

Limited-Information Estimation of Heterogeneous Agent Models*

Laura Liu

University of Pittsburgh

Mikkel Plagborg-Møller

University of Chicago

Nelson Matthew P. Tan

Princeton University

August 13, 2026

Abstract: We develop a method for estimating and testing a single block of a macroeconomic model with heterogeneous agents, without placing assumptions on the structure of the rest of the economy. In a large class of models, individual agents’ decisions depend on the macroeconomy only through their expectations of the evolution of a finite-dimensional vector of “sufficient statistics” (e.g., asset returns or aggregate earnings). Our estimator selects the structural parameters that provide the best model-consistent fit between empirical impulse responses with respect to identified macro shocks of (a) cross-sectional moments of agent choices (e.g., moments of consumption) and (b) the vector of sufficient statistics. In a simulation illustration, we estimate a two-asset heterogeneous household model block without restricting production, firm investment, financial intermediation, monetary policy, trade, etc.

Keywords: heterogeneous agents, incomplete model, structural estimation

*Email: laura.liu@pitt.edu, mikkelpm@uchicago.edu, nt1847@princeton.edu. We are grateful for comments from Stéphane Bonhomme, Mikhail Golosov, Greg Kaplan, Christian Wolf, and seminar participants at Chicago, FRB Richmond, and Princeton. Plagborg-Møller acknowledges that this material is based upon work supported by the NSF under Grant #2238049 and by the Alfred P. Sloan Foundation.

1 Introduction

In recent years, there has been extensive research on the consequences of household and firm heterogeneity for macroeconomic dynamics (Kaplan, Moll, and Violante, 2018; Auclert, Rognlie, and Straub, 2025). Models with heterogeneous agents (HAs) promise not only to fit the micro data better but also to shed new light on transmission channels and policy trade-offs. Estimation of these models has proven to be challenging due to their computational complexity and the statistical difficulties involved in jointly analyzing micro and macro data. For this reason, much of the applied literature selects structural parameters by calibrating to a set of moments, typically chosen heuristically. Nevertheless, the last decade has seen continued progress in this area, and there now exist rigorous inference methods that efficiently exploit time series of macro variables and cross-sectional or panel moments, or indeed the full information content of repeated cross sections.¹

Whether matching an *ad hoc* set of moments or using full-information likelihood procedures, a familiar issue in structural estimation is that misspecification of one part of the model can contaminate the estimation of parameters belonging to other parts of the model. For example, even if a researcher is primarily interested in the consumption block of a Heterogeneous Agent New Keynesian (HANK) model, they still have to specify the full structure of the rest of the economy in order to compute arbitrary moments or the likelihood function implied by general equilibrium.

In representative agent models, there exists a well-known robust alternative to full-information inference: Generalized Method of Moments (GMM) estimation of the optimality conditions associated with a single model block, such as the consumption Euler equation (Hansen and Singleton, 1982; Yogo, 2004) or the New Keynesian Phillips Curve (Galí and Gertler, 1999). However, it is difficult to derive empirically usable, model-consistent moment conditions in typical HA macro models, since agents' decision rules are subject to occasionally binding constraints and depend on a mix of micro and macro variables, some of which are latent.

In this paper we develop a general method for estimating and testing a single block of a HA model in isolation, without restricting other parts of the model. We exploit a well-

¹Examples of the former include Mongey and Williams (2017); Winberry (2018); Papp and Reiter (2020); Auclert, Bardóczy, Rognlie, and Straub (2021); Acharya, Chen, Del Negro, Dogra, Gleich, Goyal, Matlin, Lee, Sarfati, and Sengupta (2023); Bayer, Born, and Luettticke (2024). Examples of the latter include Fernández-Villaverde, Hurtado, and Nuño (2023); Liu and Plagborg-Møller (2023).

known feature of conventional HA macro models: individual agents’ decisions depend on the macroeconomy only through a finite-dimensional vector of aggregate “sufficient statistics”, typically prices such as asset returns or wages.² The “sequence-space Jacobians” (SSJs), introduced by Boppart, Krusell, and Mitman (2018) and Auclert, Bardóczy, Rognlie, and Straub (2021, henceforth abbreviated “ABRS”), are the theoretical objects that measure how aggregates of individual decisions depend on current and future expected values of the macro sufficient statistics. These SSJs are functions of the underlying structural parameters of the model block under estimation, but they do not depend on other unmodeled features of the general equilibrium. Our estimator selects the structural parameters of the model block of interest so as to provide the best fit between, on the one hand, the empirical impulse responses of aggregated individual outcomes (such as cross-sectional moments of consumption) and, on the other hand, the combination of theoretical SSJs and empirical impulse responses of the sufficient statistics. Our estimator can therefore be viewed as a nonlinear version of “regression in impulse response space” (Barnichon and Mesters, 2020), applicable to general HA models. Given the estimated parameters, we can also carry out an over-identification test of the validity of the assumptions in the single model block.

The conceptual novelty of our approach is that we do not impose a specific process for the evolution of the macro sufficient statistics. Existing full-information estimation procedures derive the functional form of this process from the structural assumptions placed on the rest of the economy, such as the firm, finance, international trade, and government blocks. If some of these extraneous assumptions are wrong, all elements of the full-information estimator could be biased. Adopting a reduced-form model for the dynamics of the sufficient statistics does not easily solve this problem, since such a model could theoretically depend on lags of the entire infinite-dimensional distribution of idiosyncratic state variables, in addition to (latent) macro state variables. Our procedure sidesteps the issue by merely estimating the impulse responses of the sufficient statistics with respect to a set of observed economic shocks, without requiring a full specification of the transmission of every possible shock. Focusing on these particular moments allows our procedure to be fully consistent with general equilibrium, without having to impose a full equilibrium model (in the spirit of Hansen and Singleton, 1982).

Our ability to relax the modeling assumptions required by full-information estimation procedures relies on having access to time series data on the entire vector of macro sufficient

²This fact has also been instrumental to the work of Auclert, Bardóczy, Rognlie, and Straub (2021); McKay and Wolf (2023); Pallotti, Paz-Pardo, Slacalek, Tristani, and Violante (2024).

statistics for the model block of interest, as well as a set of identified economic shocks to serve as instruments (with respect to which the impulse responses are computed). It is straightforward to obtain data on the prices and other aggregates in the vector of sufficient statistics for many existing HA models. As for instruments, researchers can exploit the menu of existing identified shocks in the applied literature (Ramey, 2016; Barnichon and Mesters, 2020). The instruments must induce sufficient dynamic variation in the macro sufficient statistics to ensure identification of the structural parameters. Moreover, they must be actual “shocks” in the sense of being orthogonal to any lagged macro variables, and in particular to the pre-determined distribution of idiosyncratic state variables. Conveniently, we do *not* need the identified shocks to have a unique economic interpretation, since any functions of multiple contemporaneous shocks can equally well serve as instruments for our moment conditions. For example, we can use high-frequency monetary shocks as instruments even if these are in fact an amalgam of true interest rate surprises and information shocks (Bauer and Swanson, 2023).

Our method allows researchers to freely choose which cross-sectional moments of agents’ choices to fit. These moments could be simple, such as the mean, variance, or interquartile range of the consumption distribution, or more complicated, such as the share of illiquid savings held by households above the 90th percentile of the wealth distribution.

Of course, by focusing on a single model block, our method does not permit estimation of arbitrary general equilibrium counterfactuals; only those that are functions of parameters within the estimated block. This may suffice for some questions: in some HANK models the steady-state Lorenz curve for wealth, say, is a function only of parameters in the household block. Moreover, our method allows researchers to test the validity of individual blocks one by one. Finally, the limited-information parameter estimates can serve as a robust benchmark for subsequent full-model calibrations.

In addition to the assumptions already discussed, our method shares the following requirements with existing full-information estimation procedures: the model block of interest must be correctly specified (this assumption is testable within our framework); the linearization of micro aggregates with respect to macro variables must be accurate, as in Reiter (2009) and ABRS; and agents must have rational expectations. For computational feasibility, we rely on the rapid and numerically stable software for computing SSJs developed by ABRS.³

We illustrate the empirical usefulness of our approach by applying it to data simulated from an empirically calibrated two-asset HANK model (Kaplan, Moll, and Violante, 2018;

³Specifically, we use their excellent `sequence-jacobian` Python package off the shelf.

ABRS). The estimation procedure is fed time series data on portfolio returns, aggregate earnings, moments of the cross-sectional distributions of household consumption and savings, and monetary and fiscal shocks. The procedure exploits only the structure of the model’s heterogeneous household block, up to several unknown parameters; it does not place any restrictions on production, corporate investment, financial intermediation, monetary policy, trade, etc. We estimate parameters related to household preferences and portfolio adjustment costs. Even in a moderate sample size of 120 quarters, the parameter estimators have only modest mean squared error, and the bootstrap confidence intervals and model specification test have accurate coverage/size.

Econometrically, we express our estimator as a minimum distance procedure in the frequency domain. This facilitates econometric analysis of the infinite decision horizon of the agents in the model, and it also suggests a natural and effective bootstrap procedure. We prove consistency and asymptotic normality for a general class of “spectral GMM” estimators under weak assumptions on the time series dependence of the data. Exploiting the bilinear form of our minimum distance moment function, we are able to relax the existing assumptions in the literature used to prove stochastic equicontinuity of the relevant empirical spectral means.

LITERATURE. Our approach is most closely related to the moment-matching and likelihood estimation procedures for HA models cited earlier. All these procedures require a full specification of general equilibrium. We drop this requirement, at the expense of having to observe the macro sufficient statistics and a set of identified shocks. While superficially similar, impulse response matching in the tradition of Rotemberg and Woodford (1997) generally requires a full equilibrium specification (though non-matched shocks need not be specified); our strategy instead considers a *combination* of impulse responses that is provably invariant with respect to the auxiliary model blocks.

Though the classic literature on estimating consumption Euler equations from micro data has wrestled with challenges arising from borrowing constraints (e.g., Deaton, 1985; Hayashi, 1985; Carroll, 2001), none of those estimation procedures appear to be valid under the kinds of data generating processes considered in the recent HANK literature. By contrast, our method can handle a general class of models with essential nonlinearities due to occasionally binding constraints, possibly latent micro and macro state variables, and portfolio adjustment costs. However, unlike the earlier literature, our method’s ability to exploit micro panel data is limited (see Section 3.6).

Our paper is concerned with structural estimation of a HA model block, unlike the semi-structural modeling approaches of [Chang, Chen, and Schorfheide \(2024\)](#); [Ettmeier, Kim, and Schorfheide \(2024\)](#); [Almuzara, Arellano, Blundell, and Bonhomme \(2025\)](#); [Chang, Kim, and Park \(2025\)](#). In these papers, the economic structure is primarily used to identify shocks and characterize the income process, rather than to estimate the deep structural parameters. By contrast, we estimate these parameters directly. The benefits of structural estimation are well known: parameters have deep economic meaning, we can compute well-defined counterfactuals, and we can directly test the validity of concrete theoretical mechanisms. The downside is that our estimation results are sensitive to misspecification of the model block of interest (but not to misspecification of other aspects of the general equilibrium).

A few papers have taken application-specific approaches that are similar in spirit to ours. [Gagliardone, Gertler, Lenzu, and Tielens \(2025\)](#) estimate the New Keynesian Phillips Curve from individual-firm optimality conditions, exploiting data on the “sufficient statistics” for the firm, namely marginal costs. [Debortoli and Galí \(2025\)](#) estimate linear consumption Euler equations motivated by HA models with hand-to-mouth households. [Tryphonides \(2023\)](#) partially identifies the preference parameters in a HA consumption block by deriving model-implied moment inequalities; these require particular micro variables to be observed by the econometrician. Unlike these papers, our framework covers a general class of HA models with potentially non-smooth individual decision rules, and all macro and micro variables in the decision problem are allowed to be latent except the vector of aggregate sufficient statistics.

OUTLINE. [Section 2](#) gives an overview of our procedures in the context of estimating a heterogeneous household model block. [Section 3](#) defines our general estimation and inference procedures, and [Section 4](#) states formal econometric results on their asymptotic properties. [Section 5](#) applies our methods to data simulated from a two-asset HANK model. [Section 6](#) concludes. Proofs and technical details are relegated to the appendix.

2 Motivating example

Here we describe the intuition behind our estimation approach in the context of the model that we will use in our empirically calibrated simulation study in [Section 5](#): a two-asset heterogeneous household model block.

2.1 Two-asset heterogeneous household model

We consider the household block of the two-asset HANK model of [ABRS](#) (Appendix B.3), which in turn is closely related to the model of [Kaplan, Moll, and Violante \(2018\)](#). There is a continuum of households $i \in [0, 1]$, each of which allocates their savings at discrete time t between liquid assets $b_{i,t}$ and illiquid assets $a_{i,t}$. Changes in illiquid asset holdings incur a convex portfolio cost $\Psi(a_{i,t}, (1 + r_t^a)a_{i,t-1})$. Hours worked N_t is uniform across households and determined by a labor union that is outside this model block and thus unrestricted. The household solves the problem

$$\max_{\{c_{i,t}, b_{i,t}, a_{i,t}\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_{i,t}^{1-1/\gamma}}{1-1/\gamma}$$

$$\begin{aligned} \text{s.t.} \quad & c_{i,t} + a_{i,t} + b_{i,t} = (1 - \tau_t)w_t N_t e_{i,t} + (1 + r_t^b)b_{i,t-1} + (1 + r_t^a)a_{i,t-1} - \Psi(a_{i,t}, (1 + r_t^a)a_{i,t-1}), \\ & a_{i,t} \geq 0, \quad b_{i,t} \geq \underline{b}. \end{aligned}$$

Here E_0 denotes the expectation at time 0, β is the discount factor, γ is the elasticity of intertemporal substitution (EIS), τ_t is the labor tax, w_t is the aggregate real wage, r_t^b and r_t^a are the real returns on liquid and illiquid wealth, respectively, and $\underline{b}$ is the borrowing constraint for liquid assets. $e_{i,t}$ is an idiosyncratic household productivity level, which evolves according to a Markov process, independently of any macro shocks.

Note that in this model, households' optimal choices depend on the macroeconomy only through three aggregate variables: after-tax labor earnings $(1 - \tau_t)w_t N_t$, and the two returns r_t^b, r_t^a . We collect these three variables in the vector $x_t \equiv ((1 - \tau_t)w_t N_t, r_t^b, r_t^a)'$ of macro "sufficient statistics". The household's time- t policy functions for consumption and asset allocations are functions of the current idiosyncratic states $(b_{i,t-1}, a_{i,t-1}, e_{i,t})$ as well as the household's beliefs about all the current and future values of the sufficient statistics: $(x_t, x_{t+1}, x_{t+2}, \dots)$. Crucially—and unlike existing full-information estimation procedures—we do not place any functional form restrictions on the dynamics of x_t .

We seek to estimate the vector θ of structural parameters for this model block. The vector includes the discount factor β , the EIS γ , any parameters that enter into the portfolio adjustment cost function Ψ , and any parameters that govern the dynamics of the idiosyncratic exogenous state $e_{i,t}$.

2.2 Sequence-space representation

Having defined the household model block, we next describe the sequence-space representation of the aggregated household decisions. This is the most important step in deriving the moment conditions that we will take to the data.

Let $y_{i,t}$ be a vector of “outputs” of the household’s decision problem, that is, functions of its choice variables $(c_{i,t}, b_{i,t}, a_{i,t})$, idiosyncratic states $(b_{i,t-1}, a_{i,t-1}, e_{i,t})$, and macro variables x_t . Suppose the econometrician observes the cross-sectional aggregate vector $y_t = \int_0^1 y_{i,t} di$. For example, if $y_{i,t} = (c_{i,t}, c_{i,t}^2)'$, then y_t equals the cross-sectional mean and second moment of consumption, which may be measured empirically using repeated cross sections from a household expenditure survey (measurement error is discussed below). Assume that all households share the same beliefs about the evolution of the macro sufficient statistics x_t . Then the aggregated outputs y_t are functions of two objects: households’ beliefs about $(x_t, x_{t+1}, x_{t+2}, \dots)$, and the pre-determined cross-sectional distribution μ_{t-1} of idiosyncratic states $(b_{i,t-1}, a_{i,t-1}, e_{i,t})$.⁴ Both these objects are infinite-dimensional.

While the fully nonlinear solution to the model block is computationally intractable, linearizing with respect to the macro variables restores tractability. This linearization approach, which preserves the full *micro* heterogeneity, follows [Reiter \(2009\)](#) and a large subsequent literature, and it has been the basis of most full-information estimation procedures for HA models. As in the general frameworks of [Boppart, Krusell, and Mitman \(2018\)](#) and [ABRS](#), linearization of the aggregate household outputs y_t with respect to x_t around a deterministic steady state (denoted with superscript “ss”) yields an approximately linear *sequence-space* representation

$$y_t - y^{\text{ss}} \approx E_t \left[\sum_{k=0}^{\infty} J_k(\theta)(x_{t+k} - x^{\text{ss}}) \right] + \xi_{t-1}(\theta). \quad (1)$$

Here E_t is the households’ expectation operator at time t , and $\xi_{t-1}(\theta)$ is a pre-determined term capturing the effects of the cross-sectional state distribution μ_{t-1} on current household choices. The sequence $\{J_k(\theta)\}_{k=0}^{\infty}$ of parameter-dependent matrices is called the *sequence-space Jacobians* (SSJs). These are the central ingredients in our estimation procedure, as they measure how the future expected time paths of macro sufficient statistics translate into aggregated household choices.⁵ [ABRS](#) developed a “fake news algorithm” which rapidly and

⁴The distribution is pre-determined in the sense that it only depends on macro shocks up to date $t - 1$.

