Estimation of continuous-time linear DSGE models from discrete-time measurements

Bent Jesper Christensen* Luca Neri† Juan Carlos Parra-Alvarez‡

October 2, 2024

Abstract

We provide a general state space framework for estimation of the parameters of continuous-time linear DSGE models from discrete-time data. Our approach relies on the exact discrete-time representation of the equilibrium dynamics, hence avoiding discretization errors. We construct the exact likelihood for data sampled either as stocks or flows, based on the Kalman filter, and provide necessary and sufficient conditions for local identification of the frequency-invariant structural parameters of the underlying continuous-time model. We recover the unobserved structural shocks at measurement times from the reduced-form residuals in the state space representation by exploiting the underlying causal links implied by the economic model. We illustrate our approach using an off-the-shelf real business cycle model. Extensive Monte Carlo experiments show that the finite sample properties of our estimator are superior to those of an estimator relying on a naive Euler-Maruyama discretization of the economic model. In an application to postwar U.S. macroeconomic data, we estimate the model using series sampled at mixed frequencies, and combinations of series sampled as stocks and flows, and we provide a historical decomposition of the effects of shocks on observables into those stemming from structural supply and demand shocks.

Keywords: DSGE models, continuous time, exact discrete-time state space representation, local identification, structural shocks, stock and flow variables, mixed frequency data.

JEL classification: C13, C32, C68, E13, E32, J22

*Corresponding author. Department of Economics and Business Economics, Aarhus University; CREATES; Danish Finance Institute; Fuglesangs Allé 4, 8210 Aarhus V, Denmark; Email address: bjchristensen@econ.au.dk

†Department of Economic Sciences, University of Bologna; CREATES; CORE, UCLouvain, Belgium; Piazza A. Scaravilli 2, 40126 Bologna, Italy; Email address: luca.neri@uclouvain.be

‡Department of Economics and Business Economics, Aarhus University; CREATES; Danish Finance Institute; Fuglesangs Allé 4, 8210 Aarhus V, Denmark; Email address: jparra@econ.au.dk

1. Introduction

Dynamic stochastic general equilibrium (DSGE) models have become the workhorse for macroeconomic analysis. Over the last decade, continuous-time methods have been gaining increasing interest, with a large number of models being developed for the study of the transmission mechanisms of monetary and fiscal policies, and the interactions between the real and financial sides of the economy (e.g., Brunnermeier and Sannikov, 2014, Kaplan et al., 2018, Posch, 2020, Liemen and Posch, 2022, Fernández-Villaverde et al., 2023). This increase in popularity has led to the development and improvement of numerical methods for approximation of the solutions of both representative and heterogeneous agent continuous-time models (e.g., Posch and Trimborn, 2013, Ahn et al., 2018, Parra-Alvarez et al., 2021, Achdou et al., 2022). However, considerably less work has been devoted to the study of how to take these models to the data. This step requires addressing the timing mismatch, with the economic theory assuming that time evolves continuously, and the data measured at discrete points in time. The issue is particularly pressing in macroeconomics, with measurements usually only available at low frequencies, e.g., monthly, quarterly, or annually.

In this paper, we propose and analyze a method for estimation and inference from multivariate linear continuous-time structural economic models based on discrete-time observations. The approach accounts for the discrete nature of the sampling scheme, while keeping the underlying probabilistic structure unaltered. The discrete-time data generating process results from a combination of the discrete sampling scheme and the underlying continuous-time model. As the decision intervals of economic agents are not tied to the observation intervals of sampled data, the frequency of measurement is not relevant for the parameters of interest, namely, those of the continuous-time model. Relying on the latter, our approach provides a logically consistent basis for jointly accommodating variables sampled at different (mixed) frequencies, and according to different schemes, such as stocks, flows, or combinations of these. A stock is a variable measured at a given point in time, e.g., a price or an interest rate at the end of the period, whereas a flow is measured as the accumulated amount from beginning to end of the period, e.g., consumption or GDP. Upon estimation, the frequency-invariant structural parameters can, if desired, be mapped to parameters associated with any discrete frequency of interest, not only that of the observed data.

The starting point of our development is the system of equations characterizing the solution of linear(-ized) continuous-time DSGE models, usually a system of multiple linear stochastic differential equations (SDEs) describing the dynamics of the optimal state variables in the economy, together with a set of algebraic equations representing equilibrium conditions, optimal policy functions, or (static) no-arbitrage conditions. We obtain the so-

lution to the system at the discrete observation times in the form of a vector autoregressive model of order one, VAR(1). When combined with the algebraic equations evaluated at the same discrete points in time, this delivers the distribution of the observations, while maintaining the cross-equation restrictions implied by the rational expectations solution of the economic model, and without introducing discretization errors that might otherwise contaminate parameter. We express the solution as a time-invariant *exact discrete state space representation* (ED-SSR). Reliance on the ED-SSR implies that all relevant information in the data is retained, and allows evaluating the likelihood function using the Kalman filter.

For estimation, we consider a likelihood-based framework, along the lines of Harvey and Stock (1985) and Zadrozny (1988). Consistency of the maximum likelihood estimator (MLE) requires local identification of the structural parameters of the underlying continuous-time model. We derive necessary and sufficient conditions for local identification from the second-order moments of the discrete-time observations by exploiting the links between (i) the deep parameters of the model, (ii) the reduced-form parameters of the continuous-time rational expectations solution, and (iii) the reduced-form parameters of the discrete-time state space representation. Our approach to local identification extends that of Komunjer and Ng (2011) for discrete-time linear DSGE models. In the latter, the reduced-form parameters of the (discrete-time) rational expectations solution correspond closely to those of the state space representation in that setting, whereas we obtain additional dimensions for satisfying the necessary and sufficient rank condition for local identification of the parameters of the underlying continuous-time model by distinguishing (ii) and (iii). Moreover, in the absence of local aliases (e.g., under a standard condition on oscillations of Phillips, 1973), the local identification conditions are (partly or fully) recast in terms of the coefficient matrices of the underlying continuous-time system and their parameter derivatives, as opposed to the reduced-form (Kalman filter) matrices of the discrete-time state space representation.

Given the discrete-time nature of the data, the residuals from the empirical model are composites of the primitive structural shocks in the underlying continuous-time model, reflecting both time-aggregation and the contemporaneous and dynamic relations among variables between measurements. Therefore, we propose a method to approximately recover the structural shocks at measurement times from the reduced-form residuals, exploiting the causal links of the continuous-time model. Ability to back out the structural shocks is important, e.g., for understanding the sources of business cycle fluctuations, the propagation of exogenous shocks through the economy, and macroeconomic policy effects. Our strategy resembles the use of short- and long-run identifying restrictions on the variance-covariance matrix common in the structural VAR literature, as pioneered by Sims (1986), Bernanke (1986), Shapiro and Watson (1988), and Blanchard and Quah (1989).

The state space approach to inference on continuous-time models from discrete-time observations goes back to Jones (1981). The framework accommodates unobserved components, measurement errors, mixed observation frequencies, missing observations, and unequally spaced observation intervals. Alternative approaches not relying on state space representations have also been proposed, including moment-based methods (Hansen and Scheinkman, 1995), likelihood-based methods (Bergstrom, 1983, Lo, 1988, Aït-Sahalia, 2002), estimating functions (Bibby and Sørensen, 1995), and nonparametric techniques (Florens-Zmirou, 1993, Jiang and Knight, 1997). Phillips and Yu (2009) provide an overview of various approaches. Recent applications to DSGE models include Posch (2009), Christensen et al. (2016), Fernández-Villaverde et al. (2023), and Chambers et al. (2023).

To investigate the properties of our approach, we consider a continuous-time version of the RBC model with indivisible labor of Hansen (1985), with shocks to capital and total factor productivity (TFP), as an off-the-shelf benchmark. We verify our local identification conditions for this model, and conduct extensive Monte Carlo experiments to study the finite sample properties of the MLE, as well as the ability to recover the structural shocks, with data sampled as stocks, flows, or a mixture. The simulations shed light on the effects of model misspecification resulting, e.g., from using a state space representation for stock variables when the observed data are in fact sampled as flows, and the consequences of replacing the ED-SSR by an alternative state space representation derived from a naive Euler-Maruyama (EM) discretization of the equilibrium dynamics. The results show that our approach delivers accurate estimates, even when applying the estimation procedure based on the linearized model to data from the DGP for the true nonlinear model. It is preferred over the EM alternative, which suffers from discretization error, implying that the estimates in effect are of discrete-time parameters only associated with the continuous-time parameters of interest up to the degree of approximation, and this introduces substantial biases. Our results on the directions and relative magnitudes of finite sample biases in our multivariate state space case are consistent with previous results obtained for simpler (e.g., univariate, and/or perfectly observed) processes. Similarly, while our approach recovers the unobserved structural shocks with great precision, the EM approximation performs poorly in comparison, due to discretization error in the variance-covariance structure. Although the approximation is popular in finance due to the availability of relatively high-frequency (e.g., daily, intradaily) data, the comparatively low frequency of macroeconomic data is particularly damaging to the EM method.

We provide an empirical illustration of our approach by estimating the benchmark model using quarterly data on aggregate real consumption and hours worked for the U.S economy over the period from 1960:Q1 through 2019:Q4. The application verifies the feasibility of our method, and estimates are consistent with those in the business cycle literature. A historical

shock decomposition indicates that the U.S. business cycle has mainly been driven by aggregate supply shocks, as indicated by the dominant contribution of recovered TFP shocks to consumption growth, whereas deviations in hours worked from the steady state mainly have reflected aggregate demand shocks, as identified with the recovered shocks to capital accumulation. Drawing on the frequency-invariance of the parameters of the underlying continuous-time model, the application illustrates the use of mixed-frequency data series by including monthly consumption expenditures, along with quarterly hours, with consistent results across common and mixed frequencies. Finally, the use of mixed stock and flow variables is illustrated by including the real interest rate as proxy for the rental rate of capital.

The paper is laid out as follows. Section 2 introduces the mapping between the continuous-time DSGE model and the exact discrete-time state space representation, as well as the construction of the likelihood function. Section 3 presents the local identification analysis, and Section 4 our method for recovering structural shocks. Section 5 shows the results of the Monte Carlo study, and Section 6 the empirical application. Section 7 concludes. All proofs, derivations, and some additional results are provided in the Online Appendix.[1]

2. The econometric framework

2.1 The economic model

Let $\mathbf{y}(t) \in \mathbb{R}^{m_y}$ denote a vector of control or jump variables at time t, and $\mathbf{x}(t) \in \mathbb{R}^{m_x}$ a vector of possibly unobserved state or predetermined variables, $t \in \mathbb{R}$. Further, let $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{m_\theta}$ be the vector of structural parameters characterizing preferences, technology, and/or endowments, with $\boldsymbol{\Theta}$ the parameter space of all theoretically admissible values of $\boldsymbol{\theta}$. We consider a class of continuous-time linear(-ized) DGSE models whose rational expectations solution can be represented as

$$\mathrm{d}\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t)\mathrm{d}t + \mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(t)\,, \quad \text{given } \mathbf{x}(t_0) = \mathbf{x}_0\,, \tag{2.1}$$

$$\mathbf{y}(t) = \mathbf{C}(\boldsymbol{\theta})\mathbf{x}(t)\,, \tag{2.2}$$

where $\mathbf{w}(t) \in \mathbb{R}^{m_w}$ is a vector of independent standard Brownian motions, $\mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, t\mathbf{I})$. By (2.1), $\mathbf{x}(t)$ is governed by a multivariate Ornstein-Uhlenbeck (OU) diffusion process describing the *optimal* dynamics of the state variables, with local drift matrix $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{R}^{m_x \times m_x}$, diffusion matrix $\mathbf{B}(\boldsymbol{\theta}) \in \mathbb{R}^{m_x \times m_w}$, $m_x \geq m_w$, and positive semi-definite instantaneous covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top \in \mathbb{R}^{m_x \times m_x}$. The interpretation of $\mathrm{d}\mathbf{w}(t)$ is that of primitive

[1] Codes for the empirical analysis, including a toolbox for checking our local identification conditions, are available from authors' homepages.

structural shocks driving the economy. Equation (2.2) determines the *optimal* values of the control variables as functions of the state variables, with reaction or loading matrix $\mathbf{C}(\boldsymbol{\theta}) \in \mathbb{R}^{m_y \times m_x}$. The entries of $\mathbf{A}(\boldsymbol{\theta})$, $\mathbf{B}(\boldsymbol{\theta})$, and $\mathbf{C}(\boldsymbol{\theta})$ are time-invariant reduced-form parameters and, in turn, nonlinear functions of the structural parameters $\boldsymbol{\theta}$, with the map

$$\boldsymbol{\theta} \mapsto \boldsymbol{\Lambda}_c(\boldsymbol{\theta}) = (\mathbf{A}(\boldsymbol{\theta}), \mathbf{B}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta})) \tag{2.3}$$

embodying the cross-equation restrictions imposed by the rational expectations solution. We adopt the following assumption, such that equations (2.1)-(2.2) define a stable linear system.

Assumption 1 (Stability)**.** For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the system matrix $\mathbf{A}(\boldsymbol{\theta})$ is full rank and stable, i.e., every eigenvalue of $\mathbf{A}(\boldsymbol{\theta})$ has strictly negative real part.

Two issues arise when considering statistical inference on the continuous-time model (2.1)-(2.2). First, measurements are available in discrete time, only, so it is important to distinguish variables measured as stocks and flows. Secondly, some variables in the model may be latent, or unobservable. We illustrate our approach by assuming that all variables in $\mathbf{x}(t)$ are unobserved, while $\mathbf{y}(t)$ contains a mixture of stock and flow measurements.

2.2 State space representation

Let $\mathbf{x}_\tau^s = \mathbf{x}(t_\tau)$ denote the n_x^s-vector containing the realizations of the continuous-time state variables $\mathbf{x}(t)$ at measurement time t_τ, $\tau \in \mathbb{Z}^+$, and let $\mathbf{x}^f(t_\tau) = \int_{t_{\tau-1}}^{t_\tau} \mathbf{x}(u)\mathrm{d}u$ be the cumulated values over the time interval $(t_{\tau-1}, t_\tau]$.[2] Since each observed variable is sampled either as a stock or a flow, we may order the n_y measurements as $\mathbf{y}_\tau = \left[\mathbf{y}_\tau^{s\top}, \mathbf{y}_\tau^{f\top}\right]^\top$, with $\mathbf{y}_\tau^s = \mathbf{y}^s(t_\tau)$ the n_y^s stocks, and $\mathbf{y}_\tau^f = \int_{t_{\tau-1}}^{t_\tau} \mathbf{y}^f(u)\mathrm{d}u$ the n_y^f flows, $n_y = m_y = n_y^s + n_y^f$. Thus, $\mathbf{y}_\tau^s = \mathbf{S}_y^s \mathbf{y}_\tau$, and $\mathbf{y}_\tau^f = \mathbf{S}_y^f \mathbf{y}_\tau$, using the $n_y^s \times n_y$ and $n_y^f \times n_y$ matrices

$$\mathbf{S}_y^s = \left[\begin{array}{cc} \mathbf{I}_{n_y^s} & \mathbf{0}_{n_y^s \times n_y^f} \end{array}\right], \qquad \mathbf{S}_y^f = \left[\begin{array}{cc} \mathbf{0}_{n_y^f \times n_y^s} & \mathbf{I}_{n_y^f} \end{array}\right]$$

selecting the stock and flow variables. By (2.2), $\mathbf{y}_\tau^f = \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) \mathbf{x}^f(t_\tau)$, and we let $\mathbf{x}_\tau^f = \mathbf{S}_x^f \mathbf{x}^f(t_\tau)$, with the $n_x^f \times n_x^s$ matrix $\mathbf{S}_x^f$ selecting the n_x^f elements of $\mathbf{x}^f(t_\tau)$ that the flow measurements $\mathbf{y}_\tau^f$ depend on, through $\mathbf{C}(\boldsymbol{\theta})$.[3] Thus, $\mathbf{y}_\tau^f = \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) \mathbf{S}_x^{f\top} \mathbf{S}_x^f \mathbf{x}^f(t_\tau) = \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) \mathbf{S}_x^{f\top} \mathbf{x}_\tau^f$.

To match the continuous-time linear model with the discrete-time nature of the data, we introduce an exact time-invariant discrete-time Gaussian state space representation that is consistent with the observations being generated by the system (2.1)-(2.2) at a fixed sampling

[2] We use m for the dimensions in the continuous-time model, and n when referring to the discrete-time realizations. Thus, $n_x^s = m_x$.

[3] $\mathbf{S}_x^f$ selects the entries of $\mathbf{x}^f(t_\tau)$ corresponding to non-zero columns of $\mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta})$.

frequency determined by the spacing between observations, $h = t_\tau - t_{\tau-1} > 0$. First, using arguments similar to those in Phillips (1959, 1973), Bergstrom (1966, 1984), and Harvey (1990), we derive a discrete-time VAR(1) process for the n_x^s state variables $\mathbf{x}_\tau^s$ from the exact solution to the SDE system (2.1) at measurement times t_τ. Next, we compute the n_x^f flows $\mathbf{x}_\tau^f$ by aggregating the stocks over the time interval $(t_{\tau-1}, t_\tau]$, and collect all state variables into the $n_x = n_x^s + n_x^f$ vector $\mathbf{x}_\tau = \left[\mathbf{x}_\tau^{s\top}, \mathbf{x}_\tau^{f\top}\right]^\top$.[4] Finally, we evaluate the policy function (2.2) at t_τ to compute the corresponding stock and flow measurements of the control variables, $\mathbf{y}_\tau = \left[\mathbf{y}_\tau^{s\top}, \mathbf{y}_\tau^{f\top}\right]^\top$. The resulting representation is summarized in the following proposition.

Proposition 1 (Exact discrete state space representation)**.** *Let* $\mathbf{x}(t)$ *and* $\mathbf{y}(t)$ *be generated by the continuous-time system* (2.1)-(2.2)*. Under Assumption 1, the stocks and flows at measurement times,* $\{\mathbf{x}_\tau^s, \mathbf{y}_\tau^s\}$ *and* $\{\mathbf{x}_\tau^f, \mathbf{y}_\tau^f\}$*, satisfy the discrete-time state-space representation*

$$\begin{bmatrix} \mathbf{x}_\tau^s \\ \mathbf{x}_\tau^f \end{bmatrix} = \begin{bmatrix} \mathbf{A}_h(\boldsymbol{\theta}) & \mathbf{0}_{n_x^s \times n_x^f} \\ \mathbf{S}_x^f \mathbf{A}(\boldsymbol{\theta})^{-1}(\mathbf{A}_h(\boldsymbol{\theta}) - \mathbf{I}_{n_x^s}) & \mathbf{0}_{n_x^f \times n_x^f} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\tau-1}^s \\ \mathbf{x}_{\tau-1}^f \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_\tau^s \\ \boldsymbol{\eta}_\tau^f \end{bmatrix}, \tag{2.4}$$

$$\begin{bmatrix} \mathbf{y}_\tau^s \\ \mathbf{y}_\tau^f \end{bmatrix} = \begin{bmatrix} \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}) & \mathbf{0}_{n_y^s \times n_x^f} \\ \mathbf{0}_{n_y^f \times n_x^s} & \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) \mathbf{S}_x^{f\top} \end{bmatrix} \begin{bmatrix} \mathbf{x}_\tau^s \\ \mathbf{x}_\tau^f \end{bmatrix}, \tag{2.5}$$

where $\mathbf{A}_h(\boldsymbol{\theta}) \in \mathbb{R}^{n_x^s \times n_x^s}$ *is given by the matrix exponential*

$$\mathbf{A}_h(\boldsymbol{\theta}) = \exp\left(\mathbf{A}(\boldsymbol{\theta})h\right) = \sum_{i=0}^{\infty} \frac{(\mathbf{A}(\boldsymbol{\theta})h)^i}{i!} = \mathbf{I}_{n_x^s} + \mathbf{A}(\boldsymbol{\theta})h + \tfrac{1}{2}\mathbf{A}(\boldsymbol{\theta})^2 h^2 + \dots, \tag{2.6}$$

and the n_x*-vector of reduced-form disturbances is*

$$\boldsymbol{\eta}_\tau = \begin{bmatrix} \boldsymbol{\eta}_\tau^s \\ \boldsymbol{\eta}_\tau^f \end{bmatrix} = \begin{bmatrix} \int_{t_{\tau-1}}^{t_\tau} \exp(\mathbf{A}(\boldsymbol{\theta})(t_\tau - u))\mathbf{B}(\boldsymbol{\theta})\, d\mathbf{w}(u) \\ \mathbf{S}_x^f \int_{t_{\tau-1}}^{t_\tau} \mathbf{A}(\boldsymbol{\theta})^{-1} \left(\exp\left(\mathbf{A}(\boldsymbol{\theta})(t_\tau - u)\right) - \mathbf{I}_{n_x^s}\right) \mathbf{B}(\boldsymbol{\theta})\, d\mathbf{w}(u) \end{bmatrix}, \tag{2.7}$$

with mean $\mathbb{E}\left[\boldsymbol{\eta}_\tau\right] = \mathbf{0}_{n_x \times 1}$*, covariance matrix*

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta},h}(\boldsymbol{\theta}) = \mathbb{E}\left[\boldsymbol{\eta}_\tau \boldsymbol{\eta}_\tau^\top\right] = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}) \\ \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta})^\top & \boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}) \end{bmatrix}, \tag{2.8}$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}) = \int_0^h \exp\left(\mathbf{A}(\boldsymbol{\theta})(h-u)\right) \boldsymbol{\Sigma}(\boldsymbol{\theta}) \exp\left(\mathbf{A}(\boldsymbol{\theta})^\top (h-u)\right) du\,, \tag{2.9}$$

[4] Thus, $n_x \geq m_x$, since $m_x = n_x^s$.

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}) = \int_0^h \int_0^u \exp(\mathbf{A}(\boldsymbol{\theta})(u-r))\boldsymbol{\Sigma}(\boldsymbol{\theta})\exp(\mathbf{A}(\boldsymbol{\theta})^\top r)\,dr\,du\;\mathbf{S}_x^{f\top}, \tag{2.10}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}) = \mathbf{S}_x^f \int_0^h \int_0^u \exp(\mathbf{A}(\boldsymbol{\theta})r)\boldsymbol{\Sigma}(\boldsymbol{\theta})\exp(\mathbf{A}(\boldsymbol{\theta})^\top r)\,dr\,du\;\mathbf{S}_x^{f\top}, \tag{2.11}$$

and autocovariance matrices $\mathbb{E}\left[\boldsymbol{\eta}_\tau\boldsymbol{\eta}_{\tau-\ell}^\top\right] = \mathbf{0}_{n_x\times n_x}$, $\ell \in \mathbb{Z}\setminus\{0\}$.

We refer to (2.4)-(2.5) as the *Exact Discrete State Space Representation* (ED-SSR) with sampling frequency h. Throughout, the time unit is taken as one year, so $h = 1$ refers to annual, $h = 1/4$ to quarterly, and $h = 1/12$ to monthly observations, etc. Although the ED-SSR does not characterize the behavior of variables between measurements, it is exact in the sense that there is no discretization error in (2.4)-(2.5), regardless of h.[5]

The ED-SSR transition equation (2.4) differs from a standard VAR(1) in that the reduced-form parameters in the transition matrix are restricted in a nonlinear manner involving the matrix exponential $\mathbf{A}_h(\boldsymbol{\theta})$ from (2.6), and the covariance matrix (2.8) depends on both $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\theta})$.[6] Further, by (2.7), the error term $\boldsymbol{\eta}_\tau$ does not represent structural shocks, but instead a reduced-form disturbance—a moving average of the structural shocks $\mathrm{d}\mathbf{w}(u)$ over an interval of length h, with time-variation in weights determined by $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\theta})$.

In Proposition 1, the assumption that all the state variables $\mathbf{x}(t)$ are latent allows deriving the dynamics of the flow state variables by integrating those of the stock state variables (cf. Harvey and Stock, 1985). This in turn implies that the exact discretization of the flow state variable dynamics has a VAR(1) representation. In contrast, Bergstrom (1983, 1984) considers a system with purely observed $\mathbf{x}_\tau$ and no $\mathbf{y}_\tau$, leading to a VARMA(1,1) representation for the flows.[7] Crucially, the two approaches are equivalent, although they result in different dimensions of the state vector (see Online Appendix C).

The measurement equation (2.5) relates the observables to the latent state variables. We refer to an ED-SSR representation with all observables sampled as stocks ($n_y^f = 0$) as an *S-SSR*, one with all observables sampled as flows ($n_y^s = 0$) as an *F-SSR*, and one including both stock and flow measurements ($n_y^s > 0$, $n_y^f > 0$) as an *MX-SSR* representation.

[5] For univariate processes in which the volatility in (2.1) is state dependent, e.g., square root processes, Nowman (1997) proposes an approximate discretization by assuming that volatility changes only at measurement times, $\mathbf{B}(\mathbf{x}(t),\boldsymbol{\theta}) = \mathbf{B}(\mathbf{x}(t_{\tau-1}),\boldsymbol{\theta})$, for $t \in [t_{\tau-1}, t_\tau)$.

[6] In Online Appendix B, we show how to compute these matrices using an eigendecomposition and the matrix factorization approach of Van Loan (1978).

[7] Bergstrom (1983, 1984) labels the solution of the SDE at observation times the "exact discrete model." See McCrorie (2009) and Chambers et al. (2018) for a comprehensive review on the exact discrete model. Representations of higher-order systems of stationary and non-stationary SDEs are discussed in Chambers (1999).

The ED-SSR (2.4)-(2.5) can be compactly written in the ABCD state space form

$$\mathbf{x}_{\tau+1} = \mathsf{A}(\boldsymbol{\theta}; h)\mathbf{x}_\tau + \mathsf{B}(\boldsymbol{\theta}; h)\boldsymbol{\epsilon}_{\tau+1}\,, \tag{2.12}$$

$$\mathbf{y}_{\tau+1} = \mathsf{C}(\boldsymbol{\theta}; h)\mathbf{x}_\tau + \mathsf{D}(\boldsymbol{\theta}; h)\boldsymbol{\epsilon}_{\tau+1}\,, \tag{2.13}$$

for conformable matrices $\mathsf{A}(\boldsymbol{\theta}; h)$, $\mathsf{B}(\boldsymbol{\theta}; h)$, $\mathsf{C}(\boldsymbol{\theta}; h)$, $\mathsf{D}(\boldsymbol{\theta}; h)$ (see Online Appendix D), with the n_ϵ-vector $\boldsymbol{\epsilon}_\tau = \left[\boldsymbol{\eta}_\tau^\top, \boldsymbol{\upsilon}_\tau^\top\right]^\top$ accommodating serially uncorrelated Gaussian errors, $\boldsymbol{\upsilon}_\tau$, to account for any potential sampling or measurement error, $\boldsymbol{\upsilon}_\tau \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta})\right)$, $\mathbb{E}\left[\boldsymbol{\eta}_\tau \boldsymbol{\upsilon}_{\tau-\ell}^\top\right] = \mathbf{0}$, for all τ and ℓ.[8] Thus, $\boldsymbol{\epsilon}_\tau$ is white noise, $\mathbb{E}[\boldsymbol{\epsilon}_\tau] = \mathbf{0}$, with $\mathbb{E}[\boldsymbol{\epsilon}_\tau \boldsymbol{\epsilon}_{\tau-\ell}^\top] = \boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}; h) \cdot \mathbb{I}\{\ell = 0\}$, $\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}; h) = \mathrm{diag}(\boldsymbol{\Sigma}_{\boldsymbol{\eta},h}(\boldsymbol{\theta}), \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}))$, and n_ϵ depending on the features of the system.

Lemma 1. *Under Assumption 1, for given $h > 0$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and $z \in \mathbb{C}$, $\det(\mathbf{I} - \mathsf{A}(\boldsymbol{\theta}; h)z) = 0$ implies $|z| > 1$.*

By the lemma, if the continuous-time system is stable (Assumption 1), then so is the discrete-time ABCD system (2.12)-(2.13), regardless the measurements being stocks or flows.

Remark 2.1. The state space representation (2.4)-(2.5) is not minimal,[9] but can be recast in minimal form using dimensionality reduction techniques (see Ahn et al., 2018), or by rearranging the state variables. In Online Appendix D, we obtain the minimal representation by substituting out the flow state variables from the transition equation, so $n_x = n_x^s = m_x$, and we provide the matrices $\mathsf{A}(\boldsymbol{\theta}; h)$, $\mathsf{B}(\boldsymbol{\theta}; h)$, $\mathsf{C}(\boldsymbol{\theta}; h)$, $\mathsf{D}(\boldsymbol{\theta}; h)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}; h)$ in (2.12)-(2.13) corresponding to both the minimal and non-minimal ABCD representations. ■

2.3 Likelihood function

Let $\mathbf{y}^T = \{\mathbf{y}_\tau^s, \mathbf{y}_\tau^f\}_{\tau=0}^T$ be a sample of equidistant measurements of stock and flow variables.[10] Based on $\mathbf{y}^T$, we consider maximum likelihood (ML) estimation of the unknown parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ of the continuous-time model (2.1)-(2.2) using the ABCD representation (2.12)-(2.13) of the ED-SSR. With the state variables unobserved, the *exact* likelihood function is constructed using the Kalman filter. To ensure existence of the Kalman gain and convergence of the Kalman filter recursions, we adopt the following additional assumption.

Assumption 2 (Nonsingularity)**.** For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\mathsf{D}(\boldsymbol{\theta}; h)\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}; h)\mathsf{D}(\boldsymbol{\theta}; h)^\top$ is nonsingular.

[8] In practice, for given $h > 0$, $\boldsymbol{\theta}$ will be comprised of the frequency-invariant parameters of (2.1)-(2.2) and the measurement error covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta})$, so the latter represents an expansion of $\boldsymbol{\theta}$. The empirical model can be augmented to accommodate VAR(1) measurement errors, following Ireland (2004).

[9] A minimal representation is one in which the dimension n_x of the state equation is the smallest possible.

[10] Equidistant measurements are chosen for simplicity of exposition. Some variation in spacing between observations is easily accommodated.

Assumption 2 rules out stochastic singularity of the ABCD representation. This is equivalent to assuming existence of a unique solution to the discrete algebraic Riccati equation of the filter. Singularity arises in models with more observables than shocks, $n_y > n_\epsilon$.

The conditional log-likelihood function of the data, given $\mathbf{y}_0$, is constructed recursively via the prediction error decomposition as

$$\mathcal{L}\left(\boldsymbol{\theta}|\mathbf{y}^T, h\right) = \log f\left(\mathbf{y}^T|\boldsymbol{\theta}, h\right) = \sum_{\tau=1}^{T} \log f\left(\mathbf{y}_\tau|\mathbf{y}^{\tau-1}; \boldsymbol{\theta}, h\right) , \tag{2.14}$$

where $f\left(\mathbf{y}_\tau|\mathbf{y}^{\tau-1}; \boldsymbol{\theta}, h\right)$ is the conditional density of the time t_τ measurements $\mathbf{y}_\tau$, given the information $\mathbf{y}^{\tau-1} = \{\mathbf{y}_0, \ldots, \mathbf{y}_{\tau-1}\}$ at $t_{\tau-1}$. By the linear structure of the Gaussian state space model, the conditional density is multivariate normal, with first- and second-order moments determined by the one-step-ahead predictions of the measurements and the associated prediction error covariance matrix. Under regularity conditions, the MLE $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{L}\left(\boldsymbol{\theta}|\mathbf{y}^T, h\right)$ delivers consistent, asymptotically normal and efficient estimates of $\boldsymbol{\theta}$. Given $\hat{\boldsymbol{\theta}}$, the information content in the full sample is used to predict the latent states $\{\mathbf{x}_\tau\}_{\tau=1}^T$ and residuals (smoothed reduced-form disturbances) $\{\boldsymbol{\eta}_\tau\}_{\tau=1}^T$.[11]

For nonsingular systems, $n_y \leq n_\epsilon$, and under Assumptions 1 and 2, the ABCD representation in (2.12)-(2.13) admits the innovations representation (Anderson and Moore, 1979)

$$\mathbf{x}_{\tau+1|\tau+1} = \mathsf{A}(\boldsymbol{\theta}; h)\mathbf{x}_{\tau|\tau} + \mathsf{K}(\boldsymbol{\theta}; h)\boldsymbol{\nu}_{\tau+1|\tau} , \tag{2.15}$$

$$\mathbf{y}_{\tau+1} = \mathsf{C}(\boldsymbol{\theta}; h)\mathbf{x}_{\tau|\tau} + \boldsymbol{\nu}_{\tau+1|\tau} , \tag{2.16}$$

$\tau = 0, \ldots, T-1$, with $\mathsf{K}(\boldsymbol{\theta}; h)$ the Kalman gain, $\mathbf{x}_{\tau|\tau}$ the contemporaneous prediction of the state vector, and $\boldsymbol{\nu}_{\tau+1|\tau}$ the one-step-ahead prediction error for the measurements, each conditional on $\mathbf{y}^\tau$, and with prediction error covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$. Thus, the conditional moments in (2.14) are $\mathsf{C}(\boldsymbol{\theta}; h)\mathbf{x}_{\tau|\tau}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$. The innovations representation, with time-invariant $\mathsf{K}(\boldsymbol{\theta}; h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$, is fundamental for establishing the conditions for local identification of the model parameters $\boldsymbol{\theta}$.[12]

We collect the matrices of the system (2.15)-(2.16) in the array

$$\boldsymbol{\Lambda}\left(\boldsymbol{\theta}\right) = \left(\mathsf{A}(\boldsymbol{\theta}; h), \mathsf{K}(\boldsymbol{\theta}; h), \mathsf{C}(\boldsymbol{\theta}; h), \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)\right) , \tag{2.17}$$

with $\mathsf{K}(\boldsymbol{\theta}; h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$ functions of the matrices $\mathsf{A}(\boldsymbol{\theta}; h)$, $\mathsf{B}(\boldsymbol{\theta}; h)$, $\mathsf{C}(\boldsymbol{\theta}; h)$, $\mathsf{D}(\boldsymbol{\theta}; h)$, and

[11] See the Online Appendix E for details on the Kalman filter, the likelihood function, and the state- and disturbance-smoother.

[12] We use the time-varying prediction error covariance and Kalman gain—Equations (E.4) and (E.7), respectively, in Online Appendix E—for ML estimation purposes.

$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta};h)$ of the ABCD representation (2.12)-(2.13), obtained through the Kalman recursions. For given $h>0$, the latter matrices are, in turn, nonlinear functions (given in Online Appendix D) of the matrices $\boldsymbol{\Lambda}_c(\boldsymbol{\theta})=(\mathbf{A}(\boldsymbol{\theta}),\mathbf{B}(\boldsymbol{\theta}),\mathbf{C}(\boldsymbol{\theta}))$ from (2.3) (and $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta})$, in case of measurement error) of the continuous-time model (2.1)-(2.2). Thus, the mapping $\boldsymbol{\theta}\mapsto\boldsymbol{\Lambda}(\boldsymbol{\theta})$ is continuously differentiable with respect to $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ under the following assumption.

Assumption 3 (Differentiability)**.** The mapping $\boldsymbol{\theta}\mapsto(\boldsymbol{\Lambda}_c(\boldsymbol{\theta}),\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}))$ is continuously differentiable with respect to $\boldsymbol{\theta}\in\boldsymbol{\Theta}$.

The identification analysis requires focusing on the minimal representation of the system. This leads to the following assumption (Anderson and Moore, 1979).

Assumption 4 (Minimality)**.** The innovations representation (2.15)-(2.16) is minimal, i.e., the reachability matrix $\mathcal{R}(\boldsymbol{\theta};h)=[\mathsf{K}(\boldsymbol{\theta};h),\mathsf{A}(\boldsymbol{\theta};h)\mathsf{K}(\boldsymbol{\theta};h),...,\mathsf{A}^{n_x-1}(\boldsymbol{\theta};h)\mathsf{K}(\boldsymbol{\theta};h)]$ and the observability matrix $\mathcal{O}(\boldsymbol{\theta};h)=[\mathsf{C}(\boldsymbol{\theta};h)^\top,\mathsf{A}(\boldsymbol{\theta};h)^\top\mathsf{C}(\boldsymbol{\theta};h)^\top,...,\mathsf{A}^{n_x-1}(\boldsymbol{\theta};h)^\top\mathsf{C}(\boldsymbol{\theta};h)^\top]^\top$ are of full row rank and full column rank, respectively.

To simplify proofs, we work with the minimal ABCD representation, so $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ is a function of the matrices from this, and Assumption 4 is automatic.

3. Local identification of structural parameters

3.1 Observational equivalence

The consistency of the MLE relies on the ability to identify the true parameter vector, $\boldsymbol{\theta}_0$, given the distribution of the data. Following Rothenberg (1971), we characterize identification in terms of observational equivalence. Thus, $\boldsymbol{\theta}_0$ is (locally) identifiable if there is no other $\boldsymbol{\theta}_\ell\in\boldsymbol{\Theta}$ (in a neighborhood of $\boldsymbol{\theta}_0$) for which $f\left(\mathbf{y}^T|\boldsymbol{\theta}_\ell,h\right)=f\left(\mathbf{y}^T|\boldsymbol{\theta}_0,h\right)$, for all $\mathbf{y}^T$, and all $T\geq 1$. Given our stationary linear Gaussian model structure, the probability distribution is characterized by the first two moments. Therefore, following Komunjer and Ng (2011), who focus on recovering the parameters of a discrete-time stationary DSGE model, the identification analysis is based on the autocovariances of the discrete-time observations.[13]

The autocovarianes are determined by $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ from (2.17).[14] Since $f\left(\mathbf{y}^T|\boldsymbol{\theta},h\right)$ depends on $\boldsymbol{\theta}$ through $\boldsymbol{\Lambda}(\boldsymbol{\theta})$, local identifiability of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$ requires that no other $\boldsymbol{\theta}_\ell$ in a neighborhood of

[13] We always work with demeaned state space representations, so there is no information in the unconditional means. There is information in the conditional mean variation, as tracked by the state space model (2.15)-(2.16), but since this is around the (zero) unconditional means, it is determined by the autocovariances.

[14] The autocovariances can be computed from the transfer function associated with the VMA(∞) representation, which depends on the ABCD and residual covariance parameters or, equivalently, $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ from (2.17), and determines the spectral density, so observational equivalence can be analyzed in terms of $\boldsymbol{\Lambda}(\boldsymbol{\theta})$.

$\boldsymbol{\theta}_0$ generates a value $\boldsymbol{\Lambda}(\boldsymbol{\theta}_\ell)$ that yields the same density (likelihood) as $\boldsymbol{\Lambda}(\boldsymbol{\theta}_0)$. Heuristically, two things can go wrong. First, if $\boldsymbol{\Lambda}(\boldsymbol{\theta}_\ell) = \boldsymbol{\Lambda}(\boldsymbol{\theta}_0)$ for $\boldsymbol{\theta}_\ell \neq \boldsymbol{\theta}_0$, then obviously $\boldsymbol{\theta}$ is not identifiable in a neighborhood around $\boldsymbol{\theta}_0$ containing $\boldsymbol{\theta}_\ell$. Second, the same is true even for $\boldsymbol{\Lambda}(\boldsymbol{\theta}_\ell) \neq \boldsymbol{\Lambda}(\boldsymbol{\theta}_0)$ if the two arrays have the same implications for the observables $\mathbf{y}_{\tau+1}$ in (2.15)-(2.16). Essentially, local identifiability is equivalent to the conditions that $\boldsymbol{\theta}$ can be backed out from $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ in a neighborhood of $\boldsymbol{\theta}_0$, and that $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ can be backed out from the data in the corresponding neighborhood of $\boldsymbol{\Lambda}(\boldsymbol{\theta}_0)$. The latter is impossible if the system (2.15)-(2.16) at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ simply is the system at $\boldsymbol{\theta}_\ell$ in the neighborhood upon pre-multiplication of the state equation by an invertible matrix (similarity transform) $\mathbf{T}_h$, thereby leaving the model for $\mathbf{y}_{\tau+1}$ unchanged. Thus, a similarity transform is a full rank $n_x \times n_x$ matrix $\mathbf{T}_h$ satisfying $\mathbf{T}_h \mathsf{A}(\boldsymbol{\theta}_\ell; h)\mathbf{T}_h^{-1} = \mathsf{A}(\boldsymbol{\theta}_0; h)$, $\mathbf{T}_h \mathsf{K}(\boldsymbol{\theta}_\ell; h) = \mathsf{K}(\boldsymbol{\theta}_0; h)$, $\mathsf{C}(\boldsymbol{\theta}_\ell; h)\mathbf{T}_h^{-1} = \mathsf{C}(\boldsymbol{\theta}_0; h)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h) = \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_0; h)$ (the latter for equivalent systems). Formally, Komunjer and Ng (2011) consider the continuously differentiable mapping

$$\boldsymbol{\delta}(\boldsymbol{\theta}, \mathbf{T}_h) = \left[\mathrm{vec}(\mathbf{T}_h \mathsf{A}(\boldsymbol{\theta}; h)\mathbf{T}_h^{-1})^\top, \mathrm{vec}(\mathbf{T}_h \mathsf{K}(\boldsymbol{\theta}; h))^\top, \mathrm{vec}(\mathsf{C}(\boldsymbol{\theta}; h)\mathbf{T}_h^{-1})^\top, \mathrm{vech}(\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h))^\top\right]^\top, \quad (3.1)$$

where vec and vech are the vectorization and half-vectorization operators, and show that under Assumptions 1, 2, and 4, $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ are observationally equivalent at frequency h if and only if there exists a similarity transform $\mathbf{T}_h$ such that

$$\boldsymbol{\delta}(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{I}_{n_x}). \quad (3.2)$$

In this case, $\boldsymbol{\theta}_\ell$ and $\boldsymbol{\theta}_0$ generate identical autocovariances at all leads and lags, and hence spectral densities.[15] Conversely, there is no observationally equivalent $\boldsymbol{\theta}_\ell$ in a neighborhood of $\boldsymbol{\theta}_0$ if and only if (3.2), considered as a system of $n_\delta = n_x^2 + 2n_x n_y + n_y(n_y+1)/2$ equations in the $m_\theta + n_x^2$ unknowns $(\boldsymbol{\theta}_\ell, \mathbf{T}_h)$, has a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ (note that any $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is observationally equivalent with itself). In this case, and under Assumptions 1, 2, and 4, $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$, as shown by Komunjer and Ng (2011). Local uniqueness of $\mathbf{T}_h = \mathbf{I}_{n_x}$ secures local identifiability of $\boldsymbol{\Lambda}(\boldsymbol{\theta}_0)$ from the data, and local uniqueness of $\boldsymbol{\theta}_\ell = \boldsymbol{\theta}_0$ secures local identifiability of $\boldsymbol{\theta}_0$ from $\boldsymbol{\Lambda}(\boldsymbol{\theta}_0)$.

In our framework, given the underlying continuous-time DSGE model, the functional dependence on $\boldsymbol{\theta}$ in $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ is different than that in Komunjer and Ng (2011). In their case, $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ reflects the rational expectations solution of the given discrete-time model, and the parameter they identify is specific to the given observation frequency, h. In contrast, in

[15] Komunjer and Ng (2011) write the condition as $\boldsymbol{\delta}(\boldsymbol{\theta}_\ell, \mathbf{I}_{n_x}) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{T}_h)$ when discussing observational equivalence, and use the format (3.2) when discussing local identification. As the domain for $\mathbf{T}_h$ is the space of full rank $n_x \times n_x$ matrices, the two ways of writing the condition are equivalent. We retain the format (3.2) throughout. Further, h is fixed and hence suppressed in Komunjer and Ng (2011).

the continuous-time case, the *non-linear cross-equation restrictions* imposed by the rational expectations solution are reflected in the matrices $\mathbf{\Lambda}_c(\boldsymbol{\theta}) = (\mathbf{A}(\boldsymbol{\theta}), \mathbf{B}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$ from (2.3), whereas the matrices in $\mathbf{\Lambda}(\boldsymbol{\theta})$ in (2.17) in addition reflect the nonlinear transformation arising from the discretization and involving the matrix exponential (2.6), cf. Proposition 1. This is represented through the mappings

$$\boldsymbol{\theta} \mapsto (\mathbf{\Lambda}_c(\boldsymbol{\theta}), \mathbf{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta})) \mapsto \mathbf{\Lambda}(\boldsymbol{\theta}) \,, \tag{3.3}$$

where the left map relates the structural parameters to the reduced-form parameters implied by the rational expectations solution of the continuous-time model, and the right map relates these to the reduced-form parameters of the exact discrete-time innovations representation (2.15)-(2.16), i.e., $\mathsf{A}(\boldsymbol{\theta}; h)$, $\mathsf{K}(\boldsymbol{\theta}; h)$, $\mathsf{C}(\boldsymbol{\theta}; h)$, and $\mathbf{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$.

To study local identification in continuous-time linear DSGE models, we extend the methodology developed in Komunjer and Ng (2011) for discrete-time linear DSGE models by including the additional identifying information from the rational expectations solution of the continuous-time model, available through the matrices $\mathbf{\Lambda}_c(\boldsymbol{\theta})$.[16] In particular, we show that in case of observational equivalence, so that the discrete-time matrices in $\mathbf{\Lambda}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ are related through the similarity transform as in (3.2), parts or all of the continuous-time matrices in $\mathbf{\Lambda}_c(\boldsymbol{\theta})$ are similarly related at $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$. This implies an increase in the information content in the analogue of (3.1)-(3.2), with more equations available for solving for the same number of unknowns, hence potentially improving identifiability in the continuous-time case, relative to the discrete-time case, and in some cases allowing a recasting of (3.1)-(3.2) entirely in terms of the continuous-time matrices.

3.2 Aliasing

Whether the observational equivalence conditions (3.1)-(3.2) can be recast in terms of the continuous-time matrices depends on whether the model is subject to the *aliasing* problem, which is specific to multivariate continuous-time models, $n_x \geq 2$. Aliasing refers to the inability to distinguish proportional oscillatory behaviors generated by different SDEs based on a sample of discrete-time observations. As discussed by Phillips (1973), aliasing at $\boldsymbol{\theta}_0$ arises in case of non-injectivity of the matrix exponential (2.6) at $\mathbf{A}(\boldsymbol{\theta}_0)h$. Specifically, an

[16] Rothenberg (1971) considers additional information in the form of parameter restrictions, such as mean restrictions, and Komunjer and Ng (2011) consider means, too. Alternative conditions for local identification of discrete-time DSGE models have been studied in Iskrev (2010) and Qu and Tkachenko (2012), while Qu and Tkachenko (2017) and Kociecki and Kolasa (2018, 2023) study global identification.

alias at $\boldsymbol{\theta}_0$ is an $n_x \times n_x$ matrix $\mathbf{A}_0 \neq \mathbf{A}(\boldsymbol{\theta}_0)$ solving the system

$$\exp(\mathbf{A}_0 h) = \exp(\mathbf{A}(\boldsymbol{\theta}_0)h)\,. \tag{3.4}$$

An assumption of no (local) aliases means that the equation (3.4) in $\mathbf{A}_0$ has a (locally) unique solution at $\mathbf{A}_0 = \mathbf{A}(\boldsymbol{\theta}_0)$. In our case, as the representation is minimal, the discrete-time transition matrix in (2.12) and (2.15) is given by (cf. Equation D.3 in Online Appendix)

$$\mathsf{A}(\boldsymbol{\theta}; h) = \mathbf{A}_h(\boldsymbol{\theta}) = \exp(\mathbf{A}(\boldsymbol{\theta})h)\,. \tag{3.5}$$

Thus, in the presence of an alias $\mathbf{A}_0$, and if there exists $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}$ such that $\mathbf{A}(\boldsymbol{\theta}_1) = \mathbf{A}_0$, then

$$\mathsf{A}(\boldsymbol{\theta}_1; h) = \exp(\mathbf{A}(\boldsymbol{\theta}_1)h) = \exp(\mathbf{A}_0 h) = \exp(\mathbf{A}(\boldsymbol{\theta}_0)h) = \mathsf{A}(\boldsymbol{\theta}_0; h)\,, \tag{3.6}$$

where the first and last equalities follow from (3.5), the second from the assumption that $\mathbf{A}(\boldsymbol{\theta}_1) = \mathbf{A}_0$, and the third from aliasing (3.4). It follows from (3.6) that $\boldsymbol{\theta}_0$ is not identifiable from $\mathsf{A}(\boldsymbol{\theta}_0; h)$, as obviously $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$, because $\mathbf{A}(\boldsymbol{\theta}_1) = \mathbf{A}_0 \neq \mathbf{A}(\boldsymbol{\theta}_0)$.

