Dynamic Identification of DSGE Models

Ivana Komunjer* Serena Ng†

Second Draft: November 2010
First Draft: October 2009
Comments Welcome

Abstract

This paper studies structural identification of parameters of a DSGE model from the first and second moments of the data. Classical results for dynamic simultaneous equations do not apply because the state space solution of the model does not constitute a standard reduced form. The rank of the Jacobian matrix of derivatives with respect to the parameters is necessary but not sufficient for identification. We use restrictions implied by observational equivalence to obtain two sets of rank and order conditions: one for stochastically singular and another for non-singular models. Measurement errors, mean, long run, and a priori restrictions can be accommodated. An example is considered to illustrate the results.

Keywords: Structural identification, Stochastic singularity, Measurement errors, Similarity transform, Spectral factorization.

JEL Classification: C1, C3, E1

*University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. Email: komunjer@ucsd.edu
†Columbia University, 420 W. 118 St. MC 3308, New York, NY 10027. Email: serena.ng@columbia.edu
This paper was presented at 2009 NBER Summer Institute, NBER/NSF Time Series Conference, CESG Meeting, All UC Conference, as well as the econometrics seminars at UC Davis, Harvard/MIT, NYU, Princeton, and Duke/UNC/NC State. We thank seminar participants and Richard Davis, Nikolay Iskrev, Victor Solo, as well as three anonymous referees and the co-editor for helpful comments. Part of this work was done while the first author was visiting Columbia University whose warm hospitality is gratefully acknowledged. The authors acknowledge financial support from the National Science Foundation SES-0962473 and SES-0962431.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models have now reached the level of sophistication to permit analysis of important macroeconomic issues. Whereas the model parameters (θ) used to be calibrated, numerical advances have made it possible to estimate models with as many as a hundred parameters. Researchers are, however, aware that not all the parameters can be consistently estimated because of identification failure: that is, changes in some of the parameters may lead to indistinguishable outcomes. This paper studies local identification of a DSGE model from its linearized solution. We use the restrictions implied by observationally equivalent spectral densities to obtain rank and order conditions for identification. Mean and long-run restrictions can be incorporated in the form of a priori restrictions. The error terms can be the primitive shocks in the model but can also be specification or measurement errors.

The literature on identification of DSGE models is relatively small. Canova and Sala (2009) drew attention to the identification problem and suggest to plot some statistics of the estimated model evaluated at different parameter values. Consolo, Favero, and Paccagnini (2009) compare the properties of the DSGE model with those of a factor augmented VAR. Both approaches, while useful, do not provide formal conditions for identification. Rubio-Ramírez, Waggoner, and Zha (2007) study identification of structural VARs but not DSGE models per se. The most complete analysis to date is due to Iskrev (2010) who proposes to evaluate the derivatives with respect to θ of some $J < T$ model-implied autocovariances, where T is the sample size. His results depend on J and the autocovariance matrices need to be solved numerically. We do not compute any autocovariances. Instead, we study the implications of observational equivalence for the canonical model that generates the autocovariances. This leads to a finite system of nonlinear equations with derivative matrix $\Delta(\theta)$. The rank of $\Delta(\theta)$ and the order of the system provide the necessary and sufficient conditions for identification of θ.

Our analysis has two distinctive features. First, the restrictions used in the identification analysis are derived from the structural features of the linearized DSGE model. This *structural identification* approach appears new to the econometrics literature. It has its roots in control theory which typically assumes that both the outputs (endogenous variables) and the inputs (innovations) are observed. In such cases, restrictions on the transfer function of minimal systems are necessary and sufficient for identification. But DSGE models have latent shocks with unknown variances and the econometrician only observes some of the outputs. Thus we derive new identification conditions that also exploit restrictions on transfer functions that are left-invertible. Minimality and left-invertibility, which will be explained below, allow us to by-pass direct calculations of autocovariances and instead focus on the determinants of these autocovariances. The $\Delta(\theta)$ matrix

sheds light on whether non-identification is due to parameter dependency, or delicate interactions between the impulse and the propagating mechanism of the model. The null space of $\Delta(\theta)$ helps isolate which parameters are responsible for non-identification.

Second, the identification conditions for singular and non-singular systems are studied separately. Classical identification analysis tend to focus on models with the same number of endogenous variables as errors. The results do not usually hold when there are fewer errors than observables, a condition known as 'stochastic singularity'. Models with fewer observables than errors are non-singular; however, the innovations is such models are usually non-fundamental. As each case has specific implications about what can be uncovered from the observables, the identification restrictions are also different. Nonetheless, in the important special case when the system is 'square', the two sets of conditions coincide. The conditions depend on the number of state variables and the number of innovations irrespective of the sample size.

Before turning to the main analysis, it is useful to better understand what is unusual about the identification of DSGE models. Consider a model parameterized by θ whose linearized solution takes a PQRS form:

$$\begin{aligned} k_{t+1} &= P(\theta)k_t + Q(\theta)z_{t+1} \\ w_{t+1} &= R(\theta)k_t + S(\theta)z_{t+1} \\ z_{t+1} &= \Psi_z(\theta)z_t + \epsilon_{zt+1}, \qquad \epsilon_{zt} \sim WN(0, \Sigma_z(\theta)), \end{aligned}$$

where k_t is the endogenous state vector, w_t are jump variables, and z_t are exogenous shocks with innovations ϵ_{zt}. To begin with, observe that $y_t' = (k_t', w_t')$ is a VARMA process. Identification of θ would be futile unless the VARMA parameters were identifiable even though we are not interested in these parameters per se. This, however, is not a trivial problem because VARMA models are potentially 'exchangeable' and common factors can exist unless a so-called left-coprime condition is satisfied.[1] The echelon (canonical) VARMA model and its order, also known as McMillan degree, can be solved only when the system is small.[2] A problem that further complicates identification is that the VARMA representations of stochastically singular models involve matrices that are generally not square. This violates the usual assumptions used in Deistler (1976), for example. Thus, no attempt has been made to identify DSGE models directly from the VARMA representations.

We can also view k_{t+1} and w_{t+1} as dependent variables being expressed in terms of the predetermined variables k_t and serially correlated errors z_{t+1}. Then the equations have the flavor

[1] Exchangeable means that two processes can have identical moving average representations if there is a unimodular factor in the autoregressive or moving-average matrix polynomial. For example, Thus a VMA model $y_t = U(L)\Theta(L)\epsilon_t$ is equivalent to the VARMA model $U(L)^{-1}y_t = \Theta(L)\epsilon_t$ (see, e.g., Section 2.3 in Reinsel, 2003). The co-primeness condition rules out redundant polynomials that lead to an overparameteried model (see, e.g., Hannan, 1971).

[2] See, e.g., Solo (1986), Reinsel (2003), and Lutkepohl (2005).

of a reduced form which is convenient for discussing why classical results of Koopmans (1950), Fisher (1966), Rothenberg (1971), Hausman and Taylor (1983) are not applicable. First, many of these results apply to static models in which the innovations of the reduced form are orthogonal to the predetermined regressors. Obviously, DSGE models are dynamic, and k_t in our model is correlated with z_{t+1} whenever $\Psi_z(\theta) \neq 0$. Second, unlike iid data with time invariant distributions, the probability distributions of time series data may change over time. Thus, likelihood based identification analysis may not be appropriate. Third, results for full rank dynamic models as in Hatanaka (1975), Hannan (1971), Sargan (1977) all assume the presence of exogenous variables. However, there are no exogenous variables in DSGE models other than the latent shocks which cannot be used for identification. Fourth, results for full rank systems do not easily extend to singular systems. Though we can drop some variables so that the system is full rank, the results will not be robust unless we know which variables are ancillary for the parameters of interest.

Finally, and perhaps the most important reason why classical identification results are invalid is that the rank conditions of Rothenberg (1971) rest on the assumption that the reduced form parameters are identifiable. However, as Wallis (1977) and others have pointed out, there may be common factors in the rational polynomial matrices that relate the exogenous to the endogenous variables. In such cases, the $(P(\theta), Q(\theta), R(\theta), S(\theta))$ parameters may themselves be *not identifiable*. This non-identifiability of the 'reduced form' model should not come as a surprise because as already pointed out, the VARMA parameters for the same data may also be non-identifiable.

The rest of the paper is organized as follows. Section 2 sets up the econometric framework. Sections 3 and 4 present rank and order conditions derived for singular and non-singular models, respectively. Section 5 studies partial and conditional identification under a priori restrictions. Section 6 gives two different perspectives on our results and relates them to the information matrix. An illustration is given in Section 7. The key steps of the proofs are given in the Appendix. A detailed long proof is available on request.

2 Setup

Consider a DSGE model with (deep) parameters θ belonging to a set $\Theta \subseteq \mathbb{R}^{n_\theta}$. The variables of the model, denoted by X_t^a, are driven by structural shocks with innovations ϵ_{zt}. The model is characterized by a set of equations that define the steady state values $X_{ss}^a(\theta)$ and Euler equations that describe the transition dynamics. Linearizing around the steady state gives a system of expectational difference equations $E_t\Gamma_0(\theta)X_{t+1}^a = \Gamma_1(\theta)X_t^a + \epsilon_{zt}$ that can be solved to yield a

solution in the form of difference equations.[3] Let X_t be a $n_X \times 1$ state vector that is a subvector of X_t^a, and let Y_t be a $n_Y \times 1$ vector of observables. We allow for measurement and specification errors whose innovations are ϵ_{vt}. Collect all the innovations into a $n_\epsilon \times 1$ vector $\epsilon_t = (\epsilon_{zt}', \epsilon_{vt}')'$. The state space solution of Y_t is given by the transition and measurement equations:

$$\underbrace{X_{t+1}}_{n_X \times 1} = \underbrace{A(\theta)}_{n_X \times n_X} \underbrace{X_t}_{n_X \times 1} + \underbrace{B(\theta)}_{n_X \times n_\epsilon} \underbrace{\epsilon_{t+1}}_{n_\epsilon \times 1} \tag{1a}$$

$$\underbrace{Y_{t+1}}_{n_Y \times 1} = \underbrace{C(\theta)}_{n_Y \times n_X} \underbrace{X_t}_{n_X \times 1} + \underbrace{D(\theta)}_{n_Y \times n_\epsilon} \underbrace{\epsilon_{t+1}}_{n_\epsilon \times 1}. \tag{1b}$$

Assumption 1 *For every $\theta \in \Theta$ and $(t,s) \geqslant 1$, $E(\epsilon_t) = 0$ and $E(\epsilon_t \epsilon_s') = \delta_{t-s}\Sigma_\epsilon(\theta)$, where $\Sigma_\epsilon(\theta)$ is positive definite with Cholesky decomposition $L_\epsilon(\theta)$.*

Assumption 2 *For every $\theta \in \Theta$ and for any $z \in \mathbb{C}$, $\det(zI_{n_X} - A(\theta)) = 0$ implies $|z| < 1$.*

Assumption 1 only requires the innovations ϵ_t to be white noise with time invariant variance $\Sigma_\epsilon(\theta)$, which is weaker than iid. The shocks can be mutually correlated as in Curdia and Reis (2009). Assumption 2 is a stability condition. Under Assumptions 1 and 2, $\{Y_t\}$ is weakly stationary and has a causal VMA(∞) representation:

$$Y_t = \sum_{j=0}^{\infty} h_\epsilon(j,\theta)\epsilon_{t-j} = H_\epsilon(L^{-1};\theta)\epsilon_t, \tag{2}$$

where L is the lag operator. The $n_Y \times n_\epsilon$ matrices $h_\epsilon(j,\theta)$ are the Markov parameters defined by $h_\epsilon(0,\theta) = D(\theta)$, and $h_\epsilon(j,\theta) = C(\theta)A(\theta)^{j-1}B(\theta)$, for all $j \geqslant 1$. For $z \in \mathbb{C}$, the transfer function (the z-transform of the impulse response function) is

$$H_\epsilon(z;\theta) = D(\theta) + C(\theta)[z\mathrm{I}_{n_X} - A(\theta)]^{-1}B(\theta) = \sum_{j=0}^{\infty} h_\epsilon(j,\theta)z^{-j}.$$

Let $\Gamma_Y(j;\theta) = E(Y_t Y_{t-j}') = \Gamma_Y(-j;\theta)'$ be the autocovariance matrix at lag j. Then for all $z \in \mathbb{C}$, the $n_Y \times n_Y$ spectral density matrix is:

$$\begin{aligned}\Omega_Y(z;\theta) &\equiv \Gamma_Y(0;\theta) + \sum_{j=1}^{\infty}\Gamma_Y(j;\theta)z^{-j} + \sum_{j=1}^{\infty}\Gamma_Y(-j;\theta)z^{-j} \\ &= H_\epsilon(z;\theta)\Sigma_\epsilon(\theta)H_\epsilon(z^{-1};\theta)'.\end{aligned}$$

In addition to stability, it is quite frequent in econometric analysis to assume left-invertibility (also known as miniphase). Under this assumption, (2) is the Wold representation and ϵ_t is fundamental for $\{Y_t\}$, meaning that ϵ_t is spanned by $Y^t \equiv \{Y_{t-k}\}_{k=0}^{\infty}$, the current and past history of

[3] Solution algorithms include Anderson and Moore (1985), Uhlig (1999), Klein (2000), King and Watson (2002), and Sims (2002) among others.

Y_t. For square models with $n_\epsilon = n_Y$, left-invertibility holds when $\det H_\epsilon(z;\theta) \neq 0$ in $|z| > 1$.[4] In models that are not square, left invertibility requires that $H_\epsilon(z;\theta)$ is full column rank in $|z| > 1$, (see, e.g., Rozanov, 1967). Left-invertibility is crucial for subsequent analysis which relies on the rank of the spectral density; clearly the rank of $\Omega_Y(z;\theta)$ is a function of the ranks of $\Sigma_\epsilon(\theta)$ and $H_\epsilon(z;\theta)$. The properties of $H_\epsilon(z;\theta)$ are related to those of the (Rosenbrock) system matrix:

$$\mathcal{P}(z;\theta) = \begin{pmatrix} z\mathrm{I}_{n_X} - A(\theta) & B(\theta) \\ -C(\theta) & D(\theta) \end{pmatrix}, \qquad z \in \mathbb{C}. \tag{3}$$

Lemma 1 *Suppose Assumptions 1 and 2 hold. Then rank* $\mathcal{P}(z;\theta) = n_X +$ *rank* $H_\epsilon(z;\theta)$ *for any* $\theta \in \Theta$ *and for every* $z \in \mathbb{C}\backslash\mathbb{A}(\theta)$*, where* $\mathbb{A}(\theta)$ *is the set of eigenvalues of* $A(\theta)$*.*

Lemma 1 will be used subsequently.

The identification problem can now be stated as follows. Suppose we are given a realization of $\{Y_t\}$ generated by (1a) and (1b) at $\theta = \theta_0$ of length T. With T infinitely large, under what conditions would it be possible to uncover the value θ_0 and the model that generated $\{Y_t\}$? Since Assumptions 1 and 2 imply covariance stationarity, the first and second moments of $\{Y_t\}$ are time invariant. The objective is to characterize whether θ_0 is identifiable from the unconditional means and autocovariances. The same analysis can be alternatively motivated by assuming ϵ_t is Gaussian. We start by discussing the identifiability from the autocovariances of $\{Y_t\}$. Mean, long-run and other (nonlinear) restrictions are exploited in Section 5.

Definition 1 θ_0 *and* θ_1 *are observationally equivalent in the second order if* $\Omega_Y(z;\theta_0) = \Omega_Y(z;\theta_1)$ *for all* $z \in \mathbb{C}$*, or equivalently,* $\Gamma_Y(j,\theta_0) = \Gamma_Y(j,\theta_1)$ *at all* $j \geqslant 0$*.*

Definition 2 *The DSGE model is locally identifiable from the second moments of* $\{Y_t\}$ *at a point* $\theta_0 \in \Theta$ *if there exists an open neighborhood of* θ_0 *such that for every* θ_1 *in this neighborhood,* θ_0 *and* θ_1 *are observationally equivalent if and only if* $\theta_1 = \theta_0$*.*

In theory, a sufficient condition for uniqueness of the second moments with respect to θ is that the derivative matrix of $\Gamma_{Y,\infty}(\theta) \equiv \lim_{T\to\infty} \begin{pmatrix} \Gamma_Y(0;\theta) & \Gamma_Y(1;\theta) & \dots & \Gamma_Y(T;\theta) \end{pmatrix}$ with respect to θ has full column rank when evaluated at $\theta = \theta_0$. The autocovariance matrix at each j is

$$\Gamma_Y(j;\theta) = \sum_{k=0}^{\infty} h_\epsilon(k+j;\theta)\Sigma_\epsilon(\theta)h_\epsilon(k;\theta)' = C(\theta)A(\theta)^j\Gamma_X(0;\theta)C(\theta)' + \mathbb{1}_{j=0}D(\theta)\Sigma_\epsilon(\theta)D(\theta)',$$

where $\Gamma_X(0;\theta) = E(X_tX_t')$ solves $\Gamma_X(0;\theta) = A(\theta)\Gamma_X(0;\theta)A(\theta)' + B(\theta)\Sigma_\epsilon(\theta)B(\theta)'$. Clearly, each $\Gamma_Y(j;\theta)$ can only be approximated as a truncated sum of the Markov parameters, or $\Gamma_X(0;\theta)$ has

[4] In the borderline case when $\det H_\epsilon(z;\theta) = 0$ at $|z| = 1$, the transfer function is left-invertible and yet not invertible.

to be solved from a system of nonlinear equations. Furthermore, we can only compute $\Gamma_{Y,T}(\theta)$ for some finite T. The rank of $\Gamma_{Y,T}(\theta)$ can be sensitive to approximation and numerical errors. Observe, however that $\Gamma_Y(j;\theta)$ is defined from the parameters of the canonical (ABCD) model. Rather than establishing identification from the partial derivatives of $\Gamma_{Y,T}(\theta)$ with respect to θ, we use features of the canonical model to characterize observational equivalence through a finite system of equations directly involving the ABCD matrices. This sheds light on the identification problem without evaluating $\Gamma_{Y,T}(\theta)$. We begin with the singular case.