⁵Note that idiosyncratic risk and precautionary behavior are embedded in the steady-state distribution and SSJs. Aggregate risk is not captured directly at first order. Incorporating aggregate risk effects would generally require a higher-order expansion and is beyond the scope of this paper.

reliably computes the whole sequence of SSJs for any given structural parameters θ . In practice, their algorithm truncates the households' decision horizon at a finite number H , yielding a finite sum in (1). H can be taken to be on the order of 500 to minimize truncation bias.

While the representation (1) is merely a first-order Taylor expansion that leaves out terms of order 2 and higher in the macro variables, we follow most of the literature in ignoring the approximation error. As argued originally by [Reiter \(2009\)](#), the linearization exploits the fact that, whereas individual-household decision rules are non-smooth functions of their own idiosyncratic states as well as of the macro sufficient statistics, *integrals* of these choices across agents are often smooth functions of the sufficient statistics. If the integrated choices are smooth functions of $\{E_t[x_{t+k}]\}_k$, and the underlying macro shocks driving x_t are not very large, the first-order linearization error will be negligible. We therefore henceforth treat (1) as an exact equality, as in [ABRS](#) as well as the voluminous literature on estimating representative agent models via the Kalman filter likelihood.

Despite having access to an efficient numerical algorithm for computing the SSJs, there are two challenges involved in taking representation (1) to the data. First, in an ideal setting, we would observe exogenous variation in the household expectations $E_t[x_{t+k}]$ at all future horizons $k \geq 1$, in which case we could estimate θ through nonlinear least squares. However, in practice we do not directly observe households' expectations at all forecast horizons.⁶ Second, the pre-determined term $\xi_{t-1}(\theta)$ is generally a complicated function of the cross-sectional state distribution μ_{t-1} and is therefore not directly observable either; yet it is likely correlated with the macro sufficient statistics x_t and so cannot be treated as a harmless error term. Our approach addresses both challenges by using identified shocks as instruments.

The full-information estimation method of [ABRS](#) goes on to specify a complete equilibrium structure and a full set of shock processes. This allows those authors to compute the SSJs of all endogenous macro variables with respect to all exogenous shocks, from which they can derive the likelihood function for any vector of observed macro time series in the model. The full-information approach, while highly informative when correctly specified, implicitly places strong functional form restrictions on the dynamics of the macro sufficient statistics x_t . If we only care about estimating and testing the household model block, we now show

⁶In representative agent models, it is often possible to derive estimating equations that involve only the one-step-ahead expectation (e.g., the conventional consumption Euler equation or New Keynesian Phillips Curve). One could then directly measure this expectation using survey data (e.g., [Fuhrer, 2017](#)). However, we are not aware of a method that can be applied to general HA models that would eliminate expectations at intermediate and long horizons from the households' optimality conditions.

that we can avoid (a) placing any functional form restrictions on the dynamics of x_t and (b) specifying the full list of shocks driving the data.

2.3 Shocks as instruments

We are able to derive empirically useful limited-information moment conditions if we observe an appropriate set of instruments and impose rational expectations. Suppose we have access to a vector z_t of instruments satisfying the following two conditions:

1. The instruments are orthogonal to past shocks, and therefore orthogonal to the pre-determined term:⁷

$$\text{Cov}(z_t, \xi_{t-1}(\theta)) = 0. \quad (2)$$

2. The instruments z_t do not predict households' forecast errors at any horizon:

$$\text{Cov}(x_{t+k} - E_t[x_{t+k}], z_t) = 0 \quad \text{for } k = 0, 1, 2, \dots \quad (3)$$

A set of sufficient conditions for this second assumption is that (i) households have rational expectations and (ii) z_t is contained in the households' time- t information set. These conditions are often imposed in full-information estimation.

Combining assumptions (2)–(3) with representation (1) (viewed as an equality), we obtain a system of moment conditions:

$$\text{Cov}(y_t, z_t) = \sum_{k=0}^{\infty} J_k(\theta) \text{Cov}(x_{t+k}, z_t). \quad (4)$$

Assuming that (y_t, x_t, z_t) are all observed, this is a system of instrumental variable exogeneity conditions, but with two non-standard features: the infinite horizon of the sum, and the nonlinearity of the coefficient matrices $J_k(\theta)$ in the structural parameters θ .

Identified economic shocks are prime candidates for instruments z_t . A wide range of shock series are available in the applied literature, such as monetary shocks, fiscal shocks, oil shocks, and technology shocks; see the lists compiled by [Ramey \(2016\)](#) and [Barnichon and Mesters \(2020\)](#). By virtue of being “shocks”, these series ought to be unpredictable and thus satisfy the first assumption (2) on the instruments. Moreover, if the shocks are economically important, they are also likely to be observed by households, thus satisfying

⁷In the first-order linearization, $\xi_{t-1}(\theta)$ is linear in past shocks.

the second assumption (3) on the instruments under rational expectations. Unlike in the case of full-information likelihood estimation, (a) we do not need to take a stand on which and how many shocks are driving the data, and (b) our moment conditions (4) allow the observed data to be contaminated by a general class of measurement error processes.⁸

Conveniently, the instruments continue to satisfy our assumptions if they are given by *functions* of several true underlying economic shocks, i.e., $z_t = \text{fct}(\varepsilon_{1t}, \varepsilon_{2t}, \dots)$. For example, even if a measured “monetary shock” instrument is actually an amalgam of a pure interest rate surprise with other kinds of shocks (such as information shocks), it will still satisfy our assumptions.

Assuming valid instruments, identification of the structural parameters requires the moment conditions (4) to have a unique solution θ . It is necessary that the total number of moments $\dim(y_t) \times \dim(z_t)$ weakly exceed the number of parameters $\dim(\theta)$. Additionally, the instruments z_t must induce sufficient variation in current and future sufficient statistics x_{t+k} in order for the right-hand side of (4) to be a non-trivial function of the structural parameters θ —at a minimum, z_t must be correlated with x_{t+k} at *some* horizon k , an easily testable condition. As a heuristic example, if θ includes the household discount factor, z_t must co-vary with x_{t+k} at relatively large horizons k , so as to be able to distinguish the choices y_t made by patient and impatient households. We give precise conditions for identification in Section 4.2 below.

2.4 Estimator and over-identification test

Given a data set $\{y_t, x_t, z_t\}_{t=1}^T$, our estimator $\hat{\theta}$ of the structural parameters satisfies the sample analogue of the population moment conditions (4) as well as possible:

$$\widehat{\text{Cov}}(y_t, z_t) \approx \sum_{k=0}^{T-1} J_k(\hat{\theta}) \widehat{\text{Cov}}(x_{t+k}, z_t), \quad (5)$$

where “ $\widehat{\text{Cov}}$ ” denotes sample covariance. Notice that we truncate the sum at horizon $k = T-1$, since the data is uninformative about covariances at longer horizons.⁹ When the model is over-identified—i.e., $\dim(y_t) \times \dim(z_t) > \dim(\theta)$ —it is typically impossible to satisfy the

⁸Specifically, the moment conditions continue to hold for the contaminated processes $y_t^{\text{obs}} = y_t + \epsilon_t^y$, $x_t^{\text{obs}} = x_t + \epsilon_t^x$, $z_t^{\text{obs}} = z_t + \epsilon_t^z$, provided that (a) $\{z_t\}$ is dynamically uncorrelated with $\{\epsilon_t^y, \epsilon_t^x\}$ and (b) $\{\epsilon_t^z\}$ is dynamically uncorrelated with $\{y_t, x_t, \epsilon_t^y, \epsilon_t^x\}$.

⁹As previously discussed, the SSJs $J_k(\theta)$ are also truncated at a large horizon $k \leq H$ for computational feasibility, but it is computationally feasible to choose $H \gg T$.

above system of equations with equality in a finite sample, so we define $\hat{\theta}$ as a minimum distance estimator that minimizes a weighted quadratic form in the discrepancies between the left-hand and right-hand sides; see [Section 3.2](#) for details.

The sample estimating equation (5) shows that our estimation approach can be viewed as a nonlinear “regression in impulse response space”, using the SSJs as the key link between data and model. The left-hand side of (5) is the contemporaneous empirical impulse response of the aggregate household outputs y_t with respect to the instruments z_t , while the right-hand side involves the empirical impulse responses $\widehat{\text{Cov}}(x_{t+k}, z_t)$ of the macro sufficient statistics with respect to the instruments at all horizons $k = 0, 1, \dots, T-1$. This estimation approach, which is applicable to general nonlinear HA models, builds on the linear, representative agent approach of [Barnichon and Mesters \(2020\)](#).

In over-identified settings, we can test whether the household model block is correctly specified. This can be done by checking whether a single parameter vector $\hat{\theta}$ can approximately satisfy the entire system of sample estimating equations (5). More specifically, we carry out a conventional over-identification test for minimum distance estimators; see [Section 3.4](#) below.

2.5 Summary and outlook

To recap, our estimation procedure is a nonlinear regression in impulse response function space that exploits SSJs as the key theoretical link between (a) the responses of macro sufficient statistics and (b) the responses of aggregated household choices. For computational feasibility, our method relies on existing software for computing SSJs ([ABRS](#)). For econometric validity, our method relies on five key assumptions. First, the household block is correctly specified (this is testable in the over-identified case). Second, the household decision problem depends on the macroeconomy only through a finite-dimensional vector of “sufficient statistics”. Third, the approximation error from linearizing the aggregated optimal choices of households with respect to macro variables is negligible. Fourth, households have rational expectations. Fifth, we observe time series data on all the macro sufficient statistics as well as a vector of instruments, with the latter being functions of contemporaneous shocks. The first four of these assumptions are shared by most existing full-information estimation methods.

Our method gives the researcher flexibility to choose the moments y_t of households’ choices (e.g., cross-sectional moments of consumption and savings) and the instruments z_t (e.g., identified shocks). It is not necessary to obtain data on any of the households’

idiosyncratic state variables—if such data happened to be available, it could be incorporated in y_t .

In the following sections we will argue that the above estimation and testing approach is applicable to a large class of HA model blocks, including blocks with heterogeneous firms, financial intermediaries, etc. Moreover, we will provide general identification results, prove the consistency and asymptotic normality of the minimum distance estimator $\hat{\theta}$ under weak conditions, and propose analytical and bootstrap procedures for computing standard errors and confidence intervals. The econometric analysis is non-trivial as we allow for general serial correlation patterns in the data and explicitly handle the error imparted by the truncation of the infinite-horizon sum in the population moment conditions (4).

3 General framework

Having explained the intuition behind our approach, we now formally define our general framework and econometric procedures.

3.1 Moment conditions

We take the moment conditions (4) derived in the previous section as the starting point for our general analysis. We restate those conditions below, using the notation θ_0 to denote the true value of the structural parameter vector.

Assumption 1 (Moment conditions). *There exist a sequence of deterministic, matrix-valued functions $J_k: \Theta \rightarrow \mathbb{R}^{d_y \times d_x}$ for $k = 0, 1, \dots$ and a parameter vector $\theta_0 \in \Theta \subset \mathbb{R}^{d_\theta}$ such that*

$$\text{Cov}(y_t, z_t) = \sum_{k=0}^{\infty} J_k(\theta_0) \text{Cov}(x_{t+k}, z_t),$$

where y_t, x_t, z_t are jointly stationary random vectors of dimensions d_y, d_x, d_z , respectively.

The theoretical analysis in [Section 4](#) will impose further assumptions on the SSJs and the data to ensure that both sides of the above equation are mathematically well-defined.

We derived moment conditions of the above form for a particular heterogeneous household model in [Section 2](#), but it is clear from our arguments that such conditions will equally well obtain for any other HA model block covered by the general framework of [ABRS](#). The only differences between models are (a) which variables enter into the macro sufficient statistics x_t , (b) the functional form of the SSJs $J_k(\theta)$, and (c) which aggregated agent choices y_t

are applicable; but given these, the system of equations in [Assumption 1](#) will be satisfied (ignoring linearization error).

In particular, model blocks with heterogeneous firms or financial intermediaries also have a sequence-space representation which, when combined with valid instruments as discussed in [Section 2.3](#), will result in moment conditions of the form in [Assumption 1](#), for particular model-specific choices of x_t and $\{J_k(\theta)\}_k$. See for example [Winberry, Auclert, Rognlie, and Straub \(2025\)](#) for the case of heterogeneous firms.¹⁰

3.2 Estimator

We now define the minimum distance estimator. Rather than formulating the sample moment conditions in the time domain as in [\(5\)](#), we define the estimator in the frequency domain. This will turn out to both facilitate the theoretical econometric analysis in [Section 4](#) and motivate an effective bootstrap strategy for inference. To that end, let

$$S_{ab}(\omega) \equiv \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\omega h} \text{Cov}(a_{t+h}, b_t)$$

denote the cross-spectral density matrix for arbitrary stationary time series a_t, b_t (with absolutely summable autocovariance function), where $i = \sqrt{-1}$. Then the moment conditions in [Assumption 1](#) can be equivalently stated as

$$\int_0^{2\pi} \{S_{yz}(\omega) - J(\omega; \theta_0) S_{xz}(\omega)\} d\omega = 0_{d_y \times d_z}, \quad (6)$$

where $J(\omega; \theta) \equiv \sum_{k=0}^{\infty} e^{ik\omega} J_k(\theta)$. The standard sample analogue of $S_{ab}(\omega)$ is the cross-periodogram $\hat{S}_{ab}(\omega) \equiv \frac{1}{2\pi T} \{\sum_{t=1}^T e^{-i\omega t} (a_t - \bar{a})\} \{\sum_{t=1}^T e^{i\omega t} (b_t - \bar{b})'\}$, where $\bar{a} \equiv \frac{1}{T} \sum_{t=1}^T a_t$ and similarly for $\bar{b}$. The following minimum distance estimator simply plugs the periodogram into the population moment condition [\(6\)](#) and approximates the integral with a sum. Let $d_g \equiv d_y d_z$ denote the total number of moments.

Definition 1 (Estimator). *Given a symmetric positive definite weight matrix $\widehat{W} \in \mathbb{R}^{d_g \times d_g}$,*

¹⁰While our focus here is on traditional applications in macroeconomics, one could imagine applying the sequence-space machinery to models from structural microeconomics where agents' choices depend on aggregate variables, such as models of schooling choice given time-varying skill premia.

the minimum distance estimator is defined as

$$\hat{\theta} \equiv \underset{\theta \in \Theta}{\operatorname{argmin}} \hat{g}(\theta)' \widehat{W} \hat{g}(\theta),$$

with moment function

$$\hat{g}(\theta) \equiv \frac{2\pi}{T} \sum_{j=0}^{T-1} \operatorname{vec} \left\{ \hat{S}_{yz}(\omega_j) - J(\omega_j; \theta) \hat{S}_{xz}(\omega_j) \right\}, \quad \text{where} \quad \omega_j \equiv \frac{2\pi j}{T}.$$

In practice, we use a gradient-based numerical optimization routine with multiple starting values to compute the minimum. As is well known, the cross-periodogram can be computed efficiently using the discrete Fourier transform. Similarly, for any θ , given the SSJs $\{J_k(\theta)\}_k$ produced by the [ABRS](#) algorithm, the spectral SSJ function $J(\omega; \theta)$ can be computed rapidly by another (inverse) Fourier transform.¹¹ As defined, the moment function $\hat{g}(\theta)$ is theoretically guaranteed to be a real vector for any θ ; however, numerical errors can cause the computed $\hat{g}(\theta)$ to have a very small but nonzero imaginary part, so in practice we just retain the real part.

3.3 Inference

In [Section 4.3](#) we will show that the minimum distance estimator is consistent and asymptotically normal under weak conditions. Its asymptotic variance-covariance matrix can be estimated consistently by

$$\hat{\Sigma} \equiv (\hat{G}' \widehat{W} \hat{G})^{-1} \hat{G}' \widehat{W} \hat{\Omega} \widehat{W} \hat{G} (\hat{G}' \widehat{W} \hat{G})^{-1},$$

where

$$\begin{aligned} \hat{G} &\equiv \frac{\partial \hat{g}(\hat{\theta})}{\partial \theta'} \in \mathbb{R}^{d_g \times d_\theta}, \\ \hat{\Omega} &\equiv \sum_{|h| \leq L_T} K\left(\frac{h}{L_T}\right) \hat{\Gamma}_\psi(h) \in \mathbb{R}^{d_g \times d_g}, \quad \hat{\Gamma}_\psi(h) \equiv \begin{cases} \frac{1}{T} \sum_{t=h+1}^T \left(\hat{\psi}_t - \bar{\psi} \right) \left(\hat{\psi}_{t-h} - \bar{\psi} \right)', & h \geq 0, \\ \hat{\Gamma}_\psi(-h)', & h < 0, \end{cases} \\ \hat{\psi}_t &\equiv (z_t - \bar{z}) \otimes \left(y_t - \bar{y} - \sum_{k=0}^{T-t} J_k(\hat{\theta})(x_{t+k} - \bar{x}) \right) \in \mathbb{R}^{d_g}, \quad \bar{\psi} \equiv \frac{1}{T} \sum_{t=1}^T \hat{\psi}_t, \end{aligned}$$

¹¹As discussed in [Section 2.2](#), in practice SSJs are computed using a finite truncation horizon H . Here we abstract from this minor numerical issue and assume $H = \infty$.

