Quantitative Economics 7 (2016), 637–669 1759-7331/20160637

Perturbation methods for Markov-switching dynamic stochastic general equilibrium models

Andrew Foerster
Federal Reserve Bank of Kansas City

Juan F. Rubio-Ramírez
Emory University, Federal Reserve Bank of Atlanta, CEPR, FEDEA, and BBVA Research

Daniel F. Waggoner
Federal Reserve Bank of Atlanta

Tao Zha
Federal Reserve Bank of Atlanta, Emory University, and NBER

Markov-switching dynamic stochastic general equilibrium (MSDSGE) modeling has become a growing body of literature on economic and policy issues related to structural shifts. This paper develops a general perturbation methodology for constructing high-order approximations to the solutions of MSDSGE models. Our new method—"the partition perturbation method"—partitions the Markov-switching parameter space to keep a maximum number of time-varying parameters from perturbation. For this method to work in practice, we show how to reduce the potentially intractable problem of solving MSDSGE models to the manageable problem of solving a system of quadratic polynomial equations. This approach allows us to first obtain all the solutions and then determine how many

Andrew Foerster: andrew.foerster@kc.frb.org
Juan F. Rubio-Ramírez: juan.rubio-ramirez@emory.edu
Daniel F. Waggoner: dwaggoner@frbatlanta.org
Tao Zha: zmail@tzha.net

Detailed comments from two anonymous referees have led to significant improvement of earlier drafts. We thank Rhys Bidder, Han Chen, Seonghoon Cho, Lars Hansen, Giovanni Lombardo, Leonardo Melosi, Oreste Tristani, Harald Uhlig, as well as seminar participants at Duke University, the Federal Reserve Bank of St. Louis, the 2010 Society of Economic Dynamics Meetings, the 2011 Federal Reserve System Committee on Business and Financial Analysis Conference, the 2012 Annual Meeting of the American Economic Association, the 8th Dynare Conference, and the 2012 NBER Workshop on Methods and Applications for DSGE Models for helpful comments. Zhao Li and Tong Xu provided excellent research assistance. This research is supported in part by the National Science Foundation Grants SES-1127665 and SES-1227397 and by the National Natural Science Foundation of China Research Grants 71473168 and 71473169. Rubio-Ramírez also thanks the Institute for Economic Analysis (IAE), the Programa de Excelencia en Educacíon e Investigacíon of the Bank of Spain, and the Spanish Ministry of Science and Technology Ref. ECO2011-30323-c03-01 for support. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Banks of Atlanta and Kansas City, the Federal Reserve System, or the National Bureau of Economic Research.

Copyright © 2016 Andrew Foerster, Juan F. Rubio-Ramírez, Daniel F. Waggoner, and Tao Zha. Licensed under the Creative Commons Attribution-NonCommercial License 3.0. Available at http://www.qeconomics.org.
DOI: 10.3982/QE596

of them are stable. We illustrate the tractability of our methodology through two revealing examples.

KEYWORDS. Partition principle, naive perturbation, quadratic polynomial system, Taylor series, high-order expansion, time-varying coefficients, nonlinearity, Gröbner bases.

JEL CLASSIFICATION. C6, E3, G1.

1. INTRODUCTION

In this paper we extend the conventional perturbation method, as described in Judd (1998) and Schmitt-Grohe and Uribe (2004) and advocated recently by Lombardo (2010) and Borovička and Hansen (2013), to approximating the solutions of Markov-switching dynamic stochastic general equilibrium (MSDSGE) models. The extension poses a very challenging task because the presence of time-varying parameters in MSDSGE models makes high-order approximations potentially intractable. We advance the literature in three significant respects. First, we develop a general methodology for approximating the solution to a wide class of Markov-switching models with any order of accuracy. Second, our methodology preserves the time-varying coefficients to the maximum extent in high-order Taylor series expansions. Third, we show the feasibility and practicality of constructing high-order approximations by reducing the potentially intractable problem to the manageable problem of solving a system of quadratic polynomial equations.

The literature on Markov-switching linear rational expectations (MSLRE) models has been an active field in empirical macroeconomics (Leeper and Zha (2003), Blake and Zampolli (2006), Svensson and Williams (2007), Davig and Leeper (2007), and Farmer, Waggoner, and Zha (2009)). Building on standard linear rational expectations models, the MSLRE approach allows parameters to change over time according to discrete Markov processes. This nonlinearity has proven to be important in explaining shifts in monetary policy and macroeconomic time series (Schorfheide (2005), Davig and Doh (2008), Liu, Waggoner, and Zha (2011), and Bianchi (2010)) and in modeling the expected effects of future fiscal policy changes (Davig, Leeper, and Walker (2010, 2011),Bi and Traum (2012), Bianchi and Melosi (2013)). In particular, Markov-switching models provide a tractable way to study how agents form expectations over possible discrete changes in the economy, such as those in technology and policy.

There are, however, two major shortcomings with the MSLRE approach advocated by Farmer, Waggoner, and Zha (2011). First, the approach begins with a system of standard linear rational expectations equations that have been obtained by linearizing equilibrium conditions as though the parameters were constant over time. Discrete Markov processes are then annexed to certain parameters. As a consequence, the resultant MSLRE model may be incompatible with the optimizing behavior of agents in an original economic model with Markov-switching parameters. Second, because it builds on linear rational expectations models, the MSLRE approach does not take into account

higher-order coefficients in the approximation. Not only do higher-order approximations improve the approximation accuracy but they are essential to addressing important questions such as whether time-varying volatility is the driving force of fluctuations in the financial markets and business cycles (Bloom (2009)).

This paper develops a general perturbation methodology for constructing first-order and second-order approximations to the solutions of MSDSGE models in which certain parameters vary over time according to discrete Markov processes.[1] The key is to derive high-order approximations to the equilibrium conditions implied by the original nonlinear economic model when Markov-switching parameters are present. Our methodology, therefore, overcomes the serious shortcomings associated with the MSLRE shortcut. By working with the original MSDSGE model directly rather than taking a system of linear rational expectations equations with fixed parameters as a shortcut, we maintain the congruity between the original economic model with Markov-switching parameters and the resultant approximations to the model solution. Such congruity is necessary for researchers to derive both first-order and higher-order approximations *consistent with the original nonlinear model*. Our general methodology leads to several developments as follows.

- We show that the steady state must be independent of the realization of any regime in the discrete Markov process governing parameter changes. We follow the literature and define the steady state with the ergodic mean values of Markov-switching parameters. One natural extension of the conventional perturbation method commonly used for dynamic stochastic general equilibrium (DSGE) models with no time-varying parameters is to perturb all Markov-switching parameters around their ergodic mean values. We call this the *naive perturbation method*.

- Since certain Markov-switching parameters such as time-varying volatilities do not influence the steady state, we develop a rigorous framework called the *partition principle* for partitioning the Markov-switching parameter space such that those Markov-switching parameters are not perturbed. By not perturbing the Markov-switching parameters that have no bearing on the steady state, we preserve the original Markov-switching nonlinearity in first-order as well as higher-order approximations. This preservation improves approximation accuracy, especially at low orders, in comparison to the naive perturbation method. We call this newly developed method the *partition perturbation method*. We provide a revealing Markov-switching model to illustrate the importance of our methodology. In addition, we use a Markov-switching real business cycle (RBC) model as a more realistic example to demonstrate that the partition perturbation method delivers more accurate first-order and second-order approximations than the naive perturbation method.

- Much of the MSDSGE literature focuses on a first-order approximation with MSLRE models. One exception is Amisano and Tristani (2011), who extend the literature to a second-order approximation but with only Markov-switching shock variances. We show

[1]We show in the paper that one can extend our methodology to higher-order approximations through standard linear algebra.

that our methodology is tractable and general enough to allow for Markov-switching coefficients in the DSGE model in high-order approximations without much additional computational burden.

- We show that any finite-order approximation to the model solution can be reduced to the manageable problem of solving a system of quadratic polynomial equations. The rest of the approximation involves solving a system of linear equations recursively—a key insight of our methodology. This result is powerful because it provides a viable way of approximating the solution of an MSDSGE model at a high order without incurring much computational time. Obtaining such a result is difficult because Markov switching compounds the complexity of implicit differentiation when deriving the Taylor series expansion. The most difficult part is the potentially rampant notation that inhibits the reader from following and implementing our methodology. Our notation makes transparent to the reader (as well as us) that simple linear algebra is all researchers need to accomplish high-order approximations, even in the presence of time-varying coefficients in the Taylor series expansion.

- We first use the Gröbner-bases method to obtain all solutions and then determine how many of these solutions are stable according to the mean-square-stability criterion (Costa, Fragoso, and Marques (2005) and Farmer, Waggoner, and Zha (2009)). This procedure enables researchers to ascertain both the existence and the uniqueness of a stable solution.

The rest of the paper is organized as follows. Section 2 presents the framework for solving a general class of MSDSGE models. We outline our methodology, review the conventional perturbation method, extend this commonly used method to the naive perturbation method, and develop the partition perturbation method according to the partition principle. Section 3 derives both first-order and second-order approximations that have convenient forms for researchers to use. We show how to reduce the complex Markov-switching problem to solving a system of quadratic polynomial equations. We prove that the rest of the approximation of any order involves simple linear algebra. Section 4 discusses different approaches to solving a system of quadratic polynomial equations and reviews the concept of mean square stability to obtain a stable solution. Section 5 uses a simple Markov-switching model to illustrate why the partition perturbation method is more accurate than the naive perturbation method. Section 6 applies our methodology to a Markov-switching RBC model and compares approximation errors between the two perturbation methods. Replication files for numerical results in Sections 5 and 6 are available in a supplementary file on the journal website, http://qeconomics.org/supp/596/code_and_data.zip. Concluding remarks are offered in Section 7.

2. The framework

This section establishes the theoretical foundation of our proposed partition perturbation method for a general class of MSDSGE models. We present the class of MSDSGE models and introduce the key idea of partitioning the Markov-switching parameter

space. Based on this idea we propose the partition perturbation method and highlight the importance of our method in contrast to the naive perturbation method that derives directly from the conventional perturbation method, which has been used for DSGE models. Throughout the paper, we use a stylized real business cycle (RBC) model as an illustrative example to guide the reader through our new methodology.

2.1 *A general class of MSDSGE models*

We study a general class of MSDSGE models in which some of the parameters follow a discrete Markov process indexed by $s_t \in \{1, \dots, n_s\}$ with the transition matrix $\mathbf{P} = [p_{s_t, s_{t+1}}]$. The element $p_{s_t, s_{t+1}}$ represents the probability of s_{t+1} at time $t+1$ conditional on observing s_t at time t. We denote the time t vector of all Markov-switching parameters by $\boldsymbol{\theta}(s_t) \in \mathbb{R}^{n_\theta}$.[2] We assume that the Markov process is ergodic and denote the n_s vector of ergodic probabilities by $\bar{p}$. The ergodic mean of $\boldsymbol{\theta}(s_t)$ is $\bar{\boldsymbol{\theta}} = [\boldsymbol{\theta}(1) \cdots \boldsymbol{\theta}(n_s)]\bar{p}$.

Given the vector of state variables $(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, s_t)$, the equilibrium conditions for MSDSGE models have the general form

$$\mathbb{E}_t \mathbf{f}\big(\mathbf{y}_{t+1}, \mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_{t+1}, \boldsymbol{\varepsilon}_t, \boldsymbol{\theta}(s_{t+1}), \boldsymbol{\theta}(s_t)\big) = \mathbf{0}_{n_y+n_x}, \tag{1}$$

where $\mathbb{E}_t$ denotes the mathematical expectation operator conditional on information available at time t, $\mathbf{y}_t \in \mathbb{R}^{n_y}$ is a vector of non-predetermined (control) variables, $\mathbf{x}_t \in \mathbb{R}^{n_x}$ is a vector of (endogenous and exogenous) predetermined variables, $\mathbf{0}_{n_y+n_x}$ is an $(n_y + n_x)$ vector of zeros, and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^{n_\varepsilon}$ is a vector of independent and identically distributed (i.i.d.) innovations to the exogenous predetermined variables with $\mathbb{E}_t \boldsymbol{\varepsilon}_{t+1} = \mathbf{0}_{n_\varepsilon}$ and $\mathbb{E}_t \boldsymbol{\varepsilon}_{t+1} \boldsymbol{\varepsilon}_{t+1}^\intercal = \mathbf{I}_{n_\varepsilon}$. The superscript $\intercal$ indicates the transpose of a matrix or a vector and $\mathbf{I}_{n_\varepsilon}$ denotes the $n_\varepsilon \times n_\varepsilon$ identity matrix. The function $\mathbf{f}$ is defined on an open subset of $\mathbb{R}^{n_f}$, where $n_f = 2(n_y + n_x + n_\varepsilon + n_\theta)$, and its range is a subset of $\mathbb{R}^{n_y+n_x}$. We make the following assumptions about $\mathbf{f}$ throughout the paper. These assumptions are satisfied by almost all economic models.

ASSUMPTION 1. *The function* $\mathbf{f}$ *is infinitely differentiable with respect to all arguments.*

ASSUMPTION 2. *Integration and differentiation of* $\mathbf{f}$ *are exchangeable.*

ASSUMPTION 3. *There exist the steady-state values* $\mathbf{y}_{ss}$ *and* $\mathbf{x}_{ss}$ *such that*

$$\mathbf{f}(\mathbf{y}_{ss}, \mathbf{y}_{ss}, \mathbf{x}_{ss}, \mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, \mathbf{0}_{n_\varepsilon}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}) = \mathbf{0}_{n_y+n_x}. \tag{2}$$

We use a simple RBC model to illustrate how the equilibrium conditions can be arranged in the form of (1). Consider an economy with the representative household whose preferences over a stochastic sequence of consumption goods, c_t, are represented by the expected utility function

$$\max \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^\upsilon}{\upsilon},$$

[2]The parameters that are constant over time, which we call *constant parameters* for the rest of the paper, are not included in the vector $\boldsymbol{\theta}(s_t)$. Unless otherwise stated, all vectors in this paper are column vectors.

where β is the discount factor and υ relates to risk aversion. The resource constraint is

$$c_t + k_t = z_t^{1-\alpha} k_{t-1}^{\alpha} + (1-\delta)k_{t-1},$$

where δ is the rate of depreciation, k_t is a stock of physical capital, and z_t represents a technological process that evolves according to

$$\log \frac{z_t}{z_{t-1}} = \big(1-\rho(s_t)\big)\mu(s_t) + \rho(s_t)\log \frac{z_{t-1}}{z_{t-2}} + \sigma(s_t)\varepsilon_t,$$

where $\varepsilon_t \sim N(0,1)$ is a standard normal random variable. The drift, persistence, and volatility parameters are time-varying with $s_t \in \{1,2\}$. The three equations characterizing the equilibrium are the equation describing the technological process and the two first-order equations

$$c_t^{\upsilon-1} = \beta \mathbb{E}_t c_{t+1}^{\upsilon-1}\big[\alpha z_{t+1}^{1-\alpha} k_t^{\alpha-1} + (1-\delta)\big],$$
$$c_t + k_t = z_t^{1-\alpha} k_{t-1}^{\alpha} + (1-\delta)k_{t-1}.$$

The economy is nonstationary. To obtain a stationary equilibrium we define $\tilde{z}_t = \frac{z_t}{z_{t-1}}$, $\tilde{k}_t = \frac{k_t}{z_t}$, and $\tilde{c}_t = \frac{c_t}{z_{t-1}}$. The stationary equilibrium conditions summarized by (1) can be specifically expressed as

$$\begin{aligned}\mathbf{0}_3 &= \mathbb{E}_t \mathbf{f}\big(\mathbf{y}_{t+1}, \mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_{t+1}, \boldsymbol{\varepsilon}_t, \boldsymbol{\theta}(s_{t+1}), \boldsymbol{\theta}(s_t)\big) \\ &= \mathbb{E}_t \begin{bmatrix} \tilde{c}_t^{\upsilon-1} - \beta \tilde{z}_t^{\upsilon-1} \tilde{c}_{t+1}^{\upsilon-1}\{\alpha e^{[(1-\rho(s_{t+1}))\mu(s_{t+1}) + \rho(s_{t+1})\log(\tilde{z}_t) + \sigma(s_{t+1})\varepsilon_{t+1}](1-\alpha)} \tilde{k}_t^{\alpha-1} + 1 - \delta\} \\ \tilde{c}_t + \tilde{z}_t \tilde{k}_t - \tilde{z}_t^{1-\alpha} \tilde{k}_{t-1}^{\alpha} - (1-\delta)\tilde{k}_{t-1} \\ \log \tilde{z}_t - \big(1-\rho(s_t)\big)\mu(s_t) - \rho(s_t)\log \tilde{z}_{t-1} - \sigma(s_t)\varepsilon_t \end{bmatrix},\end{aligned} \tag{3}$$

where $\mathbf{y}_t = \tilde{c}_t$, $\mathbf{x}_t = [\tilde{k}_t \tilde{z}_t]^{\intercal}$, $\boldsymbol{\varepsilon}_t = \varepsilon_t$, and $\boldsymbol{\theta}(s_t) = [\mu(s_t)\rho(s_t)\sigma(s_t)]^{\intercal}$. The dimensions of this RBC model are $n_y = 1$, $n_x = 2$, $n_\varepsilon = 1$, $n_\theta = 3$, and $n_s = 2$.