Aliasing implies that the $\mathsf{A}(\boldsymbol{\theta}; h)$ portion of the observational equivalence condition (3.2) is satisfied for $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_1, \mathbf{I}_{n_x})$. If such solution, with $\boldsymbol{\theta}_\ell \neq \boldsymbol{\theta}_0$, extends to all of (3.2) (for $\boldsymbol{\theta}_\ell$ local to $\boldsymbol{\theta}_0$, i.e., in an arbitrarily small neighborhood), then (local) uniqueness of the solution $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ is violated, and (local) identifiability jeopardized.

If $\partial \mathrm{vec}(\mathsf{A}(\boldsymbol{\theta}; h))/\partial \boldsymbol{\theta}$ has rank m_θ at $\boldsymbol{\theta}_0$, then the mapping $\boldsymbol{\theta} \mapsto \mathsf{A}(\boldsymbol{\theta}; h)$ is locally injective at $\boldsymbol{\theta}_0$, and local aliases are ruled out. If, further, $\mathsf{A}(\boldsymbol{\theta}_0; h)$ is locally identifiable (e.g., if $\mathbf{T}_h = \mathbf{I}_{n_x}$ is the only local similarity transform), then so is $\boldsymbol{\theta}_0$. Phillips (1973) offers the following assumption, under which the continuous-time model does not generate oscillations with periods shorter than $2h$, which in turn precludes local aliasing,[17] although this is not necessary for our main results.

Assumption 5 (Oscillations)**.** The eigenvalues of $\mathbf{A}(\boldsymbol{\theta}_0)$ are distinct, and do not differ by an integer multiple of $2\pi i/h$.

Importantly, (local) aliasing does not rule out (local) identifiability of $\boldsymbol{\theta}_0$, as this in part may rely on other portions of the model, beside $\mathsf{A}(\boldsymbol{\theta}_0; h)$. Our focus is on the local identifiability conditions that apply in the presence of the underlying continuous-time model. We do not impose Assumption 5, but simply report separate local identification conditions, according to whether or not local aliasing is ruled out.

[17] This is shown in Online Appendix F.

3.3 Observational equivalence with underlying continuous-time model

Recall from (2.5) that $\mathbf{S}_y^s$ and $\mathbf{S}_y^f$ select the rows of $\mathbf{y}_\tau$ corresponding to stock and flow measurements, respectively. Write the continuous-time reaction matrix from (2.2) as

$$\mathbf{C}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}) \\ \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \mathbf{C}^s(\boldsymbol{\theta}) \\ \mathbf{C}^f(\boldsymbol{\theta}) \end{bmatrix}, \tag{3.7}$$

similarly partitioned according to stocks and flows.

Lemma 2. *Under Assumptions 1, 2, and 4,*

(a) if there exists a full rank $n_x \times n_x$ matrix $\mathbf{T}_h$ such that

$$\begin{aligned} &\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} = \mathbf{A}(\boldsymbol{\theta}_0)\,,\ \mathbf{C}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} = \mathbf{C}(\boldsymbol{\theta}_0) \\ &\mathbf{T}_h \mathsf{K}(\boldsymbol{\theta}_\ell; h) = \mathsf{K}(\boldsymbol{\theta}_0; h)\,,\ \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h) = \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_0; h)\,, \end{aligned} \tag{3.8}$$

then $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ are observationally equivalent, i.e., (3.2) applies;

(b) if $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_\ell) = \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_0)$, and there exists a full rank $n_x \times n_x$ matrix $\mathbf{T}$, and an orthogonal $m_w \times m_w$ matrix $\mathbf{U}$, such that

$$\mathbf{T}\mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}^{-1} = \mathbf{A}(\boldsymbol{\theta}_0)\,,\ \mathbf{T}\mathbf{B}(\boldsymbol{\theta}_\ell)\mathbf{U} = \mathbf{B}(\boldsymbol{\theta}_0)\,,\ \mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{T}^{-1} = \mathbf{C}(\boldsymbol{\theta}_0)\,, \tag{3.9}$$

then $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ are observationally equivalent, i.e., (3.2) applies, with $\mathbf{T}_h = \mathbf{T}$;

(c) if $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ are observationally equivalent, i.e., (3.2) applies, for some full rank $n_x \times n_x$ matrix $\mathbf{T}_h$, then

$$\mathbf{C}^s(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1} = \mathbf{C}^s(\boldsymbol{\theta}_0) \tag{3.10}$$

holds exactly;

(d) in (c), all conditions in (3.8) hold exactly in the absence of aliases, and the conditions on $\mathbf{A}(\cdot)$ and $\mathbf{C}^f(\cdot)$ in (3.8) hold to order of approximation $O(h)$ in general ((3.10) and the conditions on $\mathsf{K}(\cdot)$, $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\cdot)$ in (3.8) are still exact);

(e) in (c), if, in addition, $\mathbf{C}^s(\boldsymbol{\theta}_0)$ has full column rank, and $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_\ell) = \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_0)$, then there exists an orthogonal $m_w \times m_w$ matrix $\mathbf{U}_h$ such that all conditions in (3.9) hold exactly in the absence of aliases, with $\mathbf{T} = \mathbf{T}_h$, $\mathbf{U} = \mathbf{U}_h$, and to order of approximation $O(h)$ in general (condition (3.10) is still exact).

Lemma 2 shows the implications of the underlying continuous-time model for the observational equivalence condition (3.1)-(3.2). By (3.8), the portion of the condition relating to the transition and loading matrices $\mathsf{A}(\cdot\,;h)$, $\mathsf{C}(\cdot\,;h)$ of the discrete-time representation (2.15)-(2.16) is alternatively stated directly in terms of $\mathbf{A}(\cdot)$, $\mathbf{C}(\cdot)$ from the continuous-time system (2.1)-(2.2). By (3.9), under an additional assumption on the measurement error covariance $\mathbf{\Sigma}_{\boldsymbol{\upsilon},h}$ (e.g., if this zero, or known), the observational equivalence condition is recast purely in terms of the continuous-time matrices $\mathbf{\Lambda}_c(\cdot)$ from (2.3), including $\mathbf{B}(\cdot)$.

Conversely, regardless of possible aliasing, the Lemma provides additional exact implications of observational equivalence, beyond those in (3.2) on the discrete-time matrices $\mathbf{\Lambda}(\cdot)$, namely, the $n_y^s n_x$ conditions (3.10) on the portion of the continuous-time reaction matrix $\mathbf{C}(\cdot)$ corresponding to stock measurements. These implications reflect the rational expectations solution of the continuous-time model, and are therefore not present in Komunjer and Ng (2011). Further, in the absence of aliases, (3.8) provides necessary and sufficient conditions for observational equivalence. These are recast as (3.9), without reference to the innovations representation matrices $\mathbf{\Lambda}(\cdot)$, under additional conditions. Full column rank of $\mathbf{C}^s(\boldsymbol{\theta}_0)$ requires at least as many stock measurements as state variables. Finally, regardless of aliasing, observational equivalence implies that (3.8) holds to order $O(h)$, as determined by the observation frequency, and so does (3.9) under the additional conditions.

Remark 3.1. Condition (3.9) resembles a portion of the observational equivalence condition for singular systems ($n_y \geq n_\epsilon$) in Komunjer and Ng (2011).The main differences are that (3.9) involves $\mathbf{\Lambda}_c = (\mathbf{A}, \mathbf{B}, \mathbf{C})$ from the continuous-time model, and $\mathbf{\Sigma}_{\boldsymbol{\upsilon},h}$, whereas the Komunjer and Ng (2011) condition involves the discrete-time matrices, $(\mathsf{A}, \mathsf{B}, \mathsf{C})$, as well as $(\mathsf{D}, \mathbf{\Sigma}_{\boldsymbol{\epsilon}})$ from the ABCD representation, and that we consider the non-singular case ($n_y \leq n_\epsilon$). Essentially, our proof ties the conditions on $\mathbf{\Lambda}_c$ to the relevant discrete-time matrices, which are $(\mathsf{A}, \mathsf{K}, \mathsf{C}, \mathbf{\Sigma}_{\boldsymbol{\nu}})$ from (2.15)-(2.16) in our case, as the non-singularity requires filtering. ■

3.4 Local identification with underlying continuous-time model

Based on (3.10), we extend $\mathbf{\Lambda}(\boldsymbol{\theta})$ from (2.17) to $\mathbf{\Lambda}^+(\boldsymbol{\theta}) = (\mathbf{C}^s(\boldsymbol{\theta}), \mathbf{\Lambda}(\boldsymbol{\theta}))$. Correspondingly, we augment $\boldsymbol{\delta}(\boldsymbol{\theta}, \mathbf{T}_h)$ from (3.1) to the continuously differentiable mapping

$$\boldsymbol{\delta}^+(\boldsymbol{\theta}, \mathbf{T}_h) = \left[\operatorname{vec}\left(\mathbf{C}^s(\boldsymbol{\theta})\mathbf{T}_h^{-1}\right)^\top, \boldsymbol{\delta}(\boldsymbol{\theta}, \mathbf{T}_h)^\top\right]^\top, \tag{3.11}$$

used to formulate the $n_{\delta^+} = n_y^s n_x + n_\delta$ equations (local identification conditions)

$$\boldsymbol{\delta}^+(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = \boldsymbol{\delta}^+(\boldsymbol{\theta}_0, \mathbf{I}_{n_x}) \tag{3.12}$$

in the $m_\theta + n_x^2$ unknowns $(\boldsymbol{\theta}_\ell, \mathbf{T}_h)$ in the next lemma, hence extending (3.2). Here, based on (3.3) and (3.9), we also consider shifting the analysis from $\boldsymbol{\Lambda}$ to the more primitive $\boldsymbol{\Lambda}_c = (\mathbf{A}, \mathbf{B}, \mathbf{C})$ from (2.3), based directly on the continuous-time matrices, and the corresponding

$$\boldsymbol{\delta}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h) = \left[\text{vec}\left(\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}) \mathbf{T}_h^{-1}\right)^\top, \text{vec}\left(\mathbf{T}_h \mathbf{B}(\boldsymbol{\theta}) \mathbf{U}_h\right)^\top, \text{vec}\left(\mathbf{C}(\boldsymbol{\theta}) \mathbf{T}_h^{-1}\right)^\top\right]^\top, \tag{3.13}$$

used to formulate the $n_c = n_x^2 + n_x m_w + n_y n_x$ equations

$$\boldsymbol{\delta}_c(\boldsymbol{\theta}_\ell, \mathbf{T}_h, \mathbf{U}_h) = \boldsymbol{\delta}_c(\boldsymbol{\theta}_0, \mathbf{I}_{n_x}, \mathbf{I}_{m_w}) \tag{3.14}$$

in the $m_\theta + n_x^2 + m_w^2$ unknowns $(\boldsymbol{\theta}_\ell, \mathbf{T}_h, \mathbf{U}_h)$. The intermediate case, based on (3.8), involves $\boldsymbol{\Lambda}_a(\boldsymbol{\theta}) = (\mathbf{A}, \mathbf{C}, \mathsf{K}, \boldsymbol{\Sigma}_{\boldsymbol{\nu}})$, and

$$\boldsymbol{\delta}_a(\boldsymbol{\theta}, \mathbf{T}_h) = \left[\text{vec}(\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}; h) \mathbf{T}_h^{-1})^\top, \text{vec}(\mathbf{C}(\boldsymbol{\theta}; h) \mathbf{T}_h^{-1})^\top, \text{vec}(\mathbf{T}_h \mathsf{K}(\boldsymbol{\theta}; h))^\top, \text{vech}(\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h))^\top\right]^\top \tag{3.15}$$

is used to formulate the n_δ equations

$$\boldsymbol{\delta}_a(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = \boldsymbol{\delta}_a(\boldsymbol{\theta}_0, \mathbf{I}_{n_x}) \tag{3.16}$$

in the $m_\theta + n_x^2$ unknowns $(\boldsymbol{\theta}_\ell, \mathbf{T}_h)$. Thus, whereas the original $\boldsymbol{\delta}$ only involves the discrete-time matrices $\boldsymbol{\Lambda}$ for the non-singular case, and the similarity transform $\mathbf{T}_h$, as in Komunjer and Ng (2011), the mapping $\boldsymbol{\delta}_h^+$ in addition involves a portion of the continuous-time reaction matrix $\mathbf{C}$. The mapping $\boldsymbol{\delta}_a$ replaces A, C in $\boldsymbol{\delta}$ with $\mathbf{A}$, $\mathbf{C}$, and $\boldsymbol{\delta}_c$ further replaces K, $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}$ with $\mathbf{B}$, while involving both $\mathbf{T}_h$ and $\mathbf{U}_h$. The rotation $\mathbf{U}_h$ only appears in the singular case in Komunjer and Ng (2011), but arises in our non-singular case, as we link $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}_c$ (see also Remark 3.1), and the pure continuous-time model (without measurement error) is singular.

Lemma 3 (Local identifiability)**.** *Under Assumptions 1, 2, and 4, for given $h > 0$,*

(a) $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ if and only if the system (3.12) has a locally unique solution $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$;

(b) in the absence of local aliases (e.g., under the additional Assumption 5), $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ if and only if the system (3.16) has a locally unique solution $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$;

(c) in (b), if, in addition, $\mathbf{C}^s(\boldsymbol{\theta}_0)$ is of full column rank, and $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta})$ is constant in a neighborhood of $\boldsymbol{\theta}_0$, then $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ if and only if the system (3.14) has a locally unique solution $(\boldsymbol{\theta}_\ell, \mathbf{T}_h, \mathbf{U}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x}, \mathbf{I}_{m_w})$.

From Lemma 3.(a), regardless of possible aliasing, the parameters of the continuous-time DSGE model are locally identifiable if and only if the augmented mapping $\boldsymbol{\delta}^{+}(\boldsymbol{\theta},\mathbf{T}_h)$ from (3.11) is locally injective at $(\boldsymbol{\theta}_0,\mathbf{I}_{n_x})$. By the implicit function theorem (IFT), a sufficient condition for this is full column rank of the Jacobian $\mathcal{J}^{+}(\boldsymbol{\theta},\mathbf{T}_h)=\left(\frac{\partial\boldsymbol{\delta}^{+}(\boldsymbol{\theta},\mathbf{T}_h)}{\partial\boldsymbol{\theta}},\frac{\partial\boldsymbol{\delta}^{+}(\boldsymbol{\theta},\mathbf{T}_h)}{\partial\mathrm{vec}\mathbf{T}_h}\right)$ at $(\boldsymbol{\theta}_0,\mathbf{I}_{n_x})$. In the discrete-time case, the Jacobian considered by Komunjer and Ng (2011) corresponds to the submatrix $\mathcal{J}(\boldsymbol{\theta},\mathbf{T}_h)$ of $\mathcal{J}^{+}(\boldsymbol{\theta},\mathbf{T}_h)$ associated with $\boldsymbol{\delta}(\boldsymbol{\theta},\mathbf{T}_h)$ from (3.1), with rank and order conditions for local identification based on $\boldsymbol{\Delta}(\boldsymbol{\theta})=\mathcal{J}(\boldsymbol{\theta},\mathbf{I}_{n_x})$ at $\boldsymbol{\theta}_0$,

$$\boldsymbol{\Delta}(\boldsymbol{\theta}_0)=\left(\begin{array}{cc}\frac{\partial\mathrm{vec}(\mathsf{A}(\boldsymbol{\theta};h))}{\partial\boldsymbol{\theta}} & \mathsf{A}(\boldsymbol{\theta};h)^{\top}\otimes\mathbf{I}_{n_x}-\mathbf{I}_{n_x}\otimes\mathsf{A}(\boldsymbol{\theta};h)\\ \frac{\partial\mathrm{vec}(\mathsf{K}(\boldsymbol{\theta};h))}{\partial\boldsymbol{\theta}} & \mathsf{K}(\boldsymbol{\theta},h)^{\top}\otimes\mathbf{I}_{n_x}\\ \frac{\partial\mathrm{vec}(\mathsf{C}(\boldsymbol{\theta};h))}{\partial\boldsymbol{\theta}} & -\mathbf{I}_{n_x}\otimes\mathsf{C}(\boldsymbol{\theta};h)\\ \frac{\partial\mathrm{vech}(\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h))}{\partial\boldsymbol{\theta}} & \mathbf{0}_{n_y(n_y+1)/2\times n_x^2}\end{array}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{3.17}$$

With the underlying continuous-time DSGE model, we augment $\boldsymbol{\delta}(\boldsymbol{\theta},\mathbf{T}_h)$ to $\boldsymbol{\delta}^{+}(\boldsymbol{\theta},\mathbf{T}_h)$ in (3.11), and hence $\mathcal{J}(\boldsymbol{\theta},\mathbf{T}_h)$ to $\mathcal{J}^{+}(\boldsymbol{\theta},\mathbf{T}_h)$, affording an additional $n_y^s n_x$ rows for fulfilling the rank and order conditions for the relevant matrix $\boldsymbol{\Delta}^{+}(\boldsymbol{\theta})=\mathcal{J}^{+}(\boldsymbol{\theta},\mathbf{I}_{n_x})$ at $\boldsymbol{\theta}_0$,

$$\boldsymbol{\Delta}^{+}(\boldsymbol{\theta}_0)=\left(\begin{array}{cc}\frac{\partial\mathrm{vec}(\mathbf{C}^s(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}} & -\mathbf{I}_{n_x}\otimes\mathbf{C}^s(\boldsymbol{\theta})\\ & \boldsymbol{\Delta}(\boldsymbol{\theta})\end{array}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{3.18}$$

Thus, our extension is based on the Jacobian $\left(\frac{\partial\mathrm{vec}(\mathbf{C}^s(\boldsymbol{\theta})\mathbf{T}_h^{-1})}{\partial\boldsymbol{\theta}},\frac{\partial\mathrm{vec}(\mathbf{C}^s(\boldsymbol{\theta})\mathbf{T}_h^{-1})}{\partial\mathrm{vec}\mathbf{T}_h}\right)$ of the additional exact observational equivalence conditions (3.10) on $\mathbf{C}^s(\boldsymbol{\theta})$.[18]

When the first m_θ columns of $\boldsymbol{\Delta}^{+}(\boldsymbol{\theta}_0)$ form a submatrix of full rank, then $\boldsymbol{\Lambda}^{+}(\boldsymbol{\theta})$ is locally invertible at $\boldsymbol{\theta}_0$, hence securing local identification of $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$ if $\boldsymbol{\Lambda}^{+}(\boldsymbol{\theta}_0)$ is locally identifiable, i.e., if $\mathbf{T}_h=\mathbf{I}_{n_x}$ is the only local similarity transform. The latter condition is met if the remaining n_x^2 columns of (3.18) form a full rank submatrix.

From Lemma 3.(b), in the absence of local aliases, the arguments are repeated, with $\boldsymbol{\Lambda}^{+}(\boldsymbol{\theta})$ replaced by $\boldsymbol{\Lambda}_a(\boldsymbol{\theta})$, $\boldsymbol{\delta}^{+}(\boldsymbol{\theta},\mathbf{T}_h)$ by $\boldsymbol{\delta}_a(\boldsymbol{\theta},\mathbf{T}_h)$ from (3.15), $\mathcal{J}^{+}(\boldsymbol{\theta},\mathbf{T}_h)$ by $\mathcal{J}_a(\boldsymbol{\theta},\mathbf{T}_h)=\left(\frac{\partial\boldsymbol{\delta}_a(\boldsymbol{\theta},\mathbf{T}_h)}{\partial\boldsymbol{\theta}},\frac{\partial\boldsymbol{\delta}_a(\boldsymbol{\theta},\mathbf{T}_h)}{\partial\mathrm{vec}\mathbf{T}_h}\right)$, $\boldsymbol{\Delta}^{+}(\boldsymbol{\theta})$ by $\boldsymbol{\Delta}_a(\boldsymbol{\theta})=\mathcal{J}_a(\boldsymbol{\theta},\mathbf{I}_{n_x})$, and $\boldsymbol{\Delta}^{+}(\boldsymbol{\theta}_0)$ from (3.18) by

$$\boldsymbol{\Delta}_a(\boldsymbol{\theta}_0)=\left(\begin{array}{cc}\frac{\partial\mathrm{vec}(\mathbf{A}(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}} & \mathbf{A}(\boldsymbol{\theta})^{\top}\otimes\mathbf{I}_{n_x}-\mathbf{I}_{n_x}\otimes\mathbf{A}(\boldsymbol{\theta})\\ \frac{\partial\mathrm{vec}(\mathbf{C}(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}} & -\mathbf{I}_{n_x}\otimes\mathbf{C}(\boldsymbol{\theta})\\ \frac{\partial\mathrm{vec}(\mathsf{K}(\boldsymbol{\theta};h))}{\partial\boldsymbol{\theta}} & \mathsf{K}(\boldsymbol{\theta},h)^{\top}\otimes\mathbf{I}_{n_x}\\ \frac{\partial\mathrm{vech}(\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h))}{\partial\boldsymbol{\theta}} & \mathbf{0}_{n_y(n_y+1)/2\times n_x^2}\end{array}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{3.19}$$

Similarly, from Lemma 3.(c), in the absence of local aliases, and under additional condi-

[18] See the proof of Proposition 2 for the derivatives of $\boldsymbol{\delta}^{+}(\boldsymbol{\theta},\mathbf{T}_h)$ in $\mathcal{J}^{+}(\boldsymbol{\theta},\mathbf{T}_h)$ for arbitrary $(\boldsymbol{\theta},\mathbf{T}_h)$.

tions on $\mathbf{C}^s(\boldsymbol{\theta}_0)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}$, the arguments relating to local identifiability of the parameters of the continuous-time model may be repeated once more, now with $\boldsymbol{\Lambda}^+(\boldsymbol{\theta})$ replaced by $\boldsymbol{\Lambda}_c(\boldsymbol{\theta})$ from (2.3), $\boldsymbol{\delta}^+(\boldsymbol{\theta},\mathbf{T}_h)$ by $\boldsymbol{\delta}_c(\boldsymbol{\theta},\mathbf{T}_h,\mathbf{U}_h)$ from (3.14), $\mathcal{J}^+(\boldsymbol{\theta},\mathbf{T}_h)$ by $\mathcal{J}_c(\boldsymbol{\theta},\mathbf{T}_h) = \left(\frac{\partial\boldsymbol{\delta}_c(\boldsymbol{\theta},\mathbf{T}_h,\mathbf{U}_h)}{\partial\boldsymbol{\theta}}, \frac{\partial\boldsymbol{\delta}_c(\boldsymbol{\theta},\mathbf{T}_h,\mathbf{U}_h)}{\partial\mathrm{vec}\mathbf{T}_h}, \frac{\partial\boldsymbol{\delta}_c(\boldsymbol{\theta},\mathbf{T}_h,\mathbf{U}_h)}{\partial\mathrm{vec}\mathbf{U}_h}\right)$, $\boldsymbol{\Delta}^+(\boldsymbol{\theta})$ by $\boldsymbol{\Delta}_c(\boldsymbol{\theta}) = \mathcal{J}_c(\boldsymbol{\theta},\mathbf{I}_{n_x},\mathbf{I}_{m_w})$, and $\boldsymbol{\Delta}^+(\boldsymbol{\theta}_0)$ by

$$\boldsymbol{\Delta}_c(\boldsymbol{\theta}_0) = \begin{pmatrix} \frac{\partial\mathrm{vec}(\mathbf{A}(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}} & \mathbf{A}(\boldsymbol{\theta})^\top\otimes\mathbf{I}_{n_x} - \mathbf{I}_{n_x}\otimes\mathbf{A}(\boldsymbol{\theta}) & \mathbf{0}_{n_x^2\times m_w^2} \\ \frac{\partial\mathrm{vec}(\mathbf{B}(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}} & \mathbf{B}(\boldsymbol{\theta})^\top\otimes\mathbf{I}_{n_x} & \mathbf{I}_{m_w}\otimes\mathbf{B}(\boldsymbol{\theta}) \\ \frac{\partial\mathrm{vec}(\mathbf{C}(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}} & -\mathbf{I}_{n_x}\otimes\mathbf{C}(\boldsymbol{\theta}) & \mathbf{0}_{n_yn_x\times m_w^2} \end{pmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} . \tag{3.20}$$

In contrast to $\boldsymbol{\Delta}(\boldsymbol{\theta}_0)$ in (3.17), which involves the matrices $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ of the discrete-time representation (2.15)-(2.16) and their derivatives, following Komunjer and Ng (2011), $\boldsymbol{\Delta}_c(\boldsymbol{\theta}_0)$ in (3.20) depends exclusively on the matrices $\boldsymbol{\Lambda}_c(\boldsymbol{\theta})$ of the continuous-time model (2.1)-(2.2) and their derivatives. If the middle n_x^2 columns of $\boldsymbol{\Delta}_c(\boldsymbol{\theta}_0)$ form a full rank submatrix, then $\mathbf{T}_h = \mathbf{I}_{n_x}$ is the only local similarity transform, and the continuous-time matrices $\mathbf{A}(\boldsymbol{\theta}_0)$ and $\mathbf{C}(\boldsymbol{\theta}_0)$ are locally identifiable. If, in addition, the last m_w^2 columns form a full rank submatrix,[19] then $\mathbf{U}_h = \mathbf{I}_{m_w}$ is the only factor rotation, and $\mathbf{B}(\boldsymbol{\theta}_0)$ is locally identifiable, as well. In this case, the observational equivalence conditions (3.9) are behind the result. Without local constancy of the measurement error covariance, the relevant matrix (for developing the rank condition) is $\boldsymbol{\Delta}_a(\boldsymbol{\theta}_0)$ from (3.19), the intermediate case, with the discrete-time matrices A, C replaced by their continuous-time counterparts $\mathbf{A}$, $\mathbf{C}$, while retaining K, $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}$ as in (3.17).

Proposition 2 (Rank and order conditions)**.** *Under Assumptions 1-4, for given $h > 0$,*

(a) if $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is a regular point of $\boldsymbol{\Delta}^+(\boldsymbol{\theta})$, then the necessary and sufficient rank condition for the identification of $\boldsymbol{\theta}$ from the autocovariances of $\mathbf{y}_\tau$ in a neighborhood of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is

$$\mathrm{rank}\,\boldsymbol{\Delta}^+(\boldsymbol{\theta}_0) = m_\theta + n_x^2\,. \tag{3.21}$$

A necessary order condition is that

$$m_\theta \leq n_{\delta^+} - n_x^2 = n_y^s n_x + 2n_xn_y + n_y(n_y+1)/2\,; \tag{3.22}$$

(b) in the absence of local aliases (e.g., under the additional Assumption 5), if $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is a regular point of $\boldsymbol{\Delta}_a(\boldsymbol{\theta})$, then the necessary and sufficient rank condition for the identification of $\boldsymbol{\theta}$ from the autocovariances of $\mathbf{y}_\tau$ in a neighborhood of $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is

$$\mathrm{rank}\,\boldsymbol{\Delta}_a(\boldsymbol{\theta}_0) = m_\theta + n_x^2\,. \tag{3.23}$$

[19] Note that this requires full column rank of $\mathbf{B}(\boldsymbol{\theta}_0)$.

A necessary order condition is that

$$m_\theta \leq n_\delta - n_x^2 = 2n_x n_y + n_y(n_y+1)/2\,; \tag{3.24}$$

(c) in the absence of local aliases, if $\boldsymbol{\theta}_0$ *is a regular point of* $\boldsymbol{\Delta}_c(\boldsymbol{\theta})$, $\mathbf{C}^s(\boldsymbol{\theta}_0)$ *is of full column rank, and* $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta})$ *is constant in a neighborhood of* $\boldsymbol{\theta}_0$*, then the necessary and sufficient rank condition for the identification of* $\boldsymbol{\theta}$ *from the autocovariances of* $\mathbf{y}_\tau$ *in a neighborhood of* $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ *is*

$$\operatorname{rank} \boldsymbol{\Delta}_c(\boldsymbol{\theta}_0) = m_\theta + n_x^2 + m_w^2\,. \tag{3.25}$$

A necessary order condition is that

$$m_\theta \leq n_c - n_x^2 - m_w^2 = n_x n_y + (n_x - m_w)m_w\,. \tag{3.26}$$

In Proposition 2, a regular point of a given Jacobian is one for which the rank remains constant in a neighborhood. Relative to the condition for discrete-time linear DSGE models, the rank condition (3.21) exploits the $n_y^s n_x$ additional restrictions (3.10) on $\mathbf{C}^s$ implied by the rational expectations solution of the continuous-time DSGE model to help achieve sufficient rank in this case. The order condition (3.22) requires that the number of equations in condition (3.12) at least equal the number of unknowns in $(\boldsymbol{\theta}_\ell, \mathbf{T}_h)$, i.e., $m_\theta + n_x^2$. This, too, is more easily met in the presence of the underlying continuous-time model, for $n_y^s > 0$ (some measurements are stocks). Substituting $\boldsymbol{\Delta}^+(\boldsymbol{\theta}_0)$ from (3.18) for $\boldsymbol{\Delta}(\boldsymbol{\theta}_0)$ from (3.17) in (3.21) and $n_y^s = 0$ in (3.22) returns the Komunjer and Ng (2011) rank and order conditions.

In part (a) of Proposition 2, local identification is achieved regardless of potential aliasing. Part (b) shows that absence of local aliases, such as in the case considered by Phillips (1973), suffices for recasting the local identifiability conditions directly in terms of the continuous-time transition and loading matrices $\mathbf{A}$, $\mathbf{C}$ from (2.1)-(2.2) and their parameter derivatives, as opposed to the discrete-time derived counterparts A, C. By part (c), the remaining Kalman filter matrices in the conditions, K, $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}$, are replaced by the continuous-time diffusion matrix $\mathbf{B}$ from (2.1) under the additional conditions of full rank of $\mathbf{C}^s(\boldsymbol{\theta}_0)$ and locally constant (e.g., zero or known) measurement error covariance, and the requirement on the right side of (3.25) is higher, by m_w^2, as the similarity transform $\mathbf{T}_h$ is accompanied by the rotation $\mathbf{U}_h$. In this case, the rank and order conditions do not involve h, and hence apply for arbitrary spacing of observations—a result of reliance on the underlying continuous-time model.

The assumption of no local aliases in parts (b) and (c) (in particular, Assumption 5)

does not preclude the existence of aliases at isolated points (i.e., not local to $\boldsymbol{\theta}_0$).[20] Further, even without any assumption on aliasing, the local identification conditions in part (a) of the Proposition are exact, and the conditions in parts (b) and (c) provide useful approximate criteria for local identification that apply to order h or better in general, cf. Lemma 2.(d)-(e).

In many applications, the structural parameters $\boldsymbol{\theta}$ are restricted *a priori*. This includes circumstances in which a subset of $n_R < m_\theta$ parameters are calibrated to particular values. Following Rothenberg (1971), and the analysis of conditional identification in Komunjer and Ng (2011), the conditions in Proposition 2 can be extended in a straightforward manner by augmenting (3.11), (3.13) or (3.15) with the function $\boldsymbol{\varphi}$ defining the parameter restrictions. Again, this implies that the number of rows in each of the Jacobians (3.18)-(3.20) in the rank conditions (3.21), (3.23) and (3.25) is increased, now by n_R, and so is the upper bound in each of the order conditions (3.22), (3.24) and (3.26), hence leaving the conditions easier to fulfill. In this type of extension, the additional rows in the Jacobian involves derivatives with respect to $\boldsymbol{\theta}$, but neither $\mathbf{T}_h$ nor $\mathbf{U}_h$, unlike in our analysis.

4. Recovering structural shocks

The reduced-form disturbances $\boldsymbol{\eta}_\tau$ in (2.7) are composites of the underlying structural shocks $\mathrm{d}\mathbf{w}(u)$ occurring continuously within the time interval $(t_{\tau-1}, t_\tau]$. We introduce a simple approach to recover a sequence of structural shocks at measurement times, $\{\mathbf{u}_\tau\}_{\tau\in\mathbb{Z}^+}$, from the reduced-form disturbances, where

$$\mathbf{u}_\tau = h^{-1/2}(\mathbf{w}(t_\tau) - \mathbf{w}(t_{\tau-1})) \tag{4.1}$$

is an m_w-vector of Gaussian random variables with $\mathbb{E}[\mathbf{u}_\tau] = \mathbf{0}$, and $\mathbb{E}\left[\mathbf{u}_\tau\mathbf{u}_\tau^\top\right] = \mathbf{I}_{m_w}$.

Conceptually, our strategy partitions $(t_{\tau-1}, t_\tau]$ into $n \geq 1$ sub-intervals $(t_{i-1}^\tau, t_i^\tau]$ of length $h_n = h/n$, with $t_{\tau-1} = t_0^\tau < t_1^\tau < \cdots < t_{n-1}^\tau < t_n^\tau = t_\tau$, as depicted in Figure 1. Thus, for given n, integrated structural shocks are given by the Riemann sum $\sum_{i=1}^n \delta\mathbf{w}(t_i^\tau) = \mathbf{w}(t_\tau) - \mathbf{w}(t_{\tau-1}) = \int_{t_{\tau-1}}^{t_\tau} \mathrm{d}\mathbf{w}(u)$, where $\delta\mathbf{w}(t_i^\tau) = \mathbf{w}(t_i^\tau) - \mathbf{w}(t_{i-1}^\tau)$. For given $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the relation between the reduced-form disturbances and the structural shocks can be written as

$$\boldsymbol{\eta}_\tau = \int_{t_{\tau-1}}^{t_\tau} \mathbf{H}(\boldsymbol{\theta}; t_\tau - u)\mathrm{d}\mathbf{w}(u) = \lim_{n\to\infty} \sum_{i=1}^n \mathbf{H}(\boldsymbol{\theta}; t_\tau - t_{i-1}^\tau)\delta\mathbf{w}(t_i^\tau)\,, \tag{4.2}$$

[20] This is shown in Online Appendix F. Stronger conditions ensuring global injectivity (no aliasing) based on external restrictions on the transition matrix, e.g., in terms of eigenvalues and their multiplicity, and/or the covariance matrix, have been proposed by Phillips (1973), McCrorie (2003), Blevins (2017), and others.

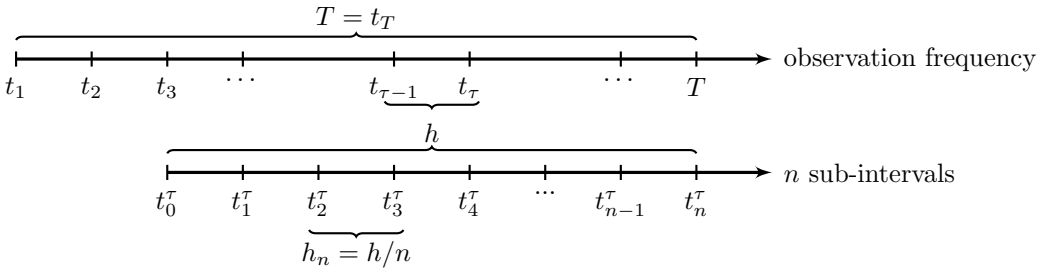


Figure 1. Partitioning of sampling interval. The figure illustrates the observation points in the sample, as well as the assumed subsampling scheme within each observation interval.

with $\mathbf{H}(\boldsymbol{\theta}; \cdot)$ a deterministic square integrable $n_x \times m_w$ matrix function over $(0, h]$,

$$\mathbf{H}(\boldsymbol{\theta}; r) = \begin{bmatrix} \exp\left(\mathbf{A}(\boldsymbol{\theta})r\right)\mathbf{B}(\boldsymbol{\theta}) \\ \mathbf{S}_x^f \mathbf{A}(\boldsymbol{\theta})^{-1}\left(\exp\left(\mathbf{A}(\boldsymbol{\theta})r\right) - \mathbf{I}_{n_x^s}\right)\mathbf{B}(\boldsymbol{\theta}) \end{bmatrix}. \tag{4.3}$$

Proposition 3. *The mapping in* (4.2) *can be written as*

$$\boldsymbol{\eta}_\tau = h^{1/2}\mathbf{H}(\boldsymbol{\theta}; h)\mathbf{u}_\tau + \mathcal{R}_\tau\,, \tag{4.4}$$

where the remainder term, $\mathcal{R}_\tau$*, is stochastically bounded,*

$$\mathcal{R}_\tau = \mathcal{O}_P(h^{3/2})\,, \;\; \textit{as } h \to 0\,. \tag{4.5}$$

Remark 4.1. Alternative characterizations of $\mathcal{R}_\tau$ in terms of the number of sub-intervals, n, or the length of each, h_n, are given in the proof of Proposition 3 in Online Appendix A. ■

The remainder $\mathcal{R}_\tau$ in Proposition 3 provides a measure of the *approximation error* incurred when backing out $\mathbf{u}_\tau$ from $\boldsymbol{\eta}_\tau$ using

$$\boldsymbol{\eta}_\tau \approx h^{1/2}\mathbf{H}(\boldsymbol{\theta}; h)\mathbf{u}_\tau\,, \tag{4.6}$$

which is equivalent to setting $n = 1$ in (4.2), so that $h_n = h$. This approximate linear relation is illustrated in the following two examples, which resemble that in Section 5.1.

Example 1 (State vector with only stocks)**.** Consider the bivariate continuous-time system

$$\begin{bmatrix} \mathrm{d}x_1(t) \\ \mathrm{d}x_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \mathrm{d}t + \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \begin{bmatrix} \mathrm{d}w_1(t) \\ \mathrm{d}w_2(t) \end{bmatrix}, \tag{4.7}$$

where the coefficients may be nonlinear functions of $\boldsymbol{\theta}$, and $n_x^f = 0$, so that $n_x = n_x^s = 2$. At measurement times τ, the state vector $\mathbf{x}_\tau = \left[x_{1,\tau}^s,\, x_{2,\tau}^s\right]^\top$ satisfies the transition equation (2.4) where, by (4.6), the reduced-form disturbances are approximately given by

$$\begin{aligned}\begin{bmatrix} \eta_{1,\tau}^s \\ \eta_{2,\tau}^s \end{bmatrix} &\approx h^{1/2} \begin{bmatrix} \exp(a_{11}h) & \frac{a_{12}\exp(a_{11}h) - a_{12}\exp(a_{22}h)}{a_{11}-a_{22}} \\ 0 & \exp(a_{22}h) \end{bmatrix} \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} \begin{bmatrix} u_{1,\tau} \\ u_{2,\tau} \end{bmatrix} \\ &= h^{1/2} \begin{bmatrix} u_{1,\tau} b_1 \exp(a_{11}h) + u_{2,\tau} a_{12} b_2 \frac{\exp(a_{11}h) - \exp(a_{22}h)}{a_{11}-a_{22}} \\ u_{2,\tau} b_2 \exp(a_{22}h) \end{bmatrix} \end{aligned} \tag{4.8}$$

in terms of the structural shocks $[u_{1,\tau}, u_{2,\tau}]^\top$. Here, $u_{2,\tau}$ impacts both $\eta_{1,\tau}^s$ and $\eta_{2,\tau}^s$, so the state variables are correlated in discrete time, although not in continuous time. ■

Example 2 (State vector with stocks and flows)**.** Assume $[x_1(t), x_2(t)]^\top$ is governed by (4.7), and let $n_x^f = 1$, with $\mathbf{S}_x^f = [1,\, 0]$, so that $n_x = n_x^s + n_x^f = 3$. In this case, the state vector in (2.4) is $\mathbf{x}_\tau = [x_{1,\tau}^s,\, x_{2,\tau}^s, x_{1,\tau}^f]^\top = [x_{1,\tau}^s,\, x_{2,\tau}^s, \int_{t_{\tau-1}}^{t_\tau} x_1(u)\mathrm{d}u]^\top$, and the reduced-form disturbances $\boldsymbol{\eta}_\tau = [\eta_{1,\tau}^s, \eta_{2,\tau}^s, \eta_{1,\tau}^f]^\top$, with $\eta_{1,\tau}^s$, $\eta_{2,\tau}^s$ as in Example 1, and

$$\eta_{1,\tau}^f \approx h^{1/2}\left(u_{1,\tau} b_1 \frac{(\exp(a_{11}h) - 1)}{a_{11}} + u_{2,\tau} a_{12} b_2 \frac{a_{11} - a_{22} - a_{11}\exp(a_{22}h) + a_{22}\exp(a_{11}h)}{a_{11}a_{22}(a_{11} - a_{22})} \right). ■$$

Remark 4.2. By (4.8), the dimension of $\boldsymbol{\eta}_\tau$ equals that of the state vector even if $b_1 = 0$ in each of Examples 1 and 2, i.e., when the endogenous state variable is not subject to structural shocks. Hence, there can be more reduced-form disturbances than structural shocks. This is particularly relevant when checking the validity of Assumption 2. ■

Since the econometrician rarely has control over n, the next proposition shows that the error using $n = 1$ is bounded in probability. Thus, writing $\mathbf{H}(\boldsymbol{\theta}) = h^{1/2}\mathbf{H}(\boldsymbol{\theta}; h)$, the structural shocks at measurement times are recovered from the reduced-form disturbances as

$$\tilde{\mathbf{u}}_\tau = \mathbf{H}(\boldsymbol{\theta})^\dagger \boldsymbol{\eta}_\tau\,, \tag{4.9}$$

based on (4.6), with superscript † denoting generalized (e.g., Moore-Penrose) inverse. The approximation implied by (4.9) resembles the identification mechanisms commonly used in the structural VAR literature. Theoretical restrictions are imposed, here through $\mathbf{H}(\boldsymbol{\theta})$, to uncover the structural shocks otherwise hidden in the correlated reduced-form disturbances $\boldsymbol{\eta}_\tau$.

Proposition 4. *For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the error in the approximation of the structural shocks is stochastically bounded,*

$$\tilde{\mathbf{u}}_\tau - \mathbf{u}_\tau = \mathcal{O}_P(1)\,.$$

Example 3 (Recovering structural shocks)**.** Consider the model in Example 1. Solving (4.8),

$$\begin{aligned}\tilde{u}_{2,\tau} &= h^{-1/2}\exp(-a_{22}h)b_2^{-1}\eta_{2,\tau}\,,\\ \tilde{u}_{1,\tau} &= h^{-1/2}\exp(-a_{11}h)b_1^{-1}\eta_{1,\tau} - h^{-1/2}\frac{a_{12}\exp(-(a_{11}+a_{22})h)(\exp(a_{11}h)-\exp(a_{22}h))}{b_1(a_{11}-a_{22})}\eta_{2,\tau}\,.\end{aligned}$$

While $\tilde{u}_{2,\tau}$ is proportional to $\eta_{2,\tau}$, $\tilde{u}_{1,\tau}$ is a linear combination of both residuals. ■

5. Monte Carlo evidence

5.1 The artificial economy

We consider a continuous-time version of the RBC model with indivisible labor of Hansen (1985), with shocks to capital and TFP. A complete derivation is in Online Appendix G.

Preferences. A representative agent maximizes expected discounted lifetime utility from consumption $C(t)$ and leisure $L(t)$, $\mathbb{E}_0\int_0^\infty e^{-\rho t}\left(\ln C(t)+\psi L(t)\right)\mathrm{d}t$, where $\rho>0$ is the subjective discount rate, and ψ the weight of leisure in utility. Hours worked per unit of time are $N(t)=1-L(t)$. The agent's income $Y(t)$ consists of wages and rents from selling labor and renting capital to firms, and is allocated between consumption and investment, $C(t)+I(t)=W(t)N(t)+r(t)K(t)$, where $W(t)$ is the real wage, $r(t)$ the real interest rate, and $K(t)$ the capital stock. The latter evolves according to

$$\mathrm{d}K(t)=\left(I(t)-\delta K(t)\right)\mathrm{d}t+\sigma_k K(t)\mathrm{d}w_k(t),\quad K\left(0\right)=K_0\,, \tag{5.1}$$

with $\delta\geq 0$ the mean depreciation rate, and $w_k(t)$ a standard Brownian motion representing shocks to the depreciation rate, marginal efficiency of investment, or future productivity of the capital stock. The diffusion parameter $\sigma_k>0$ regulates the variance of these shocks.

Technology. The representative firm produces aggregate output $Y(t)$ according to

$$Y(t)=\exp\left(Z(t)\right)K(t)^{\alpha}\left(\exp\left(\eta t\right)N(t)\right)^{1-\alpha}\,,\quad \alpha\in(0,1)\,, \tag{5.2}$$

where $\eta>1$ is the constant growth rate of labor-augmenting technological progress, and $Z(t)$ a zero-mean measure of TFP, evolving according to the Ornstein-Uhlenbeck process

$$\mathrm{d}Z(t)=-\rho_z Z(t)\mathrm{d}t+\sigma_z\mathrm{d}w_z(t)\,,\quad Z\left(0\right)=Z_0\,, \tag{5.3}$$

with mean reversion $\rho_z>0$, and shocks $\mathrm{d}w_z(t)$ independent of $\mathrm{d}w_k(t)$, with volatility $\sigma_z>0$.

Equilibrium. The equilibrium allocations $\{c(t), n(t), k(t)\}_{t\geq 0}$, detrended at rate η, satisfy

$$0 = \psi c(t)n(t) - (1-\alpha)\exp(z(t))k(t)^{\alpha}n(t)^{1-\alpha}\,, \tag{5.4}$$

$$\begin{aligned}\mathbb{E}_t\left[\mathrm{d}c(t)\right] = \Bigg[&\left(\alpha\exp\left(z(t)\right)k(t)^{\alpha-1}n(t)^{1-\alpha} - \rho - \delta - \eta\right) - \sigma_k^2\frac{k(t)\mathbf{c}_k\left(k(t),z(t)\right)}{c(t)} \\ &+ \frac{1}{2}\left(\sigma_k^2\left(\frac{k(t)\mathbf{c}_k\left(k(t),z(t)\right)}{c(t)}\right)^2 + \sigma_z^2\left(\frac{\mathbf{c}_z\left(k(t),z(t)\right)}{c(t)}\right)^2\right)\Bigg]c(t)\mathrm{d}t\,,\end{aligned} \tag{5.5}$$

$$\mathrm{d}k(t) = \left(\exp\left(z(t)\right)k(t)^{\alpha}n(t)^{1-\alpha} - c(t) - (\delta+\eta)\,k(t)\right)\mathrm{d}t + \sigma_k k(t)\mathrm{d}w_k(t)\,, \tag{5.6}$$

with $z(t)$ governed by (5.3). Equation (5.4) is the intratemporal labor supply equation, and (5.5) the Euler equation, with $\mathbf{c}_k(\cdot)$ and $\mathbf{c}_z(\cdot)$ the marginal responses of optimal consumption $\mathbf{c}(\cdot)$ to changes in capital and TFP. The solution is not available in closed form, and we approximate it by log-linearizing the equilibrium conditions around the deterministic steady state, then computing the rational expectations solution using the QZ (generalized Schur) decomposition of Sims (2002). Under the Blanchard and Kahn (1980) conditions, the solution admits the representation (2.1)-(2.2), with $\mathbf{x}(t) = [\hat{k}(t), \hat{z}(t)]^\top$, $\mathbf{y}(t) = [\hat{c}(t), \hat{n}(t)]^\top$, $\mathrm{d}\mathbf{w}(t) = [\mathrm{d}w_k(t), \mathrm{d}w_z(t)]^\top$, circumflex denoting log-deviations from steady state values, and

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix}\phi_{kk} & \phi_{kz}\\ 0 & -\rho_z\end{bmatrix}, \quad \mathbf{B}(\boldsymbol{\theta}) = \begin{bmatrix}\sigma_k & 0\\ 0 & \sigma_z\end{bmatrix}, \quad \mathbf{C}(\boldsymbol{\theta}) = \begin{bmatrix}\phi_{ck} & \phi_{cz}\\ \phi_{nk} & \phi_{nz}\end{bmatrix}, \tag{5.7}$$

where the reduced-form parameters $\boldsymbol{\phi} = (\phi_{ck}, \phi_{cz}, \phi_{nk}, \phi_{nz}, \phi_{kk}, \phi_{kz})^\top$ depend nonlinearly on the structural parameters $\boldsymbol{\theta} = (\rho, \psi, \alpha, \delta, \eta, \rho_z, \sigma_z, \sigma_k)^\top$.[21]

5.2 Finite sample properties of the MLE

We investigate the finite sample properties of the estimators of the structural parameters $\boldsymbol{\theta}$ of the continuous-time model from Section 5.1 based on discrete-time data. We generate $M = 10,000$ samples for consumption, C, and labor, N, from the data generating process (DGP) given by the nonlinear solution to the RBC model, using the parameter values in Table 1.[22] With the estimator relying on the solution to the linearized model (5.7), our analysis in effect regards the estimation of a misspecified (linearized) DSGE model. In the Online Appendix, we report results using the alternative DGP based on the linear approximation.

