Federal Reserve Bank of Dallas

Finite-Order VAR Representation of Linear Rational Expectations Models: With Some Lessons for Monetary Policy

Enrique Martínez-García

Finite-Order VAR Representation of Linear Rational Expectations Models: With Some Lessons for Monetary Policy[*]

Enrique Martínez-García[†]

September 2016 (Revised: August 1, 2018)

Abstract

This paper considers the characterization via finite-order VARs of the solution of a large class of linear rational expectations (LRE) models. I propose a unified approach that uses a companion Sylvester equation to check the existence and uniqueness of a solution to the canonical (first-order) LRE model in finite-order VAR form and a quadratic matrix equation to characterize it decoupling the backward- and forward-looking aspects of the model. I also investigate the fundamentalness of the shocks recovered. Solving LRE models by this procedure is straightforward to implement, general in its applicability, efficient in the use of computational resources, and can be handled easily with standard matrix algebra. An application to the workhorse New Keynesian model with accompanying Matlab codes is provided to illustrate the practical implementation of the methodology. I argue that existing empirical evidence on the transmission mechanism of monetary policy shocks from structural VARs (when the specification is inconsistent with theory due to the identification restrictions, lag specification, etc.) should be taken with a grain of salt as it may not have a proper structural interpretation.

Keywords: Linear Rational Expectations Models; Vector Autoregression Representation; Sylvester Matrix Equation; New Keynesian Model; Monetary Policy Shocks

JEL Classification: C32, C62, C63, E37

[†]Enrique Martínez-García, Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, TX 75201. Phone: (214) 922-5262. Fax: (214) 922-5194. E-mail: emg.economics@gmail.com. Webpage: https://sites.google.com/view/emgeconomics.

1 Introduction

The solution of linear or linearized rational expectations (LRE) models is an important part of modern macroeconomics. They are widely used to study the propagation mechanism of economic shocks, for identification, and to provide economic evaluation of policy changes. Many rational expectations macro models can be cast as a linear system of expectational difference equations. The linearity of the system may be a feature of the model itself but often is simply achieved from the first-order approximation of a Dynamic Stochastic General Equilibrium (DSGE) model. Either way, Blanchard and Kahn (1980) established the conditions under which a solution to the LRE model exists and is unique (see, among others, the related contributions of Broze et al. (1985), Broze et al. (1990), King and Watson (1998), Uhlig (1999), and Klein (2000)).

Maximum likelihood and Bayesian estimation methods can be used on LRE models with a unique solution to achieve a constrained fit of the data. However, the theory may fail in fitting the data satisfactorily because of misspecification—the cross-equation restrictions imposed may be at odds with the *true* data-generating process. Even under the null hypothesis that the LRE model is correctly specified, it may still be the case that the theory suffers from weak identification problems (see Martínez-García et al. (2012) and Martínez-García and Wynne (2014) on this point).

Structural VAR models in the spirit of Sims (1980) provide a framework with which to investigate and organize the evidence on a set of observable variables that, in principle, is largely devoid of many of the restrictions implied by theory (the LRE model) and in that way is more flexible to fit the data. The work, among others, of Fernández-Villaverde et al. (2007), Ravenna (2007), and more recently Franchi and Paruolo (2015) has shed new light on the mapping between the unique solution to the LRE model and a corresponding VAR representation. To be more precise, this strand of the literature explores conditions under which the unique solution of the LRE model—when it exists—can be properly represented in VAR form, as this facilitates the recovery of structural shocks as well as empirical inference and validation when the LRE model is brought to the data.

The structural shock innovations driving the LRE model cannot always be recovered even when a VAR representation of the unique LRE solution can be obtained due to lack of fundamentalness. Non-fundamentalness means that the observed variables do not contain enough information to recover the unobserved structural shocks (Hansen and Sargent (1980)). An LRE model solution is said to be fundamental if the structural moving average (MA) representation of the observed variables can be inverted. If the LRE model solution is fundamental (assuming the LRE model itself is correctly specified), then the observed variables have a VAR representation in the structural shocks—implying that the structural shocks can be recovered by estimating a VAR with the observed variables and that their corresponding impulse response functions can be correctly inferred.[1]

When the number of structural shocks is equal to the number of observables, the fundamentalness property of the unique solution to the LRE model can be checked with the 'poor man's invertibility condition' of Fernández-Villaverde et al. (2007). Fernández-Villaverde et al. (2007) show the conditions under which

---

[1] Even if the unique model solution has a VAR representation that is fundamental—in that it permits the exact recovery of the structural shocks and the characterization of their corresponding impulse response functions—it is worth noting that this does not ensure that the matrices of composite coefficients that describe the dynamics for the VAR form will uniquely identify all the structural parameters of a well-specified LRE model. Therefore, policy evaluation and even model comparison are still subject to the perils of identification failure noted by Martínez-García et al. (2012) and Martínez-García and Wynne (2014). In any event, the treatment of weak/partial identification and even misspecification falls outside the scope of this paper.

a VAR representation exists, while Ravenna (2007) describes the conditions that determine when such a VAR representation of the unique LRE solution is of finite order. Otherwise the inverted structural vector moving average (VMA) representation of the unique solution to the LRE model takes a VAR(∞) form and that introduces a truncation error when cast as a finite-order structural VAR (Inoue and Kilian (2002)).

Franchi and Paruolo (2015) show that if the state-space representation of the LRE solution is minimal, then both the 'poor man's invertibility condition' of Fernández-Villaverde et al. (2007) and the finite-order VAR conditions of Ravenna (2007) are necessary and sufficient to ensure fundamentalness and the existence of an exact finite-order VAR form for the solution of the LRE model. In their paper, Franchi and Paruolo (2015) argue that when the state-space representation is non-minimal, the 'poor man's invertibility condition' and the finite-order VAR conditions uncovered by Ravenna (2007) are not necessary.[2]

The main contribution of this paper with respect to the work of Fernández-Villaverde et al. (2007), Ravenna (2007), and Franchi and Paruolo (2015) is the use of theory to characterize the finite-order VAR form of the LRE model and determine the fundamentalness of the solution. While the preceding research works from a generic linear state-space representation of the solution, I restrict myself to representations of the solution that arise from theory—that is, from well-behaved LRE models. In other words, I derive my results after imposing the cross-equation restrictions that arise from the theoretical model first. Therefore, my approach derives the finite-order VAR representation of the solution and checks its existence and uniqueness directly on the known primitives of the LRE model.

In this vein, I complement the existing literature by proposing an alternative approach for determining existence and uniqueness and to characterize the solution in finite-order VAR form for a large class of LRE models via a pair of companion matrix equations—a quadratic matrix equation and a Sylvester matrix equation:

∘ First, LRE models that include backward-looking and forward-looking features with one or more lags and leads can be reduced to the canonical form of an expectational first-order system of difference equations without backward-looking terms. System reduction from the general form of the LRE model to the canonical form is achieved by solving a companion quadratic matrix equation.

∘ Second, the well-known method of undetermined coefficients can be used to solve the canonical forward-looking part of the LRE model. Conditions under which a finite-order VAR representation of the canonical LRE model solution can be obtained and a simple (yet efficient) algorithm to compute it can be derived from a companion Sylvester matrix equation.

∘ Finally, the last step simply requires reversing the transformation of the system utilized in the first step to decouple the backward- and forward-looking parts of the model in order to recover the exact representation of the solution to the general form of the LRE model.

The initial step of system reduction involves the solution of a quadratic matrix equation (as indicated before), but also permits generalizing the approach and the implementation proposed in this paper to cover a wide range of LRE models (Binder and Pesaran (1995) and Binder and Pesaran (1997)). The key contribution of the paper is that the characterization of the finite-order VAR solution of the forward-looking part of the

---

[2]The state-space form is called minimal if the dimension of the vector of forcing variables is as small as possible (Kailath (1980)). Fundamentalness and the existence of a finite-order VAR representation of the LRE model solution that is non-minimal can still be asserted by first transforming the state-space form to a minimal representation and then applying the existing conditions. Franchi and Paruolo (2015) explore necessary and sufficient conditions that are valid when the state-space form is non-minimal based on the possibility of exploiting cancellations (as in related problems from systems theory), bypassing the step of transforming the LRE model to its minimal state-space form first.

canonical LRE model arises naturally from the solution of a Sylvester matrix equation. I propose a simple approach based on this companion Sylvester equation to check for and identify unique LRE solutions in finite-order VAR form and a simple algorithm to compute such solutions. The conditions that verify existence and uniqueness of a finite-order VAR for the canonical LRE model solution also ensure the fundamentalness of the solution. Those conditions are consistent with the 'poor man's invertibility condition' of Fernández-Villaverde et al. (2007).

The tools proposed here are meant to be used by macroeconomists who deal with LRE models in their theoretical or applied work and who need to determine if the observable (endogenous) variables in their LRE model admit an exact finite-order VAR representation that is also fundamental. I illustrate the practical use of this novel approach with the workhorse New Keynesian model—showing how the procedure can be used to derive the finite-order VAR representation of the unique LRE model solution, to establish its existence and uniqueness, and to make economically-relevant inferences about the New Keynesian transmission mechanism to recover structural shocks and explore their propagation patterns. Moreover, I also contribute to the ongoing debate on the assumptions needed to correctly recover theoretically-consistent monetary policy shocks through structural VARs (Carlstrom et al. (2009)).

The rest of the paper proceeds as follows: Section 2 describes the system reduction method to decouple the backward-looking and forward-looking parts of the general form LRE model and shows how to use the method of undetermined coefficients to characterize the linear state-space form solution of the (canonical) forward-looking part of the LRE model. This section also describes the mapping of the LRE model solution into finite-order VAR form via a companion Sylvester equation and the conditions under which a unique finite-order VAR solution can be attained from the companion Sylvester equation. Section 3 applies the method to a policy-relevant illustration on the effects of monetary policy on inflation determination based on the workhorse three-equation New Keynesian model with which I also illustrate the computational efficacy of the approach. Furthermore, I discuss the fundamentalness of the finite-order VAR solution and what it means for the ongoing debate on the assumptions required to recover structural monetary shocks from VARs. Section 4 discusses the implications of the workhorse New Keynesian model for our understanding of the propagation of monetary policy shocks and highlights the estimation challenges for the identification of the deep structural parameters of the model. Section 5 then concludes.

The Appendix provides additional technical details on the system reduction approach used in this paper to isolate the forward-looking part of the LRE model in general form—including a generalized eigenvalue problem algorithm to implement it. The Appendix also discusses in some detail the available algorithms and methods for solving the companion quadratic matrix equation and the companion Sylvester matrix equation. The procedure itself is easily cast in an algorithmic form, and a collection of Matlab implementation codes is provided with the paper.[3]

---

[3] All codes for this paper are available in my website: https://sites.google.com/view/emgeconomics and can be downloaded directly using the following link: https://drive.google.com/file/d/1ZZ-zCB8z9_RLLolEX4fMKKWXV8QG8-Er/view.

Users of the codes are asked to include a citation of this paper in their work. The Matlab programs and functions appear free of errors, however I do appreciate all feedback, suggestions or corrections that you may have. While users are free to copy, modify and use the code for their work, I do not assume any responsibility for any remaining errors or for how the codes may be used or misused by users other than myself.

2 Solving Linear Rational Expectations Models

The equilibrium relationships that describe the linear or linearized rational expectations (LRE) model are given by theory. However, going from the structural relationships implied by the canonical LRE model to a reduced-form solution requires explicit assumptions on how expectations are formed and on the stochastic processes that describe the exogenous forcing variables. Under rational expectations, agents understand the structure of the economy and formulate expectations optimally incorporating all available information. Here, I consider the class of LRE models where expectations are fully rational and the exogenous forcing variables are assumed to follow a finite-order VAR process.[4] Under these assumptions, the reduced-form solution of the LRE model can be mapped into a structural VAR representation for a set of the endogenous variables of the model. That explains why VARs appear to fit the data well. The VAR representation of the reduced-form solution also provides researchers with more bite to investigate the propagation of shocks than the unrestricted VAR model does.

In this paper, I explore the connection between the reduced-form solution of the canonical LRE model and structural (finite-order) VARs to bridge the gap between theoretical and applied work. A large class of LRE models can be cast into a first-order expectational difference system of equations, featuring forward- and backward-looking dynamics. The first-order expectational difference equations capture the structural relationships between a set of $m \geq k$ endogenous variables $W_t = (w_{1t}, w_{2t}, ..., w_{mt})^T$ and $k \geq 1$ forcing variables $X_t = (x_{1t}, x_{2t}, ..., x_{kt})^T$ as follows:

$$W_t = \Phi_1 W_{t-1} + \Phi_2 \mathbb{E}_t [W_{t+1}] + \Phi_3 X_t, \tag{1}$$

$$X_t = A X_{t-1} + B \epsilon_t, \tag{2}$$

with k exogenous shock innovations $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{kt})^T$ which are independent and identically distributed (i.i.d.) with zero-mean and standard deviation of one.[5] The structural relationships of the canonical first-order LRE model given by (1) are completed with the standard (first-order) VAR(1) specification for the vector of k forcing variables X_t in (2).[6] The conforming matrices Φ_1 and Φ_2 in (1) are $m \times m$ square, Φ_3 is an $m \times k$ matrix, while A and B are both $k \times k$ square matrices.

The canonical first-order system in $(1) - (2)$ can be further generalized to include LRE models with more than one lead and one lag of the endogenous variables (W_t) and of the forcing variables (X_t) in the specification, as explained in Broze et al. (1985) and Broze et al. (1990) (see also the Appendix). Hence, the canonical first-order system given by $(1) - (2)$ can be used to investigate a large class of LRE models.

---

[4] The idea of rational expectations can be traced to the seminal work by Muth (1961). Lucas (1976) and Sargent (1980) were among the leading economists that rejected *ad hoc* assumptions on the formation of expectations and advocated the adoption of rational expectations that became prevalent in modern macroeconomics since the 80s.

[5] In practice, the literature often includes measurement errors as structural shock innovations in the vector ϵ_t as well.

[6] More generally, the full LRE model involves $m+p$ endogenous variables $\left(W_t^T, \widetilde{W}_t^T\right)^T$ where the remaining $p \geq 0$ endogenous variables given by the vector $\widetilde{W}_t = (\widetilde{w}_{1t}, \widetilde{w}_{2t}, ..., \widetilde{w}_{pt})^T$ can simply be expressed as functions of W_t and X_t. Here, the number of endogenous variables in W_t is m and this is at least equal or higher than the number of exogenous shock innovations k. I do not consider explicitly the case of partial recovery that arises when the number of endogenous variables that are observable is smaller than the number of exogenous shock innovations k, though. The selection of the subset of m endogenous variables included in the vector W_t which are observable can be significant, for instance, for the identification of the estimated structural parameters, as noted in Martínez-García et al. (2012) and Martínez-García and Wynne (2014). These questions, however, go beyond the scope of the current paper.

2.1 Decoupling Backward- and Forward-Looking Terms

Building on Broze et al. (1985), Broze et al. (1990), and the more recent contributions of Binder and Pesaran (1995) and Binder and Pesaran (1997), I propose a straightforward transformation that decouples the canonical first-order LRE model in $(1) - (2)$ in its backward- and forward-looking components to work out its full solution by parts. For a given $m \times m$ matrix Θ, the transformation of the vector W_t given by $W_t \equiv Z_t + \Theta W_{t-1}$ implies that the expectational difference system in (1) can be rewritten as:

$$Z_t + \Theta W_{t-1} = \Phi_1 W_{t-1} + \Phi_2 \mathbb{E}_t \left[Z_{t+1} + \Theta W_t \right] + \Phi_3 X_t$$
$$= \Phi_1 W_{t-1} + \Phi_2 \left[\mathbb{E}_t \left(Z_{t+1} \right) + \Theta \left(Z_t + \Theta W_{t-1} \right) \right] + \Phi_3 X_t, \tag{3}$$

which becomes,

$$\left(\mathbf{I}_m - \Phi_2 \Theta \right) Z_t = \Phi_2 \mathbb{E}_t \left(Z_{t+1} \right) + \left(\Phi_2 \Theta^2 - \Theta + \Phi_1 \right) W_{t-1} + \Phi_3 X_t. \tag{4}$$

where $\mathbf{I}_m$ is the $(m \times m)$ identity matrix. From here, this lemma follows:

Lemma 1 *A system reduction that excludes the backward-looking terms in* (1) *can be attained by choosing an* $m \times m$ *matrix* Θ *to satisfy that:*

$$P \left(\Theta \right) = \Phi_2 \Theta^2 - \Theta + \Phi_1 = \mathbf{0}_m, \tag{5}$$

where $\mathbf{0}_m$ *is an* $m \times m$ *matrix of zeroes. And, then, transforming the vector of* m *endogenous variables* W_t *as* $W_t \equiv Z_t + \Theta W_{t-1}$.

The transformation described in Lemma 1 uncouples the solution of W_t into a backward-looking part, $W_{tb} \equiv \Theta W_{t-1}$, and a forward-looking part, $W_{tf} \equiv Z_t$, such that $W_t \equiv W_{tb} + W_{tf}$. Hence, one needs to determine the matrix Θ solving the quadratic matrix equation in (5) in order to then be able to characterize the backward-looking part of the solution and reduce the expectational difference system in $(1) - (2)$ to its purely forward-looking expectational difference part. Binder and Pesaran (1995) and Binder and Pesaran (1997) establish the necessary and sufficient conditions under which real-valued solutions for Θ satisfying the quadratic matrix equation in (5) exist and provide an iterative algorithm to compute its stable solution.[7]

After decoupling the system, the vector of m transformed endogenous variables and states, $Z_t \equiv W_t - \Theta W_{t-1}$, must follow a first-order forward-looking expectational difference system of this form:

$$\Gamma_0 Z_t = \Gamma_1 \mathbb{E}_t \left[Z_{t+1} \right] + \Gamma_2 X_t, \tag{6}$$

where $\Gamma_0 \equiv \left(\mathbf{I}_m - \Phi_2 \Theta \right)$ and $\Gamma_1 \equiv \Phi_2$ are conforming $m \times m$ matrices, and $\Gamma_2 \equiv \Phi_3$ is the corresponding $m \times k$ matrix. Furthermore, whenever $\Gamma_0 \equiv \left(\mathbf{I}_m - \Phi_2 \Theta \right)$ is nonsingular, the system in (6) can be rewritten as:

$$Z_t = F \mathbb{E}_t \left[Z_{t+1} \right] + G X_t, \tag{7}$$

where $F \equiv \left(\Gamma_0 \right)^{-1} \Gamma_1 = \left(\mathbf{I}_m - \Phi_2 \Theta \right)^{-1} \Phi_2$ is an $m \times m$ matrix and $G \equiv \left(\Gamma_0 \right)^{-1} \Gamma_2 = \left(\mathbf{I}_m - \Phi_2 \Theta \right)^{-1} \Phi_3$ is an $m \times k$ matrix. The invertibility of $\Gamma_0 \equiv \left(\mathbf{I}_m - \Phi_2 \Theta \right)$ required to go from (6) to (7) depends on the choice

---

[7]An alternative algorithm to characterize the stable solution Θ based on the generalized eigenvalue problem can be found in the Appendix.

of the matrix Θ. Proposition 2 in Binder and Pesaran (1997) provides sufficient conditions under which Γ_0 would be nonsingular and invertible. The Binder and Pesaran (1997) conditions are only sufficient (not necessary), but I find that most well-specified economic LRE models indeed produce a matrix Γ_0 that is nonsingular.

In the remainder of this paper, I take (7) as the relevant benchmark with which to describe the canonical forward-looking part of the first-order LRE model.

2.2 The Finite-Order VAR Representation

Assuming that a solution to the forward-looking part of the canonical first-order LRE model given by (2) and (7) exists and is unique, then it can be written in linear state-space form as follows:

$$X_t = AX_{t-1} + B\epsilon_t, \tag{8}$$

$$Z_t = CX_{t-1} + D\epsilon_t, \tag{9}$$

where A, B, C, and D are conforming and real-valued matrices (A and B are $k \times k$ matrices while C and D are $m \times k$ matrices, with $m \geq k \geq 1$). Equation (8) simply restates (2) describing the dynamics of the vector of k forcing variables X_t, while (9) indicates that the k shock innovations ϵ_t and k lagged exogenous forcing variables X_{t-1} are mapped into the transformed m endogenous variables Z_t in the solution to the forward-looking part of the LRE model.

The following set of standard assumptions on the matrices A, B, C, and D are often considered:

Condition 1 *The linear state-space representation in* (8) − (9) *depends on the conforming and real-valued matrices A ($k \times k$), B ($k \times k$), C ($m \times k$, $m \geq k$), and D ($m \times k$, $m \geq k$) under a subset of the following conditions:*

(a) A ($k \times k$) has all its eigenvalues inside the unit circle ensuring the stationarity of the stochastic process for the k forcing variables ($rank\,(A) = k$) and is invertible. B ($k \times k$) characterizes the corresponding variance-covariance matrix $\left(B^T B\right)$ which is a symmetric and positive semi-definite matrix. Whenever B is assumed to be positive definite, then it is also invertible.

(b) C ($m \times k$, $m \geq k$) is left invertible ($rank\,(C) = k$) so there is a $k \times m$ matrix C_L^{-1} such that $C_L^{-1}C = \mathbf{I}_k$.

(c) All eigenvalues of CAC_L^{-1} ($m \times m$) are inside the unit circle.

(d) D ($m \times k$, $m \geq k$) is left invertible ($rank\,(D) = k$) so there is a $k \times m$ matrix D_L^{-1} such that $D_L^{-1}D = \mathbf{I}_k$.

(e) All eigenvalues of $\left(A - BD_L^{-1}C\right)$ ($k \times k$) are less than one in modulus (lie inside the unit circle)—the 'poor man's invertibility condition' of Fernández-Villaverde et al. (2007).