2.2 *The conventional perturbation method*

Before we propose our partition perturbation method for solving MSDSGE models, we review the conventional perturbation method used for solving constant-parameter DSGE models (Judd (1998), Schmitt-Grohe and Uribe (2004), Lombardo (2010), Holmes (2012), Borovička and Hansen (2013), Gomme and Klein (2011)). The constant-parameter model can be considered as a special Markov-switching model with either $n_s = 1$ or $\boldsymbol{\theta}(s_t) = \bar{\boldsymbol{\theta}}$ for all s_t.

The conventional perturbation method begins with positing that the solutions $\mathbf{y}_t$ and $\mathbf{x}_t$ are of the form[3]

$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi), \tag{4}$$

$$\mathbf{x}_t = \mathbf{h}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi), \tag{5}$$

[3]Some researchers may prefer to perturb $\boldsymbol{\varepsilon}_t$ in addition to $\boldsymbol{\varepsilon}_{t+1}$. To do so, one would replace equations (4), (5), and (8) with $\mathbf{y}_t = \tilde{\mathbf{g}}(\mathbf{x}_{t-1}, \chi\boldsymbol{\varepsilon}_t, \chi)$, $\mathbf{x}_t = \tilde{\mathbf{h}}(\mathbf{x}_{t-1}, \chi\boldsymbol{\varepsilon}_t, \chi)$, and

$$\mathbf{0}_{n_y+n_x} = \int_{\mathbb{R}^{n_\varepsilon}} \mathbf{f}\big(\mathbf{g}(\mathbf{x}_t, \chi\boldsymbol{\varepsilon}_{t+1}, \chi), \mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \chi\boldsymbol{\varepsilon}_{t+1}, \chi\boldsymbol{\varepsilon}_t, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}\big)\, d\mu(\boldsymbol{\varepsilon}_{t+1}).$$

where $\mathbf{g} : \mathbb{R}^{n_x+n_\varepsilon+1} \to \mathbb{R}^{n_y}$ and $\mathbf{h} : \mathbb{R}^{n_x+n_\varepsilon+1} \to \mathbb{R}^{n_x}$ are functions with the Taylor series representation about the point $(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0)$ satisfying

$$\mathbf{y}_{ss} = \mathbf{g}(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0), \tag{6}$$

$$\mathbf{x}_{ss} = \mathbf{h}(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0), \tag{7}$$

and $\chi \in \mathbb{R}$ is the perturbation parameter. The conventional perturbation is a method that recursively finds the Taylor series expansion of $\mathbf{g}$ and $\mathbf{h}$ by positing that equations (4) and (5) are a solution of

$$\begin{aligned}\mathbf{0}_{n_y+n_x} &= \mathbf{F}(\mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) \\ &\equiv \int_{\mathbb{R}^{n_\varepsilon}} \mathbf{f}\big(\mathbf{g}(\mathbf{x}_t, \chi\boldsymbol{\varepsilon}_{t+1}, \chi), \mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \chi\boldsymbol{\varepsilon}_{t+1}, \boldsymbol{\varepsilon}_t, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}\big)\, d\mu(\boldsymbol{\varepsilon}_{t+1})\end{aligned} \tag{8}$$

for all $\mathbf{x}_{t-1}$, $\boldsymbol{\varepsilon}_t$, and χ, where $\mu(\boldsymbol{\varepsilon}_{t+1})$ is a σ-finite measure on the space of $\boldsymbol{\varepsilon}_{t+1}$. When $\chi = 1$, equation (8) reduces to equation (1). By construction, $\mathbf{g}$ and $\mathbf{h}$ satisfy equation (8) when $\mathbf{x}_{t-1} = \mathbf{x}_{ss}$, $\boldsymbol{\varepsilon}_t = \mathbf{0}_{n_\varepsilon}$, and $\chi = 0$.

To form the Taylor series expansion of $\mathbf{g}$ and $\mathbf{h}$, one must be able to compute the derivatives of $\mathbf{g}$ and $\mathbf{h}$ and evaluate these derivatives at the point $(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0)$. By repeated implicit differentiation of equation (8), one can recursively solve for the derivatives of $\mathbf{g}$ and $\mathbf{h}$ evaluated at $(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0)$.

2.3 *The naive perturbation method*

It is natural and straightforward to extend the conventional perturbation method discussed in Section 2.2 to MSDSGE models. Suppose that $\mathbf{y}_t$ and $\mathbf{x}_t$ are of the form

$$\mathbf{y}_t = \mathbf{g}_{s_t}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi), \tag{9}$$

$$\mathbf{x}_t = \mathbf{h}_{s_t}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) \tag{10}$$

for all s_t, where $\mathbf{g}_{s_t} : \mathbb{R}^{n_x+n_\varepsilon+1} \to \mathbb{R}^{n_y}$ and $\mathbf{h}_{s_t} : \mathbb{R}^{n_x+n_\varepsilon+1} \to \mathbb{R}^{n_x}$ are continuously differentiable functions. In the constant-parameter case, the choice of the steady state as the approximation point is natural and one needs to perturb $\boldsymbol{\varepsilon}_{t+1}$ only. The choice of approximation point in the Markov-switching case is more involved and takes two steps. First, we show that the steady state in the Markov-switching case must be independent of regime s_t.

Suppose that the steady-state variables $\mathbf{x}_{ss}(s_t)$ depend on regime s_t. As in the constant-parameter case, we must choose the values of $\mathbf{g}_{s_t}(\mathbf{x}_{ss}(s_t), \mathbf{0}_{n_\varepsilon}, 0)$ and $\mathbf{h}_{s_t}(\mathbf{x}_{ss}(s_t), \mathbf{0}_{n_\varepsilon}, 0)$ such that

$$\begin{aligned}&\mathbf{f}\big(\mathbf{g}_{s_{t+1}}\big(\mathbf{h}_{s_t}\big(\mathbf{x}_{ss}(s_t), \mathbf{0}_{n_\varepsilon}, 0\big), \mathbf{0}_{n_\varepsilon}, 0\big), \mathbf{g}_{s_t}\big(\mathbf{x}_{ss}(s_t), \mathbf{0}_{n_\varepsilon}, 0\big), \\ &\quad \mathbf{h}_{s_t}\big(\mathbf{x}_{ss}(s_t), \mathbf{0}_{n_\varepsilon}, 0\big), \mathbf{x}_{ss}(s_t), \mathbf{0}_{n_\varepsilon}, \mathbf{0}_{n_\varepsilon}, \boldsymbol{\theta}(s_{t+1}), \boldsymbol{\theta}(s_t)\big) = \mathbf{0}_{n_y+n_x}\end{aligned} \tag{11}$$

The resultant Taylor expansion of $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{h}}$ around $\mathbf{x}_{t-1}$, $\boldsymbol{\varepsilon}_t$, and χ can be obtained from the original Taylor expansion of $\mathbf{g}$ and $\mathbf{h}$ by simply substituting $\chi\boldsymbol{\varepsilon}_t$ for $\boldsymbol{\varepsilon}_t$. When $\chi = 1$ the two techniques produce the same result, but the alternative perturbation approach requires higher-order terms to achieve the same accuracy of approximation.

for all s_t and s_{t+1}. Because the value of $\mathbf{g}_{s_{t+1}}$ is evaluated at the point $(\mathbf{x}_{ss}(s_{t+1}), \mathbf{0}_{n_\varepsilon}, 0)$, it follows that $\mathbf{x}_{ss}(s_{t+1}) = \mathbf{h}_{s_t}(\mathbf{x}_{ss}(s_t), \mathbf{0}_{n_\varepsilon}, 0)$ for all s_t and s_{t+1}. For the latter relationship to hold, it must be that $\mathbf{x}_{ss}(s_t) = \mathbf{x}_{ss}$ and $\mathbf{x}_{ss}(s_{t+1}) = \mathbf{x}_{ss}$ for all s_t and s_{t+1}. That is, the steady state must be regime independent.

Second, we show that the Markov-switching parameters $\boldsymbol{\theta}(s_{t+1})$ and $\boldsymbol{\theta}(s_t)$ must in general be perturbed. Since $\mathbf{x}_{ss}(s_t) = \mathbf{x}_{ss}$ for all s_t, the system of equations (11) becomes

$$\mathbf{f}\big(\mathbf{g}_{s_{t+1}}(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0), \mathbf{g}_{s_t}(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0), \mathbf{x}_{ss}, \mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, \mathbf{0}_{n_\varepsilon}, \boldsymbol{\theta}(s_{t+1}), \boldsymbol{\theta}(s_t)\big) = \mathbf{0}_{n_y+n_x}. \quad (12)$$

This is a system of $n_s^2(n_y + n_x)$ equations with $n_s n_y + n_x$ unknowns (n_y unknowns in each $\mathbf{g}_k(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0)$ for $1 \le k \le n_s$ and another n_x unknowns in $\mathbf{x}_{ss}$), which cannot be solved in general. We must, therefore, perturb the Markov-switching parameters to reduce the number of equations.

One natural approach is to define a perturbation function for Markov-switching parameters by

$$\boldsymbol{\theta}(k, \chi) = \chi\boldsymbol{\theta}(k) + (1 - \chi)\bar{\boldsymbol{\theta}} \quad (13)$$

for $1 \le k \le n_s$. When $\chi = 0$, we have $\boldsymbol{\theta}(k, 0) = \bar{\boldsymbol{\theta}}$; when $\chi = 1$, we have $\boldsymbol{\theta}(k, 1) = \boldsymbol{\theta}(k)$. Given $\mathbf{x}_{ss}(k) = \mathbf{x}_{ss}$ for $1 \le k \le n_s$, we have the following assumption.

ASSUMPTION 4. *The function* $\mathbf{g}_k(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0)$ *has the same value for all* $1 \le k \le n_s$. *We denote this value by* $\mathbf{y}_{ss}$.

With this perturbation and Assumption 4, system (12) becomes

$$\mathbf{f}(\mathbf{y}_{ss}, \mathbf{y}_{ss}, \mathbf{x}_{ss}, \mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, \mathbf{0}_{n_\varepsilon}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}) = \mathbf{0}_{n_y+n_x}.$$

By Assumption 3 there is a solution to this system of equations.

For illustration we return to the RBC model in which the system of equations $\mathbf{f}$ is given by (3). Let the steady state and ergodic mean values of parameters be denoted by $\mathbf{y}_{ss} = \tilde{c}_{ss}$, $\mathbf{x}_{ss} = [\tilde{k}_{ss} \quad \tilde{z}_{ss}]^\intercal$, and $\bar{\boldsymbol{\theta}} = [\bar{\mu} \quad \bar{\rho} \quad \bar{\sigma}]^\intercal$. The steady state must satisfy

$$\begin{aligned}\mathbf{0}_3 &= \mathbf{f}(\mathbf{y}_{ss}, \mathbf{y}_{ss}, \mathbf{x}_{ss}, \mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, \mathbf{0}_{n_\varepsilon}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}) \\ &= \begin{bmatrix} \tilde{c}_{ss}^{\upsilon-1} - \beta\tilde{z}_{ss}^{\upsilon-1}\tilde{c}_{ss}^{\upsilon-1}\big\{\alpha e^{[(1-\bar{\rho})\bar{\mu}+\bar{\rho}\log(\tilde{z}_{ss})](1-\alpha)}\tilde{k}_{ss}^{\alpha-1} + 1 - \delta\big\} \\ \tilde{c}_{ss} + \tilde{z}_{ss}\tilde{k}_{ss} - \tilde{z}_{ss}^{1-\alpha}\tilde{k}_{ss}^{\alpha} - (1-\delta)\tilde{k}_{ss} \\ \log\tilde{z}_{ss} - (1-\bar{\rho})\bar{\mu} - \bar{\rho}\log\tilde{z}_{ss} \end{bmatrix}.\end{aligned} \quad (14)$$

Solving for the steady state is the same as in the constant-parameter case. With the perturbation function (13), it is straightforward to write down an equation analog of the constant-parameter case (8) and obtain the Taylor series expansions for $\mathbf{g}_{s_t}$ and $\mathbf{h}_{s_t}$ around the point $(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0)$. We call this approach the *naive perturbation method*. In Section 5 we show, through a revealing example, why this method is naive in comparison to the alternative perturbation method developed below.

