ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 79, No. 6 (November, 2011), 1995–2032

DYNAMIC IDENTIFICATION OF DYNAMIC STOCHASTIC GENERAL EQUILIBRIUM MODELS

IVANA KOMUNJER
University of California, San Diego, La Jolla, CA 92093, U.S.A.

SERENA NG
Columbia University, New York, NY 10027, U.S.A.

Econometrica, Vol. 79, No. 6 (November, 2011), 1995–2032

NOTES AND COMMENTS

DYNAMIC IDENTIFICATION OF DYNAMIC STOCHASTIC GENERAL EQUILIBRIUM MODELS[1]

BY IVANA KOMUNJER AND SERENA NG[2]

This paper studies dynamic identification of parameters of a dynamic stochastic general equilibrium 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. Full rank of the Jacobian matrix of derivatives of the solution parameters with respect to the parameters of interest 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 models and another for nonsingular 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.

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 100 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 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 stationary 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 suggested

[1]This paper was previously circulated as "Dynamic Identification of DSGE Models."

[2]This paper was presented at the 2009 NBER Summer Institute, NBER/NSF Time Series Conference, CESG Meeting, All UC Conference, UC Davis, Harvard/MIT, NYU, Princeton, Duke/UNC/NC State, Wisconsin, LSE, UCL, and Cyprus. We thank seminar participants, Richard Davis, Nikolay Iskrev, Victor Solo, Jim Stock, and Frank Schorfheide as well as four 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 National Science Foundation Grants SES-0962473 and SES-0962431.

 DOI: 10.3982/ECTA8916

plotting some statistics of the estimated model evaluated at different parameter values. Consolo, Favero, and Paccagnini (2009) compared the properties of the DSGE model with those of a factor augmented vector autoregression (VAR). Both approaches, while useful, do not provide formal conditions for identification. Rubio-Ramírez, Waggoner, and Zha (2010) studied identification of structural VARs, but not DSGE models per se. The most complete analysis to date is due to Iskrev (2010), who proposed evaluating 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.

Our analysis has three distinctive features. First, 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 that admits a unique solution if and only if θ is identified. The rank and the order of the system provide the necessary and sufficient conditions for identification of θ. Our approach has roots in control theory which typically assumes that both the outputs (endogenous variables) and the inputs (innovations) are observed. In that literature, the restrictions implied by minimal systems are enough to completely characterize observational equivalence. However, this is not enough for DSGE models because the latent shocks have unknown variances and not all output variables are observed. For this reason, we derive new results that characterize observational equivalence by further restricting attention to transfer functions that are left-invertible. Minimality and left-invertibility, which will be explained below, are what allow us to establish the necessary and sufficient conditions for identification without computing any autocovariance or the spectral density.

Second, the identification conditions for singular and nonsingular systems are studied separately. Classical identification results are typically derived under the assumption that the number of endogenous variables equals the number of shocks and that the innovations are "fundamental." However, DSGE models can have fewer shocks than observables, a condition known as *stochastic singularity*. Once measurement errors are allowed, DSGE models may also have fewer observables than shocks and in such cases, the innovations are usually nonfundamental. As the singular and the nonsingular case each 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.

Third, our conditions for identification are stated in terms of a matrix $\Delta(\theta)$. The dimension of $\Delta(\theta)$ determines the order condition for identification and is a simple function of the number of variables that appear in the DSGE model. The rank of $\Delta(\theta)$ is a function of the parameters in the linearized solution to the DSGE model only and can be evaluated prior to estimation. Collinearity among the columns of $\Delta(\theta)$ sheds light on whether nonidentification 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 nonidentification.

Before turning to the main analysis under more general conditions, it is useful to better understand what is unusual about DSGE models. To begin with, the observables of the model are usually vector autoregressive moving-average (VARMA) processes. As discussed in Ravenna (2007), DSGE models have finite VAR representations only under specific conditions. Identification of θ would thus 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[3] to rule out redundant polynomials that lead to an overparameterized model (see, e.g., Hannan (1971)). The canonical VARMA model and its order, also known as McMillan degree, can be solved only when the system is small.[4] 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 assumptions used in Deistler (1976), for example. Thus, no attempt has been made to identify DSGE models directly from the VARMA representations.

Second, the solution of a DSGE model is not in the form of a simultaneous equations system for which classical results due to Koopmans (1950), Fisher (1966), Rothenberg (1971), and Hausman and Taylor (1983) apply. First, these results are derived for static models in which the errors of the reduced form are orthogonal to the predetermined regressors. Obviously, DSGE models are dynamic, and any serial correlation in the shocks will violate the orthogonality conditions. Furthermore, traditional analysis assumes independent and identically distributed (i.i.d.) innovations which rules out important features in time series data such as conditional heteroskedasticity. While results for full rank dynamic models are given by Hatanaka (1975), Hannan (1971), and Sargan (1975), they all assumed the presence of exogenous variables. However, there are no exogenous variables in DSGE models other than the latent shocks. Furthermore, results for full rank systems do not easily extend to singular systems. Although 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,

[3]Exchangeable means that two processes can have identical moving-average representations. For example, if there is a unimodular factor $U(L)$ in the autoregressive or moving-average matrix polynomial, then the vector moving-average (VMA) model $y_t = U(L)\Theta(L)\varepsilon_t$ is equivalent to the VARMA model $U(L)^{-1}y_t = \Theta(L)\varepsilon_t$ (see, e.g., Reinsel (2003, Section 2.3)).

[4]See, for example, Solo (1986), Reinsel (2003), and Lütkepohl (2005).

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, rendering the parameters of the "reduced form" model not identifiable. One of our main results is to show that this problem is a distinct possibility in DSGE models. This nonidentifiability of the parameters in the linearized (canonical) solution of DSGE models should not come as a surprise because, as already pointed out, the parameters of the VARMA representation for the same DSGE model may have common factors that prohibit identification.

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 nonsingular 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 application to An and Schorfheide (2007) model is given in Section 7. Additional examples are studied in the Supplemental Material (Komunjer and Ng (2011)). 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 of interest 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 $E_t \Gamma_0(\theta) X_{t+1}^a = \Gamma_1(\theta) X_t^a + \varepsilon_{zt}$ that can be solved to yield a solution in the form of difference equations. Solution algorithms include Anderson and Moore (1985), Uhlig (1999), Klein (2000), King and Watson (2002), and Sims (2002), among others. Numerical accuracy of these algorithms is taken as given in the identification analysis to follow.

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 where n_Y is finite. We allow for measurement and specification errors whose innovations are ε_{vt}. Collect all the innovations into a $n_\varepsilon \times 1$ vector $\varepsilon_t = (\varepsilon_{zt}', \varepsilon_{vt}')'$. The state space solution of Y_t is given by the transition and measurement equations

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

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

ASSUMPTION 1: *For every $\theta \in \Theta$ and every (t, s), $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_s') = \delta_{t-s}\Sigma_\varepsilon(\theta)$, where $\Sigma_\varepsilon(\theta)$ is positive definite with Cholesky decomposition $L_\varepsilon(\theta)$.*

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

Assumption 1 only requires the innovations ε_t to be white noise, which is weaker than i.i.d. Since $\Sigma_\varepsilon(\theta)$ is allowed to be nondiagonal, the innovations can be mutually correlated as in Curdia and Reis (2010). The conditional variance of ε_t can change over time to accommodate generalized autoregressive conditional heteroskedasticity (GARCH) or stochastic volatility effects, provided the process remains covariance stationary. In these cases, $\Sigma_\varepsilon(\theta)$ is the unconditional variance of ε_t which may be a complicated function of the underlying GARCH or stochastic volatility parameters. The first order effects of heteroskedasticity would be captured in a partially nonlinear approximation as in Justiniano and Primiceri (2008) or in a conditionally linear approximation as in Benigno, Benigno, and Nistico (2011). Nonlinearities that can only be captured by higher order approximations are outside the scope of our analysis.

Assumption 2 is a stability condition. In cases when there are permanent shocks, the variables of the model are assumed to be appropriately normalized so that the eigenvalues of $A(\theta)$ remain inside the unit circle. Under Assumptions 1 and 2, $\{Y_t\}$ is weakly stationary and its time series properties are completely characterized by its time invariant unconditional mean and autocovariances. The same result can also be motivated by assuming ε_t is i.i.d. Gaussian. The casual VMA(∞) representation for Y_t is

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

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

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

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 of $\{Y_t\}$ is

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

In addition to stability, it is quite frequent in econometric analysis to assume *left-invertibility* (also known as miniphase). Left-invertibility fails if, for

example, the shocks are anticipated. The issue is considered in Lippi and Reichlin (1994) and Leeper, Walker, and Yang (2008), among others. Under left-invertibility, (2) is the Wold representation and the innovations ε_t are fundamental, meaning that ε_t is spanned by $Y^t \equiv \{Y_{t-k}\}_{k=0}^{\infty}$, the current and past history of Y_t. For square models with $n_\varepsilon = n_Y$, left-invertibility holds when $\det H_\varepsilon(z;\theta) \neq 0$ in $|z| > 1$.[5] In models that are not square, left-invertibility requires that $H_\varepsilon(z;\theta)$ is full column rank in $|z| > 1$ (see, e.g., Rozanov (1967)). In general, the properties of $H_\varepsilon(z;\theta)$ are related to those of the (Rosenbrock) system matrix

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

LEMMA 1: *Suppose Assumptions 1 and 2 hold. Then* $\operatorname{rank}\mathcal{P}(z;\theta) = n_X + \operatorname{rank} H_\varepsilon(z;\theta)$ *for any* $\theta \in \Theta$ *and for every* $z \in \mathbb{C} \setminus \mathbb{A}(\theta)$, *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 sample of length T generated by (1a) and (1b) with $\theta = \theta_0$. With T infinitely large, under what conditions would it be possible to uncover the value θ_0 and the model that generated $\{Y_t\}$? We start by discussing 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 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 $j \geq 0$.

