The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications*

Martin M. Andreasen†
Aarhus University and CREATES

Jesús Fernández-Villaverde
University of Pennsylvania, NBER, and CEPR

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

April 21, 2016

Abstract

This paper studies the pruned state-space system for higher-order perturbation approximations to DSGE models. We show the stability of the pruned approximation up to third order and provide closed-form expressions for first and second unconditional moments and impulse response functions. Our results introduce GMM estimation and impulse-response matching for DSGE models approximated up to third order and provide a foundation for indirect inference and SMM. As an application, we consider a New Keynesian model with Epstein-Zin-Weil preferences and two novel feedback effects from long-term bonds to the real economy, allowing us to match the level and variability of the 10-year term premium in the U.S. with a low relative risk aversion of 5.

Keywords: Epstein-Zin-Weil preferences, Feedback-effects from long-term bonds, Higher-order perturbation approximation, Yield curve.

JEL: C15, C53, E30.

*We thank Mads Dang, Frank Diebold, Robert Kollmann, Dirk Krueger, Johannes Pfeifer, Eric Renault, Francisco Ruge-Murcia, Frank Schorfheide, Eric Swanson, the editor, and referees for useful comments. Remarks and suggestions from seminar participants at numerous institutions are much appreciated. Beyond the usual disclaimer, we must note that any views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. We acknowledge access to computer facilities provided by the Danish Center for Scientific Computing (DCSC). We acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Finally, we also thank the NSF for financial support.

†Corresponding author: Fuglesangs Allé 4, 8210 Aarhus V, Denmark, email: mandreasen@econ.au.dk; telephone +0045 87165982.

1 Introduction

The perturbation method approximates the solution to dynamic stochastic general equilibrium (DSGE) models by higher-order Taylor series expansions around the steady state (see Judd and Guu (1997) and Schmitt-Grohé and Uribe (2004), among others). These approximations have grown in popularity, mainly because they allow researchers to quickly and accurately solve DSGE models with many state variables and inherent non-linearities to analyze uncertainty shocks or time-varying risk premia (see Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez and Uribe (2011) and Rudebusch and Swanson (2012), among others).

Although higher-order approximations are intuitive and straightforward to compute, they often generate explosive sample paths even when the corresponding linearized solution is stable. As noted by Kim, Kim, Schaumburg and Sims (2008), these explosive sample paths arise because the higher-order terms generate unstable steady states in the approximated system. The presence of explosive behavior complicates any model evaluation because no unconditional moments exist in this approximation. It also means that any estimation method using unconditional moments, such as the generalized method of moments (GMM) or the simulated method of moments (SMM), is inapplicable because it relies on finite moments from stationary and ergodic probability distributions.[1]

For second-order approximations, Kim, Kim, Schaumburg and Sims (2008) suggest eliminating explosive sample paths by applying a pruning method that omits terms of higher-order effects than the considered approximation order when the system is iterated forward in time. To illustrate the idea, suppose we have a solution for capital k_t that depends on a quadratic function of k_{t-1}, as present in many DSGE models solved to second order. If we iterate this equation one period forward and substitute k_t for its own quadratic function of k_{t-1}, we obtain an expression for k_{t+1} that depends on k_{t-1}, k_{t-1}^2, k_{t-1}^3, and k_{t-1}^4. The pruning method omits the terms k_{t-1}^3 and k_{t-1}^4 capturing third- and fourth-order effects to obtain a second-order approximation of k_{t+1} when expressed as a function of the current state variable k_{t-1}.

This paper extends the pruning method to perturbation approximations of any order and shows

[1] Ruge-Murcia (2013) reviews the use of GMM in the context of DSGE models. Non-explosive sample paths are also required for likelihood methods, for instance, when using the particle filter outlined in Fernández-Villaverde and Rubio-Ramírez (2007).

how pruning greatly facilitates inference of DSGE models. Special attention is devoted to second- and third-order approximations, which are widely used. We first show that our pruning method ensures stable sample paths, provided the linearized solution is stable. Given this key result, we then provide closed-form solutions for first and second unconditional moments and impulse response functions (IRFs). We also derive conditions for the existence of third and fourth unconditional moments to compute skewness and kurtosis.[2]

The econometric implications of these results are significant as most of the existing moment-based estimation methods for linearized DSGE models now carry over to non-linear approximations. For models solved up to third order, this includes GMM estimation based on first and second unconditional moments and matching model-implied IRFs to their empirical counterparts. Our results are also useful when estimating DSGE models using Bayesian methods, for instance, when conducting inference using a limited information likelihood function from unconditional moments, as suggested by Kim (2002), or when doing posterior model evaluations on unconditional moments, as in An and Schorfheide (2007). If simulations are needed to calculate higher-order unconditional moments such as skewness or kurtosis, then our results provide a foundation for SMM as in Duffie and Singleton (1993) and different types of indirect inference as in Smith (1993). Finally, our results are also relevant to researchers who prefer to calibrate their models as in Cooley and Prescott (1995), because the unconditional mean of a model solved with higher-order terms generally differs from its steady-state value. Given our results, researchers can now easily correct for these higher-order effects and non-linearly calibrate their models.[3]

The suggested GMM estimation approach, its Bayesian equivalent, non-linear calibration, and IRF matching are promising because we can compute first and second unconditional moments or IRFs in a trivial amount of time for medium-size DSGE models solved up to third order. For the model described in Section 8 with seven state variables, it takes 0.75 second to find all first and second unconditional moments and only 0.08 second to compute the IRFs for 20 periods following a shock on an off-the-shelf laptop.

[2] `Matlab` codes to implement our procedures are available on the authors' home pages; see, for instance, `https://sites.google.com/site/mandreasendk/home-1`. We also note that Dynare 4.4.0. has implemented our pruning method to simulate models approximated to third order.

[3] Some papers in the literature have accounted for the difference between the steady state and the mean of the ergodic distribution by simulation; see, for instance, Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez and Uribe (2011). These simulations are, however, computationally demanding, in particular, for very persistent processes, where a long sample path is required to accurately compute unconditional moments.

An application illustrates some of the new techniques that our paper makes available for DSGE models. We consider a rich New Keynesian economy with Calvo pricing, consumption habits, and Epstein-Zin-Weil preferences, which we estimate by GMM using first and second unconditional moments for the U.S. yield curve and five macro variables. Our New Keynesian model introduces two novel mechanisms that help us to improve our understanding of the interactions between financial markets, monetary policy, and the real economy. First, households deposit their savings in a financial intermediary. This financial intermediary invests in short- and long-term bonds and creates a wedge between the policy rate set by the monetary authority and the interest rate on deposits. Second, we augment the standard Taylor rule of the monetary authority to include the excess return in a longer-term bond, which is closely related to term premia. The first mechanism captures the frictions in the financial markets that induce differences between the policy rate and the interest rate faced by private agents. The second mechanism captures the observation that central banks also react to term premia, as seen during the recent financial crisis. Our two mechanisms depend on the degree of precautionary behavior and, therefore, are only operative when the model is solved using a third-order approximation. Thus, the methods derived in the present paper are essential for the quantitative analysis of the model.

Our model matches the mean and variability of the 10-year term premium with a reasonable risk aversion of 5, while simultaneously matching key moments for standard real macro variables. We illustrate the importance of a positive steady-state inflation in driving this result, as it amplifies the non-linearities in the price dispersion index related to Calvo pricing and produces the desired conditional heteroscedasticity in the stochastic discount factor. Notably, an unpruned third-order approximation to our model gives explosive sample paths and is, therefore, unable to "see" this novel channel for term premia volatility, which we uncover when using our pruning method. Thus, our model and our pruning method go a long way in resolving the bond risk premium puzzle described in Rudebusch and Swanson (2008) without postulating highly risk-averse households, as in much of the existing literature.

The rest of the paper is structured as follows. Section 2 introduces the problem. Section 3 presents the pruning method and the pruned state-space system for approximated DSGE models. Stability and unconditional moments of the pruned state-space system for second- and third-order approximations are derived in Section 4, with the closed-form expressions for the IRFs deferred to

Section 5. Section 6 studies the accuracy of the pruning method, and we discuss the econometric implications of the pruned state-space system in Section 7. Section 8 is devoted to our empirical application. Section 9 concludes. Detailed derivations and proofs are deferred to the Appendix and a longer Online Appendix available on the authors' home pages or on request.

2 The State-Space System

We consider the following class of DSGE models. Let $\mathbf{y}_t \in \mathbb{R}^{n_y}$ be a vector of control variables, $\mathbf{x}_t \in \mathbb{R}^{n_x}$ a vector of state variables, and $\sigma \geq 0$ an auxiliary perturbation parameter. To simplify the notation below, $\mathbf{x}_t$ and $\mathbf{y}_t$ are expressed in deviations from their steady state. The exact solution to the DSGE model is given by the state-space system

$$\mathbf{y}_t = \mathbf{g}\left(\mathbf{x}_t, \sigma\right), \tag{1}$$

$$\mathbf{x}_{t+1} = \mathbf{h}\left(\mathbf{x}_t, \sigma\right) + \sigma \boldsymbol{\eta} \boldsymbol{\epsilon}_{t+1}, \tag{2}$$

where $\boldsymbol{\epsilon}_{t+1}$ contains the n_ϵ exogenous zero-mean innovations. We refer to (1) and (2) as the observation and state equations, respectively. Initially, we do not impose a distributional form for the innovations. In particular, the innovations may be non-Gaussian. We only assume that $\boldsymbol{\epsilon}_{t+1}$ is independent and identically distributed with finite second moments, denoted by $\boldsymbol{\epsilon}_{t+1} \sim \mathcal{IID}\left(\mathbf{0}, \mathbf{I}\right)$. Additional moment restrictions will be imposed later in the paper. The perturbation parameter σ scales the square root of the covariance matrix for the innovations $\boldsymbol{\eta}$ having dimension $n_x \times n_\epsilon$.[4]

In general, DSGE models do not have a closed-form solution and the functions $\mathbf{g}\left(\mathbf{x}_t, \sigma\right)$ and $\mathbf{h}\left(\mathbf{x}_t, \sigma\right)$ cannot be found explicitly. The perturbation method is a popular way to obtain Taylor series expansions to these functions around the steady state. When the functions $\mathbf{g}\left(\mathbf{x}_t, \sigma\right)$ and $\mathbf{h}\left(\mathbf{x}_t, \sigma\right)$ are solved up to first order, the state-space system is approximated by $\mathbf{g}_\mathbf{x}\mathbf{x}_t$ and $\mathbf{h}_\mathbf{x}\mathbf{x}_t$ in (1) and (2), respectively. Here, $\mathbf{g}_\mathbf{x}$ is an $n_y \times n_x$ matrix with first-order derivatives of $\mathbf{g}\left(\mathbf{x}_t, \sigma\right)$ with respect to $\mathbf{x}_t$ and $\mathbf{h}_\mathbf{x}$ is an $n_x \times n_x$ matrix with first-order derivatives of $\mathbf{h}\left(\mathbf{x}_t, \sigma\right)$ with respect to $\mathbf{x}_t$.[5] Given our assumptions about $\boldsymbol{\epsilon}_{t+1}$, this system has finite first and second unconditional moments

[4] The assumption that innovations enter linearly in (2) may appear restrictive but is without loss of generality. As shown in Appendix A.1, the state vector can be extended to deal with non-linearities between $\mathbf{x}_t$ and $\boldsymbol{\epsilon}_{t+1}$.

[5] The first-order derivatives $\mathbf{g}_\sigma$ and $\mathbf{h}_\sigma$ of $\mathbf{g}\left(\mathbf{x}_t, \sigma\right)$ and $\mathbf{h}\left(\mathbf{x}_t, \sigma\right)$ with respect to σ are known to be zero (see Schmitt-Grohé and Uribe (2004)).

if all eigenvalues of $\mathbf{h_x}$ have modulus less than one. Furthermore, the approximated state-space system fluctuates around the steady state, which also corresponds to the unconditional mean. It is, therefore, straightforward to calibrate the structural parameters in the DSGE model from unconditional first and second moments or carry out a formal estimation using existing econometric tools for Bayesian inference, maximum likelihood, GMM, SMM, etc. (see Ruge-Murcia (2007)).

When the functions $\mathbf{g}(\mathbf{x}_t, \sigma)$ and $\mathbf{h}(\mathbf{x}_t, \sigma)$ are approximated beyond linearization, we could, in principle, apply the same method to construct the approximated state-space system with their higher-order Taylor series expansions. However, the resulting approximated state-space system cannot, in general, be shown to have any finite unconditional moments and may even display explosive dynamics. This occurs even when we simulate simple versions of the New Keynesian model with few endogenous state variables. Hence, it is hard to use this approximated state-space system to calibrate or even estimate model parameters. Consequently, it is useful to construct another approximated state-space system that has well-defined statistical properties when analyzing DSGE models solved beyond linearization. We explain now how this can be done.

3 The Pruning Method

Kim, Kim, Schaumburg and Sims (2008) suggest using a pruning method to construct the approximated state-space system for DSGE models solved to second order. We will refer to this approach as the pruned state-space system. Section 3.1 reviews the pruning method and explains its logic for the second-order approximation. Section 3.2 extends the method to a third-order approximation. The general procedure for constructing the pruned state-space system for any approximation order is straightforward, but deferred to Appendix A.2 in the interest of space. We finally relate our approach to the existing literature in Section 3.3.

3.1 Second-Order Approximation

The first step when constructing the pruned state-space system for the second-order approximation is to decompose the state variables into first-order effects $\mathbf{x}_t^f$ and second-order effects $\mathbf{x}_t^s$ as follows.

We start from the second-order Taylor series expansion of the state equation

$$\mathbf{x}_{t+1}^{(2)} = \mathbf{h_x}\mathbf{x}_t^{(2)} + \frac{1}{2}\mathbf{H_{xx}}\left(\mathbf{x}_t^{(2)} \otimes \mathbf{x}_t^{(2)}\right) + \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 + \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}, \tag{3}$$

where $\mathbf{x}_t^{(2)}$ is the unpruned second-order approximation to the state variables.[6] Here, $\mathbf{H_{xx}}$ is an $n_x \times n_x^2$ matrix with the derivatives of $\mathbf{h}(\mathbf{x}_t, \sigma)$ with respect to $(\mathbf{x}_t, \mathbf{x}_t)$ and $\mathbf{h}_{\sigma\sigma}$ is an $n_x \times 1$ matrix containing derivatives taken with respect to (σ, σ). Substituting $\mathbf{x}_t^{(2)}$ with $\mathbf{x}_t^f + \mathbf{x}_t^s$ into the right-hand side of (3) gives

$$\mathbf{h_x}\left(\mathbf{x}_t^f + \mathbf{x}_t^s\right) + \frac{1}{2}\mathbf{H_{xx}}\left(\left(\mathbf{x}_t^f + \mathbf{x}_t^s\right) \otimes \left(\mathbf{x}_t^f + \mathbf{x}_t^s\right)\right) + \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 + \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}. \tag{4}$$

A law of motion for $\mathbf{x}_{t+1}^f$ is derived by preserving only first-order effects in (4). We keep the first-order effects from the previous period $\mathbf{h_x}\mathbf{x}_t^f$ and the innovations $\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}$ to obtain

$$\mathbf{x}_{t+1}^f = \mathbf{h_x}\mathbf{x}_t^f + \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}. \tag{5}$$

This expression for $\mathbf{x}_{t+1}^f$ is the standard first-order approximation to the state equation. Note that $\mathbf{x}_{t+1}^f$ is a polynomial in $\{\boldsymbol{\epsilon}_s\}_{s=1}^{t+1}$ that only includes first-order terms. The first-order approximation to the observation equation is also standard and given by

$$\mathbf{y}_t^f = \mathbf{g_x}\mathbf{x}_t^f. \tag{6}$$

Accordingly, the pruned state-space system for the first-order approximation is given by (5) and (6), meaning that the pruned and unpruned state-space systems are identical in this case.

A law of motion for $\mathbf{x}_{t+1}^s$ is derived by preserving only second-order effects in (4). Here, we include the second-order effects from the previous period $\mathbf{h_x}\mathbf{x}_t^s$, the squared first-order effects in the previous period $\frac{1}{2}\mathbf{H_{xx}}\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right)$, and the correction $\frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2$. Hence,

$$\mathbf{x}_{t+1}^s = \mathbf{h_x}\mathbf{x}_t^s + \frac{1}{2}\mathbf{H_{xx}}\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right) + \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2. \tag{7}$$

[6] Equation (3) adopts the standard assumption that the model has a unique stable first-order approximation, which implies that all second- and higher-order terms are also unique (see Judd and Guu (1997) and Lan and Meyer-Gohde (2014)).

We do not include terms with $\mathbf{x}_t^f \otimes \mathbf{x}_t^s$ and $\mathbf{x}_t^s \otimes \mathbf{x}_t^s$ because they reflect third- and fourth-order effects, respectively. Note that $\mathbf{x}_{t+1}^s$ is a polynomial in $\{\boldsymbol{\epsilon}_s\}_{s=1}^t$ that only includes second-order terms.

The final step in setting up the pruned state-space system is to derive the expression for the observation equation. Using the same approach as above, we start from the second-order Taylor series expansion of the observation equation

$$\mathbf{y}_t^{(2)} = \mathbf{g}_{\mathbf{x}}\mathbf{x}_t^{(2)} + \frac{1}{2}\mathbf{G}_{\mathbf{xx}}\left(\mathbf{x}_t^{(2)} \otimes \mathbf{x}_t^{(2)}\right) + \frac{1}{2}\mathbf{g}_{\sigma\sigma}\sigma^2, \tag{8}$$

where $\mathbf{y}_t^{(2)}$ denotes the unpruned second-order approximation to the control variables. Here, $\mathbf{G}_{\mathbf{xx}}$ is an $n_y \times n_x^2$ matrix with the corresponding derivatives of $\mathbf{g}(\mathbf{x}_t, \sigma)$ with respect to $(\mathbf{x}_t, \mathbf{x}_t)$ and $\mathbf{g}_{\sigma\sigma}$ is an $n_y \times 1$ matrix containing derivatives with respect to (σ, σ). We only want to preserve effects up to second order, meaning that the pruned approximation to the control variables is given by

$$\mathbf{y}_t^s = \mathbf{g}_{\mathbf{x}}\left(\mathbf{x}_t^f + \mathbf{x}_t^s\right) + \frac{1}{2}\mathbf{G}_{\mathbf{xx}}\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right) + \frac{1}{2}\mathbf{g}_{\sigma\sigma}\sigma^2. \tag{9}$$

Here, we leave out terms with $\mathbf{x}_t^f \otimes \mathbf{x}_t^s$ and $\mathbf{x}_t^s \otimes \mathbf{x}_t^s$ because they reflect third- and fourth-order effects, respectively. To simplify notation, we treat $\mathbf{y}_t^s$ as the sum of the first- and second-order effects, while $\mathbf{x}_t^s$ only contains the second-order effects. Hence, $\mathbf{y}_t^s$ is a polynomial in $\{\boldsymbol{\epsilon}_s\}_{s=1}^t$ that includes all first- and second-order terms.[7]

Accordingly, the pruned state-space system for the second-order approximation is given by (5), (7), and (9). The state vector in this system is thus extended to $\begin{bmatrix} \left(\mathbf{x}_t^f\right)' & (\mathbf{x}_t^s)' \end{bmatrix}'$ as we separately track first- and second-order effects. For completeness, the unpruned state-space system for the second-order approximation is given by (3) and (8).

3.2 Third-Order Approximation

We now construct the pruned state-space system for the third-order approximation. Following the steps outlined above, we start by decomposing the state variables into first-order effects $\mathbf{x}_t^f$,

[7] Lan and Meyer-Gohde (2013*b*) employ perturbation to derive a stable non-linear approximation of $\mathbf{g}(\mathbf{x}_t, \sigma)$ and $\mathbf{h}(\mathbf{x}_t, \sigma)$ in terms of past innovations. By using (5), (7), and (9), we can also express $\mathbf{y}_t^s$ as an infinite moving average in terms of past innovations.

second-order effects $\mathbf{x}_t^s$, and third-order effects $\mathbf{x}_t^{rd}$. The laws of motion for $\mathbf{x}_t^f$ and $\mathbf{x}_t^s$ are the same as in the previous section, and only the recursion for $\mathbf{x}_t^{rd}$ remains to be derived. The third-order Taylor series expansion to the state equation is (see Ruge-Murcia (2012))

$$\begin{aligned}\mathbf{x}_{t+1}^{(3)} &= \mathbf{h}_{\mathbf{x}}\mathbf{x}_t^{(3)} + \frac{1}{2}\mathbf{H}_{\mathbf{xx}}\left(\mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) + \frac{1}{6}\mathbf{H}_{\mathbf{xxx}}\left(\mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) \\ &\quad + \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 + \frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2\mathbf{x}_t^{(3)} + \frac{1}{6}\mathbf{h}_{\sigma\sigma\sigma}\sigma^3 + \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}, \end{aligned} \tag{10}$$

where $\mathbf{x}_t^{(3)}$ represents the unpruned third-order approximation to the state variables. Here, $\mathbf{H}_{\mathbf{xxx}}$ denotes an $n_x \times n_x^3$ matrix containing derivatives of $\mathbf{h}(\mathbf{x}_t, \sigma)$ with respect to $(\mathbf{x}_t, \mathbf{x}_t, \mathbf{x}_t)$, $\mathbf{h}_{\sigma\sigma\mathbf{x}}$ is an $n_x \times n_x$ matrix including derivatives with respect to $(\sigma, \sigma, \mathbf{x}_t)$, and $\mathbf{h}_{\sigma\sigma\sigma}$ is an $n_x \times 1$ matrix containing derivatives related to (σ, σ, σ). We adopt the same procedure as before and substitute $\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}$ into the right-hand side of (10) to obtain

$$\begin{aligned}&\mathbf{h}_{\mathbf{x}}\left(\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}\right) + \frac{1}{2}\mathbf{H}_{\mathbf{xx}}\left(\left(\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}\right) \otimes \left(\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}\right)\right) \\ &+ \frac{1}{6}\mathbf{H}_{\mathbf{xxx}}\left(\left(\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}\right) \otimes \left(\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}\right) \otimes \left(\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}\right)\right) \\ &+ \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 + \frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2\left(\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}\right) + \frac{1}{6}\mathbf{h}_{\sigma\sigma\sigma}\sigma^3 + \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}. \end{aligned} \tag{11}$$

A law of motion for the third-order effects is derived by preserving only third-order terms in (11):

$$\mathbf{x}_{t+1}^{rd} = \mathbf{h}_{\mathbf{x}}\mathbf{x}_t^{rd} + \mathbf{H}_{\mathbf{xx}}\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^s\right) + \frac{1}{6}\mathbf{H}_{\mathbf{xxx}}\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}^f\right) + \frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2\mathbf{x}_t^f + \frac{1}{6}\mathbf{h}_{\sigma\sigma\sigma}\sigma^3. \tag{12}$$

As in the derivation of the law of motion for x_t^s in (7), σ is interpreted as a variable when constructing (12). This means that $\frac{3}{6}h_{\sigma\sigma\mathbf{x}}\sigma^2 x_t^s$ and $\frac{3}{6}h_{\sigma\sigma\mathbf{x}}\sigma^2 x_t^{rd}$ represent fourth- and fifth-order effects, respectively, and are therefore omitted. Note that $\mathbf{x}_{t+1}^{rd}$ is a polynomial in $\{\boldsymbol{\epsilon}_s\}_{s=1}^t$ that only includes third-order terms.

The final step is to set up the expression for the observation equation. Using results in Ruge-Murcia (2012), the third-order Taylor series expansion is given by

$$\begin{aligned}\mathbf{y}_t^{(3)} &= \mathbf{g}_{\mathbf{x}}\mathbf{x}_t^{(3)} + \frac{1}{2}\mathbf{G}_{\mathbf{xx}}\left(\mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) + \frac{1}{6}\mathbf{G}_{\mathbf{xxx}}\left(\mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) \\ &\quad + \frac{1}{2}\mathbf{g}_{\sigma\sigma}\sigma^2 + \frac{3}{6}\mathbf{g}_{\sigma\sigma\mathbf{x}}\sigma^2\mathbf{x}_t^{(3)} + \frac{1}{6}\mathbf{g}_{\sigma\sigma\sigma}\sigma^3, \end{aligned} \tag{13}$$

where $\mathbf{y}_t^{(3)}$ represents the unpruned third-order approximation to the control variables. In (13), $\mathbf{G}_{\mathbf{xxx}}$ denotes an $n_y \times n_x^3$ matrix containing derivatives of $\mathbf{g}(\mathbf{x}_t, \sigma)$ with respect to $(\mathbf{x}_t, \mathbf{x}_t, \mathbf{x}_t)$, $\mathbf{g}_{\sigma\sigma\mathbf{x}}$ is an $n_y \times n_x$ matrix including derivatives with respect to $(\sigma, \sigma, \mathbf{x}_t)$, and $\mathbf{g}_{\sigma\sigma\sigma}$ is an $n_y \times 1$ matrix containing derivatives related to (σ, σ, σ). To simplify notation, we treat $\mathbf{y}_t^{rd}$ as the sum of the first-, second-, and third-order effects, while $\mathbf{x}_t^{rd}$ is only the third-order effect. Hence, preserving effects up to third-order gives

$$\begin{aligned}\mathbf{y}_t^{rd} \quad = \quad & \mathbf{g}_{\mathbf{x}}\left(\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}\right) + \frac{1}{2}\mathbf{G}_{\mathbf{xx}}\left(\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right) + 2\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^s\right)\right) \\ & + \frac{1}{6}\mathbf{G}_{\mathbf{xxx}}\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f\right) + \frac{1}{2}\mathbf{g}_{\sigma\sigma}\sigma^2 + \frac{3}{6}\mathbf{g}_{\sigma\sigma\mathbf{x}}\sigma^2\mathbf{x}_t^f + \frac{1}{6}\mathbf{g}_{\sigma\sigma\sigma}\sigma^3,\end{aligned} \tag{14}$$

which is a polynomial in $\{\boldsymbol{\epsilon}_s\}_{s=1}^t$ that include all first-, second-, and third-order terms.

Thus, the pruned state-space system for the third-order approximation is given by (5), (7), (12), and (14). The state vector in this system is further extended to $\left[\begin{array}{ccc}\left(\mathbf{x}_t^f\right)' & \left(\mathbf{x}_t^s\right)' & \left(\mathbf{x}_t^{rd}\right)'\end{array}\right]'$, as we need to separately track first-, second-, and third-order effects. For completeness, the unpruned state-space system for the third-order approximation is given by (10) and (13).

3.3 Related Literature

Lombardo and Sutherland (2007) pioneered the idea of separately keeping track of first- and second-order effects to solve for a second-order perturbation of DSGE models. We extend their idea to approximations beyond second order. Since the first circulation of our paper, Lombardo and Uhlig (2014) have presented an alternative derivation of our pruned state-space system, but only for models without interaction between the innovations and the state variables. We find, nevertheless, that our approach allows us to easily derive many results and that the lack of interaction between the innovations and the state variables in Lombardo and Uhlig (2014) is too restrictive in many models of interest.