3 Identification: Singular Case $n_\epsilon \leqslant n_Y$

Let $\Lambda^{\mathrm{S}}(\theta)$ be the hyperparameters in the state space solution:

$$\Lambda^{\mathrm{S}}(\theta) \equiv \big((\text{vec } A(\theta))', (\text{vec } B(\theta))', (\text{vec } C(\theta))', (\text{vec } D(\theta))', (\text{vech } \Sigma_\epsilon(\theta))'\big)'.$$

The dimension of $\Lambda^{\mathrm{S}}(\theta)$ is $n_\Lambda^{\mathrm{S}} = n_X^2 + n_X n_\epsilon + n_Y n_X + n_Y n_\epsilon + n_\epsilon(n_\epsilon+1)/2$.

Assumption 3-S *The mapping $\Lambda^s : \theta \mapsto \Lambda^s(\theta)$ is continuously differentiable on Θ.*

Our point of departure is that associated with each $\theta \in \Theta$ is the transfer function $H_\epsilon(z;\Lambda^{\mathrm{S}}(\theta))$ and the covariance matrix $\Sigma_\epsilon(\theta)$. The spectral density $\Omega_Y(z;\theta)$ depends on θ only through $\Lambda^{\mathrm{S}}(\theta)$. By definition, $\Omega_Y(z;\theta_0) = \Omega_Y(z;\theta_1)$ when

$$H_\epsilon(z;\Lambda^{\mathrm{S}}(\theta_0))\Sigma_\epsilon(\theta_0)H_\epsilon(z^{-1};\Lambda^{\mathrm{S}}(\theta_0))' = H_\epsilon(z;\Lambda^{\mathrm{S}}(\theta_1))\Sigma_\epsilon(\theta_1)H_\epsilon(z^{-1};\Lambda^{\mathrm{S}}(\theta_1))'. \tag{4}$$

Equivalent spectral densities can arise because: (i) for given $\Sigma_\epsilon(\theta)$, each $H_\epsilon(z;\Lambda^{\mathrm{S}}(\theta))$ can potentially be obtained from a multitude of quadruples $(A(\theta), B(\theta), C(\theta), D(\theta))$, or (ii) there can be many pairs $(H_\epsilon(z;\Lambda^{\mathrm{S}}(\theta)), \Sigma_\epsilon(\theta))$ that jointly generate the same spectral density. In economic terms, the first problem can arise when two structures induce identical impulse responses to an innovation of a given size, and the second problem can arise when an innovation of arbitrary size can combine with the propagating mechanism to yield the same impulse responses. To make precise their implications on the canonical model, the following assumptions are required.

Assumption 4-S *For every $\theta \in \Theta$, rank $\mathcal{P}(z;\theta) = n_X + n_\epsilon$ in $|z| > 1$.*

Assumption 5-S *For every $\theta \in \Theta$, (i) the matrix $\big(B(\theta)\ A(\theta)B(\theta)\ \ldots\ A^{n_X-1}(\theta)B(\theta)\big)$ has full row rank; and (ii) the matrix $\big(C(\theta)'\ A(\theta)'C(\theta)'\ \ldots\ A^{n_X-1}(\theta)'C(\theta)'\big)'$ has full column rank.*

Assumptions 4-S and 5-S are key to structural identification of singular models. A necessary and sufficient condition for left-invertibility of transfer functions of singular systems is rank $H_\epsilon(z;\theta) = n_\epsilon$ in $|z| > 1$. Lemma 1 implies that $H_\epsilon(z;\theta)$ is left-invertible if and only if

$$n_\epsilon \leqslant n_Y \quad \text{and} \quad \text{rank } \mathcal{P}(z;\theta) = n_X + n_\epsilon, \quad \text{for all } |z| > 1.$$

Assumption 4-S ensures that $H_\epsilon(z;\theta)$ is left-invertible. When combined with Assumption 1, Assumption 4-S also ensures that the rank of the spectral density is n_ϵ almost everywhere (a.e.) in $\mathbb{C}$. Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson (2007) propose an eigenvalue test for left-invertibility when $D(\theta)$ is square. In non-square cases with $n_\epsilon < n_Y$, this can be implemented as a generalized eigenvalue test (see, e.g., Laub and Moore, 1978).

Assumptions 5-S ensures that $(A(\theta), B(\theta))$ is controllable and $(A(\theta), C(\theta))$ is observable and hence that the system is minimal. Controllability means that for any initial state, it is always possible to design an input sequence that puts the system to a desired final state. Observability means that we can always reconstruct the initial state from observing the evolution of the output, given the evolution of the input.[5] Minimality is similar to relative co-primeness in VARMA terminology, and n_X is the McMillan degree. Hannan (1971) imposed co-primeness in his seminal work on identification of dynamic simultaneous systems.[6] A 'minimal' system has the property that the state vector X_t is of the smallest dimension possible. In DSGE models, this is the smallest vector of exogenous and endogenous state variables that are rid of common factors and redundant dynamics, and yet able to fully characterize the properties of the model. Minimality thus simplifies the identification analysis. As a DSGE model is based on microfoundations, n_X is not hard to determine.

The conditions for identification are obtained as follows. First, we use the fact that $H_\epsilon(z;\theta_0) = H_\epsilon(z;\theta_1)$ if and only if there exists a full rank $n_X \times n_X$ matrix T such that

$$(A(\theta_1), B(\theta_1), C(\theta_1), D(\theta_1)) = (TA(\theta_0)T^{-1}, TB(\theta_0), C(\theta_0)T^{-1}, D(\theta_0)). \tag{5}$$

The quadruples $(A(\theta_0), B(\theta_0), C(\theta_0), D(\theta_0))$ and $(A(\theta_1), B(\theta_1), C(\theta_1), D(\theta_1))$ are said to be related by a similarity transformation. That these transformations are sufficient for transfer functions to be equivalent is obvious. That they are also necessary follows from an algebraically involved but well-known result in control theory (see, e.g., Theorem 3.10 in Antsaklis and Michel, 1997).

Second, we take into account that $H_\epsilon(z;\theta)$ can interact with $\Sigma_\epsilon(\theta)$ to give equivalent spectral densities. To characterize such interactions, let $\overline{H}_\epsilon(z;\theta) \equiv H_\epsilon(z;\theta)L_\epsilon(\theta)$. Since $\Omega_Y(z;\theta) =$

[5] See, e.g. Anderson and Moore (1979). The matrices $A(\theta)$ and $C(\theta)$ in a minimal system need not be full rank, meaning that some components of X_t can be white noise, and that identities are allowed in Y_t.

[6] Further discussion of the role of minimality in this context can be found in **?**.

$\overline{H}_\epsilon(z;\theta)\overline{H}_\epsilon(z^{-1};\theta)'$, it is not hard to see that $\widetilde{H}_\epsilon(z;\theta) = \overline{H}_\epsilon(z;\theta)V(z)$ will yield the same spectral density for any $V(z)$ satisfying $V(z)V(z^{-1})' = I_{n_\epsilon}$, even if $\Sigma_\epsilon(\theta)$ is an identity matrix. Thus, equivalent pairs $(H_\epsilon(z;\theta), \Sigma_\epsilon(\theta))$ can be related only through a polynomial matrix $V(z)$ of unknown degree. This is unlike in static models in which $V(z) = V$ is a constant matrix. Hence, the number of static models in the equivalent class is substantially smaller than in the dynamic case.

Left-invertibility is crucial for being able to narrow down the equivalent pairs $(H_\epsilon(z;\theta), \Sigma_\epsilon(\theta))$. Suppose for the moment that the system is square. It is well known that the spectral density of a full rank covariance stationary process can be factorized as $\Omega_Y(z;\theta) = W_\epsilon(z;\theta)W_\epsilon(z^{-1};\theta)'$, where $W_\epsilon(z;\theta)$ is known as the spectral factor. Much less known is that if $W_\epsilon(z;\theta)$ and $\widetilde{W}_\epsilon(z;\theta)$ are both spectral factors that are also left-invertible, then necessarily $\widetilde{W}_\epsilon(z;\theta) = W_\epsilon(z;\theta)V$ with $VV' = \mathrm{I}_{n_\epsilon}$ (see, e.g., Youla, 1961; Anderson, 1969; Kailath, Sayed, and Hassibi, 2000). Note that V is a constant matrix and is no longer a polynomial matrix in z. This means that no dynamic transformations of left-invertible factors are allowed. Importantly, this result holds even in singular models. Our spectral factor is $W_\epsilon(z;\theta) = H_\epsilon(z;\theta)L_\epsilon(\theta)$. It is left-invertible if and only if $H_\epsilon(z;\theta)$ is left-invertible which holds by Assumption 4-S. In such a case, two equivalent pairs $(H_\epsilon(z;\theta), \Sigma_\epsilon(\theta))$ must be related by a full rank matrix $U = L_\epsilon(\theta_0)VL_\epsilon(\theta_1)^{-1}$ such that

$$H_\epsilon(z;\theta_1) = H_\epsilon(z;\theta_0)U, \quad \text{and} \quad U\Sigma_\epsilon(\theta_1)U' = \Sigma_\epsilon(\theta_0), \quad \text{for every } z \in \mathbb{C}. \tag{6}$$

Combining the transformations in (5) and (6) leads to our first main result.

Proposition 1-S (Observational Equivalence $n_\epsilon \leqslant n_Y$) *Suppose Assumptions 1, 2, 4-S, and 5-S hold. Then θ_0 and θ_1 are observationally equivalent if and only if there exists a full rank $n_X \times n_X$ matrix T and a full rank $n_\epsilon \times n_\epsilon$ matrix U such that:*

$$\begin{aligned} A(\theta_1) = TA(\theta_0)T^{-1}, \quad B(\theta_1) = TB(\theta_0)U, \quad C(\theta_1) = C(\theta_0)T^{-1}, \quad D(\theta_1) = D(\theta_0)U, \\ \Sigma_\epsilon(\theta_1) = U^{-1}\Sigma_\epsilon(\theta_0)U^{-1\prime}. \end{aligned} \tag{7}$$

Proposition 1-S, proved in the Appendix, says that in singular systems with $n_\epsilon \leqslant n_Y$, there can exist no other transformation of the hyperparameters $\Lambda^{\mathrm{S}}(\theta)$ other than those defined in (7) that can give rise to equivalent spectral densities. In other words, these transformations are necessary and sufficient for observational equivalence. The crux of the proposition is to use minimality and left-invertibility to narrow down the set of observationally equivalent hyperparameters. The result also holds in the important special case of a square system.

An immediate implication of Proposition 1-S is that without restrictions $\Lambda^{\mathrm{S}}(\theta)$ may not be identifiable. In other words, the ABCD representation is not a reduced form in the sense of classical

simultaneous equations analysis. Iskrev (2007) suggests this possibility in finite samples. Here, we see that the problem can arise even if the population autocovariances are available.

Proposition 1-S can now be used to derive formal identification conditions. Define the continuously differentiable mapping $\delta^{\mathrm{S}} : \Theta \times \mathbb{R}^{n_X^2} \times \mathbb{R}^{n_\epsilon^2} \to \mathbb{R}^{n_\Lambda^{\mathrm{S}}}$ as

$$\delta^{\mathrm{S}}(\theta, T, U) \equiv \begin{pmatrix} \text{vec}\,(TA(\theta)T^{-1}) \\ \text{vec}\,(TB(\theta)U) \\ \text{vec}\,(C(\theta)T^{-1}) \\ \text{vec}\,(D(\theta)U) \\ \text{vech}\,(U^{-1}\Sigma_\epsilon(\theta)U^{-1\prime}) \end{pmatrix}. \tag{8}$$

The mapping defines n_Λ^{S} equations in $n_\theta + n_X^2 + n_\epsilon^2$ unknowns.

Lemma 2-S *Under the assumptions of Proposition 1-S, θ is locally identifiable from the second moments of $\{Y_t\}$ at a point $\theta_0 \in \Theta$ if and only if the system of equations $\delta^S(\theta_0, I_{n_X}, I_{n_\epsilon}) = \delta^S(\theta_1, T, U)$ has a locally unique solution $(\theta_1, T, U) = (\theta_0, I_{n_X}, I_{n_\epsilon})$.*

Lemma 2-S, proved in the Appendix, says that a singular DSGE model is locally identifiable at $\theta_0 \in \Theta$ if and only if $\delta^{\mathrm{S}}(\theta, T, U)$ is locally injective at $(\theta_0, I_{n_X}, I_{n_\epsilon})$. Necessity and sufficiency both rely on the minimality and left-invertibility. However, as will be explained below, the necessity argument can be modified when these assumptions are relaxed.

As δ^{S} is continuously differentiable, a sufficient condition for δ^{S} to be locally injective is that the matrix of partial derivatives of $\delta^{\mathrm{S}}(\theta, T, U)$ has full column rank at $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})$. The matrix of partial derivatives of $\delta^{\mathrm{S}}(\theta, T, U)$ evaluated at $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})$ is given by:[7]

$$\begin{aligned}
\Delta^{\mathrm{S}}(\theta_0) &\equiv \left(\tfrac{\partial \delta^{\mathrm{S}}(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})}{\partial \theta} \quad \tfrac{\partial \delta^{\mathrm{S}}(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})}{\partial \text{vec}\, T} \quad \tfrac{\partial \delta^{\mathrm{S}}(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})}{\partial \text{vec}\, U} \right)\Big|_{\theta=\theta_0} \\
&= \begin{pmatrix} \frac{\partial \text{vec}\, A(\theta)}{\partial \theta} & A(\theta)' \otimes \mathrm{I}_{n_X} - \mathrm{I}_{n_X} \otimes A(\theta) & 0_{n_X^2 \times n_\epsilon^2} \\ \frac{\partial \text{vec}\, B(\theta)}{\partial \theta} & B(\theta)' \otimes \mathrm{I}_{n_X} & \mathrm{I}_{n_\epsilon} \otimes B(\theta) \\ \frac{\partial \text{vec}\, C(\theta)}{\partial \theta} & -\mathrm{I}_{n_X} \otimes C(\theta) & 0_{n_Y n_X \times n_\epsilon^2} \\ \frac{\partial \text{vec}\, D(\theta)}{\partial \theta} & 0_{n_Y n_X \times n_X^2} & \mathrm{I}_{n_\epsilon} \otimes D(\theta) \\ \frac{\partial \text{vech}\, \Sigma_\epsilon(\theta)}{\partial \theta} & 0_{\frac{n_\epsilon(n_\epsilon+1)}{2} \times n_X^2} & -2\mathcal{E}_{n_\epsilon}\left[\Sigma_\epsilon(\theta) \otimes \mathrm{I}_{n_\epsilon}\right] \end{pmatrix}_{\theta=\theta_0} \\
&\equiv \begin{pmatrix} \Delta_\Lambda^{\mathrm{S}}(\theta_0) & \Delta_T^{\mathrm{S}}(\theta_0) & \Delta_U^{\mathrm{S}}(\theta_0) \end{pmatrix}.
\end{aligned}$$

The $n_\Lambda^{\mathrm{S}} \times n_\theta$ block defined by $\Delta_\Lambda^{\mathrm{S}}(\theta_0)$ describes the local properties of the mapping from θ to $\Lambda^{\mathrm{S}}(\theta)$. When the rank of $\Delta_\Lambda^{\mathrm{S}}(\theta_0)$ equals n_θ, the mapping is locally invertible at θ_0. Since $\Lambda^{\mathrm{S}}(\theta_0)$

[7] For an arbitrary matrix X, we let vec(X) be formed by stacking the columns of X into a single column vector. Also for any symmetric $n \times n$ matrix A, $\mathcal{E}_n$ is the left inverse of the $n \times \frac{n(n+1)}{2}$ duplication matrix $\mathcal{G}_n$, where vec $(A) = \mathcal{G}_n$ vech (A).

may not be identifiable, the rank of $\Delta^{\mathrm{S}}_{\Lambda}(\theta_0)$ alone is necessary but not sufficient for identification. The middle $n^{\mathrm{S}}_{\Lambda} \times n^2_X$ matrix $\Delta^{\mathrm{S}}_T(\theta_0)$ corresponds to the partial derivatives with respect to T evaluated at $(T, U) = (\mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})$. When rank $\Delta^{\mathrm{S}}_T(\theta_0) = n^2_X$, then the only (local) similarity transformation is provided by the identity matrix. If the transfer function were observable, one could (locally) uniquely determine $(A(\theta_0), B(\theta_0), C(\theta_0), D(\theta_0))$ that generated it. The final $n^{\mathrm{S}}_{\Lambda} \times n^2_\epsilon$ matrix $\Delta^{\mathrm{S}}_U(\theta_0)$ corresponds to the partial derivatives with respect to U, evaluated at $(T, U) = (\mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})$. When rank $\Delta^{\mathrm{S}}_U(\theta_0) = n^2_\epsilon$ then the spectral factorization (locally) uniquely determines the pair $(H_\epsilon(z;\theta), \Sigma_\epsilon(\theta))$.