It follows from equation (8) that $[\mathbf{I}_k - AL]\,X_t = B\epsilon_t$ where $\mathbf{I}_k$ is a conforming identity matrix of dimension k, and L is the lag operator. If the eigenvalues of A are less than one in modulus (as in part (a) of Condition 1), the solution of the LRE model in state-space form has a moving average (MA) representation. In that case, X_t becomes a square summable polynomial given by $X_t = \sum_{j=0}^{+\infty} [A]^j B\epsilon_{t-j}$. This expression can be shifted one period back and replaced into (9) to obtain $Z_t = C \sum_{j=0}^{+\infty} [A]^j B\epsilon_{t-1-j} + D\epsilon_t$. Whenever the

matrix D is left invertible (as in part (d) of Condition 1), replacing the vector ϵ_t in (8) using (9) gives:

$$\left[\mathbf{I}_k - \left(A - BD_L^{-1}C\right)L\right]X_t = BD_L^{-1}Z_t. \tag{10}$$

If the eigenvalues of $\left(A - BD_L^{-1}C\right)$ are strictly less than one in modulus (as in part (e) of Condition 1), that amounts to the 'poor man's invertibility condition' of Fernández-Villaverde et al. (2007). Then, the inverse of the operator on the left-hand side of (10) implies that:

$$X_t = \sum_{j=0}^{+\infty}\left[A - BD_L^{-1}C\right]^j BD_L^{-1}Z_{t-j}, \tag{11}$$

which means X_t is a square summable polynomial in L.

Shifting the expression in (11) back one period and replacing it into (9), I obtain that:

$$Z_t = C\sum_{j=0}^{+\infty}\left[A - BD_L^{-1}C\right]^j BD_L^{-1}Z_{t-1-j} + D\epsilon_t, \tag{12}$$

which takes the form of an infinite-order VAR (VAR(∞)). Inoue and Kilian (2002) show that this VAR(∞) representation can be reasonably well-approximated by a finite-order structural VAR model. However, imposing the restrictions that arise from theory on the matrices C and D, it is possible to go beyond the approximation proposed by Inoue and Kilian (2002) establishing an exact finite-order VAR representation for the solution to the first-order canonical LRE model—under mild assumptions contained among the ones listed in Condition 1.

First, the dynamics of the vector of endogenous variables Z_t can be represented in VARMA(1, 1) form.[8] I take as given that the forcing variables are described by a well-behaved, stationary process as implied by part (a) of Condition 1. In order to derive the VARMA(1, 1) representation, let me rewrite (8) as $X_t = AC_L^{-1}CX_{t-1} + B\epsilon_t$ making use of part (b) of Condition 1. Then, I can replace (9) into the reformulated equation (8) in order to obtain:

$$X_t = AC_L^{-1}\left(Z_t - D\epsilon_t\right) + B\epsilon_t. \tag{13}$$

I pre-multiply (13) with C and replace (9) into it shifted one period ahead, i.e., I derive:

$$Z_{t+1} - D\epsilon_{t+1} = CX_t = CAC_L^{-1}Z_t + \left(CB - CAC_L^{-1}D\right)\epsilon_t, \tag{14}$$

so that a straightforward re-arranging of this expression gives a VARMA(1, 1) form for the vector of endogenous variables Z_t:

$$Z_t = CAC_L^{-1}Z_{t-1} + D\epsilon_t + \left(CB - CAC_L^{-1}D\right)\epsilon_{t-1}. \tag{15}$$

This representation holds for any given set of conforming matrices A, B, C, and D that satisfies parts (a) and (b) of Condition 1.

Second, while the VARMA(1, 1) in (15) is *true* for any pair of $m \times k$ matrices C and D that satisfy parts (a)-(b) of Condition 1, theory imposes a set of structural relationships that relate the matrices C and D to the known composite matrices F and G describing the forward-looking part of the canonical first-order LRE model in (7) and to the known matrices A and B that describe the stochastic process of the forcing variables in (8). I rely on the method of undetermined coefficients of Christiano (2002) in order to characterize the

---

[8]On the solution of LRE models in VAR form see, e.g., the contribution of Morris (2016).

theoretically-implied cross-equation restrictions on C and D in (9).[9]

Step 1. Using (9) shifted one period ahead to replace Z_{t+1} in the purely forward-looking system given in (7) implies that:

$$Z_t = [FC + G] X_t \tag{16}$$

$$= [FC + G] A X_{t-1} + [FC + G] B \epsilon_t, \tag{17}$$

where the second equality arises from replacing out X_t using (8). The forward-looking LRE model solution described by (9) is matched with (17) linking the unknown solution matrices C and D to the known composite matrices that define the structural relationships imposed by the forward-looking part of the canonical first-order LRE model and the stochastic process for the forcing variables (i.e., to F, G, A, and B). Matching coefficients between (9) and (17), it follows that the conforming $m \times k$ matrices C and D must satisfy the following pair of conditions:

$$C = [FC + G] A, \tag{18}$$

$$D = [FC + G] B = CA^{-1}B. \tag{19}$$

Assuming the eigenvalues of A $(k \times k)$ are all inside the unit circle (and zero is not an eigenvalue) in line with part (a) of Condition 1 ensures that the inverse matrix A^{-1} in (19) exists and is well-defined according to the invertible matrix theorem (Strang (2016)). As a result, the existence and uniqueness of a solution to C that satisfies (18) also pins down D through (19) guaranteeing the existence and uniqueness of the solution to the forward-looking part of the LRE model given by the linear state-space representation in $(8) - (9)$.

Hence, I find that solving the companion Sylvester matrix equation given by (18) to obtain C is enough to characterize the relevant cross-equation restrictions for the full solution to the forward-looking part of the LRE model given by $(8) - (9)$—whenever it holds that $\Gamma_0 \equiv (\mathbf{I}_m - \Phi_2\Theta)$ is invertible.

Step 2. Using (8) and the invertibility of A (part (a) of Condition 1), I can write (8) as $X_{t-1} = A^{-1}(X_t - B\epsilon_t)$. Replacing this expression into (9), it follows that:

$$Z_t = CA^{-1}X_t + \left[D - CA^{-1}B\right]\epsilon_t. \tag{20}$$

The characterization of the matrix D implied by equation (19) means that the term related to the vector of innovations ϵ_t must drop from (20). As a result, the vector of the transformed variables Z_t can be written as a simple linear mapping of the vector of exogenous forcing variables X_t where,

$$Z_t = CA^{-1}X_t, \tag{21}$$

as long as the matrix C that solves the Sylvester equation in (18) exists and the matrix D is defined by equation (19). Whenever C $(m \times k$ matrix) exists and is left invertible (part (b) of Condition 1), equation (21) implies that $X_t = AC_L^{-1}Z_t$ given that A is already assumed to be invertible (part (b) of Condition 1).

---

[9] Another conventional approach is that of Blanchard and Kahn (1980), which also provides conditions to check the existence and uniqueness of the solution. Broze et al. (1985), Broze et al. (1990), King and Watson (1998), Uhlig (1999), and Klein (2000), among others, expand on it in order to characterize the solution in more general settings. Another popular solution method applied to LRE models is the method of rational expectational errors advocated by Sims (2002) (see also Lubik and Schorfheide (2003)).

Shifting this expression one period back and replacing it into (9), and then replacing D with the expression given in equation (19), I obtain the following VAR(1) specification for Z_t:

$$Z_t = CAC_L^{-1} Z_{t-1} + CA^{-1}B\epsilon_t, \tag{22}$$

which is stationary if all eigenvalues of CAC_L^{-1} are inside the unit circle (part (c) of Condition 1). However, $CA^{-1}B$ is not necessarily positive semi-definite as a conventional variance-covariance matrix would be.

I find that the matrix D is related to the matrix C by (19) and that C can be characterized from the solution to a companion Sylvester matrix equation given by (18) that depends on the known matrices (F, G) and (A, B). I derived the VARMA$(1, 1)$ representation for Z_t given by (15) assuming that parts (a)-(b) of Condition 1 hold. Then, imposing the cross-equation restrictions that arise from theory (implied by (18) and (19)) on (15), I can show that the more complex (and harder to estimate) VARMA$(1, 1)$ dynamics of the endogenous variables Z_t can be further simplified to take the form of a simpler VAR(1) process—part (c) of Condition 1, in turn, is added to ensure the stationarity of the autoregressive part of the VAR(1) process in (22). In other words, if I impose the theoretical cross-equation restrictions implied by (18) and (19) on (15), the MA part of the VARMA representation drops out and I obtain the exact same VAR(1) form given by (22).

Under (19), the existence (and the properties) of the matrix D ($m \times k$ matrix) follow naturally from the existence and properties of the matrix C that solves the companion Sylvester matrix equation given by condition (18). For instance, given the assumptions made in part (a) and (b) of Condition 1 and the cross-equation restrictions in equation (19), the matrix D would naturally be left invertible (with $D_L^{-1} = B^{-1}AC_L^{-1}$). Hence, part (d) of Condition 1 is satisfied by construction. Furthermore, given that $D_L^{-1} = B^{-1}AC_L^{-1}$, it follows that $\left(A - BD_L^{-1}C\right) = (A - A) = \mathbf{0}_k$. Hence, the 'poor man's invertibility condition' which requires all the eigenvalues of $\left(A - BD_L^{-1}C\right)$ to be strictly less than one in modulus (part (e) of Condition 1) also holds by construction in this case.[10] In summary, the characterization of the finite-order VAR solution in (22) depends on parts (a)-(c) of Condition 1 and also on the invertibility of the $m \times m$ matrix $\Gamma_0 \equiv (\mathbf{I}_m - \Phi_2\Theta)$, as indicated before. Moreover, parts (d) and (e) of Condition 1, which are employed to derive the unconstrained VAR(∞) representation in (12), follow from those same conditions (parts (a)-(c) of Condition 1).

Rewriting condition (18), the characterization of the finite-order VAR solution for the forward-looking part of the canonical first-order LRE model given by (22) can be obtained under the following terms:

Lemma 2 *If $\Gamma_0 \equiv (\mathbf{I}_m - \Phi_2\Theta)$ is invertible and parts (a)-(c) of Condition 1 hold, a VAR(1) representation of the solution to the canonical first-order (purely forward-looking) expectational difference system of equations in (7) can be obtained by solving a companion Sylvester matrix equation in C ($m \times k$ matrix):*

$$FCA - C = H, \tag{23}$$

---

[10] A similar case can be made for the conditions that Ravenna (2007) proposes to ensure the finite-order representation of the solution to the LRE model. The key difference is that Fernández-Villaverde et al. (2007) and Ravenna (2007) work out their conditions on the linear state-space representation in $(8) - (9)$ generically, while I derive those conditions after imposing the cross-equation restrictions that arise from the theoretical model. While (not surprisingly) my findings are consistent with theirs, my approach derives the finite-order VAR form and checks its existence and uniqueness directly on the known primitives of the model rather than on the matrices that generically describe the state-space representation of the solution to the LRE model.

or,

$$FC + C\left(-A^{-1}\right) = HA^{-1} = -G, \tag{24}$$

where

$$F \equiv \left(\mathbf{I}_m - \Phi_2\Theta\right)^{-1}\Phi_2, \ G \equiv \left(\mathbf{I}_m - \Phi_2\Theta\right)^{-1}\Phi_3, \ H \equiv -GA. \tag{25}$$

If a matrix C that solves (24) exists and parts (a)-(c) of Condition 1 are also satisfied, then the VAR(1) representation of the solution for the vector of endogenous variables Z_t is given by (22).

The proof of this lemma follows directly from the derivation of conditions $(18)-(19)$, as discussed above.

Step 3. The full solution of the canonical (first-order) form of the LRE model in $(1)-(2)$ can be obtained by combining its backward- and forward-looking parts. I.e., $W_t \equiv W_{tb} + W_{tf}$ where the backward-looking part, $W_{tb} \equiv \Theta W_{t-1}$, follows from the solution Θ ($m \times m$ matrix) to the quadratic matrix equation in (5) (as shown in Lemma 1) and the forward-looking part, $W_{tf} \equiv Z_t \equiv W_t - \Theta W_{t-1}$, is characterized by the solution C to the companion Sylvester matrix equation (as stated in Lemma 2) and obtained from the linear state-space solution in $(8)-(9)$ under the cross-equation restrictions on C ($m \times k$ matrix) and D ($m \times k$ matrix) given by $(18)-(19)$. The implication is that the full solution of the canonical first-order LRE model in $(1)-(2)$ is given by a VAR(2) process:

Corollary 1 *Assume a matrix Θ ($m \times m$ matrix) which solves the companion quadratic matrix equation in (5) exists. Assume that the matrix $(\mathbf{I}_m - \Phi_2\Theta)$ ($m \times m$ matrix) is invertible and parts (a)-(c) of Condition 1 hold. Assume the matrix C ($m \times k$ matrix) exists and is related to the known matrices implied by theory (A, B, F, and G) through the companion Sylvester matrix equation in (24) and that matrix D ($m \times k$ matrix) exists and is related to the theoretical restrictions via (19). Then, the VAR(2) representation of the first-order canonical LRE model solution for the vector of endogenous variables W_t is given by:*

$$W_t = \Psi_1 W_{t-1} + \Psi_2 W_{t-2} + \Psi_3 \epsilon_t, \tag{26}$$

where the corresponding $m \times m$ composite coefficient matrices are $\Psi_1 \equiv \left(\Theta + CAC_L^{-1}\right)$, $\Psi_2 \equiv -CAC_L^{-1}\Theta$, and $\Psi_3 \equiv CA^{-1}B$.

The derivation of this corollary follows from: (a) the transformation of the endogenous variables which implies that $Z_t = W_t - \Theta W_{t-1}$; and (b) the VAR(1) representation for Z_t implied under the terms of Lemma 2 (equation (22)). Corollary 1 means that Ψ_3 is a linear transformation of the $k \times k$ matrix B (the variance-covariance matrix of the stochastic process for the forcing variables) where the mapping is determined by the $m \times k$ matrix CA^{-1} (which depends on the solution C to the companion Sylvester matrix equation in (24)).

Going a step further, the k structural shock innovations in the vector ϵ_t can be recovered under certain additional conditions—making the finite-order VAR form implied by (26) consistent with fundamentalness (in the sense of Hansen and Sargent (1980)) under additional conditions. The canonical first-order forward-looking LRE model solution in (22) can be re-written in terms of the innovations as $\epsilon_t = B^{-1}A\left(C_L^{-1}Z_t - AC_L^{-1}Z_{t-1}\right)$ given that A is invertible, B is positive definite (so it is invertible too)[11], and C is left invertible (parts (a) and (b) of Condition 1 hold). Then, undoing the transformation of the variables

---

[11]Every variance-covariance matrix like B is positive semidefinite, but only positive definiteness implies invertibility.

implied by $Z_t = W_t - \Theta W_{t-1}$, I obtain the following expression for the structural shock innovations ϵ_t in terms of the observable k endogenous variables W_t:

$$\epsilon_t = B^{-1}A\left[C_L^{-1}W_t - \left(C_L^{-1}\Theta + AC_L^{-1}\right)W_{t-1} + AC_L^{-1}\Theta W_{t-2}\right]. \tag{27}$$

However, this expression does not necessarily imply that the shocks can be recovered uniquely nor the VAR(2) representation in (26) is unique.

If the matrix C ($m \times k$ matrix) is left invertible (part (b) of Condition 1), then the left inverse matrix C_L^{-1} ($k \times m$ matrix) satisfies that $C_L^{-1}C = \mathbf{I}_k$. From here it follows that C_L^{-1} has $m \times k$ unknown elements while $C_L^{-1}C = \mathbf{I}_k$ imposes only $k \times k$ linear equation restrictions on those unknown elements for any given matrix C. Given that, the linear of equations that characterize the left inverse C_L^{-1} can be viewed as an underdetermined system (having either no solution—if C is not left invertible—or infinitely many—if C is left invertible). A sufficient condition to ensure that the VAR(2) solution in (26) is unique and the recovered shock innovations from (27) are fundamental is to require that the number of endogenous variables that are observable (m) be equal to the number of forcing variables and shock innovations (k). In other words, a sufficient condition for uniqueness of (26) and to ensure fundamentalness in the sense of Hansen and Sargent (1980) is that $m = k$. That implies that the condition $rank(C) = k$ (part (b) of Condition 1)) under $m = k$ would mean that the left inverse C_L^{-1} corresponds to the inverse of the square matrix C (which exists and is unique).

Hence, the fundamentalness of the unique finite-order VAR solution can be summarized as follows:

Corollary 2 *Assume the conditions of Corollary 1 hold such that the solution to the canonical first-order LRE model in $(1) - (2)$ has a VAR(2) representation given by (26). Then, if it is the case that $m = k$, the VAR(2) representation given by (26) is unique, the vector of endogenous variables W_t is fundamental, and the realization of the structural shock innovations ϵ_t can be recovered exactly (and uniquely)—except for the first two observations due to the lags—using the expression in (27).*

If $\Psi_3 \equiv CA^{-1}B$ in (26) is such that $\Psi_3^T\Psi_3$ is positive semi-definite and symmetric, then $\Psi_3^T\Psi_3$ has also a natural interpretation as a standard variance-covariance matrix. This holds true whenever $m = k$ and Ψ_3 is therefore a $k \times k$ square (Golub and van Loan (1996)).[12] Hence, the VAR(2) representation of the first-order canonical LRE model solution for the vector of endogenous variables W_t can be re-written as:

$$W_t = \Psi_1 W_{t-1} + \Psi_2 W_{t-2} + e_t, \ e_t \sim iid\left(0, \Omega\right), \tag{28}$$

where $\Omega \equiv \Psi_3^T\Psi_3$.

The VAR(2) innovations $e_t = \Psi_3\epsilon_t$ in (28) are simply a rotation of the structural shock innovations ϵ_t. Corollary 2 implicitly relies on the idea that if a solution C to the companion Sylvester matrix equation in (24) exists, this solution will be unique. Indeed, that is the case—I discuss the characterization and uniqueness of the solution to the companion Sylvester matrix equation more extensively in the next subsection. Otherwise, the proof of Corollary 2 follows directly from the simple algebraic manipulations of the finite-order VAR(2) representation described above. I should note again that these results can be further generalized to richer

---

[12] To test whether $\Psi_3^T\Psi_3$ has the property of being positive semi-definite (and symmetric) as a conventional variance-covariance matrix, one can use the *chol* function in Matlab. If *chol* returns a second argument that is zero from $[R, p] = chol\left(\Psi_3^T\Psi_3\right)$, then the matrix is symmetric and—in this case—also positive definite.

LRE models including more than one lead and one lag of the endogenous and forcing variables as those more general environments can be easily re-cast into the canonical first-order system in $(1) - (2)$, as shown in Broze et al. (1985) and Broze et al. (1990) (see also the Appendix).

2.3   The Companion Sylvester Matrix Equation

In this paper I show how to solve a large class of LRE models that can be cast into the canonical first-order form given by $(1) - (2)$ and I establish the conditions under which the LRE model solution admits a finite-order VAR representation. The solution to the backward-looking part on which the canonical first-order LRE model can be decoupled is well-studied in Binder and Pesaran (1995) and Binder and Pesaran (1997). Therefore, here I focus on the characterization of the solution C to the companion Sylvester matrix equation in (24) from which the matrix C ($m \times k$ matrix, $m \geq k$) can be obtained.

Equation (24) proposes a companion Sylvester matrix equation—i.e., $FC + C\left(-A^{-1}\right) = -G$ with $F \equiv \left(\mathbf{I}_m - \Phi_2\Theta\right)^{-1}\Phi_2 \in \mathbb{R}^{m \times m}$, $A \in \mathbb{R}^{k \times k}$, $G \equiv \left(\mathbf{I}_m - \Phi_2\Theta\right)^{-1}\Phi_3 \in \mathbb{R}^{m \times k}$ given and $C \in \mathbb{R}^{m \times k}$ to be determined. The companion Sylvester matrix equation is well-known in stability and control theory and its applications.[13] Using the Kronecker (tensor) product notation and the properties of the vectorization operator, vec, I can re-write the companion Sylvester's matrix equation in its standard form as a linear system of equations:

$$\mathcal{A}vec\left(C\right) = vec\left(-G\right), \tag{29}$$

$$\mathcal{A} := \left[\left(\mathbf{I}_k \otimes F\right) + \left(\left(-A^{-1}\right)^T \otimes \mathbf{I}_m\right)\right], \tag{30}$$

where $\otimes$ denotes the Kronecker product.[14] In this way, the companion Sylvester matrix equation is represented by a linear system of dimension $(m \cdot k) \times (m \cdot k)$ conformed by $m \cdot k$ equations in $m \cdot k$ unknowns (where the unknowns correspond to the elements of the matrix C).

Having transformed the companion Sylvester matrix equation into the linear system given by $(29) - (30)$, well-known matrix algebra results suffice to determine the following criteria for the existence and uniqueness of a solution C to the companion Sylvester matrix equation:

Proposition 1 *Let $F \in \mathbb{R}^{m \times m}$, $A \in \mathbb{R}^{k \times k}$, $G \in \mathbb{R}^{m \times k}$. Then, it follows that:*

(a) (Existence) The Sylvester equation in (24) has at least one solution $C \in \mathbb{R}^{m \times k}$ if and only if $rank\left[\mathcal{A}\,vec\left(-G\right)\right] = rank\left[\mathcal{A}\right]$.

(b) (Uniqueness) The Sylvester equation in (24) has a unique solution $C \in \mathbb{R}^{m \times k}$ if and only if $rank\left[\mathcal{A}\right] = m \cdot k$. That is, the solution is unique if and only if $\mathcal{A}$ has full rank. Then, $\mathcal{A}$ is nonsingular and invertible implying that the unique solution to the Sylvester matrix equation can be recovered as $vec\left(C\right) = \mathcal{A}^{-1}vec\left(-G\right)$.

Proof.   (a) Trivially it follows that $rank\left[\mathcal{A}\,vec\left(-G\right)\right] \geq rank\left[\mathcal{A}\right]$. If there is a solution $vec\left(C\right) = \left[\bar{c}_1 \ \bar{c}_2 \ ... \ \bar{c}_{k^2}\right]^T$ for the linear system given by $(29)-(30)$, then $\sum_{i=1}^{m \cdot k} \mathcal{A}_{*i}\bar{c}_i = vec\left(-G\right)$ where $\mathcal{A}_{*1}, \mathcal{A}_{*2}, ..., \mathcal{A}_{*k^2}$ denote the corresponding columns of the matrix $\mathcal{A}$. Hence, $vec\left(-G\right)$ is a linear combination of the columns

---

[13]Useful references on the characterization of the solution to the Sylvester matrix equation include Chapter 12 of Lancaster and Tismenetsky (1985),, Horn and Johnson (1991), and Jiang and Wei (2003).