DEFINITION 2: The DSGE model is locally identifiable from the autocovariances 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 is that the matrix of derivatives with respect to θ of $\Gamma_{Y,\infty}(\theta) \equiv \lim_{T\to\infty}\Gamma_{Y,T}(\theta)$, where $\Gamma_{Y,T}(\theta) \equiv (\Gamma_Y(0;\theta)\ \Gamma_Y(1;\theta)\ \cdots\ \Gamma_Y(T;\theta))$, has full column rank when evaluated at $\theta = \theta_0$. The autocovariance matrix at each j is

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

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

where $\Gamma_X(0;\theta) = E(X_t X_t')$ solves $\Gamma_X(0;\theta) = A(\theta)\Gamma_X(0;\theta)A(\theta)' + B(\theta) \times \Sigma_\varepsilon(\theta)B(\theta)'$. Clearly, each $\Gamma_Y(j;\theta)$ can only be approximated as a truncated sum of the Markov parameters or $\Gamma_X(0;\theta)$ has 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 $\Gamma_Y(j;\theta)$ and $\Omega_Y(z;\theta)$ are 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. SINGULAR CASE $n_\varepsilon \leq n_Y$

Stochastic singularity pertains to the situation when there are fewer stochastic disturbances than observables. For example, stochastic growth models are singular because output, consumption, and investment are all driven by a single shock. While cointegrated systems have been widely analyzed, evidence for stochastic singularity at business cycle frequencies is unusual. In practice, measurement and specification errors are added to DSGE models prior to estimation or endogenous variables are dropped from the analysis to restore nonsingularity. Nonetheless, if a singular system is not identifiable, it is not clear whether a nonsingular system with measurement errors or few observables can be identified. It is thus of interest to begin by studying singular systems, even if one is eventually interested in estimating a nonsingular one.

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

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

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

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

Our point of departure is that associated with each $\theta \in \Theta$ is the transfer function $H_\varepsilon(z;\Lambda^{\mathrm{S}}(\theta))$ and the covariance matrix $\Sigma_\varepsilon(\theta)$. 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

$$\begin{aligned} (4)\qquad & H_\varepsilon(z;\Lambda^{\mathrm{S}}(\theta_0))\Sigma_\varepsilon(\theta_0)H_\varepsilon(z^{-1};\Lambda^{\mathrm{S}}(\theta_0))' \\ & = H_\varepsilon(z;\Lambda^{\mathrm{S}}(\theta_1))\Sigma_\varepsilon(\theta_1)H_\varepsilon(z^{-1};\Lambda^{\mathrm{S}}(\theta_1))'. \end{aligned}$$

Equivalent spectral densities can arise because (i) for given $\Sigma_\varepsilon(\theta)$, each transfer function $H_\varepsilon(z;\Lambda^S(\theta))$ can potentially be obtained from a multitude of quadruples $(A(\theta),B(\theta),C(\theta),D(\theta))$ or (ii) there can be many pairs $(H_\varepsilon(z;\Lambda^S(\theta)),\Sigma_\varepsilon(\theta))$ 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 autocovariances. To make their implications on the canonical model precise, the following assumptions are required.

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

ASSUMPTION 5-S: *For every* $\theta\in\Theta$, (i) *the matrix* $\mathcal{C}\equiv(B(\theta)\ \ A(\theta)B(\theta)\ \cdots\ A^{n_X-1}(\theta)B(\theta))$ *has full row rank and* (ii) *the matrix* $\mathcal{O}\equiv(C(\theta)'\ \ A(\theta)'C(\theta)'\ \cdots\ A^{n_X-1}(\theta)'C(\theta)')'$ *has full column rank*.

The transfer functions of singular systems are left-invertible if and only if $\operatorname{rank}H_\varepsilon(z;\theta)=n_\varepsilon$ in $|z|>1$. Lemma 1 implies that $H_\varepsilon(z;\theta)$ is left-invertible if and only if

$$\operatorname{rank}\mathcal{P}(z;\theta)=n_X+n_\varepsilon\quad\text{for all }|z|>1.$$

Thus, Assumption 4-S ensures left-invertibility in systems with $n_\varepsilon\le n_Y$. It generalizes the conditions of Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007), who proposed an eigenvalue test for left-invertibility when the system is square ($n_\varepsilon=n_Y$) and $D(\theta)$ is invertible. Combined with Assumption 1, Assumption 4-S also ensures that the rank of the spectral density is n_ε almost everywhere (a.e.) in $\mathbb{C}$. That the ranks of $\Omega_Y(z;\theta)$ and $H_\varepsilon(z;\theta)$ coincide is instrumental in the analysis that will follow.

Assumption 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 in the desired final state. Observability means that we can always reconstruct the initial state by observing the evolution of the output, given the evolution of the input.[6] Minimality is similar to relative coprimeness and in VARMA terminology, n_X is the McMillan degree. Hannan (1971) imposed coprimeness in his seminal work on identification of dynamic simultaneous equations. 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

[6]See, for example, 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.

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. This will be illustrated in the example considered in Section 7.

Assumptions 4-S and 5-S are key to obtaining a complete characterization of observational equivalence in singular models. This is achieved as follows. We first use the fact that $H_\varepsilon(z;\theta_0) = H_\varepsilon(z;\theta_1)$ 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 this transformation is sufficient for transfer functions to be equivalent is obvious. That it is 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)). In that literature when the shocks are observed, using minimality to identify θ from the system matrices $ABCD$ is referred to as *structural identification*; see, for example, Hannan and Diestler (1988, Chapter 2.7). However, in DSGE models ε_t are latent with unobserved variances. Thus, minimality is necessary but not sufficient to tie down the class of equivalent systems. To do so, we also need to take into account that $H_\varepsilon(z;\theta)$ can interact with $\Sigma_\varepsilon(\theta)$ to give equivalent spectral densities.

To characterize such interactions, let $\overline{H}_\varepsilon(z;\theta) \equiv H_\varepsilon(z;\theta)L_\varepsilon(\theta)$. Since $\Omega_Y(z;\theta) = \overline{H}_\varepsilon(z;\theta)\overline{H}_\varepsilon(z^{-1};\theta)'$, it is not hard to see that $\widetilde{H}_\varepsilon(z;\theta) = \overline{H}_\varepsilon(z;\theta)V(z)$ will yield the same spectral density for any $V(z)$ satisfying $V(z)V(z^{-1})' = \mathrm{I}_{n_\varepsilon}$, even if $\Sigma_\varepsilon(\theta)$ is an identity matrix. Thus, equivalent pairs $(H_\varepsilon(z;\theta), \Sigma_\varepsilon(\theta))$ are related 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.

Assumption 4-S now comes into play as it will be used to narrow down the equivalent pairs $(H_\varepsilon(z;\theta), \Sigma_\varepsilon(\theta))$. 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 $\Omega_Y(z;\theta) = W_\varepsilon(z;\theta)W_\varepsilon(z^{-1};\theta)'$, where $W_\varepsilon(z;\theta)$ is known as the spectral factor. Much less known is that if $W_\varepsilon(z;\theta)$ and $\widetilde{W}_\varepsilon(z;\theta)$ are both spectral factors that are also left-invertible, then necessarily $\widetilde{W}_\varepsilon(z;\theta) = W_\varepsilon(z;\theta)V$ with $VV' = \mathrm{I}_{n_\varepsilon}$ (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 $W_\varepsilon(z;\theta) = H_\varepsilon(z;\theta)L_\varepsilon(\theta)$. It is

left-invertible if and only if $H_\varepsilon(z;\theta)$ is left-invertible which holds by Assumption 4-S. In such a case, two equivalent pairs $(H_\varepsilon(z;\theta), \Sigma_\varepsilon(\theta))$ must be related by a full rank matrix $U = L_\varepsilon(\theta_0)VL_\varepsilon(\theta_1)^{-1}$ such that

$$(6)\qquad \begin{aligned} &H_\varepsilon(z;\theta_1) = H_\varepsilon(z;\theta_0)U \quad \text{and} \\ &U\Sigma_\varepsilon(\theta_1)U' = \Sigma_\varepsilon(\theta_0) \quad \text{for every } z \in \mathbb{C}. \end{aligned}$$

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

PROPOSITION 1-S—Observational Equivalence $n_\varepsilon \le n_Y$: *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 $n_\varepsilon \times n_\varepsilon$ matrix U such that*

$$(7)\qquad \begin{aligned} &A(\theta_1) = TA(\theta_0)T^{-1}, \quad B(\theta_1) = TB(\theta_0)U, \quad C(\theta_1) = C(\theta_0)T^{-1}, \\ &D(\theta_1) = D(\theta_0)U, \quad \Sigma_\varepsilon(\theta_1) = U^{-1}\Sigma_\varepsilon(\theta_0)U^{-1'}. \end{aligned}$$

Proposition 1-S (proved in the Appendix) says that in singular systems with $n_\varepsilon \le n_Y$, there can exist no 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 $\Lambda^{\mathrm{S}}(\theta)$ may not be identifiable. In other words, the *ABCD* representation is not a reduced form in the sense of classical simultaneous equations analysis. 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 $\delta^{\mathrm{S}} : \Theta \times \mathbb{R}^{n_X^2} \times \mathbb{R}^{n_\varepsilon^2} \to \mathbb{R}^{n_\Lambda^{\mathrm{S}}}$ as

$$(8)\qquad \delta^{\mathrm{S}}(\theta, T, U) \equiv \begin{pmatrix} \mathrm{vec}(TA(\theta)T^{-1}) \\ \mathrm{vec}(TB(\theta)U) \\ \mathrm{vec}(C(\theta)T^{-1}) \\ \mathrm{vec}(D(\theta)U) \\ \mathrm{vech}(U^{-1}\Sigma_\varepsilon(\theta)U^{-1'}) \end{pmatrix}.$$

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

LEMMA 2-S: *Under the assumptions of Proposition 1-S, θ is locally identifiable from the autocovariances of Y_t at a point $\theta_0 \in \Theta$ if and only if the system of equations $\delta^{\mathrm{S}}(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon}) = \delta^{\mathrm{S}}(\theta_1, T, U)$ has a locally unique solution $(\theta_1, T, U) = (\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$.*