Proposition 2-S (Rank and Order Conditions $n_\epsilon \leqslant n_Y$) *Suppose Assumptions 1, 2, 3-S, 4-S and 5-S hold. If the rank of $\Delta^S(\theta)$ remains constant in a neighborhood of θ_0, then a necessary and sufficient rank condition for θ to be locally identified from the second moments of $\{Y_t\}$ at a point θ_0 in Θ is:*

$$\textit{rank}\, \Delta^S(\theta_0) = \textit{rank}\, \begin{pmatrix} \Delta^S_\Lambda(\theta_0) & \Delta^S_T(\theta_0) & \Delta^S_U(\theta_0) \end{pmatrix} = n_\theta + n^2_X + n^2_\epsilon.$$

A necessary order condition is: $n_\theta + n^2_X + n^2_\epsilon \leqslant n^S_\Lambda$, where $n^S_\Lambda = (n_X + n_Y)(n_X + n_\epsilon) + n_\epsilon(n_\epsilon + 1)/2$.

Proposition 2-S is new to both the econometrics and control theory literature. It extends the results of Hannan (1971), Deistler (1976), and Glover and Willems (1974) to stochastically singular systems in which ϵ_t is unobserved with unknown covariance $\Sigma_\epsilon(\theta)$. The conditions allow researchers to establish the identifiability of singular and square DSGE models from the second moments prior to estimation, independent of the estimator used. These conditions have three distinctive features.

First, θ can be identified from the second moments of the data, even though the hyperparameters $\Lambda^{\mathrm{S}}(\theta)$ may not be identifiable. Second, the conditions depend on n_Y, n_X, and n_ϵ representing the structure of the economic model; the sample size T is not involved. However, the term $n_Y - n_\epsilon > 0$ reflects stochastic singularity and is specific to DSGE models. Third, numerical evaluations of the population autocovariances or of the spectral density are not necessary because we study their determinants $\Lambda^{\mathrm{S}}(\theta)$ directly.

The order condition requires the number of equations defined by δ^{S} to be at least as large as the number of unknowns in those equations. It can be rewritten as

$$n_\theta \leqslant n_Y n_X + n_\epsilon(n_X + n_Y - n_\epsilon) + \frac{n_\epsilon(n_\epsilon + 1)}{2}.$$

Notably, square models have $n_Y = n_\epsilon$ and the order condition is stronger than when the system is singular. The parameters n_Y, n_ϵ and n_X play the role of the number of endogenous and predetermined variables in classical simultaneous equations analysis. Both rank and order conditions have a classical flavor even though we work with assumptions that would not be valid in a classical setup.

As in classical analysis, the constant rank requirement in Proposition 2-S ensures that $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})$ is a regular point. Our rank condition is still sufficient for identification, even if θ_0 fails to be a regular point.[8]

The sub-matrices $\Delta^{\mathrm{S}}_{\Lambda T}(\theta_0) = \begin{pmatrix}\Delta^{\mathrm{S}}_{\Lambda}(\theta_0) & \Delta^{\mathrm{S}}_{T}(\theta_0)\end{pmatrix}$ and $\Delta^{\mathrm{S}}_{\Lambda U}(\theta_0) = \begin{pmatrix}\Delta_{\Lambda}(\theta_0) & \Delta_{U}(\theta_0)\end{pmatrix}$ reveal delicate types of non-identification that might arise even when $\Lambda^{\mathrm{S}}(\theta)$ is identified, and hence $\Delta^{\mathrm{S}}_{\Lambda}(\theta_0)$ is full rank. They shed light on identification failure due to the features of the impulse and propagating mechanism of the model. Such information about the properties of the model is useful even if estimation is not of interest. As will be shown in Section 6, the null space of $\Delta^{\mathrm{S}}(\theta_0)$ can be used to isolate the parameters that are not identifiable.

Minimality and left-invertibility are maintained assumptions of Proposition 2-S. However, rank conditions on the sub-matrices of $\Delta^{\mathrm{S}}(\theta_0)$ remain necessary for identification even if one of these assumptions fail. In particular, (i) rank $\Delta^{\mathrm{S}}_{\Lambda U}(\theta_0) = n_\theta + n_\epsilon^2$ is necessary for θ_0 to be identifiable when the system fails to be minimal but the transfer function is left-invertible, while (ii) rank $\Delta^{\mathrm{S}}_{\Lambda T}(\theta_0) = n_\theta + n_X^2$ is necessary for identification of minimal systems with transfer functions that are not left-invertible. To see why, consider case (i). Local injectivity of $\delta^{\mathrm{S}}(\theta, \mathrm{I}_{n_X}, U)$ at $(\theta, U) = (\theta_0, \mathrm{I}_{n_\epsilon})$ is still necessary for local identification of θ_0 because the transformation $A(\theta_1) = A(\theta_0)$, $B(\theta_1) = B(\theta_0)U$, $C(\theta_1) = C(\theta_0)$, $D(\theta_1) = D(\theta_0)$, and $\Sigma_\epsilon(\theta_1) = U^{-1}\Sigma_\epsilon(\theta_0)U^{-1\prime}$ always leads to observational equivalence even if the system is not minimal. However, if both minimality and left-invertibility fail to hold, then only rank $\Delta^{\mathrm{S}}_{\Lambda}(\theta_0) = n_\theta$ is necessary for identification. While none of these conditions are sufficient, they are still useful diagnostic tools.

4 Structural Identification: Non-Singular Case $n_\epsilon \geqslant n_Y$

A non-singular system occurs when there are at least as many shocks as variables. When $n_\epsilon > n_Y$ a different framework is necessary because (2) is no longer the Wold representation for $\{Y_t\}$ and ϵ_t is no longer fundamental. In consequence, $H_\epsilon(z;\theta)$ is no longer left-invertible. We work with the innovations representation:[9]

$$\widehat{X}_{t+1|t+1} = A(\theta)\widehat{X}_{t|t} + K(\theta)a_{t+1} \tag{9a}$$

$$Y_{t+1} = C(\theta)\widehat{X}_{t|t} + a_{t+1}, \tag{9b}$$

where $\widehat{X}_{t|t}$ is the optimal linear predictor of X_t based on the history Y^t, $a_{t+1} = Y_{t+1} - C(\theta)\widehat{X}_{t|t}$ is the one-step ahead forecast error of Y_{t+1}, and $K(\theta)$ is the steady state Kalman gain. The state

[8] The regular point requirement is satisfied almost everywhere in the parameter space except when the determinant of $\Delta^{\mathrm{S}}(\theta)'\Delta^{\mathrm{S}}(\theta)$ is an analytic function (see, e.g., Fisher, 1966).

[9] Details of the innovations representation can be found in Anderson and Moore (1979), Hansen and Sargent (2005).

vector in the innovations model is the conditional forecast of X_t given Y^t, and the $n_\epsilon > n_Y$ shocks are now consolidated into a vector of n_Y white noise forecast errors a_t whose variance is $\Sigma_a(\theta)$. Let $L_a(\theta)$ be the Cholesky decomposition of $\Sigma_a(\theta)$. Collect the system parameters of the innovations model into

$$\Lambda^{\text{NS}}(\theta) \equiv \big((\text{vec } A(\theta))', (\text{vec } K(\theta))', (\text{vec } C(\theta))', (\text{vech } \Sigma_a(\theta))'\big)',$$

in which the number of components equals $n_\Lambda^{\text{NS}} = n_X^2 + 2n_X n_Y + n_Y(n_Y+1)/2$.

Assumption 3-NS *The mapping $\Lambda^{NS} : \theta \mapsto \Lambda^{NS}(\theta)$ is continuously differentiable on Θ.*

Assumption 4-NS *For every $\theta \in \Theta$, $D(\theta)\Sigma_\epsilon(\theta)D(\theta)'$ is non-singular.*

Assumption 5-NS *For every $\theta \in \Theta$, (i) the matrix $\big(K(\theta)\ A(\theta)K(\theta)\ \ldots\ A^{n_X-1}(\theta)K(\theta)\big)$ has full row rank; and (ii) the matrix $\big(C(\theta)'\ A(\theta)'C(\theta)'\ \ldots\ A^{n_X-1}(\theta)'C(\theta)'\big)'$ has full column rank.*

The validity of the innovations representation hinges on the existence of a positive semi-definite solution to the so-called discrete algebraic Ricatti equation (DARE).[10] Such a solution exists if $D(\theta)$ is full row rank (Assumption 4-NS) so that each of the innovations, which can be primitive to the model or is a measurement error, affects at least one series. It rules out observables that are defined by identities and are not affected by any shock or measurement error. Under Assumption 4-NS, $K(\theta)$ and $\Sigma_a(\theta)$ are well defined functions of the hyperparameter $\Lambda^{\text{S}}(\theta)$ and their expressions are given in Equations (19) and (18) of Appendix. Moreover, $\Sigma_a(\theta)$ is nonsingular for every $\theta \in \Theta$. The transfer function according to the innovations model is given by

$$H_a(z;\theta) = \text{I}_{n_Y} + C(\theta)[z\text{I}_{n_X} - A(\theta)]^{-1}K(\theta).$$

Under the assumptions of the innovations representation, $H_a(z;\theta)$ is square and invertible in $|z| > 1$ by construction. This allows us to apply the left-invertibility arguments. This is unlike in singular models where left-invertibility holds by virtue of Assumption 4-S. Assumption 5-NS ensures that the system (9a)-(9b) is minimal. Using $H_a(z;\theta)$ and $\Sigma_a(\theta)$, the spectral density of $\{Y_t\}$ is

$$\Omega_Y(z;\theta) \quad = \quad H_a(z;\theta)\Sigma_a(\theta)H_a(z^{-1};\theta)'.$$

The following result characterizes observational equivalence in non-singular models.

Proposition 1-NS (Observational Equivalence $n_\epsilon \geqslant n_Y$) *Suppose Assumptions 1, 2, 4-NS, and 5-NS hold. Then θ_0 and θ_1 are observationally equivalent if and only if there exists a full rank $n_X \times n_X$ matrix T such that:*

$$A(\theta_1) = TA(\theta_0)T^{-1}, \quad K(\theta_1) = TK(\theta_0), \quad C(\theta_1) = C(\theta_0)T^{-1}, \quad \Sigma_a(\theta_1) = \Sigma_a(\theta_0). \qquad (10)$$

[10] See, e.g., Appendix E in Kailath, Sayed, and Hassibi (2000).

Proposition 1-NS takes as a starting point that transfer functions related by similarity transforms must also have equivalent Kalman filters. Furthermore, $H_a(z;\theta)$ is nonsingular in $|z|>1$ so that the spectral factor $W_a(\theta) = H_a(z;\theta)L_a(\theta)$ is left-invertible. Combining the minimality and left-invertibility restrictions yield (10). The restrictions no longer involve U because the innovation representation imposes the normalization $h_a(0;\theta) = \mathrm{I}_{n_X}$. Furthermore, the restrictions are now expressed in terms of $K(\theta)$ which needs to be solved numerically. This matrix is a complex function of $\Lambda^{\mathrm{S}}(\theta)$ because the forecast errors a_t are derived from Kalman filtering. In contrast, no filtering is necessary to study identification of singular models.

For any $\theta \in \Theta$ and any full rank $n_X \times n_X$ matrix T, let $\delta^{\mathrm{NS}} : \Theta \times \mathbb{R}^{n_X^2} \to \mathbb{R}^{n_\Lambda^{\mathrm{NS}}}$ be defined by

$$\delta^{\mathrm{NS}}(\theta, T) \equiv \begin{pmatrix} \mathrm{vec}\,(TA(\theta)T^{-1}) \\ \mathrm{vec}\,(TK(\theta)) \\ \mathrm{vec}\,(C(\theta)T^{-1}) \\ \mathrm{vech}\,(\Sigma_a(\theta)) \end{pmatrix}. \tag{11}$$

The mapping δ^{NS} defines n_Λ^{NS} equations in $n_\theta + n_X^2$ unknowns. Under the assumptions of Proposition 1-NS, θ is locally identifiable from the second moments of $\{Y_t\}$ at a point $\theta_0 \in \Theta$ if and only if the system of equations $\delta^{\mathrm{NS}}(\theta_0, I_{n_X}) = \delta^{\mathrm{NS}}(\theta_1, T)$ has a locally unique solution $(\theta_1, T) = (\theta_0, \mathrm{I}_{n_X})$.[11] The matrix of partial derivatives of $\delta^{\mathrm{NS}}(\theta, T)$ evaluated at $(\theta_0, \mathrm{I}_{n_X})$ is:

$$\begin{aligned} \Delta^{\mathrm{NS}}(\theta_0) &\equiv \left(\tfrac{\partial \delta(\theta, \mathrm{I}_{n_X})}{\partial \theta} \quad \tfrac{\partial \delta(\theta, \mathrm{I}_{n_X})}{\partial \mathrm{vec}\, T} \right)\Big|_{\theta=\theta_0} \\ &= \begin{pmatrix} \frac{\partial \mathrm{vec}\, A(\theta)}{\partial \theta} & A(\theta)' \otimes \mathrm{I}_{n_X} - \mathrm{I}_{n_X} \otimes A(\theta) \\ \frac{\partial \mathrm{vec}\, K(\theta)}{\partial \theta} & K(\theta)' \otimes \mathrm{I}_{n_X} \\ \frac{\partial \mathrm{vec}\, C(\theta)}{\partial \theta} & -\mathrm{I}_{n_X} \otimes C(\theta) \\ \frac{\partial \mathrm{vech}\, \Sigma_a(\theta)}{\partial \theta} & 0_{\frac{n_Y(n_Y+1)}{2} \times n_X^2} \end{pmatrix}_{\theta=\theta_0} \\ &\equiv \begin{pmatrix} \Delta_\Lambda^{\mathrm{NS}}(\theta_0) & \Delta_T^{\mathrm{NS}}(\theta_0) \end{pmatrix}. \end{aligned}$$

Proposition 2-NS (Rank and Order Conditions $n_\epsilon \geqslant n_Y$) *Suppose Assumptions 1, 2, 3-NS, 4-NS, and 5-NS hold. If the rank of $\Delta^{NS}(\theta)$ remains constant in a neighborhood of θ_0, then a necessary and sufficient rank condition for θ to be locally identified from the second moments of $\{Y_t\}$ at a point θ_0 in Θ is:*

$$rank\, \Delta^{NS}(\theta_0) = rank\, \begin{pmatrix} \Delta_\Lambda^{NS}(\theta_0) & \Delta_T^{NS}(\theta_0) \end{pmatrix} = n_\theta + n_X^2.$$

A necessary order condition is $n_\theta + n_X^2 \leqslant n_\Lambda^{NS}$, where $n_\Lambda^{NS} = n_X^2 + 2n_X n_Y + n_Y(n_Y+1)/2$.

While the rank condition appears weaker than in the singular case, Assumption 4-NS must also hold. Using the same arguments as in the singular case, rank $\Delta_\Lambda^{\mathrm{NS}}(\theta_0) = n_\theta$ remains necessary for identification when the minimality Assumption 5-NS fails.

[11] The proof of this statement can be obtained by using the same arguments as in the proof of Lemma 2-S.