[14]The vectorization operation is linear, i.e. $vec\left(\alpha X + \beta Y\right) = \alpha vec\left(X\right) + \beta vec\left(Y\right)$ for any $X, Y \in \mathbb{R}^{i \times j}$ and $\alpha, \beta \in \mathbb{R}$. Proposition 4 in Chapter 12.2 of Lancaster and Tismenetsky (1985) shows that the vectorization operation is closely related to the Kronecker product as follows: If $X \in \mathbb{R}^{i \times i}, Y \in \mathbb{R}^{i \times j}, Z \in \mathbb{R}^{j \times i}$, then $vec\left(XYZ\right) = \left(Z^T \otimes X\right)vec\left(Y\right)$.

of $\mathcal{A}$ and, as a result, the rank of the augmented matrix $[\mathcal{A}\ vec\,(-G)]$ cannot be different than the rank of $\mathcal{A}$—because for $rank\,[\mathcal{A}\ vec\,(-G)] > rank\,[\mathcal{A}]$ to be true, $vec\,(-G)$ needs to be linearly independent from the columns of $\mathcal{A}$ and that contradicts the notion that $vec\,(C)$ is a solution.

(b) If the square matrix $\mathcal{A}$ has full rank—and, therefore, is nonsigular and invertible—the linear system in $(29)-(30)$ has a unique solution given by $vec\,(C) = \mathcal{A}^{-1}vec\,(-G)$. The converse statement follows naturally as well. If the linear system has a unique solution, then $\mathcal{A}$ has full rank and is nonsingular. Otherwise, at least one column in $\mathcal{A}$ is not linearly independent from the rest of the columns and can be written as a linear combination of them. Hence, for any given solution $vec\,(C)$ defined over the linearly independent columns, another different solution exists including non-trivially the linearly dependent columns of $\mathcal{A}$. The existence of more than one solution then contradicts the uniqueness assumption. $\blacksquare$

Proposition 1 characterizes the solution C ($m \times k$ matrix, $m \geq k$) to the companion Sylvester matrix equation in (24) and, by extension, the solution to the forward-looking part of the canonical first-order LRE model. The two rank conditions stated in this proposition depend solely on the properties of the matrices $F \in \mathbb{R}^{m \times m}$, $A \in \mathbb{R}^{k \times k}$, $G \in \mathbb{R}^{m \times k}$ that describe the structural relationships of the LRE model. The uniqueness rank condition implies a solution of the form $vec\,(C) = \mathcal{A}^{-1}vec\,(-G)$ for the companion Sylvester matrix equation and, naturally, that proves existence too. If the uniqueness rank condition is violated, the existence rank condition determines whether there is no solution to the companion Sylvester matrix equation—if $rank\,[\mathcal{A}\ vec\,(-G)] \neq rank\,[\mathcal{A}]$—or whether multiple solutions exist—if $rank\,[\mathcal{A}\ vec\,(-G)] = rank\,[\mathcal{A}] < m \cdot k$. In the latter case, it can be shown that the number of linearly independent solutions is determined by the dimension of the kernel of $\mathcal{A}$. The characterization of the solution to the companion Sylvester matrix equation whenever the uniqueness condition is violated can be found in Theorem 12.5.1, in Theorem 12.5.2, and in Corollary 12.5.1 of Lancaster and Tismenetsky (1985).

Focusing on the case of interest for this paper where a solution to (24) exists and is unique, the full rank condition on $\mathcal{A}$ can be expressed in terms of the eigenvalues of F and A as follows:

Proposition 2 *Let $F \in \mathbb{R}^{m \times m}$, $A \in \mathbb{R}^{k \times k}$ be given. Let $\lambda_1, ..., \lambda_m$ be the eigenvalues of F and $\mu_1, ..., \mu_k$ the eigenvalues of A. Then, for any matrix $G \in \mathbb{R}^{m \times k}$, it follows that the companion Sylvester matrix equation in (24) has a unique solution if and only if $\lambda_i \mu_j \neq 1$ for all $i = 1, ..., m$ and all $j = 1, ..., k$. In other words, the companion Sylvester matrix equation has a unique solution $C \in \mathbb{R}^{m \times k}$ if and only if the matrices F and A^{-1} have no eigenvalues in common.*

Proof. The eigenvalues of A^{-1} are the same as those of its transpose. Given that and the properties of the Kronecker product, the eigenvalues of $\mathcal{A} := \left[(\mathbf{I}_k \otimes F) + \left(\left(-A^{-1}\right)^T \otimes \mathbf{I}_m \right) \right]$ are simply the $m \cdot k$ numbers $\lambda_i - \frac{1}{\mu_j}$ for all $i = 1, ..., m$ and all $j = 1, ..., k$. By Proposition 1, the existence and uniqueness of a solution to the companion Sylvester matrix equation in (24) requires $\mathcal{A}$ to be nonsingular (and have full rank). The matrix $\mathcal{A}$ is nonsingular if and only if all its eigenvalues are nonzero, i.e., if and only if $\lambda_i - \frac{1}{\mu_j} \neq 0$ for all $i = 1, ..., m$ and all $j = 1, ..., k$. Re-arranging the nonzero conditions on the eigenvalues, it follows that $\lambda_i \neq \frac{1}{\mu_j}$ for all $i = 1, ..., m$ and all $j = 1, ..., k$. Given that the eigenvalues of A^{-1} are $\frac{1}{\mu_1}, ..., \frac{1}{\mu_k}$ while those of F are $\lambda_1, ..., \lambda_m$, a unique solution is said to exist if and only if the matrices F and A^{-1} have no eigenvalues in common. $\blacksquare$

According to Proposition 2, the companion Sylvester matrix equation has a unique solution C for each matrix G if and only if F and A^{-1} have no eigenvalues in common. Alternatively, the Sylvester matrix operator $S : \mathbb{R}^{m \times k} \to \mathbb{R}^{m \times k}$ can be defined as $S\,(C) = FC + C\left(-A^{-1}\right)$, where $F \in \mathbb{R}^{m \times m}$ and $A \in \mathbb{R}^{k \times k}$

13

are given and $C \in \mathbb{R}^{m \times k}$ is the solution to be recovered. Then, the Sylvester matrix equation in (24) can simply be written as $S(C) = -G$ for any given matrix $G \in \mathbb{R}^{m \times k}$. The $m \cdot k$ eigenvalues of the Sylvester operator $S(C)$ are $\lambda_i - \frac{1}{\mu_j}$, for all $i = 1, ..., m$ and all $j = 1, ..., k$, where $\lambda_1, ..., \lambda_m$ are the eigenvalues of F and $\mu_1, ..., \mu_k$ are the eigenvalues of A.

Let v_i be the right eigenvector of F associated with the eigenvalue λ_i such that $F v_i = \lambda_i v_i$ for all $i = 1, ..., m$. Let w_j be the left eigenvector of A^{-1} associated with the corresponding eigenvalue $\frac{1}{\mu_j}$ such that $w_j^T A^{-1} = \frac{1}{\mu_j} w_j^T$ for all $j = 1, ..., k$. Then, for any $i = 1, ..., m$ and any $j = 1, ..., k$, $C = v_i w_j^T$ is an eigenvector matrix of the Sylvester matrix operator $S(C)$ associated with its eigenvalue $\lambda_i - \frac{1}{\mu_j}$. It follows from here that the Sylvester matrix operator can be expressed as:

$$\begin{aligned} S(C) = FC - CA^{-1} &= F\left(v_i w_j^T\right) - v_i w_j^T A^{-1} \\ &= (F v_i) w_j^T - v_i \left(w_j^T A^{-1}\right) = (\lambda_i v_i) w_j^T - v_i \left(\frac{1}{\mu_j} w_j^T\right) \\ &= \left(\frac{\lambda_i \mu_j - 1}{\mu_j}\right) v_i w_j^T = \left(\frac{\lambda_i \mu_j - 1}{\mu_j}\right) C = \left(\lambda_i - \frac{1}{\mu_j}\right) C. \end{aligned} \tag{31}$$

Hence, the Sylvester matrix operator $S(C)$ must be nonsingular whenever $\lambda_i \neq \frac{1}{\mu_j}$ for all $i = 1, ..., m$ and all $j = 1, ..., k$. In other words, $S(C)$ is nonsingular if and only if the solution to the companion Sylvester equation in (24) exists and is unique.

Imposing additional restrictions on the eigenvalues of F, an explicit form of the solution C for the companion Sylvester matrix equation in (24) can be obtained as follows:

Proposition 3 *Let $F \in \mathbb{R}^{m \times m}$, $A \in \mathbb{R}^{k \times k}$ where $\lambda_1, ..., \lambda_m$ are the eigenvalues of F and $\mu_1, ..., \mu_k$ are the eigenvalues of A. Then, for any matrix $G \in \mathbb{R}^{m \times k}$, it follows that the companion Sylvester matrix equation in (24) has a unique solution whenever $\lambda_i \mu_j < 1$ for all $i = 1, ..., m$ and all $j = 1, ..., k$ and this solution is given by:*

$$C = \sum_{s=0}^{\infty} F^s (GA) A^s. \tag{32}$$

Proof. I define the following recursion: $FC_{r-1} - C_r A^{-1} = -G$ for iterations $r = 1, 2, 3, ...$ with the initial condition $C_0 = \mathbf{0}_{m \times k}$ where $\mathbf{0}_{m \times k}$ is an $m \times k$ matrix of zeros. If this recursion converges as r goes to infinity, then by construction the limit characterizes the solution of the companion Sylvester matrix equation, i.e., $\lim_{r \to \infty} C_r = \sum_{s=0}^{\infty} F^s (GA) A^s = C$. The convergence condition is equivalent to $\lim_{s \to \infty} F^s (GA) A^s = 0$. It follows that any eigenvalue of $F^s (GA) A^s$ must be proportional to the $s-$power of the product between the eigenvalues of F and A, i.e. $(\lambda_i \mu_i)^s$ for any $i = 1, ..., m$ and any $j = 1, ..., k$. Hence, if all cross-products between the eigenvalues of F and A are strictly less than one, the corresponding eigenvalues for $F^s (GA) A^s$ must go to zero in the limit as $s \to \infty$ and this suffices to show that the recursion indeed converges (Lancaster and Tismenetsky (1985), Chapter 12.3). ∎

Proposition 3 implies that whenever the product of the spectral radii of the matrices A and F is strictly less than one, a unique solution C exists that takes the special form of an infinite sum. Hence, this special case permits the straightforward computation of the solution C to the companion Sylvester matrix equation in (24) via a recursion on the convergent sequence suggested by the proof of Proposition 3. However, often a numerical solution rather than one in closed-form form is all that is needed and that is what I use in this paper (see Anderson et al. (1996) for a discussion of numerical methods with applications in economics).

14

Even when a solution $C \in \mathbb{R}^{m \times k}$ to the companion Sylvester matrix equation in (24) exists and is unique, characterizing the solution to the forward-looking part of the LRE model with a finite-order VAR form as in (22) requires C to have a left inverse C_L^{-1} (parts (a)-(c) of Condition 1). Proving the existence of C and its uniqueness under the terms of Proposition 1 and Proposition 2 does not suffice to ensure C is also left invertible. In fact, as indicated in the previous subsection, either there is no left inverse C_L^{-1} or there are infinitely many matrices C_L^{-1} that satisfy the definition of left inverse for the matrix C whenever $m > k$. The left inverse C_L^{-1} exists and is unique only when $m = k$ and C is invertible ($rank\,(C) = k$) according to the invertible matrix theorem (Strang (2016)). Hence, in summary, the following condition must hold in order to ensure that the left inverse of C exists and is unique:

Condition 2 *Assume the conditions stated in Proposition 1 and Proposition 2 on $F \in \mathbb{R}^{m \times m}$ and $A \in \mathbb{R}^{k \times k}$ (where $m \geq k$) are satisfied so that a unique solution C for the companion Sylvester matrix equation in (24) exists. Then, for a given matrix $G \in \mathbb{R}^{m \times k}$, the solution C is said to have a unique left inverse if $m = k$ and $rank\,(C) = k$.*

Condition 2 is straightforward and follows directly from the general terms of the invertible matrix theorem (Strang (2016)). Results in stability and control theory connect the properties of the matrices $F \in \mathbb{R}^{m \times m}$, $A \in \mathbb{R}^{k \times k}$, and $G \in \mathbb{R}^{m \times k}$ to the invertibility of the unique solution C for the companion Sylvester matrix equation on a related rank identity condition (see Lancaster and Tismenetsky (1985), Chapter 12.5, on Roth's removal theorem, and Lin and Wimmer (2011) on the rank identity condition). I leave for future research the full exploration of those connections. The reason for this is purely practical. If a solution exists and is unique according to the conditions stated in Proposition 1 and Proposition 2, then computing the matrix C is all that is needed to describe the solution to the canonical forward-looking part of the first-order LRE model given by the linear state-space representation in $(8) - (9)$ under the theoretical cross-equation restrictions implied by $(18) - (19)$. Then, it is straightforward to check whether $m = k$ and $rank\,(C) = k$ for the left inverse of C to exist and be unique (Condition 2). If that rank condition is satisfied, it follows that a unique finite-order VAR representation for the canonical first-order LRE model exists given by equation (26) (Corollary 1). In that case, the VAR specification also permits the recovery of the k structural shock innovations underlying the model ϵ_t from the observed k data series in W_t—fundamentalness holds according to Corollary 2.

To wrap up, the methodological contribution of this paper is twofold: First, I propose a new approach to characterize the solution to the forward-looking part of the LRE model in $(8) - (9)$ from the solution of a companion Sylvester equation in (24) and to obtain the backward-looking part of the LRE model from the solution of a companion quadratic matrix equation in (5). Second, I derive testable conditions under which a finite-order VAR representation is exact rather than an approximation for the solution to the canonical first-order LRE model and the shock innovations recovered from the data are fundamental in the sense of Hansen and Sargent (1980). The method is very efficient at checking those conditions while simultaneously deriving the corresponding finite-order VAR representation for the solution (when one such solution exists). Moreover, the first-order canonical LRE model studied here can be further generalized to much richer settings—as discussed elsewhere in the paper (see the Appendix).

15

3   The Workhorse New Keynesian Model

Popular identifying assumptions in structural VAR studies impose certain restrictions on how monetary policy shocks (or other structural shocks for that matter) affect macroeconomic variables. The standard Cholesky assumption, for instance, is known to severely distort the impulse response functions, and can contribute to the so-called "price puzzle" and to obtain more muted responses of inflation and the output gap in response to monetary policy shocks even when the *true* data-generating process is the New Keynesian model (Carlstrom et al. (2009)).[15]

In this section, I adopt a simplified version of the model of Smets and Wouters (2003) that focuses on the role of nominal rigidities in the goods market (price stickiness) and nests the specification studied by Carlstrom et al. (2009). I impose the cross-equation restrictions that arise from theory—from this variant of the workhorse New Keynesian model—bit by bit. And, at each step, I use the toolkit developed in the paper in order to characterize the finite-order VAR representation of the solution. Doing so, I recover the structural monetary policy shocks through the lens of the New Keynesian model without relying on purely empirical identifying assumptions that are often misspecified and can distort our inferences. This application illustrates the practical significance of mapping LRE solutions into finite-order VARs not just for theoretical, but also for applied research. In the process, this exercise also contributes to deepen our understanding of the propagation mechanism for monetary policy shocks and the dynamics of inflation in the workhorse New Keynesian model.

A Univariate Model of Inflation: The Hybrid Phillips Curve.   The hybrid Phillips curve with backward- and forward-looking components, arises from the well-known Calvo (1983)-type model of price-setting behavior with indexation developed by Yun (1996), features prominently in the New Keynesian literature. The hybrid Phillips curve can be specified generically as:

$$\pi_t = \gamma_f \mathbb{E}_t\left(\pi_{t+1}\right) + \gamma_b \pi_{t-1} + e_t, \tag{33}$$

where π_t is the inflation rate, and $\mathbb{E}_t\left(\pi_{t+1}\right)$ is the expected inflation rate next period. The parameters $\gamma_f > 0$ and $\gamma_b \geq 0$ determine the sensitivity to inflation expectations (the forward-looking part) and to lagged inflation (the backward-looking part) and satisfy that $\gamma_f + \gamma_b \leq 1$. The variable e_t refers to the exogenous real marginal cost which is assumed to evolve according to a given first-order autoregressive process, i.e.,

$$e_t = \rho_e e_{t-1} + \sigma_\delta \delta_t, \tag{34}$$

where δ_t is i.i.d. white noise with mean zero and variance of one. The persistence parameter $-1 < \rho_e < 1$ is expected to be less than one in absolute value to ensure stationarity, while the parameter $\sigma_\delta > 0$ pins down the volatility of the real marginal cost shock e_t.

The simple inflation model given by the system in $(33) - (34)$ consists of just one endogenous variable, π_t, and one forcing variable, e_t. Hence, it is not difficult to obtain a closed-form solution for inflation in this case and to characterize it analytically in autoregressive form. Using the notation introduced in Section 2, the model-implied relationship between the vector of endogenous variables $W_t = \left(\pi_t\right)$ and the vector of

---

[15]The "price puzzle" refers to the paradoxical result that an empirically identified monetary shock leads to an increase in inflation which seems counterintuitive based on theory.

forcing variables $X_t = (e_t)$ can be cast in the first-order form of the LRE model given by $(1) - (2)$ with 1×1 composite matrices of the form $\Phi_1 = (\gamma_b)$, $\Phi_2 = (\gamma_f)$, $\Phi_3 = (1)$, $A = (\rho_e)$ and $B = (\sigma_\delta)$.

I split the solution of the model given in $(33) - (34)$ into a backward-looking part and a forward-looking part. From the quadratic matrix equation (5) in Lemma 1 applied to this example, I find that the decoupling depends on the roots of the following characteristic equation:

$$\theta^2 - \frac{1}{\gamma_f}\theta + \frac{\gamma_b}{\gamma_f} = 0, \tag{35}$$

i.e., $\theta_1 \equiv \frac{1 - \sqrt[2]{1 - 4\gamma_f \gamma_b}}{2\gamma_f}$ and $\theta_2 \equiv \frac{1 + \sqrt[2]{1 - 4\gamma_f \gamma_b}}{2\gamma_f}$. The solution $\Theta = (\theta_1)$ that permits splitting the backward- and forward-looking parts of the model requires the existence of a stable eigenvalue that lies within the unit circle, i.e., Θ exists if and only if $|\theta_1| < 1$. As can be easily seen, the existence of the solution Θ depends solely on the parameters γ_f and γ_b. If such a solution exists, then the transformed endogenous variable $Z_t \equiv W_t - \Theta W_{t-1} = (\pi'_t)$ takes the following form: $\pi'_t = \pi_t - \theta_1 \pi_{t-1}$ where θ_1 is the corresponding stable (real-valued) root of the quadratic equation.

As indicated by equation (6) before, the forward-looking part of the hybrid Phillips curve model for $Z_t = (\pi'_t)$ becomes:

$$\Gamma_0 \pi'_t = \Gamma_1 \mathbb{E}_t \left[\pi'_{t+1} \right] + \Gamma_2 e_t, \tag{36}$$

where $\Gamma_0 \equiv (1 - \gamma_f \theta_1)$, $\Gamma_1 \equiv (\gamma_f)$, and $\Gamma_2 \equiv (1)$ are conforming 1×1 matrices, and the forcing variable e_t remains untransformed. It follows from the properties of the roots of the quadratic equation in (35) that $1 - \gamma_f \theta_1 = \gamma_f \theta_2$. Hence, so long as θ_2 is different from zero, the 1×1 matrix Γ_0 is invertible and the canonical forward-looking part of the LRE model can be re-expressed as in equation (7), i.e.,

$$\pi'_t = F \mathbb{E}_t \left[\pi'_{t+1} \right] + G e_t, \tag{37}$$

where $F \equiv \left((\gamma_f \theta_2)^{-1} \gamma_f \right) = \left((\theta_2)^{-1} \right)$ and $G \equiv \left((\gamma_f \theta_2)^{-1} \right)$. From the Blanchard and Kahn (1980) conditions applied to the system in (34) and (37), it is straightforward to show that a solution to the canonical LRE model exists and is unique if and only if $|\theta_2| > 1$.

All of this ultimately implies that the full-fledged LRE model in $(33) - (34)$ can be split into a backward- and a forward-looking part and solved uniquely if and only if the roots of the quadratic equation in (35) satisfy that $|\theta_1| < 1$ and $|\theta_2| > 1$. Then, given equation (23) of Lemma 2, the companion Sylvester matrix equation for this model is given by $FCA - C = H$ where $F = \left(\frac{1}{\theta_2} \right)$, $A = (\rho_e)$, and $H \equiv -GA = \left(-\left(\frac{\rho_e}{\gamma_f \theta_2} \right) \right)$. The solution to this matrix equation gives $C = \left(\frac{1}{\gamma_f} \left(\frac{\rho_e}{\theta_2 - \rho_e} \right) \right)$ which is well-conditioned and invertible if and only if $\rho_e \neq 0$ and $\theta_2 \neq \rho_e$. Therefore, the closed-form solution of the canonical forward-looking part of the LRE model under rational expectations maps the real marginal cost shocks into the transformed endogenous variables as in equation (21) above and can be expressed as:

$$\pi'_t = \pi_t - \theta_1 \pi_{t-1} = \frac{1}{\gamma_f} \left(\frac{1}{\theta_2 - \rho_e} \right) e_t, \tag{38}$$

which, together with the autoregressive process specification give in (34), fully describes the inflation dynamics implied by this univariate inflation model based on the hybrid Phillips curve.