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 $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$. The matrix of partial derivatives of $\delta^{\mathrm{S}}(\theta, T, U)$ evaluated at $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$ is given by[7]

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

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 $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$. 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 $\operatorname{rank} \Delta_\Lambda^{\mathrm{S}}(\theta_0) = n_\theta$, the mapping is locally invertible at θ_0. The middle $n_\Lambda^{\mathrm{S}} \times n_X^2$ matrix $\Delta_T^{\mathrm{S}}(\theta_0)$ corresponds to the partial derivatives with respect to T evaluated at $(T, U) = (\mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$. When $\operatorname{rank} \Delta_T^{\mathrm{S}}(\theta_0) = n_X^2$, then the only (local) similarity transformation is provided by the identity matrix. The final $n_\Lambda^{\mathrm{S}} \times n_\varepsilon^2$ matrix $\Delta_U^{\mathrm{S}}(\theta_0)$ corresponds to the partial derivatives with respect to U, evaluated at $(T, U) = (\mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$. When $\operatorname{rank} \Delta_U^{\mathrm{S}}(\theta_0) = n_\varepsilon^2$, then the spectral factorization (locally) uniquely determines the pair $(H_\varepsilon(z; \theta), \Sigma_\varepsilon(\theta))$. Note that since $\Lambda^{\mathrm{S}}(\theta_0)$ may not be identifiable, full column rank of $\Delta_\Lambda^{\mathrm{S}}(\theta_0)$ alone is necessary but not sufficient for identification.

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

$$\operatorname{rank} \Delta^{\mathrm{S}}(\theta_0) = \operatorname{rank} \left(\Delta_\Lambda^{\mathrm{S}}(\theta_0) \quad \Delta_T^{\mathrm{S}}(\theta_0) \quad \Delta_U^{\mathrm{S}}(\theta_0) \right) = n_\theta + n_X^2 + n_\varepsilon^2.$$

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

[7]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 $\operatorname{vec}(A) = \mathcal{G}_n \operatorname{vech}(A)$.

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 $\Sigma_\varepsilon(\theta)$. The results allow researchers to establish identifiability of DSGE models prior to estimation and independently of the estimator used. 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

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

The condition depends on n_Y, n_X, and n_ε representing the structure of the economic model. They play the role of the number of endogenous and predetermined variables in classical simultaneous equations analysis; the sample size T is not involved. The term $n_Y - n_\varepsilon \geq 0$ reflects stochastic singularity and is, however, specific to DSGE models.

An important implication of the proposition is that identifiability of θ can be studied from the second moments of the data, even though the hyperparameters $\Lambda^{\mathrm{S}}(\theta)$ may not be identifiable. The constant rank requirement in Proposition 2-S ensures that $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$ is a regular point. Our rank condition is still sufficient for identification, even if θ_0 fails to be a regular point.[8] Both rank and order conditions have a classical flavor even though we work with assumptions that would not be valid in a classical setup.

The submatrices

$$\Delta^{\mathrm{S}}_{\Lambda T}(\theta_0) = (\,\Delta^{\mathrm{S}}_{\Lambda}(\theta_0) \quad \Delta^{\mathrm{S}}_{T}(\theta_0)\,) \quad \text{and} \quad \Delta^{\mathrm{S}}_{\Lambda U}(\theta_0) = (\,\Delta^{\mathrm{S}}_{\Lambda}(\theta_0) \quad \Delta^{\mathrm{S}}_{U}(\theta_0)\,)$$

reveal delicate types of nonidentification due to the features of the impulse and propagating mechanism of the model. It is possible that $\Delta^{\mathrm{S}}_{\Lambda}(\theta_0)$ is full rank and yet $\Delta^{\mathrm{S}}(\theta_0)$ is rank deficient. As shown in Section 6, the null space of $\Delta^{\mathrm{S}}(\theta_0)$ can be used to isolate the parameters that are not identifiable. Such information about the model is useful even if estimation is not of interest.

Minimality and left-invertibility are maintained assumptions of Proposition 2-S. However, the rank conditions on the submatrices of $\Delta^{\mathrm{S}}(\theta_0)$ remain necessary for identification even if one of these assumptions fail. In particular, (i) $\operatorname{rank}\Delta^{\mathrm{S}}_{\Lambda U}(\theta_0) = n_\theta + n_\varepsilon^2$ is necessary for θ_0 to be identifiable when the system fails to be minimal but the transfer function is left-invertible, while (ii) $\operatorname{rank}\Delta^{\mathrm{S}}_{\Lambda T}(\theta_0) = n_\theta + n_X^2$ is necessary for identification of minimal systems

[8]When the elements of $\Delta^{\mathrm{S}}(\theta)$ are analytic functions, then almost every point θ in Θ is regular (see, e.g., Corollary 5.A.1. in Fisher (1966)).

with transfer functions that are not left-invertible. To see why, consider case (i). Since 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 $\Sigma_\varepsilon(\theta_1) = U^{-1}\Sigma_\varepsilon(\theta_0)U^{-1'}$ always leads to observational equivalence whether or not the system is not minimal, local injectivity of $\delta^S(\theta, \mathrm{I}_{n_X}, U)$ at $(\theta, U) = (\theta_0, \mathrm{I}_{n_\varepsilon})$ is still necessary for local identification of θ_0. Even if both minimality and left-invertibility fail to hold, $\operatorname{rank} \Delta^S_\Lambda(\theta_0) = n_\theta$ is still a useful diagnostic and is still a necessary condition for identification.

4. NONSINGULAR CASE $n_\varepsilon \geq n_Y$

A nonsingular system occurs when there are at least as many shocks as variables. When $n_\varepsilon > n_Y$, (2) is no longer the Wold representation for Y_t and ε_t is no longer fundamental. More precisely, the transfer function of a nonsingular system with more shocks than observables is *not* left-invertible. As the foregoing results for singular systems are now invalid, a different framework is necessary. We reparameterize the state space solution in (1a)–(1b) to obtain its innovations representation, details of which can be found in Anderson and Moore (1979) and Hansen and Sargent (2005).

The innovations representation consists of a system of measurement and transition equations

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

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

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

$$\Lambda^{\mathrm{NS}}(\theta) \equiv \left((\operatorname{vec} A(\theta))', (\operatorname{vec} K(\theta))', (\operatorname{vec} C(\theta))', (\operatorname{vech} \Sigma_a(\theta))'\right)',$$

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

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

ASSUMPTION 4-NS: *For every $\theta \in \Theta$, $D(\theta)\Sigma_\varepsilon(\theta)D(\theta)'$ is nonsingular.*

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

The validity of the innovations representation hinges on the existence of a positive semidefinite solution to the so-called discrete algebraic Ricatti equation (DARE). In the literature, the existence of this solution is often assumed. To proceed with identification analysis, we need to make the primitive conditions for this result precise. Specifically, $D(\theta)$ must be full row rank (Assumption 4-NS) so that each of the innovations affects at least one series. This 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 $\Lambda^{\mathrm{S}}(\theta)$, and their expressions are given in Equations (19) and (18) in the Appendix. Importantly, Y_t now has Wold representation

$$Y_t = \sum_{j=0}^{\infty} h_a(j;\theta)a_{t-j} = H_a(L^{-1};\theta)a_t,$$

in which the innovations a_t are fundamental and the covariance matrix $\Sigma_a(\theta)$ is nonsingular for every $\theta \in \Theta$. Furthermore, the corresponding transfer function

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

is such that $H_a(z;\theta)$ is square and invertible in $|z| > 1$ by construction. In other words, the innovations representation delivers left-invertibility automatically. This is unlike in singular systems where left-invertibility needs to be assumed. Assumption 5-NS ensures that the system (9a)–(9b) is minimal. Written in terms of $H_a(z;\theta)$ and $\Sigma_a(\theta)$, the spectral density matrix of Y_t is

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

The following result characterizes observational equivalence in nonsingular models.

PROPOSITION 1-NS—Observational Equivalence $n_\varepsilon \geq n_Y$: *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*

$$\begin{aligned} &A(\theta_1) = TA(\theta_0)T^{-1}, \quad K(\theta_1) = TK(\theta_0), \\ &C(\theta_1) = C(\theta_0)T^{-1}, \quad \Sigma_a(\theta_1) = \Sigma_a(\theta_0). \end{aligned} \tag{10}$$

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 yields (10). The restrictions no longer involve U because the innovations representation imposes the normalization $h_a(0;\theta) = \mathrm{I}_{n_Y}$. Instead, the

restrictions are subsumed in $K(\theta)$, which needs to be solved numerically, because the forecast errors a_t are derived from Kalman filtering. No filtering is necessary to study identification of left-invertible singular models.

For any $\theta \in \Theta$ and any full rank $n_X \times n_X$ matrix T, let

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

The mapping $\delta^{\mathrm{NS}} : \Theta \times \mathbb{R}^{n_X^2} \to \mathbb{R}^{n_\Lambda^{\mathrm{NS}}}$ 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 autocovariances of Y_t at a point $\theta_0 \in \Theta$ if and only if the system of equations $\delta^{\mathrm{NS}}(\theta_0, \mathrm{I}_{n_X}) = \delta^{\mathrm{NS}}(\theta_1, T)$ has a locally unique solution $(\theta_1, T) = (\theta_0, \mathrm{I}_{n_X})$. The proof uses arguments analogous to those for Lemma 2-S. The matrix of partial derivatives of $\delta^{\mathrm{NS}}(\theta, T)$ evaluated at $(\theta_0, \mathrm{I}_{n_X})$ is

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

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

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

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

The order condition simplifies to $n_\theta \leq 2n_X n_Y + \frac{n_Y(n_Y+1)}{2}$. Unlike in the singular case, the condition does not depend on n_ε because the error a_t in the innovations representation is always of dimension n_Y. As in the singular case,

rank $\Delta_\Lambda^{\mathrm{NS}}(\theta_0) = n_\theta$ remains necessary for identification when the minimality fails. While the rank requirement in Proposition 2-NS appears weaker than in the singular case, the assumptions are stronger since Assumption 4-NS must also hold. Furthermore, both $K(\theta)$ and $\Sigma_a(\theta)$ are highly nonlinear functions of $(A(\theta), B(\theta), C(\theta), D(\theta))$. If one starts with a singular system and the mapping from θ to $\Lambda^{\mathrm{S}}(\theta)$ is invertible, adding measurement errors does not guarantee that the mapping from θ to $\Lambda^{\mathrm{NS}}(\theta)$ will remain invertible because filtering entails information loss. The addition of measurement errors might help restore nonsingularity but could complicate identification.