2.4 *The partition perturbation method*

The steady state expressed in (14) can be obtained in closed form as

$$\begin{aligned}
&\tilde{z}_{ss} = e^{\bar{\mu}},\\
&\tilde{k}_{ss} = \big(\alpha^{-1}e^{(\alpha-1)\bar{\mu}}\big(\beta^{-1}e^{(1-\upsilon)\bar{\mu}} - 1 + \delta\big)\big)^{1/(\alpha-1)}, \quad \text{and}\\
&\tilde{c}_{ss} = \tilde{k}_{ss}\big(1 - \delta - e^{\bar{\mu}} + \alpha^{-1}\big(\beta^{-1}e^{(1-\upsilon)\bar{\mu}} - 1 + \delta\big)\big).
\end{aligned}$$

Clearly, the steady-state solution does not depend on either $\bar{\rho}$ or $\bar{\sigma}$. As argued in Section 2.3, the purpose of perturbing the Markov-switching parameters around their ergodic mean values is to solve the steady state when the perturbation parameter χ and the innovations $\boldsymbol{\varepsilon}_t$ are set to 0. Since $\rho(s_t)$ and $\sigma(s_t)$ do not influence the steady state, perturbing these parameters generates unnecessary approximations. If we do not perturb these parameters, we maintain the Markov-switching nonlinearity along the direction of these parameters in the original model. We formalize this idea by proposing the perturbation function

$$\boldsymbol{\theta}(k,\chi) = \chi\begin{bmatrix}\boldsymbol{\theta}_1(k)\\ \boldsymbol{\theta}_2(k)\end{bmatrix} + (1-\chi)\begin{bmatrix}\bar{\boldsymbol{\theta}}_1\\ \boldsymbol{\theta}_2(k)\end{bmatrix} = \begin{bmatrix}\bar{\boldsymbol{\theta}}_1 + \chi\big(\boldsymbol{\theta}_1(k) - \bar{\boldsymbol{\theta}}_1\big)\\ \boldsymbol{\theta}_2(k)\end{bmatrix} \tag{15}$$

for $1 \le k \le n_s$ with the partition principle stated below.

Partition Principle. *Let the Markov-switching parameters be ordered and partitioned as* $\boldsymbol{\theta}^{\intercal}(s_t) = [\boldsymbol{\theta}_1^{\intercal}(s_t) \quad \boldsymbol{\theta}_2^{\intercal}(s_t)]$. *The second block* $\boldsymbol{\theta}_2(s_t)$ *is chosen to contain the maximum number of Markov-switching parameters such that*

$$\begin{aligned}
&\mathbf{f}\big(\mathbf{y}_{ss}, \mathbf{y}_{ss}, \mathbf{x}_{ss}, \mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, \mathbf{0}_{n_\varepsilon}, \boldsymbol{\theta}(s_{t+1}, 0), \boldsymbol{\theta}(s_t, 0)\big)\\
&\quad = \mathbf{f}(\mathbf{y}_{ss}, \mathbf{y}_{ss}, \mathbf{x}_{ss}, \mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, \mathbf{0}_{n_\varepsilon}, \bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}) = \mathbf{0}_{n_y+n_x}
\end{aligned} \tag{16}$$

for all s_t *and* s_{t+1}.

According to the partition principle, the second block of Markov-switching parameters is not perturbed. Since perturbation is necessary *only for approximations to the original nonlinear model*, the fewer number of Markov-switching parameters we perturb, the more accurate are finite-order approximations. We illuminate this point through examples discussed in Sections 5 and 6.

It is practicable to implement the partition principle. Whenever we write down DSGE models, we should be able to write down the steady-state equilibrium conditions and identify which Markov-switching parameters have no influence on these conditions. We group all such Markov-switching parameters into $\boldsymbol{\theta}_2(s_t)$ as long as the critical system (16) is satisfied. Verifying whether (16) holds is straightforward.

To obtain the analog of system (8), we define the continuously differentiable function $\mathbf{F}_{s_t} : \mathbb{R}^{n_y+2n_x+n_\varepsilon+1} \to \mathbb{R}^{n_y+n_x}$ as

$$\mathbf{F}_{s_t}(\mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = \sum_{s_{t+1}=1}^{n_s} p_{s_t,s_{t+1}} \int_{\mathbb{R}^{n_\varepsilon}} \mathbf{f}\big(\mathbf{g}_{s_{t+1}}(\mathbf{x}_t, \chi\boldsymbol{\varepsilon}_{t+1}, \chi), \mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \chi\boldsymbol{\varepsilon}_{t+1}, \boldsymbol{\varepsilon}_t, \boldsymbol{\theta}(s_{t+1}, \chi), \boldsymbol{\theta}(s_t, \chi)\big)\, d\mu(\boldsymbol{\varepsilon}_{t+1})$$

such that (9) and (10) are a solution to

$$\mathbf{F}_{s_t}(\mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = \mathbf{0}_{n_y+n_x} \tag{17}$$

for all s_t, $\mathbf{x}_{t-1}$, $\boldsymbol{\varepsilon}_t$, and χ. The perturbation functions $\boldsymbol{\theta}(s_{t+1}, \chi)$ and $\boldsymbol{\theta}(s_t, \chi)$ are given by (15). When $\chi = 1$, the perturbed system (17) reduces to the original system (1). By construction, system (17) is satisfied for all s_t when $\mathbf{y}_t = \mathbf{y}_{ss}$, $\mathbf{x}_t = \mathbf{x}_{t-1} = \mathbf{x}_{ss}$, $\boldsymbol{\varepsilon}_t = \mathbf{0}_{n_\varepsilon}$, and $\chi = 0$. We call this approach the partition perturbation method.

Like the conventional perturbation method or the naive perturbation method, the partition perturbation method allows one to solve recursively for the partial derivatives of $\mathbf{g}_{s_t}$ and $\mathbf{h}_{s_t}$ by repeated implicit differentiation of system (17) and evaluate these derivative at $(\mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, 0)$. Unlike those perturbation methods, the partial derivatives of $\mathbf{g}_{s_t}$ and $\mathbf{h}_{s_t}$ depend on the partial derivatives of $\mathbf{f}$ evaluated at

$$\big(\mathbf{y}_{ss}, \mathbf{y}_{ss}, \mathbf{x}_{ss}, \mathbf{x}_{ss}, \mathbf{0}_{n_\varepsilon}, \mathbf{0}_{n_\varepsilon}, \boldsymbol{\theta}(s_{t+1}, 0), \boldsymbol{\theta}(s_t, 0)\big).$$

Because the second block of Markov-switching parameters is not perturbed, the Taylor series coefficients for $\mathbf{g}_{s_t}$ and $\mathbf{h}_{s_t}$ are in general time-varying when the set containing $\boldsymbol{\theta}_2(s_t)$ is not empty. The presence of such time-varying Taylor series coefficients makes high-order approximations a potentially intractable problem. One principal contribution of this paper is to prove that the partition perturbation method can be implemented by reducing this potentially intractable problem to a recursive problem involving only simple linear algebra once we remove the bottleneck of solving a system of quadratic polynomial equations. This theoretical result is provided in Section 3. In Section 5 we provide a revealing Markov-switching dynamic equilibrium example that has closed-form solutions. Using this example we illustrate that the partition principle delivers a more accurate solution than the naive perturbation method for an approximated solution of any order.

3. First-order and second-order approximations

This section gives a detailed description of how to derive first-order and second-order approximations to the model solution by using the partition perturbation method. We present the results up to only second order to conserve space, but it is straightforward to derive higher-order approximations with a similar approach. To make our theoretical results transparent to a general reader, we develop notation that proves crucial to the clarity of our derivations; moreover, it enables us to show that Markov-switching volatility (uncertainty) has first-order effects on dynamics while the naive perturbation method nullifies such effects by construction.

3.1 *Notation*

We stack the regime-dependent solutions (9) and (10) as

$$\mathbf{Y}_t = \mathbf{G}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = \begin{bmatrix} \mathbf{g}_1(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) \\ \vdots \\ \mathbf{g}_{n_s}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) \end{bmatrix} \quad \text{and}$$

$$\mathbf{X}_t = \mathbf{H}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = \begin{bmatrix} \mathbf{h}_1(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) \\ \vdots \\ \mathbf{h}_{n_s}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) \end{bmatrix}.$$

Define $\mathbf{Y}_{ss} = \mathbf{1}_{n_s} \otimes \mathbf{y}_{ss}$ and $\mathbf{X}_{ss} = \mathbf{1}_{n_s} \otimes \mathbf{x}_{ss}$, where $\mathbf{1}_{n_s}$ is the n_s vector of 1s. It follows that $\mathbf{y}_t = \mathbf{g}_{s_t}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = (\mathbf{e}_{s_t}^\intercal \otimes \mathbf{I}_{n_y})\mathbf{Y}_t$ and $\mathbf{x}_t = \mathbf{h}_{s_t}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = (\mathbf{e}_{s_t}^\intercal \otimes \mathbf{I}_{n_x})\mathbf{X}_t$ for all s_t, where $\mathbf{e}_k$, for $1 \leq k \leq n_s$, is the kth column of the $n_s \times n_s$ identity matrix. Approximating a solution to $\mathbf{y}_t$ and $\mathbf{x}_t$ is equivalent to approximating a solution to $\mathbf{Y}_t$ and $\mathbf{X}_t$.

Define $\mathbb{F}_i : \mathbb{R}^{n_s n_y + n_s n_x + n_x + n_\varepsilon + 1} \to \mathbb{R}^{n_y + n_x}$ for $i = 1, \ldots, n_s$ by

$$\mathbb{F}_i(\mathbf{Y}_t, \mathbf{X}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = \mathbf{F}_i\big((\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_y})\mathbf{Y}_t, (\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x})\mathbf{X}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi\big)$$

and $\mathbb{F} : \mathbb{R}^{n_s n_y + n_s n_x + n_x + n_\varepsilon + 1} \to \mathbb{R}^{n_s(n_y + n_x)}$ by

$$\mathbb{F}(\mathbf{Y}_t, \mathbf{X}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = \begin{bmatrix} \mathbb{F}_1(\mathbf{Y}_t, \mathbf{X}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) \\ \vdots \\ \mathbb{F}_{n_s}(\mathbf{Y}_t, \mathbf{X}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) \end{bmatrix}.$$

With these definitions, system (17) is equivalent to

$$\mathbb{F}(\mathbf{Y}_t, \mathbf{X}_t, \mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = \mathbf{0}_{n_s(n_y + n_x)}. \tag{18}$$

We now introduce a derivative notation that is used throughout the paper. Let $\mathbf{w}(\mathbf{u})$ be a continuously differentiable function from $\mathbb{R}^{n_u}$ into $\mathbb{R}^{n_w}$. Let u_ℓ be the ℓth component of $\mathbf{u}$ for $1 \leq \ell \leq n_u$ and let $w^k(\mathbf{u})$ be the kth component of $\mathbf{w}(\mathbf{u})$ for $1 \leq k \leq n_w$. The term $D_\ell w^k(\mathbf{u})$, a real number, denotes the partial derivative of w^k with respect to u_ℓ evaluated at the point $\mathbf{u}$. The term $D\mathbf{w}(\mathbf{u})$, the $n_w \times n_u$ matrix $[D_\ell w^k(\mathbf{u})]$ for $1 \leq k \leq n_w$ and $1 \leq \ell \leq n_u$, denotes the total derivative of $\mathbf{w}$ evaluated at the point $\mathbf{u}$.

As for second-order partial derivatives, let $D_{\ell_2} D_{\ell_1} w^k(\mathbf{u})$, a real number, denote the second partial derivative of w^k with respect to u_{ℓ_1} and u_{ℓ_2} evaluated at $\mathbf{u}$. The term $D_{\ell_2} D_{\ell_1} \mathbf{w}(\mathbf{u})$ denotes the n_w vector $[D_{\ell_2} D_{\ell_1} w^k(\mathbf{u})]$ for $1 \leq k \leq n_w$. The term $D_{\ell_2} D\mathbf{w}(\mathbf{u})$ denotes the $n_w \times n_u$ matrix $[D_{\ell_2} D_\ell w^k(\mathbf{u})]$ for $1 \leq k \leq n_w$ and $1 \leq \ell \leq n_u$. It is straightforward to extend this notation to higher-order partial derivatives.

If $\mathbf{w}(\mathbf{u}, \mathbf{v})$ is a continuously differentiable function from $\mathbb{R}^{n_u + n_v}$ into $\mathbb{R}^{n_w}$, we use $D_{\mathbf{u}}\mathbf{w}(\mathbf{u}, \mathbf{v})$ to denote the $n_w \times n_u$ matrix consisting of the first n_u columns of the $n_w \times (n_u + n_v)$ matrix $D\mathbf{w}(\mathbf{u}, \mathbf{v})$. Similarly, $D_{\mathbf{v}}\mathbf{w}(\mathbf{u}, \mathbf{v})$ denotes the last n_v columns of $D\mathbf{w}(\mathbf{u}, \mathbf{v})$ and $D\mathbf{w}(\mathbf{u}, \mathbf{v}) = [D_{\mathbf{u}}\mathbf{w}(\mathbf{u}, \mathbf{v}) \quad D_{\mathbf{v}}\mathbf{w}(\mathbf{u}, \mathbf{v})]$.

3.2 *First-order approximation*

Denote $\mathbf{z}_t^\intercal = [\,\mathbf{x}_{t-1}^\intercal \quad \boldsymbol{\varepsilon}_t^\intercal \quad \chi\,]$ and $\mathbf{z}_{ss}^\intercal = [\,\mathbf{x}_{ss}^\intercal \quad \mathbf{0}_{n_\varepsilon}^\intercal \quad 0\,]$. The dimension of both $\mathbf{z}_t$ and $\mathbf{z}_{ss}$ is $n_z = n_x + n_\varepsilon + 1$. The first-order approximation of $\mathbf{G}(\mathbf{z}_t)$ and $\mathbf{H}(\mathbf{z}_t)$ is

$$\mathbf{G}(\mathbf{z}_t) \approx \mathbf{Y}_{ss} + D\mathbf{G}(\mathbf{z}_{ss})(\mathbf{z}_t - \mathbf{z}_{ss}),$$
$$\mathbf{H}(\mathbf{z}_t) \approx \mathbf{X}_{ss} + D\mathbf{H}(\mathbf{z}_{ss})(\mathbf{z}_t - \mathbf{z}_{ss}).$$

The following proposition shows that both $D\mathbf{G}(\mathbf{z}_{ss}) = [D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss}) \quad D_{\boldsymbol{\varepsilon}_t}\mathbf{G}(\mathbf{z}_{ss}) \quad D_\chi \mathbf{G}(\mathbf{z}_{ss})]$ and $D\mathbf{H}(\mathbf{z}_{ss}) = [D_{\mathbf{x}_{t-1}}\mathbf{H}(\mathbf{z}_{ss}) \quad D_{\boldsymbol{\varepsilon}_t}\mathbf{H}(\mathbf{z}_{ss}) \quad D_\chi \mathbf{H}(\mathbf{z}_{ss})]$ can be obtained by solving a system of quadratic polynomial equations and two systems of linear equations.

PROPOSITION 1. *Under Assumptions* 1–4, *the matrices* $D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})$ *and* $D_{\mathbf{x}_{t-1}}\mathbf{H}(\mathbf{z}_{ss})$ *can be obtained by solving a system of* $n_s(n_y + n_x)n_x$ *quadratic polynomial equations with* $n_s(n_y + n_x)n_x$ *unknowns. Given a solution to this quadratic polynomial system, the matrices* $D_{\boldsymbol{\varepsilon}_t}\mathbf{G}(\mathbf{z}_{ss})$ *and* $D_{\boldsymbol{\varepsilon}_t}\mathbf{H}(\mathbf{z}_{ss})$ *can be obtained by solving a system of* $n_s(n_y + n_x)n_\varepsilon$ *linear equations with* $n_s(n_y + n_x)n_\varepsilon$ *unknowns; the vectors* $D_\chi \mathbf{G}(\mathbf{z}_{ss})$ *and* $D_\chi \mathbf{H}(\mathbf{z}_{ss})$ *can be obtained by solving a system of* $n_s(n_y + n_x)$ *linear equations with* $n_s(n_y + n_x)$ *unknowns.*

The proofs for Propositions 1–3 are given in Appendix A.