[21] The analytical mapping between $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, provided in Online Appendix G, reveals that ψ is unidentified, since it vanishes from the rational expectations solution of the model.

[22] Since the model does not admit a closed-form solution, we approximate it globally using collocation methods (see Parra-Alvarez, 2018). Details on the solution and simulation are in Online Appendix G.

The parameter values are standard in the literature, and match long-run values of U.S. macroeconomic aggregates observed over the postwar period. We set $\alpha = 0.30$ to match an average labor income to GDP ratio of 70%. With a time unit of one year, the depreciation rate is set to $\delta = 0.06$, and the subjective discount rate to $\rho = 0.03$, consistent with steady-state values for the net return on capital and the investment to GDP ratio of 4% and 20%, respectively. The weight of leisure in the instantaneous utility function is fixed at $\psi = 2.686$ so that, in steady state, agents spend 1/3 of their time working. The long-run growth rate of the economy is assumed to be $\eta = 2\%$. The parameters describing TFP dynamics are set to $\rho_z = 0.2052$ and $\sigma_z = 0.0140$, in line with standard estimates from the quarterly Solow residual for the U.S. economy (e.g., Hansen, 1985). Finally, we fix the volatility of the capital stock at $\sigma_k = 0.0104$, based on the calibration in Ambler and Paquet (1994).

Each Monte Carlo sample contains 240 quarterly observations, corresponding to 60 years of data, without measurement error. For each sample, we estimate the model parameters using the ML approach from Section 2 and two observables, consumption and the fraction of hours worked, $\mathbf{y}_\tau = [C_\tau, N_\tau]^\top$. With the latent state variables given by capital and TFP, $n_x \geq 2$ in all representations of the ED-SSR in Proposition 1, so $n_\epsilon \geq n_x \geq 2 = n_y$, as required to avoid stochastic singularity (indeed, Assumption 2 is satisfied at $\boldsymbol{\theta}_0$).

We compare the ML estimates based on the different versions of the ED-SSR (2.4)-(2.5), namely, the S-SSR, F-SSR, and MX-SSR (stock, flow, and mixed) representations,[23] to those obtained from the misspecified state space representation resulting from a *naive* Euler-Maruyama (EM) discretization of the continuous-time transition equation (2.1), i.e.,

$$\mathbf{x}_\tau = (\mathbf{I}_{n_x} + \mathbf{A}(\boldsymbol{\theta})h)\,\mathbf{x}_{\tau-1} + \sqrt{h}\mathbf{B}(\boldsymbol{\theta})\boldsymbol{\zeta}_\tau\,, \tag{5.8}$$

$$\mathbf{y}_\tau = \mathbf{C}(\boldsymbol{\theta})\mathbf{x}_\tau + \boldsymbol{\upsilon}_\tau\,, \tag{5.9}$$

using $\mathbb{E}[\boldsymbol{\zeta}_\tau] = \mathbf{0}$, $\mathbb{E}\left[\boldsymbol{\zeta}_\tau\boldsymbol{\zeta}_\tau^\top\right] = \mathbf{I}_{m_w}$ for the approximation. We refer to (5.8)-(5.9) as the EM-SSR model. It differs from the ED-SSR in that (i) the transition matrix is truncated to first order, ignoring terms of order smaller than h in (2.6), (ii) the disturbances ignore the temporal aggregation (2.7) of structural shocks through the system between the discrete observations, and (iii) the likelihood function is no longer exact, even for the linearized model, since the dynamics between observations are ignored.

We group the parameters in two categories, $\boldsymbol{\theta} = \left[\boldsymbol{\theta}_{\text{ss}}^\top, \boldsymbol{\theta}_{\text{exo}}^\top\right]^\top$, with $\boldsymbol{\theta}_{\text{ss}} = [\psi, \alpha, \delta, \rho, \eta]^\top$, and $\boldsymbol{\theta}_{\text{exo}} = [\rho_z, \sigma_z, \sigma_k]^\top$. The first includes parameters readily identified from the model's steady state and available measurements, and the second the parameters characterizing the

[23] Since we consider a linearized model, the ED-SSR is only exact up to linearization, i.e., misspecified relative to the nonlinear DSGE model. The detailed ED-SSR representation in Proposition 1 is provided in Online Appendix H for each of the cases considered in the Monte Carlo analysis.

Table 1. Parameter values (time unit one year).

Parameter	Value	Source / Target
Subjective discount rate, ρ	0.0300	Long-run (net) return on capital of 4%
Leisure weight, ψ,	2.6860	Average fraction of hours worked of 1/3
Capital share in output, α	0.3000	Average WN/Y of 0.7
Depreciation rate, δ	0.0600	Average I/Y of 0.2 and K/Y of 2.5
Labor-augmenting growth, η	0.0200	Average GDP growth
Mean reversion of TFP, ρ_z	0.2052	Persistence of Solow residual
Volatility of TFP, σ_z	0.0140	Volatility of Solow residual
Volatility of depreciation, σ_k	0.0104	Ambler and Paquet (1994)

dynamics of the exogenous processes driving the economy. For the Monte Carlo experiments, we focus on estimation of $\boldsymbol{\theta}_{\text{exo}}$, with $\boldsymbol{\theta}_{\text{ss}}$ fixed at the population values from Table 1.

Remark 5.1. By Remark 2.1, all the state space representations can be rewritten such that $n_x = m_x = 2$. For the parameter values in Table 1, Assumptions 1-4 are satisfied, with $\operatorname{rank} \mathcal{R}(\boldsymbol{\theta}_0; h) = \operatorname{rank} \mathcal{O}(\boldsymbol{\theta}_0; h) = 2$. There are $m_\theta = 8$ parameters, of which $n_R = 5$ are calibrated, and $\dim \boldsymbol{\theta}_{\text{exo}} = m_\theta - n_R = 3$ is less than the upper bound in each of the order conditions from Proposition 2, namely, 15 or 11 in (3.22) in the stock and flow measurement cases, and 11 and 4 in (3.24) and (3.26), respectively. Further, using the accompanying identification toolbox, we verify the rank conditions from the Proposition numerically. In particular, in (3.21), $\operatorname{rank} \boldsymbol{\Delta}^+(\boldsymbol{\theta}_0) = 7$, which equals $m_\theta - n_R + n_x^2 = 3 + 4$, with $\boldsymbol{\Delta}^+(\boldsymbol{\theta}_0)$ from (3.18) including the Jacobian of the new exact observational equivalence conditions (3.10). Hence, given $\boldsymbol{\theta}_{\text{ss},0}$, the parameter vector $\boldsymbol{\theta}_{\text{exo}}$ is locally identified around $\boldsymbol{\theta}_{\text{exo},0}$. ■

5.2.1 Stock data

Our first experiment assumes that all observables are sampled as stocks. Hence, the S-SSR representation provides the exact discretization, the ED-SSR, and the exact likelihood for the linearized model and, potentially, a better approximate likelihood for the nonlinear DSGE model than the misspecified F-SSR and EM-SSR representations. For comparison, we consider ML estimation of $\boldsymbol{\theta}_{\text{exo}}$ based on all three representations. Panel A of Table 2 summarizes the results on bias and root mean squared error (RMSE) of estimates across replications.

In terms of bias, the S-SSR based MLE outperforms the alternatives, for each parameter generating the smallest (in magnitude) bias across methods. Both bias and RMSE are within a reasonable range, relative to the magnitudes of the assumed true parameter values. This indicates that although the S-SSR based likelihood is only exact for the linearized DSGE model, it provides a useful approximation for the nonlinear model.

For each of the three methods, the largest bias occurs for the mean reversion parameter,

Table 2. Finite sample properties I.

Panel A: Data sampled as stocks							
		F-SSR		S-SSR		EM-SSR	
$\boldsymbol{\theta}_{\text{exo},0}$		Bias	RMSE	Bias	RMSE	Bias	RMSE
ρ_z	0.2052	0.0217	0.0317	0.0161	0.0259	-0.0177	0.0242
σ_z	0.014	0.0043	0.0044	-3.83e-05	0.0007	-0.0001	0.0007
σ_k	0.0104	0.0033	0.0033	-2.94e-05	0.0005	-0.0003	0.0006

Panel B: Data sampled as flows							
		F-SSR		S-SSR		EM-SSR	
$\boldsymbol{\theta}_{\text{exo},0}$		Bias	RMSE	Bias	RMSE	Bias	RMSE
ρ_z	0.2052	0.0159	0.0258	0.0104	0.0228	-0.0210	0.0268
σ_z	0.014	-4.92e-05	0.0007	-0.0026	0.0026	-0.0026	0.0027
σ_k	0.0104	-3.25e-05	0.0005	-0.0019	0.0020	-0.0022	0.0022

Note: The table reports statistics for estimates of $\boldsymbol{\theta}_{\text{exo}}$ from $M = 10,000$ samples of quarterly consumption and hours worked, covering 60 years ($T = 240$). Simulated measurements are sampled as stocks in Panel A, and as flows in Panel B. F-SSR (S-SSR) refers to the state space representation assuming measurements sampled as flows (stocks), and EM-SSR to the Euler-Maruyama discretization. The parameters in $\boldsymbol{\theta}_{\text{ss}}$ are fixed at the values in Table 1. With $\hat{\boldsymbol{\theta}}_{\text{exo},m}$ the estimates from the m^{th} sample, Bias $= M^{-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\theta}}_{\text{exo},m} - \boldsymbol{\theta}_{\text{exo},0})$, and RMSE $= (M^{-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\theta}}_{\text{exo},m} - \boldsymbol{\theta}_{\text{exo},0})^2)^{1/2}$.

ρ_z. Using the S-SSR, ρ_z is above its population value, by 0.0161, or 7.8%, on average. Since there is no discretization error, relative to the linearized model, any bias in the S-SSR case reflects the effects of linearization and pure *estimation bias*, only. It is known that a positive estimation bias in the rate of mean reversion arises in the case of direct discrete-time observations from continuous-time processes (as opposed to our latent TFP process). In the univariate OU case, this was noted by Merton (1980), and later verified by Tang and Chen (2009). Our results are consistent with the notion that such positive estimation bias in ρ_z, corresponding to downward estimation bias in persistence, or autocorrelation, carries over to our multivariate state space case, and contributes to the positive bias in the S-SSR case.

The bias increases in the F-SSR case, incorrectly treating the stock data as flows, and is negative for the EM-SSR method, which introduces discretization error, even relative to the linearized model. Again, such *discretization bias*, on top of the linearization and pure estimation biases, is known to exist in estimation based on direct discrete-time observations from diffusions. As shown by Lo (1988) in the univariate case, the discretization bias in turn implies misspecification of the likelihood function, leading to inconsistent estimates of the

parameters of the continuous-time model. Further, Wang et al. (2011) show that in systems of linear SDEs, the discretization and estimation biases in persistence parameters are of opposite sign, with the estimation bias dominating in financial applications, characterized by data recorded at a high frequency (see also Phillips and Yu, 2005). In principle, the discretization bias could be made arbitrarily small by sampling at sufficiently high frequency, i.e., letting $h \to 0$, but this is plainly impossible in macroeconomic applications. The estimation bias is inversely proportional to the total time span, Th, and thus cannot be reduced purely by sampling at higher frequency. The negative total bias in the EM-SSR estimation of ρ_z is consistent with the notion that the negative discretization bias carries over to our multivariate state space case and, at typical macroeconomic sampling frequency, more than offsets the positive linearization and estimation bias. The latter bias is non-negligible in the S-SSR case, hence further reinforcing that the negative discretization bias in the EM-SSR case is of considerable (indeed, dominating) magnitude in the macroeconomic setting.

Turning to the diffusion coefficients, σ_z and σ_k, the S-SSR method exhibits virtually no bias. The misspecified F-SSR produces biases that are orders of magnitude larger, more than 30% of true values. EM-SSR produces bias in σ_k, too, but smaller, at around 3%, which may be attributed to discretization bias, given nearly zero bias in the S-SSR case. On the other hand, EM-SSR does not induce any appreciable discretization bias in σ_z, i.e., the (small) reported bias may as well be linearization and estimation bias (S-SSR exhibits a small bias in the same direction). It is known that the discretization bias affects the speed of mean reversion, ρ_z, more severely than the diffusion parameter, σ_z, for observed processes. Our results suggest that also the relative importance of the discretization biases in mean reversion and diffusion coefficient carries over to our multivariate state space case.[24]

While the F-SSR produces largest bias and RMSE for each parameter, RMSE for S-SSR and EM-SSR are of similar magnitude, and for ρ_z in fact smallest in the EM-SSR case. Indeed, the standard deviation of the ρ_z estimates is greater for S-SSR (0.0203) than for EM-SSR (0.0165). This result for our multivariate state space case is again in line with those for observed univariate diffusions (see Wang et al., 2011). Thus, as illustrated in Panel A of Figure 2, showing the finite sample distributions of estimates, the methods seem to compete against a bias-variance trade-off for ρ_z, with S-SSR exhibiting less bias than EM-SSR, but greater variability. With respect to σ_z and σ_k, the standard deviations are low and, although highest for F-SSR and lowest for EM-SSR, the difference across methods is negligible, and the F-SSR standard deviations are swamped by the large biases in RMSE.

[24] In Online Appendix J, we illustrate the discretization bias using the log-likelihood profile for each parameter. The maximum of the EM-SSR based likelihood function is located to the left of the true value. This downward bias is most pronounced for ρ_z, but also affects the diffusion coefficients.

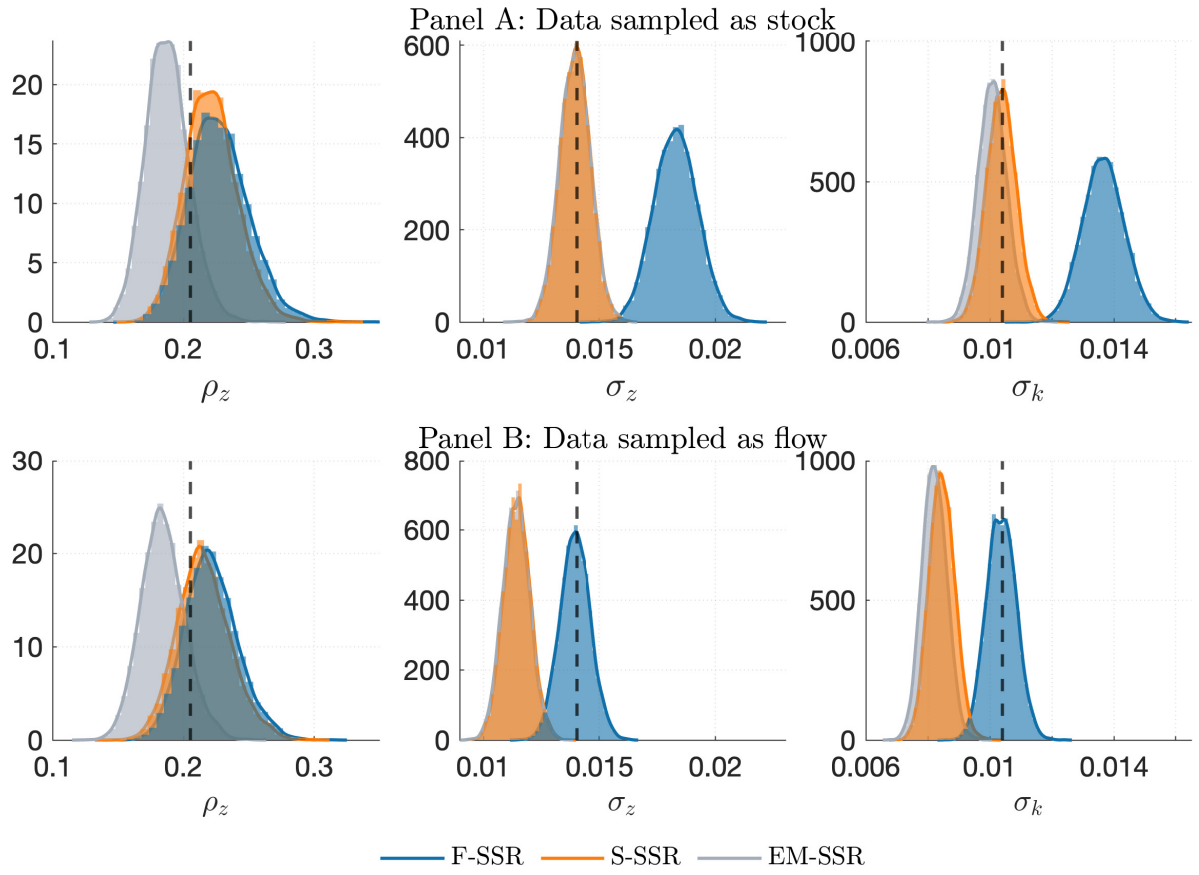


Figure 2. Finite sample distribution of parameter estimates. The figure shows the distributions of $\hat{\boldsymbol{\theta}}_{\text{exo}}$ across $M = 10,000$ samples of consumption and hours worked, using the S-SSR, F-SSR, and EM-SSR methods. Each sample includes $T = 240$ quarterly observations generated from the assumed true DGP. The data are sampled as stocks in Panel A, and as flows in Panel B.

Our conclusions are robust to using alternative observables, such as aggregate output and hours worked, or aggregate consumption and aggregate output.[25] In the absence of measurement error, the various bivariate series are equally informative. Moreover, our findings are robust to estimation of more parameters. In Online Appendix I, we estimate the discount rate, ρ, and the growth rate, η, along with $\boldsymbol{\theta}_{\text{exo}}$. The additional parameters were included after verifying our rank and order conditions for local identification.

5.2.2 Flow data

The second experiment assumes that all observables are instead sampled as flows. In this case, the F-SSR provides the ED-SSR and the exact likelihood for the linearized model and, potentially, the best approximate likelihood for the nonlinear DSGE model. The S-SSR and

[25] The results are available on request.

EM-SSR are misspecified and expected to exhibit *temporal aggregation* bias in case of flow variables, because they ignore the integral nature of measurements. Such temporal aggregation bias adds to the linearization and estimation biases and, in the EM-SSR case, the discretization bias, as well. Panel B of Table 2 reports the results of ML estimation.[26]

For ρ_z, the directions of bias agree with those for stock data, suggesting that the results on positive estimation bias for all three methods, and a larger negative discretization bias for EM-SSR, carry over to the case of flow data. For F-SSR, the bias amounts to about 7.8% in the flow case, closely corresponding to that of S-SSR in the stock case. For EM-SSR, the bias increases to 10.2% in magnitude, against 8.6% in the stock case, presumably reflecting additional temporal aggregation bias. RMSE and variability are comparable across methods.

The distributions of estimates are displayed in Panel B of Figure 2. For flow data, the bias-variance trade-off is present for all parameters, with more variability in the better centered distributions. The F-SSR method exhibits virtually no bias for the diffusion coefficients, σ_z and σ_k. In contrast, using the misspecified S-SSR and EM-SSR methods with flow data introduces substantial downward biases, of the order 20%, compared to less than 1% for stock data. These biases completely dominate the associated RMSEs in Table 2, Panel B, and the latter far exceed those of F-SSR. For S-SSR, the increase in bias cannot be attributed to discretization error, and so must be due to temporal aggregation bias. The results suggest that ignoring the state space representation that accommodates the flow nature of measurements comes at a cost in terms of properties of estimates. Although the exact state space representation for flows is slightly more complicated than that for stocks and the EM approximation, the benefits of pursuing it seem worthwhile in the context of DSGE models.

5.2.3 Mixed stock-flow data

The final experiment considers the case where the vector of observables consists of a mixture of stocks and flows. Thus, the MX-SSR method provides the ED-SSR for the linearized model. In a first exercise, we include aggregate consumption as the flow variable, and the real interest rate as the stock variable, $\mathbf{y}_\tau = [C_\tau, r_\tau]^\top$. In a second exercise, we include hours worked as an additional flow variable, so $\mathbf{y}_\tau = [C_\tau, N_\tau, r_\tau]^\top$.[27] In this case, we also study the effects of adding *iid* measurement errors, since the number of observables exceeds the number of reduced-form disturbances. The results are summarized in Table 3.

[26] Again, results are consistent across different combinations of observables. We also considered representations using the EM approximation (5.8) to the transition equation along with alternative measurement equations adjusted to capture the flow nature of the data. The Monte Carlo evidence revealed poorer performance than for the correctly specified state space representation. Results are available on request.

[27] We extend the matrix $\mathbf{C}(\boldsymbol{\theta})$ in (5.7) to include the (log-linear) equilibrium gross real interest rate as a function of the state variables. This additional equation is derived in Online Appendix G.

Table 3. Finite sample properties II.

Panel A: C sampled as flow, r sampled as stock									
		MX-SSR		F-SSR		S-SSR		EM-SSR	
$\boldsymbol{\theta}_{\text{exo}}$		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ρ_z	0.2052	0.0276	0.0434	0.3091	0.3212	0.1881	0.1976	0.1222	0.1304
σ_z	0.0140	0.0002	0.0008	0.0026	0.0027	-0.0011	0.0013	-0.0018	0.0019
σ_k	0.0104	0.0001	0.0006	0.0033	0.0034	0.0003	0.0006	-3.38e-05	0.0005

Panel B: C and N sampled as flows, r sampled as stock									
		MX-SSR		F-SSR		S-SSR		EM-SSR	
$\boldsymbol{\theta}_{\text{exo}}$		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ρ_z	0.2052	0.0141	0.0232	0.0237	0.0304	0.0194	0.0269	-0.0043	0.0165
σ_z	0.0140	-4.32e-05	0.0007	-0.0001	0.0007	-0.0026	0.0027	-0.0027	0.0028
σ_k	0.0104	-2.39e-05	0.0005	-9.21e-06	0.0013	-0.0019	0.0020	-0.0022	0.0022
σ_{υ_r}	0.0050	0.0018	0.0043	0.0101	0.0108	0.0101	0.0109	0.0104	0.0111

Note: The table reports statistics for estimates of $\boldsymbol{\theta}_{\text{exo}}$ from $M = 10,000$ samples of quarterly consumption, C, hours worked, N, and interest rates, r, covering 60 years ($T = 240$). Panel A uses C and r for estimation. Panel B uses C, N, and r, allowing for iid measurement error υ_r in r. The parameters in $\boldsymbol{\theta}_{\text{ss}}$ are fixed at the values in Table 1. With $\hat{\boldsymbol{\theta}}_{\text{exo},m}$ the estimates from the m^{th} sample, Bias $= M^{-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\theta}}_{\text{exo},m} - \boldsymbol{\theta}_{\text{exo},0})$, and RMSE $= (M^{-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\theta}}_{\text{exo},m} - \boldsymbol{\theta}_{\text{exo},0})^2)^{1/2}$.

Overall, the MX-SSR method outperforms the misspecified alternatives (F-SSR, S-SSR, and EM-SSR) in terms of bias and RMSE. From Panel A, the MX-SSR biases for bivariate data without measurement error are small and in line with those for pure stock or flow data, using the correct representation in each case. In contrast, the misspecified alternatives exhibit considerable biases, which in turn dominate RMSEs. The bias in ρ_z fluctuates between 60% and 150%, while those in σ_z and σ_k increase by up to 18%, in either direction. The deteriorating performance is attributed to misclassification of the sampling nature of the real interest rate in the F-SSR case, and consumption in the S-SSR case.

Panel B reports results using three observables, and allowing for measurement error, υ_r, in the real interest rate. Again, MX-SSR outperforms the alternatives. Bias and RMSE in the additional parameter σ_{υ_r} are small for MX-SSR, and an order of magnitude larger for the other methods. Relative to Panel A, the additional information from including N_τ data, while adding one parameter for estimation, σ_{υ_r}, reduces bias and RMSE for all structural parameters in the MX-SSR and F-SSR cases, but only for ρ_z in the S-SSR and EM-SSR cases, with the negative discretization bias resurging in the latter. Overall, our results show

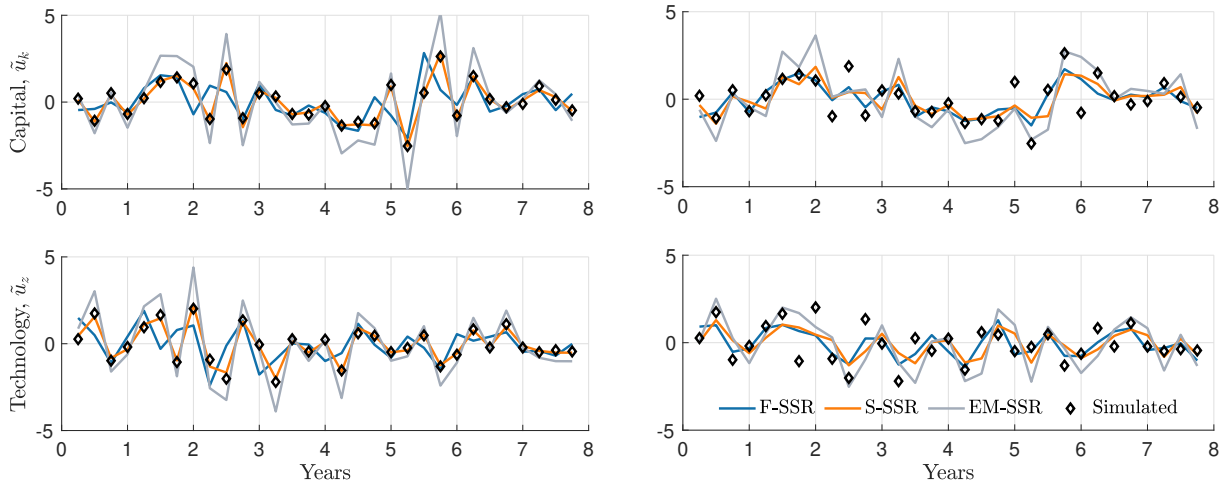
that the benefits of correct state space representation extend to the mixed stock-flow case.

5.3 Structural shocks

Next, we evaluate the accuracy of the approach in Section 4 to recover the structural shocks at measurement times, using (4.9). To control for estimation uncertainty and separate effects, we refrain from using the ML estimates, and instead compute our approximations conditional on the true DGP, fixing all parameters at the values in Table 1. We employ the same simulated data sets on consumption and hours worked as in the Monte Carlo exercises. In (4.9), the matrix $\mathbf{H}(\boldsymbol{\theta})$ is defined as $h^{1/2}\mathbf{H}(\boldsymbol{\theta}; h)$, with $\mathbf{H}(\boldsymbol{\theta}; h)$ given by (4.3) for the S-SSR and F-SSR methods, and simply by $\mathbf{B}(\boldsymbol{\theta})$ for the EM-SSR.

To gain some intuition on the mechanics of our approach, Figure 3 displays the structural shocks to the capital stock (top exhibits) and TFP (bottom exhibits) for the first eight years of a given simulated sample. Left exhibits (a) show the approximated shocks, $\tilde{u}_{k,\tau}$ and $\tilde{u}_{z,\tau}$, recovered by each of the three methods, and the true underlying structural shocks, $u_{k,\tau}$ and $u_{z,\tau}$, denoted with a $\Diamond$, for the case of stocks. Right exhibits (b) present the similar for flows. For the particular sample used, the method performs remarkably well in the case of stock data, conditional on correct state space representation, S-SSR. The somewhat poorer performance of F-SSR and EM-SSR is driven by the misspecification. While the time aggregation imposed by F-SSR alters the weighting scheme of the structural shocks over any given time interval, the EM-SSR disregards all model-based information between measurements when computing the covariance matrix. For flow data, EM-SSR produces the poorest among the three approximations, with the other two about equally accurate.

To better understand the quality of the approximations, we repeat the exercise on all $M = 10,000$ samples. The results are summarized in Figure 4, where we plot the mean squared error (MSE) and its dispersion across replications. Several conclusions emerge. First, the ED-SSR (S-SSR for stocks, F-SSR for flows) always outperforms the other two methods, both in terms of average and variability of MSE, with EM-SSR exhibiting poorest performance in all cases. Second, for each shock (capital and TFP), MSE is larger for flows than for stocks, except in the F-SSR case. Third, for stocks, MSE for shocks recovered by the S-SSR method is nearly zero. In contrast, for flows, F-SSR does not produce near-zero MSE. To understand this, note that (4.6) for the RBC model from Section 5.1 is a square system (2×2) in the stock case, and rectangular (4×2, cf. the ED-SSR in Proposition 1) for flows. With more equations than unknowns in the latter case, some are inevitably solved with error.



(a) Data sampled as stocks

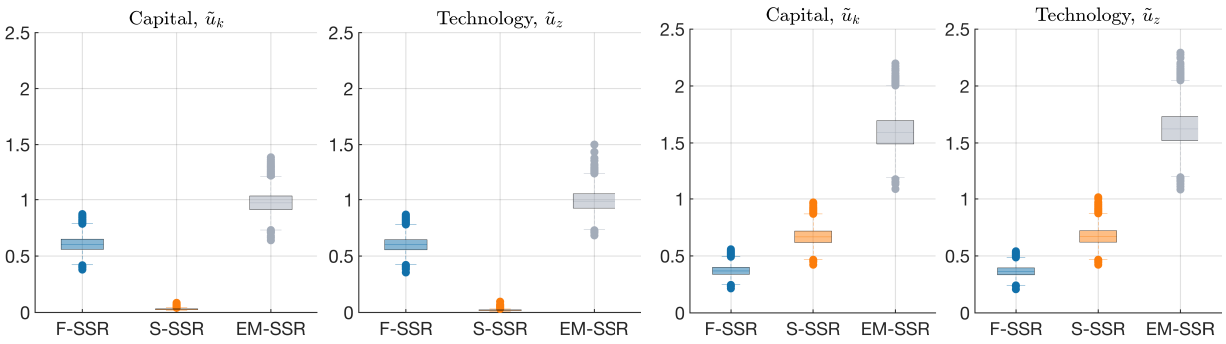
(b) Data sampled as flows

Figure 3. Structural shocks. The figure shows the time series of structural shocks to TFP, $\tilde{u}_z$, and the capital stock, $\tilde{u}_k$, recovered from a simulated sample of consumption, C, and hours worked, N. Simulated measurements are sampled as stocks in Exhibit (a), and as flows in Exhibit (b).

6. Empirical illustration

We provide an empirical illustration of the ML estimation method of Section 2, and of the structural shock recovery of Section 4, using U.S. data. We estimate the RBC model of Section 5 using quarterly data on aggregate consumption and the fraction of hours worked over the period from 1960:Q1 through 2019:Q4. Both variables are obtained from the Federal Reserve Economic Data (FRED) database. Aggregate consumption is measured by monthly nominal personal consumption expenditures (PCE), deflated by the corresponding monthly price index (PCEPI). Consumption at quarterly frequency is computed by aggregating monthly real expenditures over the quarter. Quarterly hours are those worked by wage and salary workers on nonfarm payrolls (TOTLQ). All variables are transformed to per-capita values using the civilian non-institutional population aged 16 and over (CNP16OV) from the U.S. Bureau of Labor Statistics. With the exception of population, all variables are seasonally adjusted. We assume that the observed variables are trending exponentially at a constant growth rate of 2% per year, reflecting long-run economic growth.

In a second exercise, we use the fitted residuals (smoothed reduced-form disturbances, $\boldsymbol{\eta}_\tau$) to recover the structural shocks at measurement times, $\mathbf{u}_\tau$, and from these build a historical shock decomposition of the observed variables. This exercise is usually employed in economic policy circles to build narratives around the sources of past economic fluctuations. Other potential applications include impulse-response analysis, and forecast error variance decompositions. Third, we conduct ML estimation based on a state space representation



(a) Stock variables **(b)** Flow variables

Figure 4. Approximation errors in recovered structural shocks. The figure shows distributions of mean squared error (MSE) between the true (simulated) structural shocks and their recovered (smoothed) counterparts. Each boxplot represents the distribution of MSE across replications. Results are shown for data sampled as stocks in Exhibit (a), and as flows in Exhibit (b).

that accommodates observations at mixed frequencies, using monthly (rather than quarterly) data on aggregate real consumption, along with the quarterly data on hours worked. Finally, we expand the data set with quarterly observations on the real interest rate, measured by the nominal 3-month Treasury bill rate (TB3MS) from FRED, and deflated using the PCEPI. Observations are end-of-quarter values, so we regard them as stock measurements. Using this additional variable, we conduct ML estimation for the case of a mixture of stock and flow measurements. All data series are shown in Online Appendix J.

6.1 ML estimation

As in the Monte Carlo experiments in Section 5, we report ML estimates of $\boldsymbol{\theta}_{\text{exo}}$, while fixing $\boldsymbol{\theta}_{\text{ss}}$ at the values in Table 1. We estimate the model with and without allowance for *iid* measurement errors, using both the F-SSR and S-SSR based methods, as well as the EM-SSR, hence allowing us to remain agnostic about the way in which data are sampled.

Table 4 reports empirical results based on quarterly data on aggregate consumption and hours worked. Estimated standard errors (in parentheses) are computed using the wild bootstrap algorithm, with $B = 1,000$ samples (see Angelini et al., 2021).[28] The bootstrap is chosen over asymptotic formulas to better accommodate the finite-sample nature of the data, particularly heteroskedasticity of the prediction errors, and avoids computation of numerical derivatives at estimates near the boundary of the domain. At each reported parameter vector, nonsingularity, minimality, and local identification (Assumptions 2 and 4, and Proposition 2) are verified as in Remark 5.1, using the accompanying identification toolbox.

[28] Assumptions 1–3, A4′, and 5 in Angelini et al. (2021) are verified prior to implementation.

Table 4. ML estimates (common frequency).

	Without measurement error			With measurement error		
$\boldsymbol{\theta}_{\text{exo}}$	F-SSR	S-SSR	EM-SSR	F-SSR	S-SSR	EM-SSR
ρ_z	0.0549	0.0646	0.0584	0.0633	0.0651	0.0591
	(0.0220)	(0.0251)	(0.0229)	(0.0324)	(0.0362)	(0.0351)
σ_z	0.0101	0.0093	0.0093	0.0111	0.0101	0.0101
	(0.0010)	(0.0009)	(0.0009)	(0.0013)	(0.0012)	(0.0012)
σ_k	0.0142	0.0132	0.0132	0.0163	0.0152	0.0151
	(0.0013)	(0.0011)	(0.0011)	(0.0017)	(0.0017)	(0.0016)

Note: The table reports ML estimates of $\boldsymbol{\theta}_{\text{exo}}$ using quarterly data on aggregate consumption and hours worked for the U.S. from 1960:Q1 through 2019:Q4, with $\boldsymbol{\theta}_{\text{ss}}$ fixed at the values in Table 1. Wild bootstrap standard errors computed from $B = 1,000$ samples in parentheses.

From the results, the parameters representing the driving forces behind the RBC model are similar across specifications, and quite reasonable. In the absence of measurement error, mean reversion of TFP is estimated at $\rho_z = 0.0549$ by the F-SSR method, 0.0646 by S-SSR, and 0.0584 by EM-SSR. These estimates imply a first order autocorrelation coefficient for quarterly TFP, $\exp(-\rho_z/4)$, between 0.9840 and 0.9864. Our results are close to the quarterly estimates reported in Ireland (2004) and Malley and Woitek (2010) for the U.S. economy, and thus consistent with arguments by King et al. (1988) and Hansen (1997) that TFP shocks should be highly persistent for the RBC model to match key features of U.S. data.[29] Further, the point estimates of the instantaneous volatility of TFP shocks, σ_z, imply a quarterly conditional volatility, $\sigma_z\sqrt{(1-\exp(-2\rho_z/4))/2\rho_z}$, between 0.46% and 0.50%, again close to values reported by the same authors, and of the same order of magnitude as those commonly used in the RBC literature. Moreover, our estimates suggest that the variability of shocks to the capital stock is greater than that in Ambler and Paquet (1994), and greater than for TFP shocks.[30] In particular, the estimates of σ_k imply a volatility per quarter ($\approx \sqrt{(1/4)}\sigma_k$) between 0.66% and 0.81%. Comparing across methods, F-SSR produces higher estimates of σ_z and σ_k than S-SSR and EM-SSR, in line with the Monte Carlo evidence (positive bias for F-SSR versus negligible bias for the other methods in Panel A of Table 2 and Figure 2, negligible versus negative bias in Panel B.).[31]

To assess the effects of *iid* measurement errors in the ML estimation, we fix their stan-

[29] The persistence of TFP in Table 1 is based on quarterly autoregressions for the Solow residual, with coefficients estimated at around 0.95, whereas we use state space methods to estimate ρ_z directly in Table 4.

[30] In contrast, for the parameter values in Table 1, σ_z exceeds σ_k.

[31] In Online Appendix I, we report the results from a battery of tests assessing the univariate and multivariate normality of the one-step-ahead prediction errors for the measurements in the F-SSR approach across different subsets of the data used in the estimation. The results are generally consistent with Gaussianity.

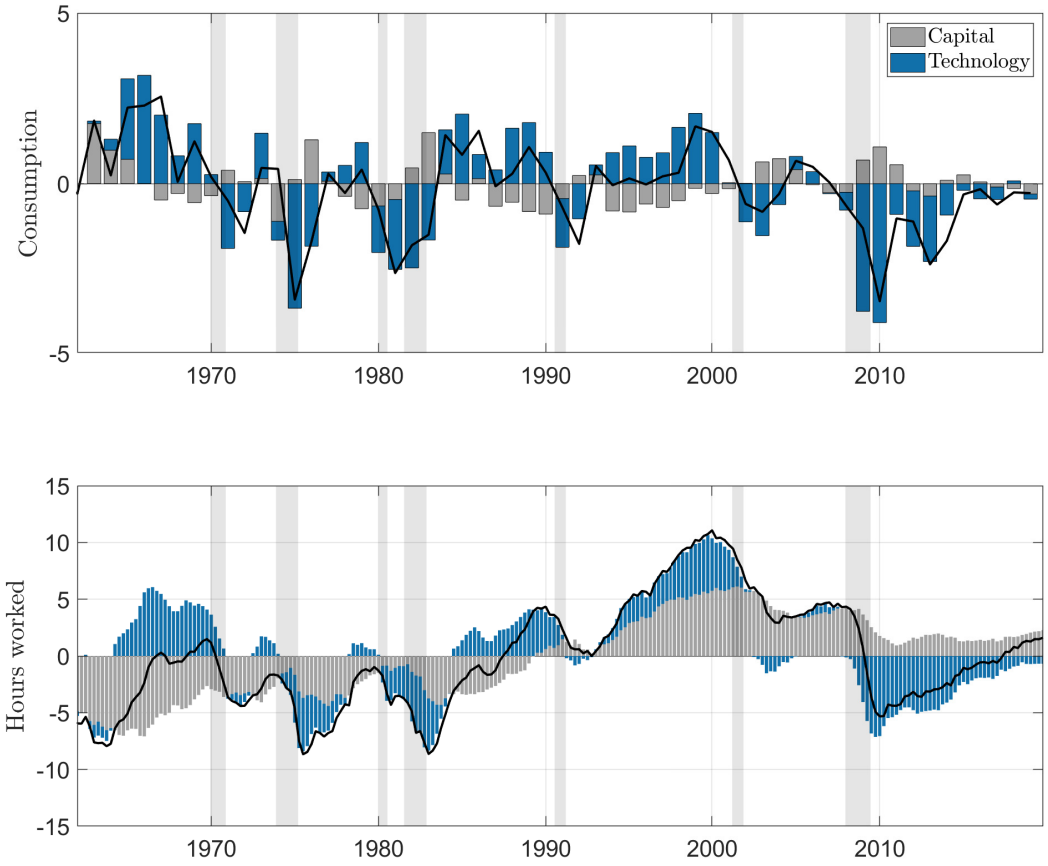


Figure 5. Historical shock decomposition (F-SSR method). The figure shows the separate historical contributions of the structural shocks to capital (demand) and technology (supply, TFP) recovered by the F-SSR method from the observed measurements of consumption and hours worked over the period from 1960:Q2 through 2019:Q4. The black solid line in the upper panel represents annual consumption growth rates, and that in the lower panel quarterly percentage deviations of hours worked from its steady state ($n^\star = 33\%$). The light gray vertical bands indicate NBER recessions.

dard deviations to the quarterly estimates reported in Ireland (2004, Table 6), namely, $\sigma_{v_C} = 0.0061$ for aggregate consumption, and $\sigma_{v_N} = 0.0073$ for hours.[32] Our results (right side of Table 4) suggest that allowing for measurement error does not have sizable effects on the structural parameter estimates. Differences are small, relative to standard errors.

[32]This corresponds to the assumption on $\mathbf{\Sigma}_{\boldsymbol{v},h}$ in Proposition 2.(c).

6.2 Historical shock decomposition

Given the ML estimates, we employ the approach of Section 4 to recover the structural shocks at measurement times. We then investigate their contribution to the observed series through the lens of the continuous-time RBC model, based on the historical shock decomposition

$$\begin{aligned}\text{HSD}(\,\mathbf{y}_\tau - \mathsf{C}(\boldsymbol{\theta};h)\mathsf{A}(\boldsymbol{\theta};h)^{\tau-1}\mathbf{x}_0\,,\,i\,) &= \mathsf{D}(\boldsymbol{\theta};h)\mathbf{H}(\boldsymbol{\theta})\tilde{\mathbf{u}}_{i,\tau} \\ &\quad + \mathsf{C}(\boldsymbol{\theta};h)\sum_{j=0}^{\tau-2}\mathsf{A}(\boldsymbol{\theta};h)^j\mathsf{B}(\boldsymbol{\theta};h)\mathbf{H}(\boldsymbol{\theta})\tilde{\mathbf{u}}_{i,\tau-1-j}, \quad (6.1)\end{aligned}$$

computed from the ABCD representation (2.12)-(2.13) associated with the F-SSR.[33] Here, $i = k, z$ indicate the components attributable to shocks to the capital stock and TFP, respectively, of the evolution of observables around the conditional trend, given the initial state vector $\mathbf{x}_0$. The derivation of (6.1) is provided in Online Appendix K.

The top panel of Figure 5 displays the contributions of $\tilde{u}_k$ and $\tilde{u}_z$ to annual consumption growth, and the bottom panel their contributions to quarterly deviations of hours worked from their steady state value. Any discrepancy between the sum of the bars and the observed series (solid line) can be attributed to approximation errors in backing out the structural shocks (4.9). For aggregate consumption, when reporting the historical decomposition of the annual growth rates, although the model is estimated using quarterly data, the shift in frequency is achieved seamlessly, due to the frequency invariance of the structural parameters in the continuous-time model. For parameters estimated at a given frequency, it is straightforward to recalibrate the ABCD representation to any frequency of interest, in this case annual.

If shocks to TFP are interpreted as "aggregate supply shocks," and shocks to the capital stock as "aggregate demand shocks," the decomposition in Figure 5 suggests that the U.S. business cycle over the period from 1960:Q2 through 2019:Q4 mainly has been driven by aggregate supply shocks to firm productivity, both during expansions and contractions. In contrast, the short-run variability of hours worked, relative to their long-run value, has mostly reflected aggregate demand shocks. This corroborates earlier evidence on the limited power of TFP shocks to explain the behavior of hours worked in RBC models, in particular their unconditional variance (see, e.g., Cooley and Prescott, 1995, and Ireland, 2004).

6.3 Mixed-frequency estimation

Here, we estimate the same parameters as in Section 6.1, now using monthly (rather than quarterly) frequency data on aggregate consumption, along with the quarterly observations

[33] Historical decompositions using S-SSR and EM-SSR are available in Online Appendix K.

Table 5. ML estimates (mixed frequencies).

	Without measurement error			With measurement error		
$\boldsymbol{\theta}_{\text{exo}}$	F-SSR	S-SSR	EM-SSR	F-SSR	S-SSR	EM-SSR
ρ_z	0.0559	0.0772	0.0749	0.0766	0.0942	0.0917
	(0.0234)	(0.0310)	(0.0294)	(0.0534)	(0.0610)	(0.0590)
σ_z	0.0108	0.0089	0.0089	0.0104	0.0097	0.0097
	(0.0008)	(0.0006)	(0.0006)	(0.0009)	(0.0009)	(0.0009)
σ_k	0.0156	0.0134	0.0134	0.0149	0.0143	0.0142
	(0.0010)	(0.0009)	(0.0009)	(0.0010)	(0.0011)	(0.0011)

Note: The table reports ML estimates of $\boldsymbol{\theta}_{\text{exo}}$ using monthly data on aggregate consumption and quarterly data on hours worked for the U.S. from 1960 through 2019, with $\boldsymbol{\theta}_{\text{ss}}$ fixed at the values in Table 1. Wild bootstrap standard errors computed from $B = 1,000$ samples in parentheses.

on hours worked. In this mixed-frequency case, the state space representations are adjusted accordingly. For observations sampled as stocks, we simply write the S-SSR and EM-SSR representations in terms of the variable(s) observed at the highest frequency, then modify the filtering algorithm to accommodate the corresponding missing observations on variable(s) observed at lower frequencies (see Durbin and Koopman, 2012, Sec. 4.10). A similar approach is adopted for data sampled as flows. In this case, we extend the state vector of the F-SSR with additional deterministic variables that allow keeping track of the time aggregation at high frequencies of the variable(s) sampled at lower frequencies.[34]

The mixed-frequency ML results for $\boldsymbol{\theta}_{\text{exo}}$ in Table 5 closely mirror those for common quarterly frequency in Table 4, especially when possible measurement error is not accounted for. When it is, point estimates of ρ_z increase by about one half in the mixed-frequency case, and standard errors by two-thirds, with the remaining parameters relatively unaffected.

To explore the frequency-invariance property of the parameter estimates based on the continuous-time model, Panel A of Table 6 reproduces the estimates obtained by the F-SSR method using common and mixed-frequency data, along with t-statistics for the null hypothesis that the parameters coincide, with standard errors of the differences (the denominators) calculated as if the estimates were independent. As they are not, we compute the lower bound on the correlation between estimates, ϱ, for the null to be rejected at level 10%. Rather high correlations would be required to reject, so the results suggest that our approach produces estimates that are statistically identical, regardless of the data frequency used. For ρ_z, the dash in the last column indicates that we fail to reject the null irrespective of the correlation between estimates, i.e., for all $\varrho \in [-1, 1]$. In particular, the smallest possible standard error of the difference, at $\varrho = 1$, yields a t-statistic of -0.7143.

[34]The details of the derivations of the mixed-frequency representations are in Online Appendix H.

Table 6. Frequency invariance of continuous time estimates.