The order condition can be rewritten as

$$n_\theta \leqslant 2n_X n_Y + \frac{n_Y(n_Y+1)}{2}.$$

Holding n_X and n_Y fixed, this is apparently stronger than in the singular case. Filtering entails information loss, and less may be identified as a result.

When the system is square ($n_Y = n_\epsilon$), the rank and order conditions derived in Proposition 2-NS coincide with those in Proposition 2-S. The reason is that the state covariance matrix of the innovations model becomes degenerate when $n_Y = n_\epsilon$. This allows $K(\theta)$ to be expressed in terms of $B(\theta)$ and $D(\theta)$.[12] Simplifying shows that the transformation (10) holds if and only if there exists a full rank $n_Y \times n_Y$ matrix U such that (7) holds. Thus, in the square case the condition rank $\Delta^{\text{NS}}(\theta_0) = n_\theta + n_X^2$ of Proposition 2-NS is equivalent to the condition rank $\Delta^{\text{S}}(\theta_0) = n_\theta + n_X^2 + n_Y^2$ of Proposition 2-S, as expected.

5 Conditional and Partial Identification

Restrictions on some components of θ are often imposed for a number of reasons. They may enable identification of the remaining parameters when the rank conditions in Proposition 2-S or 2-NS fail. A researcher may have detailed information about a parameter so that consideration of other values is unnecessary. The unconditional moments implied by second order approximations to the model may contain information about θ. More generally, consider a set of n_R a priori restrictions that when evaluated at θ_0 satisfy:

$$\varphi(\theta_0) = 0.$$

The formulation provides a convenient basis for incorporating information about the unconditional mean of Y_t since in the steady state, $EY_t^a = EY_{ss}^a(\theta)$ can be written as $\varphi(\theta_0) = 0$. Certain long run restrictions can be imposed on $H_\epsilon(1;\theta)$.

In the singular case, let $\overline{\delta}^{\text{S}}(\theta,T,U) \equiv \begin{pmatrix} \varphi(\theta) \\ \delta^{\text{S}}(\theta,T,U) \end{pmatrix}$ be the augmented vector of restrictions where δ^{S} is given in (8). Define its derivative matrix by

$$\overline{\Delta}^{\text{S}}(\theta_0) \equiv \left(\frac{\partial \overline{\delta}^{\text{S}}(\theta, \text{I}_{n_X}, \text{I}_{n_\epsilon})}{\partial \theta} \quad \frac{\partial \overline{\delta}(\theta, \text{I}_{n_X}, \text{I}_{n_\epsilon})}{\partial \text{vec}\, T} \quad \frac{\partial \overline{\delta}^{\text{S}}(\theta, \text{I}_{n_X}, \text{I}_{n_\epsilon})}{\partial \text{vec}\, U} \right)\Big|_{\theta=\theta_0} = \begin{pmatrix} \frac{\partial \varphi(\theta_0)}{\partial \theta} & 0_{n_R \times n_X^2} & 0_{n_R \times n_\epsilon^2} \\ \Delta_\Lambda^{\text{S}}(\theta_0) & \Delta_T^{\text{S}}(\theta_0) & \Delta_U^{\text{S}}(\theta_0) \end{pmatrix}.$$

In the non-singular case, let $\overline{\delta}^{\text{NS}}(\theta,T) \equiv \begin{pmatrix} \varphi(\theta) \\ \delta^{\text{NS}}(\theta,T) \end{pmatrix}$ with δ^{NS} as given in (11) and

$$\overline{\Delta}^{\text{NS}}(\theta_0) \equiv \left(\frac{\partial \overline{\delta}^{\text{NS}}(\theta, \text{I}_{n_X})}{\partial \theta} \quad \frac{\partial \overline{\delta}^{\text{NS}}(\theta, \text{I}_{n_X})}{\partial \text{vec}\, T} \right)\Big|_{\theta=\theta_0} = \begin{pmatrix} \frac{\partial \varphi(\theta_0)}{\partial \theta} & 0_{n_R \times n_X^2} \\ \Delta_\Lambda^{\text{NS}}(\theta_0) & \Delta_T^{\text{NS}}(\theta_0) \end{pmatrix}.$$

[12] Specifically, $K(\theta) = (B(\theta)\Sigma_\epsilon(\theta)D(\theta)'(D(\theta)\Sigma_\epsilon(\theta)D(\theta)')^{-1}$.

Local injectivity of $\overline{\delta}^{\mathrm{S}}$ at $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})$ or that of $\overline{\delta}^{\mathrm{NS}}$ at $(\theta_0, \mathrm{I}_{n_X})$ is necessary and sufficient for θ_0 to be locally identifiable from the second moments of $\{Y_t\}$ under the restrictions $\varphi(\theta_0) = 0$. We refer to such an analysis as *conditional identification.*

Proposition 3 (Conditional Identification) *Suppose the conditions of Proposition 2-S or Proposition 2-NS hold and let* $\overline{\Delta}(\theta)$ *denote either* $\overline{\Delta}^{S}(\theta)$ *or* $\overline{\Delta}^{NS}(\theta)$*. Assume that the* n_R *restrictions defined by* $\varphi(\theta)$ *are continuously differentiable on* Θ *and that the rank of* $\overline{\Delta}(\theta)$ *remains constant in a neighborhood of* θ_0*. A necessary and sufficient rank condition for* θ *to be locally conditionally identified at point* $\theta = \theta_0$ *is:*

$$rank\ \overline{\Delta}(\theta_0) = n_\theta + n_X^2 + \mathbb{1}_S \cdot n_\epsilon^2,$$

where $\mathbb{1}_S = 1$ *if the model is singular and zero otherwise. When* $\varphi(\theta)$ *involves the mean, the above restrictions are necessary and sufficient for local identification of* θ_0 *from the first and second moments of* $\{Y_t\}$*.*

Our a priori restrictions are incorporated in the form of additional equations for solving the *same* number of unknowns. Thus, the required rank is always the same; it is the number of rows of the $\overline{\Delta}^{\mathrm{S}}(\theta)$ or $\overline{\Delta}^{\mathrm{NS}}(\theta)$ matrix that increases with the number of restrictions. In singular models, the order condition requires $n_\theta + n_X^2 + n_\epsilon^2 \leqslant n_\Lambda^{\mathrm{S}} + n_R$; in the non-singular case, we need $n_\theta + n_X^2 \leqslant n_\Lambda^{\mathrm{NS}} + n_R$. Proposition 3 provides formal ways to check how many restrictions are mathematically necessary for identification, which is very useful in empirical work. For instance, univariate AR(i) shocks are often specified, but this entails many restrictions on the contemporaneous and past correlations amongst shocks. Proposition 3 provides a means to check their identification implications.

Situations may arise when only a subset of θ is of interest. As well, some components of θ may still be identifiable even when Proposition 2-S or 2-NS fails. To analyze such partial identification situations, partition the n_θ vector θ into two components θ_i and θ_{-i} of respective sizes $n_{\theta,i}$ and $n_{\theta,-i}$ (with $n_{\theta,i} + n_{\theta,-i} = n_\theta$). Without loss of generality, we order the components so that $\theta = ({\theta_{-i}}', {\theta_i}')'$.

Proposition 4 (Partial Identification) *Suppose the conditions of Proposition 2-S or Proposition 2-NS hold. Assume that the ranks of* $\Delta(\theta)$ *and* $\frac{\partial \delta(\theta)}{\partial \theta_{-i}}$ *remain constant in a neighborhood of* θ_0*. A necessary and sufficient rank condition for* θ_i *to be locally partially identified at point* $\theta = \theta_0$ *is:*

$$rank\ \Delta(\theta_0) \quad = \quad rank\ \left(\frac{\partial \delta(\theta_0)}{\partial \theta_{-i}}\right) + \left(n_{\theta,i} + n_X^2 + \mathbb{1}_S \cdot n_\epsilon^2\right),$$

where $\left(\Delta(\theta), \frac{\partial \delta(\theta)}{\partial \theta_{-i}}\right) = \left(\Delta^{S}(\theta), \frac{\partial \delta^{S}(\theta, I_{n_X}, I_{n_\epsilon})}{\partial \theta_{-i}}\right)$ *with* $\mathbb{1}_S = 1$ *if the model is singular, and* $\left(\Delta(\theta), \frac{\partial \delta(\theta)}{\partial \theta_{-i}}\right) = \left(\Delta^{NS}(\theta), \frac{\partial \delta^{NS}(\theta, I_{n_X})}{\partial \theta_{-i}}\right)$ *with* $\mathbb{1}_S = 0$ *in the non-singular case.*

The results generalize Propositions 2-S and 2-NS to allow θ_{-i} to be non-empty. It is important to note that even though one might be interested in a subset of parameters, its identifiability will, in general, depend on the parameters that are not of interest.

6 Relation to the Information Matrix

All identification methods must exploit the canonical solution of the DSGE model in one way or another. Iskrev (2010) uses it to numerically evaluate the T analytical autocovariances, while Qu and Tkachenko (2010) use it to evaluate the discretized spectral density. Our approach is unique in that we do not evaluate any second moments and yet obtain necessary and sufficient conditions for identification. This section provides two different perspectives on our results. The first studies the null space of $\Delta(\theta)$ and relates the proposed rank conditions to those on the Markov parameters of $\{Y_t\}$. The second relates our rank condition to the information matrix. Since the information matrix is defined only when $n_Y \leqslant n_\epsilon$ the focus of this section is on the non-singular case only.[13] It is understood that $\Lambda(\theta)$ and $\Delta(\theta)$ refer to $\Lambda^{\text{NS}}(\theta)$ and $\Delta^{\text{NS}}(\theta)$, respectively.

Let $\Delta_{\overline{h}_J}(\theta) \equiv \frac{\partial \text{vec}\, \overline{h}_J(\theta)}{\partial \theta} = \frac{\partial \text{vec}\, \overline{h}_J(\theta)}{\partial \Lambda} \cdot \frac{\partial \Lambda}{\partial \theta}$, where for any $J \geqslant 0$,

$$\overline{h}_J(\theta) \quad \equiv \quad \begin{pmatrix} h_a(0;\theta)L_a(\theta) & h_a(1;\theta)L_a(\theta) & \dots & h_a(J;\theta)L_a(\theta) \end{pmatrix}.$$

Lemma 3 *Suppose the conditions of Proposition 2-NS hold. Then for every $\theta \in \Theta$:*

(i) for every $J \geqslant 0$, $\frac{\partial vec\, \overline{h}_J(\theta)}{\partial \Lambda} \cdot \Delta_T(\theta) = 0_{(J+1)n_Y^2 \times n_X^2}$;

(ii) if $J \geqslant 2n_X - 2$, then rank $\overline{h}_J(\theta) =$ rank $\overline{h}_{2n_X-2}(\theta)$;

(iii) rank $\Delta(\theta_0) = n_\theta + n_X^2$ if and only if rank $\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = n_\theta$.

Result (i) says that the columns of $\Delta_T(\theta)$ belong to the null space of $\frac{\partial \text{vec}\, \overline{h}_J(\theta)}{\partial \Lambda}$. This is intuitive because $\Delta_T(\theta)$ is the Jacobian matrix of the transformation of $\Lambda(\theta)$ that leaves $\overline{h}_J(\theta)$ unchanged.[14] A direct consequence of result (i) is

$$\frac{\partial \text{vec}\, \overline{h}_J(\theta_0)}{\partial \Lambda} \cdot \Delta(\theta_0) = \begin{pmatrix} \Delta_{\overline{h}_J}(\theta_0) & 0_{(J+1)n_Y^2 \times n_X^2} \end{pmatrix}.$$

To use this property in a constructive way, let v be an $n_\theta + n_X^2$ vector that is in the null space of $\Delta(\theta_0)$. The above equality implies that

$$\begin{pmatrix} \Delta_{\overline{h}_J}(\theta_0) & 0_{(J+1)n_Y^2 \times n_X^2} \end{pmatrix} v = 0_{(J+1)n_Y^2 \times 1}.$$

[13] Qu and Tkachenko (2010) considered a modified information matrix and studied its rank.

[14] This follows from Proposition 1-NS which shows that for any observationally equivalent values θ_0 and θ_1, $h_a(j;\theta_1)L_a(\theta_1) = h_a(j;\theta_0)L_a(\theta_0)$ for all $j \geqslant 0$.

In particular, the first n_θ components of v form a sub-vector in the null space of $\Delta_{\overline{h}_J}(\theta_0)$. Its entries are the combinations of components of θ that leave $\overline{h}_J(\theta_0)$ unchanged. In other words, these are the elements of θ that are responsible for identification failure. This suggests a simple procedure for determining the source of non-identification: (1) if the rank test on $\Delta(\theta_0)$ fails, compute its null space; (2) for each v in the null space, find the position of its non-zero entries in the first n_θ rows. These positions correspond to the components of θ that are not identifiable without restrictions.[15]

Result (ii) of the Lemma follows from the fact that the Hankel matrix of a minimal system of order n_X consists of $2n_X - 1$ Markov parameters. These uniquely determine all $j \geqslant 2n_X - 2$ Markov parameters. As rank $\Delta_{\overline{h}_J}(\theta) =$ rank $\Delta_{\overline{h}_{2n_X-2}}(\theta)$ for any $J \geqslant 2n_X - 2$, increasing the number of Markov parameters does not increase the rank of the derivative matrix. Minimality thus allows us to use a finite number of Markov parameters containing non-redundant information for identification. This avoids direct computation of the autocovariances $\Gamma_Y(j;\theta)$ which would have involved infinite Markov parameters.

Result (iii) says that our rank condition holds if and only if $\overline{h}_{2n_X-2}(\theta)$ (locally) uniquely determines θ_0. This result allows us to link the rank condition in Proposition 2-NS to $I(\theta)$, the information matrix of the model. To make this link, let ξ be the identifiable parameters of the model and denote by $I(\xi)$ the corresponding information matrix. It is always possible to decompose $I(\theta)$ into

$$I(\theta) = \left(\frac{\partial \xi}{\partial \theta}\right)' I(\xi) \left(\frac{\partial \xi}{\partial \theta}\right).$$

Since ξ is identifiable, $I(\xi)$ is nonsingular. The rank of the information matrix is given by rank $I(\theta) =$ rank $\left(\frac{\partial \xi}{\partial \theta}\right)$. Obviously, $I(\theta_0)$ is full rank n_θ if and only if rank $\frac{\partial \xi}{\partial \theta}|_{\theta=\theta_0} = n_\theta$.[16]

For the above information matrix decomposition to be useful, we need to find an identifiable parameter ξ. By Proposition 1-NS, $\overline{h}_J(\theta)$ is identified from the second moments of $\{Y_t\}$ for any $J \geqslant 0$. Hence, if we define ξ_J to be vec $\overline{h}_J(\theta)$, then $I(\xi_J)$ is full rank. It follows that $I(\theta_0)$ is nonsingular if and only if there exists a $\overline{J} \geqslant 0$ such that rank $\frac{\partial \xi_{\overline{J}}}{\partial \theta}|_{\theta=\theta_0} = n_\theta$. Using result (ii), a necessary and sufficient condition for $I(\theta_0)$ to be nonsingular is that rank $\frac{\partial \xi_{2n_X-2}}{\partial \theta}|_{\theta=\theta_0} = n_\theta$. Combining this with result (iii) then shows that $I(\theta_0)$ is nonsingular if and only if rank $\Delta(\theta_0) = n_\theta + n_X^2$. Thus, the information matrix is full rank if and only if our rank condition is satisfied. While likelihood analysis requires ϵ_t to be iid, we require ϵ_t to be white noise and that the other assumptions in Proposition 2-NS hold.

Lemma 3 also enables a comparison of the order conditions. The dimension of $\xi_{2n_X-2} =$ vec $\overline{h}_{2n_X-2}(\theta)$ is $n_\xi = (2n_X - 1)n_Y^2$ and the order condition based on the information matrix is

[15] A similar procedure based on the null space of $\Delta^{\mathrm{S}}(\theta_0)$ can be used in the singular case.

[16] In Iskrev (2007), $I(\xi)$ is that of the (time domain) information matrix of the reduced form (ABCD) model which is not necessarily full rank. Hence, the decomposition is not useful for identification analysis.

$n_\theta \leqslant (2n_X - 1)n_Y^2$. Our order condition stated in Proposition 2-NS only requires $n_\theta \leqslant 2n_X n_Y + n_Y(n_Y+1)/2$. Since $2n_X n_Y + n_Y(n_Y+1)/2 < (2n_X - 1)n_Y^2$ whenever $n_X > 1$ and $n_Y > 1$, our order condition based on $\Delta(\theta)$ is generally tighter than the one based on the information matrix.