When the system is square, 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 $n_Y = n_\varepsilon$. This allows $K(\theta)$ to be expressed in terms of $B(\theta)$ and $D(\theta)$. Specifically, $K(\theta) = (B(\theta)\Sigma_\varepsilon(\theta)D(\theta)'(D(\theta)\Sigma_\varepsilon(\theta)D(\theta)')^{-1}$. 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 $\Delta^{\mathrm{NS}}(\theta_0) = n_\theta + n_X^2$ of Proposition 2-NS is equivalent to the condition rank $\Delta^{\mathrm{S}}(\theta_0) = n_\theta + n_X^2 + n_Y^2$ of Proposition 2-S, as it should be.

While left-invertible square systems can be analyzed by either Proposition 2-S or 2-NS, square systems that are not invertible (and hence fail to satisfy Assumption 4-S) can only be analyzed using the innovations representation. This is possible because the transfer function of the innovations model is always left-invertible. For example, the model considered Leeper, Walker, and Yang (2008) is square, but the presence of anticipated shocks violates invertibility. In such a case, identification can only be analyzed by Proposition 2-NS.

Our main identification results can be summarized as follows.

	Singular Systems	Square Systems $n_\varepsilon = n_Y$		Nonsingular Systems
	$n_\varepsilon < n_Y$	Left-Invertible	Noninvertible	$n_\varepsilon > n_Y$
Representation	*ABCD*	*ABCD*	*AKC*	*AKC*
Obs. equivalence	1-S	1-S	1-NS	1-NS
Rank and order	2-S	2-S	2-NS	2-NS

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

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

For example, steady state restrictions $EY_t^a = EY_{ss}^a(\theta)$ can be written as $\varphi(\theta_0) = 0$. Long-run restrictions can be imposed on $H_\varepsilon(1;\theta)$.

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

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

In the nonsingular case, let $\overline{\delta}^{\mathrm{NS}}(\theta, T) \equiv \binom{\varphi(\theta)}{\delta^{\mathrm{NS}}(\theta,T)}$ with δ^{NS} as given in (11) and

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

Local injectivity of $\overline{\delta}^{\mathrm{S}}$ at $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$ or 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 autocovariances 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 $\overline{\Delta}^{\mathrm{S}}(\theta)$ or $\overline{\Delta}^{\mathrm{NS}}(\theta)$. 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*

$$\operatorname{rank} \overline{\Delta}(\theta_0) = n_\theta + n_X^2 + \mathbb{1}_{\mathrm{S}} \cdot n_\varepsilon^2,$$

where $\mathbb{1}_{\mathrm{S}} = 1$ if the model is singular and $= 0$ otherwise. When $\varphi(\theta)$ involves the mean, the above restrictions are necessary and sufficient for local identification of θ_0 from the mean and autocovariances 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 $n_\theta + n_X^2 + n_\varepsilon^2 \leq n_\Lambda^{\mathrm{S}} + n_R$; in the nonsingular case, we need $n_\theta + n_X^2 \leq n_\Lambda^{\mathrm{NS}} + n_R$. 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(1) shocks are often specified, but this entails many restrictions on the contemporaneous and past correlations among 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 *partial identification* of a particular subvector θ_i, 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*

$$\operatorname{rank} \Delta(\theta_0) = \operatorname{rank}\left(\frac{\partial \delta(\theta_0)}{\partial \theta_{-i}}\right) + (n_{\theta,i} + n_X^2 + \mathbb{1}_{\mathrm{S}} \cdot n_\varepsilon^2),$$

where $(\Delta(\theta), \frac{\partial \delta(\theta)}{\partial \theta_{-i}}) = (\Delta^{\mathrm{S}}(\theta), \frac{\partial \delta^{\mathrm{S}}(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})}{\partial \theta_{-i}})$ *with* $\mathbb{1}_{\mathrm{S}} = 1$ *if the model is singular and* $(\Delta(\theta), \frac{\partial \delta(\theta)}{\partial \theta_{-i}}) = (\Delta^{\mathrm{NS}}(\theta), \frac{\partial \delta^{\mathrm{NS}}(\theta, \mathrm{I}_{n_X})}{\partial \theta_{-i}})$ *with* $\mathbb{1}_{\mathrm{S}} = 0$ *in the nonsingular case.*

The results generalize Propositions 2-S and 2-NS to allow θ_{-i} to be nonempty. 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. Straightforward modifications of the rank conditions in Propositions 3 and 4 remain necessary when minimality and/or left-invertibility fails. In such cases, appropriate columns should be removed from $\overline{\Delta}(\theta)$ and $\Delta(\theta)$, and the rank requirements adjusted accordingly.

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) used it to numerically evaluate the T analytical autocovariances, while Qu and Tkachenko (2010) used it to evaluate the discretized spectral density. Our approach is unique in that we do not perform numerical calculations of any second moments, yet we 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 $n_Y \leq n_\varepsilon$, the focus of this section is on the nonsingular case only. It is understood that $\Lambda(\theta)$ and $\Delta(\theta)$ refer to $\Lambda^{\mathrm{NS}}(\theta)$ and $\Delta^{\mathrm{NS}}(\theta)$, respectively.

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

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

are the Markov parameters in the impulse response to mutually uncorrelated unit shocks.

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

(i) *For every* $J \geq 0$, $\frac{\partial \operatorname{vec} \overline{h}_J(\theta)}{\partial \Lambda} \cdot \Delta_T(\theta) = 0_{(J+1)n_Y^2 \times n_X^2}$.

(ii) *If* $J \geq 2n_X - 2$, *then* $\operatorname{rank} \overline{h}_J(\theta) = \operatorname{rank} \overline{h}_{2n_X-2}(\theta)$.

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

Result (i), which follows directly from Proposition 1-NS, can be used to determine the source of nonidentification. Since $\Delta(\theta) = (\Delta_\Lambda(\theta) \;\; \Delta_T(\theta))$, a direct consequence of (i) is

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

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

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

In particular, the first n_θ components of v form a subvector 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 nonidentification: (a) If the rank test on $\Delta(\theta_0)$ fails, compute its null space. (b) For each v in the null space, find the position of its nonzero entries in the first n_θ rows. These positions correspond to the components of θ that are not identifiable without restrictions. A similar procedure based on the null space of $\Delta^{\mathrm{S}}(\theta_0)$ can be used in the singular case.

Result (ii) says that all the information about the system is contained in the first $2n_X - 1$ Markov parameters. This follows from the fact that in a minimal system of order n_X, the first $2n_X - 1$ Markov parameters uniquely determine

all the remaining Markov parameters of order $j \geq 2n_X - 2$. As $\operatorname{rank} \Delta_{\overline{h}_J}(\theta) = \operatorname{rank} \Delta_{\overline{h}_{2n_X-2}}(\theta)$ for any $J \geq 2n_X - 2$, 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 nonredundant 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 θ_0 can be locally identified from the first $2n_X - 1$ Markov parameters $\overline{h}_{2n_X-2}(\theta)$. 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 an identifiable parameter of the model, assume its true value ξ_0 is regular, and denote by $I(\xi_0)$ 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_0 is identifiable, $I(\xi_0)$ is nonsingular. The rank of the information matrix is given by $\operatorname{rank} I(\theta_0) = \operatorname{rank} \frac{\partial \xi}{\partial \theta}|_{\theta=\theta_0}$. Obviously, $I(\theta_0)$ is full rank n_θ if and only if $\operatorname{rank} \frac{\partial \xi}{\partial \theta}|_{\theta=\theta_0} = n_\theta$.[9]

For the above information matrix decomposition to be useful, we need to find an identifiable parameter ξ. By Proposition 1-NS, all observationally equivalent combinations of $(A(\theta), B(\theta), K(\theta), \Sigma_a(\theta))$ give equivalent Markov parameters $\overline{h}_J(\theta)$ for $J \geq 0$. Thus, $\overline{h}_J(\theta)$ are identified from the autocovariances of Y_t. Hence, if we define ξ_J to be $\operatorname{vec} \overline{h}_J(\theta)$, then $I(\xi_J)$ is full rank. It follows that $I(\theta_0)$ is nonsingular if and only if there exists a $\overline{J} \geq 0$ such that $\operatorname{rank} \frac{\partial \xi_{\overline{J}}}{\partial \theta}|_{\theta=\theta_0} = n_\theta$. Using result (ii), a necessary and sufficient condition for $I(\theta_0)$ to be nonsingular is that $\operatorname{rank} \frac{\partial \xi_{2n_X-2}}{\partial \theta}|_{\theta=\theta_0} = n_\theta$. Combining this with result (iii) then shows that $I(\theta_0)$ is nonsingular if and only if $\operatorname{rank} \Delta(\theta_0) = n_\theta + n_X^2$. Thus, the information matrix is full rank if and only if our rank condition is satisfied. While the likelihood analysis typically assumes ε_t to be i.i.d. Gaussian, we only 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 $\xi_{2n_X-2} = \operatorname{vec} \overline{h}_{2n_X-2}(\theta)$ is $n_\xi = (2n_X - 1)n_Y^2$ and the order condition based on the information matrix is $n_\theta \leq (2n_X - 1)n_Y^2$. Our order condition stated in Proposition 2-NS only requires $n_\theta \leq 2n_X n_Y + n_Y(n_Y + 1)/2$. 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.