Panel A: ML estimates from Tables 4 and 5				
$\boldsymbol{\theta}_{\text{exo}}$	M.F. F-SSR	Q. F-SSR	t-stat	ϱ
ρ_z	0.0559	0.0549	-0.0318	-
	(0.0234)	(0.0220)		
σ_z	0.0108	0.0101	-0.5799	0.9096
	(0.0008)	(0.0010)		
σ_k	0.0156	0.0142	-0.8711	0.7335
	(0.0010)	(0.0013)		
Panel B: Smoothed TFP using estimates from Table 5				
Dataset	β_h	ρ_z	$\varrho_{M=Q}$	$\varrho_{\bullet=Y}$
Monthly TFP	0.9966	0.0410	0.9966	0.7423
	(0.0032)	(0.0386)		
Quarterly TFP	0.9878	0.0493		0.8064
	(0.0105)	(0.0423)		
Annual TFP	0.8862	0.1208		
	(0.0617)	(0.0697)		

Note: Panel A reports ML estimates (with bootstrap standard errors) of $\boldsymbol{\theta}_{\text{exo}}$ using the F-SSR method, along with t-statistics comparing common-frequency and mixed-frequency estimates. Panel B shows OLS estimates of an AR(1) process for TFP at different frequencies, and the associated continuous-time mean reversion. Standard errors in parentheses. The variable ϱ is the minimum correlation between estimates required to reject the null of equality at level 10%.

We construct smoothed time series for latent capital and TFP at the monthly ($h = 1/12$), quarterly ($h = 1/4$), and annual ($h = 1$) frequencies. Quarterly latent series using quarterly consumption and hours as measurements are shown in the Online Appendix, along with monthly latent series based on monthly consumption and quarterly hours. In general, the resulting series are very similar, whether using common or mixed-frequency data and state space representations, hence reinforcing the value of the mixed-frequency approach.

Estimates from a standard discrete-time model could, if desired, be transformed to the underlying frequency-invariant continuous-time model parameters, but results might not agree with those obtained by estimating the continuous-time model directly, as in our case. To illustrate, we estimate the AR(1) process $\hat{z}_{t+1} = \beta_h \hat{z}_t + e_t$ by Ordinary Least Squares (OLS), for each of the constructed TFP series. Using $\hat{\beta}_h$ for each frequency h, we recover the underlying continuous-time parameter as $\hat{\rho}_z = -\log(\hat{\beta}_h)/h$. In Panel B of Table 6, we report the OLS estimates $\hat{\beta}_h$, with standard errors in parentheses, and the resulting estimates $\hat{\rho}_z$, with standard errors computed using the delta method. The last two columns report the minimum pairwise correlation between estimates of ρ_z from samples of different frequencies

Table 7. ML estimates (mixed stock-flow sampling).

	Measurement error in r				Measurement error in C, N, and r			
$\boldsymbol{\theta}_{\text{exo}}$	MX-SSR	F-SSR	S-SSR	EM-SSR	MX-SSR	F-SSR	S-SSR	EM-SSR
ρ_z	0.2192	0.1990	0.1896	0.1794	0.2720	0.2480	0.2203	0.2080
	(0.0431)	(0.0402)	(0.0475)	(0.0395)	(0.0742)	(0.0741)	(0.0799)	(0.0724)
σ_z	0.0093	0.0092	0.0089	0.0087	0.0102	0.0101	0.0095	0.0092
	(0.0009)	(0.0009)	(0.0008)	(0.0008)	(0.0010)	(0.0011)	(0.0010)	(0.0010)
σ_k	0.0153	0.0151	0.0138	0.0138	0.0168	0.0167	0.0153	0.0153
	(0.0012)	(0.0012)	(0.0011)	(0.0011)	(0.0017)	(0.0018)	(0.0017)	(0.0018)
σ_{υ_r}	0.1157	0.1280	0.1294	0.1311	0.1045	0.1168	0.1199	0.1216
	(0.0075)	(0.0088)	(0.0086)	(0.0089)	(0.0074)	(0.0082)	(0.0083)	(0.0089)

Note: The table reports ML estimates of $\boldsymbol{\theta}_{\text{exo}}$ and σ_{υ_r} using quarterly data on aggregate consumption, C, hours worked, N, and the real interest rate, r, for the U.S. from 1960:Q1 through 2019:Q4. The remaining parameters, $\boldsymbol{\theta}_{\text{ss}}$, are fixed at the values in Table 1, and the standard deviations of the measurement errors on C and N are calibrated to $\sigma_{\upsilon_C} = 0.0061$ and $\sigma_{\upsilon_N} = 0.0073$. F-SSR, S-SSR, and MX-SSR refer to the state space representations assuming measurements sampled as flows, stocks, and a mixture of flows and stocks, respectively, and EM-SSR to the Euler-Maruyama discretization. Wild bootstrap standard errors computed from $B = 1,000$ samples in parentheses.

(monthly vs. quarterly, monthly vs. annual, and quarterly vs. annual) needed to reject the null of equality at the 10% level. Researchers estimating monthly, quarterly, and annual frequency discrete-time models would back out continuous-time ρ_z values ranging between 0.04 and 0.12. Estimates are likely strongly correlated, and if the correlation exceeds 80%, both monthly and quarterly frequency analyses lead to results that are statistically different from annual frequency analysis. This is in contrast to the conclusion from Panel A, that direct estimation of the continuous-time model is robust to the use of different data frequencies.

Overall, we conclude that the frequency-invariance of model parameters and the robustness to differences in (possibly mixed) observation frequencies in data sets available in applications are desirable features of our approach.

6.4 Mixed stock-flow sampling estimation

Finally, we include quarterly stock measurements on the (gross) real interest rate, r. To avoid stochastic singularity, we allow for *iid* measurement errors. ML results are presented in Table 7. The left panel reports results allowing for measurement error only in r, with standard deviation σ_{υ_r} estimated along with the structural parameters, $\boldsymbol{\theta}_{\text{exo}}$. The right panel reports results allowing also for measurement errors in quarterly consumption and hours worked, with standard deviations fixed at the same values as in Table 4.

Including information on the real interest rate in the MX-SSR framework, treating r as

a stock and C, N as flows, pushes the estimate of the speed of mean reversion of TFP up, from around 0.06 in Table 4 to 0.22 in Table 7, allowing for measurement error in r, and 0.27 when allowing for measurement error in all three observables. The larger estimate of ρ_z is consistent across state space representations, whereas σ_z and σ_k remain close to earlier estimates. Finally, the large and significant estimates of σ_{υ_r} indicate the existence of a wedge between the model's implied rental rate of capital and the measured 3-month T-bill yield.

7. Conclusions

Our exact discrete state space representation in conjunction with the likelihood function based on the Kalman filter allows assessing whether the parameters of the underlying DSGE model cast in continuous time are locally identified. If they are, structural shocks at measurement times can be backed out from the reduced-form residuals. The Monte Carlo and empirical analyses indicate that our approach provides accurate parameter estimates and predicted structural shocks, based on measurements that can include a combination of stocks and flows, possibly observed at mixed frequencies, and that performance deteriorates if stocks are misclassified as flows, or vice versa. Researchers using instead a DSGE model directly formulated in discrete time could transform the received estimates to parameters corresponding to frequencies other than that of the observations, including infinite frequency (continuous-time) parameters, but the result for any frequency of interest would depend on the actual data frequency. In contrast, the result of mapping our estimates to the particular frequency of interest is robust to different data frequencies—including mixed frequencies. Further, our approach outperforms that relying on the naive Euler-Maruyama discretization. This is particularly relevant in macroeconomic applications, with data at relatively low frequencies (e.g., monthly, quarterly, annual).

While our approach works well for both the exact nonlinear and the corresponding linearized DGP, comparison with nonlinear filtering would be of interest in the former case. So would mixtures of observed and latent state and control variables, as well as the analysis of heterogeneous agent models. We leave these as exciting avenues for future research.

Acknowledgments

We are grateful to Serena Ng (the Editor), the Associate Editor, the anonymous referees, Martin Møller Andreasen, Tim Bollerslev, Federico Carlini, Leopoldo Catania, Manfred Deistler, Francis X. Diebold, Luca Fanelli, Dennis Kristensen, Alessia Paccagnini, Giovanni Pellegrino, Olaf Posch, Enrique Sentana, Peter Zadrozny, and participants at the

2022 NBER-NSF Times Series Conference (Boston), the 11th Nordic Econometric Meeting (Sandbjerg Manor, Denmark), the XII Workshop in Time Series Econometrics (Zaragoza), the Arne Ryde Workshop: Heterogeneous Agent Models in Macroeconomics - Advances in Continuous Time (Lund University), the Nordic Junior Macro Seminar, and the Quantitative Economics seminar (Hamburg University) for valuable comments, and to the Danish Social Science Research Council (grant number 2033-00137B) for research support. Luca Neri acknowledges financial support from the Carlo Giannini Association. Juan Carlos Parra-Alvarez acknowledges financial support from the Otto Mønsted Foundation.

References

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2022): "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," *The Review of Economic Studies*, 89, 45–86.

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2018): "When Inequality Matters for Macro and Macro Matters for Inequality," *NBER Macroeconomics Annual*, 32, 1–75.

Aït-Sahalia, Y. (2002): "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach," *Econometrica*, 70, 223–262.

Ambler, S. and A. Paquet (1994): "Stochastic Depreciation and the Business Cycle," *International Economic Review*, 35, 101–116.

Anderson, B. D. O. and J. B. Moore (1979): *Optimal Filtering*, Englewood Cliffs, NJ: Prentice Hall.

Angelini, G., G. Cavaliere, and L. Fanelli (2021): "Bootstrap Inference and Diagnostics in State Space Models: With Applications to Dynamic Macro Models," *Journal of Applied Econometrics*, 1–20.

Bergstrom, A. R. (1966): "Nonrecursive Models as Discrete Approximations to Systems of Stochastic Differential Equations," *Econometrica*, 34, 173–182.

——— (1983): "Gaussian Estimation of Structural Parameters in Higher Order Continuous Time Dynamic Models," *Econometrica*, 51, 117–152.

——— (1984): "Continuous Time Stochastic Models and Issues of Aggregation over Time," in *Handbook of Econometrics*, Elsevier, vol. 2, 1145–1212.

Bernanke, B. S. (1986): "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49–99.

Bibby, B. M. and M. Sørensen (1995): "Martingale Estimation Functions for Discretely Observed Diffusion Processes," *Bernoulli*, 1, 17–39.

Blanchard, O. J. and C. M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48, 1305–1311.

Blanchard, O. J. and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Economic Review*, 79, 655–673.

Blevins, J. R. (2017): "Identifying Restrictions for Finite Parameter Continuous Time Models with Discrete Time Data," *Econometric Theory*, 33, 739–754.

Brunnermeier, M. and Y. Sannikov (2014): "A Macroeconomic Model with a Financial Sector," *American Economic Review*, 104, 379–421.

Chambers, M. J. (1999): "Discrete Time Representation of Stationary and Non-Stationary Continuous Time Systems," *Journal of Economic Dynamics and Control*, 23, 619–639.

Chambers, M. J., J. R. McCrorie, and M. A. Thornton (2018): "Continuous Time Modelling Based on an Exact Discrete Time Representation," in *Continuous Time Modeling in the Behavioral and Related Sciences*, ed. by K. van Montfort, J. H. L. Oud, and M. C. Voelkle, Springer International Publishing, 317–357.

Chambers, M. J., T. Simos, and M. Tsionas (2023): "Locally Exact Discrete Time Representations of Non-Linear Continuous Time Models," Tech. rep., mimeo.

Christensen, B. J., O. Posch, and M. Van Der Wel (2016): "Estimating Dynamic Equilibrium Models using Mixed Frequency Macro and Financial Data," *Journal of Econometrics*, 194, 116–137.

Cooley, T. F. and E. C. Prescott (1995): "Economic Growth and Business Cycles," in *Frontiers of Business Cycle Research*, ed. by T. F. Cooley, Princeton, NJ: Princeton University Press, chap. 1, 1–38.

Durbin, J. and S. J. Koopman (2012): *Time Series Analysis by State Space Methods*, Oxford University Press.

Fernández-Villaverde, J., S. Hurtado, and G. Nuño (2023): "Financial Frictions and the Wealth Distribution," *Econometrica*, 91, 869–901.

Florens-Zmirou, D. (1993): "On Estimating the Diffusion Coefficient from Discrete Observations," *Journal of Applied Probability*, 30, 790–804.

Hansen, G. D. (1985): "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, 16, 309–327.

——— (1997): "Technical Progress and Aggregate Fluctuations," *Journal of Economic Dynamics and Control*, 21, 1005–1023.

Hansen, L. P. and J. A. Scheinkman (1995): "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767–804.

Harvey, A. C. (1990): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, A. C. and J. H. Stock (1985): "The Estimation of Higher-Order Continuous Time Autoregressive Models," *Econometric Theory*, 1, 97–117.

Ireland, P. N. (2004): "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control*, 28, 1205–1226.

Iskrev, N. (2010): "Local Identification in DSGE Models," *Journal of Monetary Economics*, 57, 189–202.

Jiang, G. J. and J. L. Knight (1997): "A Nonparametric Approach to the Estimation of Diffusion Processes, With an Application to a Short-Term Interest Rate Model," *Econometric Theory*, 13, 615–645.

Jones, R. H. (1981): "Fitting a Continuous Time Autoregression to Discrete Data," in *Applied Time Series Analysis II*, ed. by D. Findlay, New York: Academic Press, 651–682.

Kaplan, G., B. Moll, and G. L. Violante (2018): "Monetary Policy According to HANK," *American Economic Review*, 108, 697–743.

King, R. G., C. I. Plosser, and S. T. Rebelo (1988): "Production, Growth and Business Cycles : I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21, 195–232.

Kociecki, A. and M. Kolasa (2018): "Global Identification of Linearized DSGE Models," *Quantitative Economics*, 9, 1243–1263.

——— (2023): "A Solution to the Global Identification Problem in DSGE Models," *Journal of Econometrics*, 236.

Komunjer, I. and S. Ng (2011): "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," *Econometrica*, 79, 1995–2032.

Liemen, M. O. and O. Posch (2022): "FTPL and the Maturity Structure of Government Debt in the New Keynesian Model," CESifo Working Paper Series 9840, CESifo.

Lo, A. W. (1988): "Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data," *Econometric Theory*, 4, 231–247.

Malley, J. and U. Woitek (2010): "Technology Shocks and Aggregate Fluctuations in an Estimated Hybrid RBC Model," *Journal of Economic Dynamics and Control*, 34, 1214–1232.

McCrorie, J. R. (2003): "The Problem of Aliasing in Identifying Finite Parameter Continuous Time Stochastic Models," *Acta Applicandae Mathematicae*, 79, 9–16.

——— (2009): "Estimating Continuous-Time Models On The Basis Of Discrete Data Via An Exact Discrete Analog," *Econometric Theory*, 25, 1120–1137.

Merton, R. C. (1980): "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

Nowman, K. B. (1997): "Gaussian Estimation of Single-Factor Continuous Time Models of the Term Structure of Interest Rates," *Journal of Finance*, 52, 1695–1706.

Parra-Alvarez, J. C. (2018): "A Comparison of Numerical Methods for the Solution of Continuous-time DSGE Models," *Macroeconomic Dynamics*, 22, 1555 – 1583.

Parra-Alvarez, J. C., H. Polattimur, and O. Posch (2021): "Risk Matters: Breaking Certainty Equivalence in Linear Models," *Journal of Economics Dynamics and Control*, 133.

Phillips, A. W. (1959): "The Estimation of Parameters in Systems of Stochastic Differential Equations," *Biometrika*, 46, 67–76.

Phillips, P. C. B. (1973): "The Problem of Identification in Finite Parameter Continuous Time Models," *Journal of Econometrics*, 1, 351–362.

Phillips, P. C. B. and J. Yu (2005): "Jackknifing Bond Option Prices," *Review of Financial Studies*, 18, 707–742.

——— (2009): "Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance," in *Handbook of financial time series*, Springer, Handbook of Financial Time Series, 497–530.

Posch, O. (2009): "Structural Estimation of Jump-Diffusion Processes in Macroeconomics," *Journal of Econometrics*, 153, 196–210.

——— (2020): "Resurrecting the New-Keynesian Model: (Un)conventional Policy and the Taylor rule," CESifo Working Paper Series 6925, CESifo.

Posch, O. and T. Trimborn (2013): "Numerical Solution of Dynamic Equilibrium Models under Poisson Uncertainty," *Journal of Economic Dynamics and Control*, 37, 2602–2622.

Qu, Z. and D. Tkachenko (2012): "Identification and Frequency Domain Quasi-Maximum Likelihood Estimation of Linearized Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 3, 95–132.

——— (2017): "Global Identification in DSGE Models Allowing for Indeterminacy," *The Review of Economic Studies*, 84, 1306–1345.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, pp. 577–591.

Shapiro, M. D. and M. W. Watson (1988): "Sources of Business Cycle Fluctuations," *NBER Macroeconomics Annual*, 3, 111–148.

Sims, C. A. (1986): "Are Forecasting Models Usable for Policy Analysis?" *Quarterly Review*, 10, 2–16.

——— (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20.

Tang, C. Y. and S. X. Chen (2009): "Parameter Estimation and Bias Correction for Diffusion Processes," *Journal of Econometrics*, 149, 65–81.

Van Loan, C. (1978): "Computing Integrals Involving the Matrix Exponential," *IEEE Transactions on Automatic Control*, 23, 395–404.

Wang, X., P. C. B. Phillips, and J. Yu (2011): "Bias in Estimating Multivariate and Univariate Diffusions," *Journal of Econometrics*, 161, 228–245.

Zadrozny, P. (1988): "Gaussian Likelihood of Continuous-Time ARMAX Models when Data are Stocks and Flows at Different Frequencies," *Econometric Theory*, 4, 108–124.

Estimation of continuous-time linear DSGE models from discrete-time measurements

Online Appendix

Bent Jesper Christensen* Luca Neri† Juan Carlos Parra-Alvarez‡

October 2, 2024

Contents

*Corresponding author. Department of Economics and Business Economics, Aarhus University; CREATES; Danish Finance Institute; Fuglesangs Allé 4, 8210 Aarhus V, Denmark; Email address: bjchristensen@econ.au.dk

†Department of Economic Sciences, University of Bologna; CREATES; CORE, UCLouvain, Belgium; Piazza Scaravilli 2, Italy; Email address: luca.neri@uclouvain.be

‡Department of Economics and Business Economics, Aarhus University; CREATES; Danish Finance Institute; Fuglesangs Allé 4, 8210 Aarhus V, Denmark; Email address: jparra@econ.au.dk

A. Proofs

A.1 Proof of Proposition 1

The proof is constructed in four steps.

Step 1: State equation for stock variables. For $\mathbf{x}(t_0) = \mathbf{x}_0$ given, and $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\theta})$ time-invariant matrices, the solution to the linear SDE in (2.1) is given by

$$\mathbf{x}(t) = \exp(\mathbf{A}(\boldsymbol{\theta})h)\left[\mathbf{x}(t_0) + \int_{t_0}^{t} \exp(\mathbf{A}(\boldsymbol{\theta})(t_0 - u))\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(u)\right],$$

with $h = t - t_0$. By setting $t_0 = t_{\tau-1}$ and $t = t_\tau$, the solution for $\{\mathbf{x}(t_\tau)\}_{\tau\in\mathbb{Z}}$ can be written as

$$\mathbf{x}(t_\tau) \;=\; \mathbf{A}_h(\boldsymbol{\theta})\mathbf{x}(t_{\tau-1}) + \boldsymbol{\eta}(t_\tau)\,, \tag{A.1}$$

where

$$\boldsymbol{\eta}(t_\tau) = \int_{t_{\tau-1}}^{t_\tau} \exp\left(\mathbf{A}(\boldsymbol{\theta})(t_\tau - u)\right)\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(u)\,, \tag{A.2}$$

with $\mathbb{E}[\boldsymbol{\eta}(t_\tau)] = \mathbf{0}$. By the Itô isometry,

$$\mathbb{E}[\boldsymbol{\eta}(t_\tau)\boldsymbol{\eta}(t_\tau)^\top] = \int_{t_{\tau-1}}^{t_\tau} \exp\left(\mathbf{A}(\boldsymbol{\theta})(t_\tau - u)\right)\mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top \exp\left(\mathbf{A}(\boldsymbol{\theta})^\top(t_\tau - u)\right)\mathrm{d}u\,,$$

which is time-invariant, i.e., only depending on $h = t_\tau - t_{\tau-1}$. Therefore, a change of variable shows that

$$\mathbb{E}[\boldsymbol{\eta}(t_\tau)\boldsymbol{\eta}(t_\tau)^\top] = \int_{0}^{h} \exp\left(\mathbf{A}(\boldsymbol{\theta})(h - u)\right)\underbrace{\mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top}_{:=\boldsymbol{\Sigma}(\boldsymbol{\theta})} \exp\left(\mathbf{A}(\boldsymbol{\theta})^\top(h - u)\right)\mathrm{d}u\,,$$

corresponding to (2.9). Note that $\boldsymbol{\eta}(t_\tau)$ is serially uncorrelated, i.e., $\mathbb{E}[\boldsymbol{\eta}(t_\tau)\boldsymbol{\eta}(t_{\tau-\ell})^\top] = \mathbf{0}$, $\forall \ell \neq 0$. Now, for stock variables, let $\mathbf{x}_\tau^s = \mathbf{x}(t_\tau)$ in (A.1), and $\boldsymbol{\eta}_\tau^s = \boldsymbol{\eta}(t_\tau)$ in (A.2). This completes the stock portion of the transition equation (2.4), (2.7).

Step 2: State equation for flow variables. Introduce the time t_τ cumulator variable

$$\mathbf{x}^f(t_\tau) = \int_{t_{\tau-1}}^{t_\tau} \mathbf{x}(u)\mathrm{d}u = \int_{0}^{h} \mathbf{x}(t_{\tau-1} + u)\mathrm{d}u\,,$$

measuring the cumulated values of the state variables over the time interval $[t_{\tau-1}, t_\tau]$, with the dynamics given in (2.1). The realization at time t_τ of the continuous-time state

vector is equivalent to a stock measurement, $\mathbf{x}(t_\tau) = \mathbf{x}^s(t_\tau)$. By (A.1),

$$\begin{aligned}\mathbf{x}^f(t_\tau) &= \left[\int_0^h \exp(\mathbf{A}(\boldsymbol{\theta})u)\mathrm{d}u\right]\mathbf{x}^s(t_{\tau-1}) + \boldsymbol{\eta}^f(t_\tau) \\ &= \mathbf{A}(\boldsymbol{\theta})^{-1}(\mathbf{A}_h(\boldsymbol{\theta}) - \mathbf{I})\mathbf{x}^s(t_{\tau-1}) + \boldsymbol{\eta}^f(t_\tau)\,, \end{aligned} \tag{A.3}$$

where the last equality uses the fact that, under Assumption 1, the matrix $\mathbf{A}(\boldsymbol{\theta})$ is nonsingular to compute the definite integral. In addition,

$$\boldsymbol{\eta}^f(t_\tau) = \int_0^h \int_{t_{\tau-1}}^{t_{\tau-1}+u} \exp(\mathbf{A}(\boldsymbol{\theta})(t_{\tau-1} + u - r))\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(r)\,\mathrm{d}u$$

is a vector of normally distributed reduced-form disturbances, with mean $\mathbb{E}\left[\boldsymbol{\eta}^f(t_\tau)\right] = \mathbf{0}$, and variance-covariance matrix $\mathbb{E}\left[\boldsymbol{\eta}^f(t_\tau)\boldsymbol{\eta}_\tau^f(t_\tau)^\top\right] = \boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta})$. By redefining the bounds of the inner stochastic integral, and exchanging the order of integration, we may write

$$\boldsymbol{\eta}^f(t_\tau) = \int_0^h \left(\mathbf{A}(\boldsymbol{\theta})^{-1}(\exp(\mathbf{A}(\boldsymbol{\theta})(h - r)) - \mathbf{I})\mathbf{B}(\boldsymbol{\theta})\right)\mathrm{d}\mathbf{w}(t_{\tau-1} + r)\,.$$

Setting $u = t_{\tau-1} + r$, and using the fact that $h = t_\tau - t_{\tau-1}$, we conclude that

$$\boldsymbol{\eta}^f(t_\tau) = \int_{t_{\tau-1}}^{t_\tau} \mathbf{A}(\boldsymbol{\theta})^{-1}(\exp(\mathbf{A}(\boldsymbol{\theta})(t_\tau - u)) - \mathbf{I})\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(u)\,. \tag{A.4}$$

Now let $\mathbf{x}_\tau^f = \mathbf{S}_x^f\mathbf{x}^f(t_\tau)$, $\boldsymbol{\eta}_\tau^f = \mathbf{S}_x^f\boldsymbol{\eta}^f(t_\tau)$, and the flow portion of the transition equation (2.4), (2.7) follows.

Step 3: Measurement equation for stock variables. Evaluating (2.2) at measurement times,

$$\mathbf{y}_\tau^s = \mathbf{y}^s\left(t_\tau\right) = \mathbf{S}_y^s\mathbf{C}(\boldsymbol{\theta})\mathbf{x}\left(t_\tau\right) = \mathbf{S}_y^s\mathbf{C}(\boldsymbol{\theta})\mathbf{x}_\tau^s\,. \tag{A.5}$$

Step 4: Measurement equation for flow variables. The cumulator variable is

$$\begin{aligned}\mathbf{y}_\tau^f &= \int_{t_{\tau-1}}^{t_\tau} \mathbf{y}^f(u)\mathrm{d}u = \mathbf{S}_y^f\int_{t_{\tau-1}}^{t_\tau} \mathbf{C}(\boldsymbol{\theta})\mathbf{x}(u)\mathrm{d}u = \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta})\mathbf{x}^f\left(t_\tau\right) \\ &= \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta})\,\mathbf{S}_x^{f\top}\mathbf{S}_x^f\mathbf{x}^f(t_\tau) = \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta})\,\mathbf{S}_x^{f\top}\mathbf{x}_\tau^f\,, \end{aligned} \tag{A.6}$$

where the second equality uses (2.2). Combining (A.5) and (A.6) yields the measurement equation (2.5) for stock and flow measurements. Here, the non-zero columns of $\mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta})$ correspond to those of the latent variables that $\mathbf{y}_\tau^f$ depends on, namely, those in $\mathbf{x}_\tau^f$, whereas $\mathbf{S}_x^{f\top}\mathbf{x}_\tau^f$ returns an m_x-vector supplementing $\mathbf{x}_\tau^f$ with zeroes in positions corresponding to the remaining (zero) columns of $\mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta})$ (so dimensions match).

To compute the remaining blocks $\mathbb{E}[\boldsymbol{\eta}_\tau^s\,\boldsymbol{\eta}_\tau^{f\top}]$ and $\mathbb{E}[\boldsymbol{\eta}_\tau^f\,\boldsymbol{\eta}_\tau^{f\top}]$ of the covariance matrix (2.8), we again apply the Itô isometry, with $\boldsymbol{\eta}(t_\tau)$ and $\boldsymbol{\eta}^f(t_\tau)$ given by (A.2) and (A.4), respec-

tively, then use $\boldsymbol{\eta}_\tau^s = \boldsymbol{\eta}(t_\tau)$, $\boldsymbol{\eta}_\tau^f = \mathbf{S}_x^f \boldsymbol{\eta}^f(t_\tau)$. ■

A.2 Proof of Lemma 1

We must show that the eigenvalues of $\mathsf{A}(\boldsymbol{\theta}; h)$ lie within the unit circle.

Model with only stock state variables. Assume $\mathbf{A}(\boldsymbol{\theta})$ is diagonalizable,[1] $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$, with $\mathbf{V}$ a square matrix with columns given by the eigenvectors of $\mathbf{A}(\boldsymbol{\theta})$, and $\boldsymbol{\Lambda}$ a diagonal matrix with the corresponding eigenvalues along the diagonal. From Assumption 1, the eigenvalues of $\mathbf{A}(\boldsymbol{\theta})$ and, therefore, the elements of $\boldsymbol{\Lambda}$ all have negative real parts. Since $\mathsf{A}(\boldsymbol{\theta}; h) = \exp(\mathbf{A}(\boldsymbol{\theta})h) = \mathbf{V}\exp(\boldsymbol{\Lambda}h)\mathbf{V}^{-1}$, the eigenvalues of $\mathsf{A}(\boldsymbol{\theta}; h)$ are the diagonal elements of $\exp(\boldsymbol{\Lambda}h) = \text{diag}(\mathrm{e}^{\lambda_1 h}, ..., \mathrm{e}^{\lambda_{n_x^s} h})$, and so lie within the unit circle.

Model with stock and flow state variables. Here, $n_x = n_x^s + n_x^f$. From the previous result, the upper left block of $\mathsf{A}(\boldsymbol{\theta}; h)$ has eigenvalues within the unit circle. Given the structure of $\mathsf{A}(\boldsymbol{\theta}; h)$, and using decoupling arguments, is evident that, for $\sigma \in \mathbb{C}$, the nonzero roots of the characteristic polynomial $\det(\sigma\mathbf{I} - \mathsf{A}(\boldsymbol{\theta}; h))$ are the same as those of $\det(\sigma\mathbf{I} - \exp(\mathbf{A}(\boldsymbol{\theta})h))$ (see, e.g., Golub and Van Loan, 2013, Lemma 7.1.1). Indeed, let $\sigma(\mathbf{A}) = \{\sigma : \det(\sigma\mathbf{I} - \mathbf{A}) = 0\}$, and partition the discrete-time drift matrix as

$$\mathsf{A}(\boldsymbol{\theta}; h) = \begin{bmatrix} \mathsf{A}_{11}(\boldsymbol{\theta}; h) & \mathbf{0} \\ \mathsf{A}_{21}(\boldsymbol{\theta}; h) & \mathsf{A}_{22}(\boldsymbol{\theta}; h) \end{bmatrix}.$$

Then $\sigma(\mathsf{A}(\boldsymbol{\theta}; h)) = \sigma(\mathsf{A}_{11}(\boldsymbol{\theta}; h)) \cup \sigma(\mathsf{A}_{22}(\boldsymbol{\theta}; h))$. Since $\mathsf{A}_{22}(\boldsymbol{\theta}; h) = \mathbf{0}$ in the MX-SSR model, and $\sigma(\mathbf{0}) = \text{diag}(0, ..., 0)$, it follows that $\mathsf{A}(\boldsymbol{\theta}; h)$ has n_x^s eigenvalues given by $\mathrm{e}^{\lambda_i h}$, $i = 1, ..., n_x^s$, and n_x^f eigenvalues that are exactly zero. Therefore, all eigenvalues lie within the unit circle.

The EM-SSR model. From Assumption 1, the eigenvalues of $\mathbf{A}(\boldsymbol{\theta})$ have negative real part. Additionally, $\mathsf{A}(\boldsymbol{\theta}; h) = \mathbf{I}_{n_x^s} + \mathbf{A}(\boldsymbol{\theta})h$. The eigenvalues of $\mathsf{A}(\boldsymbol{\theta}; h)$ are of the type $1 + \lambda_i h$, $i = 1, ..., n_x^s$, which lie within the unit circle if and only if $\lambda_i > -2/h$, $i = 1, ..., n_x^s$. ■

A.3 Proof of Lemma 2

(a) We must show that under Assumptions 1, 2, and 4, if there exists a full rank $n_x \times n_x$ matrix $\mathbf{T}_h$ satisfying (3.8), then $\boldsymbol{\delta}(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, i.e., (3.2) applies, with $\boldsymbol{\delta}(\boldsymbol{\theta}, \mathbf{T}_h)$

[1] The matrix $\mathbf{A}(\boldsymbol{\theta})$ is diagonalizable if either i) it has m_x distinct eigenvalues; or ii) the sum of the geometric multiplicities of its eigenvalues is equal to m_x; or iii) the sum of the algebraic multiplicities of its eigenvalues is equal to m_x and, for each eigenvalue, the geometric multiplicity equals the algebraic multiplicity. However, the assumption is not restrictive. For non-diagonalizable matrices, the exponential can be computed using the Jordan canonical form, and the result carries over, based on a Taylor series expansion of the matrix exponential, showing that the eigenvalues of $\mathsf{A}(\boldsymbol{\theta}; h)$ still take the form $\mathrm{e}^{\lambda_i h}$, where λ_i is an eigenvalue of $\mathbf{A}(\boldsymbol{\theta})$.

from (3.1). Clearly, under (3.8), the $\mathsf{K}(\boldsymbol{\theta};h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)$ parts of (3.2) are automatic. It remains to verify the $\mathsf{A}(\boldsymbol{\theta};h)$ and $\mathsf{C}(\boldsymbol{\theta};h)$ parts of (3.2).

For the $\mathsf{A}(\boldsymbol{\theta};h)$ part of (3.2), we must show that $\mathbf{T}_h\mathsf{A}(\boldsymbol{\theta}_\ell;h)\mathbf{T}_h^{-1} = \mathsf{A}(\boldsymbol{\theta}_0;h)$. As the representation is minimal, by Assumption 4, we have from the transition equation of the state space representation (D.3) that $m_x = n_x = n_x^s$, with

$$\mathsf{A}(\boldsymbol{\theta};h) = \mathbf{A}_h(\boldsymbol{\theta}) = \exp(\mathbf{A}(\boldsymbol{\theta})h)\,. \tag{A.7}$$

It follows that

$$\begin{aligned}\mathbf{T}_h\mathsf{A}(\boldsymbol{\theta}_\ell;h)\mathbf{T}_h^{-1} &= \mathbf{T}_h\exp(\mathbf{A}(\boldsymbol{\theta}_\ell)h)\mathbf{T}_h^{-1} = \exp(\mathbf{T}_h\mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}h) = \exp(\mathbf{A}(\boldsymbol{\theta}_0)h)\\ &= \mathsf{A}(\boldsymbol{\theta}_0;h)\,,\end{aligned} \tag{A.8}$$

where the first and last equalities follow from (A.7), the second from the properties of the matrix exponential, and the third from the assumption on $\mathbf{T}_h\mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}$ in (3.8) in part (a) of the Lemma. Thus, the $\mathsf{A}(\boldsymbol{\theta};h)$ part of (3.2) is satisfied.

For the $\mathsf{C}(\boldsymbol{\theta};h)$ part of (3.2),

$$\begin{aligned}\mathsf{C}(\boldsymbol{\theta}_\ell;h)\mathbf{T}_h^{-1} &= \begin{bmatrix}\mathbf{S}_y^s\mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{A}_h(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}\\ \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{A}(\boldsymbol{\theta}_\ell)^{-1}(\mathbf{A}_h(\boldsymbol{\theta}_\ell)-\mathbf{I}_{n_x^s})\mathbf{T}_h^{-1}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{S}_y^s\mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}\mathbf{T}_h\mathbf{A}_h(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}\\ \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}\mathbf{T}_h\mathbf{A}(\boldsymbol{\theta}_\ell)^{-1}\mathbf{T}_h^{-1}(\mathbf{T}_h\mathbf{A}_h\left(\boldsymbol{\theta}_\ell\right)\mathbf{T}_h^{-1}-\mathbf{T}_h\mathbf{T}_h^{-1})\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{S}_y^s\mathbf{C}(\boldsymbol{\theta}_0)\mathbf{A}_h(\boldsymbol{\theta}_0)\\ \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta}_0)\mathbf{A}(\boldsymbol{\theta}_0)^{-1}(\mathbf{A}_h\left(\boldsymbol{\theta}_0\right)-\mathbf{I}_{n_x^s})\end{bmatrix}\\ &= \mathsf{C}(\boldsymbol{\theta}_0;h)\,,\end{aligned}$$

where the first and last equalities follow from (D.4) (as the representation is minimal, by Assumption 4), and the third from (A.8) and the assumptions on $\mathbf{T}_h\mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}$ and $\mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}$ in (3.8) in part (a) of the Lemma. Thus, the $\mathsf{C}(\boldsymbol{\theta};h)$ part of (3.2) is satisfied, too, i.e., (3.2) applies. This completes the proof of part (a) of the Lemma. We also note that the proof of the $\mathsf{A}(\boldsymbol{\theta};h)$ and $\mathsf{C}(\boldsymbol{\theta};h)$ parts of (3.2) only used the $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{C}(\boldsymbol{\theta})$ parts of (3.8), not the $\mathsf{K}(\boldsymbol{\theta};h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)$ parts.

(b) We show that under Assumptions 1, 2, and 4, if $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_\ell) = \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_0)$, and there exists a full rank $n_x \times n_x$ matrix $\mathbf{T}$ and an orthogonal $m_w \times m_w$ matrix $\mathbf{U}$ satisfying (3.9), then $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ are observationally equivalent, i.e., (3.2) applies, with $\mathbf{T}_h = \mathbf{T}$.

From the $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{C}(\boldsymbol{\theta})$ parts of (3.9), the $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{C}(\boldsymbol{\theta})$ parts of (3.8) are satisfied, with $\mathbf{T}_h = \mathbf{T}$, and because the proof of part (a) of the Lemma only used the $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{C}(\boldsymbol{\theta})$ parts of (3.8), not the $\mathsf{K}(\boldsymbol{\theta};h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)$ parts, the $\mathsf{A}(\boldsymbol{\theta};h)$ and $\mathsf{C}(\boldsymbol{\theta};h)$ parts

of (3.2) follow, with $\mathbf{T}_h = \mathbf{T}$.

For part (b), it therefore remains to verify the $\mathsf{K}(\boldsymbol{\theta}; h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$ parts of (3.2), with $\mathbf{T}_h = \mathbf{T}$. For these, we need the relation

$$\mathbf{T}\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell)\mathbf{T}^\top = \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_0) \tag{A.9}$$

and, similarly, $\mathbf{T}\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell)\mathbf{T}^\top = \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_0)$, and $\mathbf{T}\boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell)\mathbf{T}^\top = \mathbf{T}_h\boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_0)\mathbf{T}_h^\top$. We verify (A.9), and the other two follow analogously.

From (2.10), with $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top$, and $\mathbf{S}_x^f = \mathbf{I}_{n_x^s}$ in the minimal representation,

$$\mathbf{T}\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell)\mathbf{T}^\top = \int_0^h \int_0^u \mathbf{T}\exp(\mathbf{A}(\boldsymbol{\theta}_\ell)(u-r))\mathbf{B}(\boldsymbol{\theta}_\ell)\mathbf{B}(\boldsymbol{\theta}_\ell)^\top \exp(\mathbf{A}(\boldsymbol{\theta}_\ell)^\top r)\mathbf{T}^\top \mathrm{d}r\mathrm{d}u$$

Because $\mathbf{U}$ is orthogonal, this is rewritten as

$$\int_0^h \int_0^u \mathbf{T}\exp(\mathbf{A}(\boldsymbol{\theta}_\ell)(u-r))\mathbf{T}^{-1}\mathbf{T}\mathbf{B}(\boldsymbol{\theta}_\ell)\mathbf{U}\mathbf{U}^\top\mathbf{B}(\boldsymbol{\theta}_\ell)^\top\mathbf{T}^\top \left(\mathbf{T}^\top\right)^{-1} \exp(\mathbf{A}(\boldsymbol{\theta}_\ell)^\top r)\mathbf{T}^\top \mathrm{d}r\mathrm{d}u$$

By the properties of the matrix exponential, this is further rewritten as

$$\int_0^h \int_0^u \exp(\mathbf{T}\mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}^{-1}(u-r))\mathbf{T}\mathbf{B}(\boldsymbol{\theta}_\ell)\mathbf{U}\left(\mathbf{T}\mathbf{B}(\boldsymbol{\theta}_\ell)\mathbf{U}\right)^\top \exp(\left(\mathbf{T}^{-1}\right)^\top \mathbf{A}(\boldsymbol{\theta}_\ell)^\top\mathbf{T}^\top r)\mathrm{d}r\mathrm{d}u$$

Using the assumptions on $\mathbf{T}\mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}^{-1}$ and $\mathbf{T}\mathbf{B}(\boldsymbol{\theta}_\ell)\mathbf{U}$ in (3.9) in part (b) of the Lemma, this immediately reduces to

$$\int_0^h \int_0^u \exp(\mathbf{A}(\boldsymbol{\theta}_0)(u-r))\mathbf{B}(\boldsymbol{\theta}_0)\mathbf{B}(\boldsymbol{\theta}_0)^\top \exp(\mathbf{A}(\boldsymbol{\theta}_0)^\top r)\mathrm{d}r\mathrm{d}u$$

which is $\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_0)$, using (2.10), and (A.9) follows.

We need two more auxiliary results, in order to verify the $\mathsf{K}(\boldsymbol{\theta}; h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$ parts of (3.2). First, by (D.3), in the minimal representation, $\mathsf{B}(\boldsymbol{\theta}; h) = \begin{bmatrix} \mathbf{I}_{n_x^s} & \mathbf{0}_{n_x^s \times n_x^s} & \mathbf{0}_{n_x^s \times n_v} \end{bmatrix}$, and by (D.5), $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}; h) = \mathrm{diag}\big(\boldsymbol{\Sigma}_{\boldsymbol{\eta},h}(\boldsymbol{\theta}), \boldsymbol{\Sigma}_{\boldsymbol{v},h}(\boldsymbol{\theta})\big)$, so that

$$\mathsf{B}(\boldsymbol{\theta}; h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}; h)\mathsf{B}(\boldsymbol{\theta}; h)^\top = \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta})\,. \tag{A.10}$$

Therefore, we have the auxiliary result

$$\begin{aligned}\mathbf{T}\mathsf{B}(\boldsymbol{\theta}_\ell;h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_\ell;h)\mathsf{B}(\boldsymbol{\theta}_\ell;h)^\top\mathbf{T}^\top &= \mathbf{T}\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell)\mathbf{T}^\top = \dots \\ &= \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_0) = \mathsf{B}(\boldsymbol{\theta}_0;h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_0;h)\mathsf{B}(\boldsymbol{\theta}_0;h)^\top, \quad \text{(A.11)}\end{aligned}$$

where the first and last equality follow from (A.10), and the second from the discussion around (A.9).

Second, from (D.4), (D.5), and (2.8), we have that $\mathsf{D}(\boldsymbol{\theta}_\ell;h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_\ell;h)\mathsf{D}(\boldsymbol{\theta}_\ell;h)^\top$ is given by

$$\begin{bmatrix} \mathbf{S}_y^s\mathbf{C}(\boldsymbol{\theta}_\ell) & \mathbf{0}_{n_y^s\times n_x^s} & \mathbf{S}_\upsilon^s \\ \mathbf{0}_{n_y^f\times n_x^s} & \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta}_\ell) & \mathbf{S}_\upsilon^f \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell) & \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell) & \mathbf{0}_{n_x^s\times n_\upsilon} \\ \boldsymbol{\Sigma}_{\boldsymbol{\eta}^f\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell) & \boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell) & \mathbf{0}_{n_x^s\times n_\upsilon} \\ \mathbf{0}_{n_\upsilon\times n_x^s} & \mathbf{0}_{n_\upsilon\times n_x^s} & \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \mathbf{C}(\boldsymbol{\theta}_\ell)^\top\mathbf{S}_y^{s\top} & \mathbf{0}_{n_x^s\times n_y^f} \\ \mathbf{0}_{n_x^s\times n_y^s} & \mathbf{C}(\boldsymbol{\theta}_\ell)^\top\mathbf{S}_y^{f\top} \\ \mathbf{S}_\upsilon^{s\top} & \mathbf{S}_\upsilon^{f\top} \end{bmatrix},$$

writing $\mathbf{S}_\upsilon = \left(\mathbf{S}_\upsilon^{s\top}, \mathbf{S}_\upsilon^{f\top}\right)^\top$. This works out as

$$\begin{bmatrix} \mathbf{S}_y^s\mathbf{C}(\boldsymbol{\theta}_\ell)\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell)\mathbf{C}(\boldsymbol{\theta}_\ell)^\top\mathbf{S}_y^{s\top} + \mathbf{S}_\upsilon^s\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_\ell)\mathbf{S}_\upsilon^{s\top} & \mathbf{S}_y^s\mathbf{C}(\boldsymbol{\theta}_\ell)\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell)\mathbf{C}(\boldsymbol{\theta}_\ell)^\top\mathbf{S}_y^{f\top} + \mathbf{S}_\upsilon^s\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_\ell)\mathbf{S}_\upsilon^{f\top} \\ \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta}_\ell)\boldsymbol{\Sigma}_{\boldsymbol{\eta}^f\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell)\mathbf{C}(\boldsymbol{\theta}_\ell)^\top\mathbf{S}_y^{s\top} + \mathbf{S}_\upsilon^f\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_\ell)\mathbf{S}_\upsilon^{s\top} & \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta}_\ell)\boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell)\mathbf{C}(\boldsymbol{\theta}_\ell)^\top\mathbf{S}_y^{f\top} + \mathbf{S}_y^f\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_\ell)\mathbf{S}_\upsilon^{f\top} \end{bmatrix}. \quad \text{(A.12)}$$

In this, the expressions of the type $\mathbf{C}(\boldsymbol{\theta}_\ell)\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell)\mathbf{C}(\boldsymbol{\theta}_\ell)^\top$ take the form

$$\begin{aligned}\mathbf{C}(\boldsymbol{\theta}_\ell)\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell)\mathbf{C}(\boldsymbol{\theta}_\ell)^\top &= \mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{T}^{-1}\mathbf{T}\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell)\mathbf{T}^\top\left(\mathbf{T}^\top\right)^{-1}\mathbf{C}(\boldsymbol{\theta}_\ell)^\top \\ &= \mathbf{C}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_0)\mathbf{C}(\boldsymbol{\theta}_0)^\top,\end{aligned}$$

using the assumption $\mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{T}^{-1} = \mathbf{C}(\boldsymbol{\theta}_0)$ from part (a) of the Lemma, along with (A.9), and the similar relations for $\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}_\ell)$. As $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_\ell) = \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_0)$ by the assumption in part (b) of the Lemma, we conclude from the derivations that

$$\mathsf{D}(\boldsymbol{\theta}_\ell;h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_\ell;h)\mathsf{D}(\boldsymbol{\theta}_\ell;h)^\top = \mathsf{D}(\boldsymbol{\theta}_0;h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_0;h)\mathsf{D}(\boldsymbol{\theta}_0;h)^\top. \quad \text{(A.13)}$$

We have established the $\mathsf{A}(\boldsymbol{\theta};h)$ and $\mathsf{C}(\boldsymbol{\theta};h)$ parts of (3.2), as well as the auxiliary results (A.9), (A.11), and (A.13). To verify the remaining $\mathsf{K}(\boldsymbol{\theta};h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)$ parts of (3.2), we use that $\mathsf{K}(\boldsymbol{\theta};h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)$ jointly with the matrix $\mathbf{P}(\boldsymbol{\theta};h)$ in (E.10) solve the system (E.8)-(E.10). First, corresponding to the $\mathsf{A}(\boldsymbol{\theta};h)$ and $\mathsf{C}(\boldsymbol{\theta};h)$ parts of (3.2), we adopt the ansatz

$$\mathbf{T}^\top\mathbf{P}(\boldsymbol{\theta}_\ell;h)\mathbf{T}^\top = \mathbf{P}(\boldsymbol{\theta}_0;h), \quad \text{(A.14)}$$

which we subsequently verify. Given this, we have that

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h) &= \mathsf{C}(\boldsymbol{\theta}_\ell; h)\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top + \mathsf{D}(\boldsymbol{\theta}_\ell; h)\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}_\ell; h)\mathsf{D}(\boldsymbol{\theta}_\ell; h)^\top \\
&= \mathsf{C}(\boldsymbol{\theta}_\ell; h)\mathbf{T}^{-1}\mathbf{T}\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathbf{T}^\top \left(\mathbf{T}^\top\right)^{-1} \mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top + \mathsf{D}(\boldsymbol{\theta}_0; h)\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}_0; h)\mathsf{D}(\boldsymbol{\theta}_0; h)^\top \\
&= \mathsf{C}(\boldsymbol{\theta}_0; h)\mathbf{P}(\boldsymbol{\theta}_0; h)\mathsf{C}(\boldsymbol{\theta}_0; h)^\top + \mathsf{D}(\boldsymbol{\theta}_0; h)\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}_0; h)\mathsf{D}(\boldsymbol{\theta}_0; h)^\top \\
&= \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_0; h)\,, \qquad \text{(A.15)}
\end{aligned}$$

using (E.9) in the first and last equalities, (A.13) in the second, and (A.14) and the $\mathsf{C}(\boldsymbol{\theta}; h)$ part of (3.2) in the third, hence verifying the $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_0; h)$ part of (3.2). Finally, for the Kalman gain,

$$\begin{aligned}
\mathbf{T}\mathsf{K}(\boldsymbol{\theta}_\ell; h) &= \mathbf{T}\mathsf{A}(\boldsymbol{\theta}_\ell; h)\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h)^{-1} \\
&= \mathbf{T}\mathsf{A}(\boldsymbol{\theta}_\ell; h)\mathbf{T}^{-1}\mathbf{T}\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathbf{T}^\top \left(\mathbf{T}^\top\right)^{-1} \mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h)^{-1} \\
&= \mathsf{A}(\boldsymbol{\theta}_0; h)\mathbf{P}(\boldsymbol{\theta}_0; h)\mathsf{C}(\boldsymbol{\theta}_0; h)^\top \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_0; h)^{-1} \\
&= \mathsf{K}(\boldsymbol{\theta}_0; h)\,, \qquad \text{(A.16)}
\end{aligned}$$

using (E.8) in the first and last equalities, and (A.14) and the $\mathsf{A}(\boldsymbol{\theta}; h)$, $\mathsf{C}(\boldsymbol{\theta}; h)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$ parts of (3.2) in the third, hence verifying the $\mathsf{K}(\boldsymbol{\theta}; h)$ part of (3.2). However, the derivations (A.15) and (A.16) rely on the ansatz (A.14). Therefore, to complete the verification result, it remains to verify the latter, using (E.10), which involves $\mathbf{P}(\boldsymbol{\theta}; h)$ both on the left and right sides. We use what we have so far, i.e., the auxiliary result (A.11), and (3.2), which relies on the ansatz (A.14), as well as (A.14) itself, on the right side of (E.10), and show that the result satisfies (A.14), hence verifying the ansatz. In detail, we have

$$\begin{aligned}
\mathbf{T}\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathbf{T}^\top &= \mathbf{T}\mathsf{A}(\boldsymbol{\theta}_\ell; h)\left[\mathbf{P}(\boldsymbol{\theta}_\ell; h) - \mathsf{K}(\boldsymbol{\theta}_\ell; h)\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h)\mathsf{K}(\boldsymbol{\theta}_\ell; h)^\top\right]\mathsf{A}(\boldsymbol{\theta}_\ell; h)^\top\mathbf{T}^\top \\
&\quad + \mathbf{T}\mathsf{B}(\boldsymbol{\theta}_\ell; h)\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}_\ell; h)\mathsf{B}(\boldsymbol{\theta}_\ell; h)^\top\mathbf{T}^\top \\
&= \mathbf{T}\mathsf{A}(\boldsymbol{\theta}_\ell; h)\mathbf{T}^{-1}\mathbf{T}\left[\mathbf{P}(\boldsymbol{\theta}_\ell; h) - \mathsf{K}(\boldsymbol{\theta}_\ell; h)\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h)\mathsf{K}(\boldsymbol{\theta}_\ell; h)^\top\right]\mathbf{T}^\top\left(\mathbf{T}^\top\right)^{-1}\mathsf{A}(\boldsymbol{\theta}_\ell; h)^\top\mathbf{T}^\top \\
&\quad + \mathbf{T}\mathsf{B}(\boldsymbol{\theta}_\ell; h)\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}_\ell; h)\mathsf{B}(\boldsymbol{\theta}_\ell; h)^\top\mathbf{T}^\top \\
&= \mathsf{A}(\boldsymbol{\theta}_0; h)\left[\mathbf{P}(\boldsymbol{\theta}_0; h) - \mathsf{K}(\boldsymbol{\theta}_0; h)\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_0; h)\mathsf{K}(\boldsymbol{\theta}_0; h)^\top\right]\mathsf{A}(\boldsymbol{\theta}_0; h)^\top \\
&\quad + \mathsf{B}(\boldsymbol{\theta}_0; h)\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta}_0; h)\mathsf{B}(\boldsymbol{\theta}_0; h)^\top \\
&= \mathbf{P}(\boldsymbol{\theta}_0; h),
\end{aligned}$$

using (E.10) in the first and last equalities, and (A.11), (3.2), and (A.14) in the third. This verifies (A.14) and, therefore, $\boldsymbol{\delta}(\boldsymbol{\theta}_\ell, \mathbf{T}) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, so that (3.2) follows, with $\mathbf{T}_h = \mathbf{T}$, and $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ are observationally equivalent. This completes the proof of part (b) of the Lemma.