$\bar{y} \equiv T^{-1} \sum_{t=1}^T y_t$, $\bar{x}$, and $\bar{z}$ are sample averages, and $K(\cdot)$ is a heteroskedasticity and autocorrelation consistent (HAC) kernel function satisfying the standard conditions in [Appendix B](#).¹² The derivative matrix $\hat{G}$ can be computed via a finite difference approximation. As usual, standard errors for the individual parameter estimates $\hat{\theta}_j$ can be computed as $\sqrt{\hat{\Sigma}_{jj}/T}$, $j = 1, \dots, d_\theta$. In the simulation study in [Section 5](#), we use the Newey-West HAC kernel $K(u) = \max\{1 - |u|, 0\}$ with bandwidth selection rule $L_T = \lceil 2.24 \times T^{1/3} \rceil$.¹³

While the variance estimator above is computationally attractive, it is based on an asymptotic delta method linearization that can be inaccurate in finite samples whenever the SSJs are highly nonlinear in θ . Our simulation study in [Section 5](#) finds that a bootstrap inference procedure reliably delivers confidence intervals with accurate coverage. This procedure resamples the periodogram using a Gaussian multiplier bootstrap and re-runs the nonlinear optimization of the objective function on each bootstrap data set. See [Appendix A](#) for details on the algorithm.

3.4 Over-identification test

In the over-identified case $d_g > d_\theta$, a test of the specification of the model block can be carried out by computing the statistic

$$\hat{\Upsilon} \equiv T\hat{g}(\hat{\theta})' \left\{ \hat{\Omega}^{-1} - \hat{\Omega}^{-1}\hat{G}(\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}\hat{G}'\hat{\Omega}^{-1} \right\} \hat{g}(\hat{\theta}),$$

see [Newey and McFadden \(1994, Section 9.5\)](#).¹⁴ Intuitively, this statistic checks whether the single parameter vector $\hat{\theta}$ is able to approximately satisfy all moment conditions simultaneously. Under the joint null hypothesis of correct specification of the model block and instrument validity, the statistic has an asymptotic chi-squared distribution with $d_g - d_\theta$ degrees of freedom. However, our simulation study in [Section 5](#) suggests that the use of analytical critical values can cause size distortions in finite samples, whereas a bootstrap critical value controls size well even in modest sample sizes. The bootstrap procedure is described in [Appendix A](#).

¹² $\hat{\psi}_t$ is an approximation to the per-observation score process for the sample moment function (5), truncating the sum at the longest horizon available in the data.

¹³This bandwidth rule is mean-squared-error-optimal if the true score process is an AR(1) process with $\rho = 0.7$ ([Andrews, 1991](#), equations 6.2 and 6.4).

¹⁴Note that the formula plugs in the HAC estimate $\hat{\Omega}$ regardless of which weight matrix $\hat{W}$ is used to compute $\hat{\theta}$.

3.5 Weight matrix

A natural, though *ad hoc*, choice of weight matrix is $\widehat{W} = \widehat{V}_z^{-1} \otimes \widehat{V}_y^{-1}$, where $\widehat{V}_z$ is a diagonal matrix with the sample variances of the elements of z_t on the diagonal, while $\widehat{V}_y$ similarly collects the sample variances of y_t on its diagonal. This choice of weight matrix makes the entire estimation procedure invariant to the units of y_t and z_t .

Asymptotically, the efficient choice of weight matrix can be consistently estimated by $\widehat{W}^{\text{eff}} = \widehat{\Omega}^{-1}$, where $\widehat{\Omega}$ is obtained using any initial positive definite weight matrix, such as the one suggested above (Newey and McFadden, 1994, Section 5.2). However, in some applications, the number d_g of moments may be large relative to the sample size T , likely causing the purported “efficient” estimator to be poorly behaved in small samples (Altonji and Segal, 1996). We recommend that users run Monte Carlo studies calibrated to their application to compare the finite-sample performance of estimators with different choices of weight matrix.

3.6 Richer agent choice data

Our procedure can exploit richer data on agents’ choices than simple cross-sectional moments. So far we have restricted attention to aggregated outputs y_t of the agents’ decisions that can be written as integrals $y_t = \int_0^1 y_{i,t} di$, where $y_{i,t}$ is a vector of simple transformations of households’ choice variables. However, as emphasized by ABRS (Appendix A), the linearized representation (1) will also obtain for many other functionals y_t of the cross-sectional distribution of agent choices, such as quantiles, conditional expectations, etc. Their Python package readily computes SSJs for such functionals.¹⁵ As a concrete example, in the simulation study in Section 5, we consider cross-sectional moments of the form $\int a_{i,t} \mathbb{1}\{q_{p_0,t} \leq b_{i,t} + a_{i,t} \leq q_{p_1,t}\} di$, i.e., the total illiquid assets held by households who are between the $100p_0$ and $100p_1$ percentiles of the wealth distribution. By looking at various quantiles p_0, p_1 , these moments give detailed information on the dynamics of wealth inequality, which should provide identifying power for many of the parameters of the household problem.

Our method can also exploit short micro panels, though in limited ways. The Python

¹⁵Mechanically, in their `sequence-jacobian` Python package, any functional of the joint cross-sectional distribution of agent states and choices can be attached to the model block using the `add_hetoutputs` function. The code will then automatically compute the SSJs of these functionals with respect to the macro sufficient statistics, just like it would for any other aggregate variables.

package of [ABRS](#) easily allows users to compute cross-sectional moments of the entire joint distribution of agent states and choices. Some of these state variables may be lagged choice variables, such as asset holdings. For example, in the model from [Section 2.1](#), it would be straightforward to compute SSJs for any smooth moments of the form $y_t = \int_0^1 \text{fct}(a_{i,t-1}, b_{i,t-1}, a_{i,t}, b_{i,t}, c_{i,t}, x_t) di$. This allows researchers to at least exploit two-period rotating panels on liquid and illiquid asset holdings. It would be an interesting question for future research to develop an SSJ algorithm for aggregated outcomes that involve micro data for more than two consecutive time periods.

4 Theoretical results

This section presents our formal econometric assumptions and proves identification, consistency, asymptotic normality, and consistent estimation of the asymptotic variance. We encourage applied readers to skip ahead to the next section on their first reading.

The proofs in [Appendix C](#) in fact prove consistency and asymptotic normality for a more general class of “spectral GMM” estimators; the results in the present section are special cases, as formally verified in [Section C.3](#). In addition to potentially opening the door to other kinds of applications, the general framework arguably makes the proofs more transparent.

4.1 Main assumptions

Let the observed data be denoted $\zeta_t \equiv (y'_t, x'_t, z'_t)' \in \mathbb{R}^{d_\zeta}$ with $d_\zeta \equiv d_y + d_x + d_z$. All norms below are Euclidean for vectors or Frobenius for matrices, and the dimensions are fixed so norms are equivalent.

Assumption 2 (Data). *(i) ζ_t is strictly stationary with mean zero and finite fourth moments.*

(ii) The joint cumulants of orders $\ell \in \{2, 4\}$ are absolutely summable:¹⁶

$$\sum_{h_1 \in \mathbb{Z}} \sum_{h_2 \in \mathbb{Z}} \cdots \sum_{h_{\ell-1} \in \mathbb{Z}} \left\| \text{cum}(\zeta_0, \zeta_{h_1}, \dots, \zeta_{h_{\ell-1}}) \right\| < \infty.$$

Recall that [Assumption 1](#) already imposed stationarity of $\{\zeta_t\}$. The mean zero assump-

¹⁶We interpret the norm as applying to the vectorization of the cumulant tensor.

tion above is just to simplify notation; in practice, we work with de-meaned data.¹⁷ The summability assumptions on the second- and fourth-order cumulants follow Brillinger (1981) and Andrews (1991). These conditions are for example implied by mixing assumptions and hold for many stationary time series models; see Andrews (1991).

Assumption 3 (Parameter space). $\Theta \subset \mathbb{R}^{d_\theta}$ is compact, and the true parameter $\theta_0 \in \Theta$.

The compactness assumption is conventional, but could be relaxed under shape restrictions on the SSJs.

Assumption 4 (SSJ). (i) For each $k \geq 0$, $J_k(\theta)$ is continuously differentiable on Θ .

(ii) The SSJ and its derivative are uniformly summable on Θ , i.e.,

$$\sum_{k=0}^{\infty} \sup_{\theta \in \Theta} \|J_k(\theta)\| < \infty, \quad \sum_{k=0}^{\infty} \sup_{\theta \in \Theta} \|\partial \text{vec}(J_k(\theta)) / \partial \theta'\| < \infty.$$

This assumption requires the SSJs $\{J_k(\theta)\}_k$ to be sufficiently smooth and asymptote to zero sufficiently fast at long horizons $k \rightarrow \infty$. In other words, agents' *aggregated* optimal choices must depend smoothly on the structural parameters, and the dependence of current choices on expectations of macro sufficient statistics far into the future must be limited. Note that we do not impose that *individual* agent choices are smooth, nor do we require a finite decision horizon.

4.2 Identification

We now give high-level assumptions for identification of the structural parameters and subsequently discuss their economic interpretation.

Recall that Assumption 1 assumed that $\text{Cov}(y_t, z_t) = \sum_{k=0}^{\infty} J_k(\theta) \text{Cov}(x_{t+k}, z_t)$ at the true parameters $\theta = \theta_0$. The following assumption implies that the moment conditions are satisfied *only* at the true parameters.

Assumption 5 (Identification). The map $\theta \mapsto \sum_{k=0}^{\infty} J_k(\theta) \text{Cov}(x_{t+k}, z_t)$ is injective on Θ .¹⁸

¹⁷The difference between population and sample demeaning affects only the zero Fourier ordinate. As a result, the change in the sample moments $\hat{g}(\theta)$ and their derivative induced by sample de-meaning is $O_p(T^{-1})$ uniformly in θ under our assumptions, and is therefore asymptotically negligible.

¹⁸The infinite series is well-defined for all θ by Assumptions 2 and 4.

Intuitively, this assumption requires the instrument vector z_t to correlate with the macro sufficient statistics x_{t+k} at a sufficient number of future horizons k in order to distinguish the model-implied optimal agent choices $\sum_{k=0}^{\infty} J_k(\theta) \text{Cov}(x_{t+k}, z_t)$ at different parameter values θ . In particular, if the macro sufficient statistics were uncorrelated with all leads and lags of the instruments, identification would clearly fail. As a more interesting example, if a specific element in the parameter vector only governs the degree to which agents' behavior is forward-looking (that is, $J_k(\theta)$ is constant in θ at short horizons), then the instruments must predict the macro sufficient statistics at sufficiently long horizons; otherwise the exogenous variation induced by the instruments would not be helpful for identifying this specific parameter.

Example 1. Consider a stylized economic model where agents' aggregated consumption y_t (a scalar) depends on expectations of current and next-period aggregate income x_t (also a scalar), as well as a function ξ_{t-1} of lagged shocks:

$$y_t = \beta x_t + \gamma E_t[x_{t+1}] + \xi_{t-1}.$$

The two unknown parameters are $\theta = (\beta, \gamma)'$. This corresponds to the SSJs $J_0(\theta) = \beta$, $J_1(\theta) = \gamma$, and $J_k(\theta) = 0$ for $k \geq 2$. Let z_t be a vector of observed shocks. For this linear model, [Assumption 5](#) reduces to the requirement that the $2 \times d_z$ matrix

$$\text{Cov} \left(\begin{pmatrix} x_t \\ x_{t+1} \end{pmatrix}, z_t \right)$$

have full row rank 2. In particular, there must be at least two instruments, and they must be able to “perturb” the future value x_{t+1} of income differently from the current value x_t , in order to distinguish the contemporaneous coefficient β from the forward-looking coefficient γ . That is, the impulse response functions of x_t with respect to the various shocks in z_t must not only be nonzero but also differ across shocks.

Now suppose instead that our structural model imposes the restriction $\beta = \gamma$, so there is only a single unknown parameter. This is a stylized version of a permanent income model. Then [Assumption 5](#) reduces to the weaker requirement that $\text{Cov}(x_t + x_{t+1}, z_t) \neq 0$. So a single instrument suffices, and it just needs to induce variation in the *cumulated* response of income. $\square$

For the asymptotic normality result below, we additionally require local identification.

Assumption 6 (Local identification). *The $d_g \times d_\theta$ matrix*

$$G_0 \equiv - \sum_{k=0}^{\infty} \left(\text{Cov}(z_t, x_{t+k}) \otimes I_{d_y} \right) \frac{\partial \text{vec}(J_k(\theta_0))}{\partial \theta'}$$

has full column rank.

This assumption would imply [Assumption 5](#) if the SSJs were linear in the parameters θ (as in [Example 1](#)). In the nonlinear case, [Assumption 6](#) rules out pathologies where the map from parameters to outcomes is locally flat around the truth.

4.3 Consistency and asymptotic normality

We now establish the consistency and asymptotic normality of the minimum distance estimator $\hat{\theta}$ defined in [Section 3.2](#). The key step in our proofs is to show that [Assumptions 2](#) and [4](#) together imply pointwise convergence and stochastic equicontinuity of the sample moments $\hat{g}(\theta)$. The pointwise convergence is a standard result (e.g., [Brillinger, 1981](#), Chapter 7.6). Stochastic equicontinuity follows from [Dahlhaus \(1988\)](#) under strong restrictions on moments of the data of all orders. However, due to the *bilinearity* of $\hat{g}(\theta)$ in the periodogram $\hat{S}(\cdot)$ and the SSJs $J(\cdot; \theta)$, we are able to establish stochastic equicontinuity under the same fourth-moment conditions in [Assumption 2](#) that are also required for pointwise convergence.

We impose the following conventional assumption on the weight matrix $\widehat{W}$.

Assumption 7 (Weight matrix). $\widehat{W} \xrightarrow{p} W$, where W is symmetric positive definite.

Consistency then follows from the standard results for extremum estimators ([Newey and McFadden, 1994](#), Theorem 2.1).

Proposition 1 (Consistency). *Under [Assumptions 1–5](#) and [7](#),*

$$\hat{\theta} \xrightarrow{p} \theta_0.$$

To prove asymptotic normality of the estimator $\hat{\theta}$, we impose a high-level central limit theorem (CLT) for the *spectral mean* $\hat{g}(\theta_0) = \frac{2\pi}{T} \sum_{j=0}^{T-1} \text{vec}\{\hat{S}_{yz}(\omega_j) - J(\omega_j; \theta_0)\hat{S}_{xz}(\omega_j)\}$.

Assumption 8 (Spectral mean CLT). *Assume that*

$$\sqrt{T}\hat{g}(\theta_0) \xrightarrow{d} N(0, \Omega)$$

for some positive semidefinite Ω .

Recall that the population analogue of $\hat{g}(\theta_0)$ equals 0 by [Assumption 1](#). Various sets of sufficient conditions for [Assumption 8](#) are provided by [Brillinger \(1981, Theorem 7.6.1\)](#), [Dahlhaus \(1988\)](#), and [Meyer and Paparoditis \(2023\)](#). Moreover, the spectral mean CLT follows from conventional time-domain CLTs under some further assumptions; see [Remark C.8](#) in [Appendix C](#) for details.

Asymptotic normality now follows from the standard result for extremum estimators ([Newey and McFadden, 1994](#), Theorem 3.2).

Proposition 2 (Asymptotic normality). *Under [Assumptions 1–8](#), and if θ_0 lies in the interior of Θ ,*

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, (G'_0 W G_0)^{-1} G'_0 W \Omega W G_0 (G'_0 W G_0)^{-1}\right).$$

The asymptotic null distribution of the over-identification statistic $\hat{\Upsilon}$ in [Section 3.4](#) follows immediately from the calculations in [Newey and McFadden \(1994, Section 9.5\)](#).

Corollary 1 (Over-identification test). *Under the assumptions of [Proposition 2](#), and if moreover $\hat{\Omega} \xrightarrow{p} \Omega$ and Ω is non-singular, then $\hat{\Upsilon} \xrightarrow{d} \chi^2_{d_g - d_\theta}$.*

[Appendix B](#) establishes the consistency of the HAC estimator of the variance-covariance matrix $\hat{\Omega}$, defined in [Section 3.3](#).

5 Simulations

We illustrate the empirical utility of our econometric procedures by applying them to data simulated from the workhorse two-asset HANK model in [ABRS](#), which builds on [Kaplan, Moll, and Violante \(2018\)](#).

STRUCTURAL MODEL. The household block of the model is the one defined in [Section 2.1](#), and the general equilibrium is completed by specifying assumptions for financial intermediaries, firms, labor unions, monetary and fiscal policy, and market clearing. We refer to [ABRS](#) ([Appendix B.3](#)) for details.