The proof of Proposition 1 shows how to represent the first-order solution in a form that can be implemented in practice. More important is the result that reduces the potentially intractable problem of solving MSDSGE models to the manageable problem of solving a system of quadratic polynomial equations. Section 4 provides an effective way of solving this problem.

3.3 *Characterizing the first-order approximation*

As shown in the proof of Proposition 1, the slope coefficient matrices, represented by $D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})$ and $D_{\mathbf{x}_{t-1}}\mathbf{H}(\mathbf{z}_{ss})$, and the impact coefficient matrices, represented by $D_{\boldsymbol{\varepsilon}_t}\mathbf{G}(\mathbf{z}_{ss})$ and $D_{\boldsymbol{\varepsilon}_t}\mathbf{H}(\mathbf{z}_{ss})$, are functions of the partial derivatives $D_{\mathbf{y}_{t+1}}\mathbf{f}(\mathbf{u}_{ss})$, $D_{\mathbf{y}_t}\mathbf{f}(\mathbf{u}_{ss})$, $D_{\mathbf{x}_t}\mathbf{f}(\mathbf{u}_{ss})$, $D_{\mathbf{x}_{t-1}}\mathbf{f}(\mathbf{u}_{ss})$, and $D_{\boldsymbol{\varepsilon}_t}\mathbf{f}(\mathbf{u}_{ss})$, where

$$\mathbf{u}_{ss}^\intercal = \left[\mathbf{y}_{ss}^\intercal, \mathbf{y}_{ss}^\intercal, \mathbf{x}_{ss}^\intercal, \mathbf{x}_{ss}^\intercal, \mathbf{0}_{n_\varepsilon}^\intercal, \mathbf{0}_{n_\varepsilon}^\intercal, \boldsymbol{\theta}(s_{t+1}, 0)^\intercal, \boldsymbol{\theta}(s_t, 0)^\intercal\right].$$

Thus the slope and impact coefficients depend, in general, on both $\boldsymbol{\theta}_2(s_{t+1})$ and $\boldsymbol{\theta}_2(s_t)$. When the naive perturbation method is used, by contrast, the slope and impact coefficients depend only on $\bar{\boldsymbol{\theta}}$, not on $\boldsymbol{\theta}(s_{t+1})$ or $\boldsymbol{\theta}(s_t)$, as stated in the following corollary.

COROLLARY 1. *Let Assumptions* 1–4 *hold. Under the naive perturbation method, the first-order coefficients* $D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})$, $D_{\mathbf{x}_{t-1}}\mathbf{H}(\mathbf{z}_{ss})$, $D_{\boldsymbol{\varepsilon}_t}\mathbf{G}(\mathbf{z}_{ss})$, *and* $D_{\boldsymbol{\varepsilon}_t}\mathbf{H}(\mathbf{z}_{ss})$ *do not depend on* $\boldsymbol{\theta}(s_t)$, *but are functions of* $\bar{\boldsymbol{\theta}}$ *only.*

For our RBC model summarized in (3), one can see that $D_{\mathbf{y}_{t+1}}\mathbf{f}(\mathbf{u}_{ss})$ depends on $\rho(s_{t+1})$ and $\sigma(s_{t+1})$, $D_{\mathbf{x}_t}\mathbf{f}(\mathbf{u}_{ss})$ depends on $\rho(s_{t+1})$ and $\sigma(s_{t+1})$, $D_{\mathbf{x}_{t-1}}\mathbf{f}(\mathbf{u}_{ss})$ depends on $\rho(s_{t+1})$, and $D_{\boldsymbol{\varepsilon}_t}\mathbf{f}(\mathbf{u}_{ss})$ depends on $\sigma(s_{t+1})$. Thus, both the Markov-switching persistence

and volatility parameters have first-order effects. By contrast, these effects are muted by the naive perturbation method because the partial derivatives of $\mathbf{f}$ depend only on $\bar{\rho}$ and $\bar{\sigma}$.[4] Consequently the finite-order approximation becomes less accurate. In Section 6 we provide a numerical assessment of this accuracy by computing approximation errors of the Euler equations.

3.4 *Second-order approximation*

The second-order approximation is represented by

$$\begin{aligned}\mathbf{G}(\mathbf{z}_t) &\approx \mathbf{Y}_{ss} + D\mathbf{G}(\mathbf{z}_{ss})(\mathbf{z}_t - \mathbf{z}_{ss}) \\ &\quad + \frac{1}{2}\sum_{\ell_1=1}^{n_z}\sum_{\ell_2=1}^{n_z} D_{\ell_2}D_{\ell_1}\mathbf{G}(\mathbf{z}_{ss})(z_{t,\ell_1} - z_{ss,\ell_1})(z_{t,\ell_2} - z_{ss,\ell_2}), \\ \mathbf{H}(\mathbf{z}_t) &\approx \mathbf{X}_{ss} + D\mathbf{H}(\mathbf{z}_{ss})(\mathbf{z}_t - \mathbf{z}_{ss}) \\ &\quad + \frac{1}{2}\sum_{\ell_1=1}^{n_z}\sum_{\ell_2=1}^{n_z} D_{\ell_2}D_{\ell_1}\mathbf{H}(\mathbf{z}_{ss})(z_{t,\ell_1} - z_{ss,\ell_1})(z_{t,\ell_2} - z_{ss,\ell_2}),\end{aligned}$$

where $z_{t,\ell}$ and $z_{ss,\ell}$ are the ℓth components of $\mathbf{z}_t$ and $\mathbf{z}_{ss}$. The following proposition delivers a powerful result that the vector $D_{\ell_2}D_{\ell_1}\mathbf{G}(\mathbf{z}_{ss})$ and $D_{\ell_2}D_{\ell_1}\mathbf{H}(\mathbf{z}_{ss})$ can be obtained through simple linear algebra.

PROPOSITION 2. *Under Assumptions* 1–4 *and given a first-order approximation, the vectors* $D_{\ell_2}D_{\ell_1}\mathbf{G}(\mathbf{z}_{ss})$ *and* $D_{\ell_2}D_{\ell_1}\mathbf{H}(\mathbf{z}_{ss})$, *for* $1 \le \ell_1, \ell_2 \le n_z$, *can be obtained by solving a system of* $n_s(n_y + n_x)n_z^2$ *linear equations in* $n_s(n_y + n_x)n_z^2$ *unknowns.*

Because the coefficients represented by $D_{\ell_2}D_{\ell_1}\mathbf{G}(\mathbf{z}_{ss})$ and $D_{\ell_2}D_{\ell_1}\mathbf{H}(\mathbf{z}_{ss})$ can be time-varying, it is not at all obvious that Proposition 2 would hold. One of the principal developments in this paper is to reduce the potentially unmanageable complexity of the Markov-switching problem to a straightforward linear algebra problem for higher-order approximations. The Markov-switching problem is potentially unmanageable because time-varying coefficients make the model inherently nonlinear for any finite-order approximation and especially for higher-order approximation. In the proof of Proposition 2 we show that, with a careful application of implicit differentiation, the second-order approximation simply requires solving a system of linear equations *even in the presence of Markov-switching coefficients*.[5] As the second-order coefficients are functions of the first-order coefficients, Markov-switching volatility has both first-order and second-order effects on the slope and impact coefficients.

[4]The naive perturbation method resembles the existing methods for solving DSGE models with drifting parameters, where the slope and impact coefficients in the first-order approximation are *not* time-varying (Fernández-Villaverde, Guerrón-Quintana, and Rubio-Ramírez (2014)).

[5]Armed with our notation and applying the same technique, one can prove that the approximation of any higher order involves solving a system of linear equations recursively. We leave the derivation to the reader.

4. Removing the bottleneck

Propositions 1 and 2 show how to translate the complex Markov-switching DSGE problem into a simple linear algebra problem, as long as one is able to solve for $D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})$ and $D_{\mathbf{x}_{t-1}}\mathbf{H}(\mathbf{z}_{ss})$. As indicated by Proposition 1, the bottleneck involves solving a system of quadratic polynomial equations. In this section we first discuss different approaches to solving this quadratic system and then present the mean-square-stability (MSS) criterion for selecting a stable solution to the first-order Taylor series expansion of $\mathbf{G}$ and $\mathbf{H}$. Higher-order expansions can be derived recursively from a first-order stable solution as shown in Section 3.4.

4.1 *Solving polynomial equations*

When there are no Markov-switching parameters, the system of quadratic polynomial equations (see system (A.4) in Appendix A) collapses to a special form that can be solved by using the generalized Schur decomposition (Klein (2000)). When Markov-switching parameters are present, however, the system of $n_s(n_y + n_x)n_x$ quadratic polynomial equations in $n_s(n_y + n_x)n_x$ unknowns are no longer of this special form and the general Schur technique is no longer applicable.

The literature has proposed numerical methods for the solution (Svensson and Williams (2007), Farmer, Waggoner, and Zha (2011), Cho (2011)). Another approach is to apply Gröbner bases to find all the solutions (see Appendix B). This approach is a potentially powerful tool. The trade-off between the existing numerical methods and the Gröbner-bases approach is computing time. Numerical methods can be used for large DSGE models but may not find all the solutions, while the Gröbner-bases approach may be computationally costly for large DSGE models but can find all the solutions. Researchers should use their own judgment and experience in deciding which method is most efficient for their own particular application. For the two models studied in this paper, it turns out that there is a unique stable solution. In this paper we apply Gröbner bases to these models for obtaining all the solutions to the system of quadratic polynomial equations. When it is computationally feasible to apply Gröbner bases, we recommend using this approach because it does not rest on arbitrary starting points required by existing numerical methods. After we obtain all the solutions, we utilize the MSS criterion (discussed below) to ascertain the uniqueness of a stable first-order solution.

4.2 *Mean square stability*

In the case of constant-parameter DSGE models, whether the first-order approximation is stable or not can be determined by verifying whether its largest absolute generalized eigenvalue is greater than or equal to 1, a condition that holds for most concepts of stability. In the MSDSGE case, the problem is both subtle and complicated because alternative concepts of stability would imply different kinds of solutions. Given the first-order approximation, we use the concept of mean square stability (MSS) as defined in Costa, Fragoso, and Marques (2005) and advocated by Farmer, Waggoner, and Zha (2009). The

MSS criterion states that a solution is stable if and only if all the eigenvalues of the $n_s n_x^2 \times n_s n_x^2$ matrix

$$\left(P^{\intercal} \otimes I_{n_x^2}\right) \operatorname{diag}[\, D_{\mathbf{x}_{t-1}} h_1 \otimes D_{\mathbf{x}_{t-1}} h_1 \quad \dots \quad D_{\mathbf{x}_{t-1}} h_{n_s} \otimes D_{\mathbf{x}_{t-1}} h_{n_s} \,]$$

are inside the unit circle, where diag denotes the block diagonal matrix with the $D_{\mathbf{x}_{t-1}} h_k \otimes D_{\mathbf{x}_{t-1}} h_k$, for $k = 1, \dots, n_s$, along the diagonal. The $n_x \times n_x$ matrices $D_{\mathbf{x}_{t-1}} h_k$ are obtained by reading off the appropriate rows of the matrix $D_{\mathbf{x}_{t-1}} \mathbf{H}$. In particular we have

$$D_{\mathbf{x}_{t-1}} \mathbf{H} = \begin{bmatrix} D_{\mathbf{x}_{t-1}} h_1 \\ \vdots \\ D_{\mathbf{x}_{t-1}} h_{n_s} \end{bmatrix}.$$

5. Understanding the partition perturbation method

In the preceding sections we develop the partition perturbation method and show how to use it for obtaining first-order and second-order approximations to the solutions of MSDSGE models. In this section we use a simple dynamic equilibrium model to reveal the power of the partition perturbation method in comparison to the naive perturbation method. The model is particularly instructive because we can obtain a closed-form solution, which allows us to show that the naive partition method incurs needless approximation errors in the Taylor series expansion, especially in low-order expansions.

Consider a simple inflation model in which the nominal interest rate is linked to the real interest rate and the expected inflation rate by the Fisher equation

$$R_t = r + \mathbb{E}_t \pi_{t+1},$$

where R_t is the nominal interest rate at time t, π_{t+1} is the inflation rate at time $t+1$, and the steady-state real interest rate $r = R - \pi$. Monetary policy follows the rule

$$R_t = R + \phi(s_t)(\pi_t - \pi) + \sigma(s_t)\varepsilon_t,$$

where the monetary policy shock ε_t is an i.i.d. normal random variable. A positive monetary policy shock raises the nominal interest rate and lowers inflation. Denoting $\hat{\pi}_t = \pi_t - r$ and combining the previous two equations lead to

$$\phi(s_t)\hat{\pi}_t + \sigma(s_t)\varepsilon_t = \mathbb{E}_t \hat{\pi}_{t+1}. \tag{19}$$

Suppose that $s_t \in \{1, 2\}$ follows a two-state Markov process. Because of the presence of Markov-switching parameters $\phi(s_t)$ and $\sigma(s_t)$, equation (19) is in essence a nonlinear model.

To write this model in the same form as (1), we define a new variable such that $\pi_t^* = \pi_t$ and let $\mathbf{y}_t = \pi_t^*$ and $\mathbf{x}_t = \pi_t$. We thus have $\mathbf{y}_{ss} = \pi$ and $\mathbf{x}_{ss} = \pi$. To use the partition perturbation method, we follow the partition principle and partition the Markov-switching parameters so that no Markov-switching parameter is perturbed. Specifically,

$\boldsymbol{\theta}(k,\chi) = [\,\phi(k) \quad \sigma(k)\,]^{\intercal}$. The equilibrium conditions can be expressed as

$$\begin{aligned}&\mathbb{E}_t\mathbf{f}\big(\mathbf{y}_{t+1},\mathbf{y}_t,\mathbf{x}_t,\mathbf{x}_{t-1},\chi\boldsymbol{\varepsilon}_{t+1},\boldsymbol{\varepsilon}_t,\boldsymbol{\theta}(s_{t+1},\chi),\boldsymbol{\theta}(s_t,\chi)\big)\\&\quad=\mathbb{E}_t\begin{bmatrix}\big(1-\phi(s_t)\big)\pi+\phi(s_t)\pi_t-\pi^*_{t+1}-\sigma(s_t)\varepsilon_t\\ \pi^*_t-\pi_t\end{bmatrix}\end{aligned}\tag{20}$$

such that

$$\mathbf{f}\big(\mathbf{y}_{ss},\mathbf{y}_{ss},\mathbf{x}_{ss},\mathbf{x}_{ss},\mathbf{0},\mathbf{0},\boldsymbol{\theta}(j,0),\boldsymbol{\theta}(i,0)\big)=\mathbf{0}\quad\text{for } 1\le i,j\le 2.$$

PROPOSITION 3. *With the partition perturbation method, a first-order approximation to the nonlinear model* (20) *is an exact solution and there are no higher-order Taylor series expansions* (*i.e., higher-order coefficients are all* 0).