Finally, I can infer the dynamics of the transformed inflation rate in autoregressive form as in (22) in the following form:

$$\pi_t = (\theta_1 + \rho_e)\pi_{t-1} - \rho_e\theta_1\pi_{t-2} + \frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)\sigma_\delta\delta_t, \tag{39}$$

where $\theta_1 \equiv \frac{1 - \sqrt[2]{1-4\gamma_f\gamma_b}}{2\gamma_f}$ and $\theta_2 \equiv \frac{1 + \sqrt[2]{1-4\gamma_f\gamma_b}}{2\gamma_f}$.[16] From an economic point of view, this solution highlights the importance of the backward-looking component of the hybrid Phillips curve to understand the dynamics of inflation. The persistence of the inflation process is not solely determined by the persistence of the exogenous real marginal cost shock, ρ_e, but it also depends on the root θ_1 which is a composite of the backward-looking and forward-looking coefficients of the hybrid Phillips curve (that is, a composite of γ_f and γ_b). The persistence to a univariate innovation obtained as the sum of the autoregressive coefficients in (22) implies that adding a backward-looking component through γ_b tends to increase the inflation persistence—since $\rho_e \leq (1 - \theta_1)\rho_e + \theta_1 \leq 1$ whenever it holds that $\gamma_f > 0$, $\gamma_b \geq 0$, and $\gamma_f + \gamma_b \leq 1$ (which implies that $0 \leq \theta_1 \leq 1$). Similarly, inflation volatility is not solely determined by the volatility of the shock innovations, σ_δ, but it depends on the root θ_2 as well (a composite of γ_f and γ_b). Moreover, it follows that $\frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)$ increases with γ_b if $\gamma_f > 0$, $\gamma_b \geq 0$, $\gamma_f + \gamma_b \leq 1$, and $-1 < \rho_e < 1$.

The closed-form solution of the univariate hybrid Phillips curve model in (39) shows that it is possible to characterize the solution to an LRE model in finite-order autoregressive form. That, in turn, permits the identification of the fundamental economic shock innovations, δ_t, forcing inflation in the univariate model. Furthermore, this also enhances our understanding of the shock propagation mechanism and its effects—for instance, a period of declining persistence and volatility on inflation such as the Great Moderation (Martínez-García (2018)) can result from a downward shift in the backward-looking component of the hybrid Phillips curve (a decline in γ_b).

A Bivariate Monetary Model of Inflation. The method proposed in this paper provides the tools to generalize the logic behind the result in (39) to a richer setting with more than one endogenous and one forcing variables. For that purpose, a first step is to augment the hybrid Phillips curve-based model given by (33) and (34) by explicitly introducing a monetary policy rule in the determination of inflation. To be more precise, I add the following variant of the Taylor (1993) rule with inertia, i.e.,

$$i_t = \rho_i i_{t-1} + (1 - \rho_i)\psi_\pi\pi_t + \varepsilon_t^m, \tag{40}$$

where the short-term (nominal) policy rate is denoted i_t and the policy inertia is modelled with the parameter $0 \leq \rho_i < 1$. The policy rule responds to deviations of inflation alone under the conventional Taylor principle with the parameter ψ_π set to $\psi_\pi > 1$. Here, the associated monetary policy shock ε_t^m follows an exogenously given first-order autoregressive process of the following form:

$$\varepsilon_t^m = \rho_m\varepsilon_{t-1}^m + \sigma_\xi\xi_t, \tag{41}$$

---

[16] Whenever the Phillips curve is purely forward-looking and $\gamma_b = 0$, I obtain that $\theta_1 = 0$ and inflation is directly described by the first-order autoregressive in (39). Whenever $\gamma_b > 0$, the dynamics of inflation implied by (39) are easily captured with a second-order autoregressive process as in (26). The characteristic quadratic equation associated with this second-order autoregressive process is $\phi^2 - (\theta_1 + \rho_e)\phi - (-\rho_e\theta_1) = 0$, with roots given by $\phi_1 \equiv \frac{1}{2}(\theta_1 + \rho_e) - \frac{1}{2}\sqrt[2]{(\theta_1 + \rho_e)^2 - 4\rho_e\theta_1}$ and $\phi_1 \equiv \frac{1}{2}(\theta_1 + \rho_e) + \frac{1}{2}\sqrt[2]{(\theta_1 + \rho_e)^2 - 4\rho_e\theta_1}$. The properties of the roots of the characteristic quadratic equation imply that $\phi_1 + \phi_2 = (\theta_1 + \rho_e)$ and $\phi_1\phi_2 = \rho_e\theta_1$.

where ξ_t is i.i.d. white noise with mean zero and variance of one, and uncorrelated at all leads and lags with δ_t. The persistence parameter $-1 < \rho_m < 1$ is less than one in absolute value to ensure the stationarity of the process, while the parameter $\sigma_\xi > 0$ pins down the monetary shock volatility.

Then, I define the vector of endogenous variables as $W_t = (\pi_t, i_t)^T$, the vector of forcing variables as $X_t = (e_t, \varepsilon_t^m)^T$, and the vector of innovations as $\epsilon_t = (\delta_t, \xi_t)^T$. The bivariate monetary model of inflation given by (33) and (40) in matrix form, i.e.,

$$
\begin{pmatrix} 1 & 0 \\ -(1-\rho_i)\psi_\pi & 1 \end{pmatrix} \begin{pmatrix} \pi_t \\ i_t \end{pmatrix} = \begin{pmatrix} \gamma_b & 0 \\ 0 & \rho_i \end{pmatrix} \begin{pmatrix} \pi_{t-1} \\ i_{t-1} \end{pmatrix} + \begin{pmatrix} \gamma_f & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbb{E}_t[\pi_{t+1}] \\ \mathbb{E}_t[i_{t+1}] \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e_t \\ \varepsilon_t^m \end{pmatrix},
\tag{42}
$$

can be expressed in the form of (1) as follows:

$$
W_t = \Phi_1 W_{t-1} + \Phi_2 \mathbb{E}_t[W_{t+1}] + \Phi_3 X_t,
\tag{43}
$$

$$
\Phi_1 = \begin{pmatrix} \gamma_b & 0 \\ \gamma_b(1-\rho_i)\psi_\pi & \rho_i \end{pmatrix}, \quad \Phi_2 = \begin{pmatrix} \gamma_f & 0 \\ \gamma_f(1-\rho_i)\psi_\pi & 0 \end{pmatrix}, \quad \Phi_3 = \begin{pmatrix} 1 & 0 \\ (1-\rho_i)\psi_\pi & 1 \end{pmatrix}.
\tag{44}
$$

The shock processes in (34) and (41) can be cast in the form indicated by the matrix equation (2) with conforming matrices A and B given by:

$$
A = \begin{pmatrix} \rho_e & 0 \\ 0 & \rho_m \end{pmatrix}, \quad B = \begin{pmatrix} \sigma_\delta & 0 \\ 0 & \sigma_\xi \end{pmatrix}.
\tag{45}
$$

This constitutes the first-order form of the bivariate monetary model of inflation (equations (1) − (2)).

In order to solve the model, I split its backward- and forward-looking parts as indicated in Lemma 1. To solve the quadratic matrix equation in (5), I construct the following two companion matrices:

$$
D = \begin{bmatrix} 1 & 0 & -\gamma_b & 0 \\ 0 & 1 & -\gamma_b(1-\rho_i)\psi_\pi & -\rho_i \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} \gamma_f & 0 & 0 & 0 \\ \gamma_f(1-\rho_i)\psi_\pi & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},
\tag{46}
$$

and solve the corresponding generalized eigenvalue problem (see the Appendix). Hence, I obtain the following ordered matrix of generalized eigenvalues Q and their associated matrix of eigenvectors V:

$$
Q = \begin{pmatrix} \theta_1 & 0 & 0 & 0 \\ 0 & \rho_i & 0 & 0 \\ 0 & 0 & \theta_2 & 0 \\ 0 & 0 & 0 & \infty \end{pmatrix}, \quad V = \begin{pmatrix} \frac{\theta_1-\rho_i}{(1-\rho_i)\psi_\pi} & 0 & \frac{\theta_2-\rho_i}{(1-\rho_i)\psi_\pi} & 0 \\ \theta_1 & \rho_i & \theta_2 & 1 \\ \frac{\theta_1-\rho_i}{(1-\rho_i)\psi_\pi\theta_1} & 0 & \frac{\theta_2-\rho_i}{(1-\rho_i)\psi_\pi\theta_2} & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix},
\tag{47}
$$

where $\theta_1 = \frac{1-\sqrt[2]{1-4\gamma_f\gamma_b}}{2\gamma_f}$ and $\theta_2 = \frac{1+\sqrt[2]{1-4\gamma_f\gamma_b}}{2\gamma_f}$ are defined exactly as in the univariate case before. The matrices Q and V are already ordered so that the two stable eigenvalues come first.

From here it follows that $Q^1 = \begin{pmatrix} \theta_1 & 0 \\ 0 & \rho_i \end{pmatrix}$ and $V^{21} = \begin{pmatrix} \frac{\theta_1-\rho_i}{(1-\rho_i)\psi_\pi\theta_1} & 0 \\ 1 & 1 \end{pmatrix}$, so the companion quadratic

19

matrix equation (equation (5)) has the following solution:

$$\Theta = V^{21}Q^1 \left(V^{21}\right)^{-1} = \begin{pmatrix} \theta_1 & 0 \\ (1-\rho_i)\,\psi_\pi\theta_1 & \rho_i \end{pmatrix}, \tag{48}$$

which is lower triangular. The solution Θ found in (48) permits splitting the backward- and forward-looking parts of the bivariate model of inflation. To do so, the existence of two eigenvalues that are stable and lie within the unit circle are required, i.e., $|\theta_1| < 1$ and $|\rho_i| < 1$ are required. By construction I already assume that $0 < \rho_i < 1$, so the solution Θ that I seek to characterize depends solely on whether the parameters γ_f and γ_b imply also that $|\theta_1| < 1$. Then, the transformed endogenous variables $Z_t \equiv W_t - \Theta W_{t-1} = (\pi'_t, i'_t)$ take the following form: $\pi'_t = \pi_t - \theta_1\pi_{t-1}$ and $i'_t = i_t - (1-\rho_i)\,\psi_\pi\theta_1\pi_{t-1} - \rho_i i_{t-1}$, where θ_1 is the same stable root as in the univariate case. The transformed short-term interest rate i'_t is adjusted with its own lag as well as with lagged inflation. In turn, the adjustment for the inflation variable π'_t is exactly the same as in the univariate case.

The forward-looking part of the bivariate model of inflation can be expressed in the form of (6) as:

$$\Gamma_0 Z_t = \Gamma_1 \mathbb{E}_t \left[Z_{t+1}\right] + \Gamma_2 X_t, \tag{49}$$

$$\Gamma_0 \equiv (\mathbf{I}_k - \Phi_2\Theta) = \begin{pmatrix} 1 - \gamma_f\theta_1 & 0 \\ -(1-\rho_i)\,\psi_\pi\gamma_f\theta_1 & 1 \end{pmatrix} = \begin{pmatrix} \gamma_f\theta_2 & 0 \\ -(1-\rho_i)\,\psi_\pi\gamma_f\theta_1 & 1 \end{pmatrix}, \tag{50}$$

$$\Gamma_1 \equiv \Phi_2 = \begin{pmatrix} \gamma_f & 0 \\ \gamma_f (1-\rho_i)\,\psi_\pi & 0 \end{pmatrix}, \quad \Gamma_2 \equiv \Phi_3 = \begin{pmatrix} 1 & 0 \\ (1-\rho_i)\,\psi_\pi & 1 \end{pmatrix}. \tag{51}$$

Whenever Γ_0 is nonsingular, the system of structural relationships for the forward-looking part of the bivariate monetary model of inflation implied by (6) can be expressed in the form of (7) as:[17]

$$Z_t = F\mathbb{E}_t \left[Z_{t+1}\right] + GX_t, \tag{52}$$

$$F \equiv (\Gamma_0)^{-1}\Gamma_1 = \begin{pmatrix} \frac{1}{\theta_2\gamma_f} & 0 \\ \frac{\theta_1}{\theta_2}(1-\rho_i)\,\psi_\pi & 1 \end{pmatrix} \begin{pmatrix} \gamma_f & 0 \\ \gamma_f (1-\rho_i)\,\psi_\pi & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\theta_2} & 0 \\ (1-\rho_i)\,\psi_\pi \left(\frac{1}{\theta_2}\right) & 0 \end{pmatrix}, \tag{53}$$

$$G \equiv (\Gamma_0)^{-1}\Gamma_2 = \begin{pmatrix} \frac{1}{\theta_2\gamma_f} & 0 \\ \frac{\theta_1}{\theta_2}(1-\rho_i)\,\psi_\pi & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ (1-\rho_i)\,\psi_\pi & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\theta_2\gamma_f} & 0 \\ (1-\rho_i)\,\psi_\pi \left(\frac{1}{\theta_2\gamma_f}\right) & 1 \end{pmatrix}. \tag{54}$$

The solution of the bivariate model includes the matrices A and B in (45) corresponding to the bivariate stochastic process for the forcing variables and the matrix equation in (9) to describe the mapping between the lagged forcing variables and their innovations into the transformed endogenous variables in the solution of the canonical LRE model. This, in turn, requires the conforming matrices C and D to satisfy the conditions given by $(18)-(19)$.

The matrices C and D are tied to the matrices F, H, and G that arise from the canonical form of the

---

[17]Notice that $\theta_1 + \theta_2 = \frac{1 - \sqrt[2]{1-4\gamma_f\gamma_b}}{2\gamma_f} + \frac{1 + \sqrt[2]{1-4\gamma_f\gamma_b}}{2\gamma_f} = \frac{1}{\gamma_f}$.

forward-looking part of the LRE model in (52) where F and G are given above in (53) $-$ (54) and H is:[18]

$$H \equiv -GA = - \begin{pmatrix} \frac{1}{\theta_2 \gamma_f} & 0 \\ (1-\rho_i)\psi_\pi \left(\frac{1}{\theta_2 \gamma_f}\right) & 1 \end{pmatrix} \begin{pmatrix} \rho_e & 0 \\ 0 & \rho_m \end{pmatrix} = \begin{pmatrix} -\frac{\rho_e}{\theta_2 \gamma_f} & 0 \\ -\frac{\rho_e}{\theta_2}\frac{\psi_\pi}{\gamma_f}(1-\rho_i) & -\rho_m \end{pmatrix}. \tag{55}$$

From Proposition 2, I check the existence and uniqueness of a solution C via the companion Sylvester matrix equation in (23). I compute the eigenvalues of F (that is, $\lambda_1 = \frac{1}{\theta_2}$, $\lambda_2 = 0$) and the eigenvalues of A (that is, $\mu_1 = \rho_e$, $\mu_2 = \rho_m$). Then, given that $\lambda_i \mu_j \neq 1$, for all $i, j = 1, 2$ ensures a solution C to the companion Sylvester matrix equation exists and is unique, I conclude that would be the case if and only if $\theta_2 \neq \rho_e$ and $\theta_2 \neq \rho_m$.

A straightforward manipulation of the 2^2 equations implied by the Sylvester matrix equation characterizes the conforming matrices C and D as follows:

$$C = \begin{pmatrix} \frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)\rho_e & 0 \\ \psi_\pi (1-\rho_i)\frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)\rho_e & \rho_m \end{pmatrix}, \quad D \equiv CA^{-1}B = \begin{pmatrix} \frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)\sigma_\delta & 0 \\ \psi_\pi (1-\rho_i)\frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)\sigma_\delta & \sigma_\xi \end{pmatrix}. \tag{56}$$

Checking Condition 2 is straightforward to see that $rank(C) = 2$ if and only if $\rho_e \neq 0$ and $\rho \neq 0$ since $\theta_2 > 1$ and by assumption $\gamma_f > 0$, $\gamma_b > 0$, and $\psi_\pi > 1$ (the Taylor principle) must hold. All of this, in turn, implies that there exists a unique matrix C that solves the companion Sylvester matrix equation in (23) and is also invertible. Hence, the inverse of C is given as:

$$C^{-1} = \begin{pmatrix} \gamma_f(\theta_2 - \rho_e)\frac{1}{\rho_e} & 0 \\ -\psi_\pi(1-\rho_i)\frac{1}{\rho_m} & \frac{1}{\rho_m} \end{pmatrix}. \tag{57}$$

Therefore, the forward-looking part of the bivariate inflation model has a VAR(1) representation in the form of (22) which can be expressed as:

$$Z_t = CAC^{-1}Z_{t-1} + CA^{-1}B\epsilon_t, \tag{58}$$

where

$$CAC^{-1} = \begin{pmatrix} \rho_e & 0 \\ (\rho_e - \rho_m)\psi_\pi(1-\rho_i) & \rho_m \end{pmatrix}, \quad CA^{-1}B = \begin{pmatrix} \frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)\sigma_\delta & 0 \\ \psi_\pi(1-\rho_i)\frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)\sigma_\delta & \sigma_\xi \end{pmatrix}. \tag{59}$$

Then, the finite-order VAR solution of the full-fledged LRE model in (26) becomes:

$$W_t = \Psi_1 W_{t-1} + \Psi_2 W_{t-2} + \Psi_3 \epsilon_t, \tag{60}$$

---

[18] Computing the eigenvalues of H (which are $-\beta\frac{\rho_e}{\theta_2 \gamma_f}$ and $-\rho$), I find them to be non-zero if and only if $\rho_e \neq 0$ and $\rho \neq 0$ since $\theta_2 > 1$ and given that by assumption $\gamma_f, \beta > 0$ and $\gamma_b \geq 0$. Hence, the matrix H is nonsingular and invertible.

where

$$\Psi_1 \equiv \left(\Theta + CAC^{-1}\right) = \begin{pmatrix} \rho_e + \theta_1 & 0 \\ \psi_\pi \left(1 - \rho_i\right)\left(\rho_e + \theta_1 - \rho_m\right) & \rho_i + \rho_m \end{pmatrix}, \tag{61}$$

$$\Psi_2 \equiv -CAC^{-1}\Theta = \begin{pmatrix} -\rho_e\theta_1 & 0 \\ -\psi_\pi \left(1 - \rho_i\right)\rho_e\theta_1 & -\rho_i\rho_m \end{pmatrix}, \tag{62}$$

$$\Psi_3 \equiv CA^{-1}B = \begin{pmatrix} \frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)\sigma_\delta & 0 \\ \psi_\pi \left(1 - \rho_i\right)\frac{1}{\gamma_f}\left(\frac{1}{\theta_2 - \rho_e}\right)\sigma_\delta & \sigma_\xi \end{pmatrix}, \tag{63}$$

given that $\theta_1 \equiv \frac{1 - \sqrt[2]{1 - 4\gamma_f\gamma_b}}{2\gamma_f}$ and $\theta_2 \equiv \frac{1 + \sqrt[2]{1 - 4\gamma_f\gamma_b}}{2\gamma_f}$.

From an economic point of view, the solution of the bivariate monetary model of inflation presented here shows that there are no spillovers from lagged interest rates into current inflation. Spillovers are only from lagged inflation into the policy rate itself and in that case the policy parameter ψ_π and the policy inertia ρ_i determine the magnitude of the spillover—while the difference between the persistence in the real marginal cost shocks (ρ_e) plus the root θ_1 and the persistence in the monetary policy shocks (ρ_m) influences the sign of the spillover. Hence, the policy parameter ψ_π plays a key role in explaining the contribution of the monetary policy shock innovation relative to that of the real marginal cost shock innovation (the only non-monetary shock) in accounting for the fluctuations of the policy rate. In turn, current monetary policy shocks do not contribute to explain any of the inflation fluctuations.

In other words, monetary policy has no effect on inflation determination in the bivariate LRE model given by (33), (34), (40), and (41). In fact, the solution of inflation is exactly the same as that of the univariate case and could have been derived separately since there are no linkages built into the model between the dynamics of inflation and the policy rate. In other words, there is no endogenous transmission mechanism linking monetary policy (or the monetary policy shocks) to inflation. In this context, a conventional zero-restriction—similar to those imposed under a standard Cholesky identification assumption—can be used to identify the structural shock innovations of the bivariate monetary model of inflation. This is because the exogenous process for real marginal costs alone drives the dynamics of inflation via the hybrid Phillips curve. The key insight for inflation and, more generally, for monetary policy is that richer dynamics—including the possibility of transmission of shock innovations other than real marginal cost shocks onto inflation—depend at least on how the real marginal cost is (partly) endogenized in the hybrid Phillips curve.

The Workhorse New Keynesian Model of Inflation. A further extension of the model that gives monetary policy a distinct role in the determination of real marginal costs and inflation is required. To do so, I follow the logic of the workhorse New Keynesian model of Smets and Wouters (2003) to partly endogenize the real marginal costs and to connect them explicitly to monetary policy actions. For that purpose, I adopt a simplified variant of Smets and Wouters (2003) that retains nominal rigidities in goods markets (price stickiness) as the key modeling friction. I abstract from capital assuming a linear-in-labor technology, I remove entirely all labor market frictions assuming perfectly competitive and flexible wage-setting behavior, and I also exclude other ancillary features of the Smets and Wouters (2003) model such as

government expenditure shocks.[19] This workhorse model nests the one-equation and two-equation systems presented earlier, but it also encompasses the three-equation set-up considered by Carlstrom et al. (2009) (with the important addition of external additive habit formation on consumption).

Households maximize their lifetime discounted utility $\mathbb{E}_0 \sum_{t=0}^{+\infty} \beta^t U_t$ where $0 < \beta < 1$ is the discount factor, subject to a standard per-period budget constraint given by $\frac{1}{1+i_t} B_t = B_{t-1} + P_t (Y_t - C_t)$. Households hold their financial wealth in the form of nominal bonds B_t whose net nominal rate of interest is i_t. P_t is the price of final goods and aggregate final goods output Y_t equates all income sources expressed in units of the final good—including the labor income $(\frac{W_t}{P_t} L_t)$ where W_t are the nominal wages, but also all profits from the final good producer and the intermediate goods firms. The instantaneous utility function U_t is separable in consumption C_t relative to a stock of habit H_t and labor L_t, i.e., $U_t \equiv \varepsilon_t^b \left[\frac{(C_t - H_t)^{1-\sigma_c}}{1-\sigma_c} - \frac{\varepsilon_t^l}{1+\sigma_l} L_t^{1+\sigma_l} \right]$, where $\sigma_c > 0$ determines the coefficient of relative risk aversion (inverse of the intertemporal elasticity of substitution) and $\sigma_l > 0$ is the inverse of the Frisch elasticity of labor supply.