Our pruning approach is also analyzed in Haan and Wind (2012), who highlight two potential disadvantages of the method. First, pruning induces a larger vector of states than the unpruned approximation. Second, the pruned state-space system for the kth-order approximation cannot exactly fit the exact solution if it happens to be a kth-order polynomial. We do not consider the large state vector to be a problem because we find it informative to assess how important each of the second- and third-order effects is relative to the first-order effects. In addition, current computing power makes memory considerations less of a constraint. Indeed, Section 8 shows that the pruned state-space system for a third-order approximation to a medium-size DSGE model is easily obtained and stored. We also view the second disadvantage as minor because an exact fit can be obtained by raising the approximation beyond order k, as also acknowledged by Haan and Wind (2012).

Our pruning scheme differs from the alternative presented in Haan and Wind (2012) along two dimensions. First, for approximations beyond second order, these authors include terms with higher-order effects than the approximation order. For example, in the case of a third-order approximation, their first proposal for a pruning scheme includes some fourth-order effects, whereas their second proposal includes some fifth- and sixth-order effects. Second, their pruning scheme is expressed around what they refer to as the stochastic steady state, while our pruning scheme is expressed around the steady state. An advantage of our choices (i.e., omitting all higher-order effects than the approximation order and approximating around the steady state) is that they allow the derivation of unconditional moments in closed form. Furthermore, approximating around the steady state is consistent with our treatment of σ as a variable.[8]

We conclude by stressing that, if a non-linear perturbation approximation does not preserve monotonicity and convexity of the exact policy function - as seen for extreme calibrations of DSGE models - then pruning will not restore these properties. For small DSGE models, Haan and Wind (2012) propose the perturbation-plus approximation and show that it may restore these properties of the policy function. However, the perturbation-plus algorithm is numerically demanding, even for small models, and does not allow the unconditional moments to be obtained in closed form.

4 Statistical Properties of the Pruned System

This section shows that the pruned state-space system has well-defined statistical properties and presents our closed-form expressions for first and second unconditional moments.[9] We proceed as

[8]After we circulated the first version of our paper, Francisco Ruge-Murcia directed our attention to his unpublished work on pruning at third order (Kim and Ruge-Murcia (2009) and Ruge-Murcia (2012)). Ruge-Murcia's approach is broadly similar to the one in Haan and Wind (2012), but with additional approximations imposed to compute unconditional moments.

[9]Although not explicitly considered in this paper, it is straightforward to compute conditional moments for the state and control variables based on the expressions provided below.

follows. Section 4.1 extends the analysis in Kim, Kim, Schaumburg and Sims (2008) for a second-order approximation, and Section 4.2 conducts a similar analysis for a third-order approximation. Applying the steps below to higher-order approximations is conceptually transparent.

4.1 Second-Order Approximation

In this section it is convenient to consider a more compact representation of the pruned state-space system than the one provided in Section 3.1. Therefore, we introduce the vector

$$\mathbf{z}_t^{(2)} \equiv \left[\begin{array}{ccc} \left(\mathbf{x}_t^f\right)' & \left(\mathbf{x}_t^s\right)' & \left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right)' \end{array} \right]',$$

where the superscript for $\mathbf{z}_t$ denotes the approximation order. The first n_x elements in $\mathbf{z}_t^{(2)}$ are the first-order effects, while the remaining part of $\mathbf{z}_t^{(2)}$ contains second-order effects. The laws of motion for $\mathbf{x}_t^f$ and $\mathbf{x}_t^s$ are stated above and the evolution for $\mathbf{x}_t^f \otimes \mathbf{x}_t^f$ is easily derived from (5). This allows us to write the laws of motion for the first- and second-order effects in (5) and (7) by the linear law of motion in $\mathbf{z}_t^{(2)}$

$$\mathbf{z}_{t+1}^{(2)} = \mathbf{A}^{(2)}\mathbf{z}_t^{(2)} + \mathbf{B}^{(2)}\boldsymbol{\xi}_{t+1}^{(2)} + \mathbf{c}^{(2)}, \tag{15}$$

and the law of motion in (9) as

$$\mathbf{y}_t^s = \mathbf{C}^{(2)}\mathbf{z}_t^{(2)} + \mathbf{d}^{(2)}. \tag{16}$$

The expressions for $\mathbf{A}^{(2)}$, $\mathbf{B}^{(2)}$, $\boldsymbol{\xi}_{t+1}^{(2)}$, $\mathbf{c}^{(2)}$, $\mathbf{C}^{(2)}$, and $\mathbf{d}^{(2)}$ are provided in Appendix A.3. Standard properties for the Kronecker product and block matrices imply that the system in (15) is stable with all eigenvalues of $\mathbf{A}^{(2)}$ having modulus less than one, provided the same holds for $\mathbf{h}_\mathbf{x}$; see Appendix A.4. This result might also be directly inferred from (5) and (7) because $\mathbf{x}_t^f$ is stable by assumption, $\mathbf{x}_t^s$ is constructed from a stable process, and the autoregressive part of $\mathbf{x}_t^s$ is stable. The system has finite unconditional second moments if the same holds for $\boldsymbol{\xi}_{t+1}^{(2)}$, which is equivalent to $\boldsymbol{\epsilon}_{t+1}$ having finite unconditional fourth moments; see Appendix A.5.[10] This further implies that explosive sample paths do not appear in the pruned state-space system (almost surely).

[10]These results also hold for models with deterministic and stochastic trends, provided trending variables are appropriately scaled (see King and Rebelo (1999)).

The next step is to find the expressions for the first and second unconditional moments. The innovations $\boldsymbol{\xi}_{t+1}^{(2)}$ are a function of $\mathbf{x}_t^f$, $\boldsymbol{\epsilon}_{t+1}$, and $\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}$, and we directly have that $\mathbb{E}\left[\boldsymbol{\xi}_{t+1}^{(2)}\right] = \mathbf{0}$. Hence, the unconditional mean of $\mathbf{z}_t^{(2)}$ is $\mathbb{E}\left[\mathbf{z}_t^{(2)}\right] = \left(\mathbf{I}_{2n_x+n_x^2} - \mathbf{A}^{(2)}\right)^{-1}\mathbf{c}^{(2)}$. To obtain some intuition for the determinants of the mean in the pruned state-space system, we explicitly compute some of the elements in $\mathbb{E}\left[\mathbf{z}_t^{(2)}\right]$. The mean of $\mathbf{x}_t^f$ is easily seen to be zero from (5). Equation (7) implies that the mean of $\mathbf{x}_t^s$ is

$$\mathbb{E}\left[\mathbf{x}_t^s\right] = \left(\mathbf{I} - \mathbf{h}_{\mathbf{x}}\right)^{-1}\left(\frac{1}{2}\mathbf{H}_{\mathbf{xx}}\mathbb{E}\left[\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right] + \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2\right). \tag{17}$$

Adding the mean for the first- and second-order effects, we then obtain the mean of the state variables in the pruned second-order approximation:

$$\mathbb{E}\left[\mathbf{x}_t^f + \mathbf{x}_t^s\right] = \mathbb{E}\left[\mathbf{x}_t^f\right] + \mathbb{E}\left[\mathbf{x}_t^s\right]. \tag{18}$$

Equations (17) and (18) show that the second-order effects $\mathbb{E}\left[\mathbf{x}_t^s\right]$ correct the mean of the first-order effects to adjust for uncertainty in the model. The adjustment comes from the second derivative of the perturbation parameter $\mathbf{h}_{\sigma\sigma}$ and the mean of $\mathbf{x}_t^f \otimes \mathbf{x}_t^f$. The latter can be computed from (5) and is given by $\mathbb{E}\left[\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right] = \left(\mathbf{I} - \mathbf{h}_{\mathbf{x}} \otimes \mathbf{h}_{\mathbf{x}}\right)^{-1}\left(\sigma\boldsymbol{\eta} \otimes \sigma\boldsymbol{\eta}\right) vec\left(\mathbf{I}_{n_e}\right)$.

Since $\mathbb{E}\left[\mathbf{x}_t^f\right] = 0$ and $\mathbb{E}\left[\mathbf{x}_t^s\right] \neq 0$, the mean value of the variables in a first-order approximation is their steady state, while the mean of the pruned second-order approximation is corrected by the second moment of the innovations. In other words, the mean of $\mathbf{x}_t$ implied by the pruned state-space system will, in most cases, differ from the steady state. This result is crucial because it shows that we cannot, in general, ignore the term $\mathbb{E}\left[\mathbf{x}_t^s\right]$ and simply use the steady state of the model to calibrate or estimate model parameters.

Let us now consider the unconditional second moments. Standard properties of a VAR(1) system imply that the variance-covariance matrix for $\mathbf{z}_t^{(2)}$ is given by

$$\mathbb{V}\left(\mathbf{z}_t^{(2)}\right) = \mathbf{A}^{(2)}\mathbb{V}\left(\mathbf{z}_t^{(2)}\right)\left(\mathbf{A}^{(2)}\right)' + \mathbf{B}^{(2)}\mathbb{V}\left(\boldsymbol{\xi}_t^{(2)}\right)\left(\mathbf{B}^{(2)}\right)',$$

because $\mathbf{z}_t^{(2)}$ and $\boldsymbol{\xi}_{t+1}^{(2)}$ are uncorrelated as $\boldsymbol{\epsilon}_{t+1}$ is independent across time. Appendix A.5 explains how to calculate $\mathbb{V}\left(\boldsymbol{\xi}_t^{(2)}\right)$. Once $\mathbb{V}\left(\boldsymbol{\xi}_t^{(2)}\right)$ is known, we solve for $\mathbb{V}\left(\mathbf{z}_t^{(2)}\right)$ by standard methods for discrete Lyapunov equations.

Our procedure for computing $\mathbb{V}\left(\mathbf{z}_t^{(2)}\right)$ differs slightly from the one in Kim, Kim, Schaumburg and Sims (2008). They suggest using a second-order approximation to $\mathbb{V}\left(\boldsymbol{\xi}_t^{(2)}\right)$ by letting the last n_x^2 elements in $\boldsymbol{\xi}_t^{(2)}$ be zero. This eliminates all third- and fourth-order terms related to $\boldsymbol{\epsilon}_{t+1}$ and seems inconsistent with the fact that $\mathbf{A}^{(2)} \otimes \mathbf{A}^{(2)}$ in $vec\left(\mathbf{A}^{(2)}\mathbb{V}\left(\mathbf{z}_t^{(2)}\right)\left(\mathbf{A}^{(2)}\right)'\right) = \left(\mathbf{A}^{(2)} \otimes \mathbf{A}^{(2)}\right) vec\left(\mathbb{V}\left(\mathbf{z}_t^{(2)}\right)\right)$ contains third- and fourth-order terms. We prefer to compute $\mathbb{V}\left(\boldsymbol{\xi}_t^{(2)}\right)$ without further approximations, implying that $\mathbb{V}\left(\mathbf{z}_t^{(2)}\right)$ corresponds to the sample moment in a long simulation of the pruned state-space system.

The variance of the combined first- and second-order effects for the state variables is obtained by taking the variance of $x_t^s + x_t^f$, i.e.

$$\mathbb{V}\left(\mathbf{x}_t^s + \mathbf{x}_t^f\right) = \mathbb{V}\left(\mathbf{x}_t^f\right) + \mathbb{V}\left(\mathbf{x}_t^s\right) + Cov\left(\mathbf{x}_t^f, \mathbf{x}_t^s\right) + Cov\left(\mathbf{x}_t^s, \mathbf{x}_t^f\right).$$

The auto-covariances for $\mathbf{z}_t^{(2)}$ are $Cov\left(\mathbf{z}_{t+l}^{(2)}, \mathbf{z}_t^{(2)}\right) = \left(\mathbf{A}^{(2)}\right)^l \mathbb{V}\left(\mathbf{z}_t^{(2)}\right)$ for $l = 1, 2, 3, \ldots$ because $\mathbf{z}_t^{(2)}$ and $\boldsymbol{\xi}_{t+l}^{(2)}$ are uncorrelated for $l = 1, 2, 3, ...$, given that $\boldsymbol{\epsilon}_{t+1}$ is independent across time.

The closed-form expressions for all corresponding unconditional moments related to $\mathbf{y}_t^s$ follow directly from the linear relationship between $\mathbf{y}_t^s$ and $\mathbf{z}_t^{(2)}$ in (16). That is,

$$\mathbb{E}\left[\mathbf{y}_t^s\right] = \mathbf{C}^{(2)}\mathbb{E}\left[\mathbf{z}_t^{(2)}\right] + \mathbf{d}^{(2)},\ \mathbb{V}\left[\mathbf{y}_t^s\right] = \mathbf{C}^{(2)}\mathbb{V}\left[\mathbf{z}_t\right]\left(\mathbf{C}^{(2)}\right)', \text{ and}$$

$$Cov\left(\mathbf{y}_{t+l}^s, \mathbf{y}_t^s\right) = \mathbf{C}^{(2)}Cov\left(\mathbf{z}_{t+l}^{(2)}, \mathbf{z}_t^{(2)}\right)\left(\mathbf{C}^{(2)}\right)' \quad \text{for } l = 1, 2, 3, ...$$

Finally, the representation in (15) and (16) makes it straightforward to derive additional statistical properties for the system. In particular, the pruned state-space system has finite unconditional third and fourth moments if the same holds for $\boldsymbol{\xi}_{t+1}^{(2)}$, which is equivalent to $\boldsymbol{\epsilon}_{t+1}$ having finite unconditional sixth and eighth moments; see Appendix A.6.

4.2 Third-Order Approximation

As we did for the second-order approximation, we start by deriving a more compact representation for the pruned state-space system than the one in Section 3.2. This is done based on the vector

$$\mathbf{z}_t^{(3)} \equiv \left[\begin{array}{cccccc} \left(\mathbf{x}_t^f\right)' & \left(\mathbf{x}_t^s\right)' & \left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right)' & \left(\mathbf{x}_t^{rd}\right)' & \left(\mathbf{x}_t^f \otimes \mathbf{x}_t^s\right)' & \left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f\right)' \end{array} \right]',$$

where the first part reproduces $\mathbf{z}_t^{(2)}$ and the last three components denote third-order effects. The law of motion for $\mathbf{x}_t^{rd}$ was derived in Section 3.2, and recursions for $\mathbf{x}_t^f \otimes \mathbf{x}_t^s$ and $\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f$ follow from (5) and (7). Hence, the law of motion for $\mathbf{x}_t^f$, $\mathbf{x}_t^s$, and $\mathbf{x}_t^{rd}$ in (5), (7), and (12), respectively, can be represented by the linear law of motion in $\mathbf{z}_t^{(3)}$

$$\mathbf{z}_{t+1}^{(3)} = \mathbf{A}^{(3)}\mathbf{z}_t^{(3)} + \mathbf{B}^{(3)}\boldsymbol{\xi}_{t+1}^{(3)} + \mathbf{c}^{(3)}. \tag{19}$$

We also have that the control variables are linear in $\mathbf{z}_t^{(3)}$ as

$$\mathbf{y}_t^{rd} = \mathbf{C}^{(3)}\mathbf{z}_t^{(3)} + \mathbf{d}^{(3)}. \tag{20}$$

The expressions for $\mathbf{A}^{(3)}$, $\mathbf{B}^{(3)}$, $\boldsymbol{\xi}_{t+1}^{(3)}$, $\mathbf{c}^{(3)}$, $\mathbf{C}^{(3)}$, and $\mathbf{d}^{(3)}$ are provided in Appendix A.7.

Appendix A.8 shows that the system in (19) is stable, with all eigenvalues of $\mathbf{A}^{(3)}$ having modulus less than one, provided the same holds for $\mathbf{h}_\mathbf{x}$. Building on the intuition from the second-order approximation, this result follows from the fact that the new component of the state vector $\mathbf{x}_t^{rd}$ is constructed from stable processes and its autoregressive component is also stable. The stability of $\mathbf{x}_t^{rd}$ relies on σ being treated as a variable in the pruned state-space system. If, instead, we had interpreted σ as a constant and included the term $\frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2\mathbf{x}_t^{rd}$ in the law of motion for $\mathbf{x}_{t+1}^{rd}$, then $\mathbf{x}_{t+1}^{rd}$ would have the autoregressive matrix $\mathbf{h}_\mathbf{x} + \frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2$, which may imply eigenvalues with modulus greater than one even when $\mathbf{h}_\mathbf{x}$ is stable. Moreover, the system in (19) and (20) has finite unconditional second moments if the same holds for $\boldsymbol{\xi}_{t+1}^{(3)}$. The latter is equivalent to $\boldsymbol{\epsilon}_{t+1}$ having finite unconditional sixth moments; see Appendix A.9.

The next step is to compute the first and second unconditional moments. The innovations $\boldsymbol{\xi}_{t+1}^{(3)}$ in (19) are a function of $\mathbf{x}_t^f$, $\mathbf{x}_t^s$, $\boldsymbol{\epsilon}_{t+1}$, $\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}$, and $\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}$. Thus, $\mathbb{E}\left[\boldsymbol{\xi}_{t+1}^{(3)}\right] = \mathbf{0}$ and

$\mathbb{E}\left[\mathbf{z}_t^{(3)}\right] = \left(\mathbf{I}_{3n_x+2n_x^2+n_x^3} - \mathbf{A}^{(3)}\right)^{-1}\mathbf{c}^{(3)}$. It is interesting to explore the value of $\mathbb{E}\left[\mathbf{x}_t^{rd}\right]$ as it may change the mean of the state variables. From (12), we immediately have

$$\mathbb{E}\left[\mathbf{x}_t^{rd}\right] = \left(\mathbf{I}_{n_x} - \mathbf{h}_{\mathbf{x}}\right)^{-1}\left(\mathbf{H}_{\mathbf{xx}}\mathbb{E}\left[\mathbf{x}_t^f \otimes \mathbf{x}_t^s\right] + \frac{1}{6}\mathbf{H}_{\mathbf{xxx}}\mathbb{E}\left[\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f\right] + \frac{1}{6}\mathbf{h}_{\sigma\sigma\sigma}\sigma^3\right),$$

and simple algebra gives $\mathbb{E}\left[\mathbf{x}_t^f \otimes \mathbf{x}_t^s\right] = \left(\mathbf{I}_{n_x^2} - (\mathbf{h}_{\mathbf{x}} \otimes \mathbf{h}_{\mathbf{x}})\right)^{-1}\left(\mathbf{h}_{\mathbf{x}} \otimes \frac{1}{2}\mathbf{H}_{\mathbf{xx}}\right)\mathbb{E}\left[\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f\right]$ and $\mathbb{E}\left[\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f\right] = \left(\mathbf{I}_{n_x^3} - (\mathbf{h}_{\mathbf{x}} \otimes \mathbf{h}_{\mathbf{x}} \otimes \mathbf{h}_{\mathbf{x}})\right)^{-1}(\sigma\boldsymbol{\eta} \otimes \boldsymbol{\sigma\eta} \otimes \boldsymbol{\sigma\eta})\,\mathbb{E}\left[\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}\right]$. Adding the mean for the first-, second- and third-order effects, we obtain $\mathbb{E}\left[\mathbf{x}_t^f\right] + \mathbb{E}\left[\mathbf{x}_t^s\right] + \mathbb{E}\left[\mathbf{x}_t^{rd}\right]$. If we next consider the standard case where all innovations have symmetric probability distributions, then $\mathbb{E}\left[\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}\right] = \mathbf{0}$, which in turn implies $\mathbb{E}\left[\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f\right] = \mathbf{0}$ and $\mathbb{E}\left[\mathbf{x}_t^f \otimes \mathbf{x}_t^s\right] = \mathbf{0}$. Furthermore, based on the results in Andreasen (2012), $\mathbf{h}_{\sigma\sigma\sigma}$ and $\mathbf{g}_{\sigma\sigma\sigma}$ are also zero when all innovations have symmetric probability distributions. Thus, $\mathbb{E}\left[\mathbf{x}_t^{rd}\right] = \mathbf{0}$ and the unconditional mean of the state vector is not further corrected by the third-order effects when all innovations have zero third moments. A similar property holds for the control variables because they are a linear function of $\mathbf{x}_t^{rd}$, $\mathbf{x}_t^f \otimes \mathbf{x}_t^s$, and $\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f$. This result is useful when calibrating or estimating DSGE models with symmetric probability distributions. On the other hand, if one or several innovations have non-symmetric probability distributions, then $\mathbf{h}_{\sigma\sigma\sigma}$ and $\mathbf{g}_{\sigma\sigma\sigma}$ may be non-zero and $\mathbb{E}\left[\mathbf{x}_t^{rd}\right] \neq \mathbf{0}$, implying that the unconditional mean has an additional uncertainty correction compared to a second-order approximation.

Let us now consider the unconditional second moments. The expression for the variance-covariance matrix of $\mathbf{z}_t^{(3)}$ is slightly more complicated than the one for $\mathbf{z}_t^{(2)}$ because $\mathbf{z}_t^{(3)}$ is correlated with $\boldsymbol{\xi}_{t+1}^{(3)}$. This correlation arises from terms of the form $\mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}$ in $\boldsymbol{\xi}_{t+1}^{(3)}$ which are correlated with elements in $\mathbf{z}_t^{(3)}$. Hence,

$$\begin{aligned}\mathbb{V}\left(\mathbf{z}_t^{(3)}\right) &= \mathbf{A}^{(3)}\mathbb{V}\left(\mathbf{z}_t^{(3)}\right)\left(\mathbf{A}^{(3)}\right)' + \mathbf{B}^{(3)}\mathbb{V}\left(\boldsymbol{\xi}_t^{(3)}\right)\left(\mathbf{B}^{(3)}\right)' \\ &\quad + \mathbf{A}^{(3)}Cov\left(\mathbf{z}_t^{(3)}, \boldsymbol{\xi}_{t+1}^{(3)}\right)\left(\mathbf{B}^{(3)}\right)' + \mathbf{B}^{(3)}Cov\left(\boldsymbol{\xi}_{t+1}^{(3)}, \mathbf{z}_t^{(3)}\right)\left(\mathbf{A}^{(3)}\right)'.\end{aligned}$$

The expressions for $\mathbb{V}\left(\boldsymbol{\xi}_t^{(3)}\right)$ and $Cov\left(\boldsymbol{\xi}_{t+1}^{(3)}, \mathbf{z}_t^{(3)}\right)$ are provided in Appendix A.9. The variance of the combined first-, second- and third-order effects for the state variables is given by

$$\begin{aligned}\mathbb{V}\left(\mathbf{x}_t^s+\mathbf{x}_t^f+\mathbf{x}_t^{rd}\right) &= \mathbb{V}\left(\mathbf{x}_t^f\right)+\mathbb{V}\left(\mathbf{x}_t^s\right)+\mathbb{V}\left(\mathbf{x}_t^{rd}\right)+Cov\left(\mathbf{x}_t^f,\mathbf{x}_t^s\right)+Cov\left(\mathbf{x}_t^f,\mathbf{x}_t^{rd}\right)\\ &+Cov\left(\mathbf{x}_t^s,\mathbf{x}_t^f\right)+Cov\left(\mathbf{x}_t^s,\mathbf{x}_t^{rd}\right)+Cov\left(\mathbf{x}_t^{rd},\mathbf{x}_t^f\right)+Cov\left(\mathbf{x}_t^{rd},\mathbf{x}_t^s\right).\end{aligned}$$

The auto-covariances for $\mathbf{z}_t^{(3)}$ are

$$Cov\left(\mathbf{z}_{t+s}^{(3)},\mathbf{z}_t^{(3)}\right)=\left(\mathbf{A}^{(3)}\right)^s\mathbb{V}\left[\mathbf{z}_t^{(3)}\right]+\sum_{j=0}^{s-1}\left(\mathbf{A}^{(3)}\right)^{s-1-j}\mathbf{B}^{(3)}Cov\left(\boldsymbol{\xi}_{t+1+j}^{(3)},\mathbf{z}_t^{(3)}\right)$$

for $s=1,2,3,...$

The closed-form expressions for all corresponding unconditional moments related to $\mathbf{y}_t^{rd}$ follow directly from the linear relationship between $\mathbf{y}_t^{rd}$ and $\mathbf{z}_t^{(3)}$ in (20) and are given by $\mathbb{E}\left[\mathbf{y}_t^{rd}\right]=\mathbf{C}^{(3)}\mathbb{E}\left[\mathbf{z}_t^{(3)}\right]+\mathbf{d}^{(3)}$, $\mathbb{V}\left[\mathbf{y}_t^{rd}\right]=\mathbf{C}^{(3)}\mathbb{V}\left[\mathbf{z}_t^{(3)}\right]\left(\mathbf{C}^{(3)}\right)'$, and

$$Cov\left(\mathbf{y}_{t+l}^{rd},\mathbf{y}_t^{rd}\right)=\mathbf{C}^{(3)}Cov\left(\mathbf{z}_{t+l}^{(3)},\mathbf{z}_t^{(3)}\right)\left(\mathbf{C}^{(3)}\right)' \quad \text{for } l=1,2,3,..$$

Finally, the representation in (19) and (20) of the pruned state-space system and the fact that $\boldsymbol{\xi}_{t+1}^{(3)}$ is a function of $\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}$ allow us to derive additional properties for the system. For instance, the pruned state-space system has finite unconditional third and fourth moments if the same holds for $\boldsymbol{\xi}_{t+1}^{(3)}$, which is equivalent to $\boldsymbol{\epsilon}_{t+1}$ having finite ninth and twelfth moments; see Appendix A.10.

5 Generalized Impulse Response Functions

Another fruitful way to study the properties of DSGE models is to look at their IRFs. For the first-order approximation, these functions have simple expressions where the effects of shocks are scalable, symmetric, and independent of the state of the economy. For higher-order approximations, no closed-form expressions currently exist for these functions and simulation is, therefore, required. This section shows that the pruned state-space system allows us to derive closed-form solutions for these functions and avoid the use of simulation.

We consider the generalized impulse response function (GIRF) proposed by Koop, Pesaran and

Potter (1996). The GIRF for any variable in the model **var** (either a state or control variable) in period $t+l$ following a disturbance to the ith shock of size ν_i in period $t+1$ is defined as

$$GIRF_{\mathbf{var}}\left(l,\nu_i,\mathbf{w}_t\right)=\mathbb{E}\left[\mathbf{var}_{t+l}|\mathbf{w}_t,\epsilon_{i,t+1}=\nu_i\right]-\mathbb{E}\left[\mathbf{var}_{t+l}|\mathbf{w}_t\right],$$

where $\mathbf{w}_t$ denotes the required state variables in period t. As we will see below, the content of $\mathbf{w}_t$ depends on the approximation order.[11] Using this definition, the GIRFs for the first-order effects have the simple and well-known expressions

$$GIRF_{\mathbf{x}^f}\left(l,\nu_i\right)=\mathbb{E}\left[\mathbf{x}^f_{t+l}|\mathbf{x}^f_t,\epsilon_{i,t+1}=\nu_i\right]-\mathbb{E}\left[\mathbf{x}^f_{t+l}|\mathbf{x}^f_t\right]=\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu} \tag{21}$$

and

$$GIRF_{\mathbf{y}^f}\left(l,\nu_i\right)=\mathbf{g}_{\mathbf{x}}GIRF_{\mathbf{x}^f}\left(l,\nu_i\right),$$

where $\boldsymbol{\nu}$ has dimension $n_\epsilon\times 1$ and contains the size of the disturbances in period $t+1$. For (21), we have $\boldsymbol{\nu}(i,1)=\nu_i$ and $\boldsymbol{\nu}(k,1)=0$ for $k\neq i$. Here, $GIRF_{\mathbf{x}^f}$ and $GIRF_{\mathbf{y}^f}$ are scalable, symmetric, and independent of the state of the economy because the state vector $\mathbf{x}^f_t$ enters symmetrically in the two conditional expectations for computing each of these GIRFs. Momentarily, we will see how the GIRFs for second- and third-order effects will not be scalable, symmetric, and independent of the state of the economy.