(c) We show that under Assumptions 1, 2, and 4, if $\boldsymbol{\delta}(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, i.e., if (3.2)

applies, for some full rank $n_x \times n_x$ matrix $\mathbf{T}_h$, then condition (3.10) follows, with $\mathbf{C}^s(\boldsymbol{\theta}) = \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta})$ from (3.7).

From (D.4), the first n_y^s rows of $\mathsf{C}(\boldsymbol{\theta}; h)$ are given by $\mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}) \mathsf{A}(\boldsymbol{\theta}; h)$. The condition on the discrete-time reaction matrix in (3.2), namely, $\mathsf{C}(\boldsymbol{\theta}_\ell; h)\mathbf{T}_h^{-1} = \mathsf{C}(\boldsymbol{\theta}_0; h)$, for some full rank $n_x \times n_x$ matrix $\mathbf{T}_h$, therefore implies

$$\mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}_\ell) \mathsf{A}(\boldsymbol{\theta}_\ell; h) \mathbf{T}_h^{-1} = \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}_0) \mathsf{A}(\boldsymbol{\theta}_0; h)\,. \tag{A.17}$$

The left side is rewritten as

$$\mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} \mathbf{T}_h \mathsf{A}(\boldsymbol{\theta}_\ell; h) \mathbf{T}_h^{-1} = \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} \mathsf{A}(\boldsymbol{\theta}_0; h)\,, \tag{A.18}$$

using the first condition in (3.2), on the discrete-time transition matrix. Since the left sides of (A.17) and (A.18) coincide, so must the right sides. As $\mathsf{A}(\boldsymbol{\theta}_0; h)$ is full rank, it follows that $\mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1} = \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}_0)$, so (3.10) holds exactly.

(d) We show that under Assumptions 1, 2, and 4, if $\boldsymbol{\delta}(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, i.e., if (3.2) applies, for some full rank $n_x \times n_x$ matrix $\mathbf{T}_h$, then all conditions in (3.8) hold exactly in the absence of aliases, and the conditions on $\mathbf{A}(\cdot)$ and $\mathbf{C}^f(\cdot)$ in (3.8) hold to order of approximation $O(h)$ in general, regardless of possible aliases (condition (3.10) on $\mathbf{C}^s$ and the conditions on K, $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}$ in (3.8) are still exact). Clearly, under (3.2), the $\mathsf{K}(\boldsymbol{\theta}; h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$ parts of (3.8) are automatic, and the $\mathbf{C}^s(\boldsymbol{\theta}; h)$ part of (3.8) follows from part (c) of the Lemma. It remains to verify the $\mathsf{A}(\boldsymbol{\theta}; h)$ and $\mathsf{C}^f(\boldsymbol{\theta}; h)$ parts of (3.8).

As in (a) and (b), under Assumption 4, we have $\mathsf{A}(\boldsymbol{\theta}; h) = \mathbf{A}_h(\boldsymbol{\theta}) = \exp(\mathbf{A}(\boldsymbol{\theta})h)$, which is (A.7) (see also (D.3)). If $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ are observationally equivalent, the condition on the discrete-time transition matrix in (3.2), namely, $\mathbf{T}_h \mathsf{A}(\boldsymbol{\theta}_\ell; h)\mathbf{T}_h^{-1} = \mathsf{A}(\boldsymbol{\theta}_0; h)$, for some full rank $n_x \times n_x$ matrix $\mathbf{T}_h$, therefore takes the form

$$\mathbf{T}_h \exp(\mathbf{A}(\boldsymbol{\theta}_\ell)h)\mathbf{T}_h^{-1} = \exp(\mathbf{A}(\boldsymbol{\theta}_0)h)\,. \tag{A.19}$$

Using the properties of the matrix exponential, this is recast as

$$\exp(\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}h) = \exp(\mathbf{A}(\boldsymbol{\theta}_0)h)\,. \tag{A.20}$$

In the absence of aliases, the matrix exponential is injective, so (A.20) implies $\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1} = \mathbf{A}(\boldsymbol{\theta}_0)$, which is the first condition in (3.8).

In general, in the potential presence of aliases, we consider (A.20) directly. Using (2.6), this is alternatively recast as

$$\mathbf{I}_{n_x^s} + \mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}h + \tfrac{1}{2}\left[\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1}\right]^2 h^2 + \cdots = \mathbf{I}_{n_x^s} + \mathbf{A}(\boldsymbol{\theta_0})h + \tfrac{1}{2}\left[\mathbf{A}(\boldsymbol{\theta_0})\right]^2 h^2 + \ldots\,,$$

hence implying the n_x^2 approximate conditions on the continuous-time drift matrix given by

$$\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} = \mathbf{A}(\boldsymbol{\theta}_0) + O(h)\,. \tag{A.21}$$

This completes the proof of the portion of part (d) of the Lemma dealing with the transition matrix (i.e., both the exact and $O(h)$ cases).

Consider next the portion of (3.8) dealing with the reaction matrix. The remaining n_y^f rows of $\mathbf{C}(\boldsymbol{\theta})$, beyond those covered in **(c)**, are $\mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta})$, cf. (3.7). Using (D.4) again, the corresponding n_y^f rows of $\mathsf{C}(\boldsymbol{\theta}; h)$ are $\mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta})^{-1} (\mathbf{A}_h(\boldsymbol{\theta}) - \mathbf{I}_{n_x})$. The condition $\mathsf{C}(\boldsymbol{\theta}_\ell; h) \mathbf{T}_h^{-1} = \mathsf{C}(\boldsymbol{\theta}_0; h)$ in (3.2) therefore implies

$$\mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}_\ell) \mathbf{A}(\boldsymbol{\theta}_\ell)^{-1} (\mathbf{A}_h(\boldsymbol{\theta}_\ell) - \mathbf{I}_{n_x}) \mathbf{T}_h^{-1} = \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}_0) \mathbf{A}(\boldsymbol{\theta}_0)^{-1} (\mathbf{A}_h(\boldsymbol{\theta}_0) - \mathbf{I}_{n_x^s})\,. \tag{A.22}$$

The left side is rewritten as

$$\mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} \left[\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} \right]^{-1} (\mathbf{T}_h \mathbf{A}_h(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} - \mathbf{I}_{n_x})\,.$$

As $\mathsf{A}(\boldsymbol{\theta}; h) = \mathbf{A}_h(\boldsymbol{\theta})$, and using the $\mathsf{A}(\boldsymbol{\theta}; h)$ portion of (3.2), the expression is further rewritten as

$$\mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} \left[\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} \right]^{-1} (\mathbf{A}_h(\boldsymbol{\theta}_0) - \mathbf{I}_{n_x})\,. \tag{A.23}$$

As we have shown, in the absence of aliases, the first condition (on the transition matrix) in (3.8), holds exactly, implying that (A.23) takes the form

$$\mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} \mathbf{A}(\boldsymbol{\theta}_0)^{-1} (\mathbf{A}_h(\boldsymbol{\theta}_0) - \mathbf{I}_{n_x})\,, \tag{A.24}$$

exactly, and the result (A.21) implies that this applies to order $O(h)$ in general. Here,

$$\begin{aligned}
\mathbf{A}_h(\boldsymbol{\theta}_0) - \mathbf{I}_{n_x} &= \exp(\mathbf{A}(\boldsymbol{\theta}_0) h) - \mathbf{I}_{n_x} \\
&= \exp(\mathbf{V}(\boldsymbol{\theta}_0) \mathbf{D}(\boldsymbol{\theta}_0) \mathbf{V}(\boldsymbol{\theta}_0)^{-1} h) - \mathbf{I}_{n_x} \\
&= \mathbf{V}(\boldsymbol{\theta}_0) \exp(\mathbf{D}(\boldsymbol{\theta}_0) h) \mathbf{V}(\boldsymbol{\theta}_0)^{-1} - \mathbf{I}_{n_x} \\
&= \mathbf{V}(\boldsymbol{\theta}_0) \left(\exp(\mathbf{D}(\boldsymbol{\theta}_0) h) - \mathbf{I}_{n_x} \right) \mathbf{V}(\boldsymbol{\theta}_0)^{-1}\,,
\end{aligned}$$

where the columns of $\mathbf{V}(\boldsymbol{\theta}_0)$ are given by the eigenvectors of $\mathbf{A}_h(\boldsymbol{\theta}_0)$, and $\mathbf{D}(\boldsymbol{\theta}_0)$ is the diagonal matrix with the eigenvalues along the diagonal. By Assumption 1, the eigenvalues have strictly negative real parts, so $\exp(\mathbf{D}(\boldsymbol{\theta}_0) h) - \mathbf{I}_{n_x}$ is full rank and, hence, so is $\mathbf{A}_h(\boldsymbol{\theta}_0) - \mathbf{I}_{n_x}$. Since $\mathbf{A}(\boldsymbol{\theta}_0)$ is full rank, too, again using Assumption 1, so is $\mathbf{A}(\boldsymbol{\theta}_0)^{-1}(\mathbf{A}_h(\boldsymbol{\theta}_0) - \mathbf{I}_{n_x})$. Hence, from (A.22), and writing the left side as (A.24), $\mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} = \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}_0)$ in the absence of aliases, and the relation holds to order h in general. This completes the portion of part (d) of the Lemma dealing with the reaction matrix.

Combining the results on the reaction matrix in (c) and (d), under observational equivalence of $\boldsymbol{\theta}_\ell$ and $\boldsymbol{\theta}_0$, and since $\mathbf{I}_{n_y} = \mathrm{diag}(\mathbf{S}_y^s, \mathbf{S}_y^f)$, we have the $n_y m_x$ conditions on the optimal continuous-time reaction matrix

$$\mathbf{C}(\boldsymbol{\theta}_\ell)\mathbf{T}_h^{-1} = \mathbf{C}(\boldsymbol{\theta}_0)\,, \tag{A.25}$$

which are exact for the first n_y^s rows, and for the remaining n_y^f rows hold exactly in the absence of aliases, and to order h in general.

(e) For part (e) of the Lemma, it remains to verify that the conditions on $\mathbf{B}(\boldsymbol{\theta})$ in (3.8) hold exactly in the absence of aliases, with $\mathbf{T} = \mathbf{T}_h$, and to order $O(h)$ regardless of aliases. All conditions on the four matrices in (3.2) in combination, together with the expression for the Kalman gain $\mathsf{K}(\boldsymbol{\theta}; h)$ from (E.8), imply

$$\begin{aligned}
\mathbf{T}_h\mathsf{K}(\boldsymbol{\theta}_\ell; h) &= \mathsf{K}(\boldsymbol{\theta}_0; h) \\
&= \mathsf{A}(\boldsymbol{\theta}_0; h)\mathbf{P}(\boldsymbol{\theta}_0; h)\mathsf{C}(\boldsymbol{\theta}_0; h)^\top \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_0; h)^{-1} \\
&= \mathbf{T}_h\mathsf{A}(\boldsymbol{\theta}_\ell; h)\mathbf{T}_h^{-1}\mathbf{P}(\boldsymbol{\theta}_0; h)\left(\mathbf{T}_h^\top\right)^{-1}\mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h)^{-1},
\end{aligned}$$

exactly. Comparing with $\mathbf{T}_h\mathsf{K}(\boldsymbol{\theta}_\ell; h) = \mathbf{T}_h\mathsf{A}(\boldsymbol{\theta}_\ell; h)\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h)^{-1}$ from (E.8), and because $\mathsf{A}(\boldsymbol{\theta}_\ell; h)$, $\mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top$, and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h)^{-1}$ are full rank, we conclude that

$$\mathbf{T}_h\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathbf{T}_h^\top = \mathbf{P}(\boldsymbol{\theta}_0; h)\,. \tag{A.26}$$

The conditions on $\mathsf{C}(\boldsymbol{\theta}; h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$ in (3.2), along with the expression for $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)$ from (E.9), and the result (A.26) on $\mathbf{P}(\boldsymbol{\theta}_\ell; h)$, together imply

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h) &= \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_0; h) \\
&= \mathsf{C}(\boldsymbol{\theta}_0; h)\mathbf{P}(\boldsymbol{\theta}_0; h)\mathsf{C}(\boldsymbol{\theta}_0; h)^\top + \mathsf{D}(\boldsymbol{\theta}_0; h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_0; h)\mathsf{D}(\boldsymbol{\theta}_0; h)^\top \\
&= \mathsf{C}(\boldsymbol{\theta}_\ell; h)\mathbf{T}_h^{-1}\mathbf{T}_h\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathbf{T}_h^\top\left(\mathbf{T}_h^\top\right)^{-1}\mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top + \mathsf{D}(\boldsymbol{\theta}_0; h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_0; h)\mathsf{D}(\boldsymbol{\theta}_0; h)^\top \\
&= \mathsf{C}(\boldsymbol{\theta}_\ell; h)\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top + \mathsf{D}(\boldsymbol{\theta}_0; h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_0; h)\mathsf{D}(\boldsymbol{\theta}_0; h)^\top.
\end{aligned}$$

Comparing with $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}_\ell; h) = \mathsf{C}(\boldsymbol{\theta}_\ell; h)\mathbf{P}(\boldsymbol{\theta}_\ell; h)\mathsf{C}(\boldsymbol{\theta}_\ell; h)^\top + \mathsf{D}(\boldsymbol{\theta}_\ell; h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_\ell; h)\mathsf{D}(\boldsymbol{\theta}_\ell; h)^\top$ from (E.9), we conclude that

$$\mathsf{D}(\boldsymbol{\theta}_\ell; h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_\ell; h)\mathsf{D}(\boldsymbol{\theta}_\ell; h)^\top = \mathsf{D}(\boldsymbol{\theta}_0; h)\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_0; h)\mathsf{D}(\boldsymbol{\theta}_0; h)^\top\,. \tag{A.27}$$

With $\mathsf{D}(\boldsymbol{\theta}; h)$ from (D.4), the left side of (A.27) takes the form (A.12), and the right side is similar, with $\boldsymbol{\theta}_0$ replacing $\boldsymbol{\theta}_\ell$. Therefore, from the upper left corner of condition (A.27), and because $\mathbf{S}_\upsilon^s\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_\ell)\mathbf{S}_\upsilon^{s\top} = \mathbf{S}_\upsilon^s\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}_0)\mathbf{S}_\upsilon^{s\top}$, by the assumption in part (e) of

the Lemma, it then follows that

$$\mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}_\ell) \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell; h) \mathbf{C}(\boldsymbol{\theta}_\ell)^\top \mathbf{S}_y^{s\top} = \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}_0) \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_0; h) \mathbf{C}(\boldsymbol{\theta}_0)^\top \mathbf{S}_y^{s\top} . \tag{A.28}$$

Using $\mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}) = \mathbf{C}^s(\boldsymbol{\theta})$ from (3.7), and the first n_y^s exact rows of (A.25) (i.e., (3.10)), the left side of (A.28) is

$$\mathbf{C}^s(\boldsymbol{\theta}_0) \mathbf{T}_h \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell; h) \mathbf{T}_h^\top \mathbf{C}^s(\boldsymbol{\theta}_0)^\top . \tag{A.29}$$

Comparing with the right side of (A.28), and because the $n_y^s \times n_x^s$ matrix $\mathbf{C}^s(\boldsymbol{\theta}_0)$ has full column rank $n_x^s = n_x$ by the assumption in part **(e)** of the Lemma, we conclude that

$$\mathbf{T}_h \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_\ell; h) \mathbf{T}_h^\top = \boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}_0; h) \tag{A.30}$$

holds exactly. Using (2.9), this is recast in more picturesque form as

$$\mathbf{T}_h \int_0^h \exp(\mathbf{A}(\boldsymbol{\theta}_\ell)u) \mathbf{B}(\boldsymbol{\theta}_\ell) \mathbf{B}(\boldsymbol{\theta}_\ell)^\top \exp(\mathbf{A}(\boldsymbol{\theta}_\ell)^\top u) \mathrm{d}u \mathbf{T}_h^\top \\ = \int_0^h \exp(\mathbf{A}(\boldsymbol{\theta}_0)u) \mathbf{B}(\boldsymbol{\theta}_0) \mathbf{B}(\boldsymbol{\theta}_0)^\top \exp(\mathbf{A}(\boldsymbol{\theta}_0)^\top u) \mathrm{d}u \, .$$

Using the property $\mathbf{T}_h \exp(\mathbf{A}(\boldsymbol{\theta}_\ell)u) \mathbf{T}_h^{-1} = \exp(\mathbf{T}_h \mathbf{A}(\boldsymbol{\theta}_\ell) \mathbf{T}_h^{-1} u)$ of the matrix exponential, the result in (3.8) on the continuous-time transition matrix from part (d), which is exact in the absence of aliases, and holds to order $O(h)$ in general, cf. (A.21), implies that $\mathbf{T}_h \exp(\mathbf{A}(\boldsymbol{\theta}_\ell)u) \mathbf{T}_h^{-1} = \exp(\mathbf{A}(\boldsymbol{\theta}_0)u)$ (exactly resp. approximately), so the equation is

$$\int_0^h \exp(\mathbf{A}(\boldsymbol{\theta}_0)u) \mathbf{T}_h \mathbf{B}(\boldsymbol{\theta}_\ell) \mathbf{U}_h \mathbf{U}_h^\top \mathbf{B}(\boldsymbol{\theta}_\ell)^\top \mathbf{T}_h^\top \exp(\mathbf{A}(\boldsymbol{\theta}_0)^\top u) \mathrm{d}u \\ = \int_0^h \exp(\mathbf{A}(\boldsymbol{\theta}_0)u) \mathbf{B}(\boldsymbol{\theta}_0) \mathbf{B}(\boldsymbol{\theta}_0)^\top \exp(\mathbf{A}(\boldsymbol{\theta}_0)^\top u) \mathrm{d}u \, ,$$

for $\mathbf{U}_h$ orthogonal, again exactly in the absence of aliases, and to order $O(h)$ in general. As $\exp(\mathbf{A}(\boldsymbol{\theta}_\ell)u)$ is full rank, we have the $n_x m_w$ conditions on the continuous-time diffusion matrix given by

$$\mathbf{T}_h \mathbf{B}(\boldsymbol{\theta}_\ell) \mathbf{U}_h = \mathbf{B}(\boldsymbol{\theta}_0) \, ,$$

which are exact in the absence of aliases, and hold to order $O(h)$ in general. This completes the proof of part (e) of Lemma 2. ■

A.4 Proof of Lemma 3

Following Komunjer and Ng (2011), for given $h > 0$, $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ if and only if there is no observationally equivalent $\boldsymbol{\theta}_\ell$ in a neighborhood of $\boldsymbol{\theta}_0$, i.e., if and only if the system (3.2), namely,

$$\boldsymbol{\delta}(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{I}_{n_x}), \tag{A.31}$$

has a locally unique solution $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, with $\boldsymbol{\delta}$ from (3.1). Based on the underlying continuous-time model, part (a) of the Lemma extends the result to $\boldsymbol{\delta}^+$ from (3.11).

(a) Assume $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. The proof is by contradiction. Thus, suppose (3.12) does not have a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$. Clearly, $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ is a solution, so there are other local solutions. This implies that there exists a sequence satisfying $(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell) \to (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ as $\ell \to \infty$, and $\boldsymbol{\delta}^+(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell) = \boldsymbol{\delta}^+(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, for all ℓ. As $\boldsymbol{\delta}^+(\boldsymbol{\theta}, \mathbf{T}_h)$ is given as $(\text{vec}\left(\mathbf{C}^s(\boldsymbol{\theta})\mathbf{T}_h^{-1}\right)^\top, \boldsymbol{\delta}(\boldsymbol{\theta}, \mathbf{T}_h)^\top)^\top$, by (3.11), it follows that $\boldsymbol{\delta}(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, for all ℓ, with $(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell) \to (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ as $\ell \to \infty$. Thus, (A.31) does not have a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$. Then $\boldsymbol{\theta}$ is not locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at $\boldsymbol{\theta}_0$, and hence the contradiction.

For the converse, assume that the system (3.12) has a locally unique solution $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$. The proof is again by contradiction. Thus, suppose $\boldsymbol{\theta}$ is not locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at $\boldsymbol{\theta}_0$. Then (A.31) does not have a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$. Clearly, $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ is a solution, so there are other local solutions. This implies that there exists a sequence satisfying $(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell) \to (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ as $\ell \to \infty$, with $\delta(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, i.e., (A.31) and hence (3.2) is satisfied, for each ℓ. Therefore, $\boldsymbol{\theta}_\ell$ and $\boldsymbol{\theta}_0$ are observationally equivalent and, by Lemma 2.**(c)**, this implies

$$\mathbf{C}^s(\boldsymbol{\theta}_\ell)\mathbf{T}_\ell^{-1} = \mathbf{C}^s(\boldsymbol{\theta}_0), \tag{A.32}$$

for all ℓ. Hence, we have

$$\begin{aligned}
\boldsymbol{\delta}^+(\boldsymbol{\theta}_0, \mathbf{I}_{n_x}) &= \left[\text{vec}\left(\mathbf{C}^s(\boldsymbol{\theta}_0)\right)^\top, \boldsymbol{\delta}(\boldsymbol{\theta}, \mathbf{I}_{n_x})^\top\right]^\top \\
&= \left[\text{vec}\left(\mathbf{C}^s(\boldsymbol{\theta}_\ell)\mathbf{T}_\ell^{-1}\right)^\top, \boldsymbol{\delta}(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell)^\top\right]^\top \\
&= \boldsymbol{\delta}^+(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell),
\end{aligned}$$

for all ℓ, where the first and last equalities use (3.11), and the second uses (A.31) (equivalently, (3.2)) and (A.32). Thus, we have $(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell) \to (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ as $\ell \to \infty$, and $\boldsymbol{\delta}^+(\boldsymbol{\theta}_\ell, \mathbf{T}_\ell) = \boldsymbol{\delta}^+(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, so (3.12) is satisfied, for all ℓ, hence contradicting that it has

a locally unique solution $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$. This completes the proof of part (a) of the Lemma.

(b) Again, for given $h > 0$, $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at a point $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ if and only if there is no observationally equivalent $\boldsymbol{\theta}_\ell$ in a neighborhood of $\boldsymbol{\theta}_0$. Combining (a) and (d) of Lemma 2, under the stated assumptions, and in the absence of aliases, observational equivalence of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ is equivalent to existence of $\mathbf{T}_h$ such that (3.8) is satisfied. Thus, the condition that there be no observationally equivalent $\boldsymbol{\theta}_\ell$ in a neighborhood of $\boldsymbol{\theta}_0$ is equivalent to the condition that (3.16) has a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, with $\boldsymbol{\delta}_a(\boldsymbol{\theta}, \mathbf{T}_h)$ defined in (3.15), and local absence of aliases suffices. We conclude that for given $h > 0$, and in the absence of local alises, $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at a point $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ if and only if (3.16) has a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$.

(c) The proof follows along the lines of the proof of part (b) of the Lemma. Combining (b) and (e) of Lemma 2, under the stated assumptions, in the absence of aliases, and in a neighborhood of $\boldsymbol{\theta}_0$ in which $\boldsymbol{\Sigma}_{\boldsymbol{v},h}(\boldsymbol{\theta})$ is constant, observational equivalence of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_\ell$ is equivalent to existence of $\mathbf{T}_h$, $\mathbf{U}_h$ such that (3.9) is satisfied, with $(\mathbf{T}, \mathbf{U}) = (\mathbf{T}_h, \mathbf{U}_h)$. Thus, the identifiability condition that there be no observationally equivalent $\boldsymbol{\theta}_\ell$ in a neighborhood of $\boldsymbol{\theta}_0$ is equivalent to the condition that (3.14) has a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h, \mathbf{U}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x}, \mathbf{I}_{m_w})$, with $\boldsymbol{\delta}_a(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h)$ defined in (3.13), and local absence of aliases together with local constancy of $\boldsymbol{\Sigma}_{\boldsymbol{v},h}(\boldsymbol{\theta})$ suffices. ■

A.5 Proof of Proposition 2

The proof of Proposition 2 follows that of Propositions 2-S and 2-NS in Komunjer and Ng (2011). In the latter, based on the discrete-time DSGE model in the non-singular case, $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of $\mathbf{y}_\tau$ at $\boldsymbol{\theta}_0$ for given $h > 0$ if and only if the system (3.2) has a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, and Proposition 2-NS then establishes the necessary and sufficient rank condition and necessary order condition based on $\boldsymbol{\Delta}(\boldsymbol{\theta}_0)$ from (3.17). For given $h > 0$, Assumptions 1, 2 and 4 in Komunjer and Ng (2011) are implied by our Assumptions 1, 2 and 4. Further, their Assumption 3, continuous differentiability of $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ from (2.17), follows from our Assumption 3 on the underlying continuous-time matrices $\boldsymbol{\Lambda}_c(\boldsymbol{\theta}) = (\mathbf{A}(\boldsymbol{\theta}), \mathbf{B}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$ from (2.3) and $\boldsymbol{\Sigma}_{\boldsymbol{v},h}(\boldsymbol{\theta})$. In our case, based on the continuous-time underlying DSGE model, $\boldsymbol{\Lambda}(\boldsymbol{\theta})$ is extended to $\boldsymbol{\Lambda}^+(\boldsymbol{\theta})$, and modified to $\boldsymbol{\Lambda}_a(\boldsymbol{\theta})$, $\boldsymbol{\Lambda}_c(\boldsymbol{\theta})$. Correspondingly, $\boldsymbol{\delta}(\boldsymbol{\theta}, \mathbf{T}_h)$ is extended to $\boldsymbol{\delta}^+(\boldsymbol{\theta}, \mathbf{T}_h)$ in (3.11), and modified to $\boldsymbol{\delta}_a(\boldsymbol{\theta}, \mathbf{T}_h)$ in (3.15), and $\boldsymbol{\delta}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h)$ in (3.13). Continuous differentiability in $\boldsymbol{\theta}$ of $\boldsymbol{\Lambda}^+(\boldsymbol{\theta})$, $\boldsymbol{\Lambda}_a(\boldsymbol{\theta})$, and $\boldsymbol{\Lambda}_c(\boldsymbol{\theta})$ follows from our Assumption 3 on the underlying continuous-time matrices $\boldsymbol{\Lambda}_c(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_{\boldsymbol{v},h}(\boldsymbol{\theta})$. By Lemma 3, $\boldsymbol{\theta}$ is locally identifiable from the autocovariances of y_τ at $\boldsymbol{\theta}_0$ for given $h > 0$ if and only if $\boldsymbol{\delta}^+(\boldsymbol{\theta}_\ell, \mathbf{T}_h) =$

$\boldsymbol{\delta}^+(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ (i.e., (3.12)) has a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, which in the absence of local aliases (by Lemma 3.(b)) occurs if and only if $\boldsymbol{\delta}_a(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = \boldsymbol{\delta}(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ (i.e., (3.16)) has a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$, and (by Lemma 3.(c)) under the additional conditions on $\mathbf{C}^s(\boldsymbol{\theta}_0)$, $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\cdot)$ if and only if $\boldsymbol{\delta}_c(\boldsymbol{\theta}_\ell, \mathbf{T}_h, \mathbf{U}_h) = \boldsymbol{\delta}_c(\boldsymbol{\theta}_0, \mathbf{I}_{n_x}, \mathbf{I}_{m_w})$ (i.e., (3.14)) has a locally unique solution at $(\boldsymbol{\theta}_\ell, \mathbf{T}, \mathbf{U}) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x}, \mathbf{I}_{m_w})$. Therefore, the necessary and sufficient rank conditions and necessary order conditions based on the similarly extended and modified derivative matrices, $\boldsymbol{\Delta}^+(\boldsymbol{\theta}_0)$ from (3.18), $\boldsymbol{\Delta}_a(\boldsymbol{\theta}_0)$ from (3.19), and $\boldsymbol{\Delta}_c(\boldsymbol{\theta}_0)$ from (3.20), follow along the lines of Komunjer and Ng (2011). As discussed in Remark 3.1, although we consider the non-singular case, $\mathbf{U}_h$ appears as an additional argument in $\boldsymbol{\delta}_c$, in similarity with the analysis of the singular case in Komunjer and Ng (2011), the reason being that the implications of observational equivalence that we exploit stem from the underlying continuous-time model, with $\mathbf{B}(\boldsymbol{\theta}_0)$ locally identified if and only if $(\mathbf{T}_h, \mathbf{U}_h) = (\mathbf{I}_{n_x}, \mathbf{I}_{m_w})$ are the only local similarity transform and factor rotation.

For completeness, we present the relevant partial derivatives of $\boldsymbol{\delta}^+(\boldsymbol{\theta}, \mathbf{T}_h)$, $\boldsymbol{\delta}_a(\boldsymbol{\theta}, \mathbf{T}_h)$, and $\boldsymbol{\delta}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h)$, as well as the constancy of rank argument, as applied to our case. Thus, for part (a) of the Proposition,

$$\frac{\partial \boldsymbol{\delta}^+(\boldsymbol{\theta}, \mathbf{T}_h)}{\partial \boldsymbol{\theta}} = \begin{pmatrix} ((\mathbf{T}_h^{-1})^\top \otimes \mathbf{I}_{n_y^s}) \frac{\partial \mathrm{vec}\mathbf{C}^s(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ ((\mathbf{T}_h^{-1})^\top \otimes \mathbf{T}_h) \frac{\partial \mathrm{vec}\mathsf{A}(\boldsymbol{\theta};h)}{\partial \boldsymbol{\theta}} \\ (\mathbf{I}_{n_y} \otimes \mathbf{T}_h) \frac{\partial \mathrm{vec}\mathsf{K}(\boldsymbol{\theta};h)}{\partial \boldsymbol{\theta}} \\ ((\mathbf{T}_h^{-1})^\top \otimes \mathbf{I}_{n_y}) \frac{\partial \mathrm{vec}\mathsf{C}(\boldsymbol{\theta};h)}{\partial \boldsymbol{\theta}} \\ \frac{\partial \mathrm{vech}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)}{\partial \boldsymbol{\theta}} \end{pmatrix}, \tag{A.33}$$

$$\frac{\partial \boldsymbol{\delta}^+(\boldsymbol{\theta}, \mathbf{T}_h)}{\partial \mathrm{vec}(\mathbf{T}_h)} = \begin{pmatrix} -((\mathbf{T}_h^{-1})^\top \otimes \mathbf{I}_{n_y^s})[\mathbf{I}_{n_x} \otimes \mathbf{C}^s(\boldsymbol{\theta})](\mathbf{I}_{n_x} \otimes \mathbf{T}_h^{-1}) \\ ((\mathbf{T}_h^{-1})^\top \otimes \mathbf{T}_h)[\mathsf{A}(\boldsymbol{\theta};h) \otimes \mathbf{I}_{n_x} - \mathbf{I}_{n_x} \otimes \mathsf{A}(\boldsymbol{\theta};h)](\mathbf{I}_{n_x} \otimes \mathbf{T}_h^{-1}) \\ (\mathbf{I}_{n_y} \otimes \mathbf{T}_h)[\mathsf{K}(\boldsymbol{\theta};h)^\top \otimes \mathbf{I}_{n_x}](\mathbf{I}_{n_y} \otimes \mathbf{T}_h^{-1}) \\ -((\mathbf{T}_h^{-1})^\top \otimes \mathbf{I}_{n_y})[\mathbf{I}_{n_x} \otimes \mathsf{C}(\boldsymbol{\theta};h)](\mathbf{I}_{n_x} \otimes \mathbf{T}_h^{-1}) \\ \mathbf{0}_{n_y(n_y+1)/2 \times n_x^2} \end{pmatrix}. \tag{A.34}$$

From the discussion around (3.17) and (3.18), the derivatives (A.33) and (A.34) constitute the submatrices of the Jacobian $\mathcal{J}^+(\boldsymbol{\theta}, \mathbf{T}_h)$. Inserting $(\boldsymbol{\theta}, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ yields $\boldsymbol{\Delta}_h^+(\boldsymbol{\theta}_0)$ from (3.18). Note that $\mathcal{J}^+(\boldsymbol{\theta}, \mathbf{T}_h)$ can be alternatively written as $\mathbf{M}^+(\mathbf{T}_h)\boldsymbol{\Delta}^+(\boldsymbol{\theta})\mathbf{N}^+(\mathbf{T}_h)$, where $\mathbf{M}^+(\mathbf{T}_h)$ and $\mathbf{N}^+(\mathbf{T}_h)$ are square block diagonal matrices of dimensions $n_y^s n_x + n_x^2 + 2n_x n_y + n_y(n_y+1)/2$ and $m_\theta + n_x^2$, respectively, defined by

$$\mathbf{M}^+(\mathbf{T}_h) = \begin{pmatrix} (\mathbf{T}_h^{-1})^\top \otimes \mathbf{I}_{n_y^s} & & & & & \\ & (\mathbf{T}_h^{-1})^\top \otimes \mathbf{T}_h & & & & \\ & & \mathbf{I}_{n_y} \otimes \mathbf{T}_h & & & \\ & & & (\mathbf{T}_h^{-1})^\top \otimes \mathbf{I}_{n_y} & & \\ & & & & \mathbf{I}_{n_y(n_y+1)/2} & \end{pmatrix},$$

$$\mathbf{N}^{+}(\mathbf{T}_h) = \begin{pmatrix} \mathbf{I}_{n_\theta} & \\ & \mathbf{I}_{n_x} \otimes \mathbf{T}_h^{-1} \end{pmatrix}.$$

Since $\mathbf{T}_h$ is full rank, so are $\mathbf{M}^{+}(\mathbf{T}_h)$ and $\mathbf{N}^{+}(\mathbf{T}_h)$. Therefore, rank $\mathcal{J}^{+}(\boldsymbol{\theta}, \mathbf{T}_h) =$ rank $\boldsymbol{\Delta}^{+}(\boldsymbol{\theta})$. In particular, if rank $\boldsymbol{\Delta}^{+}(\boldsymbol{\theta})$ remains constant in a neighborhood of $\boldsymbol{\theta}_0$ (this is the assumption that $\boldsymbol{\theta}_0$ is a regular point of $\boldsymbol{\Delta}^{+}(\boldsymbol{\theta})$), then rank $\mathcal{J}^{+}(\boldsymbol{\theta}, \mathbf{T}_h)$ remains constant in a neighborhood of $(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$. Thus, full column rank of $\mathcal{J}^{+}(\boldsymbol{\theta}, \mathbf{T}_h)$ in a neighborhood of $(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ is equivalent to the rank condition (3.21). The order condition (3.22) simply requires at least as many rows as columns in $\boldsymbol{\Delta}^{+}(\boldsymbol{\theta})$, i.e., $n_y^s n_x + n_x^2 + 2n_x n_y + n_y(n_y+1)/2 \geq m_\theta + n_x^2$, so that full column rank is feasible.

For part (b) of the Proposition,

$$\frac{\partial \boldsymbol{\delta}_a(\boldsymbol{\theta}, \mathbf{T}_h)}{\partial \boldsymbol{\theta}} = \begin{pmatrix} ((\mathbf{T}^{-1})^\top \otimes \mathbf{T}) \frac{\partial \mathrm{vec}\mathbf{A}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ ((\mathbf{T}^{-1})^\top \otimes \mathbf{I}_{n_y}) \frac{\partial \mathrm{vec}\mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ (\mathbf{I}_{n_y} \otimes \mathbf{T}_h) \frac{\partial \mathrm{vec}\mathsf{K}(\boldsymbol{\theta}; h)}{\partial \boldsymbol{\theta}} \\ \frac{\partial \mathrm{vech}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta}; h)}{\partial \boldsymbol{\theta}} \end{pmatrix}, \tag{A.35}$$

$$\frac{\partial \boldsymbol{\delta}_a(\boldsymbol{\theta}, \mathbf{T}_h)}{\partial \mathrm{vec}(\mathbf{T}_h)} = \begin{pmatrix} ((\mathbf{T}^{-1})^\top \otimes \mathbf{T})[\mathbf{A}(\boldsymbol{\theta})^\top \otimes \mathbf{I}_{n_x} - \mathbf{I}_{n_x} \otimes \mathbf{A}(\boldsymbol{\theta})](\mathbf{I}_{n_x} \otimes \mathbf{T}^{-1}) \\ -((\mathbf{T}^{-1})^\top \otimes \mathbf{I}_{n_y})[\mathbf{I}_{n_x} \otimes \mathbf{C}(\boldsymbol{\theta})](\mathbf{I}_{n_x} \otimes \mathbf{T}^{-1}) \\ (\mathbf{I}_{n_y} \otimes \mathbf{T}_h)[\mathsf{K}(\boldsymbol{\theta}; h)^\top \otimes \mathbf{I}_{n_x}](\mathbf{I}_{n_y} \otimes \mathbf{T}_h^{-1}) \\ \mathbf{0}_{n_y(n_y+1)/2 \times n_x^2} \end{pmatrix}. \tag{A.36}$$

From the discussion around (3.19), the derivatives (A.35) and (A.36) constitute the submatrices of the Jacobian $\mathcal{J}_a(\boldsymbol{\theta}, \mathbf{T}_h)$. Inserting $(\boldsymbol{\theta}, \mathbf{T}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ yields $\boldsymbol{\Delta}_a(\boldsymbol{\theta}_0)$ from (3.19). Note that $\mathcal{J}_a(\boldsymbol{\theta}, \mathbf{T}_h)$ can be alternatively written as $\mathbf{M}_a(\mathbf{T}_h)\boldsymbol{\Delta}_a(\boldsymbol{\theta})\mathbf{N}_a(\mathbf{T}_h)$, where $\mathbf{M}_a(\mathbf{T}_h)$ and $\mathbf{N}_a(\mathbf{T}_h)$ are square block diagonal matrices of dimensions $n_x^2 + 2n_x n_y + n_y(n_y+1)/2$ and $m_\theta + n_x^2$, respectively, defined by

$$\mathbf{M}_a(\mathbf{T}_h) = \begin{pmatrix} (\mathbf{T}^{-1})^\top \otimes \mathbf{T} & & & \\ & (\mathbf{T}^{-1})^\top \otimes \mathbf{I}_{n_y} & & \\ & & \mathbf{I}_{n_y} \otimes \mathbf{T}_h & \\ & & & \mathbf{I}_{n_y(n_y+1)/2} \end{pmatrix},$$

$$\mathbf{N}_a(\mathbf{T}_h) = \begin{pmatrix} \mathbf{I}_{n_\theta} & \\ & \mathbf{I}_{n_x} \otimes \mathbf{T}_h^{-1} \end{pmatrix}.$$

Since $\mathbf{T}_h$ is full rank, so are $\mathbf{M}_a(\mathbf{T}_h)$ and $\mathbf{N}_a(\mathbf{T}_h)$ (note, $\mathbf{N}_a(\mathbf{T}_h) = \mathbf{N}^{+}(\mathbf{T}_h)$). Therefore, rank $\mathcal{J}_a(\boldsymbol{\theta}, \mathbf{T}_h) =$ rank $\boldsymbol{\Delta}_a(\boldsymbol{\theta})$. In particular, if rank $\boldsymbol{\Delta}_a(\boldsymbol{\theta})$ remains constant in a neighborhood of $\boldsymbol{\theta}_0$ (this is the assumption that $\boldsymbol{\theta}_0$ is a regular point of $\boldsymbol{\Delta}_a(\boldsymbol{\theta})$), then rank $\mathcal{J}_a(\boldsymbol{\theta}, \mathbf{T}_h)$ remains constant in a neighborhood of $(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$. Thus, full column rank of $\mathcal{J}_a(\boldsymbol{\theta}, \mathbf{T}_h)$ in a neighborhood of $(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ is equivalent to the rank condition (3.23). The order condition (3.24) simply requires at least as many rows as columns in $\boldsymbol{\Delta}_a(\boldsymbol{\theta})$, i.e.,

$n_x^2 + 2n_xn_y + n_y(n_y+1)/2 \geq m_\theta + n_x^2$, so that full column rank is feasible.

For part (c) of the Proposition,

$$\frac{\partial \boldsymbol{\delta}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h)}{\partial \boldsymbol{\theta}} = \begin{pmatrix} ((\mathbf{T}_h^{-1})^\top \otimes \mathbf{T}_h)\frac{\partial \text{vec}\mathbf{A}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ (\mathbf{U}_h^\top \otimes \mathbf{T}_h)\frac{\partial \text{vec}\mathbf{B}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ ((\mathbf{T}_h^{-1})^\top \otimes \mathbf{I}_{n_y})\frac{\partial \text{vec}\mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{pmatrix}, \tag{A.37}$$

$$\frac{\partial \boldsymbol{\delta}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h)}{\partial \text{vec}(\mathbf{T}_h)} = \begin{pmatrix} ((\mathbf{T}_h^{-1})^\top \otimes \mathbf{T}_h)[\mathbf{A}(\boldsymbol{\theta})^\top \otimes \mathbf{I}_{n_x} - \mathbf{I}_{n_x} \otimes \mathbf{A}(\boldsymbol{\theta})](\mathbf{I}_{n_x} \otimes \mathbf{T}_h^{-1}) \\ (\mathbf{U}_h^\top \otimes \mathbf{T}_h)(\mathbf{B}(\boldsymbol{\theta})^\top \otimes \mathbf{I}_{n_x})(\mathbf{I}_{n_x} \otimes \mathbf{T}_h^{-1}) \\ -((\mathbf{T}_h^{-1})^\top \otimes \mathbf{I}_{n_y})[\mathbf{I}_{n_x} \otimes \mathbf{C}(\boldsymbol{\theta})](\mathbf{I}_{n_x} \otimes \mathbf{T}_h^{-1}) \end{pmatrix}, \tag{A.38}$$

$$\frac{\partial \boldsymbol{\delta}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h)}{\partial \text{vec}(\mathbf{U}_h)} = \begin{pmatrix} \mathbf{0}_{n_x^2 \times m_w^2} \\ (\mathbf{U}_h^\top \otimes \mathbf{T}_h)(\mathbf{I}_{m_w} \otimes \mathbf{B}(\boldsymbol{\theta}))((\mathbf{U}_h^{-1})^\top \otimes \mathbf{I}_{m_w}) \\ \mathbf{0}_{n_xn_y \times m_w^2} \end{pmatrix}. \tag{A.39}$$

From the discussion around (3.20), the derivatives (A.37)-(A.39) constitute the first, middle, and last submatrix of the Jacobian $\mathcal{J}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h)$. Inserting $(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h) = (\boldsymbol{\theta}_0, \mathbf{I}_{n_x}, \mathbf{I}_{m_w})$ yields $\boldsymbol{\Delta}_c(\boldsymbol{\theta}_0)$ from (3.20). Note that $\mathcal{J}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h)$ can be written in the form $\mathbf{M}_c(\mathbf{T}_h, \mathbf{U}_h)\boldsymbol{\Delta}_c(\boldsymbol{\theta})\mathbf{N}_c(\mathbf{T}_h, \mathbf{U}_h)$, where $\mathbf{M}_c(\mathbf{T}_h, \mathbf{U}_h)$ and $\mathbf{N}_c(\mathbf{T}_h, \mathbf{U}_h)$ are square block diagonal matrices of dimensions $n_x^2+n_xn_y+n_xm_w$ and $m_\theta+n_x^2+m_w^2$, respectively, defined by

$$\mathbf{M}_c(\mathbf{T}_h, \mathbf{U}_h) = \begin{pmatrix} (\mathbf{T}_h^{-1})^\top \otimes \mathbf{T}_h & & \\ & \mathbf{U}_h^\top \otimes \mathbf{T}_h & \\ & & (\mathbf{T}_h^{-1})^\top \otimes \mathbf{I}_{n_y} \end{pmatrix},$$

$$\mathbf{N}_c(\mathbf{T}_h, \mathbf{U}_h) = \begin{pmatrix} \mathbf{I}_{n_\theta} & & \\ & \mathbf{I}_{n_x} \otimes \mathbf{T}_h^{-1} & \\ & & (\mathbf{U}_h^{-1})^\top \otimes \mathbf{I}_{m_w} \end{pmatrix}.$$

Since $\mathbf{T}_h$ and $\mathbf{U}_h$ are full rank, so are $\mathbf{M}_c(\mathbf{T}_h, \mathbf{U}_h)$ and $\mathbf{N}_c(\mathbf{T}_h, \mathbf{U}_h)$, so rank $\mathcal{J}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h) =$ rank $\boldsymbol{\Delta}_c(\boldsymbol{\theta})$. In particular, if rank $\boldsymbol{\Delta}_c(\boldsymbol{\theta})$ remains constant in a neighborhood of $\boldsymbol{\theta}_0$ (this is the assumption that $\boldsymbol{\theta}_0$ is a regular point of $\boldsymbol{\Delta}_c(\boldsymbol{\theta})$), then rank $\mathcal{J}_c(\boldsymbol{\theta}, \mathbf{T}_h, \mathbf{U}_h)$ remains constant in a neighborhood of $(\boldsymbol{\theta}_0, \mathbf{I}_{n_x}, \mathbf{I}_{m_w})$. Thus, full column rank of $\mathcal{J}_c(\boldsymbol{\theta}, \mathbf{T}_h)$ in a neighborhood of $(\boldsymbol{\theta}_0, \mathbf{I}_{n_x})$ is equivalent to the rank condition (3.25). The order condition (3.26) simply requires at least as many rows as columns in $\boldsymbol{\Delta}_c(\boldsymbol{\theta})$, i.e., $n_x^2+n_xn_y+n_xm_w \geq m_\theta + n_x^2 + m_w^2$, so that full column rank is feasible. ■

A.6 Proof of Proposition 3

Recall from (2.7) that $\boldsymbol{\eta}_\tau = \left[\boldsymbol{\eta}_\tau^{s,\top}, \boldsymbol{\eta}_\tau^{f,\top}\right]^\top$. For possibly large but finite n, the proof is constructed in three steps. First, we show that the Proposition holds for $\boldsymbol{\eta}^s$. Then, we do the same for $\boldsymbol{\eta}^f$ in isolation. Finally, the proof follows from stacking the two separate cases.