[9]In Iskrev (2008) and Bonaldi (2010), $\xi = \operatorname{vec}(A, B, C, D)$. The (time domain) information matrix of the reduced form ($ABCD$) model $I(\xi)$ is not necessarily full rank. Hence, the decomposition is not useful for identification analysis.

7. EXAMPLE

The results just presented are now used to analyze the model in An and Schorfheide (2007). The log-linearized model is given by

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

with $\varepsilon_{rt} \sim \mathrm{WN}(0, \sigma_r^2)$, $\varepsilon_{gt} \sim \mathrm{WN}(0, \sigma_g^2)$, and $\varepsilon_{zt} \sim \mathrm{WN}(0, \sigma_z^2)$ mutually uncorrelated. In the above model, $\overline{\pi}$ is the 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 last column (titled DGP) in Table 3 of An and Schorfheide (2007); the second only modifies the values of ρ_g and ρ_z so that the corresponding processes are considerably less persistent.

Let $\widetilde{X}_t \equiv (z_t, g_t, r_t, y_t, \pi_t, c_t, E_t(\pi_{t+1}), E_t(y_{t+1}))'$ and $\varepsilon_t \equiv (\varepsilon_{zt}, \varepsilon_{gt}, \varepsilon_{rt})'$. We report results for $Y_t = (r_t, y_t, \pi_t)'$. Additional results with c_t included in Y_t are reported in the Supplemental Material. Notably, adding the consumption c_t to the observables does not change the results below. This is not surprising because c_t is defined by an identity and adds no information to identification.

We use the GENSYS algorithm of Sims (2002) to solve the model. The solution evaluated at both parameter sets is determinate. However, as shown in the Supplementary Material, 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, $\operatorname{rank} \mathcal{C} = 3 < n_X = 8$ and $\operatorname{rank} \mathcal{O} = 3 < n_X = 8$, where $\mathcal{C}$ and $\mathcal{O}$ are the controllability and observability matrices in Assumption 5-S.

To be able to apply Proposition 2-S, 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 states. For example, the MINREAL function in Matlab produces the minimal state vectors from an eigenvector analysis of the nonminimal state variables. However, the eigenvectors depend on the parameters to be identified, and for our purpose, MINREAL cannot be used to find the minimal state.

Nonetheless, the problem of finding the minimal state vector is made simple by the fact that DSGE models are based on microfoundations. As such, we know which are the exogenous and endogenous state variables, and these con-

stitute the minimal state vector. In practice, the problem reduces to arranging the solution produced by numerical algorithms to be of the particular form

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

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

In other words, the key to finding a minimal representation is to rearrange the solution such that the $\widetilde{A}$ matrix has columns of zeros. This may require ordering the variables prior to solving the model or simply reordering the solution itself. Once this is accomplished, X_{1t} is immediately the minimal state vector.[10] The required $ABCD$ matrices are then defined by simple substitutions:

$$X_{1,t+1} = \underbrace{\widetilde{A}_1(\theta)}_{A(\theta)} X_{1t} + \underbrace{\widetilde{B}_1(\theta)}_{B(\theta)} \varepsilon_{t+1},$$

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

In the An–Schorfheide example, $X_{1t} \equiv (z_t, g_t, r_t)'$ and the new system has $n_X = 3$. The minimal state space form with $Y_{t+1} = (r_{t+1}, y_{t+1}, \pi_{t+1})$ is presented in Table I. As $n_\varepsilon = 3 = n_Y$, the model is square and both Propositions 2-S and 2-NS can be used to study identification. Table I shows that some of the $n_\theta = 13$ parameters of this model are not identified. An analysis of the null space of $\Delta^{\mathrm{S}}(\theta_0)$ quickly reveals that the columns corresponding to ν, ϕ, and $\overline{\pi}$ are not linearly independent. As already pointed out in An and Schorfheide (2007), these three parameters are not separately identified and rank $\Delta^{\mathrm{S}}_A(\theta_0)$ should indeed be less than $n_\theta = 13$. While nonidentifiability 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 nonzero eigenvalues. Since the magnitude of the eigenvalues may depend on the units of measurement, Anderson (1984) suggested considering the ratio of the sum of the smallest k eigenvalues to the average of all eigenvalues; see also Canova and Sala (2009). However, the choice of k remains arbitrary. Rank routines in numerical packages use a cutoff to determine whether the eigenvalues are suf-

[10]For the model of Smets and Wouters (2007), one more step is necessary because of the coexistence of a flexible and a fixed price equilibrium. This is shown in the Supplemental Material.

TABLE I

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	0.2	0.6	0.3

Minimal State Space Representation

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

$$Y_{t+1} = \begin{pmatrix} r_{t+1} \\ y_{t+1} \\ \pi_{t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 0.5450 & 0 & 0.5143 \\ 1.3377 & 0.95 & -0.8258 \\ 1.3418 & 0 & -0.5596 \end{pmatrix}}_{C(\theta)} X_t + \underbrace{\begin{pmatrix} 0.6055 & 0 & 0.6858 \\ 1.4863 & 1 & -1.1011 \\ 1.4909 & 0 & -0.7462 \end{pmatrix}}_{D(\theta)} \varepsilon_{t+1}$$

	Unrestricted Model					Model With Restrictions: Tol = 1e−3								
Tol	$\Delta^{\mathrm{S}}_{\Lambda}$	Δ^{S}_{T}	Δ^{S}_{U}	Δ^{S}	Pass	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
e−02	11	9	9	28	No	ν	–	–	12	9	9	21	20	29	No
e−03	11	9	9	28	No	ν	ϕ	–	13	9	9	22	21	30	No
e−04	11	9	9	28	No	ϕ	$\overline{\pi}$	–	13	9	9	22	21	30	No
e−05	11	9	9	28	No	ν	$\overline{\pi}$	–	13	9	9	22	21	29	No
e−06	11	9	9	28	No	β	ϕ	–	12	9	9	21	20	29	No
e−07	11	9	9	29	No	ϕ	ρ_g	–	12	9	9	21	20	29	No
e−08	11	9	9	29	No	β	ν	ϕ	13	9	9	22	21	30	No
e−09	11	9	9	29	No	β	ψ_1	ψ_2	11	9	9	20	20	29	No
e−10	12	9	9	29	No	ν	ϕ	ψ_1	13	9	9	22	22	31	Yes
e−11	12	9	9	29	No	ν	ϕ	ψ_2	13	9	9	22	22	31	Yes
Default	12	9	9	30	No	τ	ψ_1	ψ_2	11	9	9	20	20	29	No
Required	13	9	9	31					13	9	9	22	22	31	

ficiently small. Matlab uses tolerance $\mathrm{Tol} = \max(\mathrm{size}(M))\,\mathrm{EPS}(\|M\|)$, where EPS is the float point precision of M. This default tolerance does not take into account that the $\Delta^{\mathrm{S}}(\theta_0)$ 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 default (last row). Clearly, the rank of $\Delta^{\mathrm{S}}_{\Lambda}(\theta_0)$ varies with Tol. As shown in the Supplementary Material, if Tol is set to default, the rank of $\Delta^{\mathrm{S}}_{\Lambda}(\theta_0)$ in a non-minimal version of the model is 13, suggesting identification may be possible 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 grounds that the numerical derivatives are computed using a step size of 1e−3. Furthermore, the rank of $\Delta^{\mathrm{S}}_{\Lambda}(\theta_0)$ 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 to isolate the parameters that are at not well

identified even with infinite data. This flags the parameters that will likely be difficult to identify when only a finite number of observations are available.[11]

The rank of $\Delta^S(\theta_0)$ suggests that three restrictions may be necessary to identify the model. To find out which parameters fail to be identified, the rank of $\overline{\Delta}^S(\theta_0)$ is evaluated for 11 sets of restrictions. Recall that in our analysis, a restriction adds to the rows of $\overline{\Delta}^S(\theta_0)$ but leaves the number of columns unchanged. Thus, the rank required for identification is always $n_\theta + n_X^2 + n_\varepsilon^2 = 31$. It is quickly found that ν, ϕ, and either ψ_1 or ψ_2 need to be restricted. Fixing ν and ϕ always leads to full rank of $\overline{\Delta}^S_\Lambda(\theta_0)$, 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 $\overline{\Delta}^S_{\Lambda U}(\theta_0)$. The example illustrates that individual full column ranks of the matrices $\overline{\Delta}^S_\Lambda(\theta_0)$, $\overline{\Delta}^S_T(\theta_0)$, and $\overline{\Delta}^S_U(\theta_0)$ are necessary but not sufficient for identification.

Now consider a reparameterized model with $\kappa = \frac{\tau(1-\nu)}{\nu\overline{\pi}\phi}$ and $n_\theta = 11$. $\Delta^S_\Lambda(\theta_0)$, $\Delta^S_T(\theta_0)$, and $\Delta^S_U(\theta_0)$ of the reparameterized model are individually full rank but $\Delta^S(\theta_0)$ is short rank by one. This is in agreement with Table I, which shows that fixing two of ν, ϕ, and $\overline{\pi}$ is not enough for identification. Table II shows that one of the three parameters in the Taylor rule (ψ_1, ψ_2, ρ_r) needs to be further restricted. This result, which was overlooked in An and Schorfheide (2007), supports the argument of Cochrane (2011) that parameters of the Taylor rule are not identified without restrictions.

Instead of a model with highly persistent shocks in which $(\rho_g, \rho_z) = (0.95, 0.9)$, Table III takes these parameters to be equal to $(0.15, 0.1)$ while keeping the remaining components of θ_0 the same as before. At Tol $= 1e{-}3$, the three submatrices of $\overline{\Delta}^S(\theta_0)$ are all full rank. However, $\overline{\Delta}^S(\theta_0)$ always fails to be so. Thus, general statements about nonidentifiability of the Phillips curve cannot be made on the basis of rank tests evaluated at a particular θ_0. Each parameter point has to be assessed on an individual basis. Adding measurement errors to each of the three observed variables leads to the same findings.

