.4	91.8	94.2

Table 10: Estimation of the Calvo parameter $\zeta_p = 0.65$ in the baseline stylized model. Performance of the feasible and infeasible SMAS estimators with and without regularization when matching impulse responses over medium to long horizons. We consider dynamic responses obtained up to five years after the shock (20 periods) in Panel A, and up to twenty years (80 periods) in Panel B with sample sizes $T = 200$ and $T = 400$. We fit VAR(4) and consider $N = 1,000$ replications.

C Computation of (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:

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

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,

$$\begin{aligned} x_t &= \Theta(L)\Phi_0 + \Theta(L)P^{-1}Pu_t \equiv \psi_0 + \Psi(L)\epsilon_t, \qquad \epsilon_t \sim (0, I) \\ \text{with} \quad \Theta(L) &= (I - \sum_{j=1}^{p} \Phi_j L^j)^{-1}. \end{aligned}$$

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 $\hat{\Psi}(L)$.

Supplementary Appendix to:

Simulation-based estimation with many auxiliary statistics applied to long-run dynamic analysis

by Bertille Antoine and Wenqian Sun

In Supplementary Appendix S.1, we present the proofs of our theoretical results. In S.2, we explain how to computate the optimal SMAS estimator by spectral decomposition. In S.3, we present additional Monte-Carlo results obtained in the context of the medium-scale model discussed in section 5.2 of the main paper. In S.4, we present results on the direct estimation of the asymptotic variance of SMAS obtained in the context of the small-scale model discussed in section 5.1 of the main paper.

S.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 C.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 C.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 $\Big(f, (BKB')(f)\Big)$. $\square$

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

$$\hat{\theta}_{SMAS} \equiv \arg\min_{\theta\in\Theta} \left\| B_T z_T(.,\theta) \right\| = \arg\min_{\theta\in\Theta} \left(B_T z_T(.,\theta), B_T z_T(.,\theta) \right).$$

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

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

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:

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

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 C.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:

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

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

■

• **Proof of Theorem 3**:

Gonçalves and White (2004) (hereafter GW(04)) provide a unified framework for analyzing bootstrapped extremum estimators of nonlinear dynamic models for heterogeneous dependent stochastic processes. They prove the first-order asymptotic validity of the bootstrap approximation to the true distribution of quasi-maximum likelihood estimators (QMLE) for a broad class of models and data generating processes without imposing stationarity and restrictive memory conditions[24]. Even though GW(04) focus on QMLE, they explain on page 211 that their results can be applied to prove the validity of bootstrap methods for other extremum estimators. Indeed, the key lemmas (see their Lemmas A.2 and A.3) which are used to prove their main theorems (resp. Theorems 2.1 and 2.2) are written for a general objective function. We show below that, under our maintained assumptions, these lemmas apply in our framework. Accordingly, their Theorems 2.1 and 2.2 hold and we conclude that our suggested bootstrap procedure is first-order valid as stated in Theorem 3.

• Theorem 2.1 in GW(04) follows from applying Lemma A.2 in GW(04) which provides sufficient conditions to ensure the consistency of the extremum estimator and its bootstrap version. The consistency of our SMAS estimator is established in Theorem 1 under Assumptions 1 to 4. The consistency of its bootstrap version is established under three conditions: (b1) measurability of Q_T^*; (b2) continuity of Q_T^* on Θ a.s.-$\mathbb{P}$; (b3) $\sup_{\theta\in\Theta} |Q_T^*(\theta) - Q_T(\theta)| \xrightarrow{P^*}$ in prob-$\mathbb{P}$. (b1) and (b2) follow from Assumptions 2(i) and 4(i); (b3) follows from Assumption 4(ii) and the block bootstrap scheme with $\ell = o(T)$.