The true structural parameters are selected exactly as in [ABRS](#) (Table B.III). In particular, the households' idiosyncratic productivity process $e_{i,t}$ evolves as an AR(1) process in logs, and the adjustment cost function for illiquid assets has the functional form

$$\Psi(a_{i,t}, (1 + r_t^a)a_{i,t-1}) \equiv \frac{\chi_1}{2} \frac{(a_{i,t} - (1 + r_t^a)a_{i,t-1})^2}{(1 + r_t^a)a_{i,t-1} + \chi_0},$$

where $\chi_0, \chi_1 > 0$ are parameters.¹⁹

We assume the macroeconomy is driven by a subset of the shocks in the empirically estimated two-asset HANK model of [ABRS](#) (Section 5.4): a TFP shock, a monetary policy shock, and a government spending shock. We limit ourselves to this list of shocks for purely pedagogical reasons, since this allows us to use the two-asset model definition in the `sequence-jacobian` Python package entirely off the shelf; however, it would be easy to add additional shocks (e.g., to markups). The shocks are assumed to follow independent Gaussian AR(1) processes with parameters given by the posterior means reported by [ABRS](#) (Table F.III, columns labeled “Posterior (Shocks)”).

SIMULATION AND ESTIMATION SETTINGS. We assume that the researcher observes $T = 120$ quarterly observations simulated from the linearized equilibrium of the full structural model.²⁰ The econometrician exploits $d_y = 9$ household output series:

- The cross-sectional mean, variance, and centered third moment of log consumption.
- The total amount of liquid assets held by households (i) below the median of the wealth distribution, (ii) between the 50th–90th percentiles of the wealth distribution, and (iii) between the 90th–99th percentiles of the wealth distribution.²¹
- Same statistics as in the previous bullet point, but for illiquid assets.

For added realism, and to avoid numerical issues arising from dynamically singular moments, we add temporally and cross-sectionally independent Gaussian measurement errors to each of the 9 observed output series. For each series, the standard deviation of the measurement error equals 20% of the population standard deviation of the original un-contaminated series.

The econometrician additionally observes the $d_x = 3$ macro sufficient statistics: returns on liquid and illiquid assets, and the aggregate post-tax real earnings. Finally, they observe $d_z = 2$ instruments, namely the time- t innovations to the government spending and monetary policy shocks.

We estimate four parameters: the household EIS γ and discount factor β , and the illiquid asset adjustment cost parameters χ_0 and χ_1 . The researcher is assumed to have correct

¹⁹In other words, we maintain the same calibration $\chi_2 = 2$ as [ABRS](#), using their notation.

²⁰The choice of sample size is motivated by the available time spans of cross-sectional data on consumption and savings in data sources like the Consumer Expenditure Survey and Distributional Financial Accounts.

²¹Here “wealth distribution” refers to liquid plus illiquid assets.

knowledge of the true values of the remaining parameters in the household block.²² The estimation procedure only requires knowledge of the structure of the household block, and does not use any information about the rest of the general equilibrium (including the nature of any unobserved variables or shocks). We use the *ad hoc* minimum distance weight matrix proposed in Section 3.5, and compute the estimator using numerical optimization.²³

In addition to analytical confidence intervals and over-identification tests, we consider the bootstrap procedures developed in Appendix A (without the correction for potentially non-Gaussian shocks). Due to the high computational cost of repeatedly running nonlinear optimizations, we approximate the performance of bootstrap procedures using the warp-speed diagnostic of Giacomini, Politis, and White (2013).²⁴

RESULTS. Table 1 shows that our estimator is able to deliver accurate inference on the four parameters in the household block despite the moderate sample size. The finite-sample biases and standard deviations of all parameter estimators are modest relative to the scale of the true parameter values. As expected due to the limited sample and the nonlinearity of the SSJs in the underlying parameters, the analytical confidence interval developed in Section 3.3 under-covers substantially for 3 of the 4 parameters, though the extent of under-coverage may be tolerable for initial experimentation. However, the bootstrap confidence interval described in Appendix A achieves near-nominal coverage for all parameters. As for the over-identification test, the bootstrap version has near-nominal rejection rate, while the analytical critical value leads to severe over-rejection.

We conclude that the parameter estimators, as well as the bootstrap confidence intervals

²²These include the idiosyncratic productivity process parameters and the steady-state values of the macro sufficient statistics. In practice, the latter values could be estimated from long-run time series averages.

²³We use the gradient-based L-BFGS-B algorithm implemented in SciPy’s `minimize` routine. In each simulation, we first draw 5 initial parameter values at random, perform 5 optimization steps for each, and record the resulting parameter vector that yields the lowest objective function value across the 5 attempts; then the final, full optimization is initialized at this single parameter vector. The 5 random draws of initial values are uniform on $[0.5 \times \text{truth}, 1.5 \times \text{truth}]$ (independently across the four parameters), except for β which due to its natural parameter space is drawn from $[0.99 \times \text{truth}, 1.01 \times \text{truth}]$. The optimization routine enforces some very loose bounds on the parameters that never bind at the computed optimum in any simulation.

²⁴That is, we only generate a single bootstrap draw in each simulated data set. The bootstrap quantiles for the confidence interval and over-identification test are computed from the distribution of bootstrap draws across simulated data sets, rather than across bootstrap iterations in a given simulated data set. This strategy is of course infeasible in reality, where only a single data set is available. The warp-speed diagnostic checks whether the bootstrap distribution of a statistic lines up with its sampling distribution, after integrating out over the distribution of the data. As such, the diagnostic can detect various types of bootstrap failure, but not all (Giacomini, Politis, and White, 2013, p. 575).

Param.	Truth	Bias	Stdev.	RMSE	Analyt.	Bootstr.
					CI coverage	
γ	0.5000	0.0214	0.0844	0.0871	0.467	0.904
β	0.9763	0.0000	0.0032	0.0032	0.479	0.912
χ_0	0.2500	0.0374	0.0893	0.0969	0.990	0.947
χ_1	6.4164	-0.7883	1.7271	1.8985	0.621	0.857
					Rejection rate	
Over-ID test					0.949	0.115

Table 1: Simulation results across 512 Monte Carlo replications. Sample size: $T = 120$. Nominal significance level: $\alpha = 0.1$. Columns left to right: parameter name, true value, estimator bias, estimator standard deviation, estimator root mean squared error, coverage of analytical confidence interval, and coverage of bootstrap confidence interval. The last row reports the rejection rate of the over-identification test, either with analytical (left) or bootstrap (right) critical value. Bootstrap coverage and rejection rates are approximated with the warp-speed diagnostic of [Giacomini, Politis, and White \(2013\)](#).

and tests, perform well in an empirically calibrated DGP with modest sample size $T = 120$. By contrast, analytical confidence intervals should be used only as a rough guide, and the analytical critical value for the over-identification test is not reliable in our simulation DGP.

6 Conclusion

We proposed a general procedure for estimating and testing a single block of a heterogeneous agent model. Like the existing Generalized Method of Moments procedures used to estimate individual model blocks in representative agent settings, our limited-information method is fully consistent with general equilibrium but does not require imposing assumptions on the remaining blocks of the model. Our procedure is conceptually and computationally easy to implement on any heterogeneous agent model block that fits into the sequence-space framework of [ABRS](#).

The utility of estimating a single model block is threefold. First, the parameters of this block may be of direct scientific interest, or they can serve as robust calibration inputs into a subsequent fully-specified structural model. Second, one can estimate any policy counterfactuals that are functions only of the estimated model block. Third, we developed an over-identification test that can evaluate the empirical validity of the model block in isolation, without the risk of contamination from other potentially misspecified model blocks.

Appendix A Bootstrap inference

Here we describe a bootstrap inference approach that is designed to capture the nonlinearity of the SSJs in the structural parameters.

MULTIPLIER BOOTSTRAP. The bootstrap procedures build on the Gaussian multiplier bootstrap. This procedure assumes that the data is Gaussian (Meyer and Paparoditis, 2023). We discuss how to correct for non-Gaussian data below.

The bootstrap requires an estimate $\hat{f}(\omega)$ of the spectral density of the data $\zeta_t \equiv (y'_t, x'_t, z'_t)'$. Letting $\hat{S}(\omega)$ denote the periodogram of ζ_t , we choose the periodogram smoothing estimator

$$\hat{f}(\omega) \equiv \frac{\sum_{j=1}^{T-1} \widetilde{K}\left(\frac{\min\{|\omega-\omega_j|, 2\pi-|\omega-\omega_j|\}}{B}\right) \hat{S}(\omega_j)}{\sum_{j=1}^{T-1} \widetilde{K}\left(\frac{\min\{|\omega-\omega_j|, 2\pi-|\omega-\omega_j|\}}{B}\right)}$$

with Epanechnikov kernel $\widetilde{K}(u) = \max\{1 - u^2, 0\}$ and spectral bandwidth $B = T^{-0.2}$.²⁵

In each bootstrap iteration, the algorithm proceeds as follows:

1. Draw pseudo-Fourier-transforms $F_j^\dagger \sim N_c(0, \hat{f}(\omega_j), 0)$ independently across Fourier frequencies $\omega_j = 2\pi j/T$, $j = 1, \dots, \lfloor (T-1)/2 \rfloor$ (see Meyer and Paparoditis, 2023, Appendix A.1, for the definition of the complex normal distribution). Then let the pseudo-periodogram draw be given by $\hat{S}^\dagger(\omega_j) \equiv F_j^\dagger (F_j^\dagger)^*$ for all $j = 1, \dots, \lfloor (T-1)/2 \rfloor$ and $\hat{S}^\dagger(\omega_j) \equiv \overline{\hat{S}^\dagger(\omega_{T-j})}$ for the remaining $j \leq T-1$, where an asterisk denotes complex conjugate transpose and bar denotes elementwise complex conjugate. At the boundaries we set $\hat{S}^\dagger(0) = 0$ and (if T is even) $\hat{S}^\dagger(\pi) = 0$.
2. Compute the resampled estimator

$$\hat{\theta}^\dagger \equiv \underset{\theta \in \Theta}{\operatorname{argmin}} \hat{g}^\dagger(\theta)' \widehat{W} \hat{g}^\dagger(\theta),$$

²⁵This bandwidth rule is approximately integrated-MSE-optimal for an AR(1) process with $\rho = 0.7$, as can be verified by plugging into equations 6.2.108 and 6.2.119 in Priestley (1981). Details are available upon request.

using the re-centered moment function²⁶

$$\hat{g}^\dagger(\theta) \equiv \frac{2\pi}{T} \sum_{j=0}^{T-1} \text{vec} \left\{ \hat{S}_{yz}^\dagger(\omega_j) - J(\omega_j; \theta) \hat{S}_{xz}^\dagger(\omega_j) - \left[\hat{f}_{yz}(\omega_j) - J(\omega_j; \hat{\theta}) \hat{f}_{xz}(\omega_j) \right] \right\},$$

where the “ yz ” and “ xz ” subscripts indicate the (y_t, z_t) and (x_t, z_t) blocks of the cross-periodogram or cross-spectrum.²⁷

Though we do not formally prove bootstrap validity, the analysis in [Meyer and Paparoditis \(2023\)](#) suggests that the resampling distribution of $\hat{\theta}^\dagger - \hat{\theta}$ (conditional on the data) approximates the (unconditional) sampling distribution of $\hat{\theta} - \theta_0$, provided the data is Gaussian. The bootstrap standard errors are the standard deviations of the elements of $\hat{\theta}^\dagger$ across re-samples. An (asymmetric) $1 - \alpha$ confidence interval for θ_0 is given by $[2\hat{\theta} - q_{1-\alpha/2}^\dagger, 2\hat{\theta} - q_{\alpha/2}^\dagger]$, where $q_\alpha^\dagger$ is the α -th quantile of $\hat{\theta}^\dagger$ across bootstrap samples.

CORRECTION FOR NON-GAUSSIAN DATA. To allow for non-Gaussian shocks, we can rescale the multiplier bootstrap draws $\hat{\theta}^\dagger$ using the Multivariate Frequency-domain Hybrid Bootstrap (MFHB) of [Meyer and Paparoditis \(2023\)](#). The calculations in the proof of [Theorem C.2](#) show that $\hat{\theta} - \theta_0$ is asymptotically equivalent with $-(\hat{G}'\hat{W}\hat{G})^{-1}\hat{G}'\hat{W}\hat{g}(\theta_0)$, where we recall that $\hat{g}(\theta_0) = \frac{2\pi}{T} \sum_{j=0}^{T-1} \text{vec} \left\{ \hat{S}_{yz}(\omega_j) - J(\omega_j; \theta_0) \hat{S}_{xz}(\omega_j) \right\}$ is a spectral mean. We can approximate the Gaussian and non-Gaussian variance-covariance matrices of $\hat{g}(\theta_0)$ by applying the MFHB algorithm with $\theta_0 \approx \hat{\theta}$ (fixed across bootstrap iterations); denote these variance-covariance matrices by $\hat{\Omega}_G$ and $\hat{\Omega}_{NG}$, respectively.²⁸ Starting with the multiplier bootstrap draws $\hat{\theta}^\dagger$, the rescaled draws

$$\begin{aligned} \hat{\theta}_{\text{rescale}}^\dagger &\equiv \hat{\theta} + \left\{ (\hat{G}'\hat{W}\hat{G})^{-1} \hat{G}'\hat{W} \hat{\Omega}_{NG} \hat{W}\hat{G} (\hat{G}'\hat{W}\hat{G})^{-1} \right\}^{1/2} \\ &\quad \times \left\{ (\hat{G}'\hat{W}\hat{G})^{-1} \hat{G}'\hat{W} \hat{\Omega}_G \hat{W}\hat{G} (\hat{G}'\hat{W}\hat{G})^{-1} \right\}^{-1/2} (\hat{\theta}^\dagger - \hat{\theta}) \end{aligned}$$

²⁶The re-centering ensures that the sample moment function has bootstrap expectation approximately equal to zero at the real-data estimate $\hat{\theta}$.

²⁷While $\hat{g}^\dagger(\theta)$ should be real in large samples, in finite samples we drop any imaginary part.

²⁸In the notation of [Meyer and Paparoditis \(2023, p. 2379, see also Remark 3.5\)](#), these are $\mathbf{G}_n^*$ and $\mathbf{G}_n^\circ$, respectively. Note that the proofs in [Meyer and Paparoditis \(2023\)](#) do not technically allow for data-dependent spectral means (which ours is due to the dependence on $\hat{\theta}$), but we conjecture that the result can be extended to cover our setting.

then have the correct asymptotic variance-covariance matrix (as in [Proposition 2](#)) even under non-Gaussianity.²⁹ The bootstrap standard errors and confidence intervals can now be computed from the rescaled draws.

OVER-IDENTIFICATION TEST. The bootstrap critical value for the over-identification test is obtained by recomputing the estimator and over-identification statistic from [Section 3.4](#) in every bootstrap data set. As above, we use the re-centered moments $\hat{g}^\dagger(\hat{\theta}^\dagger)$ when computing the statistic, as this imposes the null hypothesis of correct specification on the bootstrap DGP. The only remaining challenge is that the HAC estimator $\hat{\Omega}$ in the formula for the statistic is defined for data in the time domain, but we bootstrap the periodogram rather than the underlying data. Hence, for every draw of the pseudo-Fourier-transforms $F_j^\dagger$, we apply the (scaled) inverse Fourier transform to compute the corresponding data $\{\zeta_t^\dagger\}$ in the time domain. Then we apply the HAC procedure to this data (just as with the real data).

Being based on Gaussian draws, this procedure is designed to work well when the real data is Gaussian, but we conjecture it remains *asymptotically* valid even if the real data is non-Gaussian. This is because as $T \rightarrow \infty$, the bootstrap distribution of the over-identification statistic converges to a χ^2 distribution with $d_g - d_\theta$ degrees of freedom, since the statistic is asymptotically pivotal. This is the correct distribution for computing asymptotic critical values whether or not the data is Gaussian, cf. [Corollary 1](#).

Appendix B Asymptotic variance estimation

Here we state the consistency of the HAC long-run variance estimator. Recall its definition from [Section 3.3](#):

$$\hat{\Omega} \equiv \sum_{|h| \leq L_T} K\left(\frac{h}{L_T}\right) \hat{\Gamma}_\psi(h), \quad \text{where} \quad \hat{\Gamma}_\psi(h) \equiv \frac{1}{T} \sum_{t=1}^T \hat{\psi}_t \hat{\psi}_{t-h}', \quad \hat{\psi}_t \equiv \psi_{t,T}(\hat{\theta}),$$

and we define the truncated triangular array score

$$\psi_{t,T}(\theta) \equiv \text{vec} \left(\left(y_t - \sum_{k=0}^{T-t} J_k(\theta) x_{t+k} \right) z_t' \right) \mathbb{1}\{1 \leq t \leq T\}.$$

²⁹We use the symmetric positive semidefinite matrix square root.

For simplicity, we here impose in the definition of the HAC estimator that the data has mean zero ([Assumption 2](#)), and we do not de-mean the sample score $\hat{\psi}_t$ itself. We conjecture that both types of sample de-meaning have negligible effects asymptotically under our assumptions (see [Section C.1.3](#)).