The proof of Proposition 3 in Appendix A shows that the implication of Proposition 3 is more general than the result specific to model (19) or (20). For MSLRE models in the Markov-switching literature, a first-order solution generated by the partition perturbation method delivers an exact solution. Indeed, applying the partition perturbation method to our example yields the exact solution as

$$\hat{\pi}_t=-\frac{\sigma(s_t)}{\phi(s_t)}\varepsilon_t.$$

By contrast, the naive perturbation method perturbs the Markov-switching parameters as

$$\boldsymbol{\theta}(k,\chi)=\begin{bmatrix}\bar{\phi}\\ \bar{\sigma}\end{bmatrix}+(1-\chi)\begin{bmatrix}\phi(s_t)-\bar{\phi}\\ \sigma(s_t)-\bar{\sigma}\end{bmatrix},$$

where $\bar{\phi}$ and $\bar{\sigma}$ are the ergodic means of $\phi(s_t)$ and $\sigma(s_t)$. The first-order approximation generated by the naive perturbation method is $\pi_t=-(\bar{\sigma}/\bar{\phi})\varepsilon_t$ for all s_t. Clearly, this solution is not exact and higher-order Taylor series expansions are needed to improve the solution accuracy.

We demonstrate these results numerically with the parameterization $p_{1,1}=0.95$, $p_{2,2}=0.85$, $\phi(1)=1.25$, $\phi(2)=0.96$, $\sigma(1)=0.1$, and $\sigma(2)=0.6$. The Gröbner-bases analysis gives four solutions for this parameterization, but only one is stable according to the MSS criterion. The first-order stable approximation generated by the partition perturbation method is

$$\hat{\pi}_t=-0.08\varepsilon_t\quad\text{for } s_t=1\quad\text{and}\quad\hat{\pi}_t=-0.625\varepsilon_t\quad\text{for } s_t=2.$$

Because all higher-order coefficients are exactly zero, the first-order approximation is the exact solution.

TABLE 1. Euler-equation errors (base-10 log absolute value).

		Perturbation Method		
		Partition	Naive	
	Exact	First Order	First Order	Second Order
EE	$-\infty$	$-\infty$	-0.5564	-1.3691

Similarly, the first-order stable approximation produced by the naive perturbation method is

$$\hat{\pi}_t = -0.191083\varepsilon_t \quad \text{for } s_t = 1 \quad \text{and} \quad \hat{\pi}_t = -0.191083\varepsilon_t \quad \text{for } s_t = 2.$$

Because all the Markov-switching parameters are perturbed according to (13), the first-order solution does not depend on the realization of a particular regime. The regime-dependent nature relies on the second-order solution

$$\hat{\pi}_t = 0.0447610\varepsilon_t \quad \text{for } s_t = 1 \quad \text{and} \quad \hat{\pi}_t = -0.8986170\varepsilon_t \quad \text{for } s_t = 2.$$

How does this approximation compare to the exact solution? To assess the accuracy of the two perturbation methods, we compute Euler-equation errors (EEs) as suggested in Judd (1998). Table 1 reports the base-10 log absolute value of the approximation error for the original nonlinear equation (19), where the initial condition is set as $\varepsilon_t = 1.0$. We discuss the reason for using the base-10 log value in Section 6.3.

Given the simplicity of this model, we can compute EEs without any simulation. Since the first-order solution generated by the partition perturbation method is the exact solution (Proposition 3), the absolute value of the approximation error is 0 (the log absolute value of the error is $-\infty$). On the other hand, the naive perturbation method relies on higher-order approximations to get a more accurate solution. As indicated in Table 1, the second-order approximation obtained by the naive perturbation method is closer to the exact solution with a much smaller approximation error than the error generated by the first-order approximation, but it is still not close to the exact solution. This example clearly illustrates the importance of the partition perturbation method in obtaining an accurate low-order approximation.

6. APPLICATION TO THE RBC MODEL

In this section we apply the partition perturbation method to the two-state Markov-switching RBC model introduced in Section 2.1. We then compare approximation errors generated by the partition perturbation method with those incurred by the naive perturbation method to asses the accuracy of both methods.

The parameterization we use is presented in Table 2 and it is motivated by business-cycle facts related to emerging markets. The value of β corresponds to a real rate of 3 percent in steady state, the value of α corresponds to a capital share of one-third,

TABLE 2. The parameterization for the Markov-switching RBC model.

α	β	υ	δ	$\mu(1)$	$\mu(2)$	$\rho(1)$	$\rho(2)$	$\sigma(1)$	$\sigma(2)$	$p_{1,1}$	$p_{2,2}$
0.33	0.9976	−1	0.025	0.0274	−0.0337	0.1	0.0	0.0072	0.0216	0.75	0.5

and the value of δ corresponds to an annual capital depreciation rate of approximately 10 percent. The growth-rate parameters $\mu(1)$ and $\mu(2)$ and the standard deviation parameters $\sigma(1)$ and $\sigma(2)$ are set to make the output growth and its unconditional variance differ across regimes in magnitudes consistent with the emerging markets such as the Argentinian economy (Fernández-Villaverde and Rubio-Ramírez (2007), Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez, and Uribe (2011)). Note that the first regime is associated with positive growth while the second is associated with negative growth. Moreover, the first regime is less volatile and more persistent than the second regime. Given this parameterization, the stationary steady-state values of consumption, capital, and technology are $\tilde{c}_{ss} = 2.08259$, $\tilde{k}_{ss} = 22.1504$, and $\tilde{z}_{ss} = 1.007$. Denote $\hat{c}_t = \tilde{c}_t - \tilde{c}_{ss}$, $\hat{k}_t = \tilde{k}_t - \tilde{k}_{ss}$, and $\hat{z}_t = \tilde{z}_t - \tilde{z}_{ss}$.

6.1 *Solution from the partition perturbation method*

For the first-order approximation, the Gröbner-bases approach delivers four solutions. According to the MSS criterion, only one of these solutions is stable. We report below the second-order approximation associated with the unique stable solution:

$$\begin{bmatrix} \hat{c}_t \\ \hat{k}_t \\ \hat{z}_t \end{bmatrix} = \begin{bmatrix} 0.0405 & 0.1264 & 0.0091 & 0.000049 \\ 0.9692 & -2.1406 & -0.1552 & -0.3720 \\ 0.0 & 0.1 & 0.0072 & 0.0184 \end{bmatrix} \begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix}$$

$$+ \frac{1}{2} \begin{bmatrix} -0.0009 & -0.0003 & 0 \\ 0.0022 & -0.0957 & 0 \\ 0.0002 & -0.0069 & 0 \\ -0.0004 & -0.0168 & 0 \\ 0.0022 & -0.0957 & 0 \\ -0.1173 & 2.3364 & -0.0894 \\ 0.0006 & 0.0153 & 0.0007 \\ 0.0008 & 0.0374 & 0.0018 \\ 0.0002 & -0.0069 & 0 \\ 0.0006 & 0.0153 & 0.0007 \\ 0.0000 & 0.0011 & 0.0001 \\ 0.0001 & 0.0027 & 0.0001 \\ -0.0004 & -0.0168 & 0 \\ 0.0008 & 0.0374 & 0.0018 \\ 0.0001 & 0.0027 & 0.0001 \\ -0.0495 & 0.0557 & 0.0003 \end{bmatrix}^{\intercal} \left[\begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix} \right]$$

if $s_t = 1$, and

$$\begin{bmatrix} \hat{c}_t \\ \hat{k}_t \\ \hat{z}_t \end{bmatrix} = \begin{bmatrix} 0.0405 & 0.0 & 0.0268 & -0.0968 \\ 0.9692 & 0.0 & -0.4649 & 0.9227 \\ 0.0 & 0.0 & 0.0217 & -0.0410 \end{bmatrix} \begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -0.0009 & -0.0003 & 0 \\ 0 & 0 & 0 \\ 0.0005 & -0.0208 & 0 \\ -0.0021 & 0.0405 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.0005 & -0.0208 & 0 \\ 0 & 0 & 0 \\ 0.0004 & 0.0100 & 0.0005 \\ -0.0012 & -0.0193 & -0.0009 \\ -0.0021 & 0.0405 & 0 \\ 0 & 0 & 0 \\ -0.0012 & -0.0193 & -0.0009 \\ -0.0467 & 0.0869 & 0.0017 \end{bmatrix}^{\intercal} \left[\begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix} \right]$$

if $s_t = 2$. For the dynamics of $\hat{c}_t$ and $\hat{k}_t$, one can see that the coefficients of $\hat{z}_{t-1}$ and ε_t are considerably different across regimes. The large difference across regimes also shows up in the coefficients of $\hat{k}_{t-1}\varepsilon_t$, $\hat{z}_{t-1}\varepsilon_t$, and ε_t^2. These differences are induced by the Markov-switching volatility parameter $\sigma(s_t)$, which has both first-order and second-order effects on the dynamics of $\hat{c}_t$ and $\hat{k}_t$.[6]

6.2 *Solution from the naive perturbation method*

The naive perturbation method, according to Corollary 1, does not have the time-varying effects as discussed in the previous section. In particular, it can be seen from the following second-order solution that the coefficients of $\hat{z}_{t-1}$, ε_t, $\hat{k}_{t-1}\hat{z}_t$, $\hat{k}_{t-1}\varepsilon_t$, $\hat{z}_{t-1}^2$, $\hat{z}_{t-1}\varepsilon_t$, and ε_t^2 are all identical across regimes:

$$\begin{bmatrix} \hat{c}_t \\ \hat{k}_t \\ \hat{z}_t \end{bmatrix} = \begin{bmatrix} 0.0406 & 0.0836 & 0.0152 & 0.0314 \\ 0.9692 & -1.4264 & -0.2586 & -0.4169 \\ 0 & 0.0667 & 0.0121 & 0.0191 \end{bmatrix} \begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix}$$

[6]The time-varying coefficients of the cross terms $\hat{k}_{t-1}$, $\hat{z}_{t-1}$, and $\hat{z}_{t-1}^2$ are related to the Markov-switching persistence parameter $\rho(s_t)$.

$$+\frac{1}{2}\begin{bmatrix} -0.0009 & -0.0003 & 0 \\ 0.0014 & -0.0638 & 0 \\ 0.0003 & -0.0116 & 0 \\ 0.0006 & -0.0185 & 0 \\ 0.0014 & -0.0638 & 0 \\ -0.0794 & 1.5100 & -0.0618 \\ 0.0007 & 0.0170 & 0.0008 \\ 0.0438 & -0.6868 & 0.0346 \\ 0.0003 & -0.0116 & 0 \\ 0.0007 & 0.0170 & 0.0008 \\ 0.0001 & 0.0031 & 0.0001 \\ -0.0057 & 0.1082 & -0.0046 \\ 0.0006 & -0.0185 & 0 \\ 0.0438 & -0.6868 & 0.0346 \\ -0.0057 & 0.1082 & -0.0046 \\ -0.1080 & 0.1431 & -0.0010 \end{bmatrix}^{\intercal} \left[\begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix} \right]$$

if $s_t = 1$, and

$$\begin{bmatrix} \hat{c}_t \\ \hat{k}_t \\ \hat{z}_t \end{bmatrix} = \begin{bmatrix} 0.0406 & 0.0836 & 0.0152 & -0.0628 \\ 0.9692 & -1.4264 & -0.2586 & 0.8339 \\ 0 & 0.0667 & 0.0121 & -0.0383 \end{bmatrix} \begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix}$$

$$+\frac{1}{2}\begin{bmatrix} -0.0009 & -0.0003 & 0 \\ 0.0014 & -0.0638 & 0 \\ 0.0003 & -0.0116 & 0 \\ -0.0011 & 0.0369 & 0 \\ 0.0014 & -0.0638 & 0 \\ -0.0794 & 1.5100 & -0.0618 \\ 0.0007 & 0.0170 & 0.0008 \\ -0.0876 & 1.3735 & -0.0692 \\ 0.0003 & -0.0116 & 0 \\ 0.0007 & 0.0170 & 0.0008 \\ 0.0001 & 0.0031 & 0.0001 \\ 0.0114 & -0.2164 & 0.0092 \\ -0.0011 & 0.0369 & 0 \\ -0.0876 & 1.3735 & -0.0692 \\ 0.0114 & -0.2164 & 0.0092 \\ -0.1124 & 0.2550 & -0.0040 \end{bmatrix}^{\intercal} \left[\begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \hat{k}_{t-1} \\ \hat{z}_{t-1} \\ \varepsilon_t \\ 1 \end{bmatrix} \right]$$

if $s_t = 2$.

The only Markov-switching effect is through the coefficient of the perturbation parameter χ. Moreover, the computed coefficients are very different, implying different magnitudes and shapes of impulse responses. For example, in the regime $s_t = 2$, the coefficients of $\tilde{z}_{t-1}$ for all the three equations are 0 for the partition solution, but the same

coefficients can be as large as -1.4264 or 1.3735 for the naive solution, where -1.4264 is the first-order coefficient of $\tilde{z}_{t-1}$ and 1.3735 is the second-order coefficient. As a result, the naive perturbation method produces less accurate approximations as shown in the following section—a result that confirms what we find in Section 5.

6.3 *Assessing approximation errors*

Using the parameterization in Table 2, we compare the accuracy of approximated solutions from the two perturbation methods. Our results confirm that the partition perturbation method is more accurate than the naive perturbation method, especially for first-order and second-order approximations.

As a basis for comparison, we solve the nonlinear model using value function iterations (Uhlig (1999)). To accomplish this task we formulate the value function problem for the Markov-switching stationary RBC model as

$$V(\tilde{k},\tilde{z},\varepsilon,s)=\max_{\tilde{c},\tilde{k}}\left\{\frac{\tilde{c}^{v}}{v}+\beta\tilde{z}^{v}\mathbb{E}V(\tilde{k}',\tilde{z}',\varepsilon',s')\right\}$$

subject to

$$\tilde{c}+\tilde{z}\tilde{k}'=\tilde{z}^{1-\alpha}\tilde{k}^{\alpha}+(1-\delta)\tilde{k}\quad\text{and}\quad\log\tilde{z}'=\big(1-\rho(s)\big)\mu(s)+\rho(s)\log\tilde{z}+\sigma(s)\varepsilon.$$

Following Aruoba, Fernández-Villaverde, and Rubio-Ramírez (2006), we solve the problem on a grid of 25,600 points for $\tilde{k}$, 51 points for $\tilde{z}$, and 51 points for ε. We use Tauchen's (1986) method to discretize the stochastic process and smooth the policy functions using the shape-preserving splines described in Judd and Solnick (1994). Since we need to find two value functions (one for each regime), the computation is very expensive. To solve the above value function problem within a reasonable amount of time, we rely on the CUDA (compute unified device architecture) of NVIDIA to build algorithms that utilize graphics processing units (GPUs). This approach leads to a remarkable improvement in computing time. Aldrich, Fernández-Villaverde, Gallant, and Rubio-Ramírez (2011) document that utilization of the GPU delivers a speed improvement of about 200 times.