• Theorem 2.2 in GW(04) follows from applying Lemma A.3 in GW(04) which provides sufficient conditions to ensure the asymptotic normality of the extremum estimator and its bootstrap version. The asymptotic normality of our SMAS estimator is established under further regularity assumptions including differentiability and functional convergence of the objective function. The formal result is provided in Theorem 2 under Assumptions 1 to 7. The asymptotic normality of its bootstrap version is established under verifying similar regularity assumptions hold on the bootstrap objective function: once again, these follow from our maintained assumptions and the block bootstrap scheme with $\ell = o(\sqrt{T})$. ■

[24]We maintain stationarity and weak dependence in our framework.

• **Proof of Theorem 4:**

We start with three preliminary results: the first one ensures that our simulation-based objective function converges to its population counterpart asymptotically and it is key to our proof. The remaining two establish asymptotic properties of the eigenvalues of the operator and of the operator; they are useful to prove Theorem C.2. All three are formally proved at the end of the proof of Theorem 4.

Theorem C.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 $T^\nu a^{3/2} \to \infty$.*

Theorem C.3. *Under the assumptions of Theorem 4, when $T/N(T) \to \zeta$ as $T \to \infty$ with $0 < \zeta < \infty$, we have:*

$$(\lambda_j^{(T)} - \lambda_j) \quad = \quad \mathcal{O}_{P^*}(\frac{1}{\sqrt{T}}) \text{ in prob-}\mathbb{P}$$

Theorem C.4. *Under the assumptions of Theorem 4, when $T/N(T) \to \zeta$ as $T \to \infty$ with $0 < \zeta < \infty$, we have: $\|K_T - K\| = \mathcal{O}_{P^*}\left(1/\sqrt{T}\right)$ in prob-$\mathbb{P}$.*

The consistency of the estimator (as $T \to \infty$, $a \to 0$, and $\sqrt{T} a^{3/4} \to \infty$) directly follows from Theorem C.2 applied to z_T^* and z under Assumption 9:

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

Following similar steps as those taken in the proof of Theorem 2, we can show that, with $G_T(\theta) \equiv \partial z_T(\theta)/\partial\theta$:

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

where $\overline{\theta}$ lies between θ_0 and $\hat{\theta}_{SMAS}^{opt}$ component by component.

We focus on the RHS term:

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

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:

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

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

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

where the last equality follows from Theorem C.2 - and intermediate results in its proof such as $\| K_{T,a}^{-1} \| \leq 1/\sqrt{a}$ for T large enough - as well as

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

which holds under the uniform convergence results maintained in Assumption 9 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 $\sqrt{T} a^{3/2} \to \infty$:

$$\sqrt{T}(\hat{\theta}_{SMAS}^{opt} - \theta_0) \;\; = \;\; -(K^{-1/2} G(\theta_0), K^{-1/2} G(\theta_0))(K^{-1} G(\theta_0), Z) + o_{\mathbb{P} \times P^*}(1)$$

with $(K^{-1/2} G(\theta_0), K^{-1/2} G(\theta_0))(K^{-1} G(\theta_0), Z) \sim \mathcal{N}(0, \| K^{-1/2} G(\theta_0) \|^{-2})$.

Finally, the optimality of $\hat{\theta}^{opt}_{SMAS}$ 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 their Theorem 8. ■

We now prove the three preliminary results stated at the beginning of the proof.

Proof of Theorem C.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:

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

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

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

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):

$$\left\|K_{T,a}^{-\frac{1}{2}}g - K_a^{-\frac{1}{2}}g\right\| \leq \left\|(K_T^2+aI)^{-\frac{1}{2}}K_T^{\frac{1}{2}}g - (K_T^2+aI)^{-\frac{1}{2}}K^{\frac{1}{2}}g\right\| \tag{B1}$$

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

- 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: $\left\|K_T^{\frac{1}{2}} - K^{\frac{1}{2}}\right\| = \mathcal{O}_{P^*}(\frac{1}{\sqrt{T}})$ in prob-$\mathbb{P}$ which follows from Theorem C.4 and the continuity of the square-root transformation. Hence, overall, (B1) goes to zero as $\sqrt{T}\sqrt{a} \to \infty$.