Define the population score process

$$\psi_t(\theta) \equiv \text{vec} \left(\left(y_t - \sum_{k=0}^{\infty} J_k(\theta) x_{t+k} \right) z_t' \right).$$

Then $g(\theta) \equiv \int_0^{2\pi} \text{vec}\{S_{yz}(\omega) - J(\omega; \theta)S_{xz}(\omega)\} d\omega = E[\psi_t(\theta)]$. Define the long-run variance

$$\Omega \equiv \sum_{h \in \mathbb{Z}} \Gamma_\psi(h), \quad \text{where} \quad \Gamma_\psi(h) \equiv E[\psi_h(\theta_0)\psi_0(\theta_0)'].$$

This is the same Ω that appears in [Assumption 8](#) whenever that assumption (and our other assumptions) hold; see [Remark C.8](#) below.

Assumption B.1 (HAC kernel and bandwidth). *Assume:*

- (i) *The kernel $K(\cdot)$ is symmetric, bounded, supported on $[-1, 1]$, and continuous at 0 with $K(0) = 1$.*
- (ii) *The bandwidth satisfies $L_T \rightarrow \infty$ and $L_T/\sqrt{T} \rightarrow 0$.*

This assumption follows [Andrews \(1991\)](#) and is satisfied by standard kernels, such as Newey-West (Bartlett).

Assumption B.2 (Stronger cumulant summability). *Assume ζ_t has finite 8th moments, and strengthen the cumulant summability condition in [Assumption 2\(ii\)](#) to hold for all $\ell = 2, 3, \dots, 8$.*

Eighth-order cumulant control on ζ_t is used to control fourth-order cumulants of the score, which is quadratic in ζ_t .

Assumption B.3 (Weighted summability of SSJ). $\sum_{k=0}^{\infty} k \|J_k(\theta_0)\| < \infty$.

The weighted summability condition ensures that the sample boundary truncation in the triangular array score $\psi_{t,T}(\theta_0)$ is asymptotically negligible.

The HAC estimation error decomposes into four terms that each vanish: error from plugging in $\hat{\theta}$ for θ_0 , stochastic fluctuation of the covariance average, kernel bias, and truncation of the infinite lag sum. Therefore, the HAC estimator is consistent.

Proposition B.1 (Consistent HAC estimator). *Under the assumptions of [Proposition 2](#), and if moreover [Assumptions B.1–B.3](#) hold, then*

$$\widehat{\Omega} \xrightarrow{p} \Omega.$$

Appendix C Proofs

In this appendix, we develop asymptotic theory for a general “spectral GMM” estimator, and then specialize it to the setup in [Sections 3](#) and [4](#). [Section C.1](#) develops consistency, asymptotic normality, and HAC variance estimation in the general setup, [Section C.2](#) collects auxiliary results, and [Section C.3](#) specializes to our applications.

Let the data $\zeta_t \in \mathbb{R}^{d_\zeta}$ be real-valued, with spectral density matrix $S(\omega) \in \mathbb{C}^{d_\zeta \times d_\zeta}$ for $\omega \in [0, 2\pi]$. Define the discrete Fourier transform and the matrix periodogram

$$F_\zeta(\omega) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^T \zeta_t e^{-i\omega t}, \quad \widehat{S}(\omega) \equiv \frac{1}{2\pi} F_\zeta(\omega) F_\zeta(\omega)^*,$$

where A^* denotes the conjugate transpose of A for a generic vector/matrix A . Let $\tilde{J}(\omega; \theta) \in \mathbb{C}^{d_g \times d_\zeta^2}$ be a function that is measurable in ω and indexed by $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$. Define the population and sample spectral means

$$g(\theta) \equiv \int_0^{2\pi} \tilde{J}(\omega; \theta) \text{vec}(S(\omega)) d\omega, \quad \widehat{g}(\theta) \equiv \frac{2\pi}{T} \sum_{j=0}^{T-1} \tilde{J}(\omega_j; \theta) \text{vec}(\widehat{S}(\omega_j)).$$

We refer to $\tilde{J}$ as a spectral weight function, and restrict attention to the following class.

Assumption C.1 (Conjugate symmetry of spectral weight function). *For any $\theta \in \Theta$, $\tilde{J}(0; \theta)$ is real, and*

$$\tilde{J}(2\pi - \omega; \theta) = \overline{\tilde{J}(\omega; \theta)} \quad \text{for all } \omega \in [0, 2\pi].$$

Here $\overline{A}$ denotes the elementwise conjugate of A for a generic matrix A .

[Assumption C.1](#) and $\zeta_t \in \mathbb{R}^{d_\zeta}$ imply that $g(\theta), \widehat{g}(\theta) \in \mathbb{R}^{d_g}$. We consider a spectral GMM estimator $\widehat{\theta}$ defined by

$$\widehat{\theta} \in \underset{\theta \in \Theta}{\text{argmin}} \widehat{Q}(\theta), \quad \text{where} \quad \widehat{Q}(\theta) \equiv \widehat{g}(\theta)' \widehat{W} \widehat{g}(\theta).$$

Define also the corresponding population objective function $Q(\theta) \equiv g(\theta)' W g(\theta)$.

For clarity, we now briefly indicate how this general framework specializes to the application in [Sections 3 and 4](#). Let $M_z \in \mathbb{R}^{d_z \times d_\zeta}$ be the selection matrix formed by the last d_z rows of the identity matrix I_{d_ζ} , so that $z_t = M_z \zeta_t$. Define the spectral weight function

$$\tilde{J}_{\text{special}}(\omega; \theta) \equiv M_z \otimes (I_{d_y}, -J(\omega; \theta), 0_{d_y \times d_z}) \in \mathbb{C}^{d_g \times d_\zeta^2}.$$

Then

$$\tilde{J}_{\text{special}}(\omega; \theta) \text{vec}(S(\omega)) = \text{vec}\{S_{yz}(\omega) - J(\omega; \theta)S_{xz}(\omega)\}.$$

C.1 Asymptotic results for general spectral GMM

C.1.1 Consistency

Assumption C.2 (Regularity of spectral weight function). *Assume:*

- (i) (Riemann integrability in ω .) For each $\theta \in \Theta$, $\omega \mapsto \tilde{J}(\omega; \theta)$ is bounded and piecewise continuous on $[0, 2\pi]$.
- (ii) (Uniform continuity in θ .) Define the modulus

$$\Delta_{\tilde{J}}(\delta) \equiv \sup_{\theta, \tilde{\theta} \in \Theta: \|\tilde{\theta} - \theta\| \leq \delta} \sup_{\omega \in [0, 2\pi]} \|\tilde{J}(\omega; \tilde{\theta}) - \tilde{J}(\omega; \theta)\|.$$

Then $\Delta_{\tilde{J}}(\delta) \rightarrow 0$ as $\delta \rightarrow 0$.

Assumption C.3 (Identification: general case). For any $\theta \in \Theta$, $g(\theta) = 0$ if and only if $\theta = \theta_0$.

Theorem C.1 (Consistency: general case). Under [Assumptions 2, 3, 7](#) and [C.1–C.3](#),

$$\hat{\theta} \xrightarrow{p} \theta_0.$$

Proof. [Assumption 3](#) gives compact Θ . Under [Assumptions 2, 3](#) and [C.2](#), [Remark C.3](#) implies $g(\theta)$ is continuous, so $Q(\theta)$ is continuous. Under [Assumptions 7](#) and [C.3](#), $Q(\theta)$ has a unique minimizer at θ_0 ([Newey and McFadden, 1994](#), Lemma 2.3).

Under [Assumptions 2, 3](#) and [C.2](#), [Lemma C.5](#) yields $\sup_{\theta \in \Theta} \|\hat{g}(\theta) - g(\theta)\| \xrightarrow{p} 0$. Since g is continuous on compact Θ , we have that $\sup_{\theta} \|g(\theta)\| < \infty$ and $\sup_{\theta} \|\hat{g}(\theta)\| = O_p(1)$. Thus, under [Assumption 7](#),

$$\sup_{\theta \in \Theta} |\hat{Q}(\theta) - Q(\theta)| \leq \|\widehat{W} - W\| \sup_{\theta} \|\hat{g}(\theta)\|^2 + \|W\| \sup_{\theta} \|\hat{g}(\theta) - g(\theta)\| \sup_{\theta} (\|\hat{g}(\theta)\| + \|g(\theta)\|)$$

$$\xrightarrow{p} 0.$$

Consistency then follows from [Newey and McFadden \(1994, Theorem 2.1\)](#). $\square$

C.1.2 Asymptotic normality

Assumption C.4 (Differentiability of spectral weight function). *There exists a compact neighborhood $\Theta_0 \subset \Theta$ of θ_0 such that:*

- (i) *For each $\omega \in [0, 2\pi]$, $\theta \mapsto \text{vec}(\tilde{J}(\omega; \theta))$ is differentiable on Θ_0 .*
- (ii) *(Riemann integrability of derivative in ω .) For each $\theta \in \Theta_0$, $\omega \mapsto \partial \text{vec}(\tilde{J}(\omega; \theta))/\partial \theta'$ is bounded and piecewise continuous on $[0, 2\pi]$.*
- (iii) *(Uniform continuity of derivative in θ .)*

$$\Delta_{\partial \tilde{J}}(\delta) \equiv \sup_{\theta, \tilde{\theta} \in \Theta_0: \|\tilde{\theta} - \theta\| \leq \delta} \sup_{\omega \in [0, 2\pi]} \left\| \frac{\partial \text{vec}(\tilde{J}(\omega; \tilde{\theta}))}{\partial \theta'} - \frac{\partial \text{vec}(\tilde{J}(\omega; \theta))}{\partial \theta'} \right\| \rightarrow 0 \quad \text{as } \delta \rightarrow 0.$$

For $\theta \in \Theta_0$, we have

$$\begin{aligned} G(\theta) &\equiv \frac{\partial g(\theta)}{\partial \theta'} = \int_0^{2\pi} \{\text{vec}(S(\omega))' \otimes I_{d_g}\} \frac{\partial \text{vec}(\tilde{J}(\omega; \theta))}{\partial \theta'} d\omega, \\ \hat{G}(\theta) &\equiv \frac{\partial \hat{g}(\theta)}{\partial \theta'} = \frac{2\pi}{T} \sum_{j=0}^{T-1} \{\text{vec}(\hat{S}(\omega_j))' \otimes I_{d_g}\} \frac{\partial \text{vec}(\tilde{J}(\omega_j; \theta))}{\partial \theta'}. \end{aligned}$$

Assumption C.5 (Local identification: general case). $G_0 \equiv G(\theta_0)$ has full column rank d_θ .

Theorem C.2 (Asymptotic normality: general case). *Under [Assumptions 2, 3, 7, 8](#) and [C.1–C.5](#),*

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, (G_0' W G_0)^{-1} G_0' W \Omega W G_0 (G_0' W G_0)^{-1}\right).$$

Proof. Under [Assumptions 2, 3, 7](#) and [C.1–C.3](#), [Theorem C.1](#) gives consistency. [Assumption C.4](#) implies $\theta_0 \in \text{int}(\Theta)$, [Assumption 8](#) gives the CLT for $\hat{g}(\theta_0)$, and [Assumptions 7](#) and [C.5](#) give invertibility of $G_0' W G_0$. [Assumptions C.4\(i\)](#) and [C.4\(iii\)](#) yield that $\hat{g}(\theta)$ is continuously differentiable on Θ_0 . Under [Assumptions 2](#) and [C.4](#), [Lemma C.6](#) yields uniform convergence of $\hat{G}(\theta)$ to $G(\theta)$ on Θ_0 , and [Remark C.4](#) yields continuity of $G(\theta)$ at θ_0 . The conclusion then follows from [Newey and McFadden \(1994, Theorem 3.2\)](#). $\square$

C.1.3 HAC

Let

$$\tilde{J}_k(\theta) \equiv \frac{1}{2\pi} \int_0^{2\pi} e^{-i\omega k} \tilde{J}(\omega; \theta) d\omega, \quad k \in \mathbb{Z},$$

denote the time-domain coefficient matrices associated with the spectral weight function $\tilde{J}(\cdot; \theta)$. By [Assumption C.1](#), each $\tilde{J}_k(\theta)$ is real-valued.

The following assumption restricts the smoothness and tail behavior of the time-domain coefficients $\{\tilde{J}_k(\theta)\}_k$. [Remark C.6](#) below argues that the assumption can be verified purely by checking smoothness conditions on the spectral weight function $\tilde{J}(\omega; \theta)$, without explicitly computing the time-domain coefficients.

Assumption C.6 (Fourier series summability). *There exists a compact neighborhood $\Theta_0 \subset \Theta$ of θ_0 such that:*

- (i) *For each $k \in \mathbb{Z}$, $\tilde{J}_k(\theta)$ is continuous on Θ . $\tilde{J}(\cdot; \theta)$ has an absolutely convergent Fourier series uniformly in Θ , i.e.,*

$$\sum_{k \in \mathbb{Z}} \sup_{\theta \in \Theta} \|\tilde{J}_k(\theta)\| < \infty,$$

and $\tilde{J}(\cdot; \theta)$ equals its Fourier series for all $\theta \in \Theta$.

- (ii) *For each $k \in \mathbb{Z}$, $\tilde{J}_k(\theta)$ is continuously differentiable on Θ_0 . $\partial \text{vec}(\tilde{J}(\cdot; \theta))/\partial \theta'$ has an absolutely convergent Fourier series uniformly in Θ_0 , i.e.,*

$$\sum_{k \in \mathbb{Z}} \sup_{\theta \in \Theta_0} \|\partial \text{vec}(\tilde{J}_k(\theta))/\partial \theta'\| < \infty.$$

- (iii) *$\partial_\omega \tilde{J}(\cdot; \theta_0)$ has an absolutely convergent Fourier series, i.e.,^{[30](#)}*

$$\sum_{k \in \mathbb{Z}} |k| \|\tilde{J}_k(\theta_0)\| < \infty.$$

Define the score process and its truncated triangular array counterpart:

$$\begin{aligned} \psi_t(\theta) &\equiv \sum_{k \in \mathbb{Z}} \tilde{J}_k(\theta) \text{vec}(\zeta_{t+k} \zeta_t'), \\ \psi_{t,T}(\theta) &\equiv \sum_{k=-(t-1)}^{T-t} \tilde{J}_k(\theta) \text{vec}(\zeta_{t+k} \zeta_t') \mathbb{1}\{1 \leq t \leq T\}. \end{aligned}$$

³⁰Note that $(\partial_\omega \tilde{J})_k(\theta_0) \equiv \frac{1}{2\pi} \int_0^{2\pi} e^{-i\omega k} \partial_\omega \tilde{J}(\omega; \theta_0) d\omega = \frac{ik}{2\pi} \int_0^{2\pi} e^{-i\omega k} \tilde{J}(\omega; \theta_0) d\omega = ik \tilde{J}_k(\theta_0)$.

The latter uses all available data $\zeta_1, \dots, \zeta_T$.

Define the population long-run variance

$$\Omega \equiv \sum_{h \in \mathbb{Z}} \Gamma_\psi(h) \quad \text{where} \quad \Gamma_\psi(h) \equiv E[\psi_h(\theta_0)\psi_0(\theta_0)'].$$

We estimate this long-run variance with the kernel HAC estimator

$$\hat{\Omega} \equiv \sum_{|h| \leq L_T} K(h/L_T) \hat{\Gamma}_{\psi,T}(h), \quad \text{where} \quad \hat{\Gamma}_{\psi,T}(h) \equiv \frac{1}{T} \sum_{t=1}^T \psi_{t,T}(\hat{\theta}) \psi_{t-h,T}(\hat{\theta})',$$

where the kernel K and bandwidth L_T satisfy [Assumption B.1](#). Note that we do not center the sample autocovariances $\hat{\Gamma}_{\psi,T}(h)$. We conjecture—but do not prove—that the centering is asymptotically negligible, since the proofs of [Theorem C.2](#) and [Lemma C.7](#) suggest that $\bar{\psi}_T(\hat{\theta}) \equiv T^{-1} \sum_{t=1}^T \psi_{t,T}(\hat{\theta}) = \hat{g}(\hat{\theta}) + O_p(T^{-1}) = O_p(T^{-1/2})$, so the difference between the centered and uncentered versions of $\hat{\Gamma}_{\psi,T}(h)$ changes $\hat{\Omega}$ by at most $O_p(L_T \|\bar{\psi}_T(\hat{\theta})\|^2) = O_p(L_T/T) = o_p(1)$ under [Assumption B.1\(ii\)](#).

Theorem C.3 (Consistent HAC estimator: general case). *Under [Assumptions 2, 3, 7](#) and [8](#) and [Assumptions B.1, B.2, C.1, C.3, C.5](#) and [C.6](#),*

$$\hat{\Omega} \xrightarrow{p} \Omega.$$

Proof. Define the following additional notation:

$$\begin{aligned} \Omega_T &\equiv \sum_{|h| < T} \Gamma_{\psi,T}(h) & \text{where} \quad \Gamma_{\psi,T}(h) &\equiv \frac{1}{T} \sum_{t=1}^T E[\psi_{t,T}(\theta_0) \psi_{t-h,T}(\theta_0)'], \\ \tilde{\Omega}_T &\equiv \sum_{|h| \leq L_T} K(h/L_T) \tilde{\Gamma}_{\psi,T}(h) & \text{where} \quad \tilde{\Gamma}_{\psi,T}(h) &\equiv \frac{1}{T} \sum_{t=1}^T \psi_{t,T}(\theta_0) \psi_{t-h,T}(\theta_0)'. \end{aligned}$$

By [Assumptions 2\(i\)](#) and [C.1](#), $\psi_t(\theta)$ and $\psi_{t,T}(\theta)$ are real-valued. Note that [Assumptions 3](#) and [C.6\(i\)](#) imply [Assumption C.2](#), and [Assumptions C.6\(i\)](#) and [C.6\(ii\)](#) imply [Assumption C.4](#), so we have the full set of assumptions for consistency and asymptotic normality in [Theorems C.1](#) and [C.2](#).