Let $\mathbf{g}_{s_t}^{\text{order}}$ and $\mathbf{h}_{s_t}^{\text{order}}$ denote the solution from the Taylor series expansion of a particular order of interest. For our Markov-switching RBC model, the dimension of $\mathbf{g}_{s_t}^{\text{order}}$ is just one (i.e., $n_y=1$) and we consider approximations up to the first three orders. Let $\mathbf{h}_{s_t}^{k,\text{order}}$ be the kth function of $\mathbf{h}_{s_t}^{\text{order}}$ (there are two functions because $n_x=2$). The EE evaluated at $\tilde{k}_{t-1}$, $\tilde{z}_{t-1}$, ε_t, and s_t is

$$\begin{aligned}
&\text{EE}^{\text{order}}(\tilde{k}_{t-1},\tilde{z}_{t-1},\varepsilon_t,s_t)\\
&\quad=1-\beta\sum_{s_{t+1}=1}^{2}p_{s_t,s_{t+1}}\int_{\mathbb{R}}\Bigg(\frac{\mathbf{g}_{s_{t+1}}^{\text{order}}\big(\mathbf{h}_{s_t}^{1,\text{order}}(\tilde{k}_{t-1},\tilde{z}_{t-1},\varepsilon_t,1),\tilde{z}_t,\varepsilon_{t+1},1\big)^{v-1}}{\mathbf{g}_{s_t}^{\text{order}}(\tilde{k}_{t-1},\tilde{z}_{t-1},\varepsilon_t,1)^{v-1}}\\
&\qquad\times\Big[\alpha\exp\big\{(1-\alpha)\mathbf{h}_{s_t}^{2,\text{order}}\big(\mathbf{h}_{s_t}^{1,\text{order}}(\tilde{k}_{t-1},\tilde{z}_{t-1},\varepsilon_t,1),\tilde{z}_t,\varepsilon_{t+1},1\big)\big\}\\
&\qquad\times\mathbf{h}_{s_t}^{1,\text{order}}(\tilde{k}_{t-1},\tilde{z}_{t-1},\varepsilon_t,1)^{\alpha-1}+1-\delta\Big]\Bigg)\mu(\varepsilon_{t+1})\,d\varepsilon_{t+1},
\end{aligned}$$

where μ denotes the unconditional probability density. The associated absolute value of the unconditional EE is

$$\mathrm{EEE}^{\mathrm{order}} = \sum_{s_t=1}^{2}\Bigg[\int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}}\left|\mathrm{EE}^{\mathrm{order}}(\tilde{k}_{t-1},\tilde{z}_{t-1},\varepsilon_t,1,s_t)\right| \times \mu(\tilde{k}_{t-1},\tilde{z}_{t-1},\varepsilon_t)\,d\tilde{k}_{t-1}\,d\tilde{z}_{t-1}d\varepsilon_t\Bigg]\bar{p}(s_t),$$

where $\bar{p}(s_t)$ is the ergodic probability of s_t. Again, $\mu(\tilde{k}_{t-1},\tilde{z}_{t-1},\varepsilon_t)$ denotes the unconditional probability density function of $\tilde{k}_{t-1}$, $\tilde{z}_{t-1}$, and ε_t.

We use the following procedure to approximate $\mathrm{EEE}^{\mathrm{order}}$ for order $\in$ {first, second, third}. We begin by simulating ε_t from the standard normal distribution and s_t from the ergodic distribution. Conditioning on each simulated set $\{\varepsilon_t, s_t\}$, we use $\mathbf{h}^{\mathrm{order}}_{s_t}$ to simulate $\tilde{k}_t$ and $\tilde{z}_t$. The length of the simulated path is 10,000 periods, with the first 1000 periods discarded as a burn-in. The remaining 9000 simulations are used to form the unconditional distribution of the variables $\tilde{k}_{t-1}$ and $\tilde{z}_{t-1}$. This procedure is justified by Santos and Peralta-Alva (2005).

For each set of $\tilde{k}_{t-1}$, $\tilde{z}_{t-1}$, ε_t, and s_t randomly selected from these 9000 simulations, we draw 10,000 values of ε_{t+1} from the standard normal distribution and 10,000 values of s_{t+1} from the transition probability $p_{s_t,s_{t+1}}$ to compute the expectation that depends on the functions $\mathbf{g}^{\mathrm{order}}_{s_{t+1}}$, $\mathbf{g}^{\mathrm{order}}_{s_t}$, and $\mathbf{h}^{\mathrm{order}}_{s_t}$. The result is 9000 values of $\mathrm{EE}^{\mathrm{order}}(\tilde{k}_{t-1},\tilde{z}_{t-1},\varepsilon_t,s_t)$. We average across these 9000 values to compute $\mathrm{EEE}^{\mathrm{order}}$. We repeat this procedure for each order $\in$ {first, second, third}, and for both the partition and naive perturbation methods. When simulating a path for second-order and third-order approximations, we use the pruning technique described in Andreasen, Fernández-Villaverde, and Rubio-Ramírez (2013). We repeat the same procedure for the value function iteration approach except there is no need for pruning.

Table 3 reports the base-10 log absolute values of EEs for each solution method.[7] Although the value function iteration method is most accurate as expected, the partition perturbation method fares remarkably well in comparison. This is an important result because, even with the advanced CUDA technology, value function iterations take about 15 minutes to find an approximation to the model solution (with the steady state as an initial starting point), while either perturbation method takes only a fraction of a second to find a third-order approximation.

For both perturbation methods, Table 3 indicates that higher-order approximations produce a higher degree of accuracy. In all cases, increasing the approximation from first order to second order delivers significant gain without taking much more computational time. The accuracy gain is much smaller when the approximation moves from second order to third order. More important is the result that the partition perturbation method is more accurate than the naive perturbation method for any order of approximation. As argued in Section 3 and illuminated in Section 5, the partition perturbation method

[7] As a reference, the base-10 log value has this interpretation: the value -4 implies an error of \$1 for each \$10,000 of consumption.

TABLE 3. Euler-equation errors (base-10 log absolute value).

Value function iteration	−4.54
Partition perturbation	
First order	−3.01
Second order	−3.59
Third order	−3.73
Naive perturbation	
First order	−2.48
Second order	−3.07
Third order	−3.16

does not take approximation along the direction of $\boldsymbol{\theta}_2(s_t)$ and thus preserves the time-varying nature of these parameters even for the first-order approximation.

Indeed, the accuracy of the first-order approximation from the partition perturbation method is almost as good as the accuracy of the second-order approximation from the naive perturbation method. For likelihood-based estimation of an MSDSGE model, a lower-order approximation with the same degree of accuracy as a higher-order is always preferred because the cost of the programming and computing time increases nonlinearly with the order of approximation. As the model becomes larger and the approximation order becomes higher, the estimation time can quickly become explosive. From both theoretical and practical points of view, therefore, the partition perturbation method is superior to the naive perturbation method.

7. CONCLUSION

Markov switching has been introduced as an essential ingredient to a large class of models usable for analyzing structural breaks in the economy and regime shifts in policy, ranging from backward-looking models (Hamilton (1989) and Sims and Zha (2006)) to forward-looking rational expectations models (Clarida, Gali, and Gertler (2000), Lubik and Schorfheide (2004), Davig and Leeper (2007), Farmer, Waggoner, and Zha (2009)). This paper expands the literature by developing a general methodology for constructing high-order approximations to the solutions of MSDSGE models. Higher-order approximations enable researchers to study many economic problems, such as how important is uncertainty in both the private sector and government policies for shaping the business cycle.

While the key developments have been extensively discussed in the Introduction, we emphasize that the contribution of this paper is not only theoretically substantive but also practically important. We show through a Markov-switching RBC model that the implementation of the partition perturbation method is not burdensome but rather straightforward, once one knows how to solve a system of quadratic polynomial equations efficiently. It is our hope that the advance made in this paper enables applied researchers to estimate MSDSGE models by focusing on particular economic problems.

Appendix A: Proofs of Propositions 1, 2, and 3

Before presenting the proofs of Propositions 1, 2, and 3, we briefly review the two forms of the chain rule in our notation and clarify the notation for the arguments of the function $\mathbf{f}$. If $\mathbf{w}:\mathbb{R}^{n_u}\to\mathbb{R}^{n_w}$, $\mathbf{u}:\mathbb{R}^{n_v}\to\mathbb{R}^{n_u}$, and $\mathbf{v}\in\mathbb{R}^{n_v}$, the chain rule for total derivatives is $D\mathbf{w}\circ\mathbf{u}(\mathbf{v})=D\mathbf{w}(\mathbf{u}(\mathbf{v}))D\mathbf{u}(\mathbf{v})$. This will be the form used for the first-order expansion. For second- and higher-order expansions, we need the form

$$D_\ell\mathbf{w}\circ\mathbf{u}(\mathbf{v})=\sum_{m=1}^{n_u}D_m\mathbf{w}\big(\mathbf{u}(\mathbf{v})\big)D_\ell\mathbf{u}^m(\mathbf{v})$$

for $1\le\ell\le n_v$. We will write the function $\mathbf{f}$ as $\mathbf{f}(\mathbf{y}_{t+1},\mathbf{y}_t,\mathbf{x}_t,\mathbf{x}_{t-1},\tilde{\boldsymbol{\varepsilon}}_{t+1},\boldsymbol{\varepsilon}_t,\boldsymbol{\theta}_{t+1},\boldsymbol{\theta}_t)$. This will prevent confusion when making the substitutions $\tilde{\boldsymbol{\varepsilon}}_{t+1}=\chi\boldsymbol{\varepsilon}_{t+1}$, $\boldsymbol{\theta}_{t+1}=\boldsymbol{\theta}(s_{t+1},\chi)$, and $\boldsymbol{\theta}_t=\boldsymbol{\theta}(s_t,\chi)$.

A.1 *Proof of Proposition* 1

Define

$$\mathbf{v}_i(\mathbf{z}_t)=\begin{bmatrix}\big(\mathbf{e}_i^\intercal\otimes\mathbf{I}_{n_x}\big)\mathbf{H}(\mathbf{z}_t)\\ \chi\boldsymbol{\varepsilon}_{t+1}\\ \chi\end{bmatrix}\quad\text{and}\quad\mathbf{u}_{i,j}(\mathbf{z}_t)=\begin{bmatrix}\big(\mathbf{e}_j^\intercal\otimes\mathbf{I}_{n_y}\big)\mathbf{G}\big(\mathbf{v}_i(\mathbf{z}_t)\big)\\ \big(\mathbf{e}_i^\intercal\otimes\mathbf{I}_{n_y}\big)\mathbf{G}(\mathbf{z}_t)\\ \big(\mathbf{e}_i^\intercal\otimes\mathbf{I}_{n_x}\big)\mathbf{H}(\mathbf{z}_t)\\ \mathbf{x}_{t-1}\\ \chi\boldsymbol{\varepsilon}_{t+1}\\ \boldsymbol{\varepsilon}_t\\ \boldsymbol{\theta}(j,\chi)\\ \boldsymbol{\theta}(i,\chi)\end{bmatrix}.$$

With this notation,

$$\mathbf{0}_{(n_y+n_x)}=\mathbb{F}_i(\mathbf{z}_t)=\sum_{j=1}^{n_s}p_{i,j}\int_{\mathbb{R}^{n_\varepsilon}}\mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_t)\big)\,d\mu(\boldsymbol{\varepsilon}_{t+1})$$

for $1\le i\le n_s$. Thus,

$$\mathbf{0}_{(n_y+n_x)\times n_z}=D\mathbb{F}_i(\mathbf{z}_t)=\sum_{j=1}^{n_s}p_{i,j}\int_{\mathbb{R}^{n_\varepsilon}}D\mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_t)\big)D\mathbf{u}_{i,j}(\mathbf{z}_t)\,d\mu(\boldsymbol{\varepsilon}_{t+1})\tag{A.1}$$

for $1\le i\le n_s$. The $n_f\times n_z$ matrix $D\mathbf{u}_{i,j}(\mathbf{z}_t)$ can computed implicitly as

$$D\mathbf{u}_{i,j}(\mathbf{z}_t)=\begin{bmatrix}\big(\mathbf{e}_j^\intercal\otimes\mathbf{I}_{n_y}\big)D\mathbf{G}\big(\mathbf{v}_i(\mathbf{z}_t)\big)D\mathbf{v}_i(\mathbf{z}_t)\\ \big(\mathbf{e}_i^\intercal\otimes\mathbf{I}_{n_y}\big)D\mathbf{G}(\mathbf{z}_t)\\ \big(\mathbf{e}_i^\intercal\otimes\mathbf{I}_{n_x}\big)D\mathbf{H}(\mathbf{z}_t)\\ \begin{bmatrix}\mathbf{I}_{n_x}&\mathbf{0}_{n_x\times n_\varepsilon}&\mathbf{0}_{n_x\times 1}\\ \mathbf{0}_{n_\varepsilon\times n_x}&\mathbf{0}_{n_\varepsilon\times n_\varepsilon}&\boldsymbol{\varepsilon}_{t+1}\\ \mathbf{0}_{n_\varepsilon\times n_x}&\mathbf{I}_{n_\varepsilon}&\mathbf{0}_{n_\varepsilon\times 1}\\ \mathbf{0}_{n_\theta\times n_x}&\mathbf{0}_{n_\theta\times n_\varepsilon}&\boldsymbol{\theta}(j,1)-\boldsymbol{\theta}(j,0)\\ \mathbf{0}_{n_\theta\times n_x}&\mathbf{0}_{n_\theta\times n_\varepsilon}&\boldsymbol{\theta}(i,1)-\boldsymbol{\theta}(i,0)\end{bmatrix}\end{bmatrix},\tag{A.2}$$

where

$$D\mathbf{v}_i(\mathbf{z}_t) = \begin{bmatrix} \left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\right) D\mathbf{H}(\mathbf{z}_t) \\ \begin{bmatrix} \mathbf{0}_{n_\varepsilon \times (n_x+n_\varepsilon)} & \boldsymbol{\varepsilon}_{t+1} \\ \mathbf{0}_{1\times(n_x+n_\varepsilon)} & 1 \end{bmatrix} \end{bmatrix}. \tag{A.3}$$

Substituting (A.2) and (A.3) into equation (A.1), evaluating at $\mathbf{z}_{ss}$, and integrating, one obtains

$$\begin{aligned}
&\mathbf{0}_{(n_y+n_x)\times n_z} \\
&\quad = D\mathbb{F}_i(\mathbf{z}_{ss}) \\
&\quad = \sum_{j=1}^{n_s} p_{i,j} D\mathbf{f}\left(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\right) \\
&\qquad \times \begin{bmatrix} \left(\mathbf{e}_j^\intercal \otimes \mathbf{I}_{n_y}\right)\left(D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})\left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\right) D\mathbf{H}(\mathbf{z}_{ss}) + \left[\mathbf{0}_{n_s n_y \times (n_x+n_\varepsilon)} D_\chi \mathbf{G}(\mathbf{z}_{ss})\right]\right) \\ \left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_y}\right) D\mathbf{G}(\mathbf{z}_{ss}) \\ \left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\right) D\mathbf{H}(\mathbf{z}_{ss}) \\ \begin{bmatrix} \mathbf{I}_{n_x} & \mathbf{0}_{n_x\times n_\varepsilon} & \mathbf{0}_{n_x\times 1} \\ \mathbf{0}_{n_\varepsilon\times n_x} & \mathbf{0}_{n_\varepsilon\times n_\varepsilon} & \mathbf{0}_{n_\varepsilon\times 1} \\ \mathbf{0}_{n_\varepsilon\times n_x} & \mathbf{I}_{n_\varepsilon} & \mathbf{0}_{n_\varepsilon\times 1} \\ \mathbf{0}_{n_\theta\times n_x} & \mathbf{0}_{n_\theta\times n_\varepsilon} & \boldsymbol{\theta}(j,1) - \boldsymbol{\theta}(j,0) \\ \mathbf{0}_{n_\theta\times n_x} & \mathbf{0}_{n_\theta\times n_\varepsilon} & \boldsymbol{\theta}(i,1) - \boldsymbol{\theta}(i,0) \end{bmatrix} \end{bmatrix}.
\end{aligned}$$