The instantaneous utility on consumption depends additively on an external habit stock $H_t = h \widetilde{C}_{t-1}$ which is proportional to the previous period aggregate consumption $\widetilde{C}_{t-1}$ with the constant of proportionality set by the parameter $0 < h < 1$.[20] Moreover, ε_t^b represents a shock to the discount rate that affects the intertemporal substitution of households (preference shock) and ε_t^l is a shock to the labor supply. They each follow a first-order autoregressive process of the following form:

$$\varepsilon_t^b = \rho_b \varepsilon_{t-1}^b + \sigma_\zeta \zeta_t, \tag{64}$$

$$\varepsilon_t^l = \rho_l \varepsilon_{t-1}^l + \sigma_\nu \nu_t, \tag{65}$$

where ζ_t and ν_t are assumed to be i.i.d. white noise with zero mean and variance of one, and uncorrelated at all leads and lags with all other shocks. The persistence parameters $-1 < \rho_b < 1$ and $-1 < \rho_l < 1$ are less than one in absolute value to ensure the stationarity of the process, while the volatility parameters $\sigma_\zeta > 0$ and $\sigma_\nu > 0$ pin down the preference shock volatility and the labor supply shock volatility.

There is a single final good used solely for household consumption, but a continuum of intermediate goods over the unit interval indexed by j. The final-good sector is perfectly competitive while the production of the intermediate goods takes place under monopolistic competition (a single producer for each intermediate good) and standard Calvo (1983) price-setting behavior. The final good is produced using the intermediate goods with the following technology: $Y_t = \left[\int_0^1 \left(Y_t^j \right)^{\frac{1}{1+\lambda_p}} dj \right]^{1+\lambda_p}$ where Y_t^j is the output of the intermediate good of type j and $1 + \lambda_p \equiv \frac{\theta}{\theta-1}$ is a transformation of the elasticity of substitution among intermediate goods $\theta > 1$. The cost minimization conditions in the final-good sector imply a demand for intermediate

---

[19]A related illustration extending the (closed-economy) workhorse New Keynesian model to a two-country setting can be found in Duncan and Martínez-García (2015).

[20]External habits à la Campbell and Cochrane (1999) are extensively studied in Dennis (2009). Dennis (2009) explores the case where the habit stock H_t obeys $H_t = h \left(c_{t-1}^D C_{t-1}^{1-D} \right)$ with $0 < h < 1$ and $0 \leq D \leq 1$. When $D = 0$, the habit formation is external to the household—it is of the "catching up with the Joneses" type used in Smets and Wouters (2003) and in this paper. Because consumption must be greater than the habit stock, additive habits imply that there is a subsistence level below which a household's consumption cannot fall. In the case considered here the subsistence level is simply a fraction h of the previous period aggregate consumption $\widetilde{C}_{t-1}$. Dennis (2009) studies also the case where habits enter into preferences multiplicatively and establishes that when habit formation is external (i.e., when $D = 0$), additive habits encompass multiplicative habits up to a log-linear approximation. When the habit formation is internal, it is always possible to parameterize a log-linearized model with additive habits to replicate the multiplicative case. The interested reader is referred to Dennis (2009) for the more general treatment of habits in the standard New Keynesian model.

goods of the following form: $Y_t^j = \left(\frac{P_t^j}{P_t}\right)^{-\left(\frac{1+\lambda_p}{\lambda_p}\right)} Y_t$. Moreover, perfect competition in the goods market

means that the final goods price P_t can be expressed as: $P_t = \left[\int_0^1 \left(P_t^j\right)^{-\frac{1}{\lambda_p}} dj\right]^{-\lambda_p}$ where P_t^j is the price of intermediate good j.

Each intermediate good j is produced using a linear in labor-in-labor technology: $Y_t^j = \varepsilon_t^a L_t^j$ where, for simplicity, I abstract from the fixed cost of production introduced in Smets and Wouters (2003). The productivity shock ε_t^a follows a first-order autoregressive process of the following form:

$$\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \sigma_\delta \delta_t, \tag{66}$$

where δ_t is i.i.d. white noise with mean zero and variance of one, and uncorrelated at all leads and lags with all other shocks. The persistence parameter $-1 < \rho_a < 1$ is less than one in absolute value to ensure stationarity, while the volatility parameter $\sigma_\delta > 0$ pins down the volatility of the productivity shock. The productivity shock is related to the real marginal cost shock specification in (34) (as it will become apparent shortly).

The nominal per-period profits of intermediate firm j are given as $Profits_t^j = \left(P_t^j - MC_t\right) \left(\frac{P_t^j}{P_t}\right)^{-\frac{1+\lambda_p}{\lambda_p}} Y_t$, where the marginal cost MC_t is independent of the intermediate good produced and equal to $MC_t = \frac{W_t}{\varepsilon_t^a}$. As in Calvo (1983), intermediate firms are not allowed to reoptimize their prices every period. Following Smets and Wouters (2003), in each period, with probability $0 \le \xi_p < 1$, the intermediate firm cannot reoptimize its price and instead partial indexation is allowed whereby the parameter $0 \le \gamma_p \le 1$ regulates the pass-through from previous period inflation. With probability $1 - \xi_p$, the intermediate firm gets to reoptimize its price. Each reoptimizing intermediate firm at time t maximizes its current and expected future profits under the set price using as discount factor the intertemporal marginal rate of substitution of the owners-households: $\beta \left(\frac{\varepsilon_{t+k}^b (C_{t+k} - H_{t+k})^{1-\sigma_c}}{\varepsilon_t^b (C_t - H_t)^{1-\sigma_c}}\right) \left(\frac{1}{P_{t+k}}\right)$ for any $k > 0$. The price-setting behavior implied by profit maximization is characterized by the following first-order condition: $\mathbb{E}_0 \sum_{i=0}^{+\infty} \beta^i \xi_p^i \varepsilon_{t+i}^b (C_{t+i} - H_{t+i})^{1-\sigma_c} Y_t^j \left(\left(\frac{\widehat{P}_t^j}{P_t}\right) \left(\frac{\left(\frac{P_{t-1+i}}{P_{t-1}}\right)^{\gamma_p}}{\frac{P_{t+i}}{P_t}}\right) - (1 + \lambda_p) \frac{MC_{t+i}}{P_{t+i}}\right) = 0$, where $\widehat{P}_t^j$ is the optimal price chosen by the reoptimizing intermediate firm j at time t. The definition of the price index P_t implies the following law of motion for the final goods price: $(P_t)^{-\frac{1}{\lambda_p}} = \xi_p \left(P_{t-1} \left(\frac{P_{t-1}}{P_{t-2}}\right)^{\gamma_p}\right)^{-\frac{1}{\lambda_p}} + \left(1 - \xi_p\right) \left(\widehat{P}_t^j\right)^{-\frac{1}{\lambda_p}}$, where $1 + \lambda_p$ defines the gross markup under monopolistic competition.

Finally, the nominal interest rate i_t is determined by a linear reaction function that describes monetary policy decisions along the lines of Taylor (1993)—an extension of $(40) - (41)$, as I will indicate shortly. The final goods market is in equilibrium if aggregate production equals household consumption in every period: $Y_t = C_t$. Consistency requires that the aggregate consumption $\widetilde{C}_t$ that sets the reference point for the external habit stock equals household consumption in equilibrium: $\widetilde{C}_t = C_t$. The labor market clears whenever the firms' demand for labor equals labor supply by the households at the prevailing wage rate: $L_t = \int_0^1 L_t^j dj$.

For the empirical exercise that follows, I log-linearize the first-order conditions and feasibility constraints that characterize the solution to the workhorse New Keynesian model around the deterministic steady state, as in Smets and Wouters (2003). The lower-case variables denote the log deviation from steady state of the corresponding variable. Then, the linear rational expectations (LRE) approximation can be described as

follows:

First, the model incorporates a dynamic Investment-Saving (IS) equation with external habit formation—an equation that did not feature in the bivariate model before—given by:

$$y_t = \frac{1}{1+h}\mathbb{E}_t\left(y_{t+1}\right) + \frac{h}{1+h}y_{t-1} - \frac{1-h}{(1+h)\,\sigma_c}\left(i_t - \mathbb{E}_t\left(\pi_{t+1}\right) - (1-\rho_b)\,\varepsilon_t^b\right), \tag{67}$$

where the dynamics of the preference shock ε_t^b given in (64) act as a shifter of the IS equation. Needless to say, whenever $h = 0$, equation (67) collapses to the familiar time-separable dynamic IS equation without habits—which is the specification used in Carlstrom et al. (2009) whose model is nested into the variant of the New Keynesian model of Smets and Wouters (2003) that I explore here.

Second, the model augments the hybrid Phillips curve equation discussed earlier with an endogenous component of the real marginal cost that is proportional to output y_t and an exogenous component that is related to the productivity shock ε_t^a and the labor supply shock ε_t^l, i.e.,

$$\pi_t = \gamma_f \mathbb{E}_t\left(\pi_{t+1}\right) + \gamma_b \pi_{t-1} + \kappa\left[\left(\sigma_l + \sigma_c \frac{1}{1-h}\right) y_t - \sigma_c \frac{h}{1-h}y_{t-1} + \varepsilon_t^l - (1+\sigma_l)\,\varepsilon_t^a\right], \tag{68}$$

where $\gamma_f \equiv \frac{\beta}{1+\beta\gamma_p} > 0$ and $\gamma_b \equiv \frac{\gamma_p}{1+\beta\gamma_p} \geq 0$ satisfying $\gamma_f + \gamma_b \leq 1$ determine the forward- and the backward-looking parts as before. In line with the New Keynesian literature, the coefficient $\kappa \equiv \frac{1}{1+\beta\gamma_p}\frac{\left(1-\beta\xi_p\right)\left(1-\xi_p\right)}{\xi_p} \geq 0$ enters anew into (68) defining the slope of the hybrid Phillips curve equation on real marginal costs. The real marginal cost $e_t \equiv \left(\sigma_l + \sigma_c \frac{1}{1-h}\right) y_t - \sigma_c \frac{h}{1-h}y_{t-1}$ is endogenous—a function of current and lagged output y_t. Abstracting from that endogenous component, the exogenous productivity shock process ε_t^a and the labor supply shock ε_t^l enter as shifters of the hybrid Phillips curve analogous to the role that the exogenous real marginal cost process for e_t in (34) played in the previous one- and two-equation specifications of the model. This exogenous component is generally referred as a cost-push shock (in this context, the cost-push shock is simply a function of the labor supply shock ε_t^l in (65) and of the productivity shock ε_t^a in (66)).

Finally, the workhorse model is completed with a standard Taylor (1993) rule with inertia—augmenting the monetary policy rule in (40) allowing the policy rate i_t to respond to the output gap $(y_t - y_t^n)$ as well as to inflation π_t—as follows:

$$i_t = \rho_i i_{t-1} + (1-\rho_i)\left[\psi_\pi \pi_t + \psi_y\left(y_t - y_t^n\right)\right] + \varepsilon_t^m, \tag{69}$$

where the exogenous process for the monetary policy shock ε_t^m is given as in (41). The corresponding policy parameters satisfy that $\psi_\pi > 1$ and $\psi_y \geq 0$ and the policy inertia is modelled with the parameter $0 \leq \rho_i < 1$. The output gap is defined as the deviation of actual output y_t relative to its potential y_t^n. The potential output y_t^n is the level of economic activity that would prevail absent all frictions (nominal rigidities).[21] A standard log-linearization of the labor supply and labor demand equations around the steady state under

---

[21] As is conventional, monetary policy needs to be supplemented with an optimal labor subsidy to offset the markup $1 + \lambda_p$ charged by intermediate firms under monopolistic competition ensuring the steady state of the model will be identical to the frictionless one (derived under flexible prices and perfect competition). This labor subsidy is financed with a lump-sum tax raised from households. However, this does not affect the short-run dynamics of the approximated model so I retain the Smets and Wouters (2003) specification—which is not explicit about it—as is.

flexible prices ($\xi_p = 0$) yields:

$$y_t^n = \frac{\sigma_c h}{\sigma_l (1-h) + \sigma_c} y_{t-1}^n + \frac{(1-h)(1+\sigma_l)}{\sigma_l (1-h) + \sigma_c} \varepsilon_t^a - \frac{(1-h)}{\sigma_l (1-h) + \sigma_c} \varepsilon_t^l, \tag{70}$$

which is a function of the productivity shocks ε_t^a and the labor supply shocks ε_t^l.

Let me define the vector of endogenous variables as $W_t = (y_t, \pi_t, i_t, y_t^n)^T$, the forcing variables as $X_t = \left(\varepsilon_t^a, \varepsilon_t^b, \varepsilon_t^l, \varepsilon_t^m\right)^T$, and the vector of innovations as $\epsilon_t = (\delta_t, \zeta_t, \nu_t, \xi_t)^T$. The forward-looking part of the three-equation New Keynesian model of inflation given by (67), (68), and (69) can be expressed as:

$$D_0 W_t = D_1 W_{t-1} + D_2 \mathbb{E}_t [W_{t+1}] + D_3 X_t, \tag{71}$$

$$D_0 = \begin{pmatrix} 1 & 0 & \left(\frac{1-h}{(1+h)\sigma_c}\right) & 0 \\ -\kappa\left(\sigma_l + \frac{\sigma_c}{1-h}\right) & 1 & 0 & 0 \\ -(1-\rho_i)\psi_y & -(1-\rho_i)\psi_\pi & 1 & (1-\rho_i)\psi_y \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$D_1 = \begin{pmatrix} \frac{h}{1+h} & 0 & 0 & 0 \\ -\kappa\sigma_c\frac{h}{1-h} & \gamma_b & 0 & 0 \\ 0 & 0 & \rho_i & 0 \\ 0 & 0 & 0 & \frac{\sigma_c h}{\sigma_l(1-h)+\sigma_c} \end{pmatrix}, \quad D_2 = \begin{pmatrix} \frac{1}{1+h} & \left(\frac{1-h}{(1+h)\sigma_c}\right) & 0 & 0 \\ 0 & \gamma_f & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{72}$$

$$D_3 = \begin{pmatrix} 0 & \frac{1-h}{(1+h)\sigma_c}(1-\rho_b) & 0 & 0 \\ -\kappa(1+\sigma_l) & 0 & \kappa & 0 \\ 0 & 0 & 0 & 1 \\ \frac{(1-h)(1+\sigma_l)}{\sigma_l(1-h)+\sigma_c} & 0 & -\frac{(1-h)}{\sigma_l(1-h)+\sigma_c} & 0 \end{pmatrix},$$

and re-written, whenever D_0 is nonsingular, in the form of (1):

$$W_t = \Phi_1 W_{t-1} + \Phi_2 \mathbb{E}_t [W_{t+1}] + \Phi_3 X_t,$$

$$\Phi_1 \equiv (D_0)^{-1} D_1, \ \Phi_2 \equiv (D_0)^{-1} D_2, \ \Phi_3 \equiv (D_0)^{-1} D_3, \tag{73}$$

where

$$(D_0)^{-1} = \begin{pmatrix} \Phi_s & \left(\frac{-(1-h)}{(1+h)\sigma_c}\right)(1-\rho_i)\psi_\pi\Phi_s & -\left(\frac{1-h}{(1+h)\sigma_c}\right)\Phi_s & \left(\frac{1-h}{(1+h)\sigma_c}\right)(1-\rho_i)\psi_y\Phi_s \\ \Phi_s\kappa\left(\sigma_l+\frac{\sigma_c}{1-h}\right) & \left(\frac{(1-h)(1-\rho_i)\psi_y}{(1+h)\sigma_c}+1\right)\Phi_s & \left(\frac{-(\sigma_l(1-h)+\sigma_c)\kappa}{(1+h)\sigma_c}\right)\Phi_s & \left(\frac{(\sigma_l(1-h)+\sigma_c)\kappa(1-\rho_i)\psi_y}{(1+h)\sigma_c}\right)\Phi_s \\ \Phi_s D_\kappa(1-\rho_i) & (1-\rho_i)\psi_\pi\Phi_s & \Phi_s & -(1-\rho_i)\psi_y\Phi_s \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\Phi_s \equiv \frac{1}{1+(1-\rho_i)\left(\frac{1-h}{(1+h)\sigma_c}\right)\left(\psi_\pi\left(\sigma_l+\frac{\sigma_c}{1-h}\right)\kappa+\psi_y\right)}, \quad D_\kappa \equiv \left(\frac{\psi_y(1-h)+(\sigma_l(1-h)+\sigma_c)\kappa\psi_\pi}{1-h}\right). \tag{74}$$

I.e.,

$$
\Phi_1 = \begin{pmatrix}
\Phi_s \frac{h(1+\kappa(1-\rho_i)\psi_\pi)}{1+h} & -\Phi_s \frac{(1-h)(1-\rho_i)\psi_\pi \gamma_b}{(1+h)\sigma_c} & -\Phi_s \frac{(1-h)\rho_i}{(1+h)\sigma_c} & \Phi_s \frac{h(1-h)(1-\rho_i)\psi_y}{(1+h)(\sigma_c+(1-h)\sigma_l)} \\
\Phi_s \frac{h\kappa\left((1-h)\sigma_l - h\sigma_c - (1-h)(1-\rho_i)\psi_y\right)}{(1+h)(1-h)} & \Phi_s \frac{\left((1+h)\sigma_c+(1-h)(1-\rho_i)\psi_y\right)\gamma_b}{(1+h)\sigma_c} & -\Phi_s \frac{(\sigma_c+\sigma_l(1-h))(1-h)\kappa\rho_i}{(1+h)\sigma_c} & \Phi_s \frac{h\kappa(1-\rho_i)\psi_y}{1+h} \\
\Phi_s \frac{h(1-\rho_i)}{(1+h)}\left(D_\kappa - \frac{\sigma_c(1+h)\kappa\psi_\pi}{1-h}\right) & \Phi_s \left(1-\rho_i\right)\psi_\pi \gamma_b & \Phi_s \rho_i & -\Phi_s \frac{h\sigma_c(1-\rho_i)\psi_y}{\sigma_c+(1-h)\sigma_l} \\
0 & 0 & 0 & \frac{h\sigma_c}{\sigma_c+(1-h)\sigma_l}
\end{pmatrix},
$$

$$
\Phi_2 = \Phi_s \begin{pmatrix}
\frac{1}{1+h} & \left(\frac{1-h}{(1+h)\sigma_c}\right)\left(1-(1-\rho_i)\psi_\pi \gamma_f\right) & 0 & 0 \\
\frac{1}{1+h}\left(\sigma_l + \frac{\sigma_c}{1-h}\right)\kappa & \left(\sigma_l + \frac{\sigma_c}{1-h}\right)\kappa\left(\frac{1-h}{(1+h)\sigma_c}\right) + \left(1+(1-\rho_i)\left(\frac{1-h}{(1+h)\sigma_c}\right)\psi_y\right)\gamma_f & 0 & 0 \\
\frac{1}{1+h}(1-\rho_i)\left(\psi_\pi\left(\sigma_l+\frac{\sigma_c}{1-h}\right)\kappa+\psi_y\right) & (1-\rho_i)\left(\left(\frac{1-h}{(1+h)\sigma_c}\right)\left(\left(\sigma_l+\frac{\sigma_c}{1-h}\right)\kappa\psi_\pi+\psi_y\right)+\psi_\pi\gamma_f\right) & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix},
$$

$$
\Phi_3 = \begin{pmatrix}
\Phi_s \frac{D_\kappa(1-h)^2(1+\sigma_l)(1-\rho_i)}{(1+h)\sigma_c(\sigma_c+(1-h)\sigma_l)} & \Phi_s\left(\frac{1-h}{(1+h)\sigma_c}\right)(1-\rho_b) & -\Phi_s \frac{D_\kappa(1-h)(1-h)(1-\rho_i)}{(1+h)\sigma_c(\sigma_c+(1-h)\sigma_l)} & -\Phi_s\left(\frac{1-h}{(1+h)\sigma_c}\right) \\
-\kappa\Phi_s\left(1+\sigma_l\right) & \Phi_s\left(\frac{\sigma_c+(1-h)\sigma_l}{(1+h)\sigma_c}\right)\kappa\left(1-\rho_b\right) & \Phi_s\kappa & -\Phi_s\left(\frac{(\sigma_c+(1-h)\sigma_l)\kappa}{(1+h)\sigma_c}\right) \\
-\Phi_s \frac{D_\kappa(1-h)(1+\sigma_l)(1-\rho_i)}{(\sigma_c+(1-h)\sigma_l)} & \Phi_s \frac{D_\kappa(1-h)(1-\rho_i)(1-\rho_b)}{(1+h)\sigma_c} & \Phi_s \frac{(1-\rho_i)D_\kappa(1-h)}{(\sigma_c+(1-h)\sigma_l)} & \Phi_s \\
\frac{(1-h)(1+\sigma_l)}{\sigma_l(1-h)+\sigma_c} & 0 & -\frac{(1-h)}{\sigma_l(1-h)+\sigma_c} & 0
\end{pmatrix}.
$$

(75)

The shock processes in (64), (65), (66), and (41) can be cast in the form indicated by the matrix equation (8) with conforming matrices A and B given by:

$$
A = \begin{pmatrix}
\rho_a & 0 & 0 & 0 \\
0 & \rho_b & 0 & 0 \\
0 & 0 & \rho_l & 0 \\
0 & 0 & 0 & \rho_m
\end{pmatrix}, \quad B = \begin{pmatrix}
\sigma_\delta & 0 & 0 & 0 \\
0 & \sigma_\zeta & 0 & 0 \\
0 & 0 & \sigma_\nu & 0 \\
0 & 0 & 0 & \sigma_\xi
\end{pmatrix}.
\tag{76}
$$

Then, the solution of the workhorse New Keynesian model can be derived following the procedure proposed in this paper by solving a companion quadratic matrix equation and a companion Sylvester matrix equation.

The implementation of the procedure laid out in this paper with the workhorse New Keynesian model articulated here has a finite-order VAR(2) representation given by (26) of the following form: $W_t = \Psi_1 W_{t-1} + \Psi_2 W_{t-2} + \Psi_3 \epsilon_t$, on the four endogenous variables in $W_t = (y_t, \pi_t, i_t, y_t^n)^T$. The corresponding composite coefficient matrices $\Psi_1 \equiv \left(\Theta + CAC^{-1}\right)$, $\Psi_2 \equiv -CAC^{-1}\Theta$, and $\Psi_3 \equiv CA^{-1}B$ have non-zero entries almost everywhere—for standard parameterizations of the model. This is in contrast with the bivariate monetary model discussed earlier where, for instance, there were no spillovers from lagged interest rates to current inflation. Partly endogenizing the real marginal costs in relation to output and tying economic activity to short-term interest rate movements through a dynamic IS equation is key for understanding how monetary policy and monetary policy shocks influence the dynamics of inflation through the lens of New Keynesian theory. From an economic perspective, the solution of the workhorse New Keynesian model shows that monetary policy shocks can contribute to drive inflation fluctuations unlike in the bivariate version of the model studied earlier.