Step 1: Innovations to stock variables, $\boldsymbol{\eta}^s$. Consider $n \in \mathbb{Z}^+$ such that

$$\boldsymbol{\eta}^s_\tau = \sum_{i=1}^{n} \exp(\mathbf{A}(\boldsymbol{\theta})(t_\tau - t^\tau_{i-1}))\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_i) + o_P(1).$$

Expanding the sum on the right-hand side yields

$$\begin{aligned}\boldsymbol{\eta}^s_\tau = \exp(\mathbf{A}(\boldsymbol{\theta})h)\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_1) + \exp(\mathbf{A}(\boldsymbol{\theta})(h-h_n))\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_2) + \dots \\ \cdots + \exp(\mathbf{A}(\boldsymbol{\theta})(h-(n-2)h_n))\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_{n-1}) \\ + \exp(\mathbf{A}(\boldsymbol{\theta})(h-(n-1)h_n))\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_n) + o_P(1),\end{aligned}$$

which using (2.6) can be rewritten as

$$\begin{aligned}\boldsymbol{\eta}^s_\tau &= (\mathbf{I} + \mathbf{A}(\boldsymbol{\theta})h + \mathbf{A}(\boldsymbol{\theta})^2h^2/2 + \dots)\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_1) \\ &\quad + (\mathbf{I} + \mathbf{A}(\boldsymbol{\theta})(h-h_n) + \mathbf{A}(\boldsymbol{\theta})^2(h-h_n)^2/2 + \dots)\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_2) + \dots \\ &\quad \cdots + (\mathbf{I} + \mathbf{A}(\boldsymbol{\theta})(h-(n-2)h_n) + \mathbf{A}(\boldsymbol{\theta})^2(h-(n-2)h_n)^2/2 + \dots)\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_{n-1}) \\ &\quad + (\mathbf{I} + \mathbf{A}(\boldsymbol{\theta})(h-(n-1)h_n) + \mathbf{A}(\boldsymbol{\theta})^2(h-(n-1)h_n)^2/2 + \dots)\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_n) + o_P(1).\end{aligned}$$

Collecting terms with similar coefficients yields

$$\begin{aligned}\boldsymbol{\eta}^s_\tau &= \left(\mathbf{I} + \mathbf{A}(\boldsymbol{\theta})h + \mathbf{A}(\boldsymbol{\theta})^2(h^2/2) + \dots\right)\mathbf{B}(\boldsymbol{\theta})\sum_{i=1}^{n}\Delta\mathbf{w}(t^\tau_i) \\ &\quad + \left[\mathbf{A}(\boldsymbol{\theta})^2h_n^2/2 - \mathbf{A}(\boldsymbol{\theta})h_n - \mathbf{A}(\boldsymbol{\theta})^2hh_n - \dots\right]\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_2) + \dots \\ &\quad \cdots + \left[(n-2)^2\mathbf{A}(\boldsymbol{\theta})^2h_n^2/2 - \mathbf{A}(\boldsymbol{\theta})(n-2)h_n - (n-2)\mathbf{A}(\boldsymbol{\theta})^2hh_n + \dots\right]\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_{n-1}) \\ &\quad + \left[(n-1)^2\mathbf{A}(\boldsymbol{\theta})^2h_n^2/2 - \mathbf{A}(\boldsymbol{\theta})(n-1)h_n - (n-1)\mathbf{A}(\boldsymbol{\theta})^2hh_n + \dots\right]\mathbf{B}(\boldsymbol{\theta})\Delta\mathbf{w}(t^\tau_n) + o_P(1) \\ &= h^{1/2}\exp(\mathbf{A}(\boldsymbol{\theta})h)\mathbf{B}(\boldsymbol{\theta})\mathbf{u}_\tau + \mathcal{R}_\tau,\end{aligned}$$

where we have used the definitions of the matrix exponential in (2.6), of $\mathbf{u}_\tau$ in (4.1), and the fact that $t^\tau_0 = t_{\tau-1}$ and $t^\tau_n = t_\tau$.

The term $\mathcal{R}_\tau$ denotes the remainder of the approximation. By recalling that $h_n = h/n$, the properties of $\mathcal{R}_\tau$ can be characterized according to the following three limiting behaviors: (i) one for increasing number of sub-intervals $n \to \infty$, while keeping h constant; (ii) one for decreasing length between sub-intervals, $h_n \to 0$; (iii) and, analogously, one for increasing frequency of data, $h \to 0$. Note that shrinking h_n or h describes a similar behavior, that is, increasing data availability.

(i) As $n \to \infty$, it follows that $\mathbf{A}(\boldsymbol{\theta})h_n = \mathcal{O}(n^{-1})$, and the term $\mathbf{A}(\boldsymbol{\theta})(n-1)h_n \to \mathbf{A}(\boldsymbol{\theta})h$. Similarly, the term $\Delta\mathbf{w}(t^\tau_i) = \mathcal{O}_P(n^{-1/2})$ for all $1 < i \leq n$. Then, the remainder

$\mathcal{R}_\tau = \mathcal{O}_P(n^{-1/2})$, i.e., it is bounded in probability by $n^{-1/2}$, a scalar that decreases with increasing number of sub-intervals.

(ii) As $h_n \to 0$, it follows that $\mathbf{A}(\boldsymbol{\theta})h_n = \mathcal{O}(h_n)$, $\mathbf{A}(\boldsymbol{\theta})(n-1)h_n = \mathcal{O}(h_n)$, and $\Delta\mathbf{w}(t_i^\tau) = \mathcal{O}_P(h_n^{1/2})$. Then, $\mathcal{R}_\tau = \mathcal{O}_P(h_n^{3/2})$.

(iii) For $h \to 0$, it follows that $\mathbf{A}(\boldsymbol{\theta})h_n = \mathcal{O}(h)$, $\mathbf{A}(\boldsymbol{\theta})(n-1)h_n = \mathcal{O}(h)$, and $\Delta\mathbf{w}(t_i^\tau) = \mathcal{O}_P(h^{1/2})$, for all $1 < i \leq n$. Then, $\mathcal{R}_\tau = \mathcal{O}_P(h^{3/2})$.

Step 2: Innovations to flow variables, $\boldsymbol{\eta^f}$. Consider $n \in \mathbb{Z}^+$ such that

$$\boldsymbol{\eta}_\tau^f = \mathbf{S}_x^f \mathbf{A}(\boldsymbol{\theta})^{-1} \sum_{i=1}^{n} [\exp(\mathbf{A}(\boldsymbol{\theta})(t_\tau - t_{i-1}^\tau)) - \mathbf{I}_{n_x^s}] \mathbf{B}(\boldsymbol{\theta}) \Delta\mathbf{w}(t_i^\tau) + o_P(1),$$

where $\mathbf{S}_x^f$ is a selection matrix. By proceeding as in Step 1, we arrive at

$$\boldsymbol{\eta}_\tau^f = h^{1/2} \mathbf{S}_x^f \mathbf{A}(\boldsymbol{\theta})^{-1} \left[\exp(\mathbf{A}(\boldsymbol{\theta})h) - \mathbf{I}_{n_x^s}\right] \mathbf{B}(\boldsymbol{\theta}) \mathbf{u}_\tau + \mathcal{R}_\tau$$

as $h \to 0$. The properties of the remainder term $\mathcal{R}_\tau$ in the expansion of $\boldsymbol{\eta}_\tau^f$ are the same as those for the expansion $\boldsymbol{\eta}_\tau^s$ in Step 1.

Step 3: Innovations to both stock and flow variables, $\boldsymbol{\eta}$. Stack $\boldsymbol{\eta}_\tau^s$ and $\boldsymbol{\eta}_\tau^f$ to obtain

$$\boldsymbol{\eta}_\tau = h^{1/2} \begin{bmatrix} \exp(\mathbf{A}(\boldsymbol{\theta})h)\mathbf{B}(\boldsymbol{\theta}) \\ \mathbf{S}_x^f \mathbf{A}(\boldsymbol{\theta})^{-1} \left[\exp(\mathbf{A}(\boldsymbol{\theta})h) - \mathbf{I}_{n_x^s}\right] \mathbf{B}(\boldsymbol{\theta}) \end{bmatrix} \mathbf{u}_\tau + \mathcal{R}_\tau,$$

where it is straightforward to show that

(i) $\mathcal{R}_\tau = \mathcal{O}_P(n^{-1/2})$ as $n \to \infty$

(ii) $\mathcal{R}_\tau = \mathcal{O}_P(h_n^{3/2})$ as $h_n \to 0$

(iii) $\mathcal{R}_\tau = \mathcal{O}_P(h^{3/2})$ as $h \to 0$. ■

A.7 Proof of Proposition 4

Using Proposition 3, and given that $\mathbf{H}(\boldsymbol{\theta}) = \mathcal{O}(h^{3/2})$, it follows that $\mathbf{H}(\boldsymbol{\theta})^{-1}\mathcal{R}_\tau = \mathcal{O}_P(1)$. ■

B. Matrix computations

B.1 Computation of $\mathbf{A}_h(\boldsymbol{\theta})$

From Proposition 1, the $m_x \times m_x$ matrix $\mathbf{A}_h(\boldsymbol{\theta})$ is defined as

$$\mathbf{A}_h(\boldsymbol{\theta}) = \exp\left(\mathbf{A}(\boldsymbol{\theta})h\right) = \mathbf{I} + \mathbf{A}(\boldsymbol{\theta})h + \tfrac{1}{2}\mathbf{A}^2(\boldsymbol{\theta})h^2 + \tfrac{1}{3!}\mathbf{A}^3(\boldsymbol{\theta})h^3 + \ldots,$$

where $\mathbf{A}^j(\boldsymbol{\theta})$ indicates right multiplication of j copies of the $m_x \times m_x$ matrix $\mathbf{A}(\boldsymbol{\theta})$. Assume $\mathbf{A}(\boldsymbol{\theta})$ is diagonalizable. Then, $\mathbf{A}(\boldsymbol{\theta})$ can be factorized as

$$\mathbf{A}(\boldsymbol{\theta}) = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1},$$

where $\mathbf{V}$ is a square matrix whose columns correspond to the eigenvectors of $\mathbf{A}(\boldsymbol{\theta})$ and $\boldsymbol{\Lambda}$ is a diagonal matrix whose elements are the corresponding eigenvalues,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_{m_x} \end{bmatrix}.$$

Therefore, the exponential matrix $\exp(\mathbf{A}(\boldsymbol{\theta})h)$ can be computed as

$$\exp(\mathbf{A}(\boldsymbol{\theta})h) = \mathbf{V}\exp(\boldsymbol{\Lambda}h)\mathbf{V}^{-1} = \mathbf{V}\begin{bmatrix} e^{\lambda_1 h} & 0 & \ldots & 0 \\ 0 & e^{\lambda_2 h} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & e^{\lambda_{m_x} h} \end{bmatrix}\mathbf{V}^{-1}.$$

B.2 Computation of $\boldsymbol{\Sigma}_{\boldsymbol{\eta},h}(\boldsymbol{\theta})$

Here, we show how to implement the matrix decomposition method in Van Loan (1978, Theorem 1) to compute the elements of the variance-covariance matrix of the reduced-form innovation, $\boldsymbol{\Sigma}_{\boldsymbol{\eta},h}(\boldsymbol{\theta})$.

Define the augmented $4m_x \times 4m_x$ upper block triangular matrix

$$\boldsymbol{\Xi}(\boldsymbol{\theta}) = \begin{bmatrix} -\mathbf{A}(\boldsymbol{\theta}) & \mathbf{I}_{m_x} & \mathbf{0}_{m_x} & \mathbf{0}_{m_x} \\ \mathbf{0}_{m_x} & -\mathbf{A}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}(\boldsymbol{\theta}) & \mathbf{0}_{m_x} \\ \mathbf{0}_{m_x} & \mathbf{0}_{m_x} & \mathbf{A}(\boldsymbol{\theta})^\top & \mathbf{I}_{m_x} \\ \mathbf{0}_{m_x} & \mathbf{0}_{m_x} & \mathbf{0}_{m_x} & \mathbf{0}_{m_x} \end{bmatrix},$$

with exponential

$$\exp(\boldsymbol{\Xi}(\boldsymbol{\theta})h) = \begin{bmatrix} \mathbf{F}_{1,h}(\boldsymbol{\theta}) & \mathbf{G}_{1,h}(\boldsymbol{\theta}) & \mathbf{H}_{1,h}(\boldsymbol{\theta}) & \mathbf{K}_{1,h}(\boldsymbol{\theta}) \\ \mathbf{0}_{m_x} & \mathbf{F}_{2,h}(\boldsymbol{\theta}) & \mathbf{G}_{2,h}(\boldsymbol{\theta}) & \mathbf{H}_{2,h}(\boldsymbol{\theta}) \\ \mathbf{0}_{m_x} & \mathbf{0}_{m_x} & \mathbf{F}_{3,h}(\boldsymbol{\theta}) & \mathbf{G}_{3,h}(\boldsymbol{\theta}) \\ \mathbf{0}_{m_x} & \mathbf{0}_{m_x} & \mathbf{0}_{m_x} & \mathbf{F}_{4,h}(\boldsymbol{\theta}) \end{bmatrix},$$

computed using the matrix exponential function `expm` in Matlab.[2] Then, it follows Equations (2.9)-(2.11) can be computed as

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s,h}(\boldsymbol{\theta}) = \mathbf{F}_{3,h}(\boldsymbol{\theta})^\top \mathbf{G}_{2,h}(\boldsymbol{\theta}), \tag{B.1}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}^s\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}) = \mathbf{F}_{3,h}(\boldsymbol{\theta})^\top \mathbf{H}_{2,h}(\boldsymbol{\theta}) \mathbf{S}_x^{f\top}, \tag{B.2}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta}) = \mathbf{S}_x^f \left(\left[\mathbf{F}_{3,h}(\boldsymbol{\theta})^\top \mathbf{K}_{1,h}(\boldsymbol{\theta}) \right] + \left[\mathbf{F}_{3,h}(\boldsymbol{\theta}) \mathbf{K}_{1,h}(\boldsymbol{\theta})^\top \right] \right) \mathbf{S}_x^{f\top}, \tag{B.3}$$

where

$$\begin{aligned} \mathbf{F}_{3,h}(\boldsymbol{\theta}) &= \exp(\mathbf{A}(\boldsymbol{\theta})^\top h) = \mathbf{A}_h(\boldsymbol{\theta})^\top \\ \mathbf{G}_{2,h}(\boldsymbol{\theta}) &= \int_0^h \exp\left(-\mathbf{A}(\boldsymbol{\theta})(h-u)\right) \boldsymbol{\Sigma}(\boldsymbol{\theta}) \exp\left(\mathbf{A}(\boldsymbol{\theta})^\top u\right) \mathrm{d}u \\ \mathbf{H}_{2,h}(\boldsymbol{\theta}) &= \mathbf{A}_h(\boldsymbol{\theta})^{-1} \int_0^h \int_0^u \exp\left(\mathbf{A}(\boldsymbol{\theta})u\right) \boldsymbol{\Sigma}(\boldsymbol{\theta}) \exp\left(\mathbf{A}(\boldsymbol{\theta})^\top r\right) \mathrm{d}r\mathrm{d}u \\ \mathbf{K}_{1,h}(\boldsymbol{\theta}) &= \mathbf{A}_h(\boldsymbol{\theta})^{-1} \int_0^h \int_0^u \int_0^r \exp\left(\mathbf{A}(\boldsymbol{\theta})r\right) \boldsymbol{\Sigma}(\boldsymbol{\theta}) \exp\left(\mathbf{A}(\boldsymbol{\theta})^\top v\right) \mathrm{d}v\mathrm{d}r\mathrm{d}u. \end{aligned}$$

The computation of $\boldsymbol{\Sigma}_{\boldsymbol{\eta}^f,h}(\boldsymbol{\theta})$ uses the fact that if $\mathbf{M}(r,v) = e^{\mathbf{A}r}\boldsymbol{\Sigma}e^{\mathbf{A}^\top v}$, then

$$\int_0^u \int_0^r \left[\mathbf{M}(r,v) + \mathbf{M}(v,r)\right] \mathrm{d}v\mathrm{d}r = \int_0^u \int_0^u \mathbf{M}(r,v) \mathrm{d}v\mathrm{d}r.$$

[2] Moler and Van Loan (1978, 2003) and Jewitt and McCrorie (2005) provide a comparison of a variety of methods for computing the matrix exponential.

C. Alternative discretization

In this section we derive an alternative discretization of the flow states variables that involves both autoregressive dynamics and a moving average (MA) component. We show that resulting discretized system is equivalent to the Exact Discrete State Space Representation (ED-SSR) in the main text.

Let $\mathbf{y}^f(t)$ and $\mathbf{y}^s(t)$ denote the time t measurements sampled, respectively, as flows and stocks at observation times. With $\mathbf{S}_x^f = \mathbf{I}_{m_x}$, the state space system (2.1)-(2.2) can be equivalently written as

$$\begin{bmatrix} \mathrm{d}\mathbf{x}^s(t) \\ \mathrm{d}\mathbf{x}^f(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{A}(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \mathbf{x}^s(t) \\ \mathbf{x}^f(t) \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \mathbf{B}(\boldsymbol{\theta}) \\ \mathbf{B}(\boldsymbol{\theta}) \end{bmatrix} \mathrm{d}\mathbf{w}(t)\,,$$
$$\begin{bmatrix} \mathbf{y}^s(t) \\ \mathbf{y}^f(t) \end{bmatrix} = \begin{bmatrix} \mathbf{S}_y^s\mathbf{C}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_y^f\mathbf{C}(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \mathbf{x}^s(t) \\ \mathbf{x}^f(t) \end{bmatrix}.$$

Taking integrals, in the Itô sense, on the left and right-hand side for $\mathrm{d}\mathbf{x}^f(t)$, we have

$$\int_{t_\tau-1}^{t_\tau} \mathrm{d}\mathbf{x}^f(v) = \mathbf{A}(\boldsymbol{\theta}) \int_{t_\tau-1}^{t_\tau} \mathbf{x}^f(v)\mathrm{d}v + \mathbf{B}(\boldsymbol{\theta}) \int_{t_\tau-1}^{t_\tau} \mathrm{d}\mathbf{w}(v)\,. \tag{C.1}$$

If Assumption 1 in the main text holds, $\mathbf{A}(\boldsymbol{\theta})$ is invertible, and Equation (C.1) can be reorganized as

$$\int_{t_{\tau-1}}^{t_\tau} \mathbf{x}^f(v)\mathrm{d}v = \mathbf{A}(\boldsymbol{\theta})^{-1} \int_{t_\tau-1}^{t_\tau} \mathrm{d}\mathbf{x}^f(v) - \mathbf{A}(\boldsymbol{\theta})^{-1}\mathbf{B}(\boldsymbol{\theta}) \int_{t_{\tau-1}}^{t_\tau} \mathrm{d}\mathbf{w}(v)\,.$$

Further, notice that by definition of the Itô's integral,

$$\int_{t_\tau-1}^{t_\tau} \mathrm{d}\mathbf{x}^f(v) = \bar{\mathbf{x}}^f(t_\tau) - \bar{\mathbf{x}}^f(t_{\tau-1})\,, \tag{C.2}$$

where $\bar{\mathbf{x}}^f(t_\tau) = \mathbf{A}_h(\boldsymbol{\theta})\bar{\mathbf{x}}^f(t_{\tau-1}) + \int_0^h \exp(\mathbf{A}(\boldsymbol{\theta})v)\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(t_{\tau-1}+v)$ (see Appendix A.1). Then we may write

$$\begin{aligned}\int_{t_\tau-1}^{t_\tau} \mathrm{d}\mathbf{x}^f(v)\mathrm{d}v = \mathbf{A}_h(\boldsymbol{\theta})[\bar{\mathbf{x}}^f(t_{\tau-1}) - \bar{\mathbf{x}}^f(t_{\tau-2})] \\ + \Bigg[\int_{t_{\tau-1}}^{t_\tau} \exp(\mathbf{A}(\boldsymbol{\theta})(t_\tau - v))\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(v) \\ - \int_{t_{\tau-2}}^{t_{\tau-1}} \exp(\mathbf{A}(\boldsymbol{\theta})(t_{\tau-1} - v))\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(v)\Bigg]\,.\end{aligned}$$

Combining (C.1) and (C.2) yields

$$\bar{\mathbf{x}}^f(t_{\tau-1}) - \bar{\mathbf{x}}^f(t_{\tau-2}) = \int_{t_{\tau-2}}^{t_{\tau-1}} \mathrm{d}\mathbf{x}^f(v) = \mathbf{A}(\boldsymbol{\theta}) \int_{t_{\tau-2}}^{t_{\tau-1}} \mathbf{x}^f(v)\mathrm{d}v + \int_{t_{\tau-2}}^{t_{\tau-1}} \mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(v)\,.$$

This, in turn, means

$$\begin{aligned}\int_{t_\tau-1}^{t_\tau} \mathbf{x}^f(v)\mathrm{d}v = \mathbf{A}(\boldsymbol{\theta})^{-1}\mathbf{A}_h(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) \int_{t_{\tau-2}}^{t_{\tau-1}} \mathbf{x}^f(v)\mathrm{d}v \\ + \mathbf{A}(\boldsymbol{\theta})^{-1}\mathbf{A}_h(\boldsymbol{\theta}) \int_{t_{\tau-2}}^{t_{\tau-1}} \mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(v) \\ + \mathbf{A}(\boldsymbol{\theta})^{-1} \int_{t_{\tau-1}}^{t_\tau} \exp(\mathbf{A}(\boldsymbol{\theta})(t_\tau - v))\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(v) \\ - \mathbf{A}(\boldsymbol{\theta})^{-1} \int_{t_{\tau-2}}^{t_{\tau-1}} \exp(\mathbf{A}(\boldsymbol{\theta})(t_{\tau-1} - v))\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(v) \\ - \mathbf{A}(\boldsymbol{\theta})^{-1}\mathbf{B}(\boldsymbol{\theta}) \int_{t_{\tau-1}}^{t_\tau} \mathrm{d}\mathbf{w}(v)\,,\end{aligned}$$

where we used $\mathrm{e}^{\mathbf{CAC}^{-1}} = \mathbf{C}\mathrm{e}^{\mathbf{A}}\mathbf{C}^{-1}$. Aggregating the terms with the same random measures, we obtain

$$\begin{aligned}\int_{t_\tau-1}^{t_\tau} \mathbf{x}^f(v)\mathrm{d}v = \mathbf{A}_h(\boldsymbol{\theta}) \int_{t_{\tau-2}}^{t_{\tau-1}} \mathbf{x}^f(s)\mathrm{d}s \\ + \int_{t_{\tau-1}}^{t_\tau} \mathbf{A}(\boldsymbol{\theta})^{-1} \left(\exp(\mathbf{A}(\boldsymbol{\theta})(t_\tau - v)) - \mathbf{I}\right) \mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(v) \\ + \int_{t_{\tau-2}}^{t_{\tau-1}} \mathbf{A}(\boldsymbol{\theta})^{-1} \big(\exp(\mathbf{A}(\boldsymbol{\theta})h) - \exp\left(\mathbf{A}(\boldsymbol{\theta})(t_{\tau-1} - v)\right)\big)\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(v)\,,\end{aligned}$$

which, using the notation of the paper, can be compactly written in the VARMA(1,1) system form

$$\mathbf{x}_\tau^f = \mathbf{A}_h(\boldsymbol{\theta})\mathbf{x}_{\tau-1}^f + \boldsymbol{\varsigma}_\tau^f\,, \tag{C.3}$$

where the vector of reduced-form residuals $\boldsymbol{\varsigma}_\tau^f = \boldsymbol{\eta}_\tau^{f,1} + \mathbf{G}(\boldsymbol{\theta})\boldsymbol{\eta}_{\tau-1}^{f,2}$ follows a vector moving average process or order one (VMA(1)). Notice that $\boldsymbol{\eta}_\tau^{f,1}$ can be rewritten as

$$\boldsymbol{\eta}_\tau^{f,1} = \int_{t_{\tau-1}}^{t_\tau} \mathbf{A}(\boldsymbol{\theta})^{-1} \big(\exp\left(\mathbf{A}(\boldsymbol{\theta})(t_\tau - u)\right) - \mathbf{I}_{m_x}\big)\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(u)\,,$$

while $\mathbf{G}(\boldsymbol{\theta})\boldsymbol{\eta}_{\tau-1}^{f,2}$ is equivalent to

$$\mathbf{G}(\boldsymbol{\theta})\boldsymbol{\eta}_{\tau-1}^{f,2} = \int_{t_{\tau-2}}^{t_{\tau-1}} \mathbf{A}(\boldsymbol{\theta})^{-1} \big(\exp\left(\mathbf{A}(\boldsymbol{\theta})(t_{\tau-1} - u)\right) - \mathbf{I}_{m_x}\big)\mathbf{B}(\boldsymbol{\theta})\mathrm{d}\mathbf{w}(u)\,.$$

The autocovariances of the reduced-form residuals are given by

$$\mathbb{E}[\boldsymbol{\varsigma}_\tau^f \boldsymbol{\varsigma}_\tau^{f\top}] = 2\int_0^h \int_0^u \exp(\mathbf{A}(\boldsymbol{\theta})r)\mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top \exp(\mathbf{A}(\boldsymbol{\theta})^\top r)\mathrm{d}r\mathrm{d}u\,,$$

$$\mathbb{E}[\boldsymbol{\varsigma}_\tau^f \boldsymbol{\varsigma}_{\tau-1}^{f\top}] = \int_0^h \int_0^u \exp(\mathbf{A}(\boldsymbol{\theta})r)\mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top \exp(\mathbf{A}(\boldsymbol{\theta})^\top r)\mathrm{d}r\mathrm{d}u\,,$$

and $\mathbb{E}[\boldsymbol{\varsigma}_\tau \boldsymbol{\varsigma}_{\tau-\ell}^\top] = \mathbf{0}_{m_x}$, for all $\ell \neq -1, 0, 1$.

Finally, notice that $\boldsymbol{\eta}_\tau^{f,1}$ and $\mathbf{G}(\boldsymbol{\theta})\boldsymbol{\eta}_\tau^{f,2}$ are the same vector of reduced-form residuals that we find in the paper, i.e, $\boldsymbol{\eta}_\tau^f = \boldsymbol{\eta}_\tau^{f,1} = \mathbf{G}(\boldsymbol{\theta})\boldsymbol{\eta}_\tau^{f,2}$. Therefore, the discretized state space form with VARMA(1,1) transition can be represented as

$$\begin{bmatrix} \mathbf{x}_\tau \\ \mathbf{x}_\tau^f \\ \boldsymbol{\eta}_\tau^f \end{bmatrix} = \begin{bmatrix} \mathbf{A}_h(\boldsymbol{\theta}) & \mathbf{0}_{n_x^s \times n_x^f} & \mathbf{0}_{n_x^s \times n_x^f} \\ \mathbf{0}_{n_x^f \times n_x^s} & \mathbf{A}_h(\boldsymbol{\theta}) & \mathbf{I}_{n_x^f} \\ \mathbf{0}_{n_x^f \times n_x^s} & \mathbf{0}_{n_x^f \times n_x^f} & \mathbf{0}_{n_x^f \times n_x^f} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\tau-1} \\ \mathbf{x}_{\tau-1}^f \\ \boldsymbol{\eta}_{\tau-1}^f \end{bmatrix} + \begin{bmatrix} \mathbf{I}_{n_x^s} & \mathbf{0}_{n_x^s \times n_x^f} \\ \mathbf{0}_{n_x^f \times n_x^s} & \mathbf{I}_{n_x^f} \\ \mathbf{0}_{n_x^f \times n_x^s} & \mathbf{I}_{n_x^f} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_\tau^s \\ \boldsymbol{\eta}_\tau^f \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{y}_\tau^s \\ \mathbf{y}_\tau^f \end{bmatrix} = \begin{bmatrix} \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}) & \mathbf{0}_{n_y^s \times n_x^f} & \mathbf{0}_{n_y^s \times n_x^f} \\ \mathbf{0}_{n_y^s \times n_x^s} & \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) & \mathbf{0}_{n_y^f \times n_x^f} \end{bmatrix} \begin{bmatrix} \mathbf{x}_\tau \\ \mathbf{x}_\tau^f \\ \boldsymbol{\eta}_\tau^f \end{bmatrix},$$

with error covariances

$$\mathbb{E}[\boldsymbol{\eta}_\tau \boldsymbol{\eta}_\tau^\top] = \begin{bmatrix} \int_0^h \mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})v}\mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top \mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})^\top v}\mathrm{d}v & \int_0^h \int_0^u \mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})(u-r)}\mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top \mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})^\top r}\mathrm{d}r\mathrm{d}u \\ \int_0^h \int_0^u \mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})r}\mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top \mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})^\top (u-r)}\mathrm{d}r\mathrm{d}u & \int_0^h \int_0^u \mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})r}\mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top \mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})^\top r}\mathrm{d}r\mathrm{d}u \end{bmatrix}.$$

C.1 Equivalence result

We now show that our proposed approach intrinsically accounts for the larger and more convoluted VARMA(1,1) system (C.3). From (2.4), recall that

$$\mathbf{x}_\tau^f = \mathbf{A}(\boldsymbol{\theta})^{-1}(\mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})h} - \mathbf{I}_{m_x})\mathbf{x}_{\tau-1}^s + \boldsymbol{\eta}_\tau^f\,,$$

whereas in the approach derived above we have $\mathbf{x}_\tau^{f,B} = \mathbf{A}_h(\boldsymbol{\theta})\mathbf{x}_{\tau-1}^{f,B} + \boldsymbol{\eta}_\tau^{f,B} + \boldsymbol{\eta}_{\tau-1}^{f,B}$, where we used the superscript B to highlight that the equations in the alternative discretization are similar to those in Bergstrom (1984). Since the structural shocks are the same and are cumulated in the system through the same integrals, $\boldsymbol{\eta}_\tau^f = \boldsymbol{\eta}_\tau^{f,B}$ for all τ, the two approaches are equivalent if and only if

$$\mathbf{A}(\boldsymbol{\theta})^{-1}(\mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})h} - \mathbf{I}_{m_x})\mathbf{x}_{\tau-1}^s = \mathbf{A}_h(\boldsymbol{\theta})\mathbf{x}_{\tau-1}^{f,B} + \boldsymbol{\eta}_{\tau-1}^f\,.$$

According to the state space representation (2.4)-(2.5), $\mathbf{A}(\boldsymbol{\theta})^{-1}(\mathrm{e}^{\mathbf{A}(\boldsymbol{\theta})h} - \mathbf{I}_{m_x})\mathbf{x}_{\tau-1}^s = \mathbf{x}_\tau^f - \boldsymbol{\eta}_\tau^f$. Then, we obtain that

$$\mathbf{x}_\tau^f = \mathbf{A}_h(\boldsymbol{\theta})\mathbf{x}_{\tau-1}^{f,B} + \boldsymbol{\eta}_\tau^f + \boldsymbol{\eta}_{\tau-1}^f \,,$$

which is the formula we found using the approach of Bergstrom (1984), Equation (C.3). Therefore, it holds that

$$\mathbf{x}_\tau^f = \mathbf{x}_\tau^{f,B}.$$

This shows that when expressing $\mathbf{x}_\tau^f$ in terms of the latent $\mathbf{x}_{\tau-1}^s$, the MA component is inherently included.

D. The ABCD representation

Define by $\mathbf{S}_\upsilon$ the $n_y \times n_\upsilon$ selection matrix that for each row corresponding to a measurement observed with error selects the latter from $\boldsymbol{\upsilon}_t$. For example, if all measurements are observed with error, then $\mathbf{S}_\upsilon = \mathbf{I}_{n_\upsilon}$. If only the last measurement is observed with error, then $n_\upsilon = 1$, and $\mathbf{S}_\upsilon = (0, \dots, 0, 1)^\top$. The matrices of the (non-minimal) ABCD state space representation (2.12)-(2.13) are given by

$$\mathsf{A}(\boldsymbol{\theta}; h) := \begin{bmatrix} \mathbf{A}_h(\boldsymbol{\theta}) & \mathbf{0}_{n_x^s \times n_x^f} \\ \mathbf{S}_x^f \mathbf{A}(\boldsymbol{\theta})^{-1}(\mathbf{A}_h(\boldsymbol{\theta}) - \mathbf{I}_{n_x^s}) & \mathbf{0}_{n_x^f \times n_x^f} \end{bmatrix}, \quad \mathsf{B}(\boldsymbol{\theta}; h) := \begin{bmatrix} \mathbf{I}_{n_x} & \mathbf{0}_{n_x \times n_{\boldsymbol{\upsilon}}} \end{bmatrix},$$

$$\mathsf{C}(\boldsymbol{\theta}; h) := \begin{bmatrix} \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}) & \mathbf{0}_{n_y^s \times n_x^f} \\ \mathbf{0}_{n_y^f \times n_x^s} & \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) \mathbf{S}_x^{f\top} \end{bmatrix} \cdot \mathsf{A}(\boldsymbol{\theta}; h),$$

$$\mathsf{D}(\boldsymbol{\theta}; h) := \begin{bmatrix} \begin{pmatrix} \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}) & \mathbf{0}_{n_y^s \times n_x^f} \\ \mathbf{0}_{n_y^f \times n_x^s} & \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) \mathbf{S}_x^{f\top} \end{pmatrix} \cdot \mathsf{B}(\boldsymbol{\theta}, h) & \mathbf{S}_\upsilon \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}; h) := \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\eta},h}(\boldsymbol{\theta}) & \mathbf{0}_{n_x \times n_\upsilon} \\ \mathbf{0}_{n_\upsilon \times n_x} & \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}) \end{bmatrix}.$$

For the case of the Euler-Maruyama (EM) discretization, the matrices are given by

$$\mathsf{A}(\boldsymbol{\theta}; h) := \mathbf{I}_{n_x^s} + \mathbf{A}(\boldsymbol{\theta}) h, \quad \mathsf{B}(\boldsymbol{\theta}; h) := [\mathbf{I}_{n_x^s}, \mathbf{0}_{n_x^s \times n_\upsilon}], \quad \mathsf{C}(\boldsymbol{\theta}; h) := \mathbf{C}(\boldsymbol{\theta}) \cdot \mathsf{A}(\boldsymbol{\theta}; h),$$

$$\mathsf{D}(\boldsymbol{\theta}; h) := [\mathbf{C}(\boldsymbol{\theta}), \mathbf{S}_\upsilon], \quad \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}; h) := \begin{bmatrix} h\mathbf{B}(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})^\top & \mathbf{0}_{n_y \times n_x}^\top \\ \mathbf{0}_{n_y \times n_x} & \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta}) \end{bmatrix}.$$

The minimal system for the ABCD state space representation is instead given by

$$\mathbf{x}_\tau^s = \underbrace{\mathbf{A}_h(\boldsymbol{\theta})}_{:=\mathsf{A}(\boldsymbol{\theta};h)} \mathbf{x}_{\tau-1}^s + \underbrace{\begin{bmatrix} \mathbf{I}_{n_x^s} & \mathbf{0}_{n_x^s \times n_x^s} & \mathbf{0}_{n_x^s \times n_\upsilon} \end{bmatrix}}_{:=\mathsf{B}(\boldsymbol{\theta};h)} \underbrace{\begin{bmatrix} \boldsymbol{\eta}_\tau^s \\ \boldsymbol{\eta}_\tau^f \\ \boldsymbol{\upsilon}_\tau \end{bmatrix}}_{:=\boldsymbol{\epsilon}_\tau}, \tag{D.3}$$

$$\begin{bmatrix} \mathbf{y}_\tau^s \\ \mathbf{y}_\tau^f \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}) \mathbf{A}_h(\boldsymbol{\theta}) \\ \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta})^{-1}(\mathbf{A}_h(\boldsymbol{\theta}) - \mathbf{I}_{n_x^s}) \end{bmatrix}}_{:=\mathsf{C}(\boldsymbol{\theta};h)} \mathbf{x}_{\tau-1}^s + \underbrace{\begin{bmatrix} \mathbf{S}_y^s \mathbf{C}(\boldsymbol{\theta}) & \mathbf{0}_{n_y^s \times n_x^s} & \mathbf{S}_\upsilon \\ \mathbf{0}_{n_y^f \times n_x^s} & \mathbf{S}_y^f \mathbf{C}(\boldsymbol{\theta}) & \end{bmatrix}}_{:=\mathsf{D}(\boldsymbol{\theta};h)} \underbrace{\begin{bmatrix} \boldsymbol{\eta}_\tau^s \\ \boldsymbol{\eta}_\tau^f \\ \boldsymbol{\upsilon}_\tau \end{bmatrix}}_{:=\boldsymbol{\epsilon}_\tau}, \tag{D.4}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}; h) = \operatorname{diag}\big(\boldsymbol{\Sigma}_{\boldsymbol{\eta},h}(\boldsymbol{\theta}), \boldsymbol{\Sigma}_{\boldsymbol{\upsilon},h}(\boldsymbol{\theta})\big), \tag{D.5}$$

using $\mathbf{S}_x^f = \mathbf{I}_{n_x^s}$ in $\boldsymbol{\eta}_\tau^f$ from (2.7) and hence in $\boldsymbol{\Sigma}_{\boldsymbol{\eta},h}(\boldsymbol{\theta})$ in the minimal representation.

E. Kalman filter recursions and the likelihood function

As argued in Section 2.2, the state-space solution for the vector of observables $\mathbf{y}^T$ admits the ABCD representation

$$\begin{aligned}\mathbf{x}_{\tau+1} &= \mathsf{A}(\boldsymbol{\theta};h)\mathbf{x}_\tau + \mathsf{B}(\boldsymbol{\theta};h)\boldsymbol{\epsilon}_{\tau+1}\\ \mathbf{y}_{\tau+1} &= \mathsf{C}(\boldsymbol{\theta};h)\mathbf{x}_\tau + \mathsf{D}(\boldsymbol{\theta};h)\boldsymbol{\epsilon}_{\tau+1}\end{aligned}$$

where $\mathbb{E}[\boldsymbol{\epsilon}_\tau\boldsymbol{\epsilon}_\tau^\top] = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta};h)$, and the matrices $\mathsf{A}(\boldsymbol{\theta};h)$, $\mathsf{B}(\boldsymbol{\theta};h)$, $\mathsf{C}(\boldsymbol{\theta};h)$, $\mathsf{D}(\boldsymbol{\theta};h)$, $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta};h)$ are defined in Appendix D.

E.1 Kalman filter

Let $\mathbf{y}^{\tau-1} = \{\mathbf{y}_1,\ldots,\mathbf{y}_{\tau-1}\}$ denote the history of measurements up to time $t_{\tau-1}$. Additionally, let $\mathbf{x}_{\tau|\tau-1} = \mathbb{E}\left[\mathbf{x}_\tau|\mathbf{y}^{\tau-1}\right]$ denote the forecast of the state vector conditional on the information available at the time $t_{\tau-1}$, and $\mathbf{P}_{\tau|\tau-1} = \mathbb{E}\left[\left(\mathbf{x}_\tau - \mathbf{x}_{\tau|\tau-1}\right)\left(\mathbf{x}_\tau - \mathbf{x}_{\tau|\tau-1}\right)^\top\right]$ the corresponding forecast error covariance matrix. Similarly, let $\mathbf{y}_{\tau|\tau-1} = \mathbb{E}\left[\mathbf{y}_\tau|\mathbf{y}^{\tau-1}\right]$ denote the forecast of the control variables conditional on past information, and $\boldsymbol{\Sigma}_{\boldsymbol{\nu},\tau|\tau-1} = \mathbb{E}\left[\left(\mathbf{y}_\tau - \mathbf{y}_{\tau|\tau-1}\right)\left(\mathbf{y}_\tau - \mathbf{y}_{\tau|\tau-1}\right)^\top\right]$ its associated forecast error covariance matrix. By exploiting the linearity of (2.12)-(2.13), the forecast of the state variables and their associated variance-covariance are

$$\mathbf{x}_{\tau|\tau-1} = \mathsf{A}(\boldsymbol{\theta};h)\mathbf{x}_{\tau-1|\tau-1}, \tag{E.1}$$

$$\mathbf{P}_{\tau|\tau-1} = \mathsf{A}(\boldsymbol{\theta};h)\mathbf{P}_{\tau-1|\tau-1}\mathsf{A}(\boldsymbol{\theta};h)^\top + \mathsf{B}(\boldsymbol{\theta};h)\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta};h)\mathsf{B}(\boldsymbol{\theta};h)^\top, \tag{E.2}$$

given initial conditions $\mathbf{x}_0$ and $\mathbf{P}_0$. Since the state vector is stationary, by Lemma 1, we use as initial values its unconditional mean $\mathbf{x}_0 = \mathbf{x}_{1|0} = \mathbb{E}\left[\mathbf{x}_1\right] = \mathbf{0}$, and its unconditional covariance matrix

$$\begin{aligned}\text{vec}\left(\mathbf{P}_0\right) = \text{vec}\left(\mathbf{P}_{1|0}\right) &= \text{vec}\left(\mathbb{E}\left[\mathbf{x}_1\mathbf{x}_1^\top\right]\right)\\ &= \left[\mathbf{I}_{n_x^2} - \left(\mathsf{A}(\boldsymbol{\theta};h)\otimes\mathsf{A}(\boldsymbol{\theta};h)\right)\right]^{-1}\left[\mathsf{B}(\boldsymbol{\theta};h)\otimes\mathsf{B}(\boldsymbol{\theta};h)\right]\text{vec}\left(\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta};h)\right),\end{aligned}$$

where vec is the vectorization operator. Given the predictions for the state variables, the Kalman filter recursively computes the one-step-ahead forecast error of the control variables and associated variance-covariance matrix

$$\boldsymbol{\nu}_{\tau|\tau-1} = \mathbf{y}_\tau - \mathbf{y}_{\tau|\tau-1} = \mathbf{y}_\tau - \mathsf{C}(\boldsymbol{\theta};h)\mathbf{x}_{\tau-1|\tau-1} \tag{E.3}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\nu},\tau|\tau-1} = \mathsf{C}(\boldsymbol{\theta};h)\mathbf{P}_{\tau-1|\tau-1}\mathsf{C}(\boldsymbol{\theta};h)^\top + \mathsf{D}(\boldsymbol{\theta};h)\boldsymbol{\Sigma}_\epsilon(\boldsymbol{\theta};h)\mathsf{D}(\boldsymbol{\theta};h)^\top. \tag{E.4}$$

Using the information above, we update the state variables according to

$$\mathbf{x}_{\tau|\tau} = \mathbf{x}_{\tau|\tau-1} + \mathbf{K}_{\tau|\tau-1}\boldsymbol{\nu}_{\tau|\tau-1} \tag{E.5}$$

$$\mathbf{P}_{\tau|\tau} = \mathbf{P}_{\tau|\tau-1} - \mathbf{K}_{\tau|\tau-1}\boldsymbol{\Sigma}_{\boldsymbol{\nu},\tau|\tau-1}\mathbf{K}_{\tau|\tau-1}^{\top}, \tag{E.6}$$

where

$$\mathbf{K}_{\tau|\tau-1} = \mathsf{A}(\boldsymbol{\theta};h)\mathbf{P}_{\tau-1|\tau-1}\mathsf{C}(\boldsymbol{\theta};h)^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\nu},\tau|\tau-1}^{-1} \tag{E.7}$$

is the Kalman gain. Equations (E.1)-(E.7), together with initial conditions $\mathbf{x}_0$ and $\mathbf{P}_0$, define the *Kalman filter recursion* for $\tau = 1, 2, \ldots, T$. If convergence occurs at a given point $t_{\tau-1}$, then $\mathbf{K} = \mathbf{K}_{s+1|s} = \mathbf{K}_{s|s-1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}} = \boldsymbol{\Sigma}_{\boldsymbol{\nu},s+1|s} = \boldsymbol{\Sigma}_{\boldsymbol{\nu},s|s-1}$ for all $s \geq \tau \in \mathbb{N}$. We define $\mathbf{K} = \mathsf{K}(\boldsymbol{\theta};h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}} = \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)$ to be the two (time-variant) matrices, after convergence. Thus, they solve the system

$$\mathsf{K}(\boldsymbol{\theta};h) = \mathsf{A}(\boldsymbol{\theta};h)\mathbf{P}(\boldsymbol{\theta};h)\mathsf{C}(\boldsymbol{\theta};h)^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)^{-1}, \tag{E.8}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h) = \mathsf{C}(\boldsymbol{\theta};h)\mathbf{P}(\boldsymbol{\theta};h)\mathsf{C}(\boldsymbol{\theta};h)^{\top} + \mathsf{D}(\boldsymbol{\theta};h)\boldsymbol{\Sigma}_{\epsilon}(\boldsymbol{\theta};h)\mathsf{D}(\boldsymbol{\theta};h)^{\top}, \tag{E.9}$$

$$\begin{aligned}\mathbf{P}(\boldsymbol{\theta};h) = {} & \mathsf{A}(\boldsymbol{\theta};h)\left[\mathbf{P}(\boldsymbol{\theta};h) - \mathsf{K}(\boldsymbol{\theta};h)\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)\mathsf{K}(\boldsymbol{\theta};h)^{\top}\right]\mathsf{A}(\boldsymbol{\theta};h)^{\top} \\ & + \mathsf{B}(\boldsymbol{\theta};h)\boldsymbol{\Sigma}_{\epsilon}(\boldsymbol{\theta};h)\mathsf{B}(\boldsymbol{\theta};h)^{\top},\end{aligned} \tag{E.10}$$

where $\mathbf{P}(\boldsymbol{\theta};h)$ is $\mathbf{P}_{\tau|\tau-1}$ upon convergence.