- 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 C.4, the second term is such that: $\|K_T - K\| = \mathcal{O}_{P^*}(\frac{1}{\sqrt{T}})$ 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 $\sqrt{T}a^{3/4} \to \infty$.

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:

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

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 $\sqrt{T} a^{\frac{3}{4}} \to \infty$. The proof of part (ii) is similar to that of part (i). $\square$

Proof of Theorem C.3:

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 9(i), 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:

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

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

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

with

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

Differentiate equation (C.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[25].
Multiply by $\phi_j(h)$ on both sides, and integrate with respect to h to get:

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

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. $\sum_h \phi_j^2(h) = 1$).

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

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 (C.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:

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

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 (C.1), we have:

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

Under the maintained regularity assumptions, a bootstrap CLT applies to

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

with rate $\sqrt{N(T)}$, and the result follows since $T/N(T) \to \zeta$ some positive constant. $\square$

• **Proof of Theorem C.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 8(i), 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:

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

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 the maintained assumptions, we get, using Theorem C.3 and results established in its proof:

$$\|K_T - K\|^2 = \mathcal{O}_{P^*}(1/T) \text{ in prob-}\mathbb{P}$$

And the expected results follow. $\square$

S.2 Extension

The operator K_T has a degenerate kernel and, contrary to K, a finite dimensional closed range with a finite number of eigenvalues and eigenfunctions (equal to $N(T)$). In Lemma 1, we show how to compute them by solving a linear system of $N(T)$ equations which builds on section 3 in Carrasco and Florens (2000)[26].

Lemma 1. *(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*

$$c_{n,n'} = \frac{1}{N(T)} \sum_h \left(z_T^{*(n)}(h, \theta_0) - \bar{z}_T^*(h, \theta_0) \right) \times \left(z_T^{*(n')}(h, \theta_0) - \bar{z}_T^*(h, \theta_0) \right).$$

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 $j = 1, \ldots, N(T)$,*

$$\phi_j^{(T)}(h) = \frac{1}{N(T)} \left(z_h^{(T)} \right)' \mathcal{B}_j \quad with \quad z_h^{(T)} \equiv \begin{bmatrix} \left(z_T^{*(1)}(h, \theta_0) - \bar{z}_T^*(h, \theta_0) \right) \\ \left(z_T^{*(2)}(h, \theta_0) - \bar{z}_T^*(h, \theta_0) \right) \\ \vdots \\ \left(z_T^{*(N(T))}(h, \theta_0) - \bar{z}_T^*(h, \theta_0) \right) \end{bmatrix}$$

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. Then, the optimal SMAS estimator is obtained as:

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

[26] An alternative approach that does not involve the computation of eigenvalues and eigenfunctions is developed in section 3.3 in Carrasco et al. (2007), which may have computational advantages, particularly in large samples. Since our sample sizes remain small to moderately large in accordance with our macro applications, we do not explore this approach.

• **Proof of Lemma 1:**

In this proof, we simplify our notations as follows, for any integer h:

$$\begin{aligned} z_{T,h}^{*(n)} &\equiv z_T^{*(n)}(h,\theta_0) \\ \bar{z}_{T,h}^{*} &\equiv \bar{z}_T^{*}(h,\theta_0) \end{aligned}$$

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

$$\begin{aligned} & (K_T\phi^{(T)})(h) = \lambda^{(T)}\phi^{(T)}(h) \\ \Leftrightarrow \quad & \frac{1}{N(T)}\sum_{n=1}^{N(T)}\big(z_{T,h}^{*(n)} - \bar{z}_{T,h}^{*}\big)\sum_{s=1}^{\infty}\big(z_{T,s}^{*(n)} - \bar{z}_{T,s}^{*}\big)\phi^{(T)}(s) = \lambda^{(T)}\phi^{(T)}(h)\,, \end{aligned}$$