5.1 Second-Order Approximation

For the second-order effects $\mathbf{x}^s_t$, we have from (7) that

$$\mathbf{x}^s_{t+l}=\mathbf{h}^l_{\mathbf{x}}\mathbf{x}^s_t+\sum_{j=1}^{l-1}\mathbf{h}_{\mathbf{x}}^{l-1-j}\frac{1}{2}\mathbf{H}_{\mathbf{xx}}\left(\mathbf{x}^f_{t+j}\otimes\mathbf{x}^f_{t+j}\right)+\frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2\sum_{j=0}^{l-1}\mathbf{h}_{\mathbf{x}}^{l-1-j}. \tag{22}$$

The GIRF for $\mathbf{x}^f_t\otimes\mathbf{x}^f_t$ is derived in Appendix A.11, showing that

$$\begin{aligned}GIRF_{\mathbf{x}^f\otimes\mathbf{x}^f}\left(l,\nu_i,\mathbf{x}^f_t\right)&=\mathbf{h}^l_{\mathbf{x}}\mathbf{x}^f_t\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}+\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\mathbf{h}^l_{\mathbf{x}}\mathbf{x}^f_t\\&\quad+\left(\mathbf{h}_{\mathbf{x}}^{l-1}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\right)\left(\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\sigma\boldsymbol{\eta}\boldsymbol{\nu}+\boldsymbol{\Lambda}\right),\end{aligned} \tag{23}$$

[11] The expressions we derive below for the GIRFs may also be used for studying the joint effects of more than one disturbance to the economy. Further details are provided in the Online Appendix.

where

$$\mathbf{\Lambda} \equiv ((\sigma\boldsymbol{\eta}(\mathbf{I}-\mathbf{S}) \otimes \sigma\boldsymbol{\eta}(\mathbf{I}-\mathbf{S})) - (\sigma\boldsymbol{\eta} \otimes \sigma\boldsymbol{\eta}))\, vec(\mathbf{I})\,. \tag{24}$$

Here, $\mathbf{S}$ is an $n_\epsilon \times n_\epsilon$ diagonal matrix with $\mathbf{S}(i,i) = 1$ and $\mathbf{S}(k,k) = 0$ for $k \neq i$. Using this expression and (22), we get the GIRF for the second-order effects

$$GIRF_{\mathbf{x}^s}\left(l, \nu_i, \mathbf{x}_t^f\right) = \sum_{j=1}^{l-1} \mathbf{h}_{\mathbf{x}}^{l-1-j} \frac{1}{2} \mathbf{H}_{\mathbf{xx}} GIRF_{\mathbf{x}^f \otimes \mathbf{x}^f}\left(j, \nu_i, \mathbf{x}_t^f\right). \tag{25}$$

The expressions in (23) to (25) reveal three implications about the GIRF for the second-order effects. First, it is not scalable as $GIRF_{\mathbf{x}^f \otimes \mathbf{x}^f}\left(l, \tau \times \nu_i, \mathbf{x}_t^f\right) \neq \tau \times GIRF_{\mathbf{x}^f \otimes \mathbf{x}^f}\left(l, \nu_i, \mathbf{x}_t^f\right)$ for $\tau \in \mathbb{R}$. Second, the term $(\sigma\boldsymbol{\eta}\boldsymbol{\nu} \otimes \sigma\boldsymbol{\eta}\boldsymbol{\nu})$ means that the GIRF is not symmetric in positive and negative shocks. Third, it depends on the first-order effects of the state variables. Adding the GIRFs for the first- and second-order effects, we obtain the pruned GIRF for the state variables in a second-order approximation.

Finally, the pruned GIRF for the control variables is easily derived from (8) and previous results:

$$\begin{aligned} GIRF_{\mathbf{y}^s}\left(l, \nu_i, \mathbf{x}_t^f\right) &= \mathbf{g}_{\mathbf{x}}\left(GIRF_{\mathbf{x}^f}\left(l, \nu_i, \mathbf{x}_t^f\right) + GIRF_{\mathbf{x}^s}\left(l, \nu_i, \mathbf{x}_t^f\right)\right) \\ &+ \frac{1}{2}\mathbf{G}_{\mathbf{xx}} GIRF_{\mathbf{x}^f \otimes \mathbf{x}^f}\left(l, \nu_i, \mathbf{x}_t^f\right). \end{aligned}$$

Another interesting result from our analytical expressions relates to the IRFs in a linearized solution for a positive or negative one-standard-deviation shock computed at the steady state. As shown in Appendix A.12, these IRFs coincide with the GIRFs in a pruned second-order approximation because $GIRF_{\mathbf{x}^f \otimes \mathbf{x}^f}\left(j, \nu_i, \mathbf{x}_t^f\right) = \mathbf{0}$, implying that these IRFs in a linearized solution are actually second-order accurate.

5.2 Third-Order Approximation

Using (12), we first note that for the third-order effects $\mathbf{x}_t^{rd}$

$$\begin{aligned} \mathbf{x}_{t+l}^{rd} &= \mathbf{h}_{\mathbf{x}}^l \mathbf{x}_t^{rd} + \sum_{j=0}^{l-1} \mathbf{h}_{\mathbf{x}}^{l-1-j}\left[\mathbf{H}_{\mathbf{xx}}\left(\mathbf{x}_{t+j}^f \otimes \mathbf{x}_{t+j}^s\right) + \frac{1}{6}\mathbf{H}_{\mathbf{xxx}}\left(\mathbf{x}_{t+j}^f \otimes \mathbf{x}_{t+j}^f \otimes \mathbf{x}_{t+j}^f\right)\right] \\ &+ \sum_{j=0}^{l-1} \mathbf{h}_{\mathbf{x}}^{l-1-j}\left[\frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2 \mathbf{x}_{t+j}^f + \frac{1}{6}\mathbf{h}_{\sigma\sigma\sigma}\sigma^3\right]. \end{aligned}$$

Simple algebra implies

$$
\begin{aligned}
GIRF_{\mathbf{x}^{rd}}\left(l,\nu_i,\left(\mathbf{x}_t^f,\mathbf{x}_t^s\right)\right) &= \sum_{j=1}^{l-1}\mathbf{h}_{\mathbf{x}}^{l-1-j}\mathbf{H}_{\mathbf{xx}}GIRF_{\mathbf{x}^f\otimes\mathbf{x}^s}\left(j,\nu_i,\left(\mathbf{x}_t^f,\mathbf{x}_t^s\right)\right)\\
&+\sum_{j=1}^{l-1}\mathbf{h}_{\mathbf{x}}^{l-1-j}\frac{1}{6}\mathbf{H}_{\mathbf{xxx}}GIRF_{\mathbf{x}^f\otimes\mathbf{x}^f\otimes\mathbf{x}^f}\left(j,\nu_i,\mathbf{x}_t^f\right)\\
&+\sum_{j=1}^{l-1}\mathbf{h}_{\mathbf{x}}^{l-1-j}\frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2 GIRF_{\mathbf{x}^f}\left(j,\nu_i\right).
\end{aligned}
$$

All terms are known except for $GIRF_{\mathbf{x}^f\otimes\mathbf{x}^s}\left(j,\nu_i,\left(\mathbf{x}_t^f,\mathbf{x}_t^s\right)\right)$ and $GIRF_{\mathbf{x}^f\otimes\mathbf{x}^f\otimes\mathbf{x}^f}\left(j,\nu_i,\mathbf{x}_t^f\right)$, which are derived in Appendix A.13. As was the case for the second-order effect, the GIRF for the third-order effect is not scalable, not symmetric, and depends on the first-order effects of the state variables $\mathbf{x}_t^f$. In addition, the GIRF for the third-order effects also depends on $\mathbf{x}_t^s$. Adding the GIRF for the first-, second-, and third-order effects, we obtain the pruned GIRF for the state variables in a third-order approximation.

The pruned GIRF for the control variables in a third-order approximation is

$$
\begin{aligned}
GIRF_{\mathbf{y}^{rd}}\left(l,\nu_i,\left(\mathbf{x}_t^f,\mathbf{x}_t^s\right)\right) &= \mathbf{g}_{\mathbf{x}}\left(GIRF_{\mathbf{x}^f}\left(l,\nu_i\right)+GIRF_{\mathbf{x}^s}\left(l,\nu_i,\mathbf{x}_t^f\right)\right)+\mathbf{g}_{\mathbf{x}}GIRF_{\mathbf{x}^{rd}}\left(l,\nu_i,\left(\mathbf{x}_t^f,\mathbf{x}_t^s\right)\right)\\
&+\frac{1}{2}\mathbf{G}_{\mathbf{xx}}\left(GIRF_{\mathbf{x}^f\otimes\mathbf{x}^f}\left(l,\nu_i,\mathbf{x}_t^f\right)+2GIRF_{\mathbf{x}^f\otimes\mathbf{x}^s}\left(l,\nu_i,\left(\mathbf{x}_t^f,\mathbf{x}_t^s\right)\right)\right)\\
&+\frac{1}{6}\mathbf{G}_{\mathbf{xxx}}GIRF_{\mathbf{x}^f\otimes\mathbf{x}^f\otimes\mathbf{x}^f}\left(l,\nu_i,\mathbf{x}_t^f\right)\\
&+\frac{3}{6}\mathbf{g}_{\sigma\sigma\mathbf{x}}\sigma^2 GIRF_{\mathbf{x}^f}\left(l,\nu_i\right),
\end{aligned}
\tag{26}
$$

where all terms are known.

5.3 Conditional Impulse Response Functions

Given that the expressions for the GIRFs in Sections 5.1 and 5.2 depend on the values of the state variables, we can use them to analyze how the responses to shocks depend on the business cycle. For example, in a simple stochastic neoclassical growth model, the economy may respond differently to a positive technological shock when capital is high than when it is low. The challenge is that in most DSGE models, the values of the state variables are unobserved. Hence, it may be challenging to specify relevant state values, except for the obvious benchmark as given by the unconditional mean. We suggest overcoming this problem by conditioning the GIRFs on some set of observables, such as the economy being in a recession (i.e., negative output growth). More concretely, consider a conditional GIRF of the form

$$GIRF_{\mathbf{var}}(l,\nu_i,A) = \int 1_A(\mathbf{w}_t)\, f(\mathbf{w}_t)\, GIRF_{\mathbf{var}}(l,\nu_i,\mathbf{w}_t)\, d\mathbf{w}_t \tag{27}$$

where A is a set defined by the criteria in the observables, $1_A(\mathbf{w}_t)$ is an indicator function, and $f(\mathbf{w}_t)$ is the unconditional density of $\mathbf{w}_t$. The integral in (27) can be evaluated by Monte Carlo integration, where draws of the states are obtained from a long simulated sample path of the pruned state-space system. The great advantage of a conditional GIRF is that it is defined on an observed set A and is therefore directly observable, in contrast to the GIRFs provided in Sections 5.1 and 5.2.

An example illustrates how we address the challenge mentioned two paragraphs ago. Imagine we are working with a simple stochastic neoclassical growth model and we want to calculate the GIRF of $\mathbf{y}_t^{rd}$ conditional on output growth being above a given threshold, say, 2% annualized. Then A is the set of $\left(\mathbf{x}_t^f, \mathbf{x}_t^s\right)$, which implies that output growth is above this annualized 2%.

6 Accuracy of Pruning

This section analyzes how pruning affects the accuracy of the approximated solution. We start with the following proposition, which studies the approximation errors when the perturbation parameter tends to zero.

Proposition 1 *For $\sigma \rightarrow 0$, the errors in $\mathbf{x}_t^f + \mathbf{x}_t^s$ and $\mathbf{y}_t^s$ are of third order, whereas the errors in $\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd}$ and $\mathbf{y}_t^{rd}$ are of fourth order.*

The proof of this proposition is provided in Appendix A.14. The consequence of this result is that, for a given approximation order and $\sigma \rightarrow 0$, the errors in the pruned and unpruned state-space systems are of the same order. When we account for uncertainty in the model by letting $\sigma = 1$, the stability of the pruned state-space system ensures that approximation errors do not accumulate over time. A similar convenient property does not necessarily hold in the unpruned state-space system, which therefore may generate explosive sample paths.

To obtain further insights into the accuracy of pruning under uncertainty, one would have to consider a particular DSGE model and study the accuracy of a pruned and an unpruned state-space system. For small models, this can be done by comparing simulated sample paths to a highly accurate projection approximation as in Lan and Meyer-Gohde (2013*a*). For larger models where the projection method or other global approximation methods are computationally infeasible, accuracy may be explored based on Euler equation errors, as in an earlier version of this paper (see Andreasen, Fernández-Villaverde and Rubio-Ramírez (2013)). Given the analyzed models, the two aforementioned papers show that our pruning scheme does not worsen accuracy (and often it improves it) when compared to the unpruned state-space system. However, these findings are model-specific.

Nevertheless, we can offer some intuition of why we found in previous versions of this paper that the Euler equation errors of the pruned approximation along the simulations were smaller than those from the unpruned one. Unpruned approximations are subject to what we call microbursts of instability. Often, the simulations are hit by relatively large innovations. These innovations push the simulation toward an explosive path. At the same time, it is also often the case that after a few periods, a large innovation of opposite sign sends the simulation back into a stable path. During these periods of transitory explosive paths (our microbursts of instability), the Euler equation errors of the unpruned approximation are poor. In comparison, pruned approximations are not subject to these microbursts. We will often have small microbursts of instability that do not reach the threshold and are kept in the simulation while triggering poor accuracy. This behavior is documented in Appendix A.15, where we show that for a simple stochastic neoclassical growth model, the pruned solution does better when we are farther away from the steady state.

In general, regardless of whether pruning improves accuracy around the steady state or not, one can also adopt the view that pruning is a simple and transparent way of eliminating explosive sample paths. One may, therefore, argue that the cautious approach is to prune perturbation approximations. Unpruned perturbation approximations may also be useful for models where explosive sample paths rarely appear, but these unpruned approximations simply may be inapplicable to many models of interest that frequently generate explosive sample paths, including the New Keynesian model presented below in Section 8.

7 Econometric Implications of the Pruning Method

Our results in Sections 4 and 5 allow us to implement standard moment matching methods in non-linearly approximated DSGE models. For approximations up to third order, these methods include GMM estimation (Hansen (1982)) based on first and second unconditional moments. More generally, we could also estimate DSGE models by matching autocorrelation functions, the spectral density, or other functions of interest. Our results are also useful for a Bayesian researcher. The work by Kim (2002) shows how to build a limited information likelihood function from optimal GMM estimation. Equipped with priors, we may then carry out a Bayesian analysis, where the asymptotic distribution for the posterior equals the limiting distribution of GMM.

Another possibility is to match model-implied GIRFs to their empirical counterparts. This approach was popularized by Christiano, Eichenbaum and Evans (2005) for a linearized model. However, for non-linear approximations, we need to move beyond a VAR to document our empirical GIRFs because the linear structure of a VAR can only produce IRFs that are scalable, symmetric, and independent of the state of the economy. The methods proposed by Jorda (2005) and Matthes and Barnichon (2014) are, thus, more appropriate for DSGE models solved non-linearly. Given that the GIRFs for non-linear approximations depend on unobserved state variables, some care is needed to ensure that the empirical and model-implied GIRFs are comparable along this dimension, i.e., that they are evaluated at the same state values. A natural possibility is to compute model-implied GIRFs at the unconditional mean of the states and compare them to the empirical GIRFs at the sample mean. An alternative is to use the conditional GIRF in (27) to consider empirical and model-implied GIRFs on some criteria for the observables, such as i) recessions vs. expansions, ii) high vs. low inflation, iii) high vs. low conditional volatility of output, etc.

A third way to obtain the empirical GIRFs is to derive these functions from an estimated auxiliary model consistent with the pruned state-space system, as in Aruoba, Bocola and Schorfheide (2013) for the scalar case. Our work provides the theoretical foundation for constructing the auxiliary model consistent with the pruned state-space system in the multivariate case and beyond a second-order approximation.

In relation to GMM estimation and IRF matching, our results enable researchers to determine the stochastic specification of the structural innovations semi-parametrically, i.e., by only estimating the moments of ε_t without assuming a given probability distribution. The ability to identify moments of ε_t, and the DSGE model in general may be examined using the procedure from linearized models in Iskrev (2010) on the pruned state-space system, as in Mutschler (2015) for a second-order approximation.

If we want to use higher-order moments such as skewness and kurtosis in the estimation, then simulations are generally needed. Although it is possible to compute closed-form expressions for skewness and kurtosis in DSGE models when the pruning method is applied, the memory requirement for such computations is extremely onerous and only applicable to small models with a few state variables.[12] But even if we use simulation, our analysis provides a foundation for SMM following Duffie and Singleton (1993) and indirect inference as considered in Smith (1993), Dridi, Guay and Renault (2007), and Creel and Kristensen (2011), among others.[13] This is because pruning ensures that the model-implied processes are stationary (possibly following a transformation), as required for the limiting distribution of these simulation-based estimators.[14]

8 An Application

We now present an empirical application to illustrate the GMM estimation methodology that our paper makes available. We also show results for the GIRFs and the conditional GIRFs shown in Section 5. We focus on a New Keynesian model with two novel features. First, we introduce a financial intermediary that trades short- and long-term government bonds. This intermediary generates a wedge between the policy rate set up by the monetary authority and the interest rate faced by private agents in the economy. This wedge arises in our model from time variation in the conditional second moments of the stochastic discount factor, whereas this premium in models with a banking sector is often motivated by steady-state frictions (see Bernanke, Gertler and Gilchrist (1999) or Gertler and Karadi (2011) among others). Thus, our model combines the macro-finance literature focusing on stochastic discount factors with the recent work on financial intermediation in DSGE models. Our second innovation is to consider a central bank that sets the policy rate

[12] Building on our work, Mutschler (2015) provides closed-form expressions for skewness and kurtosis for second-order approximations to DSGE models.

[13] See also Ruge-Murcia (2012) for a Monte Carlo study and application of SMM based on the neoclassical growth model solved up to third order.

[14] See Peralta-Alva and Santos (2012) for a summary of the literature on the relation between estimation methods and numerical errors in simulation.

not only based on the inflation and output gap, but also on a measure of term premia. This extension is motivated by the recent financial crisis, where several central banks engaged in large asset purchases to stimulate the economy by affecting term premia (see Gagnon, Raskin, Rernache and Sack (2011) and Joyce, Lasaosa, Stevens and Tong (2011)). In summary, our model displays two feedback effects from long-term bonds to the real economy: i) a wedge between the policy rate and the interest rate faced by the households and ii) a policy rate that depends on a measure of term premia. As we will show below, each of these feedback effects helps our model to overcome the counterintuitive result of Tallarini (2000) that real allocations are essentially unaffected by the amount of risk in the economy. Consequently, our model generates a much richer environment for monetary policy than the standard New Keynesian model. These features make our application of interest on its own beyond illustrating our new estimation methodology.

We proceed by outlining the model in Sections 8.1 to 8.5, describing our solution and estimation method in Sections 8.6 and 8.7, and presenting estimation results in Sections 8.8 to 8.12. We finish by showing, in Section 8.13, some of the GIRFs and conditional GIRFs of the model.

8.1 Households

We consider a representative household with recursive preferences as in Epstein and Zin (1989) and Weil (1990). Using the convenient formulation proposed by Rudebusch and Swanson (2012), the value function V_t of the household can be written as

$$V_t \equiv \begin{cases} u_t + \beta \left(\mathbb{E}_t \left[V_{t+1}^{1-\phi_3} \right] \right)^{\frac{1}{1-\phi_3}} & \text{if } u_t > 0 \text{ for all } t \\ u_t - \beta \left(\mathbb{E}_t \left[(-V_{t+1})^{1-\phi_3} \right] \right)^{\frac{1}{1-\phi_3}} & \text{if } u_t < 0 \text{ for all } t \end{cases}, \tag{28}$$

where $\mathbb{E}_t$ is the conditional expectation given information in period t and $\beta \in (0,1)$ is the subjective discount factor. For higher values of $\phi_3 \in \mathbb{R} \setminus \{1\}$, these preferences imply higher levels of risk aversion if the utility kernel u_t is always positive, and vice versa for $u_t < 0$. The main benefit of the Epstein-Zin-Weil preferences is to disentangle risk aversion from the intertemporal elasticity of substitution (IES) when $\phi_3 \neq 0$; otherwise (28) simplifies to standard expected utility.

We let the utility kernel display separability between consumption c_t and hours worked h_t

$$u_t \equiv \frac{d_t}{1-\phi_2}\left(\left(\frac{c_t - bc_{t-1}}{z_t^*}\right)^{1-\phi_2} - 1\right) + \phi_0 \frac{(1-h_t)^{1-\phi_1}}{1-\phi_1}, \tag{29}$$

where b controls the degree of internal habit formation.[15] The variable $d_t \equiv \exp\{\sigma_d \epsilon_{d,t}\}$ with $\epsilon_{d,t} \sim \mathcal{NID}(0,1)$ introduces preference shocks, where we omit the traditional AR(1) term to ensure that variations in long-term interest rates and term premia in our model are explained by consumption dynamics and not by persistent preference shocks. As in An and Schorfheide (2007), utility from habit-adjusted consumption in (29) is expressed relative to the deterministic trend in the economy z_t^* to guarantee the existence of a balanced growth path. We simultaneously include habit formation and Epstein-Zin-Weil preferences because the literature has shown that the New Keynesian model needs both features to jointly match various macro and financial moments (see Hordahl, Tristani and Vestin (2008) and Binsbergen, Fernandez-Villaverde, Koijen and Rubio-Ramirez (2012)).

The budget constraint at time t reads:

$$c_t + \frac{i_t}{\Upsilon_t} + b_t + T_t = w_t h_t + r_t^k k_t + \frac{b_{t-1}\exp\left\{r_{t-1}^b\right\}}{\pi_t} + div_t^h. \tag{30}$$

Resources are spent on consumption, investment i_t, a one-period deposit b_t in the financial intermediary at the net nominal risk-free deposit rate r_t^b, and a lump-sum tax T_t. The variable Υ_t denotes a deterministic trend in the real relative price of investment: $\log \Upsilon_{t+1} = \log \Upsilon_t + \log \mu_{\Upsilon,ss}$. Letting w_t denote the real wage and r_t^k the real price of capital k_t, resources consist of real labor income $w_t h_t$, real income from capital services sold to firms $r_t^k k_t$, real returns from deposits in the previous period, and firm dividends to households div_t^h. Here, $\pi_t \equiv P_t/P_{t-1}$ is gross inflation.

The law of motion for k_t is

$$k_{t+1} = (1-\delta)\,k_t + i_t - \frac{\kappa}{2}\left(\frac{i_i}{k_t} - \psi\right)^2 k_t, \tag{31}$$

where $\kappa \geq 0$ introduces capital adjustment costs as in Jermann (1998). The constant ψ ensures

[15] The constant $-1/(1-\phi_2)$ ensures a stable level of u_t and V_t in the steady state when ϕ_2 is close to one. As shown in Table 1, for all estimated models, the steady-state value of u_t, u_{ss}, is substantially below zero.

that these adjustment costs are zero along the balanced growth path of the economy.

8.2 The Financial Intermediary

As mentioned above, the representative household makes one-period deposits b_t in a perfectly competitive financial intermediary, which invests deposits in short- and long-term government bonds. The household may also overdraw this deposit, i.e., b_t can be negative, in which case the financial intermediary shorts these bonds. The short-term bond, for simplicity, is assumed to be the one-period bond, whereas the maturity of the long-term bond is denoted by $L > 1$. In our implementation, we set L to reflect the 10-year interest rate, but other maturities may be considered.

The behavior of the financial intermediary is solely determined by the deposit rate r_t^b. To state its expression, let the ex ante holding period return on the kth bond be

$$hr_{t,k} \equiv \mathbb{E}_t \left[\log P_{t+1,k-1} - \log P_{t,k} \right], \tag{32}$$

where $P_{t,k}$ is the nominal price in period t of a zero-coupon bond maturing in period $t+k$. The excess holding period return is then $xhr_{t,k} \equiv hr_{t,k} - r_t$, where r_t is the one-period nominal policy rate set by the central bank. We then assume that the deposit rate is equal to the ex ante holding period return on the invested bond portfolio, i.e.,

$$r_t^b \equiv (1-\omega) \times hr_{t,1} + \omega \times hr_{t,L} = r_t + \omega \times xhr_{t,L} \tag{33}$$

because $hr_{t,1} = r_t$ and $xhr_{t,1} = 0$. Here, $\omega \in [0,1]$ denotes the fraction invested by the financial intermediary in the long-term government bond. The value of ω is determined by factors exogenous to the model. For example, financial regulation forces many mutual funds to keep large shares of their bonds in short maturities, regardless of their preferred investment strategies. Endogenizing this and other factors determining ω is well beyond the scope of this paper. Nevertheless, we will treat ω as a free parameter in our estimation procedure and infer the average portfolio weight from our model.

To clarify the behavior of this financial intermediary, suppose for a moment that $\omega = 0$. In this case, the financial intermediary only holds the one-period government bond and (33) simplifies to

$r_t^b = r_t$. Thus, our framework recovers the standard specification considered in most New Keynesian models where the deposit rate equals the one-period policy rate set by the central bank.

Another possibility is to assume that $\omega > 0$. This introduces a feedback effect from long-term government bonds to the real economy as the excess holding period return affects r_t^b and the household's consumption decision. For instance, an increase in $xhr_{t,L}$ due to a higher term premium during a recession will tend to increase the deposit rate and encourage the household to postpone consumption. Given that $xhr_{t,L}$ is non-zero due to uncertainty, this feedback effect from long-term government bonds to the real economy operates through a precautionary saving channel. We will exploit this insight below to derive an efficient perturbation solution to our model.

A careful inspection of our framework reveals that it is related to the risk-premium shocks in Smets and Wouters (2007), where an *exogenous* shock drives a wedge between the policy rate and the interest rate faced by the households. When (33) is substituted into the consumption Euler equation, i.e., $\mathbb{E}_t \left[\beta\lambda_{t+1} \exp\left\{r_t^b\right\} / \pi_{t+1}\right] = \lambda_t$ with λ_t, denoting the marginal utility of habit-adjusted consumption, we obtain a similar wedge, except that this wedge is *endogenously* generated within our model. Note also that, if we were to follow Smets and Wouters (2007) and use a standard log-linear approximation to our model, then $xhr_{t,L} = 0$ for all t, implying that (33) would reduce to the standard specification where $r_t^b = r_t$, even when $\omega > 0$.