E.2 The innovations representation

Under Assumptions 1 and 2, the ABCD representation admits the time-invariant innovations representation

$$\begin{aligned}\mathbf{x}_{\tau+1|\tau+1} &= \mathsf{A}(\boldsymbol{\theta};h)\mathbf{x}_{\tau|\tau} + \mathsf{K}(\boldsymbol{\theta};h)\boldsymbol{\nu}_{\tau+1|\tau} \\ \mathbf{y}_{\tau+1} &= \mathsf{C}(\boldsymbol{\theta};h)\mathbf{x}_{\tau|\tau} + \boldsymbol{\nu}_{\tau+1|\tau}\end{aligned}$$

where $\mathbb{E}[\boldsymbol{\nu}_{\tau+1|\tau}\boldsymbol{\nu}_{\tau+1|\tau}^{\top}] = \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)$ for all $\tau \in \mathbb{N}$, and $\mathbf{x}_{\tau|\tau}$ are the filtered states.

The innovations representation involves the matrices from (2.17),

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}) = (\mathsf{A}(\boldsymbol{\theta};h), \mathsf{K}(\boldsymbol{\theta};h), \mathsf{C}(\boldsymbol{\theta};h), \boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)) \tag{E.11}$$

given by $\mathsf{A}(\boldsymbol{\theta};h)$ and $\mathsf{C}(\boldsymbol{\theta};h)$ from the ABCD representation in the Appendix D, and $\mathsf{K}(\boldsymbol{\theta};h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)$ from system (E.8)-(E.10). Thus, $\mathsf{K}(\boldsymbol{\theta};h)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}(\boldsymbol{\theta};h)$ are functions of the matrices $\mathsf{A}(\boldsymbol{\theta};h)$, $\mathsf{B}(\boldsymbol{\theta};h)$, $\mathsf{C}(\boldsymbol{\theta};h)$, $\mathsf{D}(\boldsymbol{\theta};h)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta};h)$ of the ABCD representation, too.

E.3 Likelihood function

Given the linear structure of the Gaussian state-space model (2.12)-(2.13), it follows that the (joint) probability density function of the discrete measurements can be written as

$$f(\mathbf{y}^T;\boldsymbol{\theta}) = f(\mathbf{y}_1,\ldots,\mathbf{y}_T;\boldsymbol{\theta}) = f(\mathbf{y}_0;\boldsymbol{\theta})\prod_{\tau=1}^{T} f(\mathbf{y}_\tau|\mathbf{y}^{\tau-1};\boldsymbol{\theta})\,,$$

where $f\left(\mathbf{y}_\tau|\mathbf{y}^{\tau-1};\boldsymbol{\theta};h\right) = \mathcal{N}\left(\mathsf{C}(\boldsymbol{\theta},h)\mathbf{x}_{\tau-1|\tau-1},\boldsymbol{\Sigma}_{\boldsymbol{\nu},\tau|\tau-1}\right)$. Then, using the Kalman filter recursion, the conditional log-likelihood function, given $\mathbf{y}_0$, can be constructed recursively via the prediction error decomposition as (see Durbin and Koopman, 2012)

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\theta}|\mathbf{y}^T\right) &= \sum_{\tau=1}^{T}\ln f\left(\mathbf{y}_\tau|\mathbf{y}^{\tau-1};\boldsymbol{\theta}\right)\\ &= -\frac{n_y T}{2}\ln(2\pi) - \frac{1}{2}\sum_{\tau=1}^{T}\ln|\boldsymbol{\Sigma}_{\boldsymbol{\nu},\tau|\tau-1}| - \frac{1}{2}\sum_{\tau=1}^{T}\boldsymbol{\nu}_{\tau|\tau-1}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\nu},\tau|\tau-1}^{-1}\boldsymbol{\nu}_{\tau|\tau-1}\,,\end{aligned}$$

and the maximum-likelihood (ML) estimator of $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\ \mathcal{L}\left(\boldsymbol{\theta}|\mathbf{y}^T\right).$$

E.4 State and disturbance smoothing

For $\tau = T,\ldots,1$, we smooth the state variables $\mathbf{x}_\tau$, given the observations $\{\mathbf{y}_0,\ldots,\mathbf{y}_T\}$, using the two-step fast state-smoothing recursion in Durbin and Koopman (2012, Chapter 4). In the first step, we run a backward smoother algorithm. In particular, let $\mathbf{r}_T = \mathbf{0}$, then compute

$$\begin{aligned}\mathbf{L}_\tau &= \mathsf{A}(\boldsymbol{\theta};h) - \mathbf{K}_{\tau|\tau-1}\mathsf{C}(\boldsymbol{\theta};h)\\ \mathbf{r}_{\tau-1} &= \mathsf{C}(\boldsymbol{\theta};h)^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\nu},\tau|\tau-1}^{-1}\boldsymbol{\nu}_{\tau|\tau-1} + \mathbf{L}_\tau^{\top}\mathbf{r}_\tau\\ \mathbf{N}_{\tau-1} &= \mathsf{C}(\boldsymbol{\theta};h)^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\nu},\tau|\tau-1}^{-1}\mathsf{C}(\boldsymbol{\theta};h) + \mathbf{L}_\tau^{\top}\mathbf{N}_\tau\mathbf{L}_\tau\\ \hat{\mathbf{x}}_\tau &= \mathbf{x}_{\tau|\tau} + \mathbf{P}_{\tau|\tau}\mathbf{r}_{\tau-1}\\ \mathbf{V}_\tau &= \mathbf{P}_{\tau|\tau} - \mathbf{P}_{\tau|\tau}\mathbf{N}_{\tau-1}\mathbf{P}_{\tau|\tau}\,.\end{aligned}$$

In the second step, for $\tau = 1,...,T$, compute the recursion for the smoothed reduced-form residuals

$$\hat{\boldsymbol{\eta}}_\tau = \hat{\mathbf{x}}_\tau - \mathsf{A}(\boldsymbol{\theta};h)\hat{\mathbf{x}}_{\tau-1}\,.$$

F. Oscillations

An alias is a matrix $\mathbf{A}_0 \neq \mathbf{A}(\boldsymbol{\theta}_0)$ solving (3.4). Here, we show that Assumption 5 rules out local aliases, but not aliases at isolated points. Thus, assume that the eigenvalues of $\mathbf{A}(\boldsymbol{\theta}_0)$ are distinct, and do not differ by an integer multiple of $2\pi i/h$. This imposes a simple structure on the data sampling process, whereby the underlying continuous-time model does not generate oscillations with periods shorter than twice the interval between observations, h. Following Phillips (1973), the oscillations are characterized by the complex eigenvalues of $\mathbf{A}(\boldsymbol{\theta}_0)$, that occur in complex conjugate pairs—say, q pairs, $2q \leq n_x$. In particular, the solutions of (3.4) are given by

$$\mathbf{A}_k = \mathbf{A}(\boldsymbol{\theta}_0) + \frac{2\pi i}{h}\mathbf{V}(\boldsymbol{\theta}_0)\mathbf{D}_k\mathbf{V}(\boldsymbol{\theta}_0)^{-1}\,, \tag{F.1}$$

a countably infinite set of solutions $\mathbf{A}_k$, for $q > 0$. Here, $\mathbf{V}(\boldsymbol{\theta}_0)$ is the nonsingular matrix with columns given by the of eigenvectors of $\mathbf{A}(\boldsymbol{\theta}_0)$. Further, $\mathbf{D}_k = \text{diag}(\mathbf{0}_{n_x-2q}, \mathbf{M}_k, -\mathbf{M}_k)$, where $\mathbf{M}_k$ is a diagonal matrix with diagonal elements $m_{k,j} \in \mathbb{Z}^+$, $j = 1, \ldots, q$. If the eigenvalues of $\mathbf{A}(\boldsymbol{\theta}_0)$ are real, $q = 0$, then $\mathbf{A}_k = \mathbf{A}(\boldsymbol{\theta}_0)$ in (F.1), and there are no aliases. With complex eigenvalues, $q > 0$, the aliases $\mathbf{A}_k$ are not local, but arise at isolated points in the space of $n_x \times n_x$ matrices, because $h > 0$ is fixed, as determined by the sampling frequency, $m_{k,j}$ are integers, and $\mathbf{V}(\boldsymbol{\theta}_0)$ is nonsingular. In particular,

$$||\mathbf{A}_k - \mathbf{A}(\boldsymbol{\theta}_0)||_{\mathbf{F}}^2 = \left(\frac{2\pi i}{h}\right)^2 \text{trace } \mathbf{V}(\boldsymbol{\theta}_0)\mathbf{D}_k^2\mathbf{V}(\boldsymbol{\theta}_0)^{-1} = \left(\frac{2\pi i}{h}\right)^2 \text{trace } \mathbf{D}_k^2 \geq \left(\frac{2\pi i}{h}\right)^2 q\,,$$

since $m_{k,j} \geq 1$, with $||\cdot||_{\mathbf{F}}$ the Frobenius norm, so there is a lower bound $2\pi/h > 0$ on the distance between $\mathbf{A}_k$ and $\mathbf{A}(\boldsymbol{\theta}_0)$. Therefore, the solution $\mathbf{A}_0 = \mathbf{A}(\boldsymbol{\theta}_0)$ of (3.4) is unique in a neighborhood around $\mathbf{A}(\boldsymbol{\theta}_0)$, i.e., there are no local aliases.

G. The real business cycle model

G.1 The HJB equation and the first-order conditions

The social planner chooses paths for consumption and the fraction of hours worked that maximize expected discounted lifetime utility, subject to the law of motion for the capital stock (5.1), the production function (5.2), and the evolution of TFP (5.3). There is no population growth, and both population size and endowment of available time are normalized to unity. In addition, the aggregate resource constraint $Y(t) = C(t) + I(t)$ must hold, at all points in time.

Both welfare theorems hold in this single-good economy. Hence, it is possible to solve the social planner's problem directly. It is given by the dynamic optimization problem

$$J(K_0, Z_0) = \max_{\{C(t),N(t)\}_{t=0}^{\infty}} \mathbb{E}_0 \left[\int_0^{\infty} e^{-\rho t} \left(\ln C(t) + \psi \left(1 - N(t) \right) \right) \mathrm{d}t \right] ,$$

subject to

$$\begin{aligned} \mathrm{d}K(t) &= \left(\exp\left(Z(t) \right) K(t)^{\alpha} \left(\exp\left(\eta t \right) N(t) \right)^{1-\alpha} - C(t) - \delta K(t) \right) \mathrm{d}t + \sigma_k K(t) \mathrm{d}w_k(t) , \\ \mathrm{d}Z(t) &= -\rho_z Z(t) \mathrm{d}t + \sigma_z \mathrm{d}w_z(t) , \end{aligned}$$

in which $C(t) \in \mathbb{R}^+$ and $N(t) \in [0, 1]$ are the control variables at instant $t > 0$, $K(t) \in \mathbb{R}^+$ and $Z(t) \in \mathbb{R}$ are the state variables at instant t, and $J\left(K, Z\right)$ is the value of the optimal program (value function) given the initial conditions $K\left(0\right) = K_0$ and $Z\left(0\right) = Z_0$.

The economy exhibits a balanced growth path, i.e., over the long run, the variables in the economy, with the exception of hours worked and TFP, will grow at the gross rate $\eta > 1$. A stationary version of the model can be obtained by defining $y(t) := Y(t)/\exp\left(\eta t\right)$, $c(t) := C(t)/\exp\left(\eta t\right)$, $k(t) := K(t)/\exp\left(\eta t\right)$ to be the de-trended values of the macroeconomic variables. For notational consistency, we also define $n(t) := N(t)$ and $z(t) := Z(t)$. Using these definitions, the planner's optimal control problem can be rewritten as[3]

$$J(k_0, z_0) = \max_{\{c(t),n(t)\}_{t=0}^{\infty}} \mathbb{E}_0 \left[\int_0^{\infty} e^{-\rho t} \left(\ln c(t) + \psi \left(1 - n(t) \right) \right) \mathrm{d}t \right] ,$$

subject to

$$\begin{aligned} \mathrm{d}k(t) &= \left(\exp\left(z(t) \right) k(t)^{\alpha} n(t)^{1-\alpha} - c(t) - \left(\delta + \eta \right) k(t) \right) \mathrm{d}t + \sigma_k k(t) \mathrm{d}w_k(t), \quad k\left(0\right) = k_0 , \\ \mathrm{d}z(t) &= -\rho_z z(t) \mathrm{d}t + \sigma_z \mathrm{d}w_z(t) , \quad z\left(0\right) = z_0 . \end{aligned}$$

[3] We use the fact that $\int_{t=0}^{\infty} e^{-\rho t} \eta t \mathrm{d}t = \frac{\eta}{\rho^2}$ for $\rho > 0$, and hence it is just a constant that we omit without affecting the optimization problem. In discrete time, this is equivalent to omitting $\sum_{t=0}^{\infty} \beta^t \eta t = \frac{\eta\beta}{(1-\beta)^2}$, as long as $\beta \in (0, 1)$.

A recursive representation of the planner's problem is given by the Hamilton-Jacobi-Bellman (HJB) equation,[4]

$$\rho J(k,z) = \max_{c,n} \Bigg\{ (\ln c + \psi(1-n)) + \left(\exp(z)k^{\alpha}n^{1-\alpha} - c - (\delta+\eta)k\right) J_k(k,z) \\ - \rho_z z J_z(k,z) + \frac{1}{2}\sigma_k^2 k^2 J_{kk}(k,z) + \frac{1}{2}\sigma_z^2 J_{zz}(k,z) \Bigg\}, \quad \text{(G.1)}$$

where subscripts denote partial derivatives. The first order conditions for an interior solution are given by

$$c = \left(J_k(k,z)\right)^{-1},$$
$$\psi = (1-\alpha)\exp(z)k^{\alpha}n^{-\alpha}J_k(k,z),$$

which implicitly define optimal consumption and hours worked as functions of the state variables of the economy, $c = \mathbf{c}(k,z)$ and $n = \mathbf{n}(k,z)$.

The maximized HJB equation reads

$$\rho J(k,z) = \ln \mathbf{c}(k,z) + \psi(1-\mathbf{n}(k,z)) \\ + \left(\exp(z)k^{\alpha}\mathbf{n}(k,z)^{1-\alpha} - \mathbf{c}(k,z) - (\delta+\eta)k\right) J_k(k,z) \\ - \rho_z z J_z(k,z) + \frac{1}{2}\sigma_k^2 k^2 J_{kk}(k,z) + \frac{1}{2}\sigma_z^2 J_{zz}(k,z), \quad \text{(G.2)}$$

from which it is follows that the co-state variable associated with the capital stock must satisfy (using the envelope condition)

$$\rho J_k(k,z) = \left(\alpha\exp(z)k^{\alpha-1}\mathbf{n}(k,z)^{1-\alpha} - (\delta+\eta)\right) J_k(k,z) \\ + \left(\exp(z)k^{\alpha}\mathbf{n}(k,z)^{1-\alpha} - \mathbf{c}(k,z) - (\delta+\eta)k\right) J_{kk}(k,z) \\ - \rho_z z J_{kz}(k,z) + \sigma_k^2 k J_{kk}(k,z) + \frac{1}{2}\sigma_k^2 k^2 J_{kkk}(k,z) + \frac{1}{2}\sigma_z^2 J_{kkz}(k,z).$$

Collecting terms yields

$$\left(\rho - \alpha\exp(z)k^{\alpha-1}\mathbf{n}(k,z)^{1-\alpha} + \delta + \eta\right) J_k(k,z) = \Big(\exp(z)k^{\alpha}\mathbf{n}(k,z)^{1-\alpha} \\ - \mathbf{c}(k,z) - (\delta+\eta)k\Big) J_{kk}(k,z) - \rho_z z J_{kz}(k,z) \\ + \sigma_k^2 k J_{kk}(k,z) + \frac{1}{2}\sigma_k^2 k^2 J_{kkk}(k,z) + \frac{1}{2}\sigma_z^2 J_{kkz}(k,z). \quad \text{(G.3)}$$

[4] See Chang (2009) for a formal derivation of the HJB equation.

Using Ito's formula, the co-state variable evolves according to

$$\begin{aligned}\mathrm{d}J_k\left(k,z\right) = \Bigg[& \left(\exp(z)k^{\alpha}n^{1-\alpha} - c - \left(\delta+\eta\right)k\right) J_{kk}\left(k,z\right) \\ & - \rho_z z J_{kz}\left(k,z\right) + \frac{1}{2}\sigma_k^2 k^2 J_{kkk}\left(k,z\right) + \frac{1}{2}\sigma_z^2 J_{kkz}\left(k,z\right)\Bigg]\mathrm{d}t \\ & + \sigma_k k J_{kk}\left(k,z\right)\mathrm{d}w_k + \sigma_z J_{kz}\left(k,z\right)\mathrm{d}w_z\,,\end{aligned}$$

where, substituting for the optimal costate in (G.3), we obtain the equilibrium dynamics of marginal utility of consumption,

$$\begin{aligned}\mathrm{d}J_k\left(k,z\right) = & \left[\left(\rho - \alpha\exp\left(z\right)k^{\alpha-1}n^{1-\alpha} + \delta + \eta\right) J_k\left(k,z\right) - \sigma_k^2 k J_{kk}\left(k,z\right)\right]\mathrm{d}t \\ & + \sigma_k k J_{kk}\left(k,z\right)\mathrm{d}w_k + \sigma_z J_{kz}\left(k,z\right)\mathrm{d}w_z\,. \quad \text{(G.4)}\end{aligned}$$

After some algebra, this leads to the Euler equation for consumption,

$$\begin{aligned}\frac{\mathrm{d}c}{c} = & \Bigg[\left(\alpha\exp\left(z\right)k^{\alpha-1}n^{1-\alpha} - \rho - \delta - \eta\right) - \sigma_k^2\frac{k\mathbf{c}_k\left(k,z\right)}{c} \\ & + \frac{1}{2}\left(\sigma_k^2\left(\frac{k\mathbf{c}_k\left(k,z\right)}{c}\right)^2 + \sigma_z^2\left(\frac{\mathbf{c}_z\left(k,z\right)}{c}\right)^2\right)\Bigg]\mathrm{d}t \\ & + \sigma_k\frac{k\mathbf{c}_k\left(k,z\right)}{c}\mathrm{d}w_k + \sigma_z\frac{\mathbf{c}_z\left(k,z\right)}{c}\mathrm{d}w_z\,, \quad \text{(G.5)}\end{aligned}$$

where $\mathbf{c}_k\left(k(t),z(t)\right)$ and $\mathbf{c}_z\left(k(t),z(t)\right)$ denote the marginal responses of optimal consumption to changes in the capital stock and TFP.

Given the properties of stochastic integrals with respect to Brownian motion, the Euler equation for consumption can be alternatively written in expected terms as

$$\begin{aligned}\frac{1}{\mathrm{d}t}\mathbb{E}_t\left[\frac{\mathrm{d}c}{c}\right] = & \left(\alpha\exp\left(z\right)k^{\alpha-1}\mathbf{n}\left(k,z\right)^{1-\alpha} - \rho - \delta - \eta\right) - \sigma_k^2\frac{k\mathbf{c}_k\left(k,z\right)}{c} \\ & + \frac{1}{2}\sigma_k^2\left(\frac{k\mathbf{c}_k\left(k,z\right)}{c}\right)^2 + \frac{1}{2}\sigma_z^2\left(\frac{\mathbf{c}_z\left(k,z\right)}{c}\right)^2\,. \quad \text{(G.6)}\end{aligned}$$

G.2 Equilibrium

The general equilibrium in this economy can be characterized in the time domain by the system of nonlinear stochastic differential equations

$$\mathbb{E}_t\left[\frac{\mathrm{d}c}{c}\right] = \left[\left(\alpha \exp\left(z\right) k^{\alpha-1} n^{1-\alpha} - \rho - \delta - \eta\right) - \sigma_k^2 \frac{k\mathbf{c}_k\left(k,z\right)}{c}\right. \tag{G.7}$$
$$\left. + \frac{1}{2}\left(\sigma_k^2\left(\frac{k\mathbf{c}_k\left(k,z\right)}{c}\right)^2 + \sigma_z^2\left(\frac{\mathbf{c}_z\left(k,z\right)}{c}\right)^2\right)\right]\mathrm{d}t\,,$$
$$\mathrm{d}k = \left(\exp\left(z\right) k^{\alpha} n^{1-\alpha} - c - \left(\delta+\eta\right) k\right)\mathrm{d}t + \sigma_k k \mathrm{d}w_k\,, \quad k\left(0\right) = k_0\,, \tag{G.8}$$
$$\mathrm{d}z = -\rho_z z \mathrm{d}t + \sigma_z \mathrm{d}w_z\,, \quad z\left(0\right) = z_0\,, \tag{G.9}$$

together with the algebraic (static) equation for the optimal fraction of hours worked,

$$\psi c n = \left(1-\alpha\right)\exp(z) k^{\alpha} n^{1-\alpha}\,. \tag{G.10}$$

Collecting the model variables in the vector $\check{\mathbf{x}} = \left[c, k, z, n\right]^{\top}$, and using the properties of stochastic integrals with respect to Brownian motion, we compactly write the nonlinear equilibrium as

$$\mathrm{d}\check{\mathbf{x}}(t) = \mathbf{G}_0\left(\check{\mathbf{x}}(t)\right)\mathrm{d}t + \mathbf{G}_1\left(\check{\mathbf{x}}(t)\right)\mathrm{d}\mathbf{w}(t) + \widetilde{\mathbf{\Pi}}\mathrm{d}\varepsilon(t)\,, \tag{G.11}$$

where $\mathbf{w}(t) = \left[w_k(t), w_z(t)\right]^{\top}$ is the vector of structural shocks, $\varepsilon(t)$ is an expectation error defined as the difference between the actual and unexpected change in consumption, i.e., $\mathrm{d}\varepsilon(t) = \mathbb{E}_t\left[\mathrm{d}c(t)\right] - \mathrm{d}c(t)$, satisfying $\mathbb{E}_t\left[\mathrm{d}\varepsilon(t)\right] = 0$, and $\widetilde{\mathbf{\Pi}}$ is a selection matrix.

G.3 Deterministic steady state

In the absence of uncertainty, the (de-trended) economy converges over time to a fixed point, or steady-state equilibrium, in which all variables are idle. We denote such point by $\check{\mathbf{x}}^{\star} = \left(c^{\star}, n^{\star}, k^{\star}, z^{\star}\right)^{\top}$. Therefore, imposing $\sigma_k = \sigma_z = 0$, together with the no-growth condition $\mathrm{d}\check{\mathbf{x}}(t)/\mathrm{d}t = \mathbf{0}$, on the system (G.11), we find the deterministic steady state

$$z^{\star} = 0\,, \qquad n^{\star} = \left(1-\alpha\right)\left(\psi\left(1 - \tfrac{\alpha(\delta+\eta)}{\rho+\delta+\eta}\right)\right)^{-1}\,,$$

$$k^{\star} = \left(\frac{\alpha}{\rho+\delta+\eta}\right)^{\frac{1}{1-\alpha}} n^{\star}\,, \text{ and } c^{\star} = \left(k^{\star}\right)^{\alpha}\left(n^{\star}\right)^{1-\alpha} - \left(\delta+\eta\right)k^{\star}\,.$$

G.4 Log-linearized equilibrium

The nonlinear system formed by (G.7)-(G.10) can be linearized in order to study the dynamic behavior of the stationary variables as they fluctuate in close proximity to their deterministic steady-state values. Let $\hat{c} = \ln c - \ln c^{\star}$, $\hat{n} = \ln n - \ln n^{\star}$, $\hat{k} = \ln k - \ln k^{\star}$,

and $\hat{z} = z - z^\star$ denote log-deviations of the variables with respect to their steady-state values. Then, a first-order Taylor expansion of (G.11) yields

$$\begin{bmatrix} \mathrm{d}\hat{c} \\ \mathrm{d}\hat{k} \\ \mathrm{d}\hat{z} \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \xi_{ck} & \xi_{cz} & \xi_{cn} \\ \xi_{kc} & \xi_{kk} & \xi_{kz} & \xi_{kn} \\ 0 & 0 & -\rho_z & 0 \\ \xi_{nc} & \xi_{nk} & \xi_{nz} & -1 \end{bmatrix}}_{\equiv \widetilde{\mathbf{\Gamma}}} \begin{bmatrix} \hat{c} \\ \hat{k} \\ \hat{z} \\ \hat{n} \end{bmatrix} \mathrm{d}t + \underbrace{\begin{bmatrix} 0 & 0 \\ \sigma_k & 0 \\ 0 & \sigma_z \\ 0 & 0 \end{bmatrix}}_{\equiv \widetilde{\mathbf{\Psi}}} \begin{bmatrix} \mathrm{d}w_k \\ \mathrm{d}w_z \end{bmatrix} + \underbrace{\begin{bmatrix} -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\equiv \widetilde{\mathbf{\Pi}}} \mathrm{d}\varepsilon \,,$$

where $\widetilde{\mathbf{\Gamma}}$ is the Jacobian matrix of the log-transformed equilibrium evaluated at the deterministic steady state, and $\widetilde{\mathbf{\Psi}}$ is the corresponding diffusion matrix. The log-transformation is obtained via an application of Itô's formula to (G.11). The coefficients in $\widetilde{\mathbf{\Gamma}}$ are given by $\xi_{ck} = (\alpha{-}1)(\rho{+}\delta{+}\eta)$, $\xi_{cz} = (\rho{+}\delta{+}\eta)$, $\xi_{cn} = (1{-}\alpha)(\rho{+}\delta{+}\eta)$, $\xi_{kc} = -(\rho{+}(1{-}\alpha)(\delta{+}\eta))/\alpha$, $\xi_{kk} = \rho$, $\xi_{kz} = (\rho{+}\delta{+}\eta)/\alpha$, $\xi_{kn} = (1{-}\alpha)(\rho{+}\delta{+}\eta)/\alpha$, $\xi_{nc} = -1/\alpha$, $\xi_{nk} = 1$ and $\xi_{nz} = 1/\alpha$.

Next, we substitute out the intratemporal labor supply condition $\hat{n} = \xi_{nc}\hat{c}{+}\xi_{nk}\hat{k}{+}\xi_{nz}\hat{z}$, to obtained a linearized equilibrium consisting of the 3×3 system of linear stochastic differential equations

$$\begin{bmatrix} \mathrm{d}\hat{c} \\ \mathrm{d}\hat{k} \\ \mathrm{d}\hat{z} \end{bmatrix} = \underbrace{\begin{bmatrix} \xi_{cn}\xi_{nc} & 0 & \xi_{cz} + \xi_{cn}\xi_{nz} \\ \xi_{kc} + \xi_{kn}\xi_{nc} & \xi_{kk} + \xi_{kn}\xi_{nk} & \xi_{kz} + \xi_{kn}\xi_{nz} \\ 0 & 0 & -\rho_z \end{bmatrix}}_{\equiv \mathbf{\Gamma}} \begin{bmatrix} \hat{c} \\ \hat{k} \\ \hat{z} \end{bmatrix} \mathrm{d}t$$
$$+ \underbrace{\begin{bmatrix} 0 & 0 \\ \sigma_k & 0 \\ 0 & \sigma_z \end{bmatrix}}_{\equiv \mathbf{\Psi}} \begin{bmatrix} \mathrm{d}w_k \\ \mathrm{d}w_z \end{bmatrix} + \underbrace{\begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}}_{\equiv \mathbf{\Pi}} \mathrm{d}\varepsilon \,,$$

which can be compactly written as

$$\mathrm{d}\tilde{\hat{\mathbf{x}}}(t) = \mathbf{\Gamma}\tilde{\hat{\mathbf{x}}}(t)\mathrm{d}t + \mathbf{\Psi}\mathrm{d}\mathbf{w}(t) + \mathbf{\Pi}\mathrm{d}\varepsilon(t) \,, \tag{G.12}$$

where $\tilde{\hat{\mathbf{x}}} = [\hat{c}, \hat{k}, \hat{z}]^\top$ denotes the vector of variables in deviations from their deterministic steady state, and where $\mathbf{\Gamma}$, $\mathbf{\Psi}$, and $\mathbf{\Pi}$ are the adjusted versions of $\widetilde{\mathbf{\Gamma}}$, $\widetilde{\mathbf{\Psi}}$, and $\widetilde{\mathbf{\Pi}}$. Note that the volatility parameters σ_k and σ_z do not affect the matrix $\mathbf{\Gamma}$ characterizing the endogenous persistence in the linearized equilibrium system. Therefore, they will not have any effects on the implied optimal decision rules, and hence our approximated solution exhibits certainty equivalent in the sense of Simon (1956) and Theil (1957) (see Ahn et al., 2018; Parra-Alvarez et al., 2021).

G.5 Rational expectations solution

Following Sims (2002), let us assume that the matrix $\mathbf{\Gamma}$ can be diagonalized according to

$$\mathbf{\Gamma} = \mathbf{T}\mathbf{\Upsilon}\mathbf{T}^{-1}\,, \tag{G.13}$$

where $\mathbf{T}$ is a 3×3 matrix of right-eigenvectors of $\mathbf{\Gamma}$, and $\mathbf{\Upsilon}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{\Gamma}$. Premultiplying (G.12) by $\mathbf{T}^{-1}$ and defining $\mathbf{z}(t) = \mathbf{T}^{-1}\tilde{\hat{\mathbf{x}}}(t)$ yields

$$\mathrm{d}\mathbf{z}(t) = \mathbf{\Upsilon}\mathbf{z}(t)\mathrm{d}t + \mathbf{T}^{-1}\mathbf{\Psi}\mathrm{d}\mathbf{w}(t) + \mathbf{T}^{-1}\mathbf{\Pi}\mathrm{d}\varepsilon(t)\,. \tag{G.14}$$

The eigenvalues of the matrix $\mathbf{\Gamma}$ solve the characteristic equation $|\mathbf{\Gamma} - \upsilon\mathbf{I}_3| = 0$. Thus, it follows that the eigenvalues of $\mathbf{\Gamma}$ are given by $\upsilon_1 = -\rho_z$ and the roots of the quadratic equation

$$a_0\upsilon^2 + a_1\upsilon + a_2 = 0\,,$$

with $a_0 = 1$, $a_1 = -\left(\xi_{cn}\xi_{nc} + \xi_{kk} + \xi_{kn}\xi_{nk}\right)$, and

$$a_2 = \left(\xi_{cn}\xi_{nc}\xi_{kk} + \xi_{cn}\xi_{nc}\xi_{kn}\xi_{nk}\right)\,.$$

After some algebra, it is possible to show that

$$a_1 = -\rho < 0\,, \quad \text{and} \quad a_2 = -\frac{(1-\alpha)\,(\rho+\delta+\eta)}{\alpha}\left(\frac{\rho + (1-\alpha)\,(\delta+\eta)}{\alpha}\right) < 0\,.$$

Since $a_2^2 - 4a_0a_1 > 0$ (the discriminant of the quadratic equation) and $a_2 < 0$, the quadratic equation has two distinct real roots of opposite sign, given by

$$\upsilon_2 = -\frac{(1-\alpha)(\delta+\eta+\rho)}{\alpha} < 0\,, \text{ and } \upsilon_3 = \frac{(1-\alpha)(\delta+\eta)+\rho}{\alpha} > 0\,.$$

Hence, the linearized system has two stable roots (non-positive eigenvalues, υ_1 and υ_2) and one unstable root (positive eigenvalue, υ_3). Since the reduced model in (G.12) has two state variables and one control/jump variable, the Blanchard and Kahn conditions are satisfied, and the model has a unique rational expectation solution (see Buiter, 1984). The eigenvectors of $\mathbf{\Gamma}$ associated with each of its eigenvalues are given by multiples of the vectors

$$T_1 = \frac{1}{\iota}\begin{bmatrix} -(\delta+\eta+\rho) \\ -\frac{(\delta+\eta+\rho)((1-\alpha)(\delta+\eta)+\rho+\rho_z)}{(1-\alpha)(\delta+\eta)+\rho+\alpha\rho_z} \\ \iota \end{bmatrix}, T_2 = \begin{bmatrix} \frac{\alpha(2(1-\alpha)(\delta+\eta)+(2-\alpha)\rho)}{(1-\alpha^2)(\delta+\eta)+\rho} \\ 1 \\ 0 \end{bmatrix}, \text{ and } T_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

where $\iota = (\alpha-1)(\delta+\eta+\rho) + \alpha\rho_z$.

Let $\mathbf{M}_+$ be a 1×3 vector that selects the rows of $\mathbf{T}^{-1}$ corresponding to eigenvalues with positive real parts, and $\mathbf{M}_-$ a 2×3 matrix that selects the rows of $\mathbf{T}^{-1}$ corresponding to eigenvalues with non-positive real parts. It follows that $\left(\mathbf{M}_+^\top\mathbf{M}_+ + \mathbf{M}_-^\top\mathbf{M}_-\right) = \mathbf{I}_3$. Then

$$\mathbf{M}_+\mathrm{d}\mathbf{z}(t) = \mathbf{M}_+\boldsymbol{\Upsilon}\mathbf{z}(t)\mathrm{d}t + \mathbf{M}_+\mathbf{T}^{-1}\boldsymbol{\Psi}\mathrm{d}\mathbf{w}(t) + \mathbf{M}_+\mathbf{T}^{-1}\boldsymbol{\Pi}\mathrm{d}\varepsilon(t) \tag{G.15}$$

defines the equation associated with the unstable eigenvalue. To rule out explosive paths, i.e., to ensure that $\lim_{s\to\infty}\mathbb{E}_t\left[\mathbf{z}(s)\right] < \infty$ for $s > t$, and thus to satisfy the model's transversality condition, we impose

$$\mathbf{M}_+\mathbf{z}(t) = 0\,,\ \forall t\,, \tag{G.16}$$

implying that

$$\mathrm{d}\varepsilon(t) = -\left[\mathbf{M}_+\mathbf{T}^{-1}\boldsymbol{\Pi}\right]^{-1}\mathbf{M}_+\mathbf{T}^{-1}\boldsymbol{\Psi}\mathrm{d}\mathbf{w}(t)\,. \tag{G.17}$$

In other words, the stability condition imposes an exact relationship between the vector of structural shocks and the expectation error, such that the system does not exhibit explosive paths.

Once we impose the stability conditions (G.16) and (G.17), it is possible to compute the solution associated with the stable eigenvalues by computing

$$\mathbf{M}_-\mathrm{d}\mathbf{z}(t) = \mathbf{M}_-\boldsymbol{\Upsilon}\mathbf{z}(t)\mathrm{d}t + \mathbf{M}_-\mathbf{T}^{-1}\boldsymbol{\Psi}\mathrm{d}\mathbf{w}(t) + \mathbf{M}_-\mathbf{T}^{-1}\boldsymbol{\Pi}\mathrm{d}\varepsilon(t)\,,$$

which in turn implies that

$$\mathrm{d}\mathbf{z}(t) = \boldsymbol{\Upsilon}^\star\mathbf{z}(t)\mathrm{d}t + \boldsymbol{\Psi}^\star\mathrm{d}\mathbf{w}(t)\,, \tag{G.18}$$

where $\boldsymbol{\Upsilon}^\star = \mathbf{M}_-^\top\mathbf{M}_-\boldsymbol{\Upsilon}\mathbf{M}_-^\top\mathbf{M}_-$ is the 3×3 matrix of eigenvalues with zeros in the position of the explosive paths, and

$$\boldsymbol{\Psi}^\star = \mathbf{M}_-^\top\mathbf{M}_-\mathbf{T}^{-1}\left[\mathbf{I}_3 - \boldsymbol{\Pi}\left[\mathbf{M}_+\mathbf{T}^{-1}\boldsymbol{\Pi}\right]^{-1}\mathbf{M}_+\mathbf{T}^{-1}\right]\boldsymbol{\Psi}$$

is a 3×2 matrix. Finally, we use the definition $\mathbf{z}(t) = \mathbf{T}^{-1}\tilde{\mathbf{x}}(t)$ to obtain the autoregressive representation of the rational expectation solution in the original variables,

$$\mathrm{d}\tilde{\mathbf{x}}(t) = \widetilde{\mathbf{A}}\tilde{\mathbf{x}}(t)\mathrm{d}t + \widetilde{\mathbf{B}}\mathrm{d}\mathbf{w}(t)\,, \tag{G.19}$$

where $\widetilde{\mathbf{A}} = \mathbf{T}\boldsymbol{\Upsilon}^\star\mathbf{T}^{-1}$ and $\widetilde{\mathbf{B}} = \mathbf{T}\boldsymbol{\Psi}^\star$.

From the stability condition (G.16) and the definition of the transformed variable

$\mathbf{z}(t)$, we find the optimal policy for consumption as

$$\hat{c}(t) = \phi_{ck}\hat{k}(t) + \phi_{cz}\hat{z}(t)\,, \tag{G.20}$$

where $\phi_{ck} = -(T_{32}/T_{31})$, and $\phi_{cz} = -(T_{33}/T_{31})$, with T_{ij} the (i,j)-th element of the matrix of $\mathbf{T}^{-1}$. Using the linearized condition for hours worked, the optimal policy for labor is given by

$$\hat{n}(t) = \phi_{nk}\hat{k}(t) + \phi_{nz}\hat{z}(t), \tag{G.21}$$

where $\phi_{nk} = \left(\xi_{nk} - \xi_{nc}\frac{T_{32}}{T_{33}}\right)$, and $\phi_{nz} = \left(\xi_{nz} - \xi_{nc}\frac{T_{33}}{T_{31}}\right)$.

As a final step, we eliminate the dependence of the system in (G.19) on the control variables to obtain a system of SDEs that only describes the optimal dynamics of the state variables. After some algebra, we obtain

$$\begin{bmatrix} \mathrm{d}\hat{k}(t) \\ \mathrm{d}\hat{z}(t) \end{bmatrix} = \begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} \begin{bmatrix} \hat{k}(t) \\ \hat{z}(t) \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \sigma_k & 0 \\ 0 & \sigma_z \end{bmatrix} \begin{bmatrix} \mathrm{d}w_k(t) \\ \mathrm{d}w_z(t) \end{bmatrix}, \tag{G.22}$$

with

$$\begin{aligned} \phi_{kk} &= -(\tilde{a}_{21}T_{32})/T_{31} + \tilde{a}_{22} = -\frac{(1-\alpha)(\delta+\eta+\rho)}{\alpha} < 0\,, \\ \phi_{kz} &= -(\tilde{a}_{21}T_{33})/T_{31} + \tilde{a}_{23} = \frac{(\delta+\eta+\rho)((1-\alpha)(\delta+\eta)+\rho+\rho_z)}{\alpha((1-\alpha)(\delta+\eta)+\rho+\alpha\rho_z)} > 0\,, \end{aligned}$$

where $\tilde{a}_{ij}$ is the (i,j)-th element of the matrix of $\tilde{\mathbf{A}}$.[5]

Let $\hat{\mathbf{y}}_t = [\hat{c}_t, \hat{n}_t]^\top$ and $\hat{\mathbf{x}}_t = [\hat{k}_t, \hat{z}_t]^\top$ denote, respectively, the vector of control and state variables in log-deviations from their steady state values. Then (G.22), together with (G.20) and (G.21), have the continuous-time state space representation in (2.1) and (2.2), i.e.,

$$\begin{aligned} \mathrm{d}\hat{\mathbf{x}}(t) &= \mathbf{A}\hat{\mathbf{x}}(t)\mathrm{d}t + \mathbf{B}\mathrm{d}\mathbf{w}(t)\,, \\ \hat{\mathbf{y}}(t) &= \mathbf{C}\hat{\mathbf{x}}(t)\,. \end{aligned}$$

Real interest rate. Some of our simulation and empirical estimation exercises include the real interest rate as an observable. In equilibrium, its value is determined by the marginal productivity of capital, i.e.,

$$r(t) = \alpha \exp(z(t))k(t)^{\alpha-1}n(t)^{1-\alpha} - \delta\,,$$

[5] We compute the solution to the loglinearized equilibrium conditions using the PHACT toolbox developed by Ahn et al. (2018), available at https://github.com/gregkaplan/phact/tree/master. See the `Readme.txt` file in the accompanying replication codes for further details.

with steady state value of $r^\star = \alpha\left(k^\star\right)^{\alpha-1}\left(n^\star\right)^{1-\alpha} - \delta$. Log-linearization leads to

$$\hat{r}(t) = \phi_{rk}\hat{k}(t) + \phi_{rz}\hat{z}(t)\,,$$

where $\phi_{rk} = ((\alpha-1) + (1-\alpha)\phi_{nk})\left(\frac{r^\star+\delta}{r^\star}\right)$ and $\phi_{rz} = (1-(1-\alpha)\phi_{nz})\left(\frac{r^\star+\delta}{r^\star}\right)$.

G.6 Data generating process: Solution and simulation

Solution. The data generating process (DGP) used in the Monte Carlo section corresponds to the nonlinear solution to the RBC model described above. We approximate the solution using collocation methods. In particular, we approximate the unknown policy functions for consumption, $c = \mathbf{c}(k,z)$, and hours worked, $n = \mathbf{n}(k,z)$, with Chebyshev polynominals of degree 7 over a grid of 7 Chebyshev nodes in each dimension of the state space. The grid points lie between 50% and 150% of the deterministic steady-state values of the state variables. We use a root-finding algorithm to determine the coefficients of the polynomials that make the approximated policy functions satisfy the maximized Hamilton-Jacobi-Bellman equation in (G.2). We initialize the algorithm with the polynomial coefficients implied by the linearized policy functions (G.20)-(G.21). Let $\tilde{\mathbf{c}}\left(k,z\right)$ and $\tilde{\mathbf{n}}\left(k,z\right))$ denote the approximated policy functions.

Simulation. We simulate data from the model using the approximated policy functions. We start the simulation at the model's deterministic steady state, $(k(0),z(0)) = (k_0,z_0) = (k^\star,z^\star)$. We use an Euler discretization of the (controlled) stochastic differential equations (5.3) and (5.6) with the control variables replaced by the approximations to simulate $T = 60$ years of quarterly $h = 1/4$ data using a within-quarter time step equal to $\Delta = 1/120$. In particular, we compute

$$\begin{aligned}
k(t_{i+1}) &= k(t_i) + \Big(\exp\left(z(t_i)\right)k(t_i)^\alpha\tilde{\mathbf{n}}\left(k(t_i),z(t_i)\right)^{1-\alpha} \\
&\quad -\tilde{\mathbf{c}}(k(t_i),z(t_i)) - (\delta+\eta)\,k(t_i)\Big)\left(t_{i+1}-t_i\right) \\
&\quad +\sigma_k k_i\sqrt{t_{i+1}-t_i}\,\epsilon_k(t_{i+1})\,, \\
z(t_{i+1}) &= z(t_i) - \rho_z z(t_i)\left(t_{i+1}-t_i\right) + \sigma_z\sqrt{t_{i+1}-t_i}\,\epsilon_z(t_{i+1})\,,
\end{aligned}$$

for $i = 1,\ldots,T/(\Delta\times h)$, where ϵ_k and ϵ_z are independent draws from a $\mathcal{N}(0,1)$ random variable, and $t_{i+1} - t_i = \Delta\times h$.

Hence, the total number of simulated values is $T/(\Delta\times h) = 28,800$ (note that $\Delta\times h = h_n$ in Section 4). The sample of stock measurements is then constructed by selecting all the end-of-quarter values of the variables. The sample of flows measurements is constructed using the average of simulated values within a quarter (average over the 120 simulated data points in each quarter).

H. State space representation

H.1 Common frequency data

S-SSR. In the S-SSR, the state and measurement vectors are given by $\mathbf{x}_\tau = \mathbf{x}_\tau^s = [k_\tau^s, z_\tau^s]^\top$ and $\mathbf{y}_\tau = \mathbf{y}_\tau^s = [c_\tau^s, n_\tau^s]^\top$, so $n_x^s = 2$, $n_x^f = 0$, $n_y^s = 2$, and $n_y^f = 0$. We next compute the selection matrices $\mathbf{S}_x^f$, $\mathbf{S}_y^f$, and $\mathbf{S}_y^s$. The state-space representation (2.4)-(2.5) in Proposition 1 becomes

$$\begin{bmatrix} k_\tau^s \\ z_\tau^s \end{bmatrix} = \exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} h\right) \begin{bmatrix} k_{\tau-1}^s \\ z_{\tau-1}^s \end{bmatrix} + \begin{bmatrix} \eta_{k,\tau}^s \\ \eta_{z,\tau}^s \end{bmatrix},$$
$$\begin{bmatrix} c_\tau^s \\ n_\tau^s \end{bmatrix} = \begin{bmatrix} \phi_{ck} & \phi_{cz} \\ \phi_{nk} & \phi_{nz} \end{bmatrix} \begin{bmatrix} k_\tau^s \\ z_\tau^s \end{bmatrix},$$

with

$$\begin{bmatrix} \eta_{k,\tau}^s \\ \eta_{z,\tau}^s \end{bmatrix} = \int_{t_{\tau-1}}^{t_\tau} \exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} (t_\tau - u)\right) \begin{bmatrix} \sigma_k & 0 \\ 0 & \sigma_z \end{bmatrix} \begin{bmatrix} \mathrm{d}w_k(u) \\ \mathrm{d}w_z(u) \end{bmatrix}.$$

F-SSR. In the F-SSR, the state and measurement vectors are given by $\mathbf{x}_\tau = [\mathbf{x}_\tau^{s\top}, \mathbf{x}_\tau^{f\top}]^\top = [k_\tau^s, z_\tau^s, k_\tau^f, z_\tau^f]^\top$ and $\mathbf{y}_\tau = \mathbf{y}_\tau^f = [c_\tau^f, n_\tau^f]^\top$, so $n_x^s = 2$, $n_x^f = 2$, $n_y^s = 0$, and $n_y^f = 2$. We next compute the selection matrices $\mathbf{S}_x^f$, $\mathbf{S}_y^f$, and $\mathbf{S}_y^s$. The state-space representation (2.4)-(2.5) in Proposition 1 becomes

$$\begin{bmatrix} k_\tau^s \\ z_\tau^s \\ k_\tau^f \\ z_\tau^f \end{bmatrix} = \left[\begin{array}{cc:cc} & \exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} h\right) & 0 & 0 \\ & & 0 & 0 \\ \hdashline \multicolumn{2}{c:}{\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix}^{-1} \left(\exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} h\right) - \mathbf{I}_2\right)} & 0 & 0 \\ & & 0 & 0 \end{array}\right] \begin{bmatrix} k_{\tau-1}^s \\ z_{\tau-1}^s \\ k_{\tau-1}^f \\ z_{\tau-1}^f \end{bmatrix} + \begin{bmatrix} \eta_{k,\tau}^s \\ \eta_{z,\tau}^s \\ \eta_{k,\tau}^f \\ \eta_{z,\tau}^f \end{bmatrix},$$
$$\begin{bmatrix} c_\tau^f \\ n_\tau^f \end{bmatrix} = \left[\begin{array}{cc:cc} 0 & 0 & \phi_{ck} & \phi_{cz} \\ 0 & 0 & \phi_{nk} & \phi_{nz} \end{array}\right] \begin{bmatrix} k_\tau^s \\ z_\tau^s \\ k_\tau^f \\ z_\tau^f \end{bmatrix},$$

with

$$\begin{bmatrix} \eta_{k,\tau}^s \\ \eta_{z,\tau}^s \\ \eta_{k,\tau}^f \\ \eta_{z,\tau}^f \end{bmatrix} = \int_{t_{\tau-1}}^{t_\tau} \begin{bmatrix} & \exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} (t_\tau - u)\right) \begin{bmatrix} \sigma_k & 0 \\ 0 & \sigma_z \end{bmatrix} \\ \begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix}^{-1} \left(\exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} (t_\tau - u)\right) - \mathbf{I}_2\right) & \begin{bmatrix} \sigma_k & 0 \\ 0 & \sigma_z \end{bmatrix} \end{bmatrix} \begin{bmatrix} \mathrm{d}w_k(u) \\ \mathrm{d}w_z(u) \end{bmatrix}.$$

MX-SSR. In the MX-SSR, the state and measurement vectors are given by $\mathbf{x}_\tau = [\mathbf{x}_\tau^{s\top}, \mathbf{x}_\tau^{f\top}]^\top =$

$[k_\tau^s, z_\tau^s, k_\tau^f, z_\tau^f]^\top$ and $\mathbf{y}_\tau = [\mathbf{y}_\tau^{s,\top}, \mathbf{y}_\tau^{f\top}]^\top = [r_\tau^s, c_\tau^f, n_\tau^f]^\top$, so $n_x^s = 2$, $n_x^f = 2$, $n_y^s = 1$, and $n_y^f = 2$. We next compute the selection matrices $\mathbf{S}_x^f$, $\mathbf{S}_y^f$, and $\mathbf{S}_y^s$. The state-space representation (2.4)-(2.5) in Proposition 1 becomes

$$\begin{bmatrix} k_\tau^s \\ z_\tau^s \\ k_\tau^f \\ z_\tau^f \end{bmatrix} = \left[\begin{array}{cc:cc} & \exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} h\right) & 0 & 0 \\ \begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix}^{-1} & \left(\exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} h\right) - \mathbf{I}_2\right) & 0 & 0 \\ & & 0 & 0 \\ & & 0 & 0 \end{array}\right] \begin{bmatrix} k_{\tau-1}^s \\ z_{\tau-1}^s \\ k_{\tau-1}^f \\ z_{\tau-1}^f \end{bmatrix} + \begin{bmatrix} \eta_{k,\tau}^s \\ \eta_{z,\tau}^s \\ \eta_{k,\tau}^f \\ \eta_{z,\tau}^f \end{bmatrix},$$

$$\begin{bmatrix} r_\tau^s \\ c_\tau^f \\ n_\tau^f \end{bmatrix} = \left[\begin{array}{cc:cc} \phi_{rk} & \phi_{rz} & 0 & 0 \\ \hdashline 0 & 0 & \phi_{ck} & \phi_{cz} \\ 0 & 0 & \phi_{nk} & \phi_{nz} \end{array}\right] \begin{bmatrix} k_\tau^s \\ z_\tau^s \\ k_\tau^f \\ z_\tau^f \end{bmatrix} + \left[\begin{array}{c} 1 \\ \hdashline 0 \\ 0 \end{array}\right] \upsilon_{r,\tau},$$

with

$$\begin{bmatrix} \eta_{k,\tau}^s \\ \eta_{z,\tau}^s \\ \eta_{k,\tau}^f \\ \eta_{z,\tau}^f \end{bmatrix} = \int_{t_{\tau-1}}^{t_\tau} \left[\begin{array}{cc} & \exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} (t_\tau - u)\right) \begin{bmatrix} \sigma_k & 0 \\ 0 & \sigma_z \end{bmatrix} \\ \begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix}^{-1} & \left(\exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} (t_\tau - u)\right) - \mathbf{I}_2\right) \begin{bmatrix} \sigma_k & 0 \\ 0 & \sigma_z \end{bmatrix} \end{array}\right] \begin{bmatrix} \mathrm{d}w_k(u) \\ \mathrm{d}w_z(u) \end{bmatrix}.$$

The values of $\{\phi_{kk}, \phi_{kz}, \phi_{ck}, \phi_{cz}, \phi_{nk}, \phi_{nz}, \phi_{rk}, \phi_{rz}\}$ can be found in Section G of this Online Appendix.