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

$$\sum_{s=1}^{\infty}\big(z_{T,s}^{*(n)} - \bar{z}_{T,s}^{*}\big)\phi^{(T)}(s)$$

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

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

Overall, we can write:

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

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 $n = 1, \ldots, N(T)$:

$$\frac{1}{N(T)}\sum_{n'=1}^{N(T)}\beta^{n'}\sum_{s=1}^{\infty}\big(z_{T,s}^{*(n')} - \bar{z}_{T,s}^{*}\big)\big(z_{T,s}^{*(n)} - \bar{z}_{T,s}^{*}\big) \quad = \quad \lambda^{(T)}\beta^n\,.$$

And, with the $(N(T), N(T))$-matrix C as defined in Lemma 1, 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 $j = 1, \ldots, N(T)$,

$$\phi_j^{(T)}(h) = \frac{1}{N(T)}\big(z_h^{(T)}\big)'\mathcal{B}_j \quad \text{with} \qquad z_h^{(T)} \equiv \begin{bmatrix} \big(z_{T,h}^{*(1)} - \bar{z}_{T,h}^*\big) \\ \big(z_{T,h}^{*(2)} - \bar{z}_{T,h}^*\big) \\ \vdots \\ \big(z_{T,h}^{*(N(T))} - \bar{z}_{T,h}^*\big) \end{bmatrix}.$$

■

S.3 Medium-scale New Keynesian model

Consider the medium-scale New Keynesian model with seven indices introduced in Section 5.2 of the main text.

In Table 11, we present the estimation results of the degree of price indexation alone when matching either up to $H = 20$ or $H = 80$ with sample size $T = 236$. We report the performance of two SMAS estimators, respectively with the optimal operator and the diagonal one, with regularization in terms of Monte-Carlo average (mean), standard deviation (SD), Root Mean Squared Error (RMSE), Mean Absolute Deviation (MAD), and effective coverage rates of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

Overall, the performance of the optimal SMAS estimator dominates that of the diagonal one; however, the differences remain small.

Matching IR over medium to long horizons on all 7 indices						
	Horizons 1 to 20		Horizons 1 to 80		Horizons 21 to 80	
	optimal	diagonal	optimal	diagonal	optimal	diagonal
mean	0.5078	0.5089	0.5656	0.5702	0.5052	0.5156
MAD	0.0388	0.0399	0.0965	0.1012	0.0362	0.0466
SD	0.2710	0.2780	0.2657	0.2642	0.2778	0.2797
RMSE	0.2738	0.2809	0.2827	0.2829	0.2802	0.2835
Coverage						
95%	100.0	100.0	100.0	100.0	100.0	100.0
90%	94.8	92.9	94.8	92.9	94.6	93.2

Table 11: 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. 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 $T = 232$; we consider VAR(4), $N = 1,000$ replications.

Next, we consider deviations from the model assumptions by redoing our simulation study when the error terms in the real DGP are not normally distributed, but rather t-distributed - with either 4 or 20 degrees of freedom. Everything else remains exactly as before. In Table 12, we present the estimation results of the Calvo parameter when matching either up to $H = 20$ or $H = 80$ with sample size either $T = 236$ or $T = 944$. We report the performance of two SMAS estimators, respectively with the optimal operator and the diagonal one, with and without regularization in terms of Monte-Carlo average (mean), standard deviation (SD), Root Mean Squared Error (RMSE), Mean Absolute Deviation (MAD), and effective coverage rates of 95% and 90% confidence intervals obtained over 1,000 Monte-Carlo replications.

Overall, the results are on-par with those obtained in Table 7 in the main paper. In particular, we notice that the performance of the regularized SMAS estimator is very good and dominates that of the non-regularized one according to all reported measures of performance. We do notice some small size distortions, which tend to disappear when the sample size increases. Focusing now of the performance of regularized optimal SMAS, it is better than that of diagonal SMAS as it delivers smaller SD and RMSE, and improved coverage rates - especially with $T = 236$. In addition, its performance does improve when adding IR at long horizons according to all measures of performance. Finally, when considering deviations from normality that are more severe (e.g. with t(4)), its performance deteriorates slightly with increased MAD and SD.