7 Example

An and Schorfheide (2007) consider a model whose log-linearized solution is given by:[17]

$$\begin{aligned}
y_t &= E_t y_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau}\big(r_t - E_t\pi_{t+1} - E_t z_{t+1}\big) \\
\pi_t &= \beta E_t\pi_{t+1} + \frac{\tau(1-\nu)}{\nu\overline{\pi}^2\phi}(y_t - g_t) \\
c_t &= y_t - g_t \\
r_t &= \rho_r r_{t-1} + (1-\rho_r)\psi_1\pi_t + (1-\rho_r)\psi_2(y_t - g_t) + \epsilon_{rt} \\
g_t &= \rho_g g_{t-1} + \epsilon_{gt} \\
z_t &= \rho_z z_{t-1} + \epsilon_{zt}
\end{aligned}$$

with $\epsilon_{rt} \sim WN(0,\sigma_r^2)$, $\epsilon_{gt} \sim WN(0,\sigma_g^2)$, and $\epsilon_{zt} \sim WN(0,\sigma_z^2)$ mutually uncorrelated. In the above model $\overline{\pi}$ is steady state inflation rate. The parameter vector of interest is of dimension $n_\theta = 13$. We consider two sets of θ_0. The first is taken from the file SIMPAR1.TXT distributed by the authors, while the second set corresponds to the Para (2) column in Table 2 of An and Schorfheide (2007).

Let $\widetilde{X}_t \equiv \big(z_t, g_t, r_t, y_t, \pi_t, c_t, E_t(\pi_{t+1}), E_t(y_{t+1})\big)'$, $\epsilon_t \equiv (\epsilon_{zt}, \epsilon_{gt}, \epsilon_{rt})'$, and $Y_t = (r_t, y_t, \pi_t, c_t)'$. Sims' (2002) GENSYS solution evaluated at both parameter sets is determinate. However, the solution is not minimal because the state vector $\widetilde{X}_t$ consists of the expectational variables $E_t\pi_{t+1}$, $E_t y_{t+1}$, and identities. In consequence, rank $\mathcal{C} = 3 < n_X = 8$ and rank $\mathcal{O} = 6 < n_X = 8$ where $\mathcal{C}$ and $\mathcal{O}$ are the controllability and observability matrices in Assumption 5-S.

While the GENSYS solution does not yield a minimal representation of the solution, full column ranks of the submatrices $\Delta_{\Lambda T}(\theta_0)$ and $\Delta_{\Lambda U}(\theta_0)$ remain necessary for identification. Table 1 shows that for this example, the necessary condition fails. But there could be examples in which these sub-matrices are full rank. To be able to use Proposition 2-S to check if the necessary and sufficient conditions for identification are satisfied, the state vector will need to be of the smallest dimension possible.

Minimality is not a restrictive assumption as Kalman's decomposition theorem assures that a minimal state system can always be constructed by eliminating the uncontrollable and unobservable

[17] An appendix that contains additional examples explains how to take the numerical results from several solution algorithms to the form in (1a)-(1b) and is available upon request.

states.[18] The problem of finding a minimal representation of the solution is made simple by the fact that DSGE models are based on microfoundations, and we know which are the exogenous and endogenous state variables. Algorithms such as that of Klein (2000) and Uhlig (1999) require a user to specify the size of the state vector. These algorithms usually yield a solution of the form

$$\widetilde{X}_{t+1} = \begin{pmatrix} X_{1,t+1} \\ X_{2,t+1} \end{pmatrix} = \begin{pmatrix} \widetilde{A}_1(\theta) & 0 \\ \widetilde{A}_2(\theta) & 0 \end{pmatrix} \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} + \begin{pmatrix} \widetilde{B}_1(\theta) \\ \widetilde{B}_2(\theta) \end{pmatrix} \epsilon_{t+1}$$
$$Y_{t+1} = \begin{pmatrix} \widetilde{C}_1(\theta) & \widetilde{C}_2(\theta) \end{pmatrix} \begin{pmatrix} X_{1,t+1} \\ X_{2,t+1} \end{pmatrix}.$$

By suitable ordering of the variables, the GENSYS solution will also have this form. Note that the $\widetilde{A}$ matrix has columns of zeros. In most cases, X_{1t} is the minimal state vector.[19] The ABCD matrices are then defined upon simple substitutions:

$$X_{1,t+1} = \underbrace{\widetilde{A}_1(\theta)}_{A(\theta)} X_{1t} + \underbrace{\widetilde{B}_1(\theta)}_{B(\theta)} \epsilon_{t+1}$$
$$Y_{t+1} = \underbrace{\left(\widetilde{C}_1(\theta)\widetilde{A}_1(\theta) + \widetilde{C}_2(\theta)\widetilde{A}_2(\theta)\right)}_{C(\theta)} X_{1t} + \underbrace{\left(\widetilde{C}_1(\theta)\widetilde{B}_1(\theta) + \widetilde{C}_2(\theta)\widetilde{B}_2(\theta)\right)}_{D(\theta)} \epsilon_{t+1}.$$

In the An-Schorfheide example, $X_{1t} \equiv \left(z_t, g_t, r_t\right)'$ and the new system has $n_X = 3$. The minimal state space form with $Y_{t+1} = (r_{t+1}, y_{t+1}, \pi_{t+1}, c_{t+1})$ is presented in Table 1. As $n_\epsilon = 3 < n_Y = 4$, the model is singular. Table 1 shows that some of the $n_\theta = 13$ parameters of this model are not identified. An analysis of the null space of $\Delta^{\text{S}}(\theta_0)$ quickly reveals that the columns corresponding to $\nu, \phi, \overline{\pi}$ are not linearly independent. As already pointed out in An and Schorfheide (2007), these three parameters are not separately identified and rank $\Delta^{\text{S}}_\Lambda(\theta_0)$ should indeed be less than $n_\theta = 13$. While non-identifiability at this 13 dimensional θ_0 is a foregone conclusion, it provides a useful case study to examine some numerical issues involved.

The rank of any matrix M is determined by the number of its non-zero eigenvalues. Any identification analysis must confront the problem of determining how small an eigenvalue is deemed small. Matlab uses tolerance TOL=max(size(M))EPS($\|M\|$), where EPS is the float point precision of M. This default tolerance does not take into account that the $\Delta^{\text{S}}(\theta_0)$ matrix is often sparse and can lead to misleading results. The present example bears this out. We consider 11 values of TOL ranging from 1e-2 to 1e-11, along with the Matlab's default (last row). Clearly, the rank of $\Delta^{\text{S}}_\Lambda(\theta_0)$ varies with TOL. If TOL is set to default, the rank is 13, suggesting identification may be possible

[18] The MINREAL function in Matlab produces the minimal state vectors from an eigenvector analysis of the non-minimal state variables. However, the eigenvectors depend on the parameters to be identified and cannot be used directly for identification without carefully defining the ABCD matrices.

[19] In other cases, the endogenous state variables defined by identities will also need to be removed from X_{1t}. This is straightforward because their corresponding rows in $\widetilde{B}_1(\theta)$ will be zero.

even though we know that the model is not identifiable! An overly tight tolerance here clearly gives the wrong result.

How should we set TOL? We use TOL=1e-3 in the baseline analysis on the ground that the numerical derivatives are computed using a step size of 1e-3. Furthermore, the rank of $\Delta_\Lambda^{\text{S}}(\theta_0)$ is unchanged for a range of smaller and larger values of TOL. Since TOL=1e-3 is by no means the magic number, we also use the change in rank as TOL tightens as an indication that some parameters of the model are at not well identified even with infinite data. This flags the parameters that will be difficult to identify when only a finite number of observations are available.

The rank of $\Delta^{\text{S}}(\theta_0)$ suggests that three restrictions may be necessary to identify the model. To proceed with conditional identification, the rank of $\overline{\Delta}^{\text{S}}(\theta_0)$ is evaluated for 11 sets of restrictions. Recall that in our analysis, a restriction adds to the rows of $\overline{\Delta}^{\text{S}}(\theta_0)$ but leaves the number of columns unchanged. Thus, the rank required for identification is always $n_\theta + n_X^2 + n_\epsilon^2 = 31$. As expected, three restrictions are necessary. It is quickly found that ν, ϕ and one of ψ_1 or ψ_2 needs to be restricted for identification. Fixing ν and ϕ always leads to full rank of $\overline{\Delta}_\Lambda^{\text{S}}(\theta_0)$, as we would expect. What is more surprising is that not every choice of third restriction leads to identification. For example, fixing ν, ϕ and β leads to a rank deficient $\overline{\Delta}_{\Lambda\, U}^{\text{S}}(\theta_0)$. The example illustrates that full column rank of the matrices $\overline{\Delta}_\Lambda^{\text{S}}(\theta_0), \overline{\Delta}_T^{\text{S}}(\theta_0), \overline{\Delta}_U^{\text{S}}(\theta_0)$ are individually necessary but not sufficient for identification.

Now consider a reparameterized model with $\kappa = \frac{\tau(1-\nu)}{\nu\overline{\pi}\phi}$ and $n_\theta = 11$. Table 2 shows that $\Delta_\Lambda^{\text{S}}(\theta_0), \Delta_T^{\text{S}}(\theta_0), \Delta_U^{\text{S}}(\theta_0)$ of the reparameterized model are individually full rank but $\Delta^{\text{S}}(\theta_0)$ is short rank by one. This is in agreement with Table 1 that fixing two of ν, ϕ, and $\overline{\pi}$ is not enough for identification. One of the three parameters in the Taylor rule (ψ_1, ψ_2, ρ_r) needs to be further restricted. This finding is reinforced when measurement errors are added to each of the four observed variables so that the model becomes non-singular.

Instead of the posterior mean, Table 3 takes θ_0 to be the mean of the priors. The autoregressive parameters for the shocks are now smaller in this configuration of θ, but the shocks have larger dispersion. At TOL=1e-3, the three sub-matrices of $\Delta^{\text{S}}(\theta_0)$ are all full rank. However, $\Delta^{\text{S}}(\theta_0)$ is full rank only in three cases, none of which involves parameters in the Phillips curve. Thus, general statements about non-identifiability of the Phillips curve cannot be made on the basis of rank test evaluated at a particular θ_0. Each parameter point has to be assessed on an individual basis. Adding measurement errors to each of the four observed variables leads to the same findings.

The rank conditions defined in Propositions 2-S and 2-NS do not require solving for the autocovariances whose numerical errors could affect the rank conditions. In the singular case, the only component of the $\Delta^{\text{S}}(\theta_0)$ matrix that is subject to numerical errors is $\Delta_\Lambda^{\text{S}}(\theta_0)$. The remaining

submatrices can be computed exactly in a few lines of code. In the non-singular case, the Kalman gain matrix needs to be solved numerically. This too is a simple iterative procedure. The code, along with results for larger and more complex models solved using other solution algorithms, are available in a supplementary file.

8 Concluding Remarks

Almost every empirical DSGE exercise estimates a subset of the parameters and fixes many others. At issue is how many restrictions are truly necessary. This paper uses the structural properties of the canonical solution to DSGE models to obtain identification results that do not require knowledge of infinite autocovariances or Markov parameters. These conditions are easy to compute; they do not depend on the data or the choice of the estimator. Because the identification conditions are based on structure of the model, the results also help us uncover features of the model that are not immediately transparent.

References

An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26: 2-4, 113–172.

Anderson, B. D. O. (1969): "The Inverse Problem of Stationary Covariance Generation," *Journal of Statistical Physics*, 1, 133–147.

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

Anderson, G., and G. Moore (1985): "A linear algebraic procedure for solving linear perfect foresight models," *Economic Letters*, 17(3), 247–252.

Antsaklis, P., and A. Michel (1997): *Linear Sytems*. McGraw-Hill, New York.

Canova, F., and L. Sala (2009): "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, 56, 431–449.

Christiano, L., M. Eichenbaum, and C. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Jornal of Political Economy*.

Consolo, A., C. Favero, and A. Paccagnini (2009): "On the Statistical Identification of DSGE Models," *Journal of Econometrics*, 150, 99–115.

Curdia, V., and R. Reis (2009): "Correlated Disturbances and U.S. Business Cycles," unpublished manuscript.

Deistler, M. (1976): "The Identifiability of Linear Econometric Models with Autocorrelated Errors," *International Economic Review*, 17, 26–46.

Fernandez-Villaverde, F., J. Rubio-Ramirez, T. Sargent, and M. Watson (2007): "A,B,Cs and (D)'s for Understanding VARs," *American Economic Review*, 97:3, 1021–1026.

Fisher, F. (1966): *The Identification Problem in Econometrics*. McGraw-Hill.

Glover, K., and J. Willems (1974): "Parameterizations of Linear Dynamical Systems: Canonical Forms and Identifiability," *IEEE Transactions on Automatic Control*, 19:6, 640–646.

Hannan, E. (1971): "The Identification Problem for Multiple Equation Systems with Moving Average Errors," *Econometrica*, 39:5, 751–765.

Hansen, L. P., and T. Sargent (2005): *Recursive Methods of Dynamic Linear Economies*. Unpublished manuscript.

Hatanaka, M. (1975): "On the Global Identification of the Dynamic Simultaneous Equations Model with Stationary Disturbances," *International Economic Review*, 16:3, 545–553.

Hausman, J., and W. Taylor (1983): "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Model Interpretation," *Econometrica*, 51:5, 1527–1549.

Iskrev, N. (2007): "Evaluating the Information Matrix in Linearized DSGE models," *Economics Letters*, 99:3, 607–610.

——— (2010): "Local Identification in DSGE Models," *Journal of Monetary Economics*, 57:2, 189–202.

Kailath, T., A. Sayed, and B. Hassibi (2000): *Linear Estimation*. Prentice Hall, New Jersey.

King, R. G., and M. W. Watson (2002): "System Reduction and Solution Algorithms for Singular Linear Difference Systems under Rational Expectations," *Computational Economics*, 20, pp. 57–86.

Klein, P. (2000): "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Jornal of Economic Dynamics and Control*, 4:2, 257–271.

Koopmans, T. C. (ed.) (1950): *Statistical Inference in Dynamic Economic Models*, vol. 10 of *Cowles Commission Monograph*. John Wiley & Sons, Inc.

Laub, A., and B. Moore (1978): "Calculation of Transmission Zeros Using QZ Techniques," *Automatica*, 14, 557–566.

Lutkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis*. Springer Verlag, Berlin.

Qu, Z., and D. Tkachenko (2010): "Identification and Frequency Domain QML Estimation of Linearized DSGE Models," mimeo, Boston University.

Reinsel, G. (2003): *Elements of Multivariate Time Series Analysis*. Springer, New York, 2nd edn.

Rothenberg, T. (1971): "Identification in Parametric Models," *Econometrica*, 39:3, 577–591.

Rozanov, Y. A. (1967): *Stationary Random Processes*. Holen, San Francisco.

Rubio-Ramírez, J. F., D. F. Waggoner, and T. Zha (2007): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economics and Studies*, Forthcoming.

Sargan, D. (1977): "The Spectral Estimation of Simultaneous Equation Systems with Lagged Endogenous Variables," *International Economic Review*, 18:3, 583–605.

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

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *The American Economic Review*, 97:3, 586–606.

Solo, V. (1986): "Topics in Advanced Time Series," in *Lectures in Probability and Statistics*, ed. by G. del Pino, and R. Rebolledo, vol. 1215. Springer-Verglag.

Uhlig, H. (1999): "A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily," in *Computational Methods for the Study of Dynamic Economies*, ed. by R. Marimon, and A. Scott, pp. 30–61. Oxford University Press.

Wallis, K. (1977): "Multiple Time Series Analysis and the Final Form of Econometric Models," *Econometrica*, 45:6, 1481–1497.

Youla, D. C. (1961): "On the Factorization of Rational Matrices," *IRE Transactions on Information Theory*, 16:3, 172–189.