(i) *Plug-in:* $\hat{\Omega} - \tilde{\Omega}_T \xrightarrow{p} 0$. Let $\Delta_t \equiv \psi_{t,T}(\hat{\theta}) - \psi_{t,T}(\theta_0)$. For each h , by Cauchy-Schwarz,

$$\|\hat{\Gamma}_{\psi,T}(h) - \tilde{\Gamma}_{\psi,T}(h)\| = \left\| \frac{1}{T} \sum_{t=h+1}^T \left(\Delta_t \psi_{t-h,T}(\theta_0)' + \psi_{t,T}(\theta_0) \Delta_{t-h}' + \Delta_t \Delta_{t-h}' \right) \right\|$$

$$\leq 2 \left(\frac{1}{T} \sum_{t=1}^T \|\Delta_t\|^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^T \|\psi_{t,T}(\theta_0)\|^2 \right)^{1/2} + \frac{1}{T} \sum_{t=1}^T \|\Delta_t\|^2.$$

Since the kernel $K(\cdot)$ is bounded and supported on $[-1, 1]$ by [Assumption B.1\(i\)](#), only $|h| \leq L_T$ contribute, so for some constant $C < \infty$,

$$\begin{aligned} \|\hat{\Omega} - \tilde{\Omega}_T\| &\leq C \sum_{|h| \leq L_T} \|\hat{\Gamma}_{\psi,T}(h) - \tilde{\Gamma}_{\psi,T}(h)\| \\ &\leq CL_T \left[\left(\frac{1}{T} \sum_{t=1}^T \|\Delta_t\|^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^T \|\psi_{t,T}(\theta_0)\|^2 \right)^{1/2} + \frac{1}{T} \sum_{t=1}^T \|\Delta_t\|^2 \right]. \end{aligned}$$

Under [Assumption 2\(i\)](#) and [Assumption C.6\(i\)](#) at θ_0 , $E\|\psi_{t,T}(\theta_0)\|^2$ is bounded uniformly in t, T . Hence,

$$\frac{1}{T} \sum_{t=1}^T \|\psi_{t,T}(\theta_0)\|^2 = O_p(1).$$

Moreover, by the mean value theorem,

$$\frac{1}{T} \sum_{t=1}^T \|\Delta_t\|^2 \leq \frac{1}{T} \sum_{t=1}^T \sup_{\theta \in \Theta_0} \|\partial \psi_{t,T}(\theta) / \partial \theta'\|^2 \|\hat{\theta} - \theta_0\|^2.$$

[Theorem C.2](#) implies that $\|\hat{\theta} - \theta_0\| = O_p(T^{-1/2})$. Under [Assumption 2\(i\)](#) and [Assumption C.6\(ii\)](#), we have that $E[\sup_{\theta \in \Theta_0} \|\partial \psi_{t,T}(\theta) / \partial \theta'\|^2]$ is bounded uniformly in t, T . Therefore,

$$\|\hat{\Omega} - \tilde{\Omega}_T\| = O_p\left(\frac{L_T}{\sqrt{T}}\right) + O_p\left(\frac{L_T}{T}\right) = o_p(1),$$

because $L_T/\sqrt{T} \rightarrow 0$ by [Assumption B.1\(ii\)](#).

(ii) *Bias*: $E[\tilde{\Omega}_T] - \Omega_T \rightarrow 0$. Define the scalar weights

$$a_T(h) \equiv K\left(\frac{h}{L_T}\right) \mathbb{1}\{|h| \leq L_T\}.$$

Then

$$E[\tilde{\Omega}_T] - \Omega_T = \sum_h (a_T(h) - 1) \Gamma_{\psi,T}(h).$$

For each fixed h , as $L_T \rightarrow \infty$ by [Assumption B.1\(ii\)](#), we have $\mathbb{1}\{|h| \leq L_T\} \rightarrow 1$. Also, $K(h/L_T) \rightarrow K(0) = 1$ by [Assumption B.1\(i\)](#), so $a_T(h) \rightarrow 1$ pointwise in h . Moreover, $|a_T(h) - 1|$ is uniformly bounded since K is.

Decompose $\Gamma_{\psi,T}(h) = \check{\Gamma}_{\psi,T}(h) + \frac{1}{T} \sum_{t=h+1}^T E[\eta_{t,T}]E[\eta_{t-h,T}]'$, where $\check{\Gamma}_{\psi,T}(h)$ is the centered part of [Lemma C.8](#), $\eta_{t,T}$ is the edge-truncation error of [Lemma C.7](#), and we used that $E[\psi_{t,T}(\theta_0)] = E[\eta_{t,T}]$ by [Assumption C.3](#) (which gives $E[\psi_t(\theta_0)] = g(\theta_0) = 0$). For the centered part, [Lemma C.8](#) (under [Assumptions 2](#) and [C.6\(i\)](#) at θ_0) gives $\|\check{\Gamma}_{\psi,T}(h)\| \leq C(h)$ with $C(h)$ absolutely summable, so by dominated convergence $\sum_h |a_T(h) - 1| \|\check{\Gamma}_{\psi,T}(h)\| \rightarrow 0$. For the mean part, [Lemma C.7](#) gives $\sum_{t=1}^T \|E[\eta_{t,T}]\| \leq \sum_{t=1}^T (E\|\eta_{t,T}\|^2)^{1/2} = O(1)$, so

$$\sum_h |a_T(h) - 1| \left\| \frac{1}{T} \sum_{t=h+1}^T E[\eta_{t,T}]E[\eta_{t-h,T}]' \right\| \leq \frac{\sup_u |K(u)| + 1}{T} \left(\sum_{t=1}^T \|E[\eta_{t,T}]\| \right)^2 = O(1/T).$$

Combining the two parts, $\|E[\tilde{\Omega}_T] - \Omega_T\| \rightarrow 0$ as $T \rightarrow \infty$.

(iii) *Stochastic term:* $\tilde{\Omega}_T - E[\tilde{\Omega}_T] \xrightarrow{p} 0$. We show $\text{Var}(\text{vec}(\tilde{\Omega}_T)) \rightarrow 0$, which gives the claim by Chebyshev's inequality. Since the kernel is bounded by [Assumption B.1\(i\)](#), there is a constant $C < \infty$ such that

$$\|\text{Var}(\text{vec}(\tilde{\Omega}_T))\| \leq \frac{C}{T^2} \sum_{|h| \leq L_T} \sum_{|\tau| \leq L_T} \sum_{t=1}^T \sum_{s=1}^T \|\text{Cov}(\text{vec}(\psi_{t,T}(\theta_0)\psi_{t-h,T}(\theta_0)'), \text{vec}(\psi_{s,T}(\theta_0)\psi_{s-\tau,T}(\theta_0)'))\|.$$

Under [Assumptions 2\(i\)](#), [B.2](#) and [C.6\(i\)](#), [Lemma C.9](#) implies that the inner sum over s is bounded uniformly in t, h, τ, T , so

$$\|\text{Var}(\text{vec}(\tilde{\Omega}_T))\| \leq \frac{C}{T^2} (2L_T + 1)^2 T \Lambda = O(L_T^2/T) \rightarrow 0$$

by [Assumption B.1\(ii\)](#). Hence $\tilde{\Omega}_T - E[\tilde{\Omega}_T] \xrightarrow{p} 0$.

(iv) *Edge truncation:* $\Omega_T \rightarrow \Omega$. Let $\eta_{t,T} = \psi_{t,T}(\theta_0) - \psi_t(\theta_0)$ be the edge-truncation error of [Lemma C.7](#). By that lemma, $\sum_{t=1}^T (E\|\eta_{t,T}\|^2)^{1/2} = O(1)$, so by Cauchy-Schwarz,

$$E \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^T \eta_{t,T} \right\|^2 \leq \frac{1}{T} \left(\sum_{t=1}^T (E\|\eta_{t,T}\|^2)^{1/2} \right)^2 = O(1/T) \rightarrow 0.$$

Since

$$\frac{1}{\sqrt{T}} \sum_t \psi_{t,T}(\theta_0) = \frac{1}{\sqrt{T}} \sum_t \psi_t(\theta_0) + \frac{1}{\sqrt{T}} \sum_t \eta_{t,T},$$

the second moments of $\frac{1}{\sqrt{T}} \sum_t \psi_{t,T}(\theta_0)$ and $\frac{1}{\sqrt{T}} \sum_t \psi_t(\theta_0)$ have the same limit. Moreover, by [Assumptions 2](#), [C.3](#) and [C.6\(i\)](#) at θ_0 , [Remark C.7](#) gives that $\sum_{h \in \mathbb{Z}} \|\Gamma_\psi(h)\| < \infty$, so by

dominated convergence,

$$\frac{1}{T} E \left[\left(\sum_{t=1}^T \psi_t(\theta_0) \right) \left(\sum_{t=1}^T \psi_t(\theta_0) \right)' \right] = \sum_{|h| < T} \left(1 - \frac{|h|}{T} \right) \Gamma_\psi(h) \rightarrow \sum_{h \in \mathbb{Z}} \Gamma_\psi(h) = \Omega.$$

Therefore, $\Omega_T \rightarrow \Omega$.

(v) *Conclusion.* Write

$$\hat{\Omega} - \Omega = (\hat{\Omega} - \tilde{\Omega}_T) + (\tilde{\Omega}_T - E[\tilde{\Omega}_T]) + (E[\tilde{\Omega}_T] - \Omega_T) + (\Omega_T - \Omega).$$

By (i)–(iv), each term is $o_p(1)$ or $o(1)$, so $\hat{\Omega} \xrightarrow{p} \Omega$. □

C.2 Auxiliary results for general spectral GMM

C.2.1 Spectral regularity

Remark C.1 (Continuity and boundedness of spectrum). Under [Assumption 2](#), we have $\sum_{h \in \mathbb{Z}} \|\Gamma_\zeta(h)\| < \infty$, where $\Gamma_\zeta(h) \equiv \text{Cov}(\zeta_h, \zeta_0)$. Hence $S(\omega) \equiv 1/(2\pi) \sum_{h \in \mathbb{Z}} \Gamma_\zeta(h) e^{-ih\omega}$ converges absolutely and uniformly in $\omega \in [0, 2\pi]$, so $S(\omega)$ is continuous, $\sup_{\omega \in [0, 2\pi]} \|S(\omega)\| < \infty$, and $\int_0^{2\pi} \|S(\omega)\| d\omega < \infty$.

Remark C.2 (Boundedness of spectral weight function). [Assumptions 3](#) and [C.2](#) yield that

$$\sup_{\theta \in \Theta} \sup_{\omega \in [0, 2\pi]} \|\tilde{J}(\omega; \theta)\| < \infty.$$

To see this, define $B(\theta) \equiv \sup_{\omega \in [0, 2\pi]} \|\tilde{J}(\omega; \theta)\|$, which is finite by [Assumption C.2\(i\)](#). For any $\tilde{\theta}, \theta \in \Theta$,

$$|B(\tilde{\theta}) - B(\theta)| \leq \sup_{\omega \in [0, 2\pi]} \|\tilde{J}(\omega; \tilde{\theta}) - \tilde{J}(\omega; \theta)\| \leq \Delta_{\tilde{J}}(\|\tilde{\theta} - \theta\|),$$

so [Assumption C.2\(ii\)](#) implies B is continuous on compact Θ , hence $\sup_{\theta \in \Theta} B(\theta) < \infty$.

Remark C.3 (Well-definedness and continuity of population moments). Under [Assumptions 2, 3](#) and [C.2](#), $g(\theta)$ is well-defined and continuous on Θ .

To see this, note that under [Assumption 2\(ii\)](#), [Remark C.1](#) yields that $\text{vec}(S(\omega))$ is

integrable on $[0, 2\pi]$. Moreover, under [Assumptions 3](#) and [C.2](#), [Remark C.2](#) implies that

$$\int_0^{2\pi} \|\tilde{J}(\omega; \theta) \text{vec}(S(\omega))\| d\omega < \infty$$

uniformly in θ , so $g(\theta)$ is well-defined.

Fix $\theta \in \Theta$ and let $\tilde{\theta} \rightarrow \theta$. Then, using [Assumption C.2\(ii\)](#),

$$\begin{aligned} \|g(\tilde{\theta}) - g(\theta)\| &\leq \int_0^{2\pi} \|\tilde{J}(\omega; \tilde{\theta}) - \tilde{J}(\omega; \theta)\| \|\text{vec}(S(\omega))\| d\omega \\ &\leq \Delta_{\tilde{J}}(\|\tilde{\theta} - \theta\|) \int_0^{2\pi} \|\text{vec}(S(\omega))\| d\omega \rightarrow 0. \end{aligned}$$

Hence, g is continuous on Θ .

C.2.2 Tools for consistency

Lemma C.1 (Mean of periodogram). *Under [Assumption 2](#),*

$$\sup_{\omega \in [0, 2\pi]} \|E\hat{S}(\omega) - S(\omega)\| \rightarrow 0.$$

This is the matrix analogue of the scalar uniform convergence result under mean zero in [Brockwell and Davis \(1991, Proposition 10.3.1\)](#).

Proof. By [Assumption 2\(ii\)](#), [Remark C.1](#) gives that $S(\omega)$ is continuous and bounded. For the periodogram, the Fejér representation yields

$$E\hat{S}(\omega) = \int_0^{2\pi} \tilde{F}_T(\omega - \lambda) S(\lambda) d\lambda,$$

where $\tilde{F}_T$ is the Fejér kernel; see for example [Stoica and Moses \(2005, Section 2.4\)](#). Since $\tilde{F}_T$ is an approximate identity and S is continuous and periodic on the compact set $[0, 2\pi]$, the claim follows. $\square$

Lemma C.2 (Covariance bound for periodogram: [Brillinger, 1981, Theorem 7.2.2](#)). *Under [Assumption 2](#), there exist a constant $C < \infty$ and a deterministic sequence $c_T \rightarrow 0$ such that, uniformly in $j, j' \in \{0, \dots, T-1\}$,*

$$\left\| \text{Cov} \left(\text{vec}(\hat{S}(\omega_j)), \text{vec}(\hat{S}(\omega_{j'})) \right) \right\| \leq \begin{cases} C, & j = j' \text{ or } j' = T - j, \\ c_T, & \text{otherwise.} \end{cases}$$

Proof. Apply the proof of Theorem 7.2.2 in Brillinger (1981, p. 433), but appeal to his Theorem 4.3.1 instead of Theorem 4.3.2 (pp. 92–93) to bound remainders. $\square$

Lemma C.3 (Pointwise convergence for spectral mean). *Let $m : [0, 2\pi] \rightarrow \mathbb{C}^{d \times d_\zeta^2}$ be bounded and piecewise continuous. Under Assumption 2, the spectral mean*

$$M_T(m) \equiv \frac{2\pi}{T} \sum_{j=0}^{T-1} m(\omega_j) \text{vec}(\hat{S}(\omega_j)) \xrightarrow{p} \int_0^{2\pi} m(\omega) \text{vec}(S(\omega)) d\omega.$$

See Brillinger (1981, Chapter 7.6) for related results with slightly different regularity conditions on m .

Proof. Write

$$M_T(m) - \int m \text{vec}(S) = \underbrace{(M_T(m) - EM_T(m))}_{\text{stochastic term}} + \underbrace{\left(EM_T(m) - \int m \text{vec}(S) \right)}_{\text{bias term}}.$$

(i) *Bias term.* By Lemma C.1,

$$\left\| EM_T(m) - \frac{2\pi}{T} \sum_{j=0}^{T-1} m(\omega_j) \text{vec}(S(\omega_j)) \right\| \leq \sup_{\omega} \|m(\omega)\| \times \frac{2\pi}{T} \sum_{j=0}^{T-1} \|E\hat{S}(\omega_j) - S(\omega_j)\| \rightarrow 0.$$

Since $\omega \mapsto m(\omega) \text{vec}(S(\omega))$ is piecewise continuous and bounded, the Riemann sum converges:

$$\frac{2\pi}{T} \sum_{j=0}^{T-1} m(\omega_j) \text{vec}(S(\omega_j)) \rightarrow \int_0^{2\pi} m(\omega) \text{vec}(S(\omega)) d\omega.$$

Therefore, $EM_T(m) - \int m \text{vec}(S) \rightarrow 0$.

(ii) *Stochastic term.* Under Assumption 2, Lemma C.2 bounds $\|\text{Cov}(\text{vec}(\hat{S}(\omega_j)), \text{vec}(\hat{S}(\omega_{j'})))\|$ uniformly in j, j' . Using boundedness of m , we have

$$\begin{aligned} E \|\text{vec}(M_T(m) - EM_T(m))\|^2 &\leq \left(\frac{2\pi}{T} \right)^2 \sum_{j,j'} \|m(\omega_j)\| \|m(\omega_{j'})\| \left\| \text{Cov}(\text{vec}(\hat{S}(\omega_j)), \text{vec}(\hat{S}(\omega_{j'}))) \right\| \\ &= o(1). \end{aligned}$$

Thus, $M_T(m) - EM_T(m) \xrightarrow{p} 0$ by Chebyshev's inequality.