Here we have used the fact that $\int_{\mathbb{R}^{n_\varepsilon}} \boldsymbol{\varepsilon}_{t+1}\, d\mu(\boldsymbol{\varepsilon}_{t+1}) = \mathbb{E}_t \boldsymbol{\varepsilon}_{t+1} = \mathbf{0}_{n_\varepsilon}$. Since there is an explicit expression for $\mathbf{f}$ and $\mathbf{u}_{i,j}(\mathbf{z}_{ss})$, the $(n_y + n_x) \times n_f$ matrix $D\mathbf{f}(\mathbf{u}_{i,j}(\mathbf{z}_{ss}))$ also has an explicit representation. The above system can be written as

$$\begin{aligned}
\mathbf{0}_{(n_y+n_x)\times n_x} = \sum_{j=1}^{n_s} p_{i,j} \big\{ & D_{\mathbf{x}_{t-1}} \mathbf{f}\left(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\right) \\
& + D_{\mathbf{y}_{t+1}} \mathbf{f}\left(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\right)\left(\mathbf{e}_j^\intercal \otimes \mathbf{I}_{n_y}\right) D_{\mathbf{x}_{t-1}} \mathbf{G}(\mathbf{z}_{ss})\left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\right) D_{\mathbf{x}_{t-1}} \mathbf{H}(\mathbf{z}_{ss}) \\
& + D_{\mathbf{y}_t} \mathbf{f}\left(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\right)\left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_y}\right) D_{\mathbf{x}_{t-1}} \mathbf{G}(\mathbf{z}_{ss}) \\
& + D_{\mathbf{x}_t} \mathbf{f}\left(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\right)\left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\right) D_{\mathbf{x}_{t-1}} \mathbf{H}(\mathbf{z}_{ss}) \big\},
\end{aligned} \tag{A.4}$$

$$\begin{aligned}
\mathbf{0}_{(n_y+n_x)\times n_\varepsilon} = \sum_{j=1}^{n_s} p_{i,j} \big\{ & D_{\boldsymbol{\varepsilon}_t} \mathbf{f}\left(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\right) \\
& + D_{\mathbf{y}_{t+1}} \mathbf{f}\left(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\right)\left(\mathbf{e}_j^\intercal \otimes \mathbf{I}_{n_y}\right) D_{\mathbf{x}_{t-1}} \mathbf{G}(\mathbf{z}_{ss})\left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\right) D_{\boldsymbol{\varepsilon}_t} \mathbf{H}(\mathbf{z}_{ss}) \\
& + D_{\mathbf{y}_t} \mathbf{f}\left(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\right)\left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_y}\right) D_{\boldsymbol{\varepsilon}_t} \mathbf{G}(\mathbf{z}_{ss}) \\
& + D_{\mathbf{x}_t} \mathbf{f}\left(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\right)\left(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\right) D_{\boldsymbol{\varepsilon}_t} \mathbf{H}(\mathbf{z}_{ss}) \big\},
\end{aligned} \tag{A.5}$$

$$
\begin{aligned}
\mathbf{0}_{(n_y+n_x)\times 1} = \sum_{j=1}^{n_s} p_{i,j} \big\{ & D_{\boldsymbol{\theta}_{t+1}} \mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\big)\big(\boldsymbol{\theta}(j,1)-\boldsymbol{\theta}(j,0)\big) \\
& + D_{\boldsymbol{\theta}_t} \mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\big)\big(\boldsymbol{\theta}(i,1)-\boldsymbol{\theta}(i,0)\big) \\
& + D_{\mathbf{y}_{t+1}} \mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\big)\big(\mathbf{e}_j^\intercal \otimes \mathbf{I}_{n_y}\big)\big(D_{\mathbf{x}_{t-1}} \mathbf{G}(\mathbf{z}_{ss})\big(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\big) D_\chi \mathbf{H}(\mathbf{z}_{ss}) \\
& + D_\chi \mathbf{G}(\mathbf{z}_{ss})\big) \\
& + D_{\mathbf{y}_t} \mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\big)\big(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_y}\big) D_\chi \mathbf{G}(\mathbf{z}_{ss}) \\
& + D_{\mathbf{x}_t} \mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_{ss})\big)\big(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\big) D_\chi \mathbf{H}(\mathbf{z}_{ss})\big\}
\end{aligned} \tag{A.6}
$$

for $1 \le i \le n_s$. From this representation, it is easy to see that equation (A.4) represents a system of $n_s(n_y+n_x)n_x$ quadratic equations in the $n_s(n_y+n_x)n_x$ unknowns $D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})$ and $D_{\mathbf{x}_{t-1}}\mathbf{H}(\mathbf{z}_{ss})$. For each solution of the quadratic system (A.4), equation (A.5) represents a linear system in the unknowns $D_{\boldsymbol{\varepsilon}_t}\mathbf{G}(\mathbf{z}_{ss})$ and $D_{\boldsymbol{\varepsilon}_t}\mathbf{H}(\mathbf{z}_{ss})$, and equation (A.6) represents a linear system in the unknowns $D_\chi \mathbf{G}(\mathbf{z}_{ss})$ and $D_\chi \mathbf{H}(\mathbf{z}_{ss})$. This completes the proof of Proposition 1.

A.2 *Proof of Proposition* 2

The $n_s(n_y+n_x)n_z^2$ unknowns $D_{\ell_2}D_{\ell_1}\mathbf{G}(\mathbf{z}_{ss})$ and $D_{\ell_2}D_{\ell_1}\mathbf{H}(\mathbf{z}_{ss})$ can be found by solving the system of equations

$$D_{\ell_2}D_{\ell_2}\mathbb{F}_i(\mathbf{z}_{ss}) = 0$$

for $1 \le i \le n_s$ and $1 \le \ell_1, \ell_2 \le n_z$. Since

$$D_{\ell_1}\mathbb{F}_i(\mathbf{z}_t) = \sum_{j=1}^{n_s} p_{i,j} \int_{\mathbb{R}^{n_\varepsilon}} \sum_{m_1=1}^{n_f} D_{m_1}\mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_t)\big) D_{\ell_1}\mathbf{u}_{i,j}^{m_1}(\mathbf{z}_t)\, d\mu(\boldsymbol{\varepsilon}_{t+1}),$$

we obtain

$$
\begin{aligned}
& D_{\ell_2}D_{\ell_1}\mathbb{F}_i(\mathbf{z}_t) \\
& \quad = \sum_{i=1}^{n_s} p_{i,j} \int_{\mathbb{R}^{n_\varepsilon}} \sum_{m_1=1}^{n_f} D_{m_1}\mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_t)\big) D_{\ell_2}D_{\ell_1}\mathbf{u}_{i,j}^{m_1}(\mathbf{z}_t)\, d\mu(\boldsymbol{\varepsilon}_{t+1}) \\
& \qquad + \sum_{i=1}^{n_s} p_{i,j} \int_{\mathbb{R}^{n_\varepsilon}} \sum_{m_2=1}^{n_f} \sum_{m_1=1}^{n_f} D_{m_2}D_{m_1}\mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_t)\big) D_{\ell_2}\mathbf{u}_{i,j}^{m_2}(\mathbf{z}_t) D_{\ell_1}\mathbf{u}_{i,j}^{m_1}(\mathbf{z}_t)\, d\mu(\boldsymbol{\varepsilon}_{t+1}).
\end{aligned} \tag{A.7}
$$

Each of the terms in the second summation in equation (A.7) can either be explicitly computed or is known from the first-order expansion. All that remains is to compute the term $D_{\ell_2}D_{\ell_1}\mathbf{u}_{i,j}^{m_1}(\mathbf{z}_t)$, which is the m_1th component of

$$D_{\ell_2}D_{\ell_1}\mathbf{u}_{i,j}(\mathbf{z}_t) = \begin{bmatrix} \big(\mathbf{e}_j^\intercal \otimes \mathbf{I}_{n_y}\big)D_{\ell_2}D_{\ell_1}\mathbf{G} \circ \mathbf{v}_i(\mathbf{z}_t) \\ \big(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_y}\big)D_{\ell_2}D_{\ell_1}\mathbf{G}(\mathbf{z}_t) \\ \big(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\big)D_{\ell_2}D_{\ell_1}\mathbf{H}(\mathbf{z}_t) \\ \mathbf{0}_{n_x+2n_\varepsilon+2n_\theta} \end{bmatrix}. \tag{A.8}$$

The term $D_{\ell_2}D_{\ell_1}\mathbf{G} \circ \mathbf{v}_i(\mathbf{z}_t)$ is equal to

$$\begin{aligned} &\big(\mathbf{e}_j^\intercal \otimes \mathbf{I}_{n_y}\big)\Bigg(\sum_{k_1=1}^{n_z} D_{k_1}\mathbf{G}(\mathbf{z}_t)D_{\ell_2}D_{\ell_1}\mathbf{v}_i^{k_1}(\mathbf{z}_t) \\ &\quad + \sum_{k_2=1}^{n_z}\sum_{k_1=1}^{n_z} D_{k_2}D_{k_1}\mathbf{G}(\mathbf{z}_t)D_{\ell_2}\mathbf{v}_i^{k_2}(\mathbf{z}_t)D_{\ell_1}\mathbf{v}_i^{k_1}(\mathbf{z}_t)\Bigg), \end{aligned}$$

where

$$D_{\ell_2}D_{\ell_1}\mathbf{v}_i(\mathbf{z}_t) = \begin{bmatrix} \big(\mathbf{e}_i^\intercal \otimes \mathbf{I}_{n_x}\big)D_{\ell_2}D_{\ell_1}\mathbf{H}(\mathbf{z}_t) \\ \mathbf{0}_{n_\varepsilon+1} \end{bmatrix}.$$

Substituting this into equation (A.7) and evaluating at $\mathbf{z}_{ss}$, it is easy to see that this will be linear in the unknowns $D_{\ell_2}D_{\ell_1}\mathbf{G}(\mathbf{z}_{ss})$ and $D_{\ell_2}D_{\ell_1}\mathbf{H}(\mathbf{z}_{ss})$. This completes the proof of Proposition 2.

A.3 *Proof of Proposition* 3

Proposition 3 follows directly from the more general version given below. While there is no constant term in equation (A.9), this case can easily be handled by appending a variable $\tilde{x}_t$ to the vector of predetermined variables $\mathbf{x}_t$ and adding an equation of the form $\tilde{x}_t - \tilde{x}_{t-1} = 0$. While this introduces an additional unit root into the system, this will not pose any problems for the solutions techniques discussed in this paper.

PROPOSITION 4. *With the partition perturbation method, the first-order solution of*

$$\begin{aligned} E_t\big[&\mathbf{A}_1\big(\boldsymbol{\theta}(s_t), \boldsymbol{\theta}(s_{t+1})\big)y_{t+1} + \mathbf{A}_2\big(\boldsymbol{\theta}(s_t), \boldsymbol{\theta}(s_{t+1})\big)y_t + \mathbf{A}_3\big(\boldsymbol{\theta}(s_t), \boldsymbol{\theta}(s_{t+1})\big)x_t \\ &+ \mathbf{A}_4\big(\boldsymbol{\theta}(s_t), \boldsymbol{\theta}(s_{t+1})\big)x_{t-1} + \mathbf{A}_5\big(\boldsymbol{\theta}(s_t), \boldsymbol{\theta}(s_{t+1})\big)\varepsilon_{t+1} + \mathbf{A}_6\big(\boldsymbol{\theta}(s_t), \boldsymbol{\theta}(s_{t+1})\big)\varepsilon_t\big] = 0 \end{aligned} \tag{A.9}$$

is exact and all higher-order terms are zero, where $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *are* $(n_y + n_x) \times n_y$, $\mathbf{A}_3$ *and* $\mathbf{A}_4$ *are* $(n_y + n_x) \times n_x$, *and* $\mathbf{A}_5$ *and* $\mathbf{A}_6$ *are* $(n_y + n_x) \times n_\varepsilon$.

PROOF. It is easy to see that the steady-state is $\mathbf{y}_{ss} = \mathbf{0}_{n_y}$ and $\mathbf{x}_{ss} = \mathbf{0}_{n_x}$, which is independent of all the parameters. This implies that none of the parameters needs to be

perturbed and the perturbation function is $\boldsymbol{\theta}(k,\chi)=\boldsymbol{\theta}(k)$. We first show that the first-order Taylor expansion of $\mathbf{G}$ and $\mathbf{H}$ exactly solves equation (A.9) and then show that all terms of order 2 or greater in the full Taylor series expansion of $\mathbf{G}$ and $\mathbf{H}$ are zero.

The first-order Taylor expansion, evaluated at $\chi=1$, is

$$\begin{aligned}
y_t &= (\mathbf{e}_{s_t}\otimes\mathbf{I}_{n_y})\big(D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})\mathbf{x}_{t-1}+D_{\boldsymbol{\varepsilon}_t}\mathbf{G}(\mathbf{z}_{ss})\boldsymbol{\varepsilon}_t+D_{\chi}\mathbf{G}(\mathbf{z}_{ss})\big),\\
x_t &= (\mathbf{e}_{s_t}\otimes\mathbf{I}_{n_x})\big(D_{\mathbf{x}_{t-1}}\mathbf{H}(\mathbf{z}_{ss})\mathbf{x}_{t-1}+D_{\boldsymbol{\varepsilon}_t}\mathbf{H}(\mathbf{z}_{ss})\boldsymbol{\varepsilon}_t+D_{\chi}\mathbf{H}(\mathbf{z}_{ss})\big).
\end{aligned}$$

Substituting this into the left hand side of equation (A.9), taking expectations, and gathering like terms, we obtain

$$\begin{aligned}
&\sum_{j=1}^{n_s} p_{i,j}\big\{\mathbf{A}_4(i,j)+\mathbf{A}_1(i,j)\big(\mathbf{e}_j^{\intercal}\otimes\mathbf{I}_{n_y}\big)D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})\big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_x}\big)D_{\mathbf{x}_{t-1}}\mathbf{H}(\mathbf{z}_{ss})\\
&\quad+\mathbf{A}_2(i,j)\big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_y}\big)D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})+\mathbf{A}_3(i,j)\big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_x}\big)D_{\mathbf{x}_{t-1}}\mathbf{H}(\mathbf{z}_{ss})\big\}\mathbf{x}_{t-1}\\
&\quad+\sum_{j=1}^{n_s} p_{i,j}\big\{\mathbf{A}_6(i,j)+\mathbf{A}_1(i,j)\big(\mathbf{e}_j^{\intercal}\otimes\mathbf{I}_{n_y}\big)D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})\big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_x}\big)D_{\boldsymbol{\varepsilon}_t}\mathbf{H}(\mathbf{z}_{ss})\\
&\quad+\mathbf{A}_2(i,j)\big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_y}\big)D_{\boldsymbol{\varepsilon}_t}\mathbf{G}(\mathbf{z}_{ss})+\mathbf{A}_3(i,j)\big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_x}\big)D_{\boldsymbol{\varepsilon}_t}\mathbf{H}(\mathbf{z}_{ss})\big\}\boldsymbol{\varepsilon}_t\\
&\quad+\sum_{j=1}^{n_s} p_{i,j}\big\{\mathbf{A}_1(i,j)\big(\mathbf{e}_j^{\intercal}\otimes\mathbf{I}_{n_y}\big)\big(D_{\chi}\mathbf{G}(\mathbf{z}_{ss})+D_{\mathbf{x}_{t-1}}\mathbf{G}(\mathbf{z}_{ss})\big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_x}\big)D_{\chi}\mathbf{H}(\mathbf{z}_{ss})\big)\\
&\quad+\mathbf{A}_2(i,j)\big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_y}\big)D_{\chi}\mathbf{G}(\mathbf{z}_{ss})+\mathbf{A}_3(i,j)\big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_x}\big)D_{\chi}\mathbf{H}(\mathbf{z}_{ss})\big\},
\end{aligned}$$

where $A_k(i,j)$ is shorthand notation for $A_k(\boldsymbol{\theta}(i),\boldsymbol{\theta}(j))$. Since equations (A.4)–(A.6) must hold, the above expression is equal to 0. Thus the first-order expansion is an exact solution of (A.9).