Table 1: Full Model: $n_\theta = 13$

τ	β	ν	ϕ	$\overline{\pi}$	ψ_1	ψ_2	ρ_r	ρ_g	ρ_z	$100\sigma_r$	$100\sigma_g$	$100\sigma_z$
2	0.9975	0.1	53.6797	1.008	1.5	0.125	0.75	0.95	0.9	.2	.6	.3

Minimal State Space Representation

$$X_{t+1} = \begin{pmatrix} z_{t+1} \\ g_{t+1} \\ r_{t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 0.9 & 0 & 0 \\ 0 & 0.95 & 0 \\ 0.5450 & 0 & 0.5143 \end{pmatrix}}_{A(\theta)} X_{1t} + \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.6055 & 0 & 0.6858 \end{pmatrix}}_{B(\theta)} \underbrace{\begin{pmatrix} \epsilon_{zt+1} \\ \epsilon_{gt+1} \\ \epsilon_{rt+1} \end{pmatrix}}_{\epsilon_{t+1}}$$

$$Y_{t+1} = \begin{pmatrix} r_{t+1} \\ y_{t+1} \\ \pi_{t+1} \\ c_{t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} .5450 & 0 & 0.5143 \\ 1.3377 & 0.95 & -0.8258 \\ 1.3418 & 0 & -0.5596 \\ 1.3377 & 0 & -0.8258 \end{pmatrix}}_{C(\theta)} X_{1t} + \underbrace{\begin{pmatrix} 0.6055 & 0 & 0.6858 \\ 1.4863 & 1 & -1.1011 \\ 1.4909 & 0 & -0.7462 \\ 1.4863 & 0 & -1.1011 \end{pmatrix}}_{D(\theta)} \epsilon_{t+1}$$

	Non-Minimal Model				Minimal Model						
Tol	$\Delta^{\mathrm{S}}_{\Lambda}$	Δ^{S}_{U}	$\Delta^{\mathrm{S}}_{\Lambda U}$	pass	$\Delta^{\mathrm{S}}_{\Lambda}$	Δ^{S}_{T}	Δ^{S}_{U}	$\Delta^{\mathrm{S}}_{\Lambda T}$	$\Delta^{\mathrm{S}}_{\Lambda U}$	Δ^{S}	pass
e-02	11	9	19	0	11	9	9	20	19	28	0
e-03	11	9	19	0	11	9	9	20	19	28	0
e-04	11	9	19	0	11	9	9	20	19	28	0
e-05	11	9	19	0	11	9	9	20	19	28	0
e-06	11	9	19	0	11	9	9	20	19	28	0
e-07	12	9	21	0	11	9	9	20	20	29	0
e-08	12	9	21	0	11	9	9	20	20	29	0
e-09	12	9	21	0	11	9	9	20	20	29	0
e-10	12	9	21	0	12	9	9	21	21	29	0
e-11	12	9	21	0	12	9	9	21	21	29	0
default	13	9	22	1	12	9	9	21	21	30	0
Required	13	9	22	1	13	9	9	22	22	31	1

Full Minimal Model with Restrictions: Tol=1e-3

Restriction			$\overline{\Delta}^{\mathrm{S}}_{\Lambda}$	$\overline{\Delta}^{\mathrm{S}}_{T}$	$\overline{\Delta}^{\mathrm{S}}_{U}$	$\overline{\Delta}^{\mathrm{S}}_{\Lambda,T}$	$\overline{\Delta}^{\mathrm{S}}_{\Lambda,U}$	$\overline{\Delta}^{\mathrm{S}}$	pass
ν	-	-	12	9	9	21	20	29	0
ν	ϕ	-	13	9	9	22	21	30	0
ϕ	$\overline{\pi}$	-	13	9	9	22	21	30	0
ν	$\overline{\pi}$	-	13	9	9	22	21	30	0
β	ϕ	-	12	9	9	21	20	29	0
ϕ	ρ_g	-	12	9	9	21	20	29	0
β	ν	ϕ	13	9	9	22	21	30	0
β	ψ_1	ψ_2	11	9	9	20	20	29	0
ν	ϕ	ψ_1	13	9	9	22	22	31	1
ν	ϕ	ψ_2	13	9	9	22	22	31	1
τ	ψ_1	ψ_2	11	9	9	20	20	29	0
Required			13	9	9	22	22	31	1

Table 2: Reparameterized Model with $\kappa = \frac{\tau(1-\nu)}{\nu\bar{\pi}^2\phi}$

τ	β	κ	ψ_1	ψ_2	ρ_r	ρ_g	ρ_z	$100\sigma_r$	$100\sigma_g$	$100\sigma_z$	$100\sigma_{vr}$	$100\sigma_{vy}$	$100\sigma_{v\pi}$	$100\sigma_{vc}$
2	.9975	.33	1.5	.125	.75	.95	.9	.2	.6	.3	.2	.2	.2	.2

Tol = 1e-3	No Measurement Errors, $n_\theta = 11$							With Errors, $n_\theta = 15$			
Restriction	$\overline{\Delta}^{S}_{\Lambda}$	$\overline{\Delta}^{S}_{T}$	$\overline{\Delta}^{S}_{U}$	$\overline{\Delta}^{S}_{\Lambda,T}$	$\overline{\Delta}^{S}_{\Lambda,U}$	$\overline{\Delta}^{S}$	pass	$\overline{\Delta}^{NS}_{\Lambda}$	$\overline{\Delta}^{NS}_{T}$	$\overline{\Delta}^{NS}$	pass
τ	11	9	9	20	19	28	0	14	9	23	0
β	11	9	9	20	19	28	0	14	9	23	0
κ	11	9	9	20	19	28	0	14	9	23	0
ψ_1	11	9	9	20	20	29	1	15	9	24	1
ψ_2	11	9	9	20	20	29	1	15	9	24	1
ρ_r	11	9	9	20	20	29	1	15	9	24	1
ρ_g	11	9	9	20	19	28	0	14	9	23	0
ρ_z	11	9	9	20	19	28	0	14	9	23	0
σ_r^2	11	9	9	20	19	28	0	14	9	23	0
σ_g^2	11	9	9	20	19	28	0	14	9	23	0
σ_z^2	11	9	9	20	19	28	0	14	9	23	0
Required	11	9	9	20	20	29	1	15	9	24	1

Table 3: Reparameterized Model with $\kappa = \frac{\tau(1-\nu)}{\nu\bar{\pi}^2\phi}$

τ	β	κ	ψ_1	ψ_2	ρ_r	ρ_g	ρ_z	$100\sigma_r$	$100\sigma_g$	$100\sigma_z$	$100\sigma_{vr}$	$100\sigma_{vy}$	$100\sigma_{v\pi}$	$100\sigma_{vc}$
.5	.9988	.1	.25	.25	.2	.1	.15	4	4	4	.2	.2	.2	.2

Tol = 1e-3	No Measurement Errors, $n_\theta = 11$							With Errors, $n_\theta = 15$			
Restriction	$\overline{\Delta}^{S}_{\Lambda}$	$\overline{\Delta}^{S}_{T}$	$\overline{\Delta}^{S}_{U}$	$\overline{\Delta}^{S}_{\Lambda,T}$	$\overline{\Delta}^{S}_{\Lambda,U}$	$\overline{\Delta}^{S}$	pass	$\overline{\Delta}^{NS}_{\Lambda}$	$\overline{\Delta}^{NS}_{T}$	$\overline{\Delta}^{NS}$	pass
τ	11	9	9	20	20	29	1	15	9	24	1
β	11	9	9	20	20	29	1	15	9	24	1
κ	11	9	9	20	20	29	1	15	9	24	1
ψ_1	11	9	9	20	20	28	0	14	9	23	0
ψ_2	11	9	9	20	20	28	0	14	9	23	0
ρ_r	11	9	9	20	20	28	0	14	9	23	0
ρ_g	11	9	9	20	20	28	0	14	9	23	0
ρ_z	11	9	9	20	20	28	0	14	9	23	0
σ_r^2	11	9	9	20	20	28	0	14	9	23	0
σ_g^2	11	9	9	20	20	28	0	14	9	23	0
σ_z^2	11	9	9	20	20	28	0	14	9	23	0
Required	11	9	9	20	20	29	1	15	9	24	1

Appendix

Proof of Lemma 1

For any $\theta \in \Theta$, let $\mathbb{A}(\theta) \subset \mathbb{C}$ be the set of eigenvalues of $A(\theta)$. Note that the set $\mathbb{A}(\theta)$ contains at most n_X distinct points in $\mathbb{C}$. There are two cases to consider.

Case $n_Y \geqslant n_\epsilon$. For any $\theta \in \Theta$ and any $z \in \mathbb{C}\backslash\mathbb{A}(\theta)$, consider the identity

$$\begin{pmatrix} I_{n_X} & 0_{n_X \times n_Y} \\ C(\theta)[zI_{n_X} - A(\theta)]^{-1} & I_{n_Y} \end{pmatrix} \underbrace{\begin{pmatrix} zI_{n_X} - A(\theta) & B(\theta) \\ -C(\theta) & D(\theta) \end{pmatrix}}_{\mathcal{P}(z;\theta)} = \begin{pmatrix} zI_{n_X} - A(\theta) & B(\theta) \\ 0_{n_Y \times n_X} & H_\epsilon(z;\theta) \end{pmatrix}.$$

Note that $\det\left(zI_{n_X} - A(\theta)\right) \neq 0$ for any $z \notin \mathbb{A}(\theta)$ so the first matrix is well defined. Thus, for any $z \in \mathbb{C}\backslash\mathbb{A}(\theta)$,

$$\begin{aligned} \text{rank} \begin{pmatrix} z\mathrm{I}_{n_X} - A(\theta) & B(\theta) \\ -C(\theta) & D(\theta) \end{pmatrix} &= \text{rank} \begin{pmatrix} zI_{n_X} - A(\theta) & B(\theta) \\ 0_{n_Y \times n_X} & H_\epsilon(z;\theta) \end{pmatrix} \\ &= \text{rank } (zI_{n_X} - A(\theta)) + \text{rank } H_\epsilon(z;\theta). \end{aligned}$$

Since $\det(zI_{n_X} - A(\theta)) \neq 0$ for any $z \in \mathbb{C}\backslash\mathbb{A}(\theta)$, the results follows.

Case $n_Y < n_\epsilon$. For any $z \in \mathbb{C}\backslash\mathbb{A}(\theta)$, $\det(z\mathrm{I}_{n_X} - A(\theta)) \neq 0$ and the following matrix equality holds:

$$\begin{pmatrix} z\mathrm{I}_{n_X} - A(\theta) & B(\theta) \\ -C(\theta) & D(\theta) \end{pmatrix} \underbrace{\begin{pmatrix} -[z\mathrm{I}_{n_X} - A(\theta)]^{-1}B(\theta) & [z\mathrm{I}_{n_X} - A(\theta)]^{-1} \\ \mathrm{I}_{n_\epsilon} & 0_{n_\epsilon \times n_X} \end{pmatrix}}_{\mathcal{R}(z;\theta)} = \underbrace{\begin{pmatrix} 0_{n_X \times n_\epsilon} & \mathrm{I}_{n_X} \\ H_\epsilon(z;\theta) & -C(\theta)[z\mathrm{I}_{n_X} - A(\theta)]^{-1} \end{pmatrix}}_{\mathcal{Q}(z;\theta)}$$

with rank $\mathcal{R}(z;\theta) = n_X + n_\epsilon$. Hence rank $\mathcal{P}(z;\theta) =$ rank $\mathcal{Q}(z;\theta) =$ rank $H_\epsilon(z;\theta) + n_X$. □

When $D(\theta)$ is invertible and $\mathcal{P}(z;\theta)$ is square ($n_Y = n_\epsilon$),

$$\mathcal{P}(z;\theta) \begin{pmatrix} I_{n_X} & 0_{n_X \times n_\epsilon} \\ D^{-1}(\theta)C(\theta) & I_{n_\epsilon} \end{pmatrix} = \begin{pmatrix} zI_{n_X} - (A(\theta) - B(\theta)D^{-1}(\theta)C(\theta)) & B(\theta) \\ 0_{n_Y \times n_X} & D(\theta) \end{pmatrix}, \quad \forall z \in \mathbb{C}.$$

Since $\det(\mathcal{P}(z;\theta)) = \det(D(\theta)) \det(zI_{n_X} - (A(\theta) - B(\theta)D^{-1}(\theta)C(\theta))$, the zeros of $\det(\mathcal{P}(z;\theta))$ are the eigenvalues of $A(\theta) - B(\theta)D^{-1}(\theta)C(\theta)$, which is the test proposed in Fernandez-Villaverde, Rubio-Ramirez, Sargent, and Watson (2007). In general, finding the values of z where a matrix of the form $zM - P(\theta)$ drops rank and M is possibly singular or non-square is a generalized eigenvalue problem, with z being the generalized eigenvalues (see, e.g., Laub and Moore, 1978). In our case, $M = \begin{bmatrix} I_{n_X} & 0_{n_X \times n_\epsilon} \\ 0_{n_Y \times n_X} & 0_{n_Y \times n_\epsilon} \end{bmatrix}$, $P(\theta) = \begin{bmatrix} A(\theta) & -B(\theta) \\ C(\theta) & -D(\theta) \end{bmatrix}$, and $\mathcal{P}(z;\theta) = zM - P(\theta)$. Thus, the rank test can also be formulated as a generalized eigenvalue test.

Proof of Proposition 1-S

The proof combines two results: the spectral factorization result and the similarity transformation.

STEP 1. SPECTRAL FACTORIZATION. The key argument is the following (e.g., Youla, 1961; Anderson, 1969; Kailath, Sayed, and Hassibi, 2000, p.205). Let r be the rank a.e. of the spectral density $\Omega_Y(z;\theta)$. If $W(z;\theta)$ is an $n_Y \times r$ matrix such that for all $z \in \mathbb{C}$, $\Omega_Y(z;\theta) = W(z;\theta)W(z^{-1};\theta)'$, and rank $W(z;\theta) = r$ for all $|z| > 1$, then $W(z;\theta)$ is a *left-invertible* (or minimum phase) spectral factor that is unique up to a right multiplication by a constant orthogonal $r \times r$ matrix V. That is to say, if $W(z;\theta_0)$ and $W(z;\theta_1)$ are two left-invertible spectral factors that satisfy $W(z;\theta_0)W(z^{-1};\theta_0)' = W(z;\theta_1)W(z^{-1};\theta_1)'$, then necessarily $W(z;\theta_1) = W(z;\theta_0)V$ with $V'V = VV' = \mathrm{I}_r$.

Under Assumptions 2 and 4-S, the transfer function $H_\epsilon(z;\theta)$ is left-invertible in $|z| > 1$. Combing this with Assumption 1 gives rank $W(z;\theta) =$ rank $(H_\epsilon(z;\theta)L_\epsilon(\theta)) = n_\epsilon$, for all $|z| > 1$. By Lemma 1 and Assumption 1, $\Omega_Y(z;\theta) = H_\epsilon(z;\theta)\Sigma_\epsilon(\theta)H_\epsilon(z^{-1};\theta)'$ is of rank n_ϵ a.e. in $\mathbb{C}$. Hence, $W(z;\theta)$ a left-invertible spectral factor. Using the above spectral factorization result it then follows that $\Omega_Y(z;\theta_1) = \Omega_Y(z;\theta_0)$ for all $z \in \mathbb{C}$ if and only if there is an orthogonal $n_\epsilon \times n_\epsilon$ matrix V, such that

$$H_\epsilon(z;\theta_1)L_\epsilon(\theta_1) = H_\epsilon(z;\theta_0)L_\epsilon(\theta_0)V, \quad \text{for every } z \in \mathbb{C}. \tag{12}$$

STEP 2. Necessity and sufficiency of the similarity transforms follow from Theorem 3.10 Antsaklis and Michel (1997). It remains to combine the result of (12) with the similarity transform. From (12)

$$\underbrace{D(\theta_1)L_\epsilon(\theta_1)}_{\mathbb{D}(\theta_1)} + C(\theta_1)[z\mathrm{I}_{n_X} - A(\theta_1)]^{-1}\underbrace{B(\theta_1)L_\epsilon(\theta_1)}_{\mathbb{B}(\theta_1)} = \underbrace{D(\theta_0)L_\epsilon(\theta_0)V}_{\mathbb{D}(\theta_0)} + C(\theta_0)[z\mathrm{I}_{n_X} - A(\theta_0)]^{-1}\underbrace{B(\theta_0)L_\epsilon(\theta_0)V}_{\mathbb{B}(\theta_0)}.$$

The system $(A(\theta), \mathbb{B}(\theta), C(\theta), \mathbb{D}(\theta))$ is minimal whenever $(A(\theta), B(\theta), C(\theta), D(\theta))$ is minimal, which holds under Assumption 5-S. Thus the above equality can only hold if there exists a full rank $n_X \times n_X$ matrix T such that $\mathbb{D}(\theta_1) = \mathbb{D}(\theta_0)$, $A(\theta_1) = TA(\theta_0)T^{-1}$, $\mathbb{B}(\theta_1) = T\mathbb{B}(\theta_0)$, $C(\theta_1) = C(\theta_0)T^{-1}$, that is $D(\theta_1)L_\epsilon(\theta_1) = D(\theta_0)L_\epsilon(\theta_0)V$, $A(\theta_1) = TA(\theta_0)T^{-1}$, $B(\theta_1)L_\epsilon(\theta_1) = TB(\theta_0)L_\epsilon(\theta_0)V$, $C(\theta_1) = C(\theta_0)T^{-1}$. Letting $U \equiv L_\epsilon(\theta_0)VL_\epsilon(\theta_1)^{-1}$, be a full rank $n_\epsilon \times n_\epsilon$ matrix so that $U\Sigma_\epsilon(\theta_1)U' = \Sigma_\epsilon(\theta_0)$ gives the desired result. □

Proof of Lemma 2-S

The proof is in two steps.