The special case without habit formation also has an exact VAR(2) representation—analogous to that recovered in Carlstrom et al. (2009). However, Carlstrom et al. (2009) consider only three distinct structural shocks and can therefore work out their solution focusing solely on three endogenous variables (not including output potential, y_t^n) while the workhorse model I work with here requires y_t^n to be included in the vector W_t

and to be observable. Given that the finite-order VAR(2) solution of the workhorse New Keynesian model exists, under mild conditions one can identify and recover the structural monetary policy shock itself—as well as all other fundamental economic shocks driving the economy—directly from the observable vector of data W_t.

Therefore, I argue that empirical evidence which hinges upon Cholesky (and, in general, zero) restrictions should be interpreted with great caution as it may not have a structural interpretation—as noted, among others, by Carlstrom et al. (2009). I provide further discussion on the implications of the cross-equation restrictions implied by theory for the estimation of and empirical inferences based on VARs in the next section of the paper.

4   Some Lessons for Monetary VARs

In the remainder of the paper, I illustrate the solution of the workhorse New Keynesian model numerically taking advantage of the set of Matlab codes and functions that implement the procedure described in the paper.[22] The emphasis is placed on recovering and interpreting the structural monetary policy shocks from the data, through the lens of workhorse New Keynesian model. The finite-order VAR(2) in (26) can be estimated with the cross-equations restrictions imposed by theory using Bayesian techniques or simply by minimum distance (a related minimum distance estimation toolkit can be found in Jordà and Kozicki (2011)). I do not attempt the full estimation of the workhorse model here—instead, I use the framework to further understand the role of monetary policy and the propagation of monetary shocks adopting a conventional parameterization based on the parameter estimates reported by Smets and Wouters (2003). In this sense, I assume the parameters of the New Keynesian model specified for this illustration take the parameterization presented in Table 1:

---

[22] In terms of implementation, the solution of the canonical first-order LRE model requires solving a quadratic matrix equation as in (5) and a companion Sylvester matrix equation as in (24). In the univariate and bivariate illustrations presented before, the solution of the model can be obtained analytically with standard matrix algebra. However, the paper includes a straightforward Matlab code implementation to compute numerically the solution of the workhorse New Keynesian model. I use those codes in the rest of this section and make them available upon request.

Table 1. Parameterization of the Workhorse New Keynesian Model

Structural Parameter	Notation	Value	Source
Discount factor	$0 < \beta < 1$	0.990	Smets and Wouters (2003)
Degree of price stickiness	$0 \leq \xi_p < 1$	0.905	Smets and Wouters (2003)
Degree of price indexation	$0 \leq \gamma_p \leq 1$	0.472	Smets and Wouters (2003)
▷ Forward-looking weight on Phillips curve	$\gamma_f \equiv \frac{\beta}{1+\beta\gamma_p}$	0.675	Composite
▷ Backward-looking weight on Phillips curve	$\gamma_b \equiv \frac{\gamma_p}{1+\beta\gamma_p}$	0.322	Composite
▷ Slope of the Phillips curve	$\kappa \equiv \frac{\left(1-\beta\xi_p\right)\left(1-\xi_p\right)}{\left(1+\beta\gamma_p\right)\xi_p}$	0.007	Composite
Inverse intertemporal elasticity of substitution	$\sigma_c > 0$	1.371	Smets and Wouters (2003)
Inverse of the Frisch elasticity of labor supply	$\sigma_l > 0$	2.491	Smets and Wouters (2003)
External habit formation parameter	$0 < h < 1$	0.595	Smets and Wouters (2003)
Monetary Policy			
Policy inertia	$0 \leq \rho_i < 1$	0.958	Smets and Wouters (2003)
Response to inflation deviations	$\psi_\pi > 1$	1.688	Smets and Wouters (2003)
Response to output gap deviations	$\psi_y \geq 0$	0.095	Smets and Wouters (2003)
Exogenous Shock Parameters			
Persistence of the productivity shock	$-1 < \rho_a < 1$	0.815	Smets and Wouters (2003)
Volatility of the productivity shock	$\sigma_\delta > 0$	0.345**	Smets and Wouters (2003)
Persistence of the preference shock	$-1 < \rho_b < 1$	0.842	Smets and Wouters (2003)
Volatility of the preference shock	$\sigma_\zeta > 0$	0.089**	Smets and Wouters (2003)
Persistence of the labor supply shock	$-1 < \rho_l < 1$	0.891	Smets and Wouters (2003)
Volatility of the labor supply shock	$\sigma_\nu > 0$	1.244**	Smets and Wouters (2003)
Persistence of the monetary shock	$-1 < \rho_m < 1$	0.750*	Parameterized
Volatility of the monetary shock	$\sigma_\xi > 0$	0.001**	Smets and Wouters (2003)

* Introduces persistence unlike the parameterization in Smets and Wouters (2003) ensuring also that the matrix A is invertible.
** The volatility parameters are chosen such that the standard deviation of each shock innovation recovered from the data is equal to 1.

Given the parameterization of Table 1, the solution Θ to the companion quadratic matrix equation has its roots inside the unit circle. Furthermore, the code also reports that the solution C to the companion Sylvester matrix equation exists and is both unique and invertible. Therefore, the workhorse New Keynesian model has a finite-order VAR(2) representation given by (26) where the corresponding coefficient matrices $\Psi_1 \equiv \left(\Theta + CAC^{-1}\right)$, $\Psi_2 \equiv -CAC^{-1}\Theta$, and $\Psi_3 \equiv CA^{-1}B$ take the following form under the parameterization

reported in Table 1:

$$\Psi_1 = \begin{pmatrix} 1.5057 & -0.5355 & -3.2133 & -0.0607 \\ -0.0648 & 1.4821 & 0 & 0.0648 \\ -0.0039 & 0.0576 & 1.7171 & 0.0043 \\ 0.5847 & -1.9717 & -5.4962 & 0.7552 \end{pmatrix},$$

$$\Psi_2 = \begin{pmatrix} -0.5253 & 0.2655 & 2.4934 & 0.0179 \\ 0.0222 & -0.4768 & 0 & -0.0222 \\ 0.0006 & -0.0367 & -0.7254 & -0.0009 \\ -0.2740 & 1.0036 & 4.2124 & -0.1464 \end{pmatrix}, \tag{77}$$

$$\Psi_3 = \begin{pmatrix} 0.0282 & 0.0203 & -0.0405 & -0.0132 \\ -0.0263 & 0.0048 & 0.0312 & -0.0055 \\ -0.0026 & 0.0004 & 0.0029 & 0.0006 \\ 0.2050 & 0 & -0.2117 & 0 \end{pmatrix},$$

with non-zero entries almost everywhere.

The parameter values are chosen to illustrate the qualitative features of the workhorse New Keynesian model and the performance of the algorithms for the solution of the companion quadratic matrix equation and the companion Sylvester matrix equation. I use a laptop with Intel(R) Core(TM) i7 with 2.7GHz, 4 cores and 32GB of installed memory (RAM). The wall-clock time elapsed in computing and reporting the numerical results given the defaults of the code is 10.3351 seconds (CPU time is: 0.0044187 seconds). Using the iterative algorithm to compute the solution to the quadratic matrix equation and Matlab's own implementation of the Hessenberg-Schur algorithm for Lyapunov equations to speed up the computation, the elapsed time barely fall to 9.3604 seconds (CPU time is: 0.0038391 seconds).[23]

What Does the Parameterized Workhorse Model Interpret the Observed Data?   Mapping the endogenous variables of the workhorse New Keynesian model to the observed data requires some additional clarification—for this, I largely rely on the guidelines of Pfeifer (2015). Most business cycle models—like the New Keynesian model—abstract from movements in the labor force and assume output (and, by analogy, output potential) to be expressed in per-capita terms. Hence, the empirical counterpart to per-capita output in the New Keynesian model is calculated as the ratio of U.S. Real GDP in millions of chained Dollars (from the Bureau of Economic Analysis) over the U.S. civilian labor force 16-years and older in millions (from the Bureau of Labor Statistics). Analogously, per-capita output potential is the ratio of U.S. Real Potential GDP in millions of chained Dollars (from the Congressional Budget Office) over the U.S. civilian labor force 16-years and older in millions (from the Bureau of Labor Statistics).

The price index used is the quarterly U.S. Consumer Price Index (CPI) for all urban consumers–all items (from the Bureau of Labor Statistics). Finally, the short-term policy rate is constructed using the effective Federal Funds rate (% p.a.) (from the Federal Reserve Board) but replacing the observations in the aftermath of the 2008 recession when policy rates were constrained at the zero-lower bound with the Wu-Xia Shadow Federal Funds rate (as reported by the Federal Reserve Bank of Atlanta). As customary, I transform the per-capita output and the price index taking logs to make the log-transformed series scale invariant. All four series are reported at quarterly frequency.

---

[23] The elapsed and CPU times reported here are average times over calculated with 10000 repeated draws of the parameterized workhorse model.

The workhorse model laid out earlier does not explicitly incorporate trends and assumes the endogenous variables to be stationary, so the empirical counterparts for the endogenous variables are filtered outside of the model. A conventional way of getting the output and output potential series is by removing the trend out of the per-capita series in logs applying a first-difference filter to the series. I apply the same filter (first-differencing) to the price index as well in order to compute the corresponding measure of inflation. Data in first differences still has non-zero mean average growth over the sample—accordingly, I demean the first-differenced log-per-capita output, the first-differenced log-per-capita potential, and the first-differenced price level. I also demean the short-term interest rate for the same reason—divided by 400 to express the policy rate in the same units as the other variables.

The impact of monetary policy on the volatility, cyclicality and persistence of the endogenous variables depends in nonlinear ways on the policy parameters, ψ_π and ψ_y, that describe the systematic part of the monetary policy rule. Here, I exploit the finite-order VAR(2) representation of the solution in (26) in order to compute the theoretical moments of the workhorse New Keynesian model for the benchmark parameterization of ψ_π, but also for a range of values of the anti-inflation bias ψ_π that goes from 1 (the lower bound on the Taylor principle) up to 6. I summarize the key findings describing the business cycle facts in the data since the onset of the Great Moderation period (1984:Q1 till 2017:Q4) and the implications of the parameterized workhorse model in Table 2.

As can be seen in Table 2, the monetary policy parameter ψ_π has a substantial impact over the business cycle. Generally, a higher value of ψ_π signals a stronger anti-inflation bias on the part of the central bank and is associated with significant declines in inflation volatility according to the theoretical standard deviations—albeit not for the output and the policy rate. As expected, output potential is completely unaffected by the policy rule. Moreover, higher values of the policy parameter ψ_π lead to a stronger negative (contemporaneous) correlation between output and inflation and to a stronger positive (contemporaneous) correlation between output and output potential—the latter partly reflects that in this context the higher anti-inflation bias leads the economy closer to its frictionless counterpart ('first best'). I also find a strong negative comovement between output and the policy rate and a significant weakening of the persistence of inflation as measured by its first-order autocorrelation.

Table 2. Key Business Cycle Moments at Different Degrees of Anti-Inflationary Bias

Std. Deviation	Data	$\psi_\pi = 1.688$	$\psi_\pi = 1$	$\psi_\pi = 3$	$\psi_\pi = 6$
Output	0.60	0.52	0.51	0.55	0.60
Inflation	0.50	0.08	0.09	0.07	0.05
Policy Rate	0.85	0.04	0.03	0.05	0.07
Potential	0.28	0.84	0.84	0.84	0.84
Contemporaneous Comovement					
Output	1	1	1	1	1
Inflation	0.18	-0.14	-0.05	-0.21	-0.25
Policy Rate	0.06	-0.97	-0.97	-0.97	-0.97
Potential	0.40	0.86	0.84	0.89	0.93
First-Order Autocorrelation					
Output	0.13	0.98	0.99	0.98	0.98
Inflation	0.33	0.83	0.84	0.82	0.79
Policy Rate	0.99	0.98	0.98	0.98	0.98
Potential	0.14	0.93	0.93	0.93	0.93

Note: This table summarizes the key theoretical business cycle moments of the workhorse New Keynesian model keeping the parameterization invariant as in Table 1 except for the policy parameter ψ_π.

Note: All data is demeaned, expressed in units and not annualized. The sample period covers from 1984:Q1 till 2017:Q4. The empirical standard deviations are scaled by 100 for comparability.

When compared against the business cycle features obtained from the filtered data, the model predictions tend to undershoot the volatility of inflation and the policy rate and overshoot the persistence of output and inflation. The negative comovement between output and inflation and, notably, between output and the policy rate is also inconsistent with the empirical evidence. Broadly speaking, these discrepancies between the theoretical predictions and the empirical evidence suggest that the workhorse model may be insufficient to capture the dynamics of inflation and their linkages to monetary policy actions. A richer environment might be needed to incorporate those fully.

Furthermore, the findings in Table 2 also show how shifts in the patterns of endogenous volatility, cyclicality, and persistence do not necessarily reflect changes in the underlying shock processes driving the dynamics of the endogenous variables. In turn, they can simply be the result of changes in the monetary policy regime altering the monetary policy transmission mechanism itself. I illustrate those potential changes in the transmission mechanism in Figure 1 with the theoretical one-standard deviation impulse response functions (IRFs) at different degrees of the anti-inflation bias ψ_π. For starters, I observe that the paradoxical empirical finding that a monetary shock could lead to an increase in inflation (the "price puzzle") seems inconsistent with the predictions of the workhorse New Keynesian model. Interestingly, I also show that the decline on inflation and on output resulting from a one-standard deviation positive (contractionary) shock to monetary policy gets attenuated the higher the anti-inflation bias stance on monetary policy ($\uparrow \psi_\pi$) becomes.

As expected, the policy rule has no influence on the dynamics of output potential. However, I note that a higher anti-inflation bias stance on monetary policy ($\uparrow \psi_\pi$) tends to dampen the endogenous output response to preference and monetary policy shock innovations while it amplifies the response to productivity and labor supply shock innovations. At the same time, the response of endogenous inflation to all types of shocks is shown to become more muted as ψ_π increases. Inflation and output move in the same direction in

response to preference shocks and monetary policy shocks, but they move in opposite directions in response to productivity shocks and labor supply shocks. Hence, changes in the contribution of the different shocks can result in changes in the unconditional patterns of comovement predicted by the model suggesting that the discrepancies between the model and the empirical evidence reported in Table 2 can be partly due to an inconsistent mix of the different shocks implied by the parameterization in Table 1.

Figure 1. Theoretical IRFs at Different Degrees of Anti-Inflationary Bias
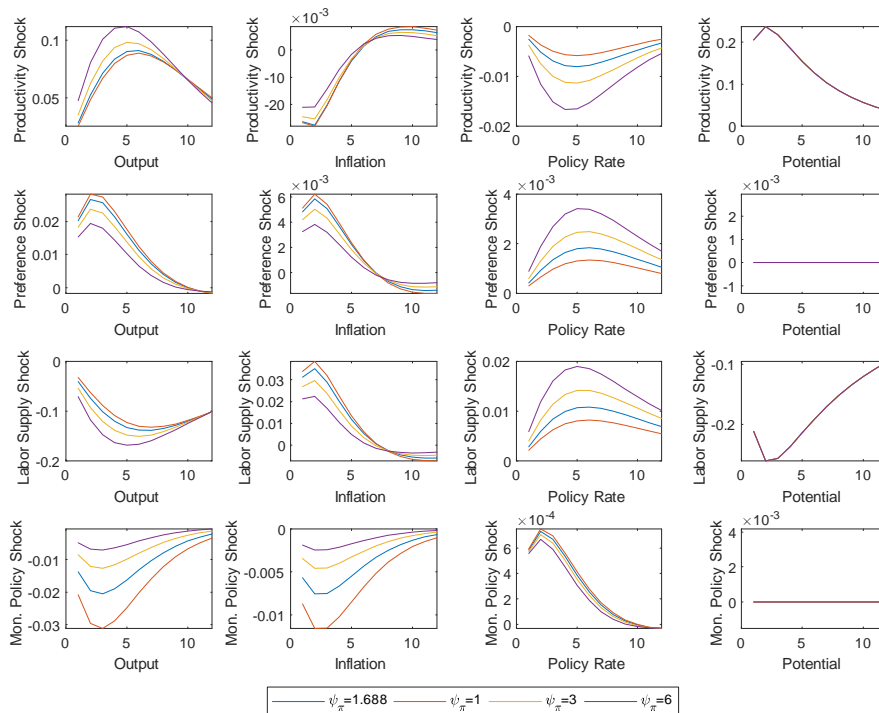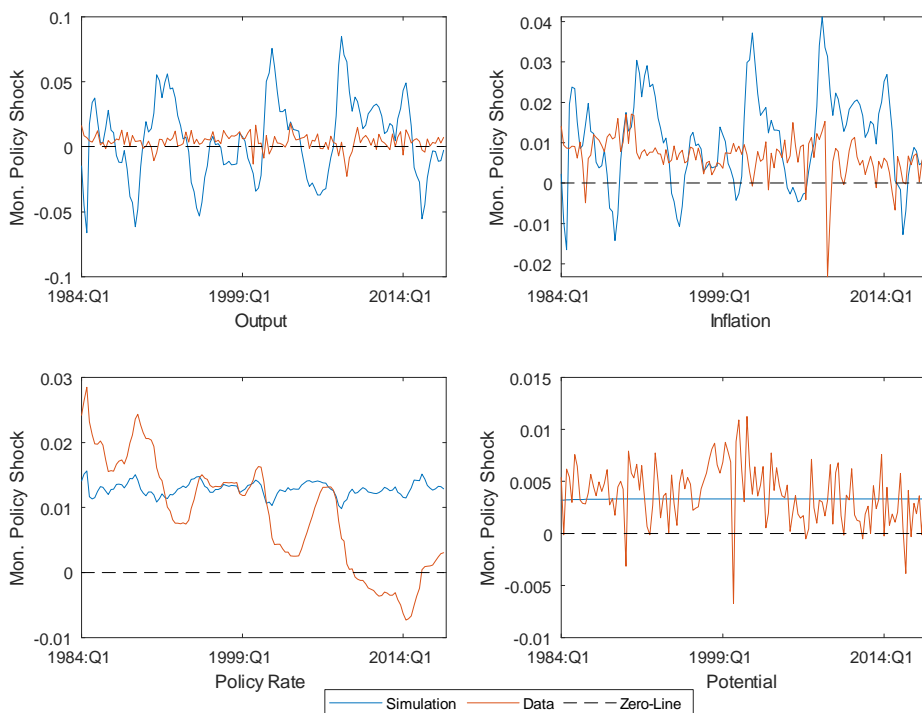

Note: This figure displays the theoretical impulse response functions (IRFs) of the workhorse New Keynesian model keeping the parameterization invariant as in Table 1 except for the policy parameter ψ_π.

Furthermore, the findings in Figure 1 show that the transmission mechanism of structural shocks—monetary policy shocks in particular—and their spillovers ultimately depend in nonlinear ways on the features of the prevailing monetary policy regime. That is, they depend on the policy parameters of the Taylor (1993) rule in the workhorse New Keynesian model as well as on other deep structural parameters related to preferences, technology, etc.

However, understanding the propagation of structural shocks is only part of the task. Exactly recovering the shock innovations when fundamental—in particular, monetary policy shocks in the New Keynesian model—is also crucial to understand their contribution over the business cycle. Through the lens of the workhorse New Keynesian model parameterized as in Table 1, the findings in Figure 2 illustrate this point comparing the observed data on per-capita output growth, inflation, and the short-term policy rate against model-consistent simulations for the monetary policy shock separately (where the structural shocks are recovered from the observed data itself).

Taking the observed data on per-capita output growth, per-capita output potential growth, inflation, and the short-term policy rate as given, I use equation (27) from Corollary 2 to recover the structural shocks for the workhorse New Keynesian model. Then, I simulate the endogenous variables feeding the recovered structural monetary policy shock through the solution in (26) from Corollary 1 in order to assess the monetary policy shock's contribution to account for the observed data. The comparison between these one-shock-only simulation and the actual observed data is plotted in Figure 2.

Figure 2. Contribution of the Monetary Policy Shock to the Observed U.S. Data

Note: This figure displays the observed data on the output gap, inflation and the short-term interest rate vs. model-consistent simulations of the corresponding variables based on the three-equation workhorse New Keynesian model keeping the parameterization invariant as in Table 1. The simulations correspond to the endogenous variables generated from the solution of the model feeding one of the structural shocks recovered at a time. The simulated series are modified with the addition of the corresponding sample mean of the corresponding observed variables for comparability.

The time series on this data starts in 1968:Q3 and ends in 2017:Q4 covering the entire Great Moderation period starting in 1984:Q1 as well as the 2008 recession and its aftermath (with conventional Federal Funds Rate-based monetary policy stuck at the zero-lower bound). The sample mean is added to the model simulations in Figure 2 to make them comparable with the observed data that contains a non-zero sample mean—all series are reported in units (rather than percentages) and not annualized.[24] The evidence reported

---

[24] The sample mean of first-differenced output per capita (in logs) is 0.0032, the sample average for first-differenced headline CPI (in logs) is 0.0099, 0.0129 is the period average for the nominal short-term interest rate, and the average of first-differenced output potential per capita (in logs) is 0.0033.

shows—not surprisingly—that monetary policy shocks are an important driver helping us understand infla-tion in the New Keynesian model. Moreover, monetary shocks play a significant role for per-capita output growth but have little influence on the short-term interest rate—which is largely driven by non-monetary policy shocks under the given parameterization in Table 1—and have no role on per-capita output potential growth.