It should be emphasized that our identification analysis does not require solving for the autocovariances whose numerical errors could affect the rank conditions. In the singular case, the only component of the $\Delta^S(\theta_0)$ matrix that is subject to numerical errors is $\Delta^S_\Lambda(\theta_0)$. The remaining submatrices can be computed exactly in a few lines of code. In the nonsingular case, the Kalman gain matrix needs to be solved numerically. This too is a simple iterative procedure. The code and the results for larger and more complex models solved using other solution algorithms are available in the Supplemental Material.

[11]The null space of $\Delta^S(\theta_0)$ is also used to find the parameters that are difficult to identify. The null space depends on the rank of $\Delta^S(\theta_0)$ which in turn depends on Tol. We start with a large Tol. The nonzero entries are the positions of the parameters that are most difficult to identify. Then Tol is reduced to find additional problematic parameters. Restrictions on subsets of these parameters are imposed to formally test conditional identification.

TABLE II

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	$100\sigma_r$	$100\sigma_g$	$100\sigma_z$	$100\sigma_{vr}$	$100\sigma_{vy}$	$100\sigma_{v\pi}$
2	0.9975	0.33	1.5	0.125	0.75	0.95	0.9	0.2	0.6	0.3	0.2	0.2	0.2

Tol = 1e−3	No Measurement Errors, $n_\theta = 11$							With Errors, $n_\theta = 14$			
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	$\overline{\Delta}^{\mathrm{NS}}_{\Lambda}$	$\overline{\Delta}^{\mathrm{NS}}_{T}$	$\overline{\Delta}^{\mathrm{NS}}$	Pass
τ	11	9	9	20	19	28	No	13	9	22	No
β	11	9	9	20	19	28	No	13	9	22	No
κ	11	9	9	20	19	28	No	13	9	22	No
ψ_1	11	9	9	20	20	29	Yes	14	9	23	Yes
ψ_2	11	9	9	20	20	29	Yes	14	9	23	Yes
ρ_r	11	9	9	20	20	29	Yes	14	9	23	Yes
ρ_g	11	9	9	20	19	28	No	13	9	22	No
ρ_z	11	9	9	20	19	28	No	13	9	22	No
σ_r^2	11	9	9	20	19	28	No	13	9	22	No
σ_g^2	11	9	9	20	19	28	No	13	9	22	No
σ_z^2	11	9	9	20	19	28	No	13	9	22	No
Required	11	9	9	20	20	29		14	9	23	

TABLE III

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	$100\sigma_r$	$100\sigma_g$	$100\sigma_z$	$100\sigma_{vr}$	$100\sigma_{vy}$	$100\sigma_{v\pi}$
2	0.9975	0.33	1.5	0.125	0.75	0.15	0.1	0.2	0.6	0.3	0.2	0.2	0.2

Tol = 1e−3	No Measurement Errors, $n_\theta = 11$							With Errors, $n_\theta = 14$			
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	$\overline{\Delta}^{\mathrm{NS}}_{\Lambda}$	$\overline{\Delta}^{\mathrm{NS}}_{T}$	$\overline{\Delta}^{\mathrm{NS}}$	Pass
τ	11	9	9	20	19	27	No	13	9	21	No
β	11	9	9	20	19	27	No	13	9	22	No
κ	11	9	9	20	19	27	No	13	9	22	No
ψ_1	11	9	9	20	20	28	No	14	9	22	No
ψ_2	11	9	9	20	19	28	No	13	9	21	No
ρ_r	11	9	9	20	19	28	No	14	9	22	No
ρ_g	11	9	9	20	19	27	No	13	9	21	No
ρ_z	11	9	9	20	19	27	No	13	9	21	No
σ_r^2	11	9	9	20	19	27	No	13	9	21	No
σ_g^2	11	9	9	20	19	27	No	13	9	21	No
σ_z^2	11	9	9	20	19	27	No	13	9	21	No
Required	11	9	9	20	20	29		14	9	23	

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 the structure of the model, the results also help us uncover features of the model that are not immediately transparent.

The results of this paper are local which leaves open the question of global parameter identification. Given that DSGE models are highly nonlinear in θ, establishing primitive conditions for global identification is not trivial. The model may well be locally identified at every point of the parameter space and yet fail to be globally identified. However, as discussed in Komunjer (2011), there are special cases when adding the condition of "properness" could ensure that everywhere local identification results become global. Whether such conditions are useful in the context of DSGE models and whether higher order approximations can facilitate identification of θ remain to be examined.

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 $\mathbb{C}$ and $\det(z\mathrm{I}_{n_X} - A(\theta)) \neq 0$ for any $z \notin \mathbb{A}(\theta)$. There are two cases to consider.

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

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

Thus, for any $z \in \mathbb{C} \setminus \mathbb{A}(\theta)$,

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

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

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

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

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

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

$$\mathcal{P}(z;\theta)\begin{pmatrix} \mathrm{I}_{n_X} & 0_{n_X \times n_\varepsilon} \\ D^{-1}(\theta)C(\theta) & \mathrm{I}_{n_\varepsilon} \end{pmatrix}$$

$$= \begin{pmatrix} z\mathrm{I}_{n_X} - (A(\theta) - B(\theta)D^{-1}(\theta)C(\theta)) & B(\theta) \\ 0_{n_Y \times n_X} & D(\theta) \end{pmatrix} \quad \forall z \in \mathbb{C}.$$

Since $\det(\mathcal{P}(z;\theta)) = \det(D(\theta))\det(z\mathrm{I}_{n_X} - (A(\theta) - B(\theta)D^{-1}(\theta)C(\theta))$, 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 Fernández-Villaverde et al. (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 nonsquare, is a generalized eigenvalue problem, with z being the generalized eigenvalues (see, e.g., Laub and Moore (1978)). In our case, $M = \begin{bmatrix} \mathrm{I}_{n_X} & 0_{n_X \times n_\varepsilon} \\ 0_{n_Y \times n_X} & 0_{n_Y \times n_\varepsilon} \end{bmatrix}$, $P(\theta) = \begin{bmatrix} A(\theta) & -B(\theta) \\ C(\theta) & -D(\theta) \end{bmatrix}$, and $\mathcal{P}(z;\theta) = zM - P(\theta)$. Thus, the rank test can also be formulated as a generalized eigenvalue test. *Q.E.D.*

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 a $n_Y \times r$ matrix such that for all $z \in \mathbb{C}$, $\Omega_Y(z;\theta) = W(z;\theta)W(z^{-1};\theta)'$ and $\operatorname{rank}W(z;\theta) = r$ 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 $H_\varepsilon(z;\theta)$ is left-invertible in $|z| > 1$. Combing this with Assumption 1 gives $\operatorname{rank}W(z;\theta) = \operatorname{rank}(H_\varepsilon(z;\theta)L_\varepsilon(\theta)) = n_\varepsilon$ for all $|z| > 1$. By Lemma 1 and Assumption 1,

$\Omega_Y(z;\theta) = H_\varepsilon(z;\theta)\Sigma_\varepsilon(\theta)H_\varepsilon(z^{-1};\theta)'$ is of rank n_ε a.e. in $\mathbb{C}$. Hence, $W(z;\theta)$ is 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 $n_\varepsilon \times n_\varepsilon$ matrix V such that

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

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

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

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 $\mathbb{D}(\theta_1) = \mathbb{D}(\theta_0)$, $A(\theta_1) = TA(\theta_0)T^{-1}$, $\mathbb{B}(\theta_1) = T\mathbb{B}(\theta_0)$, and $C(\theta_1) = C(\theta_0)T^{-1}$, that is, $D(\theta_1)L_\varepsilon(\theta_1) = D(\theta_0)L_\varepsilon(\theta_0)V$, $A(\theta_1) = TA(\theta_0)T^{-1}$, and $B(\theta_1)L_\varepsilon(\theta_1) = TB(\theta_0)L_\varepsilon(\theta_0)V$, $C(\theta_1) = C(\theta_0)T^{-1}$. Letting $U \equiv L_\varepsilon(\theta_0)VL_\varepsilon(\theta_1)^{-1}$ be a full rank $n_\varepsilon \times n_\varepsilon$ matrix so that $U\Sigma_\varepsilon(\theta_1)U' = \Sigma_\varepsilon(\theta_0)$ gives the desired result.

Q.E.D.

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 $n_\varepsilon \times n_\varepsilon$ matrices $\{U_1, \ldots, U_k, \ldots\}$ such that $T_kA(\theta_k)T_k^{-1} = A(\theta_0)$, $T_kB(\theta_k)U_k = B(\theta_0)$, $C(\theta_k)T_k^{-1} = C(\theta_0)$, $D(\theta_k)U_k = D(\theta_0)$, and $U_k^{-1}\Sigma(\theta_k)U_k^{-1'} = \Sigma(\theta_0)$, that is, $\delta^{\mathrm{S}}(\theta_k, T_k, U_k) = \delta^{\mathrm{S}}(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$. 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 $\mathrm{I}_{n_\varepsilon}$, 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, $\operatorname{rank}\mathcal{O}(\theta_0) = n_X$ and a left inverse exists which gives $T_k = [\mathcal{O}(\theta_0)'\mathcal{O}(\theta_0)]^{-1}\mathcal{O}(\theta_0)'\mathcal{O}(\theta_k)$. By continuity of $\mathcal{O}(\theta)$, $\mathcal{O}(\theta_k) \to \mathcal{O}(\theta_0)$ as $\theta_k \to \theta_0$, so $T_k \to \mathrm{I}_{n_X}$. To show that $U_k \to \mathrm{I}_{n_\varepsilon}$, take any $|z| > 1$ and note that

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

Since $\operatorname{rank}\mathcal{P}(z;\theta_k) = n_X + n_\varepsilon$, a left inverse exists and

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

It follows from continuity that $U_k \to \mathrm{I}_{n_\varepsilon}$ as $\theta_k \to \theta_0$. This shows that δ^{S} is not injective in the neighborhood of $(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$.