All in all, and similar to the results obtained in the small-scale model, our analysis suggests that the SMAS estimator can indeed withstand deviations from the model assumptions, including deviations in the distributions of the error terms.

Panel A: the errors are t(20) and $H = 20$								
	$T = 236$				$T = 944$			
	Diagonal		Optimal		Diagonal		Optimal	
	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.
mean	0.7105	0.6736	0.7634	0.7402	0.8396	0.7961	0.8572	0.8024
MAD	0.1975	0.2344	0.1446	0.1678	0.0684	0.1119	0.0508	0.1056
SD	0.2248	0.2425	0.2108	0.2251	0.1493	0.1940	0.1354	0.1908
RMSE	0.2992	0.3373	0.2556	0.2807	0.1642	0.2240	0.1446	0.2180
Coverage								
95%	86.3	82.7	89.7	88.2	94.7	91.5	93.9	92.5
90%	85.8	80.7	88.5	86.2	90.2	90.4	90.8	91.7
Panel B: the errors are t(20) and $H = 80$								
	$T = 236$				$T = 944$			
	Diagonal		Optimal		Diagonal		Optimal	
	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.
mean	0.6858	0.6536	0.7604	0.7387	0.8271	0.7790	0.8476	0.7913
MAD	0.2222	0.2544	0.1476	0.1693	0.0809	0.1290	0.0604	0.1167
SD	0.2272	0.2410	0.1955	0.2199	0.1558	0.1983	0.1414	0.1998
RMSE	0.3178	0.3504	0.2450	0.2776	0.1756	0.2366	0.1537	0.2314
Coverage								
95%	83.2	80.4	90.6	88.2	95.2	91.3	92.7	92.2
90%	82.0	77.2	90.2	86.6	90.1	89.2	90.0	91.6
Panel C: the errors are t(4) and $H = 20$								
	$T = 236$				$T = 944$			
	Diagonal		Optimal		Diagonal		Optimal	
	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.
mean	0.6994	0.6684	0.7528	0.7150	0.8267	0.7849	0.8526	0.7887
MAD	0.2086	0.2396	0.1552	0.1930	0.0813	0.1231	0.0554	0.1193
SD	0.2201	0.2312	0.2120	0.2428	0.1621	0.2062	0.1375	0.1954
RMSE	0.3033	0.3330	0.2627	0.3101	0.1814	0.2401	0.1482	0.2289
Coverage								
95%	85.9	81.8	88.3	84.5	94.8	90.7	93.3	92.0
90%	83.8	80.7	86.0	82.3	90.1	89.5	90.2	90.7
Panel D: the errors are t(4) and $H = 80$								
	$T = 236$				$T = 944$			
	Diagonal		Optimal		Diagonal		Optimal	
	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.	reg.	non-reg.
mean	0.6853	0.6530	0.7570	0.7292	0.8245	0.7792	0.8456	0.7820
MAD	0.2227	0.2550	0.1510	0.1788	0.0835	0.1288	0.0624	0.1260
SD	0.2245	0.2450	0.1950	0.2225	0.1580	0.1942	0.1459	0.2018
RMSE	0.3162	0.3536	0.2467	0.2855	0.1787	0.2330	0.1587	0.2379
Coverage								
95%	84.8	80.4	91.0	87.4	95.8	91.0	94.9	90.8
90%	83.4	77.0	89.8	86.4	90.3	90.2	91.3	90.1

Table 12: Estimation of the Calvo parameter $\zeta_p = 0.908$ in the medium-scale model when the error terms are t-distributed, either as t(20) in Panels A and B, or t(4) in Panels C and D. Performance of SMAS estimators (with and without regularization) when matching impulse responses over medium and long horizons in all indices with sample size $T = 236$ and 944. We fit VAR(4) and consider $N = 1,000$ replications.