SUFFICIENCY: Consider the contrapositive. Suppose that θ_0 is not locally identifiable. Then there exists an infinite sequence of parameter vectors $\{\theta_1, \ldots, \theta_k, \ldots\}$ (of dimension n_θ) approaching θ_0 such that $\Omega_Y(z;\theta_k) = \Omega_Y(z;\theta_0)$ for all $z \in \mathbb{C}$. By Proposition 1-S, this implies that there exist infinite sequences of full rank $n_X \times n_X$ matrices $\{T_1, \ldots, T_k, \ldots\}$ and full rank $n_\epsilon \times n_\epsilon$ matrices $\{U_1, \ldots, U_k, \ldots\}$ such that: $T_kA(\theta_k)T_k^{-1} = A(\theta_0)$, $T_kB(\theta_k)U_k = B(\theta_0)$, $C(\theta_k)T_k^{-1} = C(\theta_0)$, $D(\theta_k)U_k = D(\theta_0)$, $U_k^{-1}\Sigma(\theta_k){U_k^{-1}}' = \Sigma(\theta_0)$, i.e. $\delta^{\mathrm{S}}(\theta_k, T_k, U_k) = \delta^{\mathrm{S}}(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})$. In order to show

that the mapping δ^{S} is not locally injective, it suffices to show that the sequences $\{T_1, \ldots, T_k, \ldots\}$ and $\{U_1, \ldots, U_k, \ldots\}$ approach I_{n_X} and I_{n_ϵ}, respectively. For this, note that:

$$\mathcal{O}(\theta_k) \equiv \begin{pmatrix} C(\theta_k) \\ C(\theta_k)A(\theta_k) \\ \vdots \\ C(\theta_k)A^{n_X-1}(\theta_k) \end{pmatrix} = \begin{pmatrix} C(\theta_0)T_k \\ C(\theta_0)T_kT_k^{-1}A(\theta_0)T_k \\ \vdots \\ C(\theta_0)T_kT_k^{-1}A^{n_X-1}(\theta_0)T_k \end{pmatrix} = \mathcal{O}(\theta_0)T_k,$$

where $\mathcal{O}(\theta)$ is the observability matrix of $(A(\theta), C(\theta))$. Since for all θ, the system is observable, rank $\mathcal{O}(\theta_0) = n_X$ and a left inverse exists which gives $T_k = \left[\mathcal{O}(\theta_0)'\mathcal{O}(\theta_0)\right]^{-1}\mathcal{O}(\theta_0)'\mathcal{O}(\theta_k)$. By continuity of $\mathcal{O}(\theta)$, $\mathcal{O}(\theta_k)$ approaches $\mathcal{O}(\theta_0)$ as θ_k approaches θ_0, so T_k approaches I_{n_X}. To show that U_k approaches I_{n_ϵ}, take any $|z| > 1$ and note that

$$\mathcal{P}(z;\theta_0) = \begin{pmatrix} z\text{I}_{n_X} - A(\theta_0) & B(\theta_0) \\ -C(\theta_0) & D(\theta_0) \end{pmatrix} = \begin{pmatrix} z\text{I}_{n_X} - A(\theta_k) & B(\theta_k)U_k \\ -C(\theta_k) & D(\theta_k)U_k \end{pmatrix} = \mathcal{P}(z;\theta_k)\begin{pmatrix} \text{I}_{n_X} & 0 \\ 0 & U_k \end{pmatrix}.$$

Since rank $\mathcal{P}(z;\theta_k) = n_x + n_\epsilon$ a left inverse exists and:

$$\begin{pmatrix} \text{I}_{n_X} & 0 \\ 0 & U_k \end{pmatrix} = \left[\mathcal{P}(z;\theta_k)'\mathcal{P}(z;\theta_k)\right]^{-1}\left[\mathcal{P}(z;\theta_k)'\mathcal{P}(z;\theta_0)\right].$$

It follows from continuity that U_k approaches I_{n_ϵ} as θ_k approaches θ_0. This shows that δ^{S} is not injective in the neighborhood of $(\theta_0, \text{I}_{n_X}, \text{I}_{n_\epsilon})$.

NECESSITY: To show that θ_0 locally identifiable implies local injectivity of δ^{S}, consider (θ_1, T, U) with $\theta_1 \in \Theta$, T and U full rank $n_X \times n_X$ and $n_\epsilon \times n_\epsilon$ matrices, respectively, such that $\delta^{\text{S}}(\theta_1, T, U) = \delta^{\text{S}}(\theta_0, \text{I}_{n_X}, \text{I}_{n_\epsilon})$. That is, $TA(\theta_1)T^{-1} = A(\theta_0)$, $TB(\theta_1)U = B(\theta_0)$, $C(\theta_1)T^{-1} = C(\theta_0)$, $D(\theta_1)U = D(\theta_0)$, and $U^{-1}\Sigma(\theta_1)U^{-1\prime} = \Sigma(\theta_0)$. This implies that $\Omega_Y(z;\theta_1) = \Omega_Y(z;\theta_0)$ for all $z \in \mathbb{C}$. Since θ_0 is locally identifiable, there exists a neighborhood in which $\theta_1 = \theta_0$. To show that δ^{S} is locally injective it suffices to show that $\theta_1 = \theta_0$ implies $T = \text{I}_{n_X}$ and $U = \text{I}_{n_\epsilon}$. For this, consider again the observability matrices and note that $\mathcal{O}(\theta_1)T^{-1} = \mathcal{O}(\theta_0)$. Under observability, rank $\mathcal{O}(\theta_1) =$ rank $\mathcal{O}(\theta_0) = n_X$ so $\theta_1 = \theta_0$ implies $T = \text{I}_{n_X}$. We can then use left-invertibility of $\mathcal{P}(z;\theta_1)$ and $\mathcal{P}(z;\theta_0)$ in $|z| > 1$ to show that $U = \text{I}_{n_\epsilon}$. Hence, δ^{S} is locally injective. □

Proof of Proposition 2-S

The proof consists of two parts: the first establishes the rank condition; the second derives the order condition. Let vech be the operator for the column-wise vectorization with the upper portion excluded. In order to 'invert' the vech operator, we use an $n^2 \times n(n+1)/2$ duplication matrix $\mathcal{G}_n$ which is a matrix of 0s and 1s, with a single 1 in each row. Thus for any $n \times n$ symmetric matrix S, vec $(S) = \mathcal{G}_n$vech (S). The matrix $\mathcal{G}_n$ is full column rank and we let $\mathcal{E}_n = (\mathcal{G}_n'\mathcal{G}_n)^{-1}\mathcal{G}_n'$ be its left-inverse. Then $\mathcal{E}_n\mathcal{G}_n = \text{I}_{n(n+1)/2}$ and $\mathcal{G}_n\mathcal{E}_n = (1/2)[\text{I}_{n^2} + \mathcal{P}_{n,n}]$, where $\mathcal{P}_{n,n}$ is the $n^2 \times n^2$ permutation matrix that transforms vec X into vec X', i.e. $\mathcal{P}_{n,n}$vec $X =$ vec X'. Note that $\mathcal{P}_{n,n} = \mathcal{P}_{n,n}^{-1}$ and $\mathcal{P}_{n,n} = \mathcal{P}_{n,n}'$. In addition, rank$(\text{I}_{n^2} + \mathcal{P}_{n,n}) = \frac{n(n+1)}{2}$.

Rank condition Direct computations of the partial derivatives of $\delta^{\mathrm{S}}(\theta, T, U)$ give:

$$\frac{\partial \delta^{\mathrm{S}}(\theta,T,U)}{\partial \theta} = \begin{pmatrix} (T^{-1\prime} \otimes T)\frac{\partial \mathrm{vec}\, A(\theta)}{\partial \theta} \\ (U' \otimes T)\frac{\partial \mathrm{vec}\, B(\theta)}{\partial \theta} \\ (T^{-1\prime} \otimes \mathrm{I}_{n_Y})\frac{\partial \mathrm{vec}\, C(\theta)}{\partial \theta} \\ (U' \otimes \mathrm{I}_{n_Y})\frac{\partial \mathrm{vec}\, D(\theta)}{\partial \theta} \\ \mathcal{E}_{n_\epsilon}(U^{-1} \otimes U^{-1})\mathcal{G}_{n_\epsilon}\frac{\partial \mathrm{vech}\, \Sigma(\theta)}{\partial \theta} \end{pmatrix}, \tag{13}$$

$$\frac{\partial \delta^{\mathrm{S}}(\theta,T,U)}{\partial \mathrm{vec}\, T} = \begin{pmatrix} \big(T^{-1\prime} \otimes T\big)\big[\big(A(\theta)' \otimes \mathrm{I}_{n_X}\big) - \big(\mathrm{I}_{n_X} \otimes A(\theta)\big)\big]\big(\mathrm{I}_{n_X} \otimes T^{-1}\big) \\ \big(U' \otimes T\big)\big[B(\theta)' \otimes \mathrm{I}_{n_X}\big]\big(\mathrm{I}_{n_X} \otimes T^{-1}\big) \\ -\big(T^{-1\prime} \otimes \mathrm{I}_{n_Y}\big)\big[\mathrm{I}_{n_X} \otimes C(\theta)\big]\big(\mathrm{I}_{n_X} \otimes T^{-1}\big) \\ 0_{n_Y n_\epsilon \times n_X^2} \\ 0_{\frac{n_\epsilon(n_\epsilon+1)}{2} \times n_X^2} \end{pmatrix}, \tag{14}$$

and

$$\frac{\partial \delta(\theta,T,U)}{\partial \mathrm{vec}\, U} = \begin{pmatrix} 0_{n_X^2 \times n_\epsilon^2} \\ \big(U' \otimes T\big)\big[\mathrm{I}_{n_\epsilon} \otimes B(\theta)\big]\big(U^{-1\prime} \otimes \mathrm{I}_{n_\epsilon}\big) \\ 0_{n_Y n_X \times n_\epsilon^2} \\ \big(U' \otimes \mathrm{I}_{n_Y}\big)\big[\mathrm{I}_{n_\epsilon} \otimes D(\theta)\big]\big(U^{-1\prime} \otimes \mathrm{I}_{n_\epsilon}\big) \\ -\mathcal{E}_{n_\epsilon}\big(U^{-1} \otimes U^{-1}\big)\mathcal{G}_{n_\epsilon}\big[2\mathcal{E}_{n_\epsilon}(\Sigma(\theta) \otimes \mathrm{I}_{n_\epsilon})\big]\big(U^{-1\prime} \otimes \mathrm{I}_{n_\epsilon}\big) \end{pmatrix}. \tag{15}$$

Now let $\Delta(\theta) \equiv \left(\frac{\partial \delta(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})}{\partial \theta} \quad \frac{\partial \delta(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})}{\partial \mathrm{vec}\, T} \quad \frac{\partial \delta(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})}{\partial \mathrm{vec}\, U}\right)$, that is,

$$\Delta(\theta) = \begin{pmatrix} \frac{\partial \mathrm{vec}\, A(\theta)}{\partial \theta} & \big[\big(A(\theta)' \otimes \mathrm{I}_{n_X}\big) - \big(\mathrm{I}_{n_X} \otimes A(\theta)\big)\big] & 0_{n_X^2 \times n_\epsilon^2} \\ \frac{\partial \mathrm{vec}\, B(\theta)}{\partial \theta} & \big[B(\theta)' \otimes \mathrm{I}_{n_X}\big] & \big[\mathrm{I}_{n_\epsilon} \otimes B(\theta)\big] \\ \frac{\partial \mathrm{vec}\, C(\theta)}{\partial \theta} & -\big[\mathrm{I}_{n_X} \otimes C(\theta)\big] & 0_{n_Y n_X \times n_\epsilon^2} \\ \frac{\partial \mathrm{vec}\, D(\theta)}{\partial \theta} & 0_{n_Y n_\epsilon \times n_X^2} & \big[\mathrm{I}_{n_\epsilon} \otimes D(\theta)\big] \\ \frac{\partial \mathrm{vech}\, \Sigma(\theta)}{\partial \theta} & 0_{\frac{n_\epsilon(n_\epsilon+1)}{2} \times n_X^2} & -2\mathcal{E}_{n_\epsilon}\big[\Sigma(\theta) \otimes \mathrm{I}_{n_\epsilon}.\big] \end{pmatrix}. \tag{16}$$

We can write $\left(\frac{\partial \delta(\theta,T,U)}{\partial \theta} \quad \frac{\partial \delta(\theta,T,U)}{\partial \mathrm{vec}\, T} \quad \frac{\partial \delta(\theta,T,U)}{\partial \mathrm{vec}\, U}\right) = M(T,U)\Delta(\theta)N(T,U)$, where $M(T,U)$ and $N(T,U)$ are, respectively, an $n_\Lambda^{\mathrm{S}} \times n_\Lambda^{\mathrm{S}}$ diagonal matrix and an $(n_\theta + n_X^2 + n_\epsilon^2) \times (n_\theta + n_X^2 + n_\epsilon^2)$ diagonal matrix defined as:

$$M(T,U) \equiv \begin{pmatrix} T^{-1\prime} \otimes T & & & & \\ & U' \otimes T & & & \\ & & T^{-1\prime} \otimes \mathrm{I}_{n_Y} & & \\ & & & U' \otimes \mathrm{I}_{n_Y} & \\ & & & & \mathcal{E}_{n_\epsilon}(U^{-1} \otimes U^{-1})\mathcal{G}_{n_\epsilon} \end{pmatrix}$$

$$N(T,U) \equiv \begin{pmatrix} \mathrm{Id}_{n_\theta} & & \\ & \mathrm{Id}_{n_X} \otimes T^{-1} & \\ & & U^{-1\prime} \otimes \mathrm{I}_{n_\epsilon} \end{pmatrix}.$$

The $\frac{n_\epsilon(n_\epsilon+1)}{2} \times \frac{n_\epsilon(n_\epsilon+1)}{2}$ matrix $\mathcal{E}_{n_\epsilon}(U^{-1} \otimes U^{-1})\mathcal{G}_{n_\epsilon}$ is nonsingular if and only if U is nonsingular. Since T and U are full rank, both $M(T,U)$ and $N(T,U)$ are full rank,

$$\operatorname{rank}\left(\begin{array}{ccc}\frac{\partial\delta(\theta,T,U)}{\partial\theta} & \frac{\partial\delta(\theta,T,U)}{\partial\operatorname{vec}T} & \frac{\partial\delta(\theta,T,U)}{\partial\operatorname{vec}U}\end{array}\right) = \operatorname{rank}\Delta(\theta).$$

If the rank of $\Delta(\theta)$ remains constant in a neighborhood of θ_0, then the rank of the partial derivatives of δ remains constant in a neighborhood of $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\epsilon})$.

Order Condition The necessary (order) condition is established by counting the number of rows in the matrix $\Delta(\theta)$ in (16). This yields rank $\Delta(\theta) \leqslant n_\Lambda^{\mathrm{S}} = n_X^2 + n_X n_\epsilon + n_Y n_X + n_Y n_\epsilon + \frac{n_\epsilon(n_\epsilon+1)}{2}$, hence a necessary order condition is that $n_\theta \leqslant n_Y n_X + n_\epsilon(n_X + n_Y - n_\epsilon) + \frac{n_\epsilon(n_\epsilon+1)}{2}$. □

Proof of Proposition 1-NS

Unlike in the singular case, the spectral factor $W(z;\theta)$ is now a square matrix of size $n_Y \times n_Y$. We can no longer directly consider $H_\epsilon(z;\theta)L_\epsilon(\theta)$. Instead, we work with the innovations representation for Y_t. We proceed in three steps.