Necessity. To show that local identifiability of θ_0 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 $n_\varepsilon \times n_\varepsilon$ matrices, respectively, such that $\delta^{\mathrm{S}}(\theta_1, T, U) = \delta^{\mathrm{S}}(\theta_0, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})$. 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 $U^{-1}\Sigma(\theta_1)U^{-1'} = \Sigma(\theta_0)$. 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 the only solution is $\theta_1 = \theta_0$ implies $T = \mathrm{I}_{n_X}$ and $U = \mathrm{I}_{n_\varepsilon}$. For this, consider again the observability matrices and note that $\mathcal{O}(\theta_1)T^{-1} = \mathcal{O}(\theta_0)$. Under observability, $\operatorname{rank}\mathcal{O}(\theta_1) = \operatorname{rank}\mathcal{O}(\theta_0) = n_X$, so $\theta_1 = \theta_0$ implies $T = \mathrm{I}_{n_X}$. We can then use left-invertibility of $\mathcal{P}(z;\theta_1)$ and $\mathcal{P}(z;\theta_0)$ in $|z| > 1$ to show that $U = \mathrm{I}_{n_\varepsilon}$. Hence, δ^{S} is locally injective. *Q.E.D.*

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 columnwise vectorization with the upper portion excluded. 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, $\operatorname{vec}(S) = \mathcal{G}_n \operatorname{vech}(S)$. 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 $\mathcal{E}_n\mathcal{G}_n = \mathrm{I}_{n(n+1)/2}$ and $\mathcal{G}_n\mathcal{E}_n = (1/2)[\mathrm{I}_{n^2} + \mathcal{P}_{n,n}]$, where $\mathcal{P}_{n,n}$ is the $n^2 \times n^2$ permutation matrix that

transforms $\operatorname{vec} X$ into $\operatorname{vec} X'$, that is, $\mathcal{P}_{n,n}\operatorname{vec} X = \operatorname{vec} X'$. Note that $\mathcal{P}_{n,n} = \mathcal{P}_{n,n}^{-1}$ and $\mathcal{P}_{n,n} = \mathcal{P}_{n,n}'$. In addition, $\operatorname{rank}(\mathrm{I}_{n^2} + \mathcal{P}_{n,n}) = \frac{n(n+1)}{2}$.

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

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

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

and

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

Now let

$$\Delta^{\mathrm{S}}(\theta) \equiv \Big(\frac{\partial \delta^{\mathrm{S}}(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})}{\partial \theta} \quad \frac{\partial \delta^{\mathrm{S}}(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})}{\partial \operatorname{vec} T} \quad \frac{\partial \delta^{\mathrm{S}}(\theta, \mathrm{I}_{n_X}, \mathrm{I}_{n_\varepsilon})}{\partial \operatorname{vec} U} \Big),$$

that is,

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

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

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

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

The $\frac{n_\varepsilon(n_\varepsilon+1)}{2} \times \frac{n_\varepsilon(n_\varepsilon+1)}{2}$ matrix $\mathcal{E}_{n_\varepsilon}(U^{-1} \otimes U^{-1})\mathcal{G}_{n_\varepsilon}$ 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:

$$\operatorname{rank}\left(\frac{\partial \delta^{\mathrm{S}}(\theta,T,U)}{\partial \theta} \quad \frac{\partial \delta^{\mathrm{S}}(\theta,T,U)}{\partial \operatorname{vec} T} \quad \frac{\partial \delta^{\mathrm{S}}(\theta,T,U)}{\partial \operatorname{vec} U}\right) = \operatorname{rank} \Delta^{\mathrm{S}}(\theta).$$

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

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

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 $H_\varepsilon(z;\theta)L_\varepsilon(\theta)$. 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 semidefinite solutions to the discrete algebraic Ricatti equation (DARE)

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

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 semidefinite solution $\overline{\Sigma}(\theta)$ to the DARE (17). Moreover, if

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

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

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 innovations representation

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

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

where $\widehat{X}_{t|t}$ is 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}, $a_t \sim \mathrm{WN}(0, \Sigma_a(\theta))$. 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 $H_a(z;\theta) = \mathrm{I}_{n_Y} + C(\theta)(z\mathrm{I}_{n_X} - A(\theta))^{-1}K(\theta)$.

Step 2—Spectral Factorization. It must first be shown that $\operatorname{rank} H_a(z;\theta) = n_Y$ 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, and d with a and 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 $\det(z\mathrm{I}_{n_X} - A(\theta))\det(\mathrm{I}_{n_X} + C(\theta)[z\mathrm{I}_{n_X} - A(\theta)]^{-1}K(\theta)) = \det(z\mathrm{I}_{n_X} - A(\theta) + K(\theta)C(\theta))$. Equivalently,

$$\begin{aligned}(21)\qquad \det H_a(z;\theta) &= \det\big(\mathrm{I}_{n_X} + C(\theta)\big[z\mathrm{I}_{n_X} - A(\theta)\big]^{-1}K(\theta)\big) \\ &= \frac{\det(z\mathrm{I}_{n_X} - [A(\theta) - K(\theta)C(\theta)])}{\det(z\mathrm{I}_{n_X} - A(\theta))}.\end{aligned}$$

Since $\det(z\mathrm{I}_{n_X} - [A(\theta) - K(\theta)C(\theta)]) \neq 0$ for all $|z| > 1$, it follows that $\operatorname{rank} H_a(z;\theta) = n_Y$ 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 $\operatorname{rank} W(z;\theta) = \operatorname{rank}(H_a(z;\theta)L_a(\theta)) = n_Y$ 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

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

Step 3. From (22),

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

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 same 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 $\mathbb{D}(\theta_1) = \mathbb{D}(\theta_0)$, $A(\theta_1) = TA(\theta_0)T^{-1}$, $\mathbb{B}(\theta_1) = T\mathbb{B}(\theta_0)$, and $C(\theta_1) = C(\theta_0)T^{-1}$. Equivalently, $L_a(\theta_1) = L_a(\theta_0)V$, $A(\theta_1) = TA(\theta_0)T^{-1}$, $K(\theta_1)L_a(\theta_1) = TK(\theta_0)L_a(\theta_0)V$, and $C(\theta_1) = C(\theta_0)T^{-1}$. Now uniqueness of Cholesky decomposition implies $V = \mathrm{I}_{n_Y}$. Thus $L_a(\theta_1) = L_a(\theta_0)$ and the result follows. *Q.E.D.*

The proof of Proposition 2-NS follows directly from the proof of Proposition 2-S and is hence omitted. The proofs of Propositions 3 and 4 are analogous to the poofs of Propositions 2-S and 2-NS.

PROOF OF LEMMA 3: Recall that $\Lambda \equiv \Lambda^{\mathrm{NS}} = ((\operatorname{vec} A)', (\operatorname{vec} K)', (\operatorname{vec} C)', (\operatorname{vech} \Sigma_a)')'$. For any $j \geq 0$ and $J \geq 0$, let

$$\overline{h}(j;\theta) \equiv h_a(j;\theta)L_a(\theta),$$
$$\overline{h}_J(\theta) \equiv \big(\, h_a(0;\theta)L_a(\theta) \quad h_a(1;\theta)L_a(\theta) \quad \cdots \quad h_a(J;\theta)L_a(\theta)\,\big),$$
$$\frac{\partial \operatorname{vec} \overline{h}_J(\theta)}{\partial \Lambda} = \begin{pmatrix} \dfrac{\partial \operatorname{vec} \overline{h}(0;\theta)}{\partial \Lambda} \\ \vdots \\ \dfrac{\partial \operatorname{vec} \overline{h}(J;\theta)}{\partial \Lambda} \end{pmatrix},$$
$$\Delta_{\overline{h}}(j;\theta) \equiv \frac{\partial \operatorname{vec} \overline{h}(j;\theta)}{\partial \theta}.$$

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

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

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

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

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

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

where the second equality follows by direct computation. This establishes result (i) of the lemma. The result is quite intuitive: By Proposition 1-NS, 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 $j \geq 0$. This implies that the columns of $\Delta_T(\theta)$ belong to the null space of $\frac{\partial \operatorname{vec} \overline{h}_J(\theta)}{\partial \Lambda}$ because $\Delta_T(\theta)$ is the Jacobian matrix of the similarity transformation

$(A(\theta_1), K(\theta_1), C(\theta_1), \Sigma_a(\theta_1)) = (TA(\theta_0)T^{-1}, TK(\theta_0), C(\theta_0)T^{-1}, \Sigma_a(\theta_0))$ that leaves $\overline{h}_J(\theta)$ unchanged.

(ii) To show that $\operatorname{rank}\overline{h}_J(\theta) = \operatorname{rank}\overline{h}_{2n_X-2}(\theta)$ for all $J \geq 2n_X - 2$, we use the Cayley–Hamilton theorem which ensures that for any minimal system $(A(\theta), K(\theta), C(\theta), \mathrm{I}_{n_Y})$,

$$\operatorname{rank}(\mathcal{C}_N(\theta)'\mathcal{O}_N(\theta)') = \operatorname{rank}\big(\mathcal{C}_{n_X}(\theta)'\mathcal{O}_{n_X}(\theta)'\big) \quad \text{for all } N \geq n_X,$$

where $\mathcal{C}_N(\theta)$ and $\mathcal{O}_N(\theta)$ are the controllability and the observability matrices of order N, that is,

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

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

(iii) Combining $\Delta(\theta) = (\Delta_\Lambda(\theta) \;\; \Delta_T(\theta))$ 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 $\operatorname{rank}\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = n_\theta$ implies $\operatorname{rank}\Delta(\theta_0) = n_\theta + n_X^2$. Now, $\operatorname{rank}\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = n_\theta$ implies that $\operatorname{rank}\Delta_\Lambda(\theta_0) = n_\theta$ because from (23), $\operatorname{rank}\Delta_{\overline{h}_{2n_X-2}}(\theta_0) \leq \operatorname{rank}\Delta_\Lambda(\theta_0) \leq n_\theta$. Then proceed by contradiction: assume that $\operatorname{rank}\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = n_\theta$ and that $\operatorname{rank}\Delta(\theta_0) < n_\theta + n_X^2$. Then $\operatorname{rank}\Delta_\Lambda(\theta_0) = n_\theta$ and $\operatorname{rank}\Delta(\theta_0) < n_\theta + n_X^2$. 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 $\operatorname{rank}\Delta(\theta_0) = n_\theta + n_X^2$ implies $\operatorname{rank}\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = n_\theta$. Now, $\operatorname{rank}\Delta(\theta_0) = n_\theta + n_X^2$ implies $\operatorname{rank}\Delta_\Lambda(\theta_0) = n_\theta$ and $\operatorname{rank}\Delta_T(\theta_0) = n_X^2$. First, we show that when the system is minimal and left-invertible, then