S.4 Estimation of the asymptotic variance

We now present new simulation results to illustrate the validity of Theorem 2 in the main text - especially the accuracy of finite sample standard errors and their convergence towards their asymptotic equivalents. Specifically, we compute finite sample standard errors and their asymptotic equivalents in the context of our benchmark (small-scale) DSGE model presented in Section 5.1 in the main text.

We focus on the baseline experiment highlighted in Section 5.1 in the main text. We estimate one unknown parameter denoted α, the probability of the price stickiness. We fit a VAR(2) model to the sample of observations, match IR up to horizon $H = 80$, and consider 500 Monte-Carlo replications. As proven in Theorem 2 in the main text, the asymptotic variance of the SMAS estimator (computed with operator B) of α is a real number defined as,

$$\|BG(\alpha_0)\|^{-2}\,,$$

when B is chosen as the optimal operator, and defined as,

$$\|BG(\alpha_0)\|^{-2}(BG(\alpha_0),(BKB')BG(\alpha_0))\|BG(\alpha_0)\|^{-2}\,,$$

otherwise. The function $G(.)$ - the derivative of the distance function $z(.)$ with respect to α - is approximated numerically as follows,

$$G_e(\alpha) = \frac{z(\alpha+e) - z(\alpha-e)}{2e}\,,$$

with e a small (positive) constant. In practice, we choose e equal to the increment of the grid of candidate values used to solve the optimization problem: that is, $e = 0.005$.

The results are presented in Table 13 where we report the percentage difference between the finite sample variance (denoted $\hat{\sigma}^2$) of the SMAS estimator $\hat{\alpha}$ and its asymptotic equivalent (denoted σ_0^2), that is:

$$\frac{|(\sigma_0^2 - \hat{\sigma}^2)|}{\sigma_0^2} \times 100\,.$$

We consider different sample sizes that are comparable to the ones used in the Monte-Carlo study conducted in the main text: that is, $T = 116$, 232 and 464. We report the percentage differences in the variances of two SMAS estimators, the optimal one and the diagonal one. Overall, the differences between the finite sample variance and its asymptotic equivalent can be large, especially for smaller sample sizes: for example,

differences up to 19% for the smallest sample size, $T = 116$, for both optimal and diagonal SMAS estimators. That being said, these differences tend to decrease when the sample size increases: for example, this difference drops to 7.6% when $T = 464$ for the optimal SMAS, and to 10.6% for the diagonal SMAS.

Matching IR over long horizons			
Sample size T	116	232	464
% difference with optimal operator	18.94	18.06	7.64
% difference with diagonal operator	18.84	19.00	10.60

Table 13: Estimation of the price indexation $\alpha = 0.75$ in the small-scale model when matching IR up to $H = 80$. We report the percentage difference between the finite sample variance of the SMAS estimator and its asymptotic equivalent for samples of size $T = 116$, 232, and 464. The SMAS estimator is either computed with the optimal operator or the diagonal one. We fit VAR(2), and consider 500 replications.

In practice, we do not recommend the above-mentioned direct estimation of the asymptotic variance for two main reasons. First, it depends on the derivative of a function that is not known in closed-form. This is actually related to our framework of interest where auxiliary statistics are not known in closed-form. Second, even in a simpler (or more standard) framework, it is well-documented that the asymptotic distribution of standard minimum distance estimators such as GMM or Indirect Inference may not accurately reflect finite-sample performance. Further, it is not only about the estimation of the asymptotic variance, but also about deviations from the true value. Ultimately, what matters is that estimated variances appropriately *adjust* to deviations from the true value to ensure associated confidence intervals and tests deliver reliable inference on the parameter of interest - e.g. without severe size distortions. And this is what we have focused on in the Monte-Carlo experiments presented in the main text. To this end, we propose a bootstrap approach. Theorem 2 (and Theorem 4) in the main text establish asymptotic properties of the SMAS estimator that are key to validate our bootstrap approach.