Having outlined how the deposit rate is determined, we next describe how the financial intermediary prices government bonds. Given that the financial intermediary is owned by the households and, therefore, acts in their interest, we determine the price of these bonds by the stochastic discount factor of the representative household. That is,

$$P_{t,k} = \mathbb{E}_t \left[\beta \frac{\lambda_{t+1}}{\lambda_t} \frac{1}{\pi_{t+1}} P_{t+1,k-1}\right], \tag{34}$$

for $k = 2, 3..., \mathcal{K}$ with $P_{t,1} = \exp\{-r_t\}$. The nominal yield curve with continuous compounding is then given by $r_{t,k} = -\frac{1}{k} \log P_{t,k}$ for $k = 2, 3, ..., \mathcal{K}$.

8.3 Firms

A perfectly competitive representative firm produces final output y_t by aggregating a continuum of intermediate goods $y_{i,t}$ using the production function $y_t = \left(\int_0^1 y_{i,t}^{\frac{\eta-1}{\eta}} di\right)^{\frac{\eta}{\eta-1}}$ with $\eta > 1$. This generates the demand function $y_{i,t} = \left(\frac{P_{i,t}}{P_t}\right)^{-\eta} y_t$, with aggregate price level $P_t \equiv \left[\int_0^1 P_{i,t}^{1-\eta} di\right]^{\frac{1}{1-\eta}}$.

The intermediate good i is produced by a monopolistic competitor using the production function $y_{i,t} = a_t k_{i,t}^{\theta} (z_t h_{i,t})^{1-\theta}$. Here, z_t is a deterministic trend that follows $\log z_{t+1} = \log z_t + \log \mu_{z,ss}$, and $\log a_{t+1} = \rho_a \log a_t + \sigma_a \epsilon_{a,t+1}$ where $\epsilon_{a,t} \sim \mathcal{NID}(0,1)$. As in Altig, Christiano, Eichenbaum and Linde (2011), we define $z_t^* \equiv \Upsilon_t^{\frac{\theta}{1-\theta}} z_t$, which denotes the technological trend in the economy.

The intermediate firms maximize the net present value of real profit with respect to capital, labor, and prices given a nominal rigidity. We consider price-setting à la Calvo (1983), where contracts expire with probability $1-\alpha$ in each period. Whenever a contract expires, firms set their optimal nominal prices, which otherwise are equal to past prices, i.e., $P_{i,t} = P_{i,t-1}$.

8.4 Monetary and Fiscal Policy

A central bank sets the policy rate r_t based on a desire to stabilize the inflation gap $\log(\pi_t/\pi_{ss})$ and the output gap $\log(y_t/(z_t^* Y_{ss}))$, subject to smoothing changes in r_t. Here, π_{ss} refers to steady-state inflation. As in Justiniano and Primiceri (2008) and Rudebusch and Swanson (2012), the output gap is measured in deviation from the deterministic trend in output, which equals z_t^* times production in the normalized steady state Y_{ss}. As we argued above, with a financial intermediary investing in long-term government bonds, the deposit rate offered to households is no longer fully determined by the central bank's policy rate due to changes in $xhr_{t,L}$. Thus, the central bank may find it useful also to account for variability in $xhr_{t,L}$ when setting its policy rate. For instance, term premia typically increase during recessions and this generates upward pressure on $xhr_{t,L}$ and the deposit rate within our framework. Then, a central bank may consider a larger reduction in the policy rate than required with $xhr_{t,L} = 0$ to offset the negative impact from higher term premia on economic activity. Also, the central bank may provide different policy responses to shocks that create the same inflation and output gaps on impact, but affect $xhr_{t,L}$ and term premia asymmetrically, for instance, because the shocks differ in their persistence. More concretely, we postulate that monetary policy follows an augmented Taylor rule of the form

$$\begin{aligned} r_t &= (1-\rho_r) r_{ss} + \rho_r r_{t-1} + (1-\rho_r)\left(\beta_\pi \log\left(\frac{\pi_t}{\pi_{ss}}\right) + \beta_y \log\left(\frac{y_t}{z_t^* Y_{ss}}\right)\right) \\ &\quad + (1-\rho_r)\beta_{xhr}\left(xhr_{t,L} - \mathbb{E}[xhr_{t,L}]\right) \end{aligned} \tag{35}$$

where we omit monetary policy shocks as the literature has documented that they have a tiny effect on term premia.[16]

When implementing (35), we approximate $\mathbb{E}[xhr_{t,L}]$ by $\mathbb{E}_t\left[(1-\gamma)\sum_{l=0}^{\infty}\gamma^l xhr_{t+l,L}\right] \equiv X_{t,L}$ with $\gamma = 0.9999$, as it has the convenient representation $X_{t,L} = (1-\gamma)\,xhr_{t,L} + \gamma\mathbb{E}_t[X_{t+1,L}]$. In contrast, the steady-state value of $xhr_{t,L}$ equals zero and is a poor approximation of $\mathbb{E}[xhr_{t,L}] \neq 0$. Finally, note that if we were to solve our model by a log-linearization, then $xhr_{t,L} = 0$ for all t and (35) would reduce to the standard Taylor rule even if $\beta_{xhr} \neq 0$.

Government consumption $g_t \equiv G_t z_t^*$ grows with the economy as in Rudebusch and Swanson (2012), where

$$\log\left(\frac{G_{t+1}}{G_{ss}}\right) = \rho_G \log\left(\frac{G_t}{G_{ss}}\right) + \sigma_G \epsilon_{G,t+1}$$

and $\epsilon_{G,t+1} \sim \mathcal{NID}(0,1)$. Government consumption and the interest on government debt are paid with lump-sum taxes. Given that a version of Ricardian equivalence holds in our economy, we do not need to specify the timing of these taxes and simply write the resource constraint of the economy as $y_t = c_t + i_t\Upsilon_t^{-1} + g_t$.

8.5 Model Aggregation

The aggregated resource constraint in the goods market is $a_t k_t^{\theta}(z_t h_t)^{1-\theta} = y_t s_{t+1}$, where s_t is the price dispersion index. The dynamic of this endogenous state variable is

$$s_{t+1} = (1-\alpha)\,\tilde{p}_t^{-\eta} + \alpha\pi_t^{\eta} s_t, \tag{36}$$

where $\tilde{p}_t \equiv \tilde{P}_t/P_t$ and $\tilde{P}_t$ denotes the optimal nominal price in period t. The relation between inflation and the newly optimized prices is

$$1 = (1-\alpha)\,\tilde{p}_t^{1-\eta} + \alpha\pi_t^{\eta-1}. \tag{37}$$

Also, household deposits are zero in equilibrium (their net assets are in terms of capital investment), implying that net profit by the financial intermediary is also zero. Thus, the net worth of the

[16] This is illustrated in Rudebusch and Swanson (2012), who also show that it is the systematic part of monetary policy that has a large impact on term premia. We will find a similar result below.

financial intermediary does not change across time and may be ignored when solving the model.

There is an alternative, yet fully equivalent, representation of our model with complete markets, which is described in Appendix A.17. In that representation of the model, the households do not need to rely on the financial intermediary to invest their savings in government bonds, but the Taylor rule depends instead on $\omega * xhr_{t,L}$ and $\omega * \rho_r * xhr_{t-1,L}$.

8.6 An Efficient Perturbation Approximation

To solve the model, we first induce stationarity by eliminating trending variables with appropriate transformations; see Appendix A.16. The desired policy functions that characterize the equilibrium dynamics of the model are then obtained by employing a third-order perturbation approximation. We require at least a third-order approximation to generate variation in $xhr_{t,L}$ and capture the feedback effects from long-term government bonds to the real economy. Our quarterly model with L set to reflect the 10-year interest rate has seven state variables and 54 control variables.[17]

The standard approach in the literature to efficiently compute a higher-order perturbation of DSGE models with a yield curve exploits the fact that bond prices beyond the policy rate typically do not affect allocations and prices (i.e., consumption, inflation, etc.). Taking advantage of that property, these models are approximated by a two-step procedure, where the first step solves the model *without* bond prices exceeding one period, after which, in a second step, all remaining bond prices are computed recursively based on (34). This two-step procedure reduces the size of the simultaneous equation systems to be solved and, with it, the computation burden of the approximation (see Hordahl, Tristani and Vestin (2008), Binsbergen, Fernandez-Villaverde, Koijen and Rubio-Ramirez (2012), and Andreasen and Zabczyk (2015)).

We cannot apply this two-step procedure to our model when $\omega > 0$ or $\beta_{xhr} \neq 0$ because the long-term bond price affects the deposit rate r_t^b and the policy rate through $xhr_{t,L}$ and, hence, all allocations and prices. Fortunately, the terms associated with the perfect foresight solution of our model -i.e., $(\mathbf{g_x}, \mathbf{G_{xx}}, \mathbf{G_{xxx}})$ and $(\mathbf{h_x}, \mathbf{H_{xx}}, \mathbf{H_{xxx}})$- may be found with the standard two-step procedure even when $\omega > 0$ or $\beta_{xhr} \neq 0$ because $xhr_{t,L}$ is equal to zero under perfect foresight. Once we have computed these terms, we only need to find the derivatives involving the perturbation

[17] The relatively large number of control variables is needed to compute all bond prices within the 10-year maturity range.

parameter σ using the full model. The whole three-step procedure is formally described in Appendix A.18 and constitutes a new numerical contribution to the literature. Accordingly, our three-step procedure allows us to compute a third-order solution to our model with feedback effects in just 3.7 seconds, whereas it takes 6.2 seconds when using the standard one-step perturbation algorithm of Binning (2013). This improvement in computational speed of more than 40% greatly facilitates the estimation, as the perturbation approximation must be computed for many different parameter values.[18]

8.7 Data and Moments for GMM

We employ the following quarterly time series to estimate our model: i) consumption growth Δc_t, ii) investment growth Δi_t, iii) inflation π_t, iv) the 1-quarter nominal interest rate r_t, v) the 10-year nominal interest rate $r_{t,40}$, vi) the 10-year ex post excess holding period return $xhr_{t,40} \equiv \log(P_{t,39}/P_{t-1,40}) - r_{t-1}$, vii) the log ratio of government spending to GDP $\log(g_t/y_t)$, and viii) the log of hours $\log h_t$. The presence of a short- and long-term interest rate captures the slope of the yield curve, whereas the excess holding period return is included as a noisy proxy for the 10-year term premium. All series are stored in $\mathbf{data}_t$ with dimension 8×1. Our sample goes from 1961.Q3 to 2007.Q4. The end date is set to avoid the complications created by the zero lower bound of the nominal interest rate. See Appendix A.19 for a description of the data series.

We want to explore whether our model can match the mean, the variance, the contemporaneous covariances, and the persistence in the data. Hence, we let

$$\mathbf{q}_t \equiv \begin{bmatrix} \mathbf{data}_t \\ diag\left(\mathbf{data}_t \mathbf{data}_t'\right) \\ vech\left(\widetilde{\mathbf{data}}_t \widetilde{\mathbf{data}}_t'\right) \\ diag\left(\mathbf{data}_t \mathbf{data}_{t-1}'\right) \end{bmatrix}, \tag{38}$$

where $diag(\cdot)$ denotes the diagonal elements of a matrix and $\widetilde{\mathbf{data}}_t$ refers to the first six elements of $\mathbf{data}_t$. We omit moments on the contemporaneous correlation relating to $\log(g_t/y_t)$ and $\log h_t$ due to the parsimonious specification of government spending and the labor market in our model.

[18] These computations are done in `Matlab` 2014a on a Fujitsu laptop with an Intel(R) Core(TM) i5-4200M CPU @ 2.50 GHz.

Letting $\boldsymbol{\theta}$ contain the structural parameters, our GMM estimator is given by

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{q}_t - \mathbb{E}\left[\mathbf{q}_t\left(\boldsymbol{\theta}\right)\right]\right)' \mathbf{W} \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{q}_t - \mathbb{E}\left[\mathbf{q}_t\left(\boldsymbol{\theta}\right)\right]\right).$$

Here, $\mathbf{W}$ is a positive definite weighting matrix and $\mathbb{E}\left[\mathbf{q}_t\left(\boldsymbol{\theta}\right)\right]$ contains the model-implied moments computed in closed form using the above formulas. We use the conventional two-step implementation of GMM by letting $\mathbf{W}_T = diag\left(\hat{\mathbf{S}}_{mean}^{-1}\right)$ in a preliminary first step to obtain $\hat{\boldsymbol{\theta}}^{step1}$, where $\hat{\mathbf{S}}_{mean}$ denotes the long-run variance of $\frac{1}{T}\sum_{t=1}^{T}\mathbf{q}_t$ when re-centered around its sample mean. Our final estimates $\hat{\boldsymbol{\theta}}^{step2}$ are obtained using the optimal weighting matrix $\mathbf{W}_T = \hat{\mathbf{S}}_{\boldsymbol{\theta}^{step1}}^{-1}$, where $\hat{\mathbf{S}}_{\boldsymbol{\theta}^{step1}}$ denotes the long-run variance of our moments re-centered around $\mathbb{E}\left[\mathbf{q}_t\left(\hat{\boldsymbol{\theta}}^{step1}\right)\right]$. The long-run variances in both steps are estimated by the Newey-West estimator using 10 lags, but our results are robust to using more lags.

We estimate all structural parameters in our model except for a few poorly identified parameters. That is, we let $\delta = 0.025$ and $\theta = 0.36$ as typically considered for the U.S. economy. We also impose $\eta = 6$ to get an average markup of 20%, and we let $\phi_1 = 4$ to obtain a Frisch labor supply elasticity in the neighborhood of 0.5.[19]

8.8 Estimation Results I: The Benchmark Model

As a convenient benchmark, we first estimate our model without feedback effects from long-term bonds by imposing $\omega = 0$ and $\beta_{xhr} = 0$. This version of our model is denoted $\mathcal{M}_0$. The estimated parameters in Table 1 are fairly standard with investment adjustment costs ($\hat{\kappa} = 5.40$), little curvature in the periodic utility of consumption $\left(\hat{\phi}_2 = 0.98\right)$, and sizeable habits $\left(\hat{b} = 0.67\right)$. The latter implies a relatively low steady-state intertemporal elasticity of substitution ($IES_{ss} = 0.053$), which in the presence of internal habits is

$$IES_{ss} = \frac{1}{\phi_2}\left[\frac{\left(1 - \frac{b}{\mu_{z^*,ss}}\right)\left(\mu_{z^*,ss} - \beta b\right)}{\mu_{z^*,ss} + b\beta + \beta b^2 \mu_{z^*,ss}^{-1}}\right], \tag{39}$$

[19] The Frisch labor supply in our model is $\frac{1}{\phi_1}\left(\frac{1}{h_{ss}} - 1\right)$ and hence is affected by the steady-steady labor supply h_{ss}, which is close to 1/3 (see Table 1).

or simply $IES_{ss} \approx \frac{1}{\phi_2}\frac{(1-b)^2}{1+b+b^2}$ with $\mu_{z^*,ss} \approx 1$ and $\beta \approx 1$. As in much of the existing macro-finance literature, we find extreme levels of relative risk aversion ($\widehat{RRA} = 615.7$), even when accounting for a variable labor supply as in Swanson (2012). Using the general formulas provided in Swanson (2013), our utility function in (29) implies

$$RRA = \frac{\phi_2}{\frac{1-b\mu_{z^*,ss}^{-1}}{1-\beta b} + \frac{\phi_2}{\phi_1}\frac{W_{ss}(1-h_{ss})}{C_{ss}}} + \phi_3 \frac{1-\phi_2}{\frac{1-b\mu_{z^*,ss}^{-1}}{1-\beta b} - \frac{\left(1-b\mu_{z^*,ss}^{-1}\right)^{\phi_2}}{1-\beta b} C_{ss}^{\phi_2 - 1} + \frac{W_{ss}(1-h_{ss})}{C_{ss}}\frac{1-\phi_2}{1-\phi_1}}, \tag{40}$$

where W_{ss} and C_{ss} are the real wage and consumption in the normalized steady state.[20] Our estimated level of risk aversion is clearly too high to be consistent with the micro-evidence. For instance, Barsky, Juster, Kimball and Shapiro (1997) find a RRA between 3.8 and 15.7 in surveys, and Mehra and Prescott (1985) argue that a plausible level of relative risk aversion should not exceed 10. However, a key contribution of the present model is to demonstrate the sizeable reduction in risk aversion that follows when introducing feedback effects from long-term bonds. We also find a moderate degree of nominal frictions with prices being re-optimized roughly every fifth quarter ($\hat{\alpha} = 0.81$), and a central bank assigning more weight to stabilize inflation than output ($\hat{\beta}_\pi = 1.27$ vs. $\hat{\beta}_y = 0.03$), subject to smoothing changes in the policy rate ($\hat{\rho}_r = 0.65$).

Table 2 shows that our benchmark model $\mathcal{M}_0$ reproduces all means, in particular the short- and long-term interest rates of 5.6% and 7.0%, respectively. We only match the mean inflation rate of 3.8% due to a large precautionary saving correction that lowers the annual steady-state inflation rate of $4 \log \pi_{ss} = 4.8\%$ to obtain a model-implied inflation rate of 3.4%. The model is also successful in matching the variability in the data, except for a too low standard deviation in the 10-year excess holding period return (12.93% vs. 22.98%). A satisfying performance is also seen for the first-order autocorrelations and the contemporaneous correlations (bottom of Table 2).

Of considerable interest is the implied term premia from our model. Following Rudebusch and Swanson (2012), we define term premia as $TP_{t,k} = r_{t,k} - \tilde{r}_{t,k}$, where $\tilde{r}_{t,k}$ is the yield-to-maturity on a zero-coupon bond $\tilde{P}_{t,k}$ under risk-neutral valuation by the financial intermediary, i.e., $\tilde{P}_{t,k} = e^{-r_t}\mathbb{E}_t\left[\tilde{P}_{t+1,k-1}\right]$. Our benchmark model has a 10-year term premium $TP_{t,40}$ with a

[20] Household wealth is measured by the present value of lifetime consumption in (40) as recommended by Swanson (2013). Given that W_{ss} and C_{ss} are unaffected by ϕ_3, we then use (40) to back out the value of the Epstein-Zin-Weil coefficient ϕ_3 for a given value of the RRA during the estimation. Rudebusch and Swanson (2012) explain why the benchmark model requires high risk aversion to match post-war U.S. data.

mean of 145 basis points, which is close to the average slope of the yield curve (139 basis points) that serves as an observable proxy for the average term premium. We also find substantial variation in the 10-year term premium having a standard deviation of 116 basis points. This is in line with the variability in the 10-year term premium obtained in various Gaussian affine term structure models for our sample: i) the three- and four-factor models of Andreasen and Meldrum (2014) with bias-adjusted factor dynamics have a standard deviation of 105 and 115 basis points, respectively, and ii) the five-factor model of Adrian, Crump and Moench (2013) has a standard deviation of 121 basis points. Finally, our model is consistent with another noisy measure of term premia variability, namely the standard deviation of the slope for the 10-year yield curve, which equals 139 basis points.

8.9 Understanding the Volatility of the Term Premium

Although high risk aversion helps to increase the mean term premium, it does not necessarily generate a highly volatile term premium. To understand the main mechanism behind the variability in $TP_{t,40}$, recall that its volatility is directly related to the degree of heteroscedasticity in the stochastic discount factor $M_{t,t+1} \equiv \beta\lambda_{t+1}/\left(\lambda_t\pi_{t+1}\right)$, i.e., the variation of $\mathbb{V}_t\left(M_{t,t+1}\right)$. The three shocks in our model are all homoscedastic. Thus, the model is endogenously generating heteroscedasticity in $M_{t,t+1}$, as captured by our third-order perturbation. But, what is the source of this large heteroscedasticity? A possibility is to consider the effect of the price dispersion index s_t, which is an endogenous state variable. Combining (36) and (37), its law of motion is then given by

$$s_{t+1} = (1-\alpha)^{\frac{1}{1-\eta}}\left[1-\alpha\pi_t^{\eta-1}\right]^{\frac{\eta}{\eta-1}} + \alpha\pi_t^{\eta}s_t,$$

which is highly non-linear to ensure $s_t \geq 1$, as shown by Schmitt-Grohe and Uribe (2007). The first term in the expression for s_{t+1} does not generate much heteroscedasticity with $1-\alpha\pi_t^{\eta-1}$ being well below one given our estimates. The second term $\alpha\pi_t^{\eta}s_t$, on the other hand, may generate extreme levels of heteroscedasticity because $s_t \geq 1$ and we generally also have $\pi_t \geq 1$. Also, the degree of heteroscedasticity is increasing in the mean of both variables. A higher value of π_{ss} clearly increases π_t, but also the steady state of s_t, given that $\partial s_{ss}/\partial\pi_{ss} \geq 0$ for $\pi_{ss} \geq 1$. This effect is illustrated in Figure 1 by considering a sample path with positive steady-state inflation $\left(\pi_{ss} = \hat{\pi}_{ss}^{GMM}\right)$ and one without ($\pi_{ss} = 1$). In the case of positive steady-state inflation (left panels), we see more extreme observations and hence more heteroscedasticity compared to the case of no steady-state inflation (right panels). Note also how the capital stock and the price dispersion with positive steady-state inflation attain very low and high values, respectively, just before observation $9,000$, at which point the unpruned state-space system explodes. A similar divergence in sample path does not appear for $\pi_{ss} = 1$, where the two approximations generate nearly identical time series. Accordingly, positive steady-state inflation serves as a channel to generate heteroscedasticity in s_t and, hence, variation in $\mathbb{V}_t(M_{t,t+1})$, as required to produce the volatile 10-year term premium in our model.[21]

The first column in Table 4 shows that this channel has a large effect, as the standard deviation of the 10-year term premium falls from 116 basis points with positive steady-state inflation to just 1.42 basis points when $\pi_{ss} = 1.00$. To further decompose the effects of positive steady-state inflation, we adopt the standard decomposition of risk premia into the market price of risk MPR_t times the quantity of risk. As in Cochrane (2001), we let $MPR_t = \mathbb{V}_t(M_{t,t+1}) / \mathbb{E}_t[M_{t,t+1}]$, implying that the quantity of risk $QoR_{t,k}$ equals $TP_{t,k}/MPR_t$. Table 4 shows that omitting positive steady-state inflation lowers the standard deviation in the MPR_t by a factor of 100, whereas the standard deviation of $QoR_{t,k}$ falls by a factor of 10^5. Hence, positive steady-state inflation mainly generates a volatile term premium in our model by increasing the variability in the quantity of risk. Table 4 further shows that positive steady-state inflation also affects the mean term premium, which falls from 145 basis points to just 32 basis points when $\pi_{ss} = 1.00$, although RRA equals 615.7! This fall is due to a reduction in the mean of $\mathbb{V}_t(M_{t,t+1})$, which lowers the MPR_t, whereas the level for the $QoR_{t,k}$ is nearly unaffected.

Thus, accounting for positive steady-state inflation serves as a key new channel to endogenously generate heteroscedasticity in the New Keynesian model and produce a 10-year term premium with the desired level and variability. Importantly, an unpruned third-order approximation to our model results in explosive sample paths and is unable to "detect" this novel channel, which we uncover by using our pruning scheme for a third-order perturbation.

[21] Swanson (2015) also emphasizes the importance of the price dispersion index as a source of heteroscedasticity in the New Keynesian model but without noticing the importance of positive steady-state inflation for this channel.

8.10 Estimation Results II: The First Feedback Effect

Our next step is to introduce the first feedback effect from long-term bonds by allowing $\omega \geq 0$, while still maintaining that the central bank does not respond to the excess holding period return ($\beta_{xhr} = 0$). This version of our model is referred to as $\mathcal{M}_{FB}$. Table 1 shows that the financial intermediary is estimated to allocate a large fraction of its investments to long-term bonds with $\hat{\omega} = 0.91$. A standard $t-$test clearly rejects the null hypothesis of $\omega = 0$ at conventional significant levels, which provide support for our first feedback channel from long-term bonds. Another important property of $\mathcal{M}_{FB}$ relates to the estimated degree of RRA, which is only 23, and thus substantially lower than in the benchmark model. For the remaining parameters, we find minor changes compared to our benchmark model, except for the policy rule and investment adjustment costs.

Table 2 shows that $\mathcal{M}_{FB}$ delivers a satisfying fit to the considered moments despite its lower risk aversion. To quantify the performance of $\mathcal{M}_{FB}$ compared to the benchmark model, Table 3 reports objective functions from our two-step GMM procedure. Only the objective functions from the first step use the same weighting matrix and are, therefore, comparable across models. They show that $\mathcal{M}_{FB}$ fits the data better than the benchmark model (14.546 vs. 16.929).[22]

However, risk aversion in $\mathcal{M}_{FB}$ is estimated very imprecisely with a large standard error of 31, and it is likely that RRA can be lowered further with only a minor reduction in the goodness of fit. Consistent with the micro-evidence provided in Barsky, Juster, Kimball and Shapiro (1997), we restrict RRA to 5 and re-estimate our model with the first feedback effect. Table 2 verifies our conjecture as this restricted model $\mathcal{M}_{FB}^{RRA}$ with low risk aversion provides nearly the same fit as the unrestricted model. In particular, $\mathcal{M}_{FB}^{RRA}$ matches the slope of the yield curve, while simultaneously fitting key moments for the five macro variables. We also note from Table 3 that $\mathcal{M}_{FB}^{RRA}$ provides a better overall fit to the data than our benchmark model. Here, we report the $P-$value from the $J-$test for model misspecification, showing that we are unable to reject $\mathcal{M}_{FB}^{RRA}$ (and all the other models). That is, the observed differences between empirical and model-implied moments in Table 2 are not unusual given the sample variation in the empirical moments. However, this finding should be interpreted with caution as the $J-$test has low power due to our sample size ($T = 186$). Finally, $\mathcal{M}_{FB}^{RRA}$ generates a realistic 10-year term premium with a mean of 140 basis points and a

[22] The estimates in step 1 are very similar to those reported for step 2 in Table 1, implying that the objective functions in step 1 serve as a good metric for model comparison.

standard deviation of 93 basis points, as seen from Table 4.

Thus, allowing the financial intermediary to invest in long-term bonds goes a long way in resolving the bond risk premium puzzle described in Rudebusch and Swanson (2008) without postulating highly risk-averse households as in much of the existing literature (see Andreasen (2012), Binsbergen, Fernandez-Villaverde, Koijen and Rubio-Ramirez (2012), and Rudebusch and Swanson (2012), among others).