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

where $\mathcal{N}(\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda})$ is the null space of $\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda}$ and $\text{span}(\Delta_T(\theta_0))$ denotes the subspace spanned by the columns of $\Delta_T(\theta_0)$. For this, consider a Taylor expansion of $\text{vec}\,\overline{h}_{2n_X-2}(\theta)$ around $\theta = \theta_0$,

$$\text{vec}\,\overline{h}_{2n_X-2}(\theta_0+\delta) = \text{vec}\,\overline{h}_{2n_X-2}(\theta_0) + \frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda}\Delta_\Lambda(\theta_0)\delta^*,$$

where $\delta^* \in (0, \delta)$. Under minimality and left-invertibility, equalities (10) are the only transformations that lead to the same $\overline{h}_{2n_X-2}$. That is, $\text{vec}\,\overline{h}_{2n_X-2}(\theta_0 + \delta) = \text{vec}\,\overline{h}_{2n_X-2}(\theta_0)$ if and only if $\Delta_\Lambda(\theta_0)\delta^* \in \text{span}(\Delta_T(\theta_0))$. Combining this with the Taylor expansion gives

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

so (24) holds. To show that this implies that $\text{rank}\,\Delta_{\overline{h}_{2n_X-2}}(\theta_0) = \text{rank}(\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda} \cdot \Delta_\Lambda(\theta_0)) = n_\theta$, we proceed by contradiction. Suppose that $\text{rank}(\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda} \cdot \Delta_\Lambda(\theta_0)) < n_\theta$. Then there exists a nonzero linear combination of the columns of $\Delta_\Lambda(\theta_0)$ which belongs to the null space $\mathcal{N}(\frac{\partial \overline{h}_{2n_X-2}(\theta_0)}{\partial \Lambda})$. Then using (24), 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 $\text{rank}\,\Delta(\theta_0) = n_\theta + n_X^2$ and thus leads to a contradiction. *Q.E.D.*

REFERENCES

AN, S., AND F. SCHORFHEIDE (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26, 113–172. [1998,2015,2016,2018]

ANDERSON, B. D. O. (1969): "The Inverse Problem of Stationary Covariance Generation," *Journal of Statistical Physics*, 1, 133–147. [2003,2021,2027]

ANDERSON, B. D. O., AND J. MOORE (1979): *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall. [2002,2007]

ANDERSON, G., AND G. MOORE (1985): "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models," *Economic Letters*, 17, 247–252. [1998]

ANDERSON, T. W. (1984): *An Introduction to Multivariate Statistical Analysis*. New York: Wiley. [2016]

ANTSAKLIS, P., AND A. MICHEL (1997): *Linear Systems*. New York: McGraw-Hill. [2003,2022]

BENIGNO, G., P. BENIGNO, AND S. NISTICO (2011): "Second Order Approximation of Dynamic Models With Time Varying Risk," Working Paper 16633, NBER. [1999]

BONALDI, J. P. (2010): "Identification Problems in the Solution of Linearized DSGE Models," Working Paper 593, Banco de la Republica. [2014]

CANOVA, F., AND L. SALA (2009): "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, 56, 431–449. [1995,2016]

COCHRANE, J. (2011): "Determinancy and Identification With Taylor Rules," Mimeo, University of Chicago. [2018]

CONSOLO, A., C. FAVERO, AND A. PACCAGNINI (2009): "On the Statistical Identification of DSGE Models," *Journal of Econometrics*, 150, 99–115. [1996]

CURDIA, V., AND R. REIS (2010): "Correlated Disturbances and U.S. Business Cycles," Working Paper 15774, NBER. [1999]

DEISTLER, M. (1976): "The Identifiability of Linear Econometric Models With Autocorrelated Errors," *International Economic Review*, 17, 26–46. [1997,2006]

FERNÁNDEZ-VILLAVERDE, F., J. RUBIO-RAMÍREZ, T. SARGENT, AND M. WATSON (2007): "A,B,Cs and (D)'s for Understanding VARs," *American Economic Review*, 97, 1021–1026. [2002,2021]

FISHER, F. (1966): *The Identification Problem in Econometrics*. New York: McGraw-Hill. [1997, 2006]

GLOVER, K., AND J. WILLEMS (1974): "Parameterizations of Linear Dynamical Systems: Canonical Forms and Identifiability," *IEEE Transactions on Automatic Control*, 19, 640–646. [2006]

HANNAN, E. (1971): "The Identification Problem for Multiple Equation Systems With Moving Average Errors," *Econometrica*, 39, 751–765. [1997,2002,2006]

HANNAN, E., AND M. DIESTLER (1988): *The Statistical Theory of Linear Systems*. New York: Wiley. [2003]

HANSEN, L. P., AND T. SARGENT (2005): "Recursive Methods of Dynamic Linear Economies," Unpublished Manuscript, University of Chicago. Available at http://home.uchicago.edu/~lhansen/mbook2.pdf. [2007,2026]

HATANAKA, M. (1975): "On the Global Identification of the Dynamic Simultaneous Equations Model With Stationary Disturbances," *International Economic Review*, 16, 545–554. [1997]

HAUSMAN, J., AND W. TAYLOR (1983): "Identification in Linear Simultaneous Equations Models With Covariance Restrictions: An Instrumental Variables Interpretation," *Econometrica*, 51, 1527–1550. [1997]

ISKREV, N. (2008): "Evaluating the Information Matrix in Linearized DSGE Models," *Economics Letters*, 99, 607–610. [2014]

——— (2010): "Local Identification in DSGE Models," *Journal of Monetary Economics*, 57, 189–202. [1996,2012]

JUSTINIANO, A., AND G. PRIMICERI (2008): "The Time Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98, 604–641. [1999]

KAILATH, T., A. SAYED, AND B. HASSIBI (2000): *Linear Estimation*. Upper Saddle River, NJ: Prentice Hall. [2003,2021,2026]

KING, R. G., AND M. W. WATSON (2002): "System Reduction and Solution Algorithms for Singular Linear Difference Systems Under Rational Expectations," *Computational Economics*, 20, 57–86. [1998]

KLEIN, P. (2000): "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Journal of Economic Dynamics and Control*, 24, 1405–1423. [1998]

KOMUNJER, I. (2011): "Global Identification in Nonlinear Models With Moment Restrictions," *Econometric Theory* (forthcoming). [2020]

KOMUNJER, I., AND S. NG (2011): "Supplement to 'Dynamic Identification of DSGE Models'," *Econometrica Supplemental Material*, 79, http://www.econometricsociety.org/ecta/Supmat/8916_extensions.pdf; http://www.econometricsociety.org/ecta/Supmat/8916_data and programs.zip. [1998]

KOOPMANS, T. C. (ED.) (1950): *Statistical Inference in Dynamic Economic Models*. Cowles Commission Monograph, Vol. 10. New York: Wiley. [1997]

LAUB, A., AND B. MOORE (1978): "Calculation of Transmission Zeros Using QZ Techniques," *Automatica*, 14, 557–566. [2021]

LEEPER, E., T. WALKER, AND S. YANG (2008): "Fiscal Foresight: Analytics and Econometrics," Working Paper 14028, NBER. [2000,2010]

LIPPI, M., AND L. REICHLIN (1994): "VAR Analysis, Non-Fundamental Representation, Blaschke Matrices," *Journal of Econometrics*, 63, 307–325. [2000]

LÜTKEPOHL, H. (2005): *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag. [1997]

QU, Z., AND D. TKACHENKO (2010): "Identification and Frequency Domain QML Estimation of Linearized DSGE Models," Mimeo, Boston University. [2012]

RAVENNA, F. (2007): "Vector Autoregressions and Reduced Form Representations of DSGE Models," *Journal of Monetary Economics*, 54, 2048–2064. [1997]

REINSEL, G. (2003): *Elements of Multivariate Time Series Analysis* (Second Ed.). New York: Springer. [1997]

ROTHENBERG, T. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591. [1997]

ROZANOV, Y. A. (1967): *Stationary Random Processes*. San Francisco: Holden-Day. [2000]

RUBIO-RAMÍREZ, J. F., D. F. WAGGONER, AND T. ZHA (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economics and Studies*, 77, 665–696. [1996]

SARGAN, J. D. (1975): "The Identification and Estimation of Sets of Simultaneous Stochastic Equations," reprinted in SARGAN, J. D. (1988): *Contributions to Econometrics*, Vol. 1, ed. by E. Maasoumi. Cambridge: Cambridge University Press, 236–249, Chap. 12. [1997]

SIMS, C. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20. [1998,2015]

SMETS, F., AND R. WOUTERS (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606. [2016]

SOLO, V. (1986): "Topics in Advanced Time Series," in *Lectures in Probability and Statistics*, Vol. 1215, ed. by G. del Pino and R. Rebolledo. Berlin: Springer-Verlag. [1997]

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. Oxford: Oxford University Press, 30–61. [1998]

WALLIS, K. (1977): "Multiple Time Series Analysis and the Final Form of Econometric Models," *Econometrica*, 45, 1481–1497. [1998]

YOULA, D. C. (1961): "On the Factorization of Rational Matrices," *IRE Transactions on Information Theory*, 7, 172–189. [2003,2021,2027]

Dept. of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, U.S.A.; komunjer@ucsd.edu

and

Dept. of Economics, Columbia University, 420 West 118 Street, MC 3308, New York, NY 10027, U.S.A.; serena.ng@columbia.edu.

Manuscript received November, 2009; final revision received May, 2011.