We now show that all the higher-order terms must be 0. Because none of the parameters is perturbed, one sees that the last $2n_\theta$ rows of the expression for $D\mathbf{u}_{i,j}(\mathbf{z}_t)$ given in equation (A.2) are 0. So, if $m>2(n_y+n_x+n_\varepsilon)$, then $D_\ell\mathbf{u}_{i,j}^m(\mathbf{z}_t)=0$ for $1\le\ell\le n_z$. It is also easy to see that $D_{m_2}D_{m_1}\mathbf{f}(\mathbf{u}_{i,j}(\mathbf{z}_t))=0$ if both m_1 and m_2 are less than or equal to $2(n_y+n_x+n_\varepsilon)$. Thus, an easy induction argument on q shows that

$$\begin{aligned}
&D_{\ell_q}\cdots D_{\ell_1}\mathbb{F}_i(\mathbf{z}_t)\\
&\quad=\sum_{j=1}^{n_s} p_{i,j}\int_{\mathbb{R}^{n_\varepsilon}}\sum_{m_1=1}^{2(n_y+n_x+n_\varepsilon)} D_{m_1}\mathbf{f}\big(\mathbf{u}_{i,j}(\mathbf{z}_t)\big)D_{\ell_q}\cdots D_{\ell_1}\mathbf{u}_{i,j}^{m_1}(\mathbf{z}_t)\,d\mu(\boldsymbol{\varepsilon}_{t+1}).
\end{aligned}$$

Finally, it follows from equation (A.8) that

$$D_{\ell_q}\cdots D_{\ell_1}\mathbf{u}_{i,j}(\mathbf{z}_t)=\begin{bmatrix}\big(\mathbf{e}_j^{\intercal}\otimes\mathbf{I}_{n_y}\big)D_{\ell_q}\cdots D_{\ell_1}\mathbf{G}\circ\mathbf{v}_i(\mathbf{z}_t)\\ \big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_y}\big)D_{\ell_q}\cdots D_{\ell_1}\mathbf{G}(\mathbf{z}_t)\\ \big(\mathbf{e}_i^{\intercal}\otimes\mathbf{I}_{n_x}\big)D_{\ell_q}\cdots D_{\ell_1}\mathbf{H}(\mathbf{z}_t)\\ \mathbf{0}_{n_x+2n_\varepsilon+2n_\theta}\end{bmatrix}$$

for $q > 1$. Since $D_{\ell_q} \cdots D_{\ell_1} \mathbf{G} \circ \mathbf{v}_i(\mathbf{z}_t)$ is *linear* in $D_{\ell_q} \cdots D_{\ell_1} \mathbf{G}(\mathbf{z}_t)$ and $D_{\ell_q} \cdots D_{\ell_1} \mathbf{H}(\mathbf{z}_t)$, it follows that $D_{\ell_q} \cdots D_{\ell_1} \mathbf{G}(\mathbf{z}_t) = 0$ and $D_{\ell_q} \cdots D_{\ell_1} \mathbf{H}(\mathbf{z}_t) = 0$ will be a solution of $D_{\ell_q} \cdots D_{\ell_1} \mathbb{F}_i(\mathbf{z}_t) = 0$. Thus all the terms of order 2 or greater in the Taylor series expansion of **G** and **H** are 0. This completes the proof of Proposition 4. □

Appendix B: Application of Gröbner bases

Although the theory of Gröbner bases is well known in the mathematics literature, the existing DSGE literature has not utilized this powerful application. We apply Gröbner bases to the two models studied in this paper; this application has indeed proven very powerful. For many MSDSGE models, Gröbner bases deliver a practical means to obtain all the solutions to a system of quadratic polynomial equations. For this reason, we provide below an intuitive explanation of how to apply Gröbner bases to solving a system of multivariate polynomials.

Suppose one wishes to find all the solutions to a system of n polynomial equations in n unknowns. There exist a number of routines that transform the original system of n polynomial equations to another system of n polynomial equations with the same set of solutions. This transformed system is known as a Gröbner basis. The following theorem, known as the shape lemma, shows that in most cases there is a Gröbner basis with a particularly powerful form. The shape lemma is known in the mathematics and computational science literature, but is still an unfamiliar object in the economics literature. We therefore restate this theorem in a form that is suitable to our problem.

The Shape Lemma. *There exists an open dense subset S of all systems of n polynomial equations in n unknowns such that for every system*

$$f_1(x_1, \ldots, x_n) = 0, \ldots, f_n(x_1, \ldots, x_n) = 0$$

in S, there exists a system of n polynomial equations in n unknowns with the same set of roots of the form

$$x_1 - q_1(x_n) = 0, \ldots, x_{n-1} - q_{n-1}(x_n) = 0, q_n(x_n) = 0,$$

where each $q_i(x_n)$ is a univariate polynomial.

See Becker, Marianari, Mora, and Treverso (1993) for the proof of this result. There are several important aspects of the shape lemma. First, most polynomial systems have a Gröbner basis of this form. Second, most algorithms for obtaining a Gröbner basis returns the above form. For instance, Mathematica's `GroebnerBasis[]` command implements the shape lemma. Third, it is straightforward to find all the roots of the univariate polynomial $q_n(x_n)$. With these values of x_n in hand, it is trivial to find $x_1, \ldots, x_{n-1}$.

A large strand of literature has dealt with the computation of Gröbner bases in the shape lemma. Buchberger's (1998) algorithm is the original technique. A number of more efficient variants have been subsequently proposed. We refer the interested reader

to Cox, Little, and O'Shea (1997). In this paper we use Mathematica to find a Gröbner basis.

To illustrate how powerful the shape lemma is, consider the following example featuring a system of quadratic polynomial equations in four unknown variables $x_1, \ldots, x_4$:

$$x_1x_2 + x_3x_4 + 2 = 0, \qquad x_1x_2 + x_2x_3 + 3 = 0,$$
$$x_1x_3 + x_4x_1 + x_4x_2 + 6 = 0, \quad \text{and} \quad x_1x_3 + 2x_1x_2 + 3 = 0.$$

A Gröbner basis of the form given in the shape lemma is

$$x_1 - \frac{1}{28}\left(9x_4^5 + 6x_4^3 - 15x_4\right) = 0, \qquad x_2 - \frac{1}{28}\left(-9x_4^5 - 6x_4^3 + 99x_4\right) = 0,$$
$$x_3 - \frac{1}{14}\left(-3x_4^5 - 9x_4^3 - 2x_4\right) = 0, \quad \text{and} \quad 3x_4^6 + 9x_4^4 - 19x_4^2 - 49 = 0.$$

The last polynomial is univariate of degree 6 in x_4. There are six roots for this polynomial. Each of these roots can be substituted into the first three equations to obtain all six solutions. The theory of Gröbner bases ensures that these solutions are the same as those of the original system. This example illustrates the multiple-solution nature of a system of quadratic polynomial equations.

References

Aldrich, E., J. Fernández-Villaverde, R. Gallant, and J. Rubio-Ramírez (2011), "Tapping the supercomputer under your desk: Solving dynamic equilibrium models with graphics processors." *Journal of Economic Dynamics and Control*, 35 (3), 386–393. [657]

Amisano, G. and O. Tristani (2011), "Exact likelihood computation for nonlinear DSGE models with heteroskedastic innovations." *Journal of Economic Dynamics and Control*, 35 (12), 2167–2185. [639]

Andreasen, M., J. Fernández-Villaverde, and J. Rubio-Ramírez (2013), "The pruned state-space system for non-linear DSGE models: Theory and empirical applications." Working Paper 18983, NBER. [658]

Aruoba, S. B., J. Fernández-Villaverde, and J. F. Rubio-Ramírez (2006), "Comparing solution methods for dynamic equilibrium economies." *Journal of Economic Dynamics and Control*, 30 (12), 2477–2508. [657]

Becker, E., M. G. Marianari, T. Mora, and C. Treverso (1993), "The shape of the shape lemma." In *International Symposium on Symbolic and Algebraic Computation*, 129–133, ACM Press, New York. [665]

Bi, H. and N. Traum (2012), "Estimating sovereign default risk." *European Economic Review*, 102 (3), 161–166. [638]

Bianchi, F. (2010), "Regime switches, agents' beliefs, and post-World War II US macroeconomic dynamics." Working Paper 12-04, Duke University. [638]

Bianchi, F. and L. Melosi (2013), "Modeling the evolution of expectations and uncertainty in general equilibrium." Manuscript. [638]

Blake, A. and F. Zampolli (2006), "Optimal monetary policy in Markov-switching models with rational expectations agents." Working Paper 298, Bank of England. [638]

Bloom, N. (2009), "The impact of uncertainty shocks." *Econometrica*, 77 (3), 623–685. [639]

Borovička, J. and L. P. Hansen (2013), "Examining macroeconomic models through the lens of asset pricing." Manuscript. [638, 642]

Buchberger, B. (1998), "An algorithmic criterion for the solvability of algebraic systems of equations." *Gröbner Bases and Applications*, 251, 535–545. [665]

Cho, S. (2011), "Characterizing Markov-switching rational expectation models." Working paper. [650]

Clarida, R., J. Gali, and M. Gertler (2000), "Monetary policy rules and macroeconomic stability: Evidence and some theory." *Quarterly Journal of Economics*, 115 (1), 147–180. [659]

Costa, O., M. Fragoso, and R. Marques (2005), *Discrete-Time Markov Jump Linear Systems*. Springer, London. [640, 650]

Cox, D. A., J. B. Little, and D. B. O'Shea (1997), *Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, second edition. Undergraduate Texts in Mathematics, Springer-Verlag, New York. [666]

Davig, T. and T. Doh (2008), "Monetary policy regime shifts and inflation persistence." Research Working Paper 08-16, Federal Reserve Bank of Kansas City. [638]

Davig, T. and E. Leeper (2007), "Generalizing the Taylor principle." *American Economic Review*, 97 (3), 607–635. [638, 659]

Davig, T., E. M. Leeper, and T. B. Walker (2010), "Unfunded liabilities and uncertain fiscal financing." *Journal of Monetary Economics*, 57 (5), 600–619. [638]

Davig, T., E. M. Leeper, and T. B. Walker (2011), "Inflation and the fiscal limit." *European Economic Review*, 55 (1), 31–47. [638]

Farmer, R., D. Waggoner, and T. Zha (2009), "Understanding Markov-switching rational expectations models." *Journal of Economic Theory*, 144 (5), 1849–1867. [638, 640, 650, 659]

Farmer, R., D. Waggoner, and T. Zha (2011), "Minimal state variable solutions to Markov-switching rational expectations models." *Journal of Economic Dynamics and Control*, 35 (12), 2150–2166. [638, 650]

Fernández-Villaverde, J., P. Guerrón-Quintana, and J. F. Rubio-Ramírez (2014), "Estimating dynamic equilibrium models with stochastic volatility." Manuscript. [649]

Fernández-Villaverde, J., P. Guerrón-Quintana, J. F. Rubio-Ramírez, and M. Uribe (2011), "Risk matters: The real effects of volatility shocks." *American Economic Review*, 101 (6), 2530–2561. [654]

Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007), "Estimating macroeconomic models: A likelihood approach." *Review of Economic Studies*, 74 (4), 1059–1087. [654]

Gomme, P. and P. Klein (2011), "Second-order approximation of dynamic models without the use of tensors." *Journal of Economic Dynamics and Control*, 35, 604–615. [642]

Hamilton, J. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle." *Econometrica*, 57, 357–384. [659]

Holmes, M. H. (2012), *Introduction to Perturbation Methods*, second edition. Springer, New York. [642]

Judd, K. (1998), *Numerical Methods in Economics*. MIT Press, Cambridge, MA. [638, 642, 653]

Judd, K. and A. Solnick (1994), "Numerical dynamic programming with shape-preserving splines." Report. [657]

Klein, P. (2000), "Using the generalized Schur form to solve a multivariate linear rational expectations model." *Journal of Economic Dynamics and Control*, 24, 1405–1423. [650]

Leeper, E. and T. Zha (2003), "Modest policy interventions." *Journal of Monetary Economics*, 50 (8), 1673–1700. [638]

Liu, Z., D. F. Waggoner, and T. Zha (2011), "Sources of macroeconomic fluctuations: A regime-switching DSGE approach." *Quantitative Economics*, 2 (2), 251–301. [638]

Lombardo, G. (2010), "On approximating DSGE models by series expansions." European Central Bank Working Paper 1264. [638, 642]

Lubik, T. A. and F. Schorfheide (2004), "Testing for indeterminacy: An application to US monetary policy." *American Economic Review*, 94 (1), 190–217. [659]

Santos, M. S. and A. Peralta-Alva (2005), "Accuracy of simulations for stochastic dynamic models." *Econometrica*, 73 (6), 1939–1976. [658]

Schmitt-Grohe, S. and M. Uribe (2004), "Solving dynamic general equilibrium models using a second-order approximation to the policy function." *Journal of Economic Dynamics and Control*, 28 (4), 755–775. [638, 642]

Schorfheide, F. (2005), "Learning and monetary policy shifts." *Review of Economic Dynamics*, 8, 392–419. [638]

Sims, C. A. and T. Zha (2006), "Were there regime switches in US monetary policy?" *American Economic Review*, 96 (1), 54–81. [659]

Svensson, L. and N. Williams (2007), "Monetary policy with model uncertainty: Distribution forecast targeting." Discussion Paper 6331, CEPR. [638, 650]

Tauchen, G. (1986), "Finite state Markov-chain approximations to univariate and vector autoregressions." *Economics Letters*, 20 (2), 177–181. [657]

Uhlig, H. (1999), "A toolkit for analyzing nonlinear dynamic stochastic models easily." In *Computational Methods for the Study of Dynamic Economies* (R. Marimon and A. Scott, eds.), 30–61, Oxford University Press, Oxford, England. [657]

Co-editor Karl Schmedders handled this manuscript.

Submitted July, 2015. Final version accepted November, 2015.