It is worth pointing out here that monetary shocks tend to generate more volatile output and inflation series than what is observed in the data. This parameterized New Keynesian model attributes a expansionary path to monetary policy in the aftermath of the 2008 recession (during the zero-lower bound episode until around 2014) pushing inflation and output upwards. This is consistent with a significant body of the theoretical literature that has built a narrative about the recent U.S. zero-lower bound experience partly around the perceived accommodation provided by the implementation of a broad range of unconventional monetary policy actions aimed at supporting the economy.

Some Final Thoughts on Estimation. The method proposed in this paper to solve LRE models makes an important contribution to better understand the transmission mechanism, the role of structural economic shocks and its mapping into finite-order VAR specifications—stylized ones like the workhorse New Keynesian model explored here or richer ones (as indicated before)—that can be recast in the first-order canonical form given by $(1)-(2)$. The VAR(2) solution in (28) can be expressed in companion form as follows:

$$\mathbb{E}\left(W_t \mid W_{t-1}, W_{t-2}\right) = \Pi^T X_t, \tag{78}$$

$$X_t = \left[\begin{array}{c} W_{t-1} \\ W_{t-2} \end{array}\right], \Pi^T \equiv \left[\begin{array}{cc} \Psi_1 & \Psi_2 \end{array}\right],$$

$$\text{and } \theta = \left(vec\left(\Psi_1\right)^T, vec\left(\Psi_2\right)^T, vec\left(\Omega\right)^T\right)^T \text{ where } \Omega \equiv \Psi_3^T \Psi_3.$$

Assuming Gaussianity of the shock innovations, the log-likelihood function can be expressed as:

$$\mathcal{L}\left(\theta\right) = -\frac{nk}{2}\log\left(2\pi\right) + \frac{n}{2}\log\left|\Omega^{-1}\right| - \frac{1}{2}\sum_{j=1}^{n}\left[\left(W_j - \Pi^T X_j\right)^T \Omega^{-1}\left(W_j - \Pi^T X_j\right)\right]. \tag{79}$$

where $k\left(=m\right)$ is the number of observable variables while n defines the number of observations of each observable.[25] The maximum likelihood estimates of Π and Ω are given by:

$$\widehat{\Pi}^T = \left[\sum_{j=1}^{n} W_j X_j^T\right]\left[\sum_{j=1}^{n} X_j X_j^T\right]^{-1}, \tag{80}$$

$$\widehat{\Omega} = \frac{1}{T}\sum_{j=1}^{n}\widehat{e}_j\widehat{e}_j^T, \text{ where } \widehat{e}_j = W_j - \widehat{\Pi}^T X_j. \tag{81}$$

The maximum likelihood estimate $\widehat{\Pi}$ is the estimated coefficient vector from an OLS regression of W_t on X_t^T. The maximum likelihood estimates $\widehat{\Pi}$ and $\widehat{\Omega}$, under standard conditions, will give consistent estimates of the population parameters even if the true innovations are non-gaussian (Hamilton (1994)).

However, the challenge for model estimation is identification—the problem arises from the fact that the solution of the LRE model involves a nonlinear mapping between the deep structural parameters of the model and the elements in Π and Ω. This means that even if the reduced-form representation of the VAR(2) in (1)

---

[25]This is the log-likelihood function assuming $\Omega \equiv \Psi_3^T \Psi_3$ is positive definite and therefore invertible.

can be easily estimated, identification of the deep structural parameters is not guaranteed. As an example, let me recall the AR(2) representation of the one-equation model of inflation given in (39). Estimating that AR(2) model by maximum likelihood, I obtain two reduced-form autoregressive coefficients and one variance estimates with which to recover four structural parameters (i.e., $(\gamma_f, \gamma_b, \rho_e, \sigma_\delta)$). This poses an under-identified system and, hence, at least one of the structural parameters cannot be recovered.

Assuming that σ_δ is known and equal to 1, I end up with a just identified system between the remaining three deep structural parameters (i.e., $(\gamma_f, \gamma_b, \rho_e)$) and the estimated reduced-form coefficients. Even in this case, either multiple solutions or no solution could exist. For instance, if $\rho_e = 0$ the solution in (39) reduces to $\pi_t = \left(\frac{1 - \sqrt[2]{1 - 4\gamma_f \gamma_b}}{2\gamma_f} \right) \pi_{t-1} + \left(\frac{1}{\gamma_f \left(\frac{1 + \sqrt[2]{1 - 4\gamma_f \gamma_b}}{2\gamma_f} \right)} \right) \delta_t$ and if $\gamma_b = 0$ to $\pi_t = \rho_e \pi_{t-1} + \left(\frac{1}{1 - \gamma_f \rho_e} \right) \delta_t$.

Therefore, if our reduced-form estimates of the AR(2) model imply that the second lag of the autoregressive is zero, I may not be able to distinguish between a solution where the exogenous real marginal cost shock has no persistence but some form of price indexation ($\rho_e = 0$, $\gamma_b \neq 0$) or a solution where the exogenous real marginal cost shock is persistent but there is no price indexation ($\rho_e \neq 0$, $\gamma_b = 0$). In turn, if the true model is one with $\gamma_b = 0$ and $\sigma_\delta = 1$, and I estimate the corresponding AR(1) process, I find that the first lag of the reduced-form determines ρ_e exactly. However, the structural model solution implies that $0 < \gamma_f \leq 1$ and, therefore, the implied standard deviations go between 1 and $\frac{1}{1 - \rho_e} \geq 1$ (assuming $\rho_e \geq 0$). Hence, for estimated values of the standard deviation obtained from the reduced-form that fall outside that range, no solution exist that can recover simultaneously both structural parameters.

Going beyond maximum likelihood estimates, Bayesian techniques may be useful to incorporate additional information through the priors that can lessen the identification problems. Minimum distance estimators such as those proposed by Jordà and Kozicki (2011) can also be helpful—for instance, whenever a unique solution for the deep structural parameters that maximizes the log-likelihood of the reduced-form representation exists (as in my last example above). However, in general, structural parameter identification remains a concern for estimation and the best approach to deal with it (at least partly) is likely to depend on the data available and on the empirical question that the model is aimed to address. In any event, the treatment of weak/partial identification and even misspecification falls outside the scope of this paper.

5 Concluding Remarks

I propose a straightforward approach to solve a large class of LRE models and show under mild conditions that the corresponding solution—whenever one exists and is unique—can be represented in finite-order VAR form. For this, I represent the LRE model in its first-order canonical form as a system of expectational first-order difference equations. Then, I reduce the first-order form by splitting the backward-looking and forward-looking parts of the model by transforming the endogenous variables with the solution to a companion quadratic matrix equation. The forward-looking part on the transformed variables is solved through the method of undetermined coefficients. The unique solution to the forward-looking part—when one exists—can be described by means of a linear state-space form. Imposing the cross-equation restrictions that arise from theory on the state-space representation, I show that solving a companion Sylvester matrix equation suffices to: check the existence and uniqueness of the solution to the purely forward-looking part of the LRE model and to also characterize the reduced-form solution of the transformed variables in finite-order VAR

form. Finally, I undo the transformation of the variables obtaining the solution to the first-order canonical LRE model in VAR form as well.

An important contribution of the paper is the derivation of conditions under which the finite-order VAR solution of the canonical forward-looking LRE model via the well-known Sylvester matrix equation is well-defined. The approach proposed here not only provides a way to decouple the canonical form of the model from its first-order specification and to characterize the corresponding solution in finite-order VAR form, but it also checks its properties—existence, uniqueness, and invertibility. Furthermore, the paper also makes a contribution to the computational economics literature with the development of an integrated, unified algorithm to solve numerically LRE models for which a unique finite-order VAR representation exists (illustrated with an application based on the workhorse New Keynesian model).

Solving LRE models with a finite-order VAR representation by this method is straightforward to implement, efficient, and can be handled easily with standard matrix algebra and conventional computational methods. The paper provides a number of functions for the solution of the companion quadratic matrix equation and the companion Sylvester matrix equation with an economically-relevant application to the study of the transmission mechanism of monetary policy in the context of the workhorse New Keynesian model. For this purpose, the paper showcases the procedure building up the workhorse New Keynesian model from the hybrid Phillips curve upwards progressively incorporating monetary policy and a channel to endogenize the real marginal costs that responds to policy actions (through the dynamic IS equation). This opens up a feedback channel for monetary policy and for the monetary policy shocks to affect the determination of inflation.

The paper analyzes the fundamental features of the workhorse New Keynesian model that permit the propagation of monetary policy shocks to inflation and output—with particular attention to the role of the systematic part of the monetary policy rule (which determines the response to inflation and output deviations from their targets) and to the recovery of fundamental shocks (particularly the monetary policy shock). The paper also shows how under the conditions that ensure the existence of a finite-order VAR representation of the solution, the identification of fundamental shocks for empirical research—including the recovery of monetary shocks—can be made tractable. However, the paper also shows that the identification of the deep structural parameters of the model for estimation purposes remains a challenge.

References

Anderson, E. W., E. R. McGrattan, L. P. Hansen, and T. J. Sargent (1996). Mechanics of Forming and Estimating Dynamic Linear Economies. *Handbook of Computational Economics 1*, 171–252. Elsevier Science, Amsterdam. https://doi.org/10.1016/S1574-0021(96)01006-4.

Bartels, R. H. and G. W. Stewart (1972). Solution of the Matrix Equation AX+XB=C. *Communications of the Association for Computing Machinery (ACM) 15*(9), 820–826. https://doi.org/10.1145/361573.361582.

Binder, M. and M. H. Pesaran (1995). Multivariate Rational Expectations Models and Macroeconomic Modelling: A Review and Some New Results. In M. H. Pesaran and M. Wickens (Eds.), *Handbook of Applied Econometrics: Macroeconomics*, pp. 139–187. Oxford: Basil Blackwell.

Binder, M. and M. H. Pesaran (1997). Multivariate Linear Rational Expectations Models: Characterization of the Nature of the Solutions and Their Fully Recursive Computation. *Econometric Theory 13*(6), 877–888. https://doi.org/10.1017/s0266466600006307.

Blanchard, O. J. and C. M. Kahn (1980). The Solution of Linear Difference Models Under Rational Expectations. *Econometrica 48*(5), 1305–13011.

Broze, L., C. Gouriéroux, and A. Szafarz (1985). Solutions of Dynamic Linear Rational Expectations Models. *Econometric Theory 1*(3), 341–368. https://doi.org/10.1017/S0266466600011257.

Broze, L., C. Gouriéroux, and A. Szafarz (1990). *Reduced Forms of Rational Expectations Models*. Chur, Switzerland: Harwood Academic Publishers. Macroeconomic Theory. https://doi.org/10.4324/9781315014685.

Calvo, G. A. (1983). Staggered Prices in a Utility-Maximizing Framework. *Journal of Monetary Economics 12*(3), 383–398. https://doi.org/10.1016/0304-3932(83)90060-0.

Campbell, H. and J. H. Cochrane (1999). By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy 107*(2), 205–251. https://doi.org/10.1086/250059.

Carlstrom, C. T., T. S. Fuerst, and M. Paustian (2009). Monetary Policy Shocks, Cholesky Identification, and DNK Models. *Journal of Monetary Economics 56*(7), 1014–1021. https://doi.org/10.1016/j.jmoneco.2009.09.009.

Christiano, L. J. (2002). Solving Dynamic Equilibrium Models by a Method of Undetermined Coefficients. *Computational Economics 20*(1), 21–55. October. https://doi.org/10.1023/A:1020534927853.

Dennis, R. (2009). Consumption Habits in a New Keynesian Business Cycle Model. *Journal of Money, Credit and Banking 41*(5), 1015–1030. https://doi.org/10.1111/j.1538-4616.2009.00242.x.

Duncan, R. and E. Martínez-García (2015). Forecasting Local Inflation with Global Inflation: When Economic Theory Meets the Facts. Federal Reserve Bank of Dallas *Globalization and Monetary Policy Institute Working Paper No. 235*. https://doi.org/10.24149/gwp235.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. Sargent, and M. Watson (2007). ABCs (and Ds) of Understanding VARs. *American Economic Review 97*(3), 1021–1026. https://doi.org/10.1257/aer.97.3.1021.

Franchi, M. and P. Paruolo (2015). Minimality of State Space Solutions of DSGE Models and Existence Conditions for Their VAR Representation. *Computational Economics 46*(4), 613–626. https://doi.org/10.1007/s10614-014-9465-4.

Golub, G. H., S. Nash, and C. F. van Loan (1979). A Hessenberg-Schur Method for the Problem AX+XB=C. *IEEE Transactions on Automatic Control AC-24*(6), 909–913. https://doi.org/10.1109/tac.1979.1102170.

Golub, G. H. and C. F. van Loan (1996). The Hessenberg and Real Schur Forms. In *Matrix Computations* (3rd ed.)., pp. 361–372. Baltimore, MD: Johns Hopkins University Press.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, L. and T. Sargent (1980). Formulating and Estimating Dynamic Linear Rational Expectations Models. *Journal of Economic Dynamics and Control 2*(1), 7–46. https://doi.org/10.1016/0165-1889(80)90049-4.

Horn, R. A. and C. R. Johnson (1991). *Topics in Matrix Analysis*. Cambridge University Press. https://doi.org/10.1017/cbo9780511840371.

Inoue, A. and L. Kilian (2002). Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in VAR(∞) Models. *International Economic Review 43*(2), 309–332. https://doi.org/10.1111/1468-2354.t01-1-00016.

Jiang, T. and M. Wei (2003). On Solutions of the Matrix Equations X-AXB=C and X-A$\bar{X}$B=C. *Linear Algebra and Its Applications 367*, 225–233. https://doi.org/10.1016/s0024-3795(02)00633-x.

Jordà, Ò. and S. Kozicki (2011). Estimation and Inference by the Method of Projection Minimum Distance: An Application to the New Keynesian Hybrid Phillips Curve. *International Economic Review 52*(2), 461–487. https://doi.org/10.1111/j.1468-2354.2011.00635.x.

Kailath, T. (1980). *Linear Systems*. Englewood Cliffs, NJ: Prentice Hall.

King, R. G. and M. W. Watson (1998). The Solution of Singular Linear Difference Systems Under Rational Expectations. *International Economic Review 39*(4), 1015–1026. https://doi.org/10.2307/2527350.

Klein, P. (2000). Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model. *Journal of Economic Dynamics and Control 1405-1423*(10), 1405–1423. https://doi.org/10.1016/s0165-1889(99)00045-7.

Lancaster, P. and M. Tismenetsky (1985). *The Theory of Matrices with Applications. Second Ed.* New York: Academic Press.

Lin, M. and H. K. Wimmer (2011). The Generalized Sylvester Matrix Equation, Rank Minimization and Roth's Equivalence Theorem. *Bulletin of the Australian Mathematical Society 84*, 441–443. https://doi.org/10.1017/s0004972711002334.

Lubik, T. A. and F. Schorfheide (2003). Computing Sunspot Equilibria in Linear Rational Expectations Models. *Journal of Economic Dynamics and Control 28*(2), 273–285. https://doi.org/10.1016/s0165-1889(02)00153-7.

Lucas, R. E. J. (1976). Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy 1*, 19–46. https://doi.org/10.1016/s0167-2231(76)80003-6.

Martínez-García, E. (2018, July). Modeling Time-Variation Over the Business Cycle (1960-2017): An International Perspective. *Federal Reserve Bank of Dallas, Globalization and Monetary Policy Institute Working Paper no. 343*.

Martínez-García, E., D. Vilán, and M. A. Wynne (2012). Bayesian Estimation of NOEM Models: Identification and Inference in Small Samples. *Advances in Econometrics 28*, 137–199. DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments (Advances in Econometrics, Volume 28). https://doi.org/10.1108/s0731-9053(2012)0000028007.

Martínez-García, E. and M. A. Wynne (2014). Assessing Bayesian Model Comparison in Small Samples. *Advances in Econometrics 34*, 71–115. Bayesian Model Comparison (Advances in Econometrics, Volume 34). https://doi.org/10.1108/s0731-905320140000034006.

Morris, S. D. (2016). VARMA Representation of DSGE Models. *Economic Letters 138*, 30–33.

Muth, J. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica 29*(3), 315–335. https://doi.org/10.2307/1909635.

Pfeifer, J. (2015). A Guide to Specifying Observation Equations for the Estimation of DSGE Models. *Mimeo, University of Mannheim*.

Ravenna, F. (2007). Vector Autorregressions and Reduced Form Representations of DSGE Models. *Journal of Monetary Economics 54* (7), 2048–2064. https://doi.org/10.1016/j.jmoneco.2006.09.002.

Sargent, T. (1980). Rational Expectations and the Reconstruction of Macroeconomics. Federal Reserve Bank of Minneapolis *Quarterly Review Summer*, 15–19.

Sima, V. and P. Benner (2015). Solving Linear Matrix Equations with SLICOT. *National Institute for Research and Development in Informatics, Bucharest, Romania*. Bucharest, Romania.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica 48*(1), 1–48. https://doi.org/10.2307/1912017.

Sims, C. A. (2002). Solving Linear Rational Expectations Models. *Computational Economics 20*(1), 1–20. https://doi.org/10.1023/A:1020517101123.

Smets, F. and R. Wouters (2003). An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association 1*, 1123–1175. https://doi.org/10.1162/154247603770383415.

Strang, G. (2016). *Introduction to Linear Algebra (5th Ed.)*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).

Taylor, J. B. (1993). Discretion versus Policy Rules in Practice. *Carnegie-Rochester Conference Series on Public Policy 39*, 195–214. https://doi.org/10.1016/0167-2231(93)90009-l.

Uhlig, H. (1999). A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily. In R. Marimon and A. Scott (Eds.), *Computational Methods for the Study of Dynamic Economies*, pp. 30–61. Oxford and New York: Oxford University Press. https://doi.org/10.1093/0199248273.003.0003.

Yun, T. (1996). Nominal Price Rigidity, Money Supply Endogeneity, and Business Cycles. *Journal of Monetary Economics 37*(2), 345–370. https://doi.org/10.1016/s0304-3932(96)90040-9.

Appendix

A The General Form of the LRE Model

The general form of a large class of multivariate LRE models can be written as

$$\Phi_{00} Y_t = \sum_{i=1}^{p} \Phi_{i0} Y_{t-i} + \sum_{i=0}^{p} \sum_{j=1}^{n} \Phi_{ij} \mathbb{E}_{t-i} \left(Y_{t+j-i} \right) + \Omega e_t, \tag{82}$$

where Y_t and e_t are the $r \times 1$ vectors of the endogenous and forcing variables of the LRE model, respectively. The vector Y_t collects the relevant subset r of endogenous variables and the r forcing variables of the LRE model, while the vector e_t are the innovations corresponding to the exogenous forcing variables. The matrices Φ_{ij} for all $i = 0, ..., p$ and $j = 0, ..., n$ and Ω are all $r \times r$ conforming matrices, and $\mathbb{E}_t \left(\cdot \right)$ represents the conditional expectations operator based on all current and lagged values of Y_t and e_t.[26] I assume that Φ_{00} in (82) is nonsingular and, without loss of generality, set it equal to the $r \times r$ identity matrix $\mathbf{I}_r$.

As in Broze et al. (1985) and Broze et al. (1990), the more general form given by (82) can be rewritten in more compact form as follows:

$$N_t = \Phi_1^* N_{t-1} + \Phi_2^* \mathbb{E}_t \left[N_{t+1} \right] + \Phi_3^* \overline{u}_t, \tag{83}$$

where

$$N_t = \left(U_t^T, U_{t-1}^T, ..., U_{t-p+1}^T \right)^T, \ U_t^T = \left(Y_t^T, \mathbb{E}_t \left(Y_{t+1}^T \right), ..., \mathbb{E}_t \left(Y_{t+n}^T \right) \right)^T,$$

$$\overline{u}_t = \left(u_t^T, 0_{l \times 1}^T, ..., 0_{l \times 1}^T \right)^T, \ u_t = \left(\left(\Omega e_t \right)^T, 0_{r \times 1}^T, ..., 0_{r \times 1}^T \right)^T,$$

$$\Phi_1^* = -D_0^{-1} D_1, \ \Phi_2^* = -D_0^{-1} D_1, \ \Phi_3^* = D_0^{-1}.$$

Here, $\overline{u}_t$ is a column-vector of dimension $q = p \left(n + 1 \right) r$ and u_t is a column-vector of dimension $l = \left(n + 1 \right) r$. The $q \times q$ square matrices D_i, $i = -1, 0, 1$ are defined as:

$$D_{-1} = \begin{pmatrix} \Psi_{-1} & 0_l & ... & 0_l \\ 0_l & 0_l & ... & 0_l \\ & & \ddots & \\ 0_l & 0_l & ... & 0_l \end{pmatrix}, \ D_0 = \begin{pmatrix} \Psi_0 & \Psi_1 & ... & \Psi_{p-1} \\ 0_l & \mathbf{I}_l & ... & 0_l \\ & & \ddots & \\ 0_l & 0_l & ... & \mathbf{I}_l \end{pmatrix},$$

$$D_1 = \begin{pmatrix} 0_l & 0_l & ... & 0_l & \Psi_p \\ -\mathbf{I}_l & 0_l & ... & 0_l & 0_l \\ & & \ddots & & \\ 0_l & 0_l & ... & -\mathbf{I}_l & 0_l \end{pmatrix},$$

---

[26] In case the processes have non-zero means, one should add a constant as well to the specification in (82).

where Ψ_i, $i = -1, 0, ..., p$, are square matrices of dimension l given by:

$$\Psi_{-1} = \begin{pmatrix} 0_r & 0_r & ... & 0_r & 0_r \\ -\mathbf{I}_r & 0_r & ... & 0_r & 0_r \\ & & \ddots & & \\ 0_r & 0_r & ... & -\mathbf{I}_r & 0_r \end{pmatrix}, \quad \Psi_0 = \begin{pmatrix} \mathbf{I}_r & -\Phi_{01} & ... & -\Phi_{0n} \\ 0_r & \mathbf{I}_r & ... & 0_r \\ & & \ddots & \\ 0_r & 0_r & ... & \mathbf{I}_r \end{pmatrix},$$

$$\Psi_i = \begin{pmatrix} -\Phi_{i0} & -\Phi_{i1} & ... & -\Phi_{in} \\ 0_r & 0_r & ... & 0_r \\ & & \ddots & \\ 0_r & 0_r & ... & 0_r \end{pmatrix} \quad \text{for } i = -1, 0, ..., p.$$

In the general form of the LRE model in (82), the endogenous variables are bundled together with the forcing variables in the vector Y_t. Naturally, the compact form derived in (83) inherits the same feature and the column-vector N_t of dimension $q = 2k$ combines k current and lagged forcing variables with k current and lagged endogenous variables (where both forcing and endogenous variables have the same finite lag order). Assume the vectors N_t and $\bar{u}_t$ are re-ordered such that $N_t = \left(W_t^T, X_t^T\right)^T$, where W_t is the vector of endogenous variables and X_t the vector of the forcing variables, and $\bar{u}_t$ follows the same consistent order. Then, the compact form in (83) can be rewritten as in the expectational difference system $(1) - (2)$ introduced in Section 2.