Step 1. The existence of the innovations representation depends on the existence of the positive semi-definite solutions to the discrete algebraic Ricatti equation (DARE):

$$\begin{aligned}&\overline{\Sigma}(\theta) = A(\theta)\overline{\Sigma}(\theta)A(\theta)' + B(\theta)\Sigma_\epsilon(\theta)B(\theta)' \qquad (17)\\ &- \left[A(\theta)\overline{\Sigma}(\theta)C(\theta)' + B(\theta)\Sigma_\epsilon(\theta)D(\theta)'\right]\left[C(\theta)\overline{\Sigma}(\theta)C(\theta)' + D(\theta)\Sigma_\epsilon(\theta)D(\theta)'\right]^{-1}\left[C(\theta)\overline{\Sigma}(\theta)A(\theta)' + D(\theta)\Sigma_\epsilon(\theta)B(\theta)'\right]\end{aligned}$$

Under Assumptions 2 (stability) and 4-NS (positive definiteness), Lemma E.3.2 of Kailath, Sayed, and Hassibi (2000) shows that there always exists a maximal positive semi-definite solution $\overline{\Sigma}(\theta)$ to the DARE (17). Moreover, if

$$\Sigma_a(\theta) \equiv C(\theta)\overline{\Sigma}(\theta)C(\theta)' + D(\theta)\Sigma_\epsilon(\theta)D(\theta)' \qquad (18)$$

$$K(\theta) \equiv \left[A(\theta)\overline{\Sigma}(\theta)C(\theta)' + B(\theta)\Sigma_\epsilon(\theta)D(\theta)'\right]\Sigma_a^{-1}(\theta) \qquad (19)$$

then all the eigenvalues of $A(\theta) - K(\theta)C(\theta)$ lie inside the closed unit disc (Lemma E.4.1 in Kailath, Sayed, and Hassibi, 2000). Thus Y_t has the following innovations representation:

$$\widehat{X}_{t+1|t+1} = A(\theta)\widehat{X}_{t|t} + K(\theta)a_{t+1} \qquad (20a)$$

$$Y_{t+1} = C(\theta)\widehat{X}_{t|t} + a_{t+1} \qquad (20b)$$

where $\widehat{X}_{t|t}$ be the optimal linear predictor of X_t based on the history Y^t, and $a_{t+1} = Y_{t+1} - C(\theta)\widehat{X}_{t|t}$ is the one-step ahead forecast error of Y_{t+1}, $a_t \sim WN(0, \Sigma_a(\theta))$. Hence, for all $z \in \mathbb{C}$, $\Omega_Y(z;\theta) = H_a(z;\theta)\Sigma_a(\theta)H_a(z^{-1};\theta)'$, with the transfer function $H_a(z;\theta) = \mathrm{I}_{n_Y} + C(\theta)\big(z\mathrm{I}_{n_X} - A(\theta)\big)^{-1}K(\theta)$

Step 2. Spectral Factorization It must first be shown that rank $H_a(z;\theta) = n_Y$ for all $|z| > 1$. As in Hansen and Sargent (2005), the proof is based on the property that for any conformable matrices a, b, c, d with a, d invertible, $\det(a)\det(d + ca^{-1}b) = \det(d)\det(a + bd^{-1}c)$. Now, let $a \equiv z\mathrm{I}_{n_X} - A(\theta)$, $b \equiv K(\theta)$, $c \equiv C(\theta)$, and $d \equiv \mathrm{I}_{n_X}$. Since $A(\theta)$ is stable, a is invertible (so

is d) and $\det(z\mathrm{I}_{n_X} - A(\theta))\det\left(\mathrm{I}_{n_X} + C(\theta)[z\mathrm{I}_{n_X} - A(\theta)]^{-1}K(\theta)\right) = \det(z\mathrm{I}_{n_X} - A(\theta) + K(\theta)C(\theta))$. Equivalently,

$$\det H_a(z;\theta) = \det\left(\mathrm{I}_{n_X} + C(\theta)[z\mathrm{I}_{n_X} - A(\theta)]^{-1}K(\theta)\right) = \frac{\det\left(z\mathrm{I}_{n_X} - [A(\theta) - K(\theta)C(\theta)]\right)}{\det(z\mathrm{I}_{n_X} - A(\theta))}. \quad (21)$$

Since $\det\left(z\mathrm{I}_{n_X} - [A(\theta) - K(\theta)C(\theta)]\right) \neq 0$ for all $|z| > 1$, it follows that rank $H_a(z;\theta) = n_Y$ for all $|z| > 1$. Now, under Assumption 4-NS, $\Sigma_a(\theta)$ in (18) is positive definite with $L_a(\theta)$ as its Cholesky decomposition. Then rank $W(z;\theta) =$ rank $(H_a(z;\theta)L_a(\theta)) = n_Y$, for all $|z| > 1$. In addition, it follows from (21) and $\Omega_Y(z;\theta) = H_a(z;\theta)\Sigma_a(\theta)H_a(z^{-1};\theta)'$ that the rank of $\Omega_Y(z;\theta)$ equals n_Y a.e. in $\mathbb{C}$. Hence, $W(z;\theta) = H_a(z;\theta)L_a(\theta)$ is a left-invertible spectral factor. By the spectral factorization result of Youla (1961) and Anderson (1969), $\Omega_Y(z;\theta_1) = \Omega_Y(z;\theta_0)$ for all $z \in \mathbb{C}$ if and only if there is an orthogonal $n_Y \times n_Y$ matrix V such that

$$H_a(z;\theta_1)L_a(\theta_1) = H_a(z;\theta_0)L_a(\theta_0)V, \quad \text{for every } z \in \mathbb{C}. \quad (22)$$

Step 3. From (22),

$$\underbrace{L_a(\theta_1)}_{\mathbb{D}(\theta_1)} + C(\theta_1)[z\mathrm{I}_{n_X} - A(\theta_1)]^{-1}\underbrace{K(\theta_1)L_a(\theta_1)}_{\mathbb{B}(\theta_1)} = \underbrace{L_a(\theta_0)V}_{\mathbb{D}(\theta_0)} + C(\theta_0)[z\mathrm{I}_{n_X} - A(\theta_0)]^{-1}\underbrace{K(\theta_0)L_a(\theta_0)V}_{\mathbb{B}(\theta_0)}.$$

Minimality of $(A(\theta), K(\theta), C(\theta), \mathrm{I}_{n_Y})$ which is ensured by Assumption 5-NS, implies minimality of $(A(\theta), \mathbb{B}(\theta), C(\theta), \mathbb{D}(\theta))$. Using the argument as in the singular case, the above equality holds only if there exists a full rank $n_X \times n_X$ matrix T such that $\mathbb{D}(\theta_1) = \mathbb{D}(\theta_0)$, $A(\theta_1) = TA(\theta_0)T^{-1}$, $\mathbb{B}(\theta_1) = T\mathbb{B}(\theta_0)$, $C(\theta_1) = C(\theta_0)T^{-1}$. Equivalently, $L_a(\theta_1) = L_a(\theta_0)V$, $A(\theta_1) = TA(\theta_0)T^{-1}$, $K(\theta_1)L_a(\theta_1) = TK(\theta_0)L_a(\theta_0)V$, $C(\theta_1) = C(\theta_0)T^{-1}$. Now uniqueness of Cholesky decomposition implies $V = \mathrm{I}_{n_Y}$. Thus $L_a(\theta_1) = L_a(\theta_0)$, and the result follows. □

Proof of Proposition 2-NS

The proof follows directly from the proof of Proposition 2-S and is hence omitted.

Proofs of Propositions 3 and 4

The proofs are analogous to those of Propositions 2-S and 2-NS.

Proof of Lemma 3

Recall that $\Lambda \equiv \Lambda^{\mathrm{NS}} = ((\mathrm{vec}\ A)', (\mathrm{vec}\ K)', (\mathrm{vec}\ C)', (\mathrm{vech}\ \Sigma_a)')'$. For any $j \geqslant 0$ and $J \geqslant 0$, let

$$\begin{aligned} \overline{h}(j;\theta) &\equiv h_a(j;\theta)L_a(\theta) \\ \overline{h}_J(\theta) &\equiv \begin{pmatrix} h_a(0;\theta)L_a(\theta) & h_a(1;\theta)L_a(\theta) & \dots & h_a(J;\theta)L_a(\theta) \end{pmatrix}, \quad \frac{\partial \mathrm{vec}\ \overline{h}_J(\theta)}{\partial \Lambda} = \begin{pmatrix} \frac{\partial \mathrm{vec}\ \overline{h}(0;\theta)}{\partial \Lambda} \\ \vdots \\ \frac{\partial \mathrm{vec}\ \overline{h}(J;\theta)}{\partial \Lambda} \end{pmatrix}. \\ \Delta_{\overline{h}}(j;\theta) &\equiv \frac{\partial \mathrm{vec}\ \overline{h}(j;\theta)}{\partial \theta} \end{aligned}$$

Direct computations show that for any $j \geqslant 0$,

$$\frac{\partial \mathrm{vec}\ \overline{h}(j;\theta)}{\partial \Lambda} = \left((L_a(\theta)' \otimes \mathrm{I}_{n_Y})\mathbb{H}_a(j;\theta) \quad (\mathrm{I}_{n_Y} \otimes h_a(j;\theta))\frac{\partial \mathrm{vec}\ L_a(\theta)}{\partial \mathrm{vech}\ \Sigma_a(\theta)}\right),$$

where $\mathbb{H}_a(j;\theta)$ is the j-th row of $\mathbb{H}_a(\theta)$ defined by

$$\mathbb{H}_\epsilon(\theta) \;=\; \begin{pmatrix} 0_{n_Y^2\times n_X^2} & 0_{n_Y^2\times n_X n_Y} & 0_{n_Y^2\times n_Y n_X} \\ 0_{n_Y^2\times n_X^2} & \mathrm{I}_{n_Y}\otimes C(\theta) & K(\theta)'\otimes \mathrm{I}_{n_Y} \\ \vdots & \vdots & \vdots \\ \sum_{k=1}^{j} K(\theta)'A^{j-k}(\theta)'\otimes C(\theta)A(\theta)^{k-1} & \mathrm{I}_{n_Y}\otimes C(\theta)A(\theta)^{j} & K(\theta)'A^{j}(\theta)'\otimes \mathrm{I}_{n_Y} \\ \vdots & \vdots & \vdots \end{pmatrix}.$$

Proof of (i) To show that for any $j \geqslant 0$, $\frac{\partial \mathrm{vec}\,\overline{h}(j;\theta)}{\partial \Lambda}\cdot \Delta_T(\theta) = 0_{n_Y^2\times n_X^2}$ (where $\Delta_T(\theta) = \Delta_T^{\mathrm{NS}}(\theta)$), write:

$$\begin{aligned}\frac{\partial \mathrm{vec}\,\overline{h}(j;\theta)}{\partial \Lambda}\cdot \Delta_T(\theta) &= \big[(L_a(\theta)'\otimes \mathrm{I}_{n_Y})\mathbb{H}_a(j;\theta)\big]\cdot \begin{pmatrix} \big[\big(A(\theta)'\otimes \mathrm{I}_{n_X}\big) - \big(\mathrm{I}_{n_X}\otimes A(\theta)\big)\big] \\ \big[K(\theta)'\otimes \mathrm{I}_{n_X}\big] \\ -\big[\mathrm{I}_{n_X}\otimes C(\theta)\big] \end{pmatrix} \\ &\quad + \left[(\mathrm{I}_{n_Y}\otimes h_a(j;\theta))\frac{\partial \mathrm{vec}\,L_a(\theta)}{\partial \mathrm{vech}\,\Sigma_a(\theta)}\right]\cdot 0_{\frac{n_Y(n_Y+1)}{2}\times n_X^2} = 0_{n_Y^2\times n_X^2},\end{aligned}$$

where the second equality follows by direct computations. This establishes result (i) of the lemma.

Proof of (ii) To show that rank $\overline{h}_J(\theta) = $ rank $\overline{h}_{2n_X-2}(\theta)$ for all $J \geqslant 2n_X-2$ we use the Cayley-Hamilton theorem which ensures that for any minimal system $(A(\theta), K(\theta), C(\theta), \mathrm{I}_{n_Y})$, rank $(\mathcal{C}_N(\theta)'\mathcal{O}_N(\theta)') =$ rank $(\mathcal{C}_{n_X}(\theta)'\mathcal{O}_{n_X}(\theta)')$, for all $N \geqslant n_X$, where $\mathcal{C}_N(\theta)$ and $\mathcal{O}_N(\theta)$ are the controllability and the observability matrices of order N, i.e.

$$\mathcal{C}_N(\theta) \equiv \begin{pmatrix} K(\theta) & \dots & A(\theta)^{N-1}K(\theta)\end{pmatrix} \quad \text{and} \quad \mathcal{O}_N(\theta) \equiv \begin{pmatrix} C(\theta) \\ \vdots \\ C(\theta)A(\theta)^{N-1}\end{pmatrix}.$$

Take $N = n_X+1$. Then $K(\theta)'A^{2n_X-1}(\theta)C(\theta)'$ is a linear combination of $(K(\theta)'A^{n_X-1}(\theta)'C(\theta)', \dots, K(\theta)'A^{2n_X-2}(\theta)'C(\theta)')$. Therefore, $L_a(\theta)'K(\theta)'A^{2n_X-1}(\theta)'C(\theta)'$ is a linear combination of $(L_a(\theta)'K(\theta)'A^{n_X-1}(\theta)'C(\theta)', \dots, L_a(\theta)'K(\theta)'A^{2n_X-2}(\theta)'C(\theta)')$. Thus, rank $\overline{h}_{2n_X-1}(\theta) =$ rank $\overline{h}_{2n_X-2}(\theta)$. The result holds recursively for any $J \geqslant 2n_X-2$.

Proof of (iii) Combining $\Delta(\theta) = \begin{pmatrix}\Delta_\Lambda(\theta) & \Delta_T(\theta)\end{pmatrix}$ and result (i) gives:

$$\frac{\partial \overline{h}_{2n_X-2}(\theta)}{\partial \Lambda}\cdot \Delta(\theta) = \begin{pmatrix}\Delta_{\overline{h}_{2n_X-2}}(\theta) & 0_{(2n_X-1)n_Y^2\times n_X^2}\end{pmatrix}. \tag{23}$$

Necessity: We need to show that rank $\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = n_\theta$ implies rank $\Delta(\theta_0) = n_\theta + n_X^2$. Now, rank $\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = n_\theta$ implies that rank $\Delta_\Lambda(\theta_0) = n_\theta$ because from (23), rank $\Delta_{\overline{h}_{2n_X-2}}(\theta_0) \leqslant$ rank $\Delta_\Lambda(\theta_0) \leqslant n_\theta$. Then proceed by contradiction: assume that rank $\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = n_\theta$ and that rank $\Delta(\theta_0) < n_\theta + n_X^2$. Then, rank $\Delta_\Lambda(\theta_0) = n_\theta$ and rank $\Delta(\theta_0) < n_\theta + n_X^2$. This means that at least one column of $\Delta_\Lambda(\theta_0)$, say $C_\Lambda(\theta_0)$, can be written as a linear combination of the columns of

$\Delta_T(\theta_0)$. Using (23), it follows that $\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda} C_\Lambda(\theta_0) = 0_{(2n_X-1)n_Y^2 \times 1}$. This implies $\Delta_{\overline{h}_{2n_X-2}}(\theta_0)$ has one zero column, which is a contradiction.

SUFFICIENCY: We need to show that rank $\Delta(\theta_0) = n_\theta + n_X^2$ implies rank $\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = n_\theta$. Now, rank $\Delta(\theta_0) = n_\theta + n_X^2$ implies rank $\Delta_\Lambda(\theta_0) = n_\theta$, and rank $\Delta_T(\theta_0) = n_X^2$. First, we show that when the system is minimal state and left-invertible,

$$\mathcal{N}\left(\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda}\right) = \operatorname{span}\big(\Delta_T(\theta_0)\big), \tag{24}$$

where $\mathcal{N}\big(\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda}\big)$ is the null space of $\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda}$ and $\operatorname{span}\big(\Delta_T(\theta_0)\big)$ denotes the subspace spanned by the columns of $\Delta_T(\theta_0)$. For this, consider a Taylor expansion of vec $\overline{h}_{2n_X-2}(\theta)$ around $\theta = \theta_0$: vec $\overline{h}_{2n_X-2}(\theta_0 + \delta) =$ vec $\overline{h}_{2n_X-2}(\theta_0) + \frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda} \Delta_\Lambda(\theta_0)\delta^*$, where $\delta^* \in (0, \delta)$. Under minimality and left-invertibility, (10) are the only transformations leading to the same $\overline{h}_{2n_X-2}$. That is, vec $\overline{h}_{2n_X-2}(\theta_0 + \delta) =$ vec $\overline{h}_{2n_X-2}(\theta_0)$ if and only if $\Delta_\Lambda(\theta_0)\delta^* \in \operatorname{span}\big(\Delta_T(\theta_0)\big)$. Combining this with the Taylor expansion gives:

$$\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda} \Delta_\Lambda(\theta_0)\delta^* = 0_{(2n_X-1)n_Y^2} \iff \Delta_\Lambda(\theta_0)\delta^* \in \operatorname{span}\big(\Delta_T(\theta_0)\big),$$

so (24) holds. To show that this implies that rank $\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = \operatorname{rank}\big(\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda} \cdot \Delta_\Lambda(\theta_0)\big) = n_\theta$, we proceed by contradiction. Suppose that rank $\big(\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda} \cdot \Delta_\Lambda(\theta_0)\big) < n_\theta$. Then there exists a nonzero linear combination of the columns of $\Delta_\Lambda(\theta_0)$ which belongs to the null space $\mathcal{N}\big(\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda}\big)$. Using (24) then there exists a nonzero linear combination of the columns of $\Delta_\Lambda(\theta_0)$ which can be written as a linear combination of the columns of $\Delta_T(\theta_0)$. This violates the assumption that rank $\Delta(\theta_0) = n_\theta + n_X^2$ and thus leads to a contradiction. □