H.2 Mixed-frequency data

Consider the case where measurements are sampled at two different frequencies, i.e., a high frequency (short space between observations) $\bar{h}$, and a low frequency $\underline{h}$, with $\underline{h} > \bar{h}$.

Stock data. Rewrite the transition equation for the S-SSR, or the EM-SSR, in terms of the time step corresponding to the highest frequency available, i.e., $\bar{h}$. Similarly, rewrite the measurement equation as $\mathbf{y}_\tau = \mathbf{W}_\tau \mathbf{C}(\boldsymbol{\theta})$, where $\mathbf{W}_\tau$ is a known time-varying matrix whose rows at a given point in time are a subset of the rows of $\mathbf{I}_{n_y}$. More specifically, the number of rows at time t_τ is determined by the number of variables for which observations are available at t_τ.

For the RBC model discussed in the main text, we consider monthly observations ($\bar{h} = 1/12$) on aggregate consumption, c_τ^s, and quarterly observations ($\underline{h} = 1/4$) on the fraction of hours worked, n_τ^s. Assume that there is no measurement error, i.e., $\boldsymbol{\upsilon}_\tau = \mathbf{0}$

for all τ. Then the S-SSR for mixed-frequency data sampling is given by

$$\begin{bmatrix} k_\tau^s \\ z_\tau^s \end{bmatrix} = \exp\left(\begin{bmatrix} \phi_{kk} & \phi_{kz} \\ 0 & -\rho_z \end{bmatrix} \bar{h}\right) \begin{bmatrix} k_{\tau-1}^s \\ z_{\tau-1}^s \end{bmatrix} + \boldsymbol{\eta}_\tau^s \,, \tag{H.1}$$

$$\begin{bmatrix} c_\tau^s \\ n_\tau^s \end{bmatrix} = \mathbf{W}_\tau \begin{bmatrix} \phi_{ck} & \phi_{cz} \\ \phi_{nk} & \phi_{nz} \end{bmatrix} \begin{bmatrix} k_\tau^s \\ z_\tau^s \end{bmatrix} , \tag{H.2}$$

with $\mathbf{W}_\tau = [\ 1 \quad 0\ ]$ when τ coincides with observations recorded at points in time within a quarter (so only c_τ is available), and $\mathbf{W}_\tau = \mathbf{I}_{n_y}$ when τ coincides with observations recorded at the end of a quarter (so both c_τ^s and n_τ^s are available). The Kalman filter recursions proceed accordingly by accommodating the missing values of the variables sampled at low frequencies, see Durbin and Koopman (2012, Chp. 4.10).[6]

Notice that the state space representation (H.1)-(H.2) is the same as that for the S-SSR, or the EM-SSR, with $\mathbf{W}_\tau = \mathbf{I}_{n_y}$ for all τ, in the case of a common sampling frequency.

Flow data. The state space representation for flow variables is more involved than that for stock data, because we need to keep track of the unobserved time aggregation occurring at "high" frequencies of the variables sampled at lower frequencies. This can be achieved by introducing a number of additional deterministic states that measure the unobserved behavior of the flow variables within the time intervals for which observations are not available. As for the stock case, the state space representation is written in terms of the time step associated with the highest frequency, $\bar{h}$.

For the RBC model discussed in the main text, we consider monthly observations ($\bar{h} = 1/12$) on aggregate consumption, c_τ^f, and quarterly observations ($\underline{h} = 1/4$) on the fraction of hours worked, n_τ^f. Assume that there is no measurement error, i.e., $\boldsymbol{\upsilon}_\tau = \mathbf{0}$ for all τ. Define by $\mathbf{C}(\boldsymbol{\theta})_{n,\bullet}$ the row of $\mathbf{C}(\boldsymbol{\theta})$ related to measurement n^f. Then the F-SSR

[6]The case of mixed-frequency sampling frequency has been addressed in the context of state space models by Harvey and Pierse (1984), Zadrozny (1988), Harvey (1990), Zadrozny (1990), Mariano and Murasawa (2003), Aruoba et al. (2009), Ghysels and Wright (2009), Kuzin et al. (2011), and Bai et al. (2013), among others.

for mixed-frequency sampling is given by

$$\left[\begin{array}{c} k_\tau^s \\ z_\tau^s \\ k_\tau^f \\ z_\tau^f \\ \hdashline n_{\tau-1}^f \\ n_{\tau-2}^f \end{array}\right] = \left[\begin{array}{c:cccc} \mathbf{A}_{\bar{h}}(\boldsymbol{\theta}) & \multicolumn{4}{c}{\multirow{2}{*}{$\mathbf{0}_{4\times 4}$}} \\ \mathbf{A}(\boldsymbol{\theta})^{-1}(\mathbf{A}_{\bar{h}}(\boldsymbol{\theta}) - \mathbf{I}) & & & & \\ \hdashline \multirow{2}{*}{$\mathbf{0}_{2\times 2}$} & 0 & \mathbf{C}(\boldsymbol{\theta})_{n,\bullet} & 0 & 0 \\ & 0 & 0 & 1 & 0 \end{array}\right] \left[\begin{array}{c} k_{\tau-1}^s \\ z_{\tau-1}^s \\ k_{\tau-1}^f \\ z_{\tau-1}^f \\ \hdashline n_{\tau-2}^f \\ n_{\tau-3}^f \end{array}\right] + \left[\begin{array}{c} \boldsymbol{\eta}_\tau^s \\ \boldsymbol{\eta}_\tau^f \\ \hdashline 0 \\ 0 \end{array}\right] \tag{H.3}$$

$$\left[\begin{array}{c} c_\tau^f \\ n_\tau^f \end{array}\right] = \mathbf{W}_\tau \left[\begin{array}{cc:cc} \multicolumn{2}{c}{\multirow{2}{*}{$\mathbf{0}_{2\times 2}$}} & \multicolumn{2}{c}{\multirow{2}{*}{$\mathbf{C}(\boldsymbol{\theta})$}} \\ & & & \\ \hdashline 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{array}\right] \left[\begin{array}{c} k_\tau^s \\ z_\tau^s \\ k_\tau^f \\ z_\tau^f \\ \hdashline n_{\tau-1}^f \\ n_{\tau-2}^f \end{array}\right], \tag{H.4}$$

where $\mathbf{W}_\tau$ is defined as in the stock case. Notice that at any point in time, t_τ, the observed fraction of hours worked in (H.4), sampled at a lower frequency $\underline{h}$, n_τ^f, is related to the higher frequency state equations according to

$$\begin{aligned} n_\tau^f = n^f(t_\tau) &= \int_{t_\tau - \underline{h}}^{t_\tau} n(u)\mathrm{d}u \\ &= \int_{t_\tau - \overline{h}}^{t_\tau} n(u)\mathrm{d}u + \int_{t_\tau - 2\overline{h}}^{t_\tau - \overline{h}} n(u)\mathrm{d}u + \int_{t_\tau - \underline{h}}^{t_\tau - 2\overline{h}} n(u)\mathrm{d}u \\ &= n_\tau^f + n_{\tau-1}^f + n_{\tau-2}^f \,, \end{aligned}$$

where $n_\tau^f := \int_{t_\tau - \overline{h}}^{t_\tau} n(u)\mathrm{d}u$, $n_{\tau-1}^f := \int_{t_\tau - 2\overline{h}}^{t_\tau - \overline{h}} n(u)\mathrm{d}u$, and $n_{\tau-2}^f := \int_{t_\tau - \underline{h}}^{t_\tau - 2\overline{h}} n(u)\mathrm{d}u$.

Remark H.1 The transition equations for the additional deterministic state variables in (H.3), i.e., those associated with $n_{\tau-2}^f$ and $n_{\tau-3}^f$, do not depend on the vector of unknown parameters. Therefore, it is possible to use $\boldsymbol{\theta}$ estimated from observations sampled at a low frequency $\underline{h}$ to recover state estimates at a higher frequency $\overline{h}$ by simply using the state space representation in (H.1)-(H.2) or (H.3)-(H.4) together with any filtering algorithm that handles missing observations. This means that the model does not need to be re-estimated if the analysis entails the same variables observed at different frequencies.

I. Additional tables

I.1 Finite sample properties for extended vector of parameters

Table I.1. Finite sample properties for extended vector of parameters. The table reports statistics for estimates of $\boldsymbol{\theta} = [\rho, \eta, \rho_z, \sigma_z, \sigma_k]^\top$ from $M = 10,000$ samples of quarterly ($h = 1/4$) observations on aggregate consumption (C) and hours worked (N), generated over a period of 60 years ($T = 240$ observations in each sample). Simulated measurements in Panel A are sampled as stocks, and those in Panel B as flows. The share of capital in output, α, and the depreciation rate, δ, are calibrated to their population values in Table 1. Let $\hat{\boldsymbol{\theta}}_m$ denote the estimates from the m-th sample. The table shows bias (Bias $= M^{-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0)$) and root mean squared error (RMSE $= (M^{-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0)^2)^{1/2}$) across repetitions.

Panel A: Data sampled as stocks

		F-SSR		S-SSR		EM-SSR	
	$\boldsymbol{\theta}$	Bias	RMSE	Bias	RMSE	Bias	RMSE
ρ	0.03	0.0005	0.0016	0.0001	0.0016	4.13e-05	0.0016
η	0.02	0.0006	0.0023	4.22e-05	0.0022	-0.0001	0.0022
ρ_z	0.2052	0.0250	0.0343	0.0179	0.0274	-0.0171	0.0239
σ_z	0.014	0.0043	0.0044	-0.0001	0.0007	-0.0001	0.0007
σ_k	0.0104	0.0032	0.0033	-0.0001	0.0005	-0.0004	0.0006

Panel B: Data sampled as flows

		F-SSR		S-SSR		EM-SSR	
	$\boldsymbol{\theta}$	Bias	RMSE	Bias	RMSE	Bias	RMSE
ρ	0.03	0.0001	0.0016	-3.35e-05	0.0016	-0.0002	0.0018
η	0.02	4.53e-05	0.0022	-0.0002	0.0022	-0.0006	0.0027
ρ_z	0.2052	0.0177	0.0272	0.0117	0.0239	-0.0223	0.0282
σ_z	0.014	-0.0001	0.0007	-0.0026	0.0027	-0.0026	0.0027
σ_k	0.0104	-0.0001	0.0005	-0.002	0.002	-0.0022	0.0023

I.2 Finite sample properties for linearized DGP

Table I.2. Finite sample properties, data simulated from the linearized economy. The table reports statistics for estimates of $\boldsymbol{\theta} = [\rho_z, \sigma_z, \sigma_k]^\top$ from $M = 10,000$ samples of quarterly ($h = 1/4$) observations on aggregate consumption (C) and hours worked (N), generated over a period of 60 years ($T = 240$ observations in each sample). Simulated measurements in Panel A are sampled as stocks, and those in Panel B as flows. Remaining parameters are calibrated to their population values in Table 1. Let $\hat{\boldsymbol{\theta}}_m$ denote the estimates from the m-th sample. The table shows bias (Bias $= M^{-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0)$) and root mean squared error (RMSE $= (M^{-1}\sum_{m=1}^{M}(\hat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_0)^2)^{1/2}$) across repetitions.

Panel A: Data sampled as stocks

		F-SSR		S-SSR		EM-SSR	
	$\boldsymbol{\theta}$	Bias	RMSE	Bias	RMSE	Bias	RMSE
ρ_z	0.2052	0.0066	0.0222	0.0019	0.0189	-0.0295	0.0332
σ_z	0.014	0.0043	0.0044	-0.0001	0.0007	-0.0001	0.0007
σ_k	0.0104	0.0032	0.0033	-4.64e-05	0.0005	-0.0003	0.0006

Panel B: Data sampled as flows

		F-SSR		S-SSR		EM-SSR	
	$\boldsymbol{\theta}$	Bias	RMSE	Bias	RMSE	Bias	RMSE
ρ_z	0.2052	0.0017	0.0189	-0.0029	0.0191	-0.0321	0.0358
σ_z	0.014	-0.0001	0.0007	-0.0026	0.0027	-0.0026	0.0027
σ_k	0.0104	-4.96e-05	0.0005	-0.002	0.002	-0.0022	0.0022

I.3 Normality tests

Table I.3. Normality tests. Battery of normality tests for estimated F-SSR without measurement error (see Table 4, first column). The table reports test statistics and associated p-values (in brackets) for alternative tests assessing the normality of the prediction errors, $\boldsymbol{\nu}_{\tau|\tau-1}$. The tests are performed over different subperiods of the full estimation period, as indicated in column "Period". The tests are the Shapiro-Wilk test (SW, Shapiro and Wilk, 1965) of univariate normality of the prediction errors associated with consumption, v_C, and hours worked, v_N; the Kolmogorov-Smirnov test (KS) on v_C, and v_N; the Jarque-Bera test (JB); and the omnibus approximate normality test (DH, Doornik and Hansen, 2008). The null of Gaussianity is rejected at the 5% level for p-values less than 0.05. The first observations from 1960 are excluded from the exercise.

Period	SW		KS		JB		DH
	ν_C	ν_N	ν_C	ν_N	ν_C	ν_N	(ν_C, ν_N)
01-Jul-1960 - 01-Oct-2019	0.99 [0.035]	0.95 [4e-06]	0.08 [0.127]	0.07 [0.167]	5.71 [0.050]	55.83 [0.001]	30.44 [4e-06]
01-Jul-1960 - 01-Jul-1967	0.98 [0.796]	0.96 [0.406]	0.20 [0.195]	0.15 [0.504]	0.34 [0.500]	1.13 [0.396]	1.32 [0.858]
01-Jul-1967 - 01-Jul-1974	0.92 [0.042]	0.94 [0.142]	0.14 [0.633]	0.13 [0.700]	2.29 [0.129]	1.86 [0.184]	9.61 [0.048]
01-Jul-1974 - 01-Jul-1981	0.98 [0.806]	0.92 [0.045]	0.16 [0.400]	0.10 [0.934]	0.31 [0.500]	4.92 [0.041]	12.47 [0.014]
01-Jul-1981 - 01-Jul-1988	0.97 [0.480]	0.87 [0.004]	0.13 [0.677]	0.23 [0.084]	0.74 [0.500]	9.59 [0.013]	12.51 [0.014]
01-Jul-1988 - 01-Jul-1995	0.96 [0.308]	0.96 [0.350]	0.07 [0.996]	0.12 [0.798]	1.46 [0.283]	1.19 [0.371]	6.90 [0.141]
01-Jul-1995 - 01-Jul-2002	0.92 [0.036]	0.98 [0.726]	0.09 [0.971]	0.12 [0.774]	7.61 [0.020]	0.85 [0.500]	11.38 [0.023]
01-Jul-2002 - 01-Jul-2009	0.98 [0.786]	0.86 [0.003]	0.10 [0.928]	0.20 [0.190]	0.27 [0.500]	7.88 [0.019]	3.84 [0.427]
01-Jul-2009 - 01-Jul-2016	0.96 [0.300]	0.97 [0.465]	0.09 [0.945]	0.16 [0.412]	1.21 [0.366]	0.30 [0.500]	5.17 [0.270]

J. Additional figures

J.1 Measurements

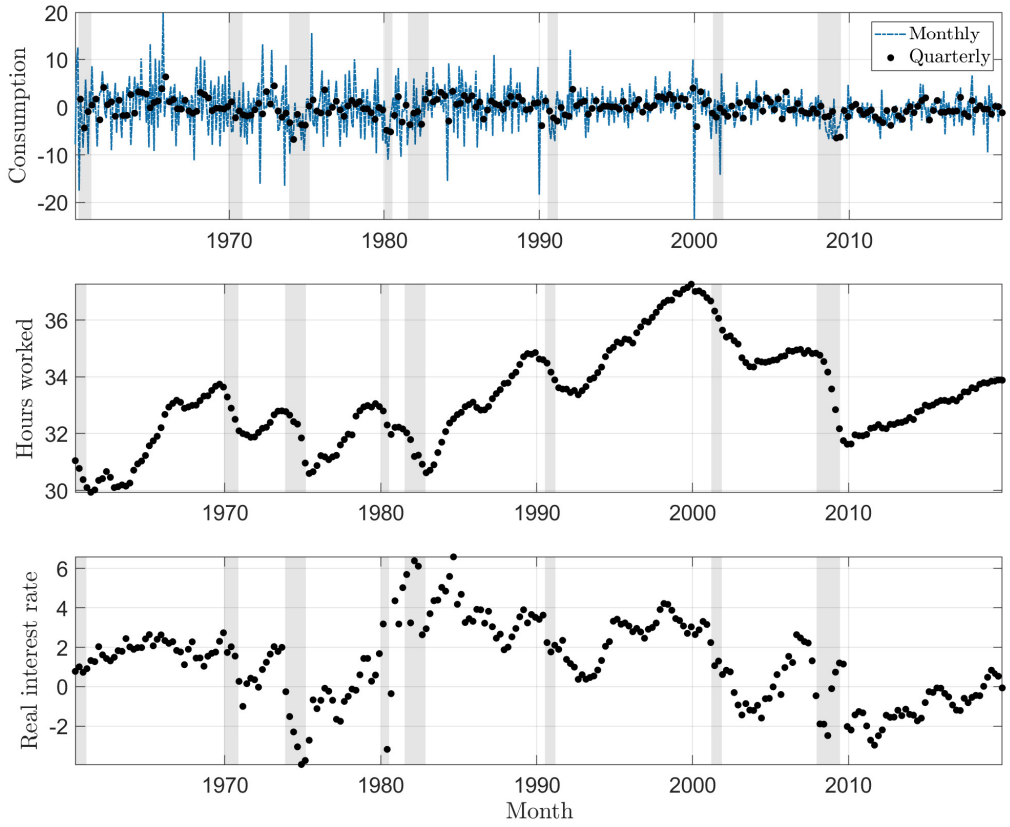
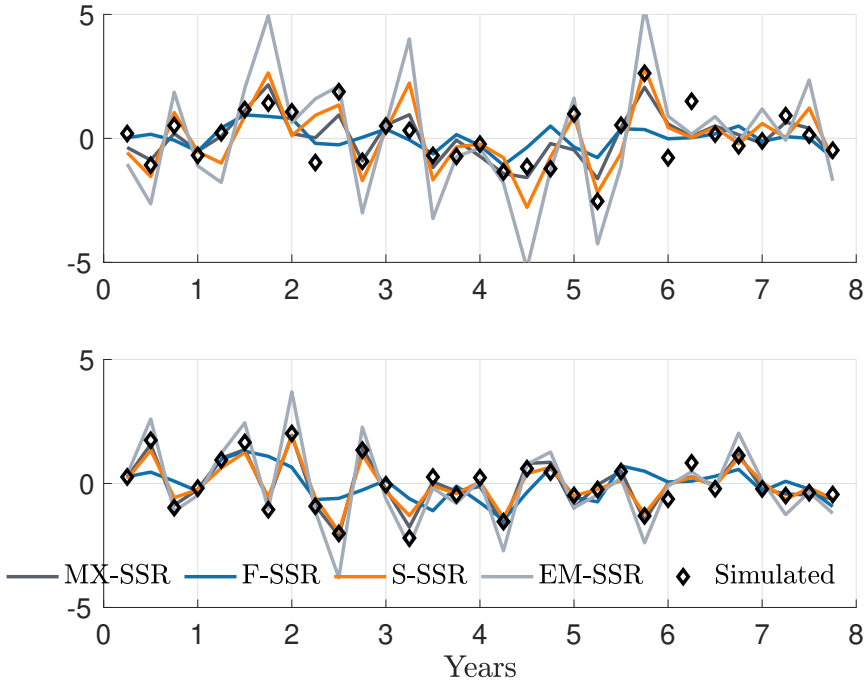
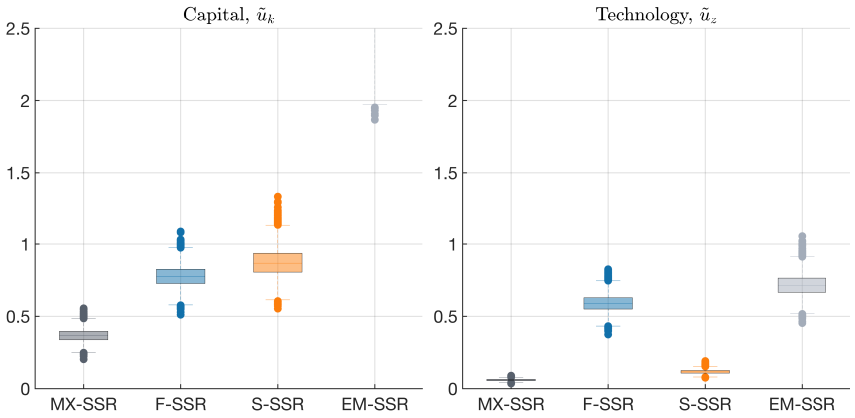


Figure J.1. Measurements. The figure shows the time series of observed measurement used in the different ML estimations reported in the paper: Annualized monthly (dashed line) and quarterly (dots) growth rate of real aggregate consumption per capita (top exhibit); logarithm of quarterly hours worked, expressed as percentage deviation from steady state (middle); and quarterly gross annual real interest rate (bottom) for the U.S., over the period from 1960 through 2019. Light gray vertical bands indicate NBER recessions.

J.2 Structural shocks, mixed sampling



(a) Structural shocks



(b) Mean Squared Errors

Figure J.2. Structural shocks, mixed sampling. Exhibit (a) shows the time series of structural shocks to TFP, $\tilde{u}_z$ (top portion of exhibit), and capital stock, $\tilde{u}_k$ (bottom portion), recovered from a simulated sample of flows for consumption, C, and stocks for the interest rate, r. Exhibit (b) shows distributions of the mean squared error (MSE) between the true (simulated) structural shocks and their recovered (smoothed) counterparts. Each boxplot represents the distribution of MSE across Monte Carlo replication.

J.3 Log-likelihood profiles

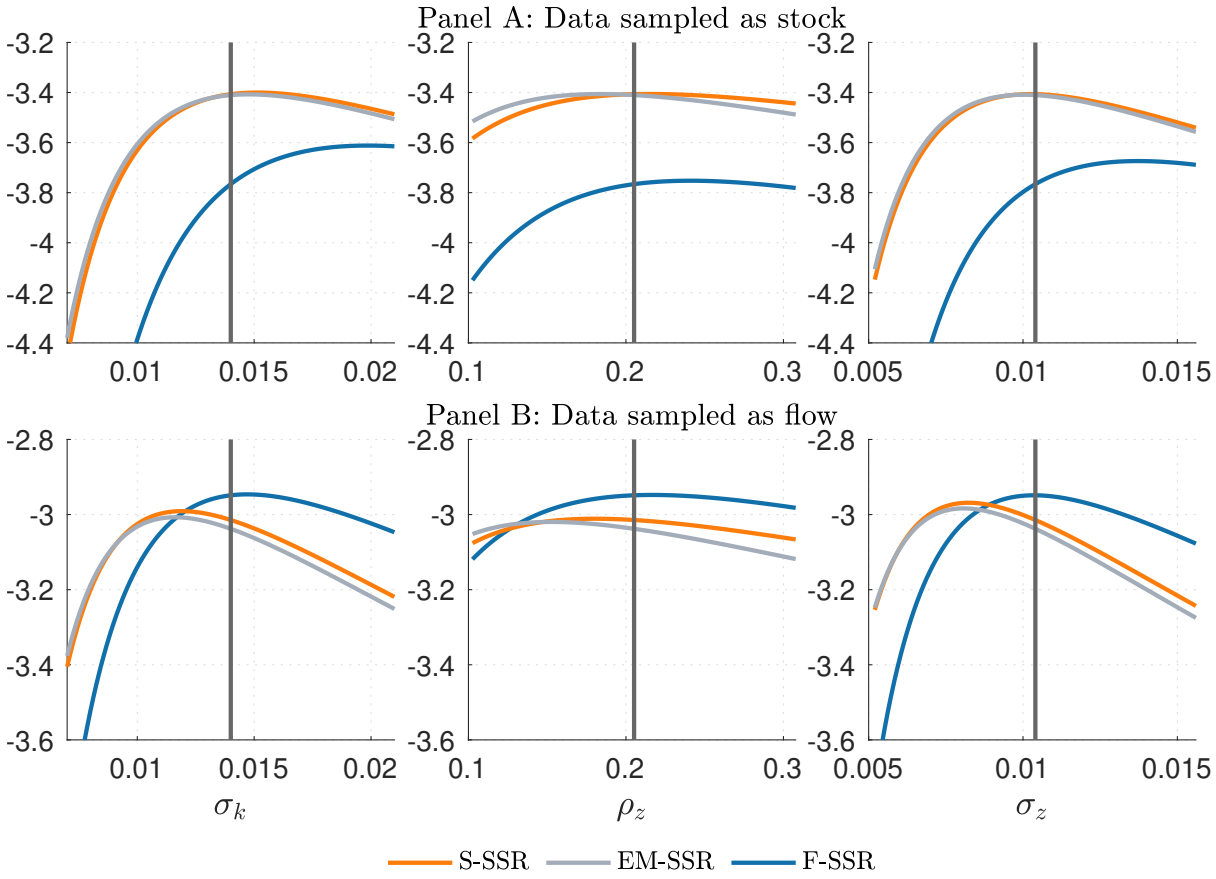


Figure J.3. Log-likelihood profiles. The figure shows the log-likelihood function $\mathcal{L}(\boldsymbol{\theta}|\mathbf{y}^T)$ for each of the parameters σ_k, ρ_z, and σ_z, keeping the remaining parameters at their population values in Table 1, and for each of the methods S-SSR, EM-SSR, and F-SSR. Simulated measurements are sampled as stocks in Panel A, and as flows in Panel B. The vertical solid line shows the true population value of the parameter. Graphs are generated using a single simulated sample of quarterly observations for a period of 60 years. Different samples provide similar conclusions.

J.4 Latent states

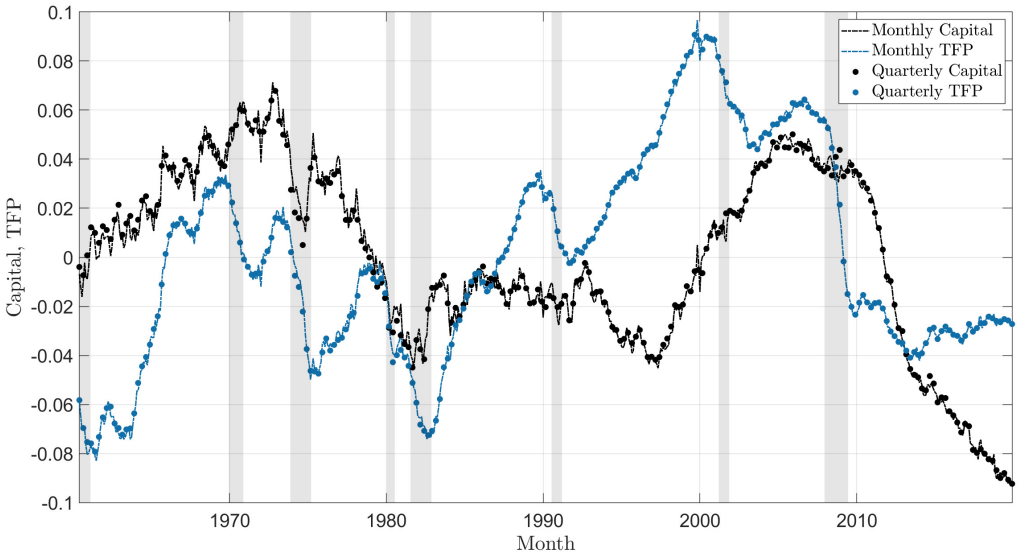


Figure J.4. Latent states (F-SSR method). Smoothed monthly and quarterly series for aggregate capital stock and TFP (% deviations from corresponding steady-state values). The sample covers the period from 1960 trough 2019. The series are obtained by the F-SSR method after parameter estimation and state smoothing, for measurements using quarterly aggregate consumption and hours worked for the quarterly estimates, monthly aggregate consumption and quarterly hours worked for the monthly estimates. Light gray vertical bands indicate NBER recessions.

K. Historical shock decomposition

K.1 Derivation of Equation (6.1)

The historical shock decomposition (see Lütkepohl, 2005 and Canova 2007) links the structural shocks of the system to the observables. Given Assumption 1 and Proposition 4, let $\boldsymbol{\epsilon} := \boldsymbol{\eta}$ in the ABCD representation (2.12)-(2.13), so that there is no measurement error, wlog. Then the transition equation can be written as

$$\mathbf{x}_\tau = (\mathbf{I}_{n_x} - \mathsf{A}(\boldsymbol{\theta}, h)L)^{-1}\mathsf{B}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_\tau + \mathcal{O}_P(h^{3/2}) = \sum_{j=0}^{\infty} \mathsf{A}(\boldsymbol{\theta}, h)^j \mathsf{B}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_{\tau-j} + \mathcal{O}_P(h^{3/2})\,.$$

We disregard the $\mathcal{O}_P(h^{3/2})$ term and substitute the definition of $\mathbf{x}_\tau$ in the measurement equation to obtain

$$\begin{aligned}\mathbf{y}_{\tau+1} &= \mathsf{C}(\boldsymbol{\theta}, h)(\mathbf{I}_{n_x} - \mathsf{A}(\boldsymbol{\theta}, h)L)^{-1}\mathsf{B}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_\tau + \mathsf{D}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_{\tau+1} \\ &= \mathsf{C}(\boldsymbol{\theta}, h)\sum_{j=0}^{\infty} \mathsf{A}(\boldsymbol{\theta}, h)^j \mathsf{B}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_{\tau-j} + \mathsf{D}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_{\tau+1}\,,\end{aligned}$$

where $L := L_h = L(h)$ is the lag operator, such that $L\mathbf{x}_\tau = \mathbf{x}_{\tau-1}$. Conditioning on the beginning of the sample, the transition equation can be alternatively written as

$$\mathbf{x}_\tau = \mathsf{A}(\boldsymbol{\theta}, h)^\tau \mathbf{x}_0 + \sum_{j=0}^{\tau-1} \mathsf{A}(\boldsymbol{\theta}, h)^j \mathsf{B}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_{\tau-j}\,,$$

so the structural shocks are related to the measurements through

$$\begin{aligned}\mathbf{y}_{\tau+1} &= \mathsf{C}(\boldsymbol{\theta}, h)\mathsf{A}(\boldsymbol{\theta}, h)^\tau \mathbf{x}_0 + \mathsf{D}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_{\tau+1} + \mathsf{C}(\boldsymbol{\theta}, h)(\mathbf{I}_{n_x} - \mathsf{A}(\boldsymbol{\theta}, h)L)^{-1}\mathsf{B}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_\tau \\ &= \mathsf{C}(\boldsymbol{\theta}, h)\mathsf{A}(\boldsymbol{\theta}, h)^\tau \mathbf{x}_0 + \mathsf{D}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_{\tau+1} + \mathsf{C}(\boldsymbol{\theta}, h)\sum_{j=0}^{\tau-1} \mathsf{A}(\boldsymbol{\theta}, h)^j \mathsf{B}(\boldsymbol{\theta}, h)\mathbf{H}(\boldsymbol{\theta})\mathbf{u}_{\tau-j}\,.\end{aligned}$$

The historical contributions of the individual structural shock u_i, $i = 1, \ldots, m_w$ to the measurements at the τth observation are then given by selecting and propagating the individual shocks one at a time. In our analysis, we use the ML estimates $\hat{\boldsymbol{\theta}}$ of the structural parameters. Additionally, we replace the unobserved $\mathbf{x}_0$ and $\mathbf{u}$ with their smoothed estimates, using approximations $\tilde{\mathbf{u}}$ instead of $\mathbf{u}$.

K.2 Decomposition using S-SSR method

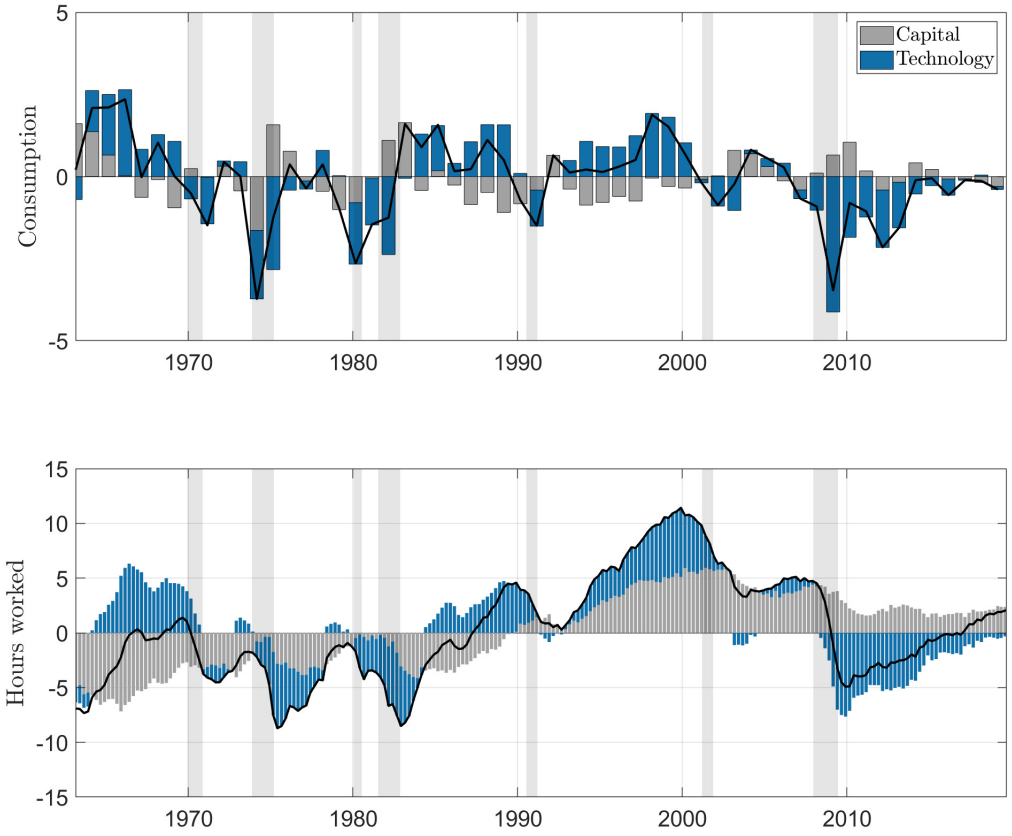


Figure K.1. Historical shock decomposition (S-SSR method). The figure shows the separate historical contributions of the structural shocks to capital (demand) and technology (supply, TFP) recovered by the S-SSR method from the observed measurements of consumption and hours worked over the period from 1948:Q1 through 2019:Q4. The black solid line in the upper panel represents annual consumption growth rates, and that in the lower panel quarterly percentage deviations of hours worked from its steady state ($n^\star = 33\%$). Light gray vertical bands indicate NBER recessions.

K.3 Decomposition using EM-SSR method

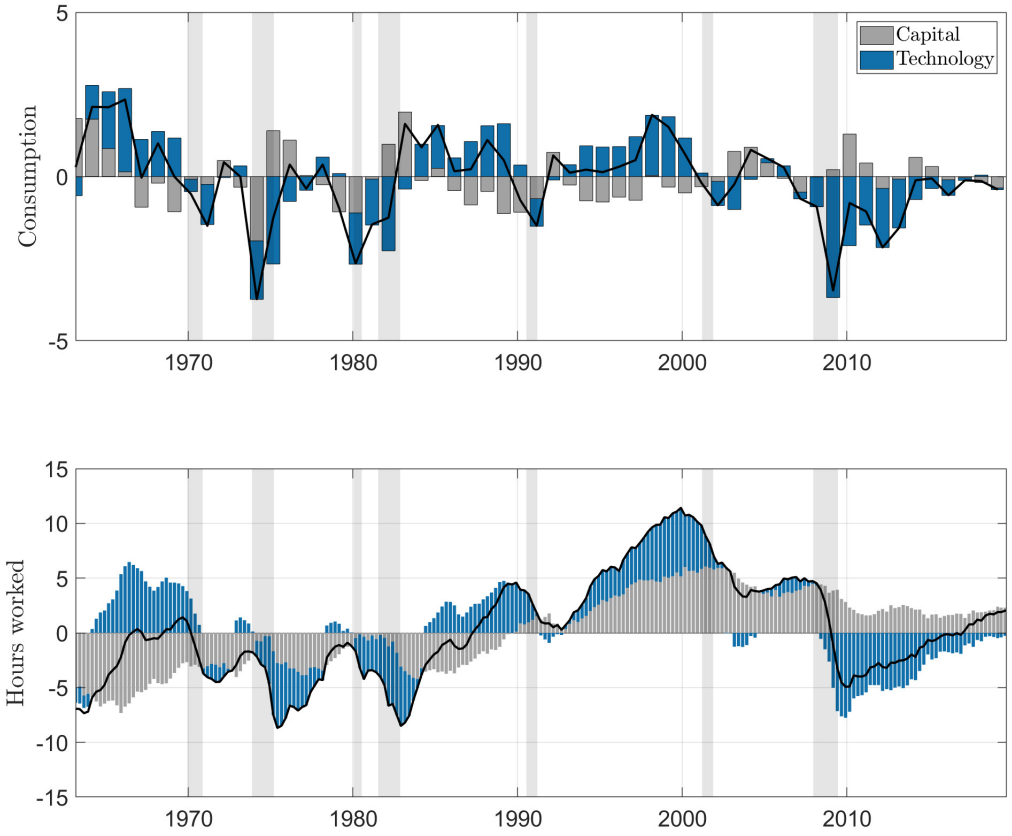


Figure K.2. Historical shock decomposition (EM-SSR method). The figure shows the separate historical contributions of the structural shocks to capital (demand) and technology (supply, TFP) recovered by the EM-SSR method from the observed measurements of consumption and hours worked over the period from 1948:Q1 through 2019:Q4. The black solid line in the upper panel represents annual consumption growth rates, and that in the lower panel quarterly percentage deviations of hours worked from its steady state ($n^\star = 33\%$). Light gray vertical bands indicate NBER recessions.

K.4 Decomposition using MX-SSR method

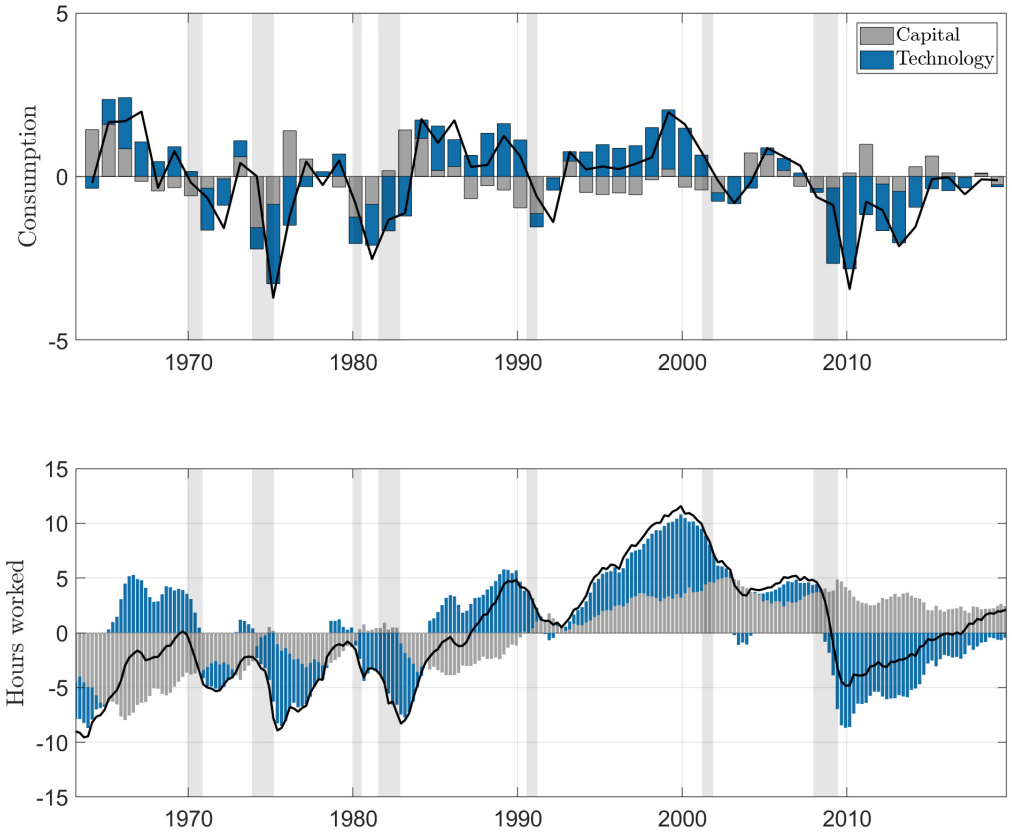


Figure K.3. Historical shock decomposition (MX-SSR method). The figure shows the separate historical contributions of the structural shocks to capital (demand) and technology (supply, TFP) recovered by the MX-SSR method from the observed measurements of consumption, hours worked, and the real interest rate over the period from 1948:Q1 through 2019:Q4. The black solid line in the upper panel represents annual consumption growth rates, and that in the lower panel quarterly percentage deviations of hours worked from its steady state ($n^\star = 33\%$). Light gray vertical bands indicate NBER recessions.

References

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2018): "When Inequality Matters for Macro and Macro Matters for Inequality," *NBER Macroeconomics Annual*, 32, 1–75.

Aruoba, S., F. X. Diebold, and C. Scotti (2009): "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 27, 417–427.

Bai, J., E. Ghysels, and J. Wright (2013): "State Space Models and MIDAS Regressions," *Econometric Reviews*, 32, 779–813.

Bergstrom, A. R. (1984): "Continuous Time Stochastic Models and Issues of Aggregation over Time," in *Handbook of Econometrics*, Elsevier, vol. 2, 1145–1212.

Buiter, W. H. (1984): "Saddlepoint Problems in Continuous Time Rational Expectations Models: A General Method and Some Macroeconomic Examples," *Econometrica*, 52, 665–680.

Canova, F. (2007): *Methods for Applied Macroeconomic Research*, Princeton University Press.

Chang, F.-R. (2009): *Stochastic Optimization in Continuous Time*, Cambridge University Press.

Doornik, J. A. and H. Hansen (2008): "An Omnibus Test for Univariate and Multivariate Normality," *Oxford Bulletin of Economics and Statistics*, 70, 927–939.

Durbin, J. and S. J. Koopman (2012): *Time Series Analysis by State Space Methods*, Oxford University Press.

Ghysels, E. and J. H. Wright (2009): "Forecasting Professional Forecasters," *Journal of Business & Economic Statistics*, 27, 504–516.

Golub, G. H. and C. F. Van Loan (2013): *Matrix Computations*, Johns Hopkins University Press, 4th ed.

Harvey, A. C. (1990): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, A. C. and R. G. Pierse (1984): "Estimating Missing Observations in Economic Time Series," *Journal of the American Statistical Association*, 79, 125–131.

Jewitt, G. and J. R. McCrorie (2005): "Computing Estimates of Continuous Time Macroeconometric Models on the Basis of Discrete Data," *Computational Statistics & Data Analysis*, 49, 397–416.

Komunjer, I. and S. Ng (2011): "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," *Econometrica*, 79, 1995–2032.

Kuzin, V., M. Marcellino, and C. Schumacher (2011): "MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area," *International Journal of Forecasting*, 27, 529–542.

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

Mariano, R. S. and Y. Murasawa (2003): "A New Coincident Index of Business Cycles based on Monthly and Quarterly Series," *Journal of Applied Econometrics*, 18, 427–443.

Moler, C. and C. Van Loan (1978): "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later," *SIAM Review*, 20, 801–836.

——— (2003): "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later," *SIAM Review*, 45, 3–49.

Parra-Alvarez, J. C., H. Polattimur, and O. Posch (2021): "Risk Matters: Breaking Certainty Equivalence in Linear Models," *Journal of Economics Dynamics and Control*, 133.

Phillips, P. C. B. (1973): "The Problem of Identification in Finite Parameter Continuous Time Models," *Journal of Econometrics*, 1, 351–362.

Shapiro, S. S. and M. B. Wilk (1965): "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, 52, 591–611.

Simon, H. A. (1956): "Rational Choice and the Structure of the Environment." *Psychological review*, 63, 129.

Sims, C. A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20.

Theil, H. (1957): "A Note on Certainty Equivalence in Dynamic Planning," *Econometrica: Journal of the Econometric Society*, 346–349.

Van Loan, C. (1978): "Computing Integrals Involving the Matrix Exponential," *IEEE Transactions on Automatic Control*, 23, 395–404.

Zadrozny, P. (1988): "Gaussian Likelihood of Continuous-Time ARMAX Models when Data are Stocks and Flows at Different Frequencies," *Econometric Theory*, 4, 108–124.

Zadrozny, P. A. (1990): "Forecasting U.S. GNP at Monthly Intervals with an Estimated Bivariate Time Series Model," *Economic Review*, 2–15.