Finally, combining (i) and (ii) yields the claim. $\square$

Lemma C.4 (Stochastic equicontinuity of sample moments). *Under [Assumptions 2](#) and [C.2](#), the process $\{\hat{g}(\theta) : \theta \in \Theta\}$ is stochastically equicontinuous, i.e., for every $\varepsilon > 0$,*

$$\lim_{\delta \rightarrow 0} \limsup_{T \rightarrow \infty} \Pr \left(\sup_{\|\tilde{\theta} - \theta\| < \delta} \|\hat{g}(\tilde{\theta}) - \hat{g}(\theta)\| > \varepsilon \right) = 0.$$

Proof. For any $\tilde{\theta}, \theta \in \Theta$, by Cauchy-Schwarz,

$$\begin{aligned} \|\hat{g}(\tilde{\theta}) - \hat{g}(\theta)\| &\leq \frac{2\pi}{T} \sum_{j=0}^{T-1} \|\tilde{J}(\omega_j; \tilde{\theta}) - \tilde{J}(\omega_j; \theta)\| \|\text{vec}(\hat{S}(\omega_j))\| \\ &\leq \left(\frac{2\pi}{T} \sum_{j=0}^{T-1} \|\tilde{J}(\omega_j; \tilde{\theta}) - \tilde{J}(\omega_j; \theta)\|^2 \right)^{1/2} \left(\frac{2\pi}{T} \sum_{j=0}^{T-1} \|\hat{S}(\omega_j)\|^2 \right)^{1/2}. \end{aligned}$$

It follows from [Lemmas C.1](#) and [C.2](#) that $E\|\hat{S}(\omega_j)\|^2$ is uniformly bounded in j . Hence, by Markov's inequality,

$$V_T \equiv \frac{2\pi}{T} \sum_{j=0}^{T-1} \|\hat{S}(\omega_j)\|^2 = O_p(1).$$

Under [Assumption C.2\(ii\)](#), $\|\tilde{J}(\omega_j; \tilde{\theta}) - \tilde{J}(\omega_j; \theta)\| \leq \Delta_{\mathcal{J}}(\delta)$ for $\|\tilde{\theta} - \theta\| < \delta$ and all j , hence

$$\sup_{\|\tilde{\theta} - \theta\| < \delta} \|\hat{g}(\tilde{\theta}) - \hat{g}(\theta)\| \leq \sqrt{2\pi} \Delta_{\mathcal{J}}(\delta) \sqrt{V_T}.$$

Fix $\varepsilon > 0$ and $M > 0$. Then

$$\Pr \left(\sup_{\|\tilde{\theta} - \theta\| < \delta} \|\hat{g}(\tilde{\theta}) - \hat{g}(\theta)\| > \varepsilon \right) \leq \Pr \left(\sqrt{V_T} > M \right) + \mathbb{1} \left(\sqrt{2\pi} \Delta_{\mathcal{J}}(\delta) M > \varepsilon \right).$$

As $V_T = O_p(1)$, choose M large to make the first term small uniformly in T . Then choose δ small so that $\sqrt{2\pi} \Delta_{\mathcal{J}}(\delta) M \leq \varepsilon$ using [Assumption C.2\(ii\)](#). This proves stochastic equicontinuity. $\square$

The bound $V_T = O_p(1)$ is sufficient here because we only need to control a fixed-weight, finite-dimensional parametric spectral mean. In contrast, results that control an empirical spectral process uniformly over a rich index class, as in [Dahlhaus \(1988\)](#), typically require stronger smoothness conditions on the spectrum and an entropy restriction on the index class.

Lemma C.5 (Uniform convergence of sample moments). *Under Assumptions 2, 3 and C.2,*

$$\sup_{\theta \in \Theta} \|\hat{g}(\theta) - g(\theta)\| \xrightarrow{P} 0.$$

Proof. Assumption 3 gives compact Θ . Under Assumptions 2, 3 and C.2, Remark C.3 gives continuity of g on Θ . Under Assumptions 2, 3 and C.2, Lemma C.4 gives stochastic equicontinuity.

Under Assumption 2, Lemma C.3 applies with $m(\omega) = \tilde{J}(\omega; \theta)$ for each θ , which is bounded and piecewise continuous by Assumption C.2(i), so $\hat{g}(\theta) \xrightarrow{P} g(\theta)$ pointwise.

Together, the uniform convergence follows from pointwise convergence plus stochastic equicontinuity on compact Θ and continuity of g (Newey and McFadden, 1994, Lemma 2.8). $\square$

C.2.3 Tools for asymptotic normality

Remark C.4 (Continuity of G). Under Assumptions 2 and C.4, $G(\theta)$ is continuous on Θ_0 . The argument is similar to Remark C.3.

Lemma C.6 (Uniform convergence of $\hat{G}$). *Under Assumptions 2 and C.4,*

$$\sup_{\theta \in \Theta_0} \|\hat{G}(\theta) - G(\theta)\| \xrightarrow{P} 0.$$

The proof is similar to that of Lemma C.5.

C.2.4 Tools for HAC and time-domain CLT

Remark C.5 (Assumption C.6 vs. Assumptions C.2 and C.4). Note that:

- Assumption C.6(i) $\Rightarrow$ Assumption C.2(i): by the Weierstrass M-test, $\tilde{J}(\cdot; \theta)$ is bounded and continuous in ω on $[0, 2\pi]$, hence piecewise continuous and Riemann integrable.
- Assumptions 3 and C.6(i) $\Rightarrow$ Assumption C.2(ii): uniform absolute summability plus coefficient continuity on compact set implies the Fourier series is uniformly continuous by the Weierstrass M-test.
- Assumptions C.6(i) and C.6(ii) $\Rightarrow$ Assumptions C.4(i)–C.4(iii): similar argument on $\partial \text{vec}(\tilde{J}(\cdot; \theta))/\partial \theta'$ for $\theta \in \Theta_0$.

- **Assumptions C.2** and **C.4** do not generally imply **Assumption C.6**: even continuous differentiability in ω does not guarantee absolute summability of Fourier coefficients. See **Remark C.6** for a set of sufficient conditions.

Remark C.6 (Sufficient conditions for **Assumption C.6**). Assume that the map $\omega \mapsto \tilde{J}(\omega; \theta)$ is twice differentiable with second derivative continuous in (ω, θ) , and that the following boundary conditions hold: $\tilde{J}(0; \theta) = \tilde{J}(2\pi; \theta)$ and $\partial \tilde{J}(0; \theta)/\partial \omega = \partial \tilde{J}(2\pi; \theta)/\partial \omega$. Then integration by parts gives

$$\tilde{J}_k(\theta) = -\frac{1}{k^2} \times \frac{1}{2\pi} \int_0^{2\pi} e^{-i\omega k} \frac{\partial^2 \tilde{J}(\omega; \theta)}{\partial \omega^2} d\omega,$$

which is easily seen to imply **Assumption C.6(i)**. Similar conditions placed on the map $\omega \mapsto \partial \text{vec}(\tilde{J}(\omega; \theta))/\partial \theta'$ yield **Assumption C.6(ii)**. Finally, under the conditions above, the display above implies (for any indices j_1, j_2)

$$\begin{aligned} \sum_{k=1}^{\infty} k |(\tilde{J}_k(\theta_0))_{j_1, j_2}| &= \sum_{k=1}^{\infty} \frac{1}{k} \times \left| \frac{1}{2\pi} \int_0^{2\pi} e^{-i\omega k} \frac{\partial^2 (\tilde{J}(\omega; \theta_0))_{j_1, j_2}}{\partial \omega^2} d\omega \right| \\ &\leq \left(\sum_{k=1}^{\infty} k^{-2} \right)^{1/2} \times \left(\sum_{k=1}^{\infty} \left| \frac{1}{2\pi} \int_0^{2\pi} e^{-i\omega k} \frac{\partial^2 (\tilde{J}(\omega; \theta_0))_{j_1, j_2}}{\partial \omega^2} d\omega \right|^2 \right)^{1/2} \\ &\leq \left(\sum_{k=1}^{\infty} k^{-2} \right)^{1/2} \times \left(\sum_{k \in \mathbb{Z}} \left| \frac{1}{2\pi} \int_0^{2\pi} e^{-i\omega k} \frac{\partial^2 (\tilde{J}(\omega; \theta_0))_{j_1, j_2}}{\partial \omega^2} d\omega \right|^2 \right)^{1/2} \\ &= \left(\sum_{k=1}^{\infty} k^{-2} \right)^{1/2} \times \left(\frac{1}{2\pi} \int_0^{2\pi} \left| \frac{\partial^2 (\tilde{J}(\omega; \theta_0))_{j_1, j_2}}{\partial \omega^2} \right|^2 d\omega \right)^{1/2}, \end{aligned}$$

where the first inequality uses Cauchy-Schwarz and the final equality uses Parseval. Since $\partial^2 \tilde{J}(\omega; \theta_0)/\partial \omega^2$ was already assumed to be bounded in ω , **Assumption C.6(iii)** follows.

Remark C.7 (Summability of $\Gamma_\psi(h)$). Under **Assumptions 2**, **C.3** and **C.6(i)** at θ_0 , we have that $\sum_{h \in \mathbb{Z}} \|\Gamma_\psi(h)\| < \infty$, so $\Omega \equiv \sum_{h \in \mathbb{Z}} \Gamma_\psi(h)$ is well-defined. Indeed, by **Assumption C.3**, $E[\psi_t(\theta_0)] = g(\theta_0) = 0$, so $\Gamma_\psi(h) = E[\psi_h(\theta_0)\psi_0(\theta_0)'] = \text{Cov}(\psi_h(\theta_0), \psi_0(\theta_0))$, and its absolute summability follows from the argument in **Lemma C.8** below applied to the untruncated score $\psi_t(\theta_0)$.

Lemma C.7 (Truncation error). Let $\mathcal{K}_{t,T}^c \equiv \{k \in \mathbb{Z} : k < -(t-1) \text{ or } k > T-t\}$ and define

the edge-truncation error

$$\eta_{t,T} \equiv \psi_{t,T}(\theta_0) - \psi_t(\theta_0) = - \sum_{k \in \mathcal{K}_{t,T}^c} \tilde{J}_k(\theta_0) \text{vec}(\zeta_{t+k} \zeta_t').$$

Under *Assumption 2(i)* and *Assumption C.6(iii)* at θ_0 ,

$$\sum_{t=1}^T \left(E \|\eta_{t,T}\|^2 \right)^{1/2} = O(1).$$

Proof. Under *Assumption 2(i)*, $E \|\eta_{t,T}\|^2 \leq C \left(\sum_{k \in \mathcal{K}_{t,T}^c} \|\tilde{J}_k(\theta_0)\| \right)^2$ for some $C < \infty$, so

$$\sum_{t=1}^T \left(E \|\eta_{t,T}\|^2 \right)^{1/2} \leq C^{1/2} \sum_{t=1}^T \sum_{k \in \mathcal{K}_{t,T}^c} \|\tilde{J}_k(\theta_0)\|.$$

Under *Assumption C.6(iii)*,

$$\begin{aligned} \sum_{t=1}^T \sum_{k \in \mathcal{K}_{t,T}^c} \|\tilde{J}_k(\theta_0)\| &= \sum_{|k| \leq T-1} |k| \|\tilde{J}_k(\theta_0)\| + T \sum_{|k| \geq T} \|\tilde{J}_k(\theta_0)\| \\ &\leq \sum_{|k| \leq T-1} |k| \|\tilde{J}_k(\theta_0)\| + \sum_{|k| \geq T} |k| \|\tilde{J}_k(\theta_0)\| = O(1), \end{aligned}$$

which gives the result. $\square$

Lemma C.8 (Summability of the centered score autocovariance). *Define*

$$\check{\Gamma}_{\psi,T}(h) \equiv \frac{1}{T} \sum_{t=1}^T \text{Cov}(\psi_{t,T}(\theta_0), \psi_{t-h,T}(\theta_0)).$$

Under *Assumption 2* and *Assumption C.6(i)* at θ_0 , there exists a positive, deterministic sequence $C(h)$, independent of T , such that $\|\check{\Gamma}_{\psi,T}(h)\| \leq C(h)$ for all T, h , and $\sum_{h \in \mathbb{Z}} C(h) < \infty$.

Proof. With $\mathcal{K}_{t,T} \equiv \{-(t-1), \dots, T-t\}$, $\psi_{t,T}(\theta_0) = \sum_{k \in \mathcal{K}_{t,T}} \tilde{J}_k(\theta_0) \text{vec}(\zeta_{t+k} \zeta_t')$. Then

$$\check{\Gamma}_{\psi,T}(h) = \frac{1}{T} \sum_{t=1}^T \sum_{k \in \mathcal{K}_{t,T}} \sum_{\ell \in \mathcal{K}_{t-h,T}} \tilde{J}_k(\theta_0) \text{Cov}(\text{vec}(\zeta_{t+k} \zeta_t'), \text{vec}(\zeta_{t-h+\ell} \zeta_{t-h}')) \tilde{J}_\ell(\theta_0)'$$

Hence, by the triangle inequality, stationarity in *Assumption 2(i)*, and dropping the finite

index restrictions,

$$\|\check{\Gamma}_{\psi,T}(h)\| \leq \sum_{k \in \mathbb{Z}} \sum_{\ell \in \mathbb{Z}} \|\tilde{J}_k(\theta_0)\| \|\tilde{J}_\ell(\theta_0)\| \left\| \text{Cov} \left(\text{vec}(\zeta_k \zeta'_0), \text{vec}(\zeta_{-h+\ell} \zeta'_{-h}) \right) \right\| \equiv C(h),$$

which is independent of T .

To show summability of $C(h)$, note that each entry of $\text{Cov}(\text{vec}(\zeta_k \zeta'_0), \text{vec}(\zeta_{-h+\ell} \zeta'_{-h}))$ is a fourth-order covariance $\text{Cov}(AB, CD)$ with A, B, C, D mean-zero components of ζ , and admits the product cumulant decomposition ([Brillinger, 1981](#), Theorem 2.3.2)

$$\text{Cov}(AB, CD) = \text{cum}(A, B, C, D) + \text{Cov}(A, C) \text{Cov}(B, D) + \text{Cov}(A, D) \text{Cov}(B, C).$$

Thus, by [Assumption 2\(ii\)](#),

$$\sum_{h \in \mathbb{Z}} \left\| \text{Cov} \left(\text{vec}(\zeta_k \zeta'_0), \text{vec}(\zeta_{-h+\ell} \zeta'_{-h}) \right) \right\| \leq M < \infty,$$

uniformly in k, ℓ . Therefore, by [Assumption C.6\(i\)](#),

$$\sum_{h \in \mathbb{Z}} C(h) \leq M \sum_{k, \ell \in \mathbb{Z}} \|\tilde{J}_k(\theta_0)\| \|\tilde{J}_\ell(\theta_0)\| = M \left(\sum_{k \in \mathbb{Z}} \|\tilde{J}_k(\theta_0)\| \right)^2 < \infty. \quad \square$$

Lemma C.9 (Fourth-order dependence of the score). *Under [Assumptions 2\(i\)](#) and [B.2](#) and [Assumption C.6\(i\)](#) at θ_0 , there is a constant $\Lambda < \infty$ such that, for all $h, \tau \in \mathbb{Z}$ and all t, T ,*

$$\sum_{s \in \mathbb{Z}} \left\| \text{Cov}(\text{vec}(\psi_{t,T}(\theta_0) \psi_{t-h,T}(\theta_0)'), \text{vec}(\psi_{s,T}(\theta_0) \psi_{s-\tau,T}(\theta_0)')) \right\| \leq \Lambda.$$

Proof. To economize on notation, we drop the θ_0 argument and pretend that $\tilde{J}$ and ζ_t are scalars (this is without essential loss of generality, since all variables are finite-dimensional anyway). As in the proof of [Lemma C.8](#), we extend all lag sums to $k \in \mathbb{Z}$ after taking absolute values, since dropping the edge restrictions $\mathcal{K}_{t,T}$ only enlarges the bound, and write $\psi_{t,T} = \sum_{k \in \mathbb{Z}} \tilde{J}_k \zeta_{t+k} \zeta_t$. Then

$$|\text{Cov}(\psi_{t,T} \psi_{t-h,T}, \psi_{s,T} \psi_{s-\tau,T})| \leq \sum_{k_1, k_2, \ell_1, \ell_2 \in \mathbb{Z}} |\tilde{J}_{k_1} \tilde{J}_{k_2} \tilde{J}_{\ell_1} \tilde{J}_{\ell_2}| |\text{Cov}(P_{t,h,k_1,k_2}, Q_{s,\tau,\ell_1,\ell_2})|,$$

where $P_{t,h,k_1,k_2} \equiv \zeta_{t+k_1} \zeta_t \zeta_{t-h+k_2} \zeta_{t-h}$ and $Q_{s,\tau,\ell_1,\ell_2} \equiv \zeta_{s+\ell_1} \zeta_s \zeta_{s-\tau+\ell_2} \zeta_{s-\tau}$. The lemma follows

from [Assumption C.6\(i\)](#) at θ_0 if we can show that

$$\sum_{s \in \mathbb{Z}} |\text{Cov}(P_{t,h,k_1,k_2}, Q_{s,\tau,\ell_1,\ell_2})| \leq \Lambda \quad (7)$$

uniformly in $h, \tau, k_1, k_2, \ell_1, \ell_2$.