8.11 Understanding the First Feedback Effect

We now explore the mechanisms that enable $\mathcal{M}_{FB}$ and $\mathcal{M}_{FB}^{RRA}$ to generate a large and variable term premium without relying on high risk aversion. Suppose $xhr_{t,L} > 0$ for $\omega = 0$ and consider increasing ω to some positive value less than one. This increase in ω lowers $\mathbb{E}_t\left[M_{t,t+1}\right]$ as $\mathbb{E}_t\left[M_{t,t+1}\right] = e^{-r_t - \omega \times xhr_{t,L}}$ and, hence, the current bond price because

$$P_{t,L} = \mathbb{E}_t\left[M_{t,t+1}\right]\mathbb{E}_t\left[P_{t+1,L-1}\right] + Cov_t\left(M_{t,t+1}, P_{t+1,L-1}\right).$$

This fall in $P_{t,L}$ increases $xhr_{t,L}$ according to (32). But a higher $xhr_{t,L}$ induces a further fall in $\mathbb{E}_t\left[M_{t,t+1}\right]$ and the current bond price $P_{t,L}$, which generates an even larger increase in $xhr_{t,L}$. That is, $\omega > 0$ generates a "feedback multiplication effect" that amplifies the level and variability in $xhr_{t,L}$. An explicit way to see the implication of this feedback loop is to use a first-order approximation of the logarithmic and exponential function in (32) to obtain (see Appendix A.20):

$$xhr_{t,L} \approx \frac{1}{1 - \omega \frac{\mathbb{E}_t\left[P_{t+1,L-1}\right]}{P_{ss,L-1}}}\left[\left(\frac{\mathbb{E}_t\left[P_{t+1,L-1}\right]}{P_{ss,L-1}} - 1\right)(r_t - r_{ss}) - Cov_t\left(\frac{M_{t,t+1}}{M_{ss,ss+1}}, \frac{P_{t+1,L-1}}{P_{ss,L-1}}\right)\right]. \quad (41)$$

The first term in (41) is the risk-neutral component of $xhr_{t,L}$, whereas $Cov_t\left(\frac{M_{t,t+1}}{M_{ss,ss+1}}, \frac{P_{t+1,L-1}}{P_{ss,L-1}}\right)$ is the required compensation by the risk averse household for carrying risk. The expression in (41) shows that both terms in $xhr_{t,L}$ are amplified by the factor $1/\left(1 - \omega\frac{\mathbb{E}_t\left[P_{t+1,L-1}\right]}{P_{ss,L-1}}\right)$ when $\omega > 0$. Therefore, our model requires lower volatility in $M_{t,t+1}$ and, correspondingly, lower risk aversion, to match short- and long-term interest rates. Finally, extending the expression for the term premium in Rudebusch and Swanson (2012) to our model, it follows that

$$
\begin{aligned}
TP_{t,k} \approx & -\frac{1}{kP_{ss,k}}\mathbb{E}_t\left[\sum_{j=0}^{k-1} e^{\left\{-\sum_{m=0}^{j-1}\left(r_{t+m}+\omega\times xhr_{t+m,L}\right)\right\}} Cov_{t+j}\left(M_{t+j,t+j+1}, P_{t+j+1,k-j-1}\right)\right] \\
& -\frac{1}{kP_{ss,k}}\mathbb{E}_t\left[e^{\left\{-\sum_{m=0}^{k-1}\omega\times xhr_{t+m,L}\right\}}-1\right]. \qquad (42)
\end{aligned}
$$

This expression demonstrates how a higher level and variability in $xhr_{t,L}$ translates into a larger and more volatile term premium, which is to be expected given that both variables measure compensation for risk.

To illustrate the magnitude of the multiplication effect from long-term bonds on term premia, we momentarily set $\omega = 0$ in $\mathcal{M}_{FB}^{RRA}$, whereas all remaining parameters are as reported in Table 1. Omitting the first feedback channel from long-term bonds reduces the mean of the 10-year term premium from 140 to 9 basis points, and similarly for the standard deviation which falls from 93 to 4 basis points. Table 4 documents that the feedback channel from long-term bonds reduces the mean and standard deviation of $\mathbb{V}_t\left[M_{t,t+1}\right]$ by a factor of 100 in $\mathcal{M}_{FB}$ and $\mathcal{M}_{FB}^{RRA}$ compared to the benchmark model. This, in turn, leads to a similar reduction in the corresponding moments for the MPR_t. Hence, $\mathcal{M}_{FB}$ and $\mathcal{M}_{FB}^{RRA}$ generate a high and volatile term premium by increasing the quantity of risk. As observed for the benchmark model, we see in Table 4 that positive steady-state inflation is essential for $\mathcal{M}_{FB}$ and $\mathcal{M}_{FB}^{RRA}$ to generate the desired level and variability of the term premium even with the first feedback from long-term bonds. Hence, it would be hard to discover this novel feedback effect without our pruning method, as an unpruned state-space system generates explosive sample paths when we assume positive steady-state inflation.

8.12 Estimation Results III: The First and Second Feedback Effects

We finally introduce our second feedback effect from long-term bonds by allowing the central bank to respond to variation in $xhr_{t,L}$, which is closely related to term premia as shown in (41) and (42). That is, we let $\beta_{xhr} \neq 0$ and refer to this model as $\mathcal{M}_{FB,Taylor}$. Table 1 shows that this second feedback effect from long-term bonds has a small effect in the model as $\hat{\beta}_{xhr} = -0.069$, a point estimate not sufficiently far from zero to be statistically significant. However, the effect from our second feedback effect is somewhat larger when RRA is restricted to 5 in $\mathcal{M}_{FB,Taylor}^{RRA}$.

Here, $\hat{\beta}_{xhr} = -0.519$ and the response of the central bank to $xhr_{t,L}$ is now significant given a standard error of 0.094 for $\hat{\beta}_{xhr}$. That is, our model implies a reduction in the policy rate when term premia and $xhr_{t,L}$ increase, as the central bank tries to offset the rise in the deposit rate with a lower policy rate. Although we end our sample in 2007.Q4, this finding is consistent with monetary policy during the recent financial crisis, where the Federal Reserve undertook vigorous policy measures to stimulate economic activity in response to elevated levels of term premia.

Table 2 shows that $\mathcal{M}^{RRA}_{FB,Taylor}$ matches most of the moments considered, in particular all mean values and the slope of the yield curve. Table 3 documents how $\mathcal{M}^{RRA}_{FB,Taylor}$ outperforms both $\mathcal{M}^{RRA}_{FB}$ and the benchmark model in terms of overall goodness of fit, although $\mathcal{M}_{FB,Taylor}$ with unrestricted risk aversion does somewhat better than $\mathcal{M}^{RRA}_{FB,Taylor}$. The term premium is also found to be consistent with empirical moments, as $\mathcal{M}^{RRA}_{FB,Taylor}$ generates a 10-year term premium with a mean of 142 basis points and a standard deviation of 132 basis points (see Table 4). As before, positive steady-state inflation is essential for $\mathcal{M}_{FB,Taylor}$ and $\mathcal{M}^{RRA}_{FB,Taylor}$ to generate the desired level and variability in the term premium by "activating" the two novel feedback effects from long-term bonds to the real economy considered in this paper.

8.13 GIRFs and Conditional GIRFs

Our next exercise is to report the GIRFs following positive one-standard-deviation shocks in $\mathcal{M}^{RRA}_{FB,Taylor}$ to technology, government spending, and preferences (Figures 2 to 4). These functions are computed for a log-linearized solution and a third-order approximation using (26) with the relevant state variables at their unconditional means. All the GIRFs have the expected pattern, and we therefore direct attention to the effects of higher-order terms, i.e., the differences between the marked and unmarked lines. Shocks to technology and government spending have substantial non-linear effects on consumption and investment, mainly because these shocks generate considerable variation in the ex ante excess holding period return and the term premium. This finding reveals that higher-order effects, and hence the amount of risk in the economy, affect real allocations in our model, which therefore overturns the result of Tallarini (2000) that risk does not matter for real allocations.

A key advantage of computing second- and third-order approximations is that we can analyze the effects of different shocks conditional on the state of the economy. This is illustrated in Figure 5, where we show how the response of consumption, investment, and $r_{t,40}$ to a positive one-standard-deviation shock to technology is larger when the economy is in a recession than when it is not.[23] The intuition is that, when the economy is in a recession, consumption and capital tend to be low, and hence the marginal utility of extra consumption and the marginal return of additional investment are higher than usual. A similar exercise is done in Figure 6, except that now we compare the situation where we condition on high vs. low inflation. When inflation is high, consumption, investment, and interest rates respond more vigorously than when inflation is low. When inflation is high, nominal rigidities are particularly damaging, since firms that are not able to change their prices are far from the price they would set under flexible prices. A positive productivity shock translates into lower inflation through lower marginal costs and, hence, it alleviates these pernicious effects of nominal rigidities. When inflation is low, nominal rigidities are less of a constraint on firm behavior and a positive technology shock is less useful for firms.[24] The asymmetries in responses to shocks documented by Figures 5 and 6 demonstrate how the methods we present in our paper allow researchers to probe deeper into the behavior of their models and uncover economic mechanisms that would otherwise remain hidden.

9 Conclusion

This paper extends the pruning method by Kim, Kim, Schaumburg and Sims (2008) to third- and higher-order approximations, with special attention devoted to models solved up to third order. Conditions for the existence of first and second unconditional moments are derived, and their values are provided in closed form. The existence of higher-order unconditional moments in the form of skewness and kurtosis is also established. We also analyze GIRFs and provide simple closed-form expressions for these functions.

The econometric implications of our findings are significant, as most of the existing moment-based estimation methods for linearized DSGE models now carry over to non-linear approximations. For approximations up to third order, this includes GMM estimation based on first and second un-

[23] We define a recession as a quarter where there is (detrended) negative output in the current and the previous two periods. Otherwise, the economy is in expansion.

[24] High inflation is defined as inflation larger than one standard deviation of inflation. Otherwise the economy is defined to be in a low inflation regime. The corresponding conditional GIRFs for government spending and preference shocks are omitted in the interest of space.

conditional moments and matching model-implied GIRFs to their empirical counterparts. When simulations are needed, our analysis also provides a foundation for different types of indirect inference and SMM. These results are not just relevant for classical inference, as the moment conditions in optimal GMM estimation may be used to build a limited information likelihood function, from which Bayesian inference may be carried out.

To illustrate one of the new estimation methods that our paper makes available, we revisit the term structure implications of the New Keynesian model. We first demonstrate a new channel to amplify the level and time variation in term premia by accounting for positive steady-state inflation. Given this more realistic term premium, we then introduce two feedback effects from long-term bonds to the real economy, and we show that they enable the New Keynesian model to generate a high and variable term premium with the same low risk aversion as found in the micro-evidence. We once again emphasize that our pruning scheme has greatly facilitated the discovery of these new channels and helped us to address the long-standing bond premium puzzle.

A Appendix

A.1 Non-linearities Between State Variables and Innovations

To illustrate how non-linearities between $\mathbf{x}_t$ and $\boldsymbol{\epsilon}_{t+1}$ can be addressed in our framework, let $\mathbf{v}_t \equiv \left[\begin{array}{cc} \mathbf{x}'_{t-1} & \boldsymbol{\epsilon}'_t \end{array} \right]'$ be an expanded state vector where the innovations now appear as state variables. The new state equation is, then, given by

$$\mathbf{v}_{t+1} = \left[\begin{array}{c} \mathbf{h}\left(\mathbf{v}_t, \sigma\right) \\ \mathbf{0} \end{array} \right] + \sigma \left[\begin{array}{c} \mathbf{0}_{n_x} \\ \mathbf{u}_{t+1} \end{array} \right],$$

where $\mathbf{u}_{t+1} \sim IID\left(\mathbf{0}, \mathbf{I}\right)$ is of dimension n_ϵ, and the new observation equation is

$$\mathbf{y}_t = \mathbf{g}\left(\mathbf{v}_t, \sigma\right).$$

Thus, any model with non-linearities between state variables and innovations may be rewritten into our notation with only linear innovations.

As an illustration, consider a neoclassical growth model with stochastic volatility. Using standard notation, the equilibrium conditions are given by:

$$c_t^{-\gamma} = \beta \mathbb{E}_t \left[c_{t+1}^{-\gamma} \left(a_{t+1} \alpha k_{t+1}^{\alpha-1} + 1 - \delta \right) \right]$$
$$c_t + k_{t+1} = a_t k_t^{\alpha} + (1-\delta) k_t$$
$$\log a_{t+1} = \rho \log a_t + \sigma_{a,t+1} \epsilon_{a,t+1}$$

and

$$\log \left(\frac{\sigma_{a,t+1}}{\sigma_{a,ss}} \right) = \rho_\sigma \log \left(\frac{\sigma_{a,t}}{\sigma_{a,ss}} \right) + \epsilon_{\sigma,t+1}.$$

We then rewrite these conditions as:

$$c_t^{-\gamma} = \mathbb{E}_t \left[\beta c_{t+1}^{-\gamma} \left(\exp^{\left\{ \rho \log a_t + \sigma_{a,ss} \exp^{\left\{ \rho_\sigma \log\left(\frac{\sigma_{a,t}}{\sigma_{a,ss}} \right) + \epsilon_{\sigma,t+1} \right\}} \epsilon_{a,t+1} \right\}} \alpha k_{t+1}^{\alpha-1} + 1 - \delta \right) \right]$$
$$c_t + k_{t+1} = a_t k_t^{\alpha} + (1-\delta) k_t$$
$$\log a_t = \rho \log a_{t-1} + \sigma_{a,t} \epsilon_{a,t}$$
$$\log \left(\frac{\sigma_{a,t}}{\sigma_{a,ss}} \right) = \rho_\sigma \log \left(\frac{\sigma_{a,t-1}}{\sigma_{a,ss}} \right) + \epsilon_{\sigma,t}$$
$$\epsilon_{a,t+1} = \sigma u_{a,t+1}$$

and

$$\epsilon_{\sigma,t+1} = \sigma u_{\sigma,t+1},$$

where the extended state vector is $\mathbf{v}_t \equiv \left[\begin{array}{ccccc} k_t & a_{t-1} & \sigma_{a,t-1} & \epsilon_{a,t} & \epsilon_{\sigma,t} \end{array} \right]$ and σ is the perturbation parameter scaling the innovations $u_{a,t+1}$ and $u_{\sigma,t+1}$.

If, instead, the volatility process is specified as a GARCH(1,1) model, then the equilibrium conditions can be expressed as:

$$c_t^{-\gamma} = \mathbb{E}_t\left[\beta c_{t+1}^{-\gamma}\left(\exp^{\{\rho\log a_t + \sigma_{a,t+1}\epsilon_{a,t+1}\}}\alpha k_{t+1}^{\alpha-1} + 1 - \delta\right)\right]$$
$$c_t + k_{t+1} = a_t k_t^{\alpha} + (1-\delta)\,k_t$$
$$\log a_t = \rho\log a_{t-1} + \sigma_{a,t}\epsilon_{a,t}$$
$$\sigma_{a,t+1}^2 = (1-\rho_1)\,\sigma_{a,ss}^2 + \rho_1\sigma_{a,t}^2 + \rho_2\sigma_{a,t}^2\epsilon_{a,t}^2$$

and

$$\epsilon_{a,t+1} = \sigma u_{t+1},$$

where the extended state vector is $\mathbf{v}_t \equiv \begin{bmatrix} k_t & \sigma_{a,t} & a_{t-1} & \epsilon_{a,t} \end{bmatrix}$ and σ is the perturbation parameter scaling u_{t+1}. As in Andreasen (2012), the constant term in the GARCH process is scaled by $(1-\rho_1)$ to ensure that $\sigma_{a,t} = \sigma_{a,ss}$ in the steady state where $\epsilon_{a,t}^2 = 0$.

A.2 Pruned State-Space Beyond Third Order

The pruned state-space system for the kth-order approximation based on the kth-order Taylor series expansions of $\mathbf{g}(\mathbf{x}_t,\sigma)$ and $\mathbf{h}(\mathbf{x}_t,\sigma)$ are obtained by: i) decomposing the state variables into first-, second-, ... , and kth-order effects, ii) setting up laws of motion for the state variables capturing only first-, second-, ... , and kth-order effects, and iii) constructing the expression for control variables by preserving only effects up to kth-order. In comparison, the unpruned state-space system for the kth-order approximation is given by the kth-order Taylor series expansions of $\mathbf{g}(\mathbf{x}_t,\sigma)$ and $\mathbf{h}(\mathbf{x}_t,\sigma)$.

A.3 Coefficients for the Pruned State-Space System at Second Order

$$\mathbf{A}^{(2)} \equiv \begin{bmatrix} \mathbf{h_x} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{h_x} & \frac{1}{2}\mathbf{H_{xx}} \\ \mathbf{0} & \mathbf{0} & \mathbf{h_x}\otimes\mathbf{h_x} \end{bmatrix},$$

$$\mathbf{B}^{(2)} \equiv \begin{bmatrix} \sigma\boldsymbol{\eta} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma\boldsymbol{\eta}\otimes\boldsymbol{\sigma\eta} & \sigma\boldsymbol{\eta}\otimes\mathbf{h_x} & \mathbf{h_x}\otimes\sigma\boldsymbol{\eta} \end{bmatrix},$$

$$\boldsymbol{\xi}_{t+1}^{(2)} \equiv \begin{bmatrix} \boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1} - vec\left(\mathbf{I}_{n_e}\right) \\ \boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f \\ \mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1} \end{bmatrix},$$

$$\mathbf{c}^{(2)} \equiv \begin{bmatrix} \mathbf{0} \\ \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 \\ (\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta})\,vec\left(\mathbf{I}_{n_e}\right) \end{bmatrix},$$

$$\mathbf{C}^{(2)} \equiv \begin{bmatrix} \mathbf{g_x} & \mathbf{g_x} & \frac{1}{2}\mathbf{G_{xx}} \end{bmatrix},$$

and

$$\mathbf{d}^{(2)} \equiv \frac{1}{2}\mathbf{g}_{\sigma\sigma}\sigma^2.$$

A.4 Second Order: Stability

First, note that all eigenvalues of $\mathbf{A}^{(2)}$ are strictly less than one. To see this, we work with

$$\begin{aligned}
p(\lambda) &= \left|\mathbf{A}-\lambda\mathbf{I}_{2n_x+n_x^2}\right| \\
&= \left|\begin{bmatrix} \mathbf{h_x}-\lambda\mathbf{I}_{n_x} & \mathbf{0}_{n_x\times n_x} & \mathbf{0}_{n_x\times n_x^2} \\ \mathbf{0}_{n_x\times n_x} & \mathbf{h_x}-\lambda\mathbf{I}_{n_x} & \frac{1}{2}\mathbf{H_{xx}} \\ \mathbf{0}_{n_x^2\times n_x} & \mathbf{0}_{n_x^2\times n_x} & \mathbf{h_x}\otimes\mathbf{h_x}-\lambda\mathbf{I}_{n_x^2} \end{bmatrix}\right| \\
&= \begin{vmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{vmatrix} \\
&= \left|\mathbf{B}_{11}\right|\left|\mathbf{B}_{22}\right|,
\end{aligned}$$

where we let

$$\begin{aligned}
\mathbf{B}_{11} &\equiv \begin{bmatrix} \mathbf{h_x}-\lambda\mathbf{I}_{n_x} & \mathbf{0}_{n_x\times n_x} \\ \mathbf{0}_{n_x\times n_x} & \mathbf{h_x}-\lambda\mathbf{I}_{n_x} \end{bmatrix}, \\
\mathbf{B}_{12} &\equiv \begin{bmatrix} \mathbf{0}_{n_x\times n_x^2} \\ \frac{1}{2}\mathbf{H_{xx}} \end{bmatrix}, \\
\mathbf{B}_{21} &\equiv \begin{bmatrix} \mathbf{0}_{n_x^2\times n_x} & \mathbf{0}_{n_x^2\times n_x} \end{bmatrix},
\end{aligned}$$

and

$$\mathbf{B}_{22} \equiv \mathbf{h_x}\otimes\mathbf{h_x}-\lambda\mathbf{I}_{n_x^2}$$

and we use the fact that

$$\begin{vmatrix} \mathbf{U} & \mathbf{C} \\ \mathbf{0} & \mathbf{Y} \end{vmatrix} = \left|\mathbf{U}\right|\left|\mathbf{Y}\right|,$$

where $\mathbf{U}$ is an $m\times m$ matrix and $\mathbf{Y}$ is an $n\times n$ matrix. Hence,

$$p(\lambda) = \left|\begin{bmatrix} \mathbf{h_x}-\lambda\mathbf{I}_{n_x} & \mathbf{0}_{n_x\times n_x} \\ \mathbf{0}_{n_x\times n_x} & \mathbf{h_x}-\lambda\mathbf{I}_{n_x} \end{bmatrix}\right| \left|\mathbf{h_x}\otimes\mathbf{h_x}-\lambda\mathbf{I}_{n_x^2}\right| = \left|\mathbf{h_x}-\lambda\mathbf{I}_{n_x}\right|\left|\mathbf{h_x}-\lambda\mathbf{I}_{n_x}\right|\left|\mathbf{h_x}\otimes\mathbf{h_x}-\lambda\mathbf{I}_{n_x^2}\right|.$$

The eigenvalues are determined from $\left|\mathbf{h_x}-\lambda\mathbf{I}_{n_x}\right|=0$ or $\left|\mathbf{h_x}\otimes\mathbf{h_x}-\lambda\mathbf{I}_{n_x^2}\right|=0$. The absolute values of all eigenvalues to the first problem are strictly less than one by assumption. That is, $|\lambda_i|<1$ $i=1,2,...,n_x$. This is also the case for the second problem because the eigenvalues to $\mathbf{h_x}\otimes\mathbf{h_x}$ are $\lambda_i\lambda_j$ for $i=1,2,...,n_x$ and $j=1,2,...,n_x$.

A.5 Second Order: Unconditional Second Moments

For the variance, we have

$$\mathbb{V}\left(\mathbf{z}_{t+1}^{(2)}\right) = \mathbf{A}^{(2)}\mathbb{V}\left(\mathbf{z}_{t}^{(2)}\right)\left(\mathbf{A}^{(2)}\right)' + \mathbf{B}^{(2)}\mathbb{V}\left(\boldsymbol{\xi}_{t+1}^{(2)}\right)\left(\mathbf{B}^{(2)}\right)'$$

as

$$\mathbb{E}\left[\mathbf{z}_t^{(2)}\left(\boldsymbol{\xi}_{t+1}^{(2)}\right)'\right]=\mathbb{E}\left[\begin{array}{cccc} \mathbf{x}_t^f\boldsymbol{\epsilon}_{t+1}' & \mathbf{x}_t^f\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}-vec\left(\mathbf{I}_{ne}\right)\right)' & \mathbf{x}_t^f\left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)' & \mathbf{x}_t^f\left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)' \\ \mathbf{x}_t^s\boldsymbol{\epsilon}_{t+1}' & \mathbf{x}_t^s\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}-vec\left(\mathbf{I}_{ne}\right)\right)' & \mathbf{x}_t^s\left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)' & \mathbf{x}_t^s\left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)' \\ \left(\mathbf{x}_t^f\otimes\mathbf{x}_t^f\right)\boldsymbol{\epsilon}_{t+1}' & \left(\mathbf{x}_t^f\otimes\mathbf{x}_t^f\right)\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}-vec\left(\mathbf{I}_{ne}\right)\right)' & \left(\mathbf{x}_t^f\otimes\mathbf{x}_t^f\right)\left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)' & \left(\mathbf{x}_t^f\otimes\mathbf{x}_t^f\right)\left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)' \end{array}\right]=\mathbf{0}$$

Now, we only need to compute $\mathbb{V}\left(\boldsymbol{\xi}_{t+1}^{(2)}\right)$:

$$\begin{aligned}\mathbb{V}\left(\boldsymbol{\xi}_{t+1}^{(2)}\right) &= \mathbb{E}\left[\left[\begin{array}{c}\boldsymbol{\epsilon}_{t+1}\\ \boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}-vec\left(\mathbf{I}_{n_e}\right)\\ \boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\\ \mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\end{array}\right]\left[\begin{array}{c}\boldsymbol{\epsilon}_{t+1}\\ \boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}-vec\left(\mathbf{I}_{n_e}\right)\\ \boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\\ \mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\end{array}\right]'\right]\\ &= \left[\begin{array}{cccc}\mathbf{I}_{n_e} & \mathbb{E}\left[\boldsymbol{\epsilon}_{t+1}\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\right)'\right] & \mathbf{0} & \mathbf{0}\\ \mathbb{E}\left[\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\right)\boldsymbol{\epsilon}_{t+1}'\right] & \mathbb{E}\left[\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}-vec\left(\mathbf{I}_{n_e}\right)\right)\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}-vec\left(\mathbf{I}_{n_e}\right)\right)'\right] & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbb{E}\left[\left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)\left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)'\right] & \mathbb{E}\left[\left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)\left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)'\right]\\ \mathbf{0} & \mathbf{0} & \mathbb{E}\left[\left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)\left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)'\right] & \mathbb{E}\left[\left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)\left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)'\right]\end{array}\right].\end{aligned}$$

This variance is finite when $\boldsymbol{\epsilon}_{t+1}$ has a finite fourth moment. All elements in this matrix can be computed element-by-element.

A.6 Second Order: Unconditional Third and Fourth Moments

We consider the system $\mathbf{x}_{t+1}=\mathbf{a}+\mathbf{A}\mathbf{x}_t+\mathbf{v}_{t+1}$, where $\mathbf{A}$ is stable and $\mathbf{v}_{t+1}$ are mean-zero innovations. Thus, the pruned state-space representations of DSGE models belong to this class. For notational convenience, the system is expressed in deviation from its mean as $\mathbf{a}=(\mathbf{I}-\mathbf{A})\,\mathbb{E}[\mathbf{x}_t]$. Therefore

$$\begin{gathered}\mathbf{x}_{t+1}=(\mathbf{I}-\mathbf{A})\,\mathbb{E}[\mathbf{x}_t]+\mathbf{A}\mathbf{x}_t+\mathbf{v}_{t+1}\Rightarrow\\ \mathbf{x}_{t+1}-\mathbb{E}[\mathbf{x}_t]=\mathbf{A}\left(\mathbf{x}_t-E[\mathbf{x}_t]\right)+\mathbf{v}_{t+1}\Rightarrow\\ \mathbf{z}_{t+1}=\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\end{gathered}$$

We then have

$$\begin{aligned}\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1} &= \left(\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\right)\otimes\left(\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\right)\\ &= \mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t+\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}+\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1},\end{aligned}$$

$$\begin{aligned}
\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1} \;=\;& \mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t+\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\\
&+\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t+\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\\
&+\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\\
&+\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1} \;=\;& \mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t+\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\\
&+\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t+\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\\
&+\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t+\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\\
&+\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t+\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\\
&+\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\\
&+\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\\
&+\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t\otimes\mathbf{v}_{t+1}\\
&+\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{A}\mathbf{z}_t+\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}\otimes\mathbf{v}_{t+1}.
\end{aligned}$$

Thus, to solve for $\mathbb{E}\left[\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1}\right]$, the innovations need to have a finite third moment. At second order, $\mathbf{v}_{t+1}$ depends on $\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}$, meaning that $\boldsymbol{\epsilon}_{t+1}$ must have a finite sixth moment. Similarly, to solve for $\mathbb{E}\left[\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1}\otimes\mathbf{z}_{t+1}\right]$, the innovations need to have finite fourth moments. At second order, $\mathbf{v}_{t+1}$ depends on $\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}$, meaning that $\boldsymbol{\epsilon}_{t+1}$ must have a finite eighth moment.

A.7 Coefficients for the Pruned State-Space System at Third Order

$$\mathbf{A}^{(3)} \equiv \begin{bmatrix}
\mathbf{h_x} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{h_x} & \frac{1}{2}\mathbf{H_{xx}} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{h_x}\otimes\mathbf{h_x} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2 & \mathbf{0} & \mathbf{0} & \mathbf{h_x} & \mathbf{H_{xx}} & \frac{1}{6}\mathbf{H_{xxx}}\\
\mathbf{h_x}\otimes\frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{h_x}\otimes\mathbf{h_x} & \mathbf{h_x}\otimes\frac{1}{2}\mathbf{H_{xx}}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{h_x}\otimes\mathbf{h_x}\otimes\mathbf{h_x}
\end{bmatrix},$$

$$\mathbf{B}^{(3)} \equiv \left[\begin{matrix}
\sigma\boldsymbol{\eta} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta} & \sigma\boldsymbol{\eta}\otimes\mathbf{h_x} & \mathbf{h_x}\otimes\sigma\boldsymbol{\eta}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\sigma\boldsymbol{\eta}\otimes\frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}
\end{matrix}\right.$$

$$\begin{matrix}
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\sigma\boldsymbol{\eta}\otimes\mathbf{h_x} & \sigma\boldsymbol{\eta}\otimes\frac{1}{2}\mathbf{H_{xx}} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \sigma\boldsymbol{\eta}\otimes\mathbf{h_x}\otimes\mathbf{h_x} & \mathbf{h_x}\otimes\mathbf{h_x}\otimes\sigma\boldsymbol{\eta} & \mathbf{h_x}\otimes\sigma\boldsymbol{\eta}\otimes\mathbf{h_x}
\end{matrix}$$

$$\left.\begin{matrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{h_x} \otimes \sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta} & \sigma\boldsymbol{\eta} \otimes \mathbf{h_x}\otimes\sigma\boldsymbol{\eta} & \sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta} \otimes \mathbf{h_x} & \sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta} \end{matrix}\right],$$

$$\boldsymbol{\xi}_{t+1}^{(3)} \equiv \begin{bmatrix} \boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} - vec\left(\mathbf{I}_{n_e}\right) \\ \boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f \\ \mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^s \\ \boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f \\ \mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1} \\ \mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f \\ \mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f \\ \left(\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}\right) - \mathbb{E}\left[\left(\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}\right)\right] \end{bmatrix},$$

$$\mathbf{c}^{(3)} \equiv \begin{bmatrix} \mathbf{0}_{n_x \times 1} \\ \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 \\ \left(\sigma\boldsymbol{\eta} \otimes \sigma\boldsymbol{\eta}\right) vec\left(\mathbf{I}_{n_e}\right) \\ \frac{1}{6}\mathbf{h}_{\sigma\sigma\sigma}\sigma^3 \\ \mathbf{0}_{n_x^2 \times 1} \\ \left(\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta}\right) \mathbb{E}\left[\left(\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}\right)\right] \end{bmatrix},$$

$$\mathbf{C}^{(3)} \equiv \begin{bmatrix} \mathbf{g_x} + \frac{3}{6}\mathbf{g}_{\sigma\sigma\mathbf{x}}\sigma^2 & \mathbf{g_x} & \frac{1}{2}\mathbf{G_{xx}} & \mathbf{g_x} & \mathbf{G_{xx}} & \frac{1}{6}\mathbf{G_{xxx}} \end{bmatrix},$$

and

$$\mathbf{d}^{(3)} \equiv \frac{1}{2}\mathbf{g}_{\sigma\sigma}\sigma^2 + \frac{1}{6}\mathbf{g}_{\sigma\sigma\sigma}\sigma^3.$$

A.8 Third Order: Stability

To prove stability:

$$p\left(\lambda\right) = \left|\mathbf{A}^{(3)} - \lambda\mathbf{I}\right|$$

$$= \left|\begin{bmatrix} \mathbf{h_x} - \lambda\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{h_x} - \lambda\mathbf{I} & \frac{1}{2}\mathbf{H_{xx}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2 & \mathbf{0} & \mathbf{0} & \mathbf{h_x} - \lambda\mathbf{I} & \mathbf{H_{xx}} & \frac{1}{6}\mathbf{H_{xxx}} \\ \mathbf{h_x} \otimes \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I} & \mathbf{h_x} \otimes \frac{1}{2}\mathbf{H_{xx}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{h_x} \otimes \mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I} \end{bmatrix}\right|$$

$$= \left|\begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}\right|,$$

where

$$\mathbf{B}_{11} \equiv \begin{bmatrix} \mathbf{h_x} - \lambda\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{h_x} - \lambda\mathbf{I} & \frac{1}{2}\mathbf{H_{xx}} \\ \mathbf{0} & \mathbf{0} & \mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I} \end{bmatrix} \quad \mathbf{B}_{12} \equiv \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$$\mathbf{B}_{21} \equiv \begin{bmatrix} \frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{h_x} \otimes \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \mathbf{B}_{22} \equiv \begin{bmatrix} \mathbf{h_x} - \lambda\mathbf{I} & \mathbf{H_{xx}} & \frac{1}{6}\mathbf{H_{xxx}} \\ \mathbf{0} & \mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I} & \mathbf{h_x} \otimes \frac{1}{2}\mathbf{H_{xx}} \\ \mathbf{0} & \mathbf{0} & \mathbf{h_x} \otimes \mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I} \end{bmatrix},$$

$= |\mathbf{B}_{11}| \, |\mathbf{B}_{22}|$
$= |\mathbf{h_x} - \lambda\mathbf{I}| \, |\mathbf{h_x} - \lambda\mathbf{I}| \, |\mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I}| \, |\mathbf{B}_{22}|$
(using the result from the proof of proposition 1)
$= |\mathbf{h_x} - \lambda\mathbf{I}| \, |\mathbf{h_x} - \lambda\mathbf{I}| \, |\mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I}| \, |\mathbf{h_x} - \lambda\mathbf{I}| \, |\mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I}| \, |\mathbf{h_x} \otimes \mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I}|$
(using the rule on block determinants repeatedly on $\mathbf{B}_{22}$).

The eigenvalue λ solves $p(\lambda) = 0$, which implies:

$$|\mathbf{h_x} - \lambda\mathbf{I}| = 0 \text{ or } |\mathbf{h_x} \otimes \mathbf{h_x} - \lambda\mathbf{I}| = 0 \text{ or } |(\mathbf{h_x} \otimes \mathbf{h_x} \otimes \mathbf{h_x}) - \lambda\mathbf{I}| = 0$$

The absolute values of all eigenvalues to the first problem are strictly less than one by assumption. That is $|\lambda_i| < 1$, $i = 1, 2, ..., n_x$. This is also the case for the second problem, because the eigenvalues to $\mathbf{h_x} \otimes \mathbf{h_x}$ are $\lambda_i \lambda_j$ for $i = 1, 2, ..., n_x$ and $j = 1, 2, ..., n_x$. The same argument ensures that the absolute values of all eigenvalues to the third problem are also less than one. This shows that all eigenvalues of $\mathbf{A}^{(3)}$ have modulus less than one.

A.9 Third Order: Unconditional Second Moments

For the variance, we have

$$\begin{aligned} \mathbb{V}\left[\mathbf{z}_{t+1}^{(3)}\right] &= \mathbf{A}^{(3)} \mathbb{V}\left[\mathbf{z}_t^{(3)}\right] \left(\mathbf{A}^{(3)}\right)' + \mathbf{B}^{(3)} \mathbb{V}\left[\boldsymbol{\xi}_{t+1}^{(3)}\right] \left(\mathbf{B}^{(3)}\right)' \\ &\quad + \mathbf{A}^{(3)} Cov\left[\mathbf{z}_t^{(3)}, \boldsymbol{\xi}_{t+1}^{(3)}\right] \left(\mathbf{B}^{(3)}\right)' + \mathbf{B}^{(3)} Cov\left[\boldsymbol{\xi}_{t+1}^{(3)}, \mathbf{z}_t^{(3)}\right] \left(\mathbf{A}^{(3)}\right)' \end{aligned}$$

Contrary to a second-order approximation, $Cov\left[\boldsymbol{\xi}_{t+1}^{(3)}, \mathbf{z}_t^{(3)}\right] \neq 0$. This is seen as follows:

$$\mathbb{E}\left[\mathbf{z}_t^{(3)} \left(\boldsymbol{\xi}_{t+1}^{(3)}\right)'\right] = \mathbb{E}\left[\begin{bmatrix} \mathbf{x}_t^f \\ \mathbf{x}_t^s \\ \mathbf{x}_t^f \otimes \mathbf{x}_t^f \\ \mathbf{x}_t^{rd} \\ \mathbf{x}_t^f \otimes \mathbf{x}_t^s \\ \mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f \end{bmatrix}\right.$$

$$\times \Big[\ \boldsymbol{\epsilon}_{t+1}' \quad (\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} - vec(\mathbf{I}_{n_e}))' \quad \left(\boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f\right)' \quad \left(\mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1}\right)' \quad (\boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^s)' \quad \left(\boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f\right)'$$
$$\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1}\right)' \quad \left(\mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f\right)' \quad \left(\mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}\right)' \quad \left(\boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f \otimes \boldsymbol{\epsilon}_{t+1}\right)'$$
$$\left.\left(\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \mathbf{x}_t^f\right)' \quad ((\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}) - E[(\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1})])' \ \Big]\right]$$

$$= \left[\begin{matrix} 0_{n_x \times n_e} & 0_{n_x \times n_e^2} & 0_{n_x \times n_e n_x} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_e n_x^2} & 0_{n_x \times n_x^2 n_e} & 0_{n_x \times n_x^2 n_e} \\ 0_{n_x \times n_e} & 0_{n_x \times n_e^2} & 0_{n_x \times n_e n_x} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_e n_x^2} & 0_{n_x \times n_x^2 n_e} & 0_{n_x \times n_x^2 n_e} \\ 0_{n_x^2 \times n_e} & 0_{n_x^2 \times n_e^2} & 0_{n_x^2 \times n_e n_x} & 0_{n_x^2 \times n_x n_e} & 0_{n_x^2 \times n_x n_e} & 0_{n_x^2 \times n_e n_x^2} & 0_{n_x^2 \times n_x^2 n_e} & 0_{n_x^2 \times n_x^2 n_e} \\ 0_{n_x \times n_e} & 0_{n_x \times n_e^2} & 0_{n_x \times n_e n_x} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_e n_x^2} & 0_{n_x \times n_x^2 n_e} & 0_{n_x \times n_x^2 n_e} \\ 0_{n_x^2 \times n_e} & 0_{n_x^2 \times n_e^2} & 0_{n_x^2 \times n_e n_x} & 0_{n_x^2 \times n_x n_e} & 0_{n_x^2 \times n_x n_e} & 0_{n_x^2 \times n_e n_x^2} & 0_{n_x^2 \times n_x^2 n_e} & 0_{n_x^2 \times n_x^2 n_e} \\ 0_{n_x^3 \times n_e} & 0_{n_x^3 \times n_e^2} & 0_{n_x^3 \times n_e n_x} & 0_{n_x^3 \times n_x n_e} & 0_{n_x^3 \times n_x n_e} & 0_{n_x^3 \times n_e n_x^2} & 0_{n_x^3 \times n_x^2 n_e} & 0_{n_x^3 \times n_x^2 n_e} \end{matrix}\right.$$
$$\left.\begin{matrix} R_{1,1} & R_{1,2} & R_{1,3} & 0_{n_x \times n_e^3} \\ R_{2,1} & R_{2,2} & R_{2,3} & 0_{n_x \times n_e^3} \\ R_{3,1} & R_{3,2} & R_{3,3} & 0_{n_x^2 \times n_e^3} \\ R_{4,1} & R_{4,2} & R_{4,3} & 0_{n_x \times n_e^3} \\ R_{5,1} & R_{5,2} & R_{5,3} & 0_{n_x^2 \times n_e^3} \\ R_{6,1} & R_{6,2} & R_{6,3} & 0_{n_x^3 \times n_e^3} \end{matrix}\right]$$

$$= \begin{bmatrix} \mathbf{0} & \mathbf{R} & \mathbf{0} \end{bmatrix}.$$

The $\mathbf{R}$ matrix can easily be computed element-by-element. To compute $\mathbb{V}\left[\boldsymbol{\xi}_{t+1}^{(3)}\right]$, we consider

$$\mathbb{E}\left[\boldsymbol{\xi}_{t+1}^{(3)}\left(\boldsymbol{\xi}_{t+1}^{(3)}\right)'\right] = \mathbb{E}\left[\begin{bmatrix} \boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1} - vec\left(\mathbf{I}_{n_e}\right) \\ \boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f \\ \mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^s \\ \boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\otimes\mathbf{x}_t^f \\ \mathbf{x}_t^f\otimes\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1} \\ \mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f \\ \mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1} \\ \boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f \\ \left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\right) - \mathbb{E}\left[\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\right)\right] \end{bmatrix}\right.$$

$$\times \left[\begin{matrix} \boldsymbol{\epsilon}_{t+1}' & \left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1} - vec\left(\mathbf{I}_{n_e}\right)\right)' & \left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)' & \left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)' & \left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^s\right)' & \left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\otimes\mathbf{x}_t^f\right)' \\ \left(\mathbf{x}_t^f\otimes\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)' & \left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)' & \left(\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\right)' & \left(\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\otimes\boldsymbol{\epsilon}_{t+1}\right)' \\ \left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\otimes\mathbf{x}_t^f\right)' & \left(\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\right) - \mathbb{E}\left[\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\right)\right]\right)' \end{matrix}\right].$$

Note that $\mathbb{V}\left[\boldsymbol{\xi}_{t+1}^{(3)}\right]$ contains $\left(\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\otimes\boldsymbol{\epsilon}_{t+1}\right)$ squared, meaning that $\boldsymbol{\epsilon}_{t+1}$ must have a finite sixth moment for $\mathbb{V}\left[\boldsymbol{\xi}_{t+1}^{(3)}\right]$ to be finite. Again, all elements in $\mathbb{V}\left[\boldsymbol{\xi}_{t+1}^{(3)}\right]$ can be computed element-by-element. For further details, we refer to the paper's Online Appendix, which also discusses how $V\left[\boldsymbol{\xi}_{t+1}^{(3)}\right]$ can be computed in a more memory-efficient manner.

For the auto-covariance, we have

$$\begin{aligned} Cov\left(\mathbf{z}_{t+1}^{(3)},\mathbf{z}_t^{(3)}\right) &= Cov\left(\mathbf{c}^{(3)}+\mathbf{A}^{(3)}\mathbf{z}_t^{(3)}+\mathbf{B}^{(3)}\boldsymbol{\xi}_{t+1}^{(3)},\mathbf{z}_t^{(3)}\right) \\ &= \mathbf{A}^{(3)}Cov\left(\mathbf{z}_t^{(3)},\mathbf{z}_t^{(3)}\right)+\mathbf{B}^{(3)}Cov\left(\boldsymbol{\xi}_{t+1}^{(3)},\mathbf{z}_t^{(3)}\right) \end{aligned}$$

and

$$\begin{aligned} Cov\left(\mathbf{z}_{t+2}^{(3)},\mathbf{z}_t^{(3)}\right) &= Cov\left(\mathbf{c}^{(3)}+\mathbf{A}^{(3)}\mathbf{z}_{t+1}^{(3)}+\mathbf{B}^{(3)}\boldsymbol{\xi}_{t+2}^{(3)},\mathbf{z}_t^{(3)}\right) \\ &= Cov\left(\mathbf{c}^{(3)}+\mathbf{A}^{(3)}\left(\mathbf{c}^{(3)}+\mathbf{A}^{(3)}\mathbf{z}_t^{(3)}+\mathbf{B}^{(3)}\boldsymbol{\xi}_{t+1}^{(3)}\right)+\mathbf{B}^{(3)}\boldsymbol{\xi}_{t+2}^{(3)},\mathbf{z}_t^{(3)}\right) \\ &= Cov\left(\mathbf{c}^{(3)}+\mathbf{A}^{(3)}\mathbf{c}^{(3)}+\left(\mathbf{A}^{(3)}\right)^2\mathbf{z}_t^{(3)}+\mathbf{A}^{(3)}\mathbf{B}^{(3)}\boldsymbol{\xi}_{t+1}^{(3)}+\mathbf{B}^{(3)}\boldsymbol{\xi}_{t+2}^{(3)},\mathbf{z}_t^{(3)}\right) \\ &= Cov\left(\left(\mathbf{A}^{(3)}\right)^2\mathbf{z}_t^{(3)},\mathbf{z}_t^{(3)}\right)+Cov\left(\mathbf{A}^{(3)}\mathbf{B}^{(3)}\boldsymbol{\xi}_{t+1}^{(3)},\mathbf{z}_t^{(3)}\right)+Cov\left(\mathbf{B}^{(3)}\boldsymbol{\xi}_{t+2}^{(3)},\mathbf{z}_t^{(3)}\right) \\ &= \left(\mathbf{A}^{(3)}\right)^2Cov\left(\mathbf{z}_t^{(3)},\mathbf{z}_t^{(3)}\right)+\mathbf{A}^{(3)}\mathbf{B}^{(3)}Cov\left(\boldsymbol{\xi}_{t+1}^{(3)},\mathbf{z}_t^{(3)}\right)+\mathbf{B}^{(3)}Cov\left(\boldsymbol{\xi}_{t+2}^{(3)},\mathbf{z}_t^{(3)}\right). \end{aligned}$$

So, for $s = 1, 2, 3...$

$$Cov\left(\mathbf{z}_{t+s}^{(3)}, \mathbf{z}_t^{(3)}\right) = \left(\mathbf{A}^{(3)}\right)^s \mathbb{V}\left[\mathbf{z}_t^{(3)}\right] + \sum_{j=0}^{s-1} \left(\mathbf{A}^{(3)}\right)^{s-1-j} \mathbf{B}^{(3)} Cov\left(\boldsymbol{\xi}_{t+1+j}^{(3)}, \mathbf{z}_t^{(3)}\right)$$

and we therefore only need to compute $Cov\left(\boldsymbol{\xi}_{t+1+j}^{(3)}, \mathbf{z}_t^{(3)}\right)$:

$$\mathbb{E}\left[\mathbf{z}_t^{(3)}\left(\boldsymbol{\xi}_{t+1+j}^{(3)}\right)'\right] = \mathbb{E}\left[\begin{bmatrix} \mathbf{x}_t^f \\ \mathbf{x}_t^s \\ \mathbf{x}_t^f \otimes \mathbf{x}_t^f \\ \mathbf{x}_t^{rd} \\ \mathbf{x}_t^f \otimes \mathbf{x}_t^s \\ \mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f \end{bmatrix}\right.$$

$$\times \left[\begin{array}{l} \boldsymbol{\epsilon}_{t+1+j}' \quad (\boldsymbol{\epsilon}_{t+1+j} \otimes \boldsymbol{\epsilon}_{t+1+j} - vec(\mathbf{I}_{n_e}))' \quad \left(\boldsymbol{\epsilon}_{t+1+j} \otimes \mathbf{x}_{t+j}^f\right)' \\ \left(\mathbf{x}_{t+j}^f \otimes \boldsymbol{\epsilon}_{t+1+j}\right)' \quad \left(\boldsymbol{\epsilon}_{t+1+j} \otimes \mathbf{x}_{t+j}^s\right)' \quad \left(\boldsymbol{\epsilon}_{t+1+j} \otimes \mathbf{x}_{t+j}^f \otimes \mathbf{x}_{t+j}^f\right)' \\ \left(\mathbf{x}_{t+j}^f \otimes \mathbf{x}_{t+j}^f \otimes \boldsymbol{\epsilon}_{t+1+j}\right)' \quad \left(\mathbf{x}_{t+j}^f \otimes \boldsymbol{\epsilon}_{t+1+j} \otimes \mathbf{x}_{t+j}^f\right)' \\ \left(\mathbf{x}_{t+j}^f \otimes \boldsymbol{\epsilon}_{t+1+j} \otimes \boldsymbol{\epsilon}_{t+1+j}\right)' \quad \left(\boldsymbol{\epsilon}_{t+1+j} \otimes \mathbf{x}_{t+j}^f \otimes \boldsymbol{\epsilon}_{t+1+j}\right)' \\ \left(\boldsymbol{\epsilon}_{t+1+j} \otimes \boldsymbol{\epsilon}_{t+1+j} \otimes \mathbf{x}_{t+j}^f\right)' \quad ((\boldsymbol{\epsilon}_{t+1+j} \otimes \boldsymbol{\epsilon}_{t+1+j} \otimes \boldsymbol{\epsilon}_{t+1+j}) - E\left[(\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1})\right])' \end{array}\right]$$

$$= \left[\begin{array}{cccccccc} 0_{n_x \times n_e} & 0_{n_x \times n_e^2} & 0_{n_x \times n_e n_x} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_e n_x^2} & 0_{n_x \times n_x^2 n_e} & 0_{n_x \times n_x^2 n_e} \\ 0_{n_x \times n_e} & 0_{n_x \times n_e^2} & 0_{n_x \times n_e n_x} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_e n_x^2} & 0_{n_x \times n_x^2 n_e} & 0_{n_x \times n_x^2 n_e} \\ 0_{n_x^2 \times n_e} & 0_{n_x^2 \times n_e^2} & 0_{n_x^2 \times n_e n_x} & 0_{n_x^2 \times n_x n_e} & 0_{n_x^2 \times n_x n_e} & 0_{n_x^2 \times n_e n_x^2} & 0_{n_x^2 \times n_x^2 n_e} & 0_{n_x^2 \times n_x^2 n_e} \\ 0_{n_x \times n_e} & 0_{n_x \times n_e^2} & 0_{n_x \times n_e n_x} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_x n_e} & 0_{n_x \times n_e n_x^2} & 0_{n_x \times n_x^2 n_e} & 0_{n_x \times n_x^2 n_e} \\ 0_{n_x^2 \times n_e} & 0_{n_x^2 \times n_e^2} & 0_{n_x^2 \times n_e n_x} & 0_{n_x^2 \times n_x n_e} & 0_{n_x^2 \times n_x n_e} & 0_{n_x^2 \times n_e n_x^2} & 0_{n_x^2 \times n_x^2 n_e} & 0_{n_x^2 \times n_x^2 n_e} \\ 0_{n_x^3 \times n_e} & 0_{n_x^3 \times n_e^2} & 0_{n_x^3 \times n_e n_x} & 0_{n_x^3 \times n_x n_e} & 0_{n_x^3 \times n_x n_e} & 0_{n_x^3 \times n_e n_x^2} & 0_{n_x^3 \times n_x^2 n_e} & 0_{n_x^3 \times n_x^2 n_e} \end{array}\right.$$

$$\left.\begin{array}{cccc} R_{1,1}^j & R_{1,2}^j & R_{1,3}^j & 0_{n_x \times n_e^3} \\ R_{2,1}^j & R_{2,2}^j & R_{2,3}^j & 0_{n_x \times n_e^3} \\ R_{3,1}^j & R_{3,2}^j & R_{3,3}^j & 0_{n_x^2 \times n_e^3} \\ R_{4,1}^j & R_{4,2}^j & R_{4,3}^j & 0_{n_x \times n_e^3} \\ R_{5,1}^j & R_{5,2}^j & R_{5,3}^j & 0_{n_x^2 \times n_e^3} \\ R_{6,1}^j & R_{6,2}^j & R_{6,3}^j & 0_{n_x^3 \times n_e^3} \end{array}\right]$$

$$= \begin{bmatrix} \mathbf{0} & \mathbf{R}^j & \mathbf{0} \end{bmatrix}.$$

The matrix $\mathbf{R}^j$ can then be computed element-by-element. For further details, see the paper's Online Appendix.

A.10 Third Order: Unconditional Third and Fourth Moments

The proof proceeds as for a second-order approximation. At third order, the only difference is that $\mathbf{v}_{t+1}$ also depends on $\boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1} \otimes \boldsymbol{\epsilon}_{t+1}$. Hence, unconditional third moments exist if $\boldsymbol{\epsilon}_{t+1}$ has a finite ninth moment, and the unconditional fourth moment exists if $\boldsymbol{\epsilon}_{t+1}$ has a finite twelfth moment.

A.11 GIRFs: Second Order

We first note that

$$
\begin{aligned}
\mathbf{x}_{t+l}^{f}\otimes\mathbf{x}_{t+l}^{f} &= \left(\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}+\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\right)\otimes\left(\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}+\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\right)\\
&= \mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\otimes\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}+\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\otimes\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\\
&+\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\otimes\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}+\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\otimes\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}.
\end{aligned}
$$

Next, let

$$
\begin{aligned}
\tilde{\mathbf{x}}_{t+l}^{f}\otimes\tilde{\mathbf{x}}_{t+l}^{f} &= \mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\otimes\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}+\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\otimes\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j}+\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j}\otimes\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\\
&+\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j}\otimes\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j},
\end{aligned}
$$

where we define $\boldsymbol{\delta}_{t+j}$ such that $\boldsymbol{\delta}_{t+1}=\boldsymbol{\nu}+(\mathbf{I}-\mathbf{S})\,\boldsymbol{\epsilon}_{t+1}$ and $\boldsymbol{\delta}_{t+j}=\boldsymbol{\epsilon}_{t+j}$ for $j\neq 1$. This means

$$
\begin{aligned}
&GIRF_{\mathbf{x}^f\otimes\mathbf{x}^f}\left(l,\nu_i,\mathbf{x}_t^f\right)=\mathbb{E}\left[\mathbf{x}_{t+l}^{f}\otimes\mathbf{x}_{t+l}^{f}|\mathbf{x}_t^f,\epsilon_{i,t+1}=\nu_i\right]-\mathbb{E}\left[\mathbf{x}_{t+l}^{f}\otimes\mathbf{x}_{t+l}^{f}|\mathbf{x}_t^f\right]\\
&=\mathbb{E}\left[\tilde{\mathbf{x}}_{t+l}^{f}\otimes\tilde{\mathbf{x}}_{t+l}^{f}|\mathbf{x}_t^f\right]-\mathbb{E}\left[\mathbf{x}_{t+l}^{f}\otimes\mathbf{x}_{t+l}^{f}|\mathbf{x}_t^f\right]\\
&=\mathbb{E}[\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+1}+\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+1}\otimes\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\\
&+\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j}\otimes\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j}\\
&-\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\otimes\sum_{j=1}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}|\mathbf{x}_t^f]\\
&=\mathbb{E}[\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\boldsymbol{\nu}+(\mathbf{I}-\mathbf{S})\,\boldsymbol{\epsilon}_{t+1}\right)+\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\boldsymbol{\nu}+(\mathbf{I}-\mathbf{S})\,\boldsymbol{\epsilon}_{t+1}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\\
&+\left(\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\boldsymbol{\nu}+(\mathbf{I}-\mathbf{S})\,\boldsymbol{\epsilon}_{t+1}\right)+\sum_{j=2}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\right)\otimes\left(\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\boldsymbol{\nu}+(\mathbf{I}-\mathbf{S})\,\boldsymbol{\epsilon}_{t+1}\right)+\sum_{j=2}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\right)\\
&-\left(\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}+\sum_{j=2}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\right)\otimes\left(\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}+\sum_{j=2}^{l}\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}\right)|\mathbf{x}_t^f]\\
&=\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}+\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}+\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}\\
&+\left(\mathbf{h}_{\mathbf{x}}^{l-1}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\right)\left(\mathbb{E}\left[\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\boldsymbol{\epsilon}_{t+1}\otimes\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\boldsymbol{\epsilon}_{t+1}\right]-\mathbb{E}\left[\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}\otimes\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}\right]\right).
\end{aligned}
$$

With $\mathbb{E}\left[\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\boldsymbol{\epsilon}_{t+1}\otimes\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\boldsymbol{\epsilon}_{t+1}\right]=\left(\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\otimes\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\right)vec\left(\mathbf{I}\right)$ and $\mathbb{E}\left[\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}\otimes\sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1}\right]$ $=\left(\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta}\right)vec\left(\mathbf{I}\right)$ we then obtain (23).

A.12 Second-Order Accuracy of Linear IRFs

Let $\mathbf{x}_t^f=\mathbf{0}$ and suppose $\nu\left(i,1\right)=\pm1$ and $\nu\left(j,1\right)=0$ for $i\neq j$. These assumptions imply

$$
\begin{aligned}
\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\sigma\boldsymbol{\eta}\boldsymbol{\nu}+\boldsymbol{\Lambda} &= \sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\sigma\boldsymbol{\eta}\boldsymbol{\nu}+\left(\left(\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\otimes\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\right)-\left(\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta}\right)\right)vec\left(\mathbf{I}\right)\\
&=\left(\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta}\right)\left\{\mathbf{S}\otimes\mathbf{S}+\left(\left(\mathbf{I}-\mathbf{S}\right)\otimes\left(\mathbf{I}-\mathbf{S}\right)\right)-\mathbf{I}\otimes\mathbf{I}\right\}vec\left(\mathbf{I}\right)\\
&=\left(\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta}\right)\left\{2\left(\mathbf{S}\otimes\mathbf{S}\right)-\mathbf{I}\otimes\mathbf{S}-\mathbf{S}\otimes\mathbf{I}\right\}vec\left(\mathbf{I}\right)
\end{aligned}
$$

because $\boldsymbol{\nu}\otimes\boldsymbol{\nu}=\left(\mathbf{S}\otimes\mathbf{S}\right)vec\left(\mathbf{I}\right)$ and $\mathbf{I}_{n_e^2}=\mathbf{I}\otimes\mathbf{I}$**,** where $\mathbf{I}$ has dimension $n_e\times n_e$. Next, let $\mathbf{D}_i\left(i,i\right)=1$ with all remaining elements of $\mathbf{D}_i$ equal to zero. Hence, $\mathbf{I}$ can be written as $\mathbf{I}=\sum_{j=1}^{n_\varepsilon}\mathbf{D}_j$ and $\mathbf{S}=\mathbf{D}_i$. This implies

$$
\begin{aligned}
\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\sigma\boldsymbol{\eta}\boldsymbol{\nu}+\boldsymbol{\Lambda} &= \left(\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta}\right)\left\{-\sum_{\substack{j=1\\ i\neq j}}^{n_\varepsilon}\mathbf{D}_j\otimes\mathbf{D}_i-\sum_{\substack{j=1\\ i\neq j}}^{n_\varepsilon}\mathbf{D}_i\otimes\mathbf{D}_j\right\}vec\left(\sum_{k=1}^{n_\varepsilon}\mathbf{D}_k\right)\\
&=\left(\sigma\boldsymbol{\eta}\otimes\sigma\boldsymbol{\eta}\right)\left\{-\sum_{\substack{j=1\\ i\neq j}}^{n_\varepsilon}\sum_{k=1}^{n_\varepsilon}\left(\mathbf{D}_j\otimes\mathbf{D}_i\right)vec\left(\mathbf{D}_k\right)-\sum_{\substack{j=1\\ i\neq j}}^{n_\varepsilon}\sum_{k=1}^{n_\varepsilon}\left(\mathbf{D}_i\otimes\mathbf{D}_j\right)vec\left(\mathbf{D}_k\right)\right\}
\end{aligned}
$$

$$= (\sigma\boldsymbol{\eta} \otimes \sigma\boldsymbol{\eta}) \left\{ -\sum_{\substack{j=1 \\ i \neq j}}^{n_\varepsilon} \sum_{k=1}^{n_\varepsilon} vec\left(\mathbf{D}_i \mathbf{D}_k \mathbf{D}_j\right) - \sum_{\substack{j=1 \\ i \neq j}}^{n_\varepsilon} \sum_{k=1}^{n_\varepsilon} vec\left(\mathbf{D}_j \mathbf{D}_k \mathbf{D}_i\right) \right\}$$
$$= \mathbf{0}$$

because $\mathbf{D}_i \mathbf{D}_k \mathbf{D}_j$ is only different from the zero matrix when $i = k = j$, but we have $i \neq j$. Thus, $GIRF_{\mathbf{x}^f \otimes \mathbf{x}^f}\left(l, \nu_i, \mathbf{x}_t^f\right) = \mathbf{0}$ and $GIRF_{\mathbf{x}^s}\left(l, \nu_i, \mathbf{x}_t^f\right) = \mathbf{0}$, which proves that GIRFs in a pruned second-order approximation reduces to the IRFs in a linearized solution.

A.13 GIRFs: Third Order

Deriving $GIRF_{\mathbf{x}^f \otimes \mathbf{x}^f \otimes \mathbf{x}^f}\left(j, \boldsymbol{\nu}, \mathbf{x}_t^f\right)$ We first note that

$$\mathbf{x}_{t+l}^f \otimes \mathbf{x}_{t+l}^f \otimes \mathbf{x}_{t+l}^f$$
$$= \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f + \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f$$
$$+ \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f + \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f$$
$$+ \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} + \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}$$
$$+ \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} + \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j} \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+j}.$$

Using the definition of $\boldsymbol{\delta}_{t+j}$ from Appendix A.11, we have

$$\tilde{\mathbf{x}}_{t+l}^f \otimes \tilde{\mathbf{x}}_{t+l}^f \otimes \tilde{\mathbf{x}}_{t+l}^f = \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f + \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f$$
$$+ \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f$$
$$+ \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j} \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f$$
$$+ \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j}$$
$$+ \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j} \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j}$$
$$+ \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j}$$
$$+ \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j} \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j} \otimes \sum_{j=1}^{l} \mathbf{h}_\mathbf{x}^{l-j} \sigma\boldsymbol{\eta}\boldsymbol{\delta}_{t+j}.$$

Simple algebra gives

$$GIRF_{\mathbf{x}^f \otimes \mathbf{x}^f \otimes \mathbf{x}^f}\left(j, \nu_i, \mathbf{x}_t^f\right) = \mathbb{E}\left[\tilde{\mathbf{x}}_{t+l}^f \otimes \tilde{\mathbf{x}}_{t+l}^f \otimes \tilde{\mathbf{x}}_{t+l}^f | \mathbf{x}_t^f\right] - \mathbb{E}\left[\mathbf{x}_{t+l}^f \otimes \mathbf{x}_{t+l}^f \otimes \mathbf{x}_{t+l}^f | \mathbf{x}_t^f\right]$$
$$= \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^{l-1} \sigma\boldsymbol{\eta}\boldsymbol{\nu} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f$$
$$+ \mathbf{h}_\mathbf{x}^{l-1} \sigma\boldsymbol{\eta}\boldsymbol{\nu} \otimes \left(\left(\mathbf{h}_\mathbf{x}^l \otimes \mathbf{h}_\mathbf{x}^l\right)\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right)\right)$$
$$+ \left(\left(\mathbf{h}_\mathbf{x}^l \otimes \mathbf{h}_\mathbf{x}^l\right)\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f\right)\right) \otimes \mathbf{h}_\mathbf{x}^{l-1} \sigma\boldsymbol{\eta}\boldsymbol{\nu}$$
$$+ \left(\mathbf{h}_\mathbf{x}^{l-1} \otimes \mathbf{h}_\mathbf{x}^{l-1}\right)\left[(\sigma\boldsymbol{\eta}\boldsymbol{\nu} \otimes \sigma\boldsymbol{\eta}\boldsymbol{\nu}) + \boldsymbol{\Lambda}\right] \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f$$
$$+ \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \left(\mathbf{h}_\mathbf{x}^{l-1} \otimes \mathbf{h}_\mathbf{x}^{l-1}\right)\left[(\sigma\boldsymbol{\eta}\boldsymbol{\nu} \otimes \sigma\boldsymbol{\eta}\boldsymbol{\nu}) + \boldsymbol{\Lambda}\right]$$
$$+ \mathbf{h}_\mathbf{x}^{l-1} \sigma\boldsymbol{\eta}\boldsymbol{\nu} \otimes \mathbf{h}_\mathbf{x}^l \mathbf{x}_t^f \otimes \mathbf{h}_\mathbf{x}^{l-1} \sigma\boldsymbol{\eta}\boldsymbol{\nu}$$

$$+\left(\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)-\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\otimes\mathbf{h}_{\mathbf{x}}^{l}\mathbf{x}_{t}^{f}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\right)vec\left(\mathbf{I}\right)$$
$$+\left(\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\right)vec\left(\mathbf{I}\right)$$
$$+\left(\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}\right)vec\left(\mathbf{I}\right)$$
$$+\left(\left(\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\right)vec\left(\mathbf{I}\right)$$
$$+\left\{\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\left(\mathbf{I}-\mathbf{S}\right)\right\}\mathbf{m}^{3}\left(\boldsymbol{\epsilon}_{t+1},\boldsymbol{\epsilon}_{t+1},\boldsymbol{\epsilon}_{t+1}\right)$$
$$+\sum_{j=2}^{l}\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\left(\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\otimes\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\right)vec\left(\mathbf{I}\right)$$
$$+\sum_{j=2}^{l}\left(\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\otimes\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\right)vec\left(\mathbf{I}\right)\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}$$
$$+\sum_{j=2}^{l}\left(\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\mathbf{h}_{\mathbf{x}}^{l-j}\sigma\boldsymbol{\eta}\right)vec\left(\mathbf{I}\right)$$
$$-\left(\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\otimes\mathbf{h}_{\mathbf{x}}^{l-1}\sigma\boldsymbol{\eta}\right)\mathbf{m}^{3}\left(\boldsymbol{\epsilon}_{t+1},\boldsymbol{\epsilon}_{t+1},\boldsymbol{\epsilon}_{t+1}\right),$$

where $\mathbf{m}^{3}\left(\boldsymbol{\epsilon}_{t+1},\boldsymbol{\epsilon}_{t+1},\boldsymbol{\epsilon}_{t+1}\right)$ has dimension $n_e^3\times 1$ and contains all the third moments of $\boldsymbol{\epsilon}_{t+1}$.

Deriving $GIRF_{\mathbf{x}^f\otimes\mathbf{x}^s}\left(j,\boldsymbol{\nu},\left(\mathbf{x}_t^f,\mathbf{x}_t^s\right)\right)$ Using the law of motion for $\mathbf{x}_t^f\otimes\mathbf{x}_t^s$, we first note that

$$\mathbf{x}_{t+l}^{f}\otimes\mathbf{x}_{t+l}^{s}=\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l}\left(\mathbf{x}_{t}^{f}\otimes\mathbf{x}_{t}^{s}\right)+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\mathbf{h}_{\mathbf{x}}\otimes\tfrac{1}{2}\mathbf{H}_{\mathbf{xx}}\right)\left(\mathbf{x}_{t+j}^{f}\otimes\mathbf{x}_{t+j}^{f}\otimes\mathbf{x}_{t+j}^{f}\right)$$
$$+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\mathbf{h}_{\mathbf{x}}\otimes\tfrac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^{2}\right)\mathbf{x}_{t+j}^{f}$$
$$+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\sigma\boldsymbol{\eta}\otimes\tfrac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^{2}\right)\boldsymbol{\epsilon}_{t+1+j}$$
$$+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\sigma\boldsymbol{\eta}\otimes\mathbf{h}_{\mathbf{x}}\right)\left(\boldsymbol{\epsilon}_{t+1+j}\otimes\mathbf{x}_{t+j}^{s}\right)$$
$$+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\sigma\boldsymbol{\eta}\otimes\tfrac{1}{2}\mathbf{H}_{\mathbf{xx}}\right)\left(\boldsymbol{\epsilon}_{t+1+j}\otimes\mathbf{x}_{t+i}^{f}\otimes\mathbf{x}_{t+j}^{f}\right)$$

Using the definition of $\boldsymbol{\delta}_{t+j}$ from Appendix A.11, we obtain

$$\tilde{\mathbf{x}}_{t+l}^{f}\otimes\tilde{\mathbf{x}}_{t+l}^{s}=\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l}\left(\mathbf{x}_{t}^{f}\otimes\mathbf{x}_{t}^{s}\right)+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\mathbf{h}_{\mathbf{x}}\otimes\tfrac{1}{2}\mathbf{H}_{\mathbf{xx}}\right)\left(\tilde{\mathbf{x}}_{t+j}^{f}\otimes\tilde{\mathbf{x}}_{t+j}^{f}\otimes\tilde{\mathbf{x}}_{t+j}^{f}\right)$$
$$+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\mathbf{h}_{\mathbf{x}}\otimes\tfrac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^{2}\right)\tilde{\mathbf{x}}_{t+j}^{f}$$
$$+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\sigma\boldsymbol{\eta}\otimes\tfrac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^{2}\right)\boldsymbol{\delta}_{t+1+j}$$
$$+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\sigma\boldsymbol{\eta}\otimes\mathbf{h}_{\mathbf{x}}\right)\left(\boldsymbol{\delta}_{t+1+j}\otimes\tilde{\mathbf{x}}_{t+j}^{s}\right)$$
$$+\sum_{j=0}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\sigma\boldsymbol{\eta}\otimes\tfrac{1}{2}\mathbf{H}_{\mathbf{xx}}\right)\left(\boldsymbol{\delta}_{t+1+j}\otimes\tilde{\mathbf{x}}_{t+j}^{f}\otimes\tilde{\mathbf{x}}_{t+j}^{f}\right)$$

Simple algebra then implies

$$GIRF_{\mathbf{x}^f\otimes\mathbf{x}^s}\left(j,\nu_i,\left(\mathbf{x}_t^f,\mathbf{x}_t^s\right)\right)=\sum_{j=1}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\mathbf{h}_{\mathbf{x}}\otimes\tfrac{1}{2}\mathbf{H}_{\mathbf{xx}}\right)GIRF_{\mathbf{x}^f\otimes\mathbf{x}^f\otimes\mathbf{x}^f}\left(j,\nu_i,\mathbf{x}_t^f\right)$$
$$+\sum_{j=1}^{l-1}\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1-j}\left(\mathbf{h}_{\mathbf{x}}\otimes\tfrac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^{2}\right)GIRF_{\mathbf{x}^f}\left(j,\nu_i\right)$$
$$+\left(\mathbf{h}_{\mathbf{x}}\otimes\mathbf{h}_{\mathbf{x}}\right)^{l-1}\left(\sigma\boldsymbol{\eta}\boldsymbol{\nu}\otimes\left(\mathbf{h}_{\mathbf{x}}\mathbf{x}_{t}^{s}+\tfrac{1}{2}\mathbf{H}_{\mathbf{xx}}\left(\mathbf{x}_{t}^{f}\otimes\mathbf{x}_{t}^{f}\right)+\tfrac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^{2}\right)\right)$$

A.14 Proof of Proposition 1

Second Order Let us first consider the state variables. Provided that the unpruned state-space system is stable, we know that

$$\mathbf{x}_{t+1}^{(2)} = \mathbf{h}_{\mathbf{x}}\mathbf{x}_t^{(2)} + \frac{1}{2}\mathbf{H}_{\mathbf{xx}}\left(\mathbf{x}_t^{(2)} \otimes \mathbf{x}_t^{(2)}\right) + \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 + \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1} + O\left(\sigma^3\right),$$

i.e., the errors are of third order when $\sigma \longrightarrow 0$. Comparing the pruned state-space system to this expression, we obtain

$$\mathbf{x}_{t+1}^{f} + \mathbf{x}_{t+1}^{s} - \mathbf{x}_{t+1}^{(2)} = \mathbf{h}_{\mathbf{x}}\left(\mathbf{x}_t^{f} + \mathbf{x}_t^{s} - \mathbf{x}_t^{(2)}\right) + \frac{1}{2}\mathbf{H}_{\mathbf{xx}}\left(\mathbf{x}_t^{f} \otimes \mathbf{x}_t^{f} - \mathbf{x}_t^{(2)} \otimes \mathbf{x}_t^{(2)}\right) + O\left(\sigma^3\right).$$

To show that $\mathbf{x}_t^{f} \otimes \mathbf{x}_t^{f} - \mathbf{x}_t^{(2)} \otimes \mathbf{x}_t^{(2)} = O\left(\sigma^3\right)$, algebra gives

$$\begin{aligned}\mathbf{x}_{t+1}^{f} \otimes \mathbf{x}_{t+1}^{f} - \mathbf{x}_{t+1}^{(2)} \otimes \mathbf{x}_{t+1}^{(2)} &= \left(\mathbf{h}_{\mathbf{x}} \otimes \mathbf{h}_{\mathbf{x}}\right)\left(\mathbf{x}_t^{f} \otimes \mathbf{x}_t^{f} - \mathbf{x}_t^{(2)} \otimes \mathbf{x}_t^{(2)}\right) + \left(\mathbf{h}_{\mathbf{x}} \otimes \boldsymbol{\eta}\right)\left(\left(\mathbf{x}_t^{f} - \mathbf{x}_t^{(2)}\right) \otimes \sigma\boldsymbol{\epsilon}_{t+1}\right) \\ &\quad + \left(\boldsymbol{\eta} \otimes \mathbf{h}_{\mathbf{x}}\right)\left(\sigma\boldsymbol{\epsilon}_{t+1} \otimes \left(\mathbf{x}_t^{f} - \mathbf{x}_t^{(2)}\right)\right) + O\left(\sigma^3\right)\end{aligned}$$

We know that $\mathbf{x}_t^{f} - \mathbf{x}_t^{(2)} = O\left(\sigma^2\right)$ and, therefore, $\sigma\left(\mathbf{x}_t^{f} - \mathbf{x}_t^{(2)}\right) = O\left(\sigma^3\right)$. This shows that $\mathbf{x}_t^{f} \otimes \mathbf{x}_t^{f} - \mathbf{x}_t^{(2)} \otimes \mathbf{x}_t^{(2)} = O\left(\sigma^3\right)$, given that all eigenvalues of $\mathbf{h}_{\mathbf{x}}$ have modulus less than one. This in turn shows that $\mathbf{x}_t^{f} + \mathbf{x}_t^{s} - \mathbf{x}_t^{(2)} = O\left(\sigma^3\right)$. For the controls we easily obtain

$$\mathbf{y}_t^{s} - \mathbf{y}_t^{(2)} = \mathbf{g}_{\mathbf{x}}\left(\mathbf{x}_t^{f} + \mathbf{x}_t^{s} - \mathbf{x}_t^{(2)}\right) + \frac{1}{2}\mathbf{G}_{\mathbf{xx}}\left(\mathbf{x}_t^{f} \otimes \mathbf{x}_t^{f} - \mathbf{x}_t^{(2)} \otimes \mathbf{x}_t^{(2)}\right) + O\left(\sigma^3\right).$$

Given that $\mathbf{x}_t^{f} + \mathbf{x}_t^{s} - \mathbf{x}_t^{(2)} = O\left(\sigma^3\right)$ and $\mathbf{x}_t^{f} \otimes \mathbf{x}_t^{f} - \mathbf{x}_t^{(2)} \otimes \mathbf{x}_t^{(2)} = O\left(\sigma^3\right)$, we have $\mathbf{y}_t^{s} - \mathbf{y}_t^{(2)} = O\left(\sigma^3\right)$ as desired.

Third Order Let us first consider the state variables. Provided that the unpruned state-space system is stable, we know that

$$\begin{aligned}\mathbf{x}_{t+1}^{(3)} &= \left(\mathbf{h}_{\mathbf{x}} + \frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2\right)\mathbf{x}_t^{(3)} + \frac{1}{2}\mathbf{H}_{\mathbf{xx}}\left(\mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) + \frac{1}{6}\mathbf{H}_{\mathbf{xx}}\left(\mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) \\ &\quad + \frac{1}{2}\mathbf{h}_{\sigma\sigma}\sigma^2 + \frac{1}{6}\mathbf{h}_{\sigma\sigma\sigma}\sigma^3 + \sigma\boldsymbol{\eta}\boldsymbol{\epsilon}_{t+1} + O\left(\sigma^4\right),\end{aligned}$$

that is, the errors are of fourth order when $\sigma \longrightarrow 0$. Comparing the pruned state-space system to this expression, we have

$$\begin{aligned}\mathbf{x}_{t+1}^{f} + \mathbf{x}_{t+1}^{s} + \mathbf{x}_{t+1}^{rd} - \mathbf{x}_{t+1}^{(3)} &= \mathbf{h}_{\mathbf{x}}\left(\mathbf{x}_t^{f} + \mathbf{x}_t^{s} + \mathbf{x}_t^{rd} - \mathbf{x}_t^{(3)}\right) \\ &\quad + \frac{1}{2}\mathbf{H}_{\mathbf{xx}}\left(\left(\mathbf{x}_t^{f} \otimes \mathbf{x}_t^{f}\right) + \left(\mathbf{x}_t^{f} \otimes \mathbf{x}_t^{s}\right) + \left(\mathbf{x}_t^{s} \otimes \mathbf{x}_t^{f}\right) - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) \\ &\quad + \frac{1}{6}\mathbf{H}_{\mathbf{xxx}}\left(\mathbf{x}_t^{f} \otimes \mathbf{x}_t^{f} \otimes \mathbf{x}_t^{f} - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) + \frac{3}{6}\mathbf{h}_{\sigma\sigma\mathbf{x}}\sigma^2\left(\mathbf{x}_t^{f} - \mathbf{x}_t^{(3)}\right)\end{aligned}$$

We know that $\mathbf{x}_t^f - \mathbf{x}_t^{(3)} = O\left(\sigma^2\right)$, and therefore $\sigma^2\left(\mathbf{x}_t^f - \mathbf{x}_t^{(3)}\right) = O\left(\sigma^4\right)$. We clearly also have $\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} = O\left(\sigma^4\right)$. For the final term, some algebra implies

$$\begin{aligned}
& \mathbf{x}_{t+1}^f \otimes \mathbf{x}_{t+1}^f + \mathbf{x}_{t+1}^f \otimes \mathbf{x}_{t+1}^s + \mathbf{x}_{t+1}^s \otimes \mathbf{x}_{t+1}^f - \mathbf{x}_{t+1}^{(3)} \otimes \mathbf{x}_{t+1}^{(3)} \\
= \; & \left(\mathbf{h_x} \otimes \mathbf{h_x}\right)\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f + \mathbf{x}_t^f \otimes \mathbf{x}_t^s + \mathbf{x}_t^s \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) \\
& + \left(\mathbf{h_x} \otimes \boldsymbol{\eta\epsilon}_{t+1} + \boldsymbol{\eta\epsilon}_{\mathbf{t+1}} \otimes \mathbf{h_x}\right)\left(\left(\mathbf{x}_t^f + \mathbf{x}_t^s - \mathbf{x}_t^{(3)}\right)\sigma\right) \\
& + \left(\mathbf{h_x} \otimes \frac{1}{2}\mathbf{H_{xx}} + \frac{1}{2}\mathbf{H_{xx}} \otimes \mathbf{h_x}\right)\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) \\
& + \left(\mathbf{h_x} \otimes \frac{1}{2}\mathbf{h}_{\sigma\sigma} + \frac{1}{2}\mathbf{h}_{\sigma\sigma} \otimes \mathbf{h_x}\right)\sigma^2\left(\mathbf{x}_t^f - \mathbf{x}_t^{(3)}\right) \\
& + \left(\boldsymbol{\eta\epsilon}_{t+1} \otimes \frac{1}{2}\mathbf{H_{xx}} + \frac{1}{2}\mathbf{H_{xx}} \otimes \boldsymbol{\eta\epsilon}_{t+1}\right)\left(\sigma\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right)\right)
\end{aligned}$$

We know that $\mathbf{x}_t^f + \mathbf{x}_t^s - \mathbf{x}_t^{(3)} = O\left(\sigma^3\right)$, so $\left(\mathbf{x}_t^f + \mathbf{x}_t^s - \mathbf{x}_t^{(3)}\right)\sigma = O\left(\sigma^4\right)$. Similarly, $\mathbf{x}_t^f - \mathbf{x}_t^{(3)} = O\left(\sigma^2\right)$, so $\sigma^2\left(\mathbf{x}_t^f - \mathbf{x}_t^{(3)}\right) = O\left(\sigma^4\right)$. Finally, $\mathbf{x}_t^f \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} = O\left(\sigma^3\right)$, and, thus, $\sigma\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) = O\left(\sigma^4\right)$. Hence, $\mathbf{x}_t^f \otimes \mathbf{x}_t^f + \mathbf{x}_t^f \otimes \mathbf{x}_t^s + \mathbf{x}_t^s \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} = O\left(\sigma^4\right)$ and, therefore, $\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd} - \mathbf{x}_t^{(3)} = O\left(\sigma^4\right)$, given that all eigenvalues of $\mathbf{h_x}$ have modulus less than one. For the controls we easily obtain

$$\begin{aligned}
\mathbf{y}_t^{rd} - \mathbf{y}_t^{(3)} \; = \; & \mathbf{g_x}\left(\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd} - \mathbf{x}_t^{(3)}\right) + \frac{1}{2}\mathbf{G_{xx}}\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f + \mathbf{x}_t^f \otimes \mathbf{x}_t^s + \mathbf{x}_t^s \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right) \\
& + \frac{1}{6}\mathbf{G_{xxx}}\left(\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)}\right).
\end{aligned}$$

Given that $\mathbf{x}_t^f + \mathbf{x}_t^s + \mathbf{x}_t^{rd} - \mathbf{x}_t^{(3)} = O\left(\sigma^4\right)$, $\mathbf{x}_t^f \otimes \mathbf{x}_t^f + \mathbf{x}_t^f \otimes \mathbf{x}_t^s + \mathbf{x}_t^s \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} = O\left(\sigma^4\right)$, and $\mathbf{x}_t^f \otimes \mathbf{x}_t^f \otimes \mathbf{x}_t^f - \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} \otimes \mathbf{x}_t^{(3)} = O\left(\sigma^4\right)$, we clearly have $\mathbf{y}_t^{rd} - \mathbf{y}_t^{(3)} = O\left(\sigma^4\right)$ as desired.

A.15 An Assessment of Accuracy

As a supplement to the accuracy studies mentioned in the main text, we briefly consider the performance of pruning on the stochastic neoclassical growth model, the workhorse of modern macroeconomics. Here, a representative household selects a sequence of consumption c_t and investment i_t to solve

$$\begin{gathered}
\max_{\{c_t, i_t\}_{t=0}^{\infty}} \mathbb{E}_t \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma} \\
\text{s.t. } c_t + i_t = a_t k_t^{\alpha} \\
k_{t+1} = (1-\delta) k_t + i_t \\
\log a_{t+1} = \rho_a \log a_t + \sigma_a \epsilon_{a,t+1}, \; \epsilon_{a,t+1} \sim \mathcal{NID}(0,1).
\end{gathered}$$

The calibration is conventional: $\beta = 0.99$, $\delta = 0.025$, $\alpha = 0.36$, $\alpha = 0.36$, $\rho_a = 0.98$, $\sigma_a = 0.01$. We consider three cases for γ: $2, 5$, and 25. The first value, $\gamma = 2$, is a standard calibration for risk aversion. The value $\gamma = 5$ is at the high end of estimated risk aversions. Finally, $\gamma = 25$ is an extreme calibration well beyond the values compatible with micro-evidence. We assess the accuracy of the pruned and unpruned state-space system based on a fourth-order projection approximation, which is sufficiently accurate to be used as a stand-in for the exact solution. In particular, we measure the root mean squared errors (RMSE) between each perturbation solution and the projection solution.[25]

Table A.1 shows that the accuracy of the pruned and unpruned state-space systems at third order is roughly the same for $\gamma = 2$ (with a trivially small advantage for the pruned solution). The pruned state-space system is a bit less accurate than the unpruned approximation when $\gamma = 5$. There is more deterioration of accuracy for the pruned solution when $\gamma = 25$. The results are, however, biased against pruning in this case, since the unpruned solution explodes in 53 out of the 500 simulated sample paths. The reason why the pruned solution loses some accuracy as γ increases is that precautionary behavior becomes larger and the terms eliminated by pruning may carry information relevant to the solution. But, of course, these terms also generate explosive paths. Note also that the pruned state-space system clearly outperforms the standard first-order approximation for all the considered values of γ.

Table A.1: Stochastic Neoclassical Growth Model: Accuracy Test

Approximation errors are computed based on an accurate fourth-order projection solution. Moments are computed from 500 sample paths of length 4,500 observations with a burn-in of 500 periods. The regression reads $\left|\hat{c}_t^{per} - \hat{c}_t^{proj}\right| = \alpha^{(s)} + \beta_k^{(s)} \left|\hat{k}_t^{proj(s)}\right| + \beta_a^{(s)} \left|\hat{a}_t^{(s)}\right| + \rho \left|\hat{c}_{t-1}^{per} - \hat{c}_{t-1}^{proj}\right| + \varepsilon_t^{(s)}$, where *per* refers to a perturbation approximation and *proj* to the projection solution. The circumflex denotes percentage deviation from steady state. N_{explode} denotes the number of explosive sample paths. The reported values are averages across non-explosive sample paths.

	$RMSE \times 10^3$	$\alpha \times 10^3$	$\beta_k \times 10^3$	$\beta_a \times 10^3$	ρ	N_{explode}
$\gamma = 2$						
1st order	1.3069	0.0032	0.1050	-0.1498	0. 9959	0
3rd order: no pruning	0.1527	0.0111	-0.0721	0.2024	0.8978	0
3rd order: pruning	0.1522	0.0110	-0.0719	0.2023	0.8982	0
$\gamma = 5$						
1st order	3.4282	0.0464	0.1592	-0.7397	0.9897	0
3rd order: no pruning	0.0840	0.0037	0.0014	-0.0018	0. 9444	0
3rd order: pruning	0.0949	0.0030	0.0097	-0.0133	0.9525	0
$\gamma = 25$						
1st order	36.8920	0.6557	-0.7711	-0.4707	0.9889	0
3rd order: no pruning	17.7337	-0.3529	0.6016	12.4716	0. 9685	53
3rd order: pruning	30.5984	0.3958	0.0132	-0.6874	0.9873	0

To obtain further insight into the accuracy of pruning, we regress the approximation errors on the distance of each state variable from the steady state and lagged pricing errors (needed to get a well-specified regression). For $\gamma = 25$, the intercepts in these regressions are higher for the pruned than the unpruned state-space system, whereas the slope coefficients are smaller with pruning. Hence, for more non-linear models, the unpruned state-space system is more accurate around the steady state, but its performance deteriorates faster away from the steady state compared to the pruned system. Of course, we should emphasize once more that the results in Table A.1 depend on

[25] We also checked the log case $\gamma = 1$. Given how linear the model is when we have a log utility function, the pruned and unpruned solutions are nearly identical and they display the same level of accuracy.

the model considered.

A.16 Making the DSGE Model Stationary

We eliminate all trending variables in the model by adopting the transformation $C_t \equiv \frac{c_t}{z_t^*}$, $R_t^k \equiv \Upsilon_t r_t^k$, $Q_t \equiv \Upsilon_t q_t$, $I_t \equiv \frac{i_t}{\Upsilon_t z_t^*}$, $W_t \equiv \frac{w_t}{z_t^*}$, $Y_t \equiv \frac{y_t}{z_t^*}$, $K_{t+1} \equiv \frac{k_{t+1}}{\Upsilon_t^{\frac{1}{1-\theta}} z_t} = \frac{k_{t+1}}{\Upsilon_t z_t^*}$, and $\Lambda_t \equiv \frac{\lambda_t}{m_t (z_t^*)^{-1}}$. Here q_t is the Lagrangian multiplier for the law of motion for capital and m_t for the value function in equation (28); see Rudebusch and Swanson (2012). Hence, $\mu_{\lambda,t+1} \equiv \frac{\lambda_{t+1}}{\lambda_t} = \frac{\Lambda_{t+1}}{\Lambda_t}\mu_{z^*,t+1}^{-1}\left(\mathbb{E}_t\left[V_{t+1}^{1-\phi_3}\right]\right)^{\frac{\phi_3}{1-\phi_3}} V_{t+1}^{-\phi_3}$, and the value of ψ that eliminates capital adjustment costs in the steady state is therefore given by $\psi \equiv \frac{I_{ss}}{K_{ss}}\mu_{\Upsilon,ss}\mu_{z^*,ss}$.

The transformed equilibrium conditions are summarized below:

Eq.	**The Households**
1	$V_t = \left[\frac{d_t}{1-\phi_2}\left(\left(C_t - bC_{t-1}\mu_{z^*,t}^{-1}\right)^{1-\phi_2} - 1\right) + d_t\phi_0\frac{(1-h_t)^{1-\phi_1}}{1-\phi_1}\right] + \beta\left(\mathbb{E}_t\left[V_{t+1}^{1-\phi_3}\right]\right)^{\frac{1}{1-\phi_3}}$
2	$\Lambda_t = d_t\left(C_t - bC_{t-1}\mu_{z^*,t}^{-1}\right)^{-\phi_2}$
	$-b\beta\mathbb{E}_t\left[\left(\frac{\left[E_t\left[V_{t+1}^{1-\phi_3}\right]\right]^{\frac{1}{1-\phi_3}}}{V_{t+1}(s)}\right)^{\phi_3} d_{t+1}\left(C_{t+1} - bC_t\mu_{z^*,t+1}^{-1}\right)^{-\phi_2}\left(\mu_{z^*,t+1}\right)^{-1}\right]$
3	$Q_t = \mathbb{E}_t\frac{\beta\mu_{\lambda,t+1}}{\mu_{\Upsilon,t+1}}[R_{t+1}^k + Q_{t+1}(1-\delta) - Q_{t+1}\frac{\kappa}{2}\left(\frac{I_{t+1}}{K_{t+1}}\mu_{\Upsilon,t+1}\mu_{z^*,t+1} - \frac{I_{ss}}{k_{ss}}\mu_{\Upsilon,ss}\mu_{z^*,ss}\right)^2$
	$+Q_{t+1}\kappa\left(\frac{I_{t+1}}{K_{t+1}}\mu_{\Upsilon,t+1}\mu_{z^*,t+1} - \frac{I_{ss}}{k_{ss}}\mu_{\Upsilon,ss}\mu_{z^*,ss}\right)\frac{I_{t+1}}{K_{t+1}}\mu_{\Upsilon,t+1}\mu_{z^*,t+1}]$
4	$d_t\phi_0(1-h_t)^{-\phi_1} = \Lambda_t W_t$
5	$1 = Q_t\left(1 - \kappa\left(\frac{I_t}{k_t}\mu_{\Upsilon,t}\mu_{z^*,t} - \frac{I_{ss}}{K_{ss}}\mu_{\Upsilon,ss}\mu_{z^*,ss}\right)\right)$
6	$1 = \mathbb{E}_t\left[\beta\mu_{\lambda t+1}\frac{\exp\{r_t^b\}}{\pi_{t+1}}\right]$
	The Firms
7	$mc_t a_t\theta\mu_{\Upsilon,t}\mu_{z,t}^{1-\theta}K_t^{\theta-1}h_t^{1-\theta} = R_t^k$
8	$mc_t(1-\theta)a_t\mu_{\Upsilon,t}^{\frac{-\theta}{1-\theta}}\mu_{z,t}^{-\theta}K_t^{\theta}h_t^{-\theta} = W_t$
9	$\frac{(\eta-1)}{\eta}X_t^2 = Y_t mc_t\tilde{p}_t^{-\eta-1} + \mathbb{E}_t\left[\alpha\beta\mu_{\lambda,t+1}\left(\frac{\tilde{p}_t}{\tilde{p}_{t+1}}\right)^{-\eta-1}\left(\frac{1}{\pi_{t+1}}\right)^{-\eta}\frac{(\eta-1)}{\eta}X_{t+1}^2\mu_{z^*,t+1}\right]$
10	$X_t^2 = Y_t\tilde{p}_t^{-\eta} + \mathbb{E}_t\left[\alpha\beta\mu_{\lambda,t+1}\left(\frac{\tilde{p}_t}{\tilde{p}_{t+1}}\right)^{-\eta}\left(\frac{1}{\pi_{t+1}}\right)^{1-\eta}X_{t+1}^2\mu_{z^*,t+1}\right]$
11	$1 = (1-\alpha)\tilde{p}_t^{1-\eta} + \alpha\left(\frac{1}{\pi_t}\right)^{1-\eta}$
	The Financial Intermediary
12	$r_t^b = r_t + \omega \times xhr_{t,L}$
13	$P_{t,1} = \frac{1}{\exp\{r_t\}}$
14	$P_{t,k} = \mathbb{E}_t\left[\beta\mu_{\lambda,t+1}\frac{1}{\pi_{t+1}}P_{t+1,k-1}\right]$ for $k = 2, 3, ..., K$

	The Central Bank
15	$r_t = r_{ss}(1-\rho_r) + \rho_r r_{t-1} + (1-\rho_r)\left(\beta_\pi \log\left(\frac{\pi_t}{\pi_{ss}}\right) + \beta_y \log\left(\frac{Y_t}{Y_{ss}}\right)\right)$
	$+ (1-\rho_r)\beta_{xhr}(xhr_{t,L} - X_{t,L})$
16	$X_{t,L} = (1-\gamma) xhr_{t,L} + \gamma \mathbb{E}_t [X_{t+1,L}]$
	Other relations
17	$a_t \left(K_t \mu_{\Upsilon,t}^{\frac{-1}{1-\theta}} \mu_{z,t}^{-1}\right)^{\theta} h_t^{1-\theta} = Y_t s_{t+1}$
18	$s_{t+1} = (1-\alpha)\tilde{p}_t^{-\eta} + \alpha \pi_t^{\eta} s_t$
19	$K_{t+1} = (1-\delta) K_t (\mu_{\Upsilon,t}\mu_{z^*,t})^{-1} + I_t$
	$-K_t (\mu_{\Upsilon,t}\mu_{z^*,t})^{-1} \frac{\kappa}{2}\left(\frac{I_i}{K_t}\mu_{\Upsilon,t}\mu_{z^*,t} - \frac{I_{ss}}{K_{ss}}\mu_{\Upsilon,ss}\mu_{z^*,ss}\right)^2$
20	$Y_t = C_t + I_t + g_t$
21	$\mu_{z^*,t} \equiv \mu_{\Upsilon,t}^{\theta/(1-\theta)} \mu_{z,t}$
	Exogenous processes
22	$\log(\mu_{z,t}) = \log(\mu_{z,ss})$ and $z_{t+1} \equiv z_t \mu_{z,t+1}$ (i.e. a deterministic trend)
23	$\log(\mu_{\Upsilon,t}) = \log \mu_{\Upsilon,ss}$ and $\Upsilon_{t+1} \equiv \Upsilon_t \mu_{\Upsilon,t+1}$ (i.e. a deterministic trend)
24	$\log a_{t+1} = \rho_a \log a_t + \sigma_a \epsilon_{a,t+1}$
25	$\log\left(\frac{G_{t+1}}{G_{ss}}\right) = \rho_G \log\left(\frac{G_t}{G_{ss}}\right) + \sigma_G \epsilon_{G,t+1}$
26	$\log d_{t+1} = \sigma_d \epsilon_{d,t+1}$

From these equilibrium conditions, it is straightforward to derive a closed-form solution for the steady state of the model.

A.17 An Alternative Interpretation

The deposit rate r_t^b only enters in equations 6, 12, and 15 of the model summary in Appendix A.16. But note that $r_t = r_t^b - \omega \times xhr_{t,L}$ and when substituted into the Taylor rule we get

$$\begin{aligned} r_t^b &= (1-\rho_r) r_{ss} + \rho_r r_{t-1}^b + (1-\rho_r)\left(\beta_\pi \log\left(\frac{\pi_t}{\pi_{ss}}\right) + \beta_y \log\left(\frac{y_t}{z_t^* Y_{ss}}\right)\right) \\ &\quad + \omega \times xhr_{t,L} - \rho_r \omega \times xhr_{t-1,L} + (1-\rho_r)\beta_{xhr}(xhr_{t,L} - X_{t,L}). \end{aligned}$$

Given this substitution, r_t^b only enters in equations 6 and 15 of the model summary in Appendix A.16. This implies that our model is equivalent to a standard New Keynesian model with market completeness, but with a Taylor rule for r_t^b that depends on past and current values of excess holding period return on the long bond.

A.18 An Efficient Perturbation Approximation

To formally present our efficient perturbation approximation, consider the decomposition $\mathbf{y}_t \equiv \begin{bmatrix} (\mathbf{y}_t^{macro})' & (\mathbf{y}_t^{bonds})' \end{bmatrix}$ and similarly for all derivatives of $\mathbf{g}(\mathbf{x}_t, \sigma)$. Here, $\mathbf{y}_t^{macro}$ refers to the control variables needed to solve the model without feedback effects from long-term bond prices to the real economy (when $\omega = 0$ and $\beta_{xhr} = 0$), whereas $\mathbf{y}_t^{bonds}$ denotes the remaining variables related to pricing government bonds and computing excess holding period returns. Our three-step perturbation approximation is then:

Step 1: Solve for $(\mathbf{g}_{\mathbf{x}}^{macro}, \mathbf{G}_{\mathbf{xx}}^{macro}, \mathbf{G}_{\mathbf{xxx}}^{macro})$ and $(\mathbf{h}_{\mathbf{x}}, \mathbf{H}_{\mathbf{xx}}, \mathbf{H}_{\mathbf{xxx}})$ by a standard perturbation algorithm using a version of our model *without* feedback effects from government bonds to the real economy. This version of our model has only 11 control variables and 18 equations and is solved using the `Matlab` codes of Binning (2013).

Step 2: Use the perturbation algorithm of Andreasen and Zabczyk (2015) to recursively solve for $\left(\mathbf{g}_{\mathbf{x}}^{bonds}, \mathbf{G}_{\mathbf{xx}}^{bonds}, \mathbf{G}_{\mathbf{xxx}}^{bonds}\right)$, given the derivatives obtained in Step 1.

Step 3: With the derivatives obtained in Steps 1 and 2, solve for $(\mathbf{g}_{\sigma\sigma}, \mathbf{g}_{\sigma\sigma\mathbf{x}}, \mathbf{g}_{\sigma\sigma\sigma})$ and $(\mathbf{h}_{\sigma\sigma}, \mathbf{h}_{\sigma\sigma\mathbf{x}}, \mathbf{h}_{\sigma\sigma\sigma})$ by the standard perturbation algorithm when using the full model with 54 control variables and 61 equations.

To maximize the efficiency of our perturbation algorithm, steps 2 and 3 are computed using a `FORTRAN` implementation accessible via MEX files in `Matlab`.

A.19 Data for the Application

We use data from the Federal Reserve Bank of St. Louis covering the period 1961.Q3 to 2007.Q4, giving a total of 186 observations. The annualized growth rate in consumption is calculated from real consumption expenditures (PCECC96). The series for real private fixed investment (FPIC96) is used to calculate the growth rate in investment. Both growth rates are expressed in per capita terms based on the total population in the US. The ratio of government spending to output is computed as government consumption expenditures and investments divided by gross domestic production. The annual inflation rate is for consumer prices. The 3-month nominal interest rate is measured by the rate in the secondary market (TB3MS), and the 10-year nominal rate is from Gürkaynak, Sack and Wright (2007). As in Rudebusch and Swanson (2012), observations for the 10-year interest rate from 1961.Q3 to 1971.Q3 are calculated by extrapolation of the estimated curves in Gürkaynak, Sack and Wright (2007). All moments related to interest rates are expressed in annualized terms. Finally, we use average weekly hours of production and non-supervisory employees in manufacturing (AWHMAN) as provided by the Bureau of Labor Statistics. The series is normalized by dividing it by five times 24 hours, giving a mean level of 0.34.

A.20 Approximate Expression for Excess Holding Period Return

First,

$$xhr_{t,L} = \mathbb{E}_t \left[\log\left(P_{t+1,L-1}\right) - \log\left\{\mathbb{E}_t\left[M_{t,t+1}\right]\mathbb{E}_t\left[P_{t+1,L-1}\right] + Cov_t\left(M_{t,t+1}, P_{t+1,L-1}\right)\right\}\right] - r_t.$$

To first order,

$$\log\left(x_t + y_t\right) \approx \log\left(x_{ss} + y_{ss}\right) + \frac{1}{x_{ss} + y_{ss}}\left(x_t - x_{ss}\right) + \frac{1}{x_{ss} + y_{ss}}\left(y_t - y_{ss}\right)$$

and let $x_t = \mathbb{E}_t\left[M_{t,t+1}\right]\mathbb{E}_t\left[P_{t+1,L-1}\right]$ and $y_t = Cov_t\left(M_{t,t+1}, P_{t+1,L-1}\right)$, implying that $x_t + y_t = P_{t,L}$.

Hence,

$$\begin{aligned} xhr_{t,L} &\approx \mathbb{E}_t\left[\log\left(P_{t+1,L-1}\right)\right] - \log P_{ss,L} - \tfrac{1}{P_{ss,L}}\left(\mathbb{E}_t\left[M_{t,t+1}\right]\mathbb{E}_t\left[P_{t+1,L-1}\right] - M_{ss,ss+1}P_{ss,L-1}\right) \\ &\quad - \tfrac{1}{P_{ss,L}} Cov_t\left(M_{t,t+1}, P_{t+1,L-1}\right) - r_t \\ &= \mathbb{E}_t\left[\log\left(P_{t+1,L-1}\right)\right] - \log P_{ss,L} - \tfrac{\mathbb{E}_t\left[P_{t+1,L-1}\right]}{P_{ss,L}} e^{-r_t - \omega\times xhr_{t,L}} + 1 - \tfrac{Cov_t\left(M_{t,t+1}, P_{t+1,L-1}\right)}{P_{ss,L}} - r_t \\ &\approx \mathbb{E}_t\left[\log\left(\tfrac{P_{t+1,L-1}}{P_{ss,L}}\right)\right] - \tfrac{\mathbb{E}_t\left[P_{t+1,L-1}\right]}{P_{ss,L}}\left(e^{-r_{ss}}\left(1 - r_t - \omega\times xhr_{t,L} + r_{ss}\right)\right) - \tfrac{Cov_t\left(M_{t,t+1}, P_{t+1,L-1}\right)}{P_{ss,L}} \end{aligned}$$

$+1 - r_t$

as

$$\mathbb{E}_t\left[M_{t,t+1}\right] = e^{-r_t - \omega \times xhr_{t,L}}, P_{ss,L} = M_{ss,ss+1} P_{ss,L-1}$$

and

$$e^{-r_t - \omega \times xhr_{t,L}} \approx e^{-r_{ss}}\left(1 - r_t - \omega \times xhr_{t,L} + r_{ss}\right).$$

Finally, using $\log\left(\frac{P_{t+1,L-1}}{P_{ss,L-1}}\right) \approx \frac{P_{t+1,L-1}}{P_{ss,L-1}} - 1$, we obtain (41).

References

Adrian, T., Crump, R. K. and Moench, E. (2013), 'Pricing the term structure with linear regressions', *Journal of Financial Economics* **110**, 110–138.

Altig, D., Christiano, L. J., Eichenbaum, M. and Linde, J. (2011), 'Firm-specific capital, nominal rigidities and the business cycle', *Review of Economic Dynamics* **14**, 225–247.

An, S. and Schorfheide, F. (2007), 'Bayesian analysis of DSGE models', *Econometric Review* **26(2-4)**, 113–172.

Andreasen, M. M. (2012), 'On the effects of rare disasters and uncertainty shocks for risk premia in non-linear DSGE models', *Review of Economic Dynamics* **15**, 295–316.

Andreasen, M. M., Fernández-Villaverde, J. and Rubio-Ramírez, J. (2013), The pruned state-space system for non-linear DSGE models: Theory and empirical applications, Working Paper 18983, National Bureau of Economic Research.

Andreasen, M. M. and Meldrum, A. (2014), Dynamic term structure models: the best way to enforce the zero lower bound, Working paper, Aarhus University.

Andreasen, M. M. and Zabczyk, P. (2015), 'Efficient bond price approximations in non-linear equilibrium-based term structure models', *Studies in Nonlinear Dynamics and Econometrics* **19**(1), 1–34.

Aruoba, S. B., Bocola, L. and Schorfheide, F. (2013), A new class of nonlinear time series models for the evaluation of DSGE models, Working paper, University of Pennsylvania.

Barsky, R. B., Juster, F. T., Kimball, M. S. and Shapiro, M. D. (1997), 'Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study', *Quarterly Journal of Economics* **112**, 537–579.

Bernanke, B. S., Gertler, M. and Gilchrist, S. (1999), 'The financial accelerator in a quantitative business cycle framework', *Handbook of Macroeconomics* **1**, 1341–1393.

Binning, A. (2013), Solving second and third-order approximations to DSGE models: A recursive Sylvester equation solution, Working paper, Norges Bank.

Binsbergen, J. H. V., Fernandez-Villaverde, J., Koijen, R. S. and Rubio-Ramirez, J. (2012), 'The term structure of interest rates in a DSGE model with recursive preferences', *Journal of Monetary Economics* **59**, 634–648.

Calvo, G. A. (1983), 'Staggered prices in a utility-maximizing framework', *Journal of Monetary Economics* **12**, 383–398.

Christiano, L. J., Eichenbaum, M. and Evans, C. L. (2005), 'Nominal rigidities and the dynamic effects of a shock to monetary policy', *Journal of Political Economy* **113**, 1–45.

Cochrane, J. H. (2001), *Asset Pricing*, Princeton, NJ.

Cooley, T. F. and Prescott, E. (1995), *Economic Growth and Business Cycles*, Chapter 1 in Frontiers of Business Cycle Research, Ed. T. F. Cooley, Princeton University Press.

Creel, M. and Kristensen, D. (2011), Indirect likelihood inference, Working paper, Barcelona GSE.

Dridi, R., Guay, A. and Renault, E. (2007), 'Indirect inference and calibration of dynamic stochastic general equilibrium models', *Journal of Econometrics* **136**, 397–430.

Duffie, D. and Singleton, K. J. (1993), 'Simulated moments estimation of Markov models of asset prices', *Econometrica* **61**(4), 929–952.

Epstein, L. G. and Zin, S. E. (1989), 'Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework', *Econometrica* **57**(4), 937–969.

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

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007), 'Estimating macroeconomic models: A likelihood approach', *Review of Economic Studies* **74**, 1–46.

Gagnon, J., Raskin, M., Rernache, J. and Sack, B. (2011), 'Large-scale asset purchases by the Federal Reserve: Did they work?', *Federal Reserve Bank of New York, Research Paper Series - Economic Policy Review* **17**(1), 41–59.

Gertler, M. and Karadi, P. (2011), 'A model of unconventional monetary policy', *Journal of Monetary economics* **58**, 17–34.

Gürkaynak, R., Sack, B. and Wright, J. (2007), 'The U.S. Treasury yield curve: 1961 to the present', *Journal of Monetary Economics* **54**, 2291–2304.

Haan, W. J. D. and Wind, J. D. (2012), 'Nonlinear and stable perturbation-based approximations', *Journal of Economic Dynamic and Control* **36**, 1477–1497.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**(4), 1029–1054.

Hordahl, P., Tristani, O. and Vestin, D. (2008), 'The yield curve and macroeconomic dynamics', *The Economic Journal* **118**, 1937–1970.

Iskrev, N. (2010), 'Local identification in DSGE models', *Journal of Monetary Economics* **57**(2), 189–202.

Jermann, U. J. (1998), 'Asset pricing in production economics', *Journal of Monetary Economics* **41**, 257–275.

Jorda, O. (2005), 'Estimation and inference of impulse responses by local projections', *American Economic Review* **95**(1), 161–182.

Joyce, M. A. S., Lasaosa, A., Stevens, I. and Tong, M. (2011), 'The financial market impact of quantitative easing in the United Kingdom', *International Journal of Central Banking* **7**(3), 113–161.

Judd, K. L. and Guu, S.-M. (1997), 'Asymptotic methods for aggregate growth models', *Journal of Economic Dynamics and Control* **21**, 1025–1042.

Justiniano, A. and Primiceri, G. E. (2008), 'The time-varying volatility of macroeconomic fluctuations', *American Economic Review* **98(3)**, 604–641.

Kim, J., Kim, S., Schaumburg, E. and Sims, C. A. (2008), 'Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models', *Journal of Economic Dynamics and Control* **32**, 3397–3414.

Kim, J. and Ruge-Murcia, F. J. (2009), 'How much inflation is necessary to grease the wheels?', *Journal of Monetary Economics* **56**, 365–377.

Kim, J.-Y. (2002), 'Limited information likelihood and Bayesian analysis', *Journal of Econometrics* **107**, 175–193.

King, R. G. and Rebelo, S. T. (1999), 'Resuscitating real business cycles', *Handbook of Macroeconomics* **1**, 927–1007.

Koop, G., Pesaran, M. H. and Potter, S. M. (1996), 'Impulse response analysis in nonlinear multivariate models', *Journal