The compact system $(1) - (2)$ used in the paper simply presents the compact solution in two block sub-systems splitting the endogenous variables W_t which can be both backward-looking and forward-looking from the exogenous forcing variables X_t which are only backward-looking but stochastic. Then, the compact system $(1) - (2)$ can be transformed into its canonical purely-forward looking form in (6) and (2) decoupling the backward-looking and the forward-looking terms of the endogenous variables, as explained in Subsection 2.1. Whenever $\Gamma_0 \equiv (\mathbf{I}_k - \Phi_2 \Theta)$ is nonsingular, the canonical system of structural relationships for the endogenous variables implied by (6) can be rewritten as in (7). The resulting system of expectational difference equations in terms of transformed variables contains only forward-looking terms.

B Supplementary Materials

B.1 A Closed-Form Solution to the Quadratic Matrix Equation

The companion quadratic matrix equation in (5) can be solved for its (stable) real-valued solution Θ with a straightforward iterative algorithm (Binder and Pesaran (1995) and Binder and Pesaran (1997)). Alternatively, the stable solution of the quadratic matrix equation in (5) can be found in closed-form from the solution to a related generalized eigenvalue problem.

I construct a pair of $2m \times 2m$ companion matrix forms (D, E) where the $m \times m$ matrices Φ_1 and Φ_2 that describe (5) enter as follows:

$$D = \begin{bmatrix} \mathbf{I}_m & -\Phi_1 \\ \mathbf{I}_m & \mathbf{0}_m \end{bmatrix}, \quad E = \begin{bmatrix} \Phi_2 & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{I}_m \end{bmatrix}. \tag{84}$$

The solution to the generalized eigenvalue problem for the matrix pair (D, E) is a set of $2m$ eigenvalues q_k and their corresponding eigenvectors v_k such that $Dv_k = Ev_kq_k$. Assuming there are at least m stable eigenvalues (those who are inside the unit circle), I can then order the eigenvalues and their corresponding eigenvectors so that the m stable eigenvalues come first. I denote the diagonal matrix with the ordered eigenvalues as Q and the matrix of corresponding eigenvectors as V.

Partitioning the ordered matrix of eigenvalues Q to collect the first m stable eigenvalues and the remaining ones into block matrices and partitioning the matrix of eigenvectors V accordingly, I can write $Q \equiv \begin{bmatrix} Q^1 & \mathbf{0}_m \\ \mathbf{0}_m & Q^2 \end{bmatrix}$ and $V \equiv \begin{bmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{bmatrix}$ where each block matrix is of dimension $m \times m$. The generalized eigenvalue problem can then be stated in matrix form as:

$$\begin{bmatrix} \mathbf{I}_m & -\Phi_1 \\ \mathbf{I}_m & \mathbf{0}_m \end{bmatrix} \begin{bmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{bmatrix} = \begin{bmatrix} \Phi_2 & \mathbf{0}_m \\ \mathbf{0}_m & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{bmatrix} \begin{bmatrix} Q^1 & \mathbf{0}_m \\ \mathbf{0}_m & Q^2 \end{bmatrix}, \tag{85}$$

and from here it follows that:

$$\begin{bmatrix} V^{11} - \Phi_1 V^{21} & V^{12} - \Phi_1 V^{22} \\ V^{11} & V^{12} \end{bmatrix} = \begin{bmatrix} \Phi_2 V^{11} & \Phi_2 V^{12} \\ V^{21} & V^{22} \end{bmatrix} \begin{bmatrix} Q^1 & \mathbf{0}_m \\ \mathbf{0}_m & Q^2 \end{bmatrix} = \begin{bmatrix} \Phi_2 V^{11} Q^1 & \Phi_2 V^{12} Q^2 \\ V^{21} Q^1 & V^{22} Q^2 \end{bmatrix}. \tag{86}$$

This block system implies that $V^{11} = V^{21}Q^1$ and $V^{11} - \Phi_1 V^{21} = \Phi_2 V^{11} Q^1$. Substituting the first expression into the second one gives that $V^{21}Q^1 - \Phi_1 V^{21} = \Phi_2 V^{21} Q^1 Q^1$. Then, post-multiplying both sides by $\left(V^{21}\right)^{-1}$ and re-arranging, it follows that:

$$\Phi_2 \left(V^{21} \left(Q^1\right)^2 \left(V^{21}\right)^{-1}\right) - \left(V^{21} Q^1 \left(V^{21}\right)^{-1}\right) + \Phi_1 = \mathbf{0}_m. \tag{87}$$

Defining Θ to be $\Theta \equiv V^{21} Q^1 \left(V^{21}\right)^{-1}$ gives $\Theta^2 = V^{21} Q^1 \left(V^{21}\right)^{-1} V^{21} Q^1 \left(V^{21}\right)^{-1} = V^{21} \left(Q^1\right)^2 \left(V^{21}\right)^{-1}$. Therefore, under this definition of Θ, the matrix equation in (87) is observationally equivalent to the quadratic matrix equation (5) introduced in Sub-section 2.1 for the purpose of decoupling the backward-looking and forward-looking terms of the canonical first LRE model given by (1)−(2). Hence, this suffices to characterize the solution of the quadratic matrix equation based on its m stable (real-valued) eigenvalues.

If such a real-valued solution of the quadratic matrix equation exists with all its eigenvalues inside the unit circle, then it is straightforward to compute Θ ($m \times m$ matrix) by constructing the matrices D and

E and finding the ordered solution to the corresponding generalized eigenvalue problem for the matrix pair (D, E). The block matrix Q^1 contains the m stable eigenvalues and the block matrix V^{21} contains the corresponding eigenvectors from the generalized eigenvalue problem after partitioning. Then, the solution Θ for the quadratic matrix is simply $\Theta \equiv V^{21} Q^1 \left(V^{21} \right)^{-1}$. To check that this transformation Θ is valid, I only need to verify that all the eigenvalues collected in Q^1 are indeed inside the unit circle.[27]

B.2 Numerical Methods to Solve the Sylvester Matrix Equation

It follows from Proposition 1 that a unique solution C of the companion Sylvester matrix equation in (24) exists if and only if the $(m \cdot k) \times (m \cdot k)$ square matrix $\mathcal{A}$ is invertible. If so, the unique solution takes the form of a linear system with $m \cdot k$ equations and $m \cdot k$ unknowns given by $vec\left(C\right) = \mathcal{A}^{-1} vec\left(-G\right)$ which can be solved in $O\left(\left(m \cdot k\right)^3\right)$ operations. Although obtaining the solution in this way is straightforward, there are algorithms and methods that can improve the efficiency of the numerical computation of C.

(a) Schur's Triangulation and Recursive Implementation. **Step 1.** The first step of the approach is to implement the generalized Schur triangulation to re-write the companion Sylvester matrix equation in (24). I find the real Schur decompositions $F = UKU^T \in \mathbb{R}^{m \times m}$ and $A = VQV^T \in \mathbb{R}^{k \times k}$ where $m \geq k$. The matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{k \times k}$ are unitary matrices of dimensions m and k , respectively, such that $UU^T = U^TU = \mathbf{I}_m$ and $VV^T = V^TV = \mathbf{I}_k$. The matrices $K \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{k \times k}$, referred to as the corresponding Schur forms for F and A, respectively, are both upper triangular.[28] The eigenvalues of F and A are the diagonal entries of the (upper) triangular matrices K and Q, respectively. Hence, I can re-write the companion Sylvester matrix equation in (24)—i.e., $FC + C\left(-A^{-1}\right) = -G$ as $FCA - C = -GA$ with $F \in \mathbb{R}^{m \times m}$, $A \in \mathbb{R}^{k \times k}$, and $G \in \mathbb{R}^{m \times k}$ given and $C \in \mathbb{R}^{m \times k}$ to be determined—as follows:

$$KYQ - Y = R, \tag{88}$$

where $Y = U^TCV$ and $R = -U^TGAV$.

Step 2. The second step of the approach is to solve the transformed Sylvester matrix equation in (88). The transformed equation can be vectorized to obtain:

$$\overline{\mathcal{A}} := \left[\left(Q^T \otimes K\right) - \mathbf{I}_{m \cdot k}\right], \tag{89}$$

$$\overline{\mathcal{A}} vec\left(Y\right) = vec\left(R\right). \tag{90}$$

Then, this can be solved directly by calculating the inverse of $\overline{\mathcal{A}}$ and using standard matrix algebra to solve the linear system $vec\left(Y\right) = \overline{\mathcal{A}}^{-1} vec\left(R\right)$ for Y.

Step 3. The last step of the approach is to recover the solution C to the companion Sylvester matrix equation in (24). For that, I simply undo Schur's triangulation as follows $C = UYV^T$.

---

[27]The accompanying codes provided with the paper include a Matlab function that implements the iterative solution of Binder and Pesaran (1995) and Binder and Pesaran (1997) together with another Matlab function that obtains the solution using alternatively the generalized eigenvalue problem algorithm described here. For a discussion of the quadratic matrix equation and the iterative algorithm, the interested reader is also refered to Binder and Pesaran (1995) and Binder and Pesaran (1997) and their accompanying Matlab codes: https://ideas.repec.org/c/dge/qmrbcd/73.html.

[28]Since K is similar to F, both have the same eigenvalues. Since Q is similar to A, their eigenvalues are the same too.

Although this three-step approach works, the solution to the transformed system in $(89) - (90)$ can be further optimized under additional assumptions on the matrix F and, particularly, on the matrix A.

(a) The solution to the transformed Sylvester matrix equation given by $(89) - (90)$ permits a more efficient recursive implementation, if A is diagonalizable. The diagonalization theorem indicates that the $k \times k$ matrix A is diagonalizable if and only if A has k linearly independent eigenvectors (Strang (2016)). If A is diagonalizable, then the matrix S of its eigenvectors is invertible and $S^{-1}AS = M = diag(\mu_1, ..., \mu_k)$ is the diagonal matrix of its eigenvalues. A sufficient (but not necessary) condition for A to be diagonalizable is that all its k eigenvalues be distinct.[29] By the principal axis theorem, it follows that if A is a real matrix (i.e., all entries of A are real numbers) and symmetric (i.e., $A^T = A$), then A is diagonalizable as well (Strang (2016)). Hence, the additional assumption that A be a diagonalizable matrix does not appear to be too restrictive in most practical applications—given that the stochastic process for the forcing variables is often assumed symmetric—i.e., A is often posited as a real symmetric square matrix. Even when the symmetry assumption is relaxed, generally the eigenvalues appear as distinct.

Assuming that the matrix A is diagonalizable, I can re-write the companion Sylvester matrix equation in (24)—i.e., $FCA - C = -GA$—with the Schur triangulation of F as before but using the diagonalization of A to obtain:

$$K\widehat{Y}M - \widehat{Y} = \widehat{R}, \tag{91}$$

where $\widehat{Y} = U^T CS$ and $\widehat{R} = -U^T GAS$. Then, the transformed Sylvester matrix equation can be vectorized as:

$$\widehat{\mathcal{A}} := \left[\left(M^T \otimes K\right) - \mathbf{I}_{m \cdot k}\right], \tag{92}$$

$$\widehat{\mathcal{A}} vec\left(\widehat{Y}\right) = vec\left(\widehat{R}\right). \tag{93}$$

The matrices M^T and $\mathbf{I}_{m \cdot k}$ are diagonal, while K is an upper triangular matrix. As a result, it follows that $\widehat{\mathcal{A}}$ itself must be an upper triangular matrix. The resulting linear system can be solved directly by calculating the inverse of $\widehat{\mathcal{A}}$ and then using standard matrix algebra to solve $vec\left(\widehat{Y}\right) = \overline{\mathcal{A}}^{-1} vec\left(\widehat{R}\right)$ for the unknown matrix $\widehat{Y}$. The inverse of an upper triangular matrix is also upper triangular. Hence, the diagonalization of A reduces the number of calculations needed to compute the solution to the transformed Sylvester matrix equation.

Moreover, the resulting linear system lends itself to a recursive implementation that does not require the computation of the inverse of $\widehat{\mathcal{A}}$ explicitly. Let me define $\widehat{\mathcal{A}} = [\widehat{a}_{i,j}]_{i,j=1}^{m \cdot k} \in \mathbb{R}^{(m \cdot k) \times (m \cdot k)}$ as well as the column-vectors $vec\left(\widehat{Y}\right) = [\widehat{y}_i]_{i=1}^{m \cdot k}$ and $vec\left(\widehat{R}\right) = [\widehat{r}_i]_{i=1}^{m \cdot k}$. Then, for any given $j = 1, ..., m \cdot k$, it holds true that $\widehat{a}_{i,j} = 0$ for all $i = 1, ..., m \cdot k$ and $i < j$. It follows from here that $\widehat{a}_{m \cdot k, m \cdot k}\widehat{y}_{m \cdot k} = \widehat{r}_{m \cdot k}$ pins down $\widehat{y}_{m \cdot k}$. Given $\widehat{y}_{m \cdot k}$, the expression for $\widehat{y}_{m \cdot k - 1}$ can be immediately obtained from $\widehat{a}_{m \cdot k - 1, m \cdot k - 1}\widehat{y}_{m \cdot k - 1} + \widehat{a}_{m \cdot k - 1, m \cdot k}\widehat{y}_{m \cdot k} = \widehat{r}_{m \cdot k - 1}$. Given $\widehat{y}_{m \cdot k}$ and $\widehat{y}_{m \cdot k - 1}$, the expression for $\widehat{y}_{m \cdot k - 2}$ is derived from $\widehat{a}_{m \cdot k - 2, m \cdot k - 2}\widehat{y}_{m \cdot k - 2} + \widehat{a}_{m \cdot k - 2, m \cdot k - 1}\widehat{y}_{m \cdot k - 1} + \widehat{a}_{m \cdot k - 2, m \cdot k}\widehat{y}_{m \cdot k} = \widehat{r}_{m \cdot k - 2}$. And so on and so forth. Then, once the matrix $\widehat{Y}$ is completed in this recursive way, the last step of the procedure is to recover the solution C to the companion Sylvester matrix equation. For that, I simply undo the transformation as follows $C = U\widehat{Y}S^{-1}$.

(b) The computation of the matrix C can be further improved whenever F and A are both diagonalizable

---

[29] A matrix A can be diagonalizable, yet have repeated eigenvalues. For example, the identity matrix I_k is diagonal (hence diagonalizable), but has one eigenvalue repeated k times (i.e., $\mu_i = 1$ for all $i = 1, ..., k$).

matrices. The diagonalization theorem implies that the $m \times m$ square matrix F and the $k \times k$ matrix A are both diagonalizable if and only these matrices have m and k linearly independent eigenvectors, respectively. In other words, F and A are diagonalizable if and only if the rank of the matrix formed by the corresponding eigenvectors is m and k, respectively. I also know that if both matrices are real-valued and symmetric, they are also diagonalizable. Furthermore, if the eigenvalues of each matrix are distinct, this is sufficient (albeit not necessary) for each matrix to be diagonalizable. Assuming matrices F and A can be diagonalized—i.e., using $F = T\Lambda T^{-1}$ and $A = SMS^{-1}$ where $\Lambda = diag\,(\lambda_1, ..., \lambda_m)$ and $M = diag\,(\mu_1, ..., \mu_k)$—I obtain the following transformation of the companion Sylvester matrix equation in (24):

$$\left(T\Lambda T^{-1}\right) C \left(SMS^{-1}\right) - C = -GA. \tag{94}$$

Multiplying the left-hand side of this matrix equation by T^{-1} and the right-hand side by S, it follows that:

$$\Lambda T^{-1} C S M - T^{-1} C S = -T^{-1} G A S. \tag{95}$$

Let $\widetilde{Y} = T^{-1}CS$ and $\widetilde{R} = -T^{-1}GAS$. Then, I find that:

$$\Lambda \widetilde{Y} M - \widetilde{Y} = \widetilde{R}. \tag{96}$$

Denoting the (i, j)-th entry of $\widetilde{Y}$ as $\widetilde{y}_{ij}$ and the (i, j)-th entry of $\widetilde{R}$ as $\widetilde{r}_{ij}$, the diagonalized Sylvester matrix equation can be rewritten simple as:

$$\lambda_i \mu_j \widetilde{y}_{ij} - \widetilde{y}_{ij} = \widetilde{r}_{ij}, \ \forall i = 1, ..., m, \ \forall j = 1, ..., k, \tag{97}$$

which means that:

$$\widetilde{y}_{ij} = \frac{\widetilde{r}_{ij}}{\lambda_i \mu_j - 1}. \tag{98}$$

Since the eigenvectors and eigenvalues of a diagonalizable matrix can be found with only $O\left(\sqrt[2]{(m \cdot k)^3}\right)$ operations, the diagonalized Sylvester matrix equation can be solved more efficiently in this way. Then, the matrix C can be immediately recovered undoing the transformation as $C = T\widetilde{Y}S^{-1}$.

(b) Bartels-Stewart Approach and Hessenberg-Schur Algorithm. A related classical numerical algorithm for solving the companion Sylvester matrix equation is the Bartels-Stewart algorithm which also makes use of the Schur decompositions of F and A in order to obtain a more efficient algorithm to compute the solution C (Bartels and Stewart (1972)). Using a Schur decomposition as before, the companion Sylvester matrix equation—i.e. $FCA - C = -GA$ with $F \in \mathbb{R}^{m \times m}$, $A \in \mathbb{R}^{k \times k}$, and $G \in \mathbb{R}^{m \times k}$ given and $C \in \mathbb{R}^{m \times k}$ to be determined—can be re-written as in equation (88). Let Q_{ij} denote a block of the upper triangular matrix Q, and let Y and R be partitioned according to a column partitioning of Q. The key step is to exploit these facts to decompose the transformed Sylvester matrix equation in (88) into smaller Sylvester matrix

equations by blocks as follows:

$$KY_1Q_{11} - Y_1 = R_1, \tag{99}$$

$$KY_jQ_{jj} - Y_j = R_j - K\sum_{i=1}^{j-1} Y_jQ_{ij}, \ \forall j = 2, ..., k. \tag{100}$$

Each of the block equations in $(99) - (100)$ takes the form of the transformed Sylvester matrix equation in (88) given that the sum that appears on the right-hand side of (100) is recursively known, as indicated above.[30]

An improved modification of the Bartels-Stewart algorithm, known as the Hessenberg-Schur algorithm, was proposed in Golub et al. (1979). This algorithm uses the Hessenberg decomposition instead of the Schur decomposition to transform the companion Sylvester matrix equation (Golub and van Loan (1996), Anderson et al. (1996)). The Hessenberg decomposition implies $F = UKU^H$ and $A = VQV^H$ where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{k \times k}$ are unitary matrices of dimension m and k, respectively. The matrices U^H and V^H denote the corresponding conjugate transpose of those matrices and $K \in \mathbb{R}^{m \times m}$ and $Q \in \mathbb{R}^{k \times k}$ are the corresponding Hessenberg matrices.

The Sylvester matrix equation is a special case of the Lyapunov equation. Hence, the *dlyap* function in the Control Systems Toolbox in Matlab which solves the discrete-time Lyapunov equation can be used to solve the companion Sylvester matrix equation in (24) as follows: $C = dlyap\,(F, A, GA)$. This function uses the SLICOT (Subroutine Library In COntrol Theory) library—with a routine that implements the Hessenberg-Schur algorithm. Starting with R2014a, the matrix C can also be computed in Matlab directly with the *sylvester* function as $C = sylvester\,(F, -inv\,(A), -GA)$.[31]

(c) Doubling Algorithm. The doubling algorithm exploits the convergence result posited in Proposition 3 which establishes that, for any matrix $G \in \mathbb{R}^{m \times k}$, the companion Sylvester matrix equation in (24) has a unique exact solution given by (32) whenever $\lambda_i \mu_j < 1$ for all $i = 1, ..., m$ and all $j = 1, ..., k$ where $\lambda_1, ..., \lambda_m$ are the eigenvalues of F and $\mu_1, ..., \mu_k$ are the eigenvalues of A. The doubling algorithm defines the following sequence:

$$\begin{aligned} \Lambda_{s+1} &= \Lambda_s \Lambda_s, \\ \Psi_{s+1} &= \Psi_s \Psi_s, \\ C_{s+1} &= C_s + \Lambda_s C_s \Psi_s, \end{aligned} \tag{101}$$

where $\Lambda_0 = F$, $\Psi_0 = A$, and $C_0 = GA$. This sequence converges to the solution C of the companion Sylvester matrix equation as $s \to \infty$. By repeated substitution, it can be shown that each iteration doubles the number of terms in the sum—hence the name of the algorithm—such that:

$$C_r = \sum_{s=0}^{2^r - 1} F^s\,(GA)\,A^s, \tag{102}$$

becomes arbitrarily close to the solution $C = \sum_{s=0}^{\infty} F^s\,(GA)\,A^s$ as r gets arbitrarily large. Further discussion of this algorithm and Matlab codes to implement it can be found, among others, in Anderson et al. (1996).

---

[30] If from the Schur decomposition Q is a real matrix, then Q_{jj} for all $j = 1, 2, ..., k$ must be either a scalar or a 2×2 matrix.

[31] Further details on standard implementation methods using Matlab can be found in Sima and Benner (2015). For further references on the Matlab function *dlyap*, see: http://www.mathworks.com/help/control/ref/dlyap.html and http://slicot.org/matlab-toolboxes/basic-control/basic-control-fortran-subroutines. For reference on the Matlab function *sylvester*, see: http://www.mathworks.com/help/matlab/ref/sylvester.html.