By the product theorem for cumulants ([Brillinger, 1981](#), Theorem 2.3.2),

$$\text{Cov}(P_{t,h,k_1,k_2}, Q_{s,\tau,\ell_1,\ell_2}) = \sum_{\nu} \text{cum}(\zeta_j; j \in \nu_1) \times \cdots \times \text{cum}(\zeta_j; j \in \nu_p),$$

where the sum is over all those (finitely many) partitions $\nu = \nu_1 \cup \cdots \cup \nu_p$ of the index list $\{t + k_1, t, t - h + k_2, t - h, s + \ell_1, s, s - \tau + \ell_2, s - \tau\}$ that are “indecomposable” in the language of [Brillinger](#): at least one of the component sets—let it be ν_1 upon reordering—must contain at least one index from the t group and at least one from the s group of indices. By [Assumption B.2](#), there exists $C \in [1, \infty)$ (depending only on $E[\zeta_0^8]$) such that $|\text{cum}(\zeta_j; j \in \nu_b)| \leq C$ for all component sets ν_b . Thus,

$$|\text{Cov}(P_{t,h,k_1,k_2}, Q_{s,\tau,\ell_1,\ell_2})| \leq C^6 \sum_{\nu} |\text{cum}(\zeta_j; j \in \nu_1)|. \quad (8)$$

By stationarity ([Assumption 2\(i\)](#)) and the above-mentioned structure of ν_1 , each $\text{cum}(\zeta_j; j \in \nu_1)$ can be written as

$$\text{cum}(\zeta_0, \zeta_{\delta_1}, \dots, \zeta_{\delta_{q_1}}, \zeta_{s-t+\delta_{q_1+1}}, \dots, \zeta_{s-t+\delta_{q_2}})$$

for some integers $\delta_1, \dots, \delta_{q_1}, \delta_{q_1+1}, \dots, \delta_{q_2}$, where $0 \leq q_1 < q_2 \leq 7$. The above cumulant is absolutely summable in s :

$$\sum_{s \in \mathbb{Z}} |\text{cum}(\zeta_0, \zeta_{\delta_1}, \dots, \zeta_{\delta_{q_1}}, \zeta_{s-t+\delta_{q_1+1}}, \dots, \zeta_{s-t+\delta_{q_2}})| \leq \sum_{\tilde{\ell}_1, \dots, \tilde{\ell}_{q_2} \in \mathbb{Z}} |\text{cum}(\zeta_0, \zeta_{\tilde{\ell}_1}, \dots, \zeta_{\tilde{\ell}_{q_2}})|,$$

and the latter sum is finite by [Assumption B.2](#); it is also independent of $h, \tau, k_1, k_2, \ell_1, \ell_2, t, T$. It follows from this result and from (8) that (7) holds. This concludes the proof. $\square$

Remark C.8 (Time-domain CLT). Standard Fourier transform calculations yield

$$\int_0^{2\pi} \tilde{J}(\omega; \theta_0) \text{vec}(\hat{S}(\omega)) d\omega = \frac{1}{T} \sum_{t=1}^T \psi_{t,T}(\theta_0).$$

Since $\hat{g}(\theta_0) = \frac{2\pi}{T} \sum_{j=0}^{T-1} \tilde{J}(\omega_j; \theta_0) \text{vec}(\hat{S}(\omega_j))$ is a Riemann approximation to the integral on the left-hand side above, it can be shown that they differ by $O_p(T^{-1})$ under [Assumptions 2](#) and [C.6](#). Moreover, by [Assumptions 2](#) and [C.6\(iii\)](#), step (iv) in the proof of [Theorem C.3](#) gives that $\frac{1}{T} \sum_{t=1}^T (\psi_t(\theta_0) - \psi_{t,T}(\theta_0)) = o_p(T^{-1/2})$. Putting these together,

$$\hat{g}(\theta_0) = \frac{1}{T} \sum_{t=1}^T \psi_t(\theta_0) + o_p(T^{-1/2}).$$

Consequently, the spectral mean CLT in [Assumption 8](#) follows from any conventional time-domain CLT for the stationary process $\{\psi_t(\theta_0)\}$.

C.3 Specialization to main application

Now we specialize the general spectral GMM results to the application in [Sections 3](#) and [4](#). Recall that this framework corresponds to the spectral weight function $\tilde{J}_{\text{special}}(\omega; \theta) \equiv M_z \otimes (I_{d_y}, -J(\omega; \theta), 0_{d_y \times d_z})$, as discussed at the beginning of [Appendix C](#).

C.3.1 Consistency

Remark C.9 (Conjugate symmetry). Since $J_k(\theta)$ is real, [Assumption C.1](#) is satisfied for $\tilde{J}_{\text{special}}(\omega; \theta)$.

Lemma C.10 (SSJ assumptions imply $\tilde{J}$ regularity). *Under [Assumptions 3](#) and [4](#), $\tilde{J}_{\text{special}}(\omega; \theta)$ satisfies [Assumption C.2](#).*

Proof. Let $\Xi \equiv [0, 2\pi] \times \Theta$. By [Assumption 3](#), Ξ is compact. Let $m_k(\omega; \theta) = e^{i\omega k} J_k(\theta)$ for each $k \geq 0$. By [Assumption 4\(i\)](#), m_k is continuous on Ξ , and by [Assumption 4\(ii\)](#),

$$\sum_{k=0}^{\infty} \sup_{(\omega, \theta) \in \Xi} \|m_k(\omega; \theta)\| = \sum_{k=0}^{\infty} \sup_{\theta \in \Theta} \|J_k(\theta)\| < \infty.$$

Hence, by the Weierstrass M-test, $J(\omega; \theta) \equiv \sum_{k=0}^{\infty} m_k(\omega; \theta) = \sum_{k=0}^{\infty} e^{i\omega k} J_k(\theta)$ converges uniformly on Ξ . Since each m_k is continuous, J is continuous on Ξ , and by Heine-Cantor it is uniformly continuous on Ξ .

For each $\theta \in \Theta$, continuity of $\omega \mapsto J(\omega; \theta)$ on $[0, 2\pi]$ implies that $J(\cdot; \theta)$ is bounded and Riemann integrable in ω . Moreover, the uniform continuity of J on Ξ implies

$$\sup_{\omega \in [0, 2\pi]} \|J(\omega; \tilde{\theta}) - J(\omega; \theta)\| \rightarrow 0, \quad \text{as } \|\tilde{\theta} - \theta\| \rightarrow 0,$$

so $J(\cdot; \theta)$ is uniformly continuous in θ uniformly over ω .

Since $\tilde{J}_{\text{special}}$ is an affine transformation of J , the claim follows. $\square$

Proof of Proposition 1. Note first that

$$g(\theta) = \int_0^{2\pi} \tilde{J}_{\text{special}}(\omega; \theta) \text{vec}(S(\omega)) d\omega = \text{vec} \left(\text{Cov}(y_t, z_t) - \sum_{k=0}^{\infty} J_k(\theta) \text{Cov}(x_{t+k}, z_t) \right),$$

after interchanging the k -sum with the ω -integral (justified by Assumptions 2 and 4). Hence, Assumptions 1 and 5 imply Assumption C.3 with $\tilde{J} = \tilde{J}_{\text{special}}$. Remark C.9 and Lemma C.10 therefore imply that all the assumptions of Theorem C.1 hold. $\square$

C.3.2 Asymptotic normality

Remark C.10 (SSJ assumptions imply $\partial\tilde{J}$ regularity). Under Assumptions 3 and 4 and interiority of θ_0 in Θ , $\tilde{J}_{\text{special}}$ satisfies Assumption C.4 with Θ_0 given by any compact neighborhood of θ_0 contained in Θ . The argument uses the same M-test and Heine-Cantor steps as in Lemma C.10.

Proof of Proposition 2. As in the proof of Proposition 1, Assumptions 1, 2, 4 and 5 imply Assumption C.3, and a similar argument shows that Assumptions 2, 4 and 6 imply Assumption C.5, both with $\tilde{J} = \tilde{J}_{\text{special}}$. Hence, Remarks C.9 and C.10 and Lemma C.10 imply that all the assumptions of Theorem C.2 hold. $\square$

C.3.3 HAC

Proof of Proposition B.1. Note first that the general HAC estimator $\hat{\Omega}$ from Section C.1.3 reduces to the special HAC estimator in Appendix B when $\tilde{J} = \tilde{J}_{\text{special}}$, since

$$\tilde{J}_{k,\text{special}}(\theta) \equiv \frac{1}{2\pi} \int_0^{2\pi} e^{-i\omega k} \tilde{J}_{\text{special}}(\omega; \theta) d\omega = M_z \otimes \left(\mathbb{1}\{k=0\} \times I, -\mathbb{1}\{k \geq 0\} \times J_k(\theta), 0 \right).$$

Next, Remark C.9 gives Assumption C.1. Moreover, with $\tilde{J} = \tilde{J}_{\text{special}}$, Assumptions 1, 2, 4 and 5 imply Assumption C.3, Assumptions 2, 4 and 6 imply Assumption C.5, Assumption 4 implies Assumptions C.6(i) and C.6(ii), and Assumption B.3 implies Assumption C.6(iii). Thus, the conclusion follows from Theorem C.3. $\square$

References

- ACHARYA, S., W. CHEN, M. DEL NEGRO, K. DOGRA, A. GLEICH, S. GOYAL, E. MATLIN, D. LEE, R. SARFATI, AND S. SENGUPTA (2023): “Estimating HANK for Central Banks,” Discussion Paper 1071, Federal Reserve Bank of New York.
- ALMUZARA, M., M. ARELLANO, R. BLUNDELL, AND S. BONHOMME (2025): “Nonlinear Micro Income Processes with Macro Shocks,” Discussion Paper 1162, Federal Reserve Bank of New York.
- ALTONJI, J. G., AND L. M. SEGAL (1996): “Small-Sample Bias in GMM Estimation of Covariance Structures,” *Journal of Business & Economic Statistics*, 14(3), 353–366.
- ANDREWS, D. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59(3), 817–858.
- AUCLERT, A., B. BARDÓCZY, M. ROGNLIE, AND L. STRAUB (2021): “Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models,” *Econometrica*, 89(5), 2375–2408.
- AUCLERT, A., M. ROGNLIE, AND L. STRAUB (2025): “Fiscal and Monetary Policy with Heterogeneous Agents,” *Annual Review of Economics*, 17, 539–562.
- BARNICHON, R., AND G. MESTERS (2020): “Identifying Modern Macro Equations with Old Shocks,” *Quarterly Journal of Economics*, 135(4), 2255–2298.
- BAUER, M. D., AND E. T. SWANSON (2023): “An Alternative Explanation for the “Fed Information Effect”,” *American Economic Review*, 113(3), 664–700.
- BAYER, C., B. BORN, AND R. LUETTICKE (2024): “Shocks, Frictions, and Inequality in US Business Cycles,” *American Economic Review*, 114(5), 1211–1247.
- BOPPART, T., P. KRUSELL, AND K. MITMAN (2018): “Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative,” *Journal of Economic Dynamics and Control*, 89, 68–92.
- BRILLINGER, D. R. (1981): *Time Series: Data Analysis and Theory*, no. 36 in Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

- BROCKWELL, P. J., AND R. A. DAVIS (1991): *Time Series: Theory and Methods*, Springer Series in Statistics. Springer, 2nd edn.
- CARROLL, C. D. (2001): “Death to the Log-Linearized Consumption Euler Equation! (And Very Poor Health to the Second-Order Approximation),” *Advances in Macroeconomics*, 1(1), 1–38.
- CHANG, M., X. CHEN, AND F. SCHORFHEIDE (2024): “Heterogeneity and Aggregate Fluctuations,” *Journal of Political Economy*, 132(12), 4021–4067.
- CHANG, Y., S. KIM, AND J. PARK (2025): “How Do Macroaggregates and Income Distribution Interact Dynamically? A Novel Structural Mixed Autoregression with Aggregate and Functional Variables,” Discussion Paper 5141761, SSRN.
- DAHLHAUS, R. (1988): “Empirical spectral processes and their applications to time series analysis,” *Stochastic Processes and their Applications*, 30(1), 69–83.
- DEATON, A. (1985): “Panel Data from Time Series of Cross-Sections,” *Journal of Econometrics*, 30(1–2), 109–126.
- DEBORTOLI, D., AND J. GALÍ (2025): “Heterogeneity and Aggregate Consumption: An Empirical Assessment,” Manuscript, CREI.
- ETTMEIER, S., C. H. KIM, AND F. SCHORFHEIDE (2024): “Measuring the Effects of Aggregate Shocks on Cross-sectional Distributions: Functional vs. Panel Approach,” Presentation slides.
- FERNÁNDEZ-VILLAYERDE, J., S. HURTADO, AND G. NUÑO (2023): “Financial Frictions and the Wealth Distribution,” *Econometrica*, 91(3), 869–901.
- FUHRER, J. (2017): “Expectations as a source of macroeconomic persistence: Evidence from survey expectations in a dynamic macro model,” *Journal of Monetary Economics*, 86, 22–35.
- GAGLIARDONE, L., M. GERTLER, S. LENZU, AND J. TIELENS (2025): “Anatomy of the Phillips Curve: Micro Evidence and Macro Implications,” *American Economic Review*, 115(11), 3941–3974.
- GALÍ, J., AND M. GERTLER (1999): “Inflation Dynamics: A Structural Econometric Analysis,” *Journal of Monetary Economics*, 44(2), 195–222.

- GIACOMINI, R., D. N. POLITIS, AND H. WHITE (2013): “A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators,” *Econometric Theory*, 29(3), 567–589.
- HANSEN, L. P., AND K. J. SINGLETON (1982): “Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models,” *Econometrica*, 50(5), 1269–1286.
- HAYASHI, F. (1985): “The Effect of Liquidity Constraints on Consumption: A Cross-Sectional Analysis,” *Quarterly Journal of Economics*, 100(1), 183–206.
- KAPLAN, G., B. MOLL, AND G. L. VIOLANTE (2018): “Monetary Policy According to HANK,” *American Economic Review*, 108(3), 697–743.
- LIU, L., AND M. PLAGBORG-MØLLER (2023): “Full-Information Estimation of Heterogeneous Agent Models Using Macro and Micro Data,” *Quantitative Economics*, 14(1), 1–35.
- MCKAY, A., AND C. K. WOLF (2023): “What Can Time-Series Regressions Tell Us About Policy Counterfactuals?,” *Econometrica*, 91(5), 1695–1725.
- MEYER, M., AND E. PAPARODITIS (2023): “A frequency domain bootstrap for general multivariate stationary processes,” *Bernoulli*, 29(3), 2367–2391.
- MONGEY, S., AND J. WILLIAMS (2017): “Firm Dispersion and Business Cycles: Estimating Aggregate Shocks Using Panel Data,” Unpublished manuscript.
- NEWKEY, W. K., AND D. L. MCFADDEN (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, chap. 36, pp. 2111–2245. Elsevier Science B.V.
- PALLOTTI, F., G. PAZ-PARDO, J. SLACALEK, O. TRISTANI, AND G. L. VIOLANTE (2024): “Who bears the costs of inflation? Euro area households and the 2021–2023 shock,” *Journal of Monetary Economics*, 148, 103671, Special issue: “Inflation in the COVID Era and Beyond”.
- PAPP, T. K., AND M. REITER (2020): “Estimating linearized heterogeneous agent models using panel data,” *Journal of Economic Dynamics and Control*, 115, 103881.
- PRIESTLEY, M. (1981): *Spectral Analysis and Time Series*, Probability and Mathematical Statistics. Elsevier Academic Press.

- RAMEY, V. A. (2016): “Macroeconomic Shocks and Their Propagation,” in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and H. Uhlig, vol. 2, chap. 2, pp. 71–162. Elsevier B.V.
- REITER, M. (2009): “Solving heterogeneous-agent models by projection and perturbation,” *Journal of Economic Dynamics and Control*, 33(3), 649–665.
- ROTEMBERG, J. J., AND M. WOODFORD (1997): “An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy,” *NBER Macroeconomics Annual*, 12, 297–346.
- STOICA, P., AND R. L. MOSES (2005): *Spectral Analysis of Signals*. Prentice Hall.
- TRYPHONIDES, A. (2023): “Identifying Preferences when Households Are Financially Constrained,” *Review of Economic Dynamics*, 51, 521–546.
- WINBERRY, T. (2018): “A Method for Solving and Estimating Heterogeneous Agent Macro Models,” *Quantitative Economics*, 9(3), 1123–1151.
- WINBERRY, T., A. AUCLERT, M. ROGNLIE, AND L. STRAUB (2025): “New Keynesian Economics with Household and Firm Heterogeneity,” Working Paper 34611, National Bureau of Economic Research.
- YOGO, M. (2004): “Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak,” *Review of Economics and Statistics*, 86(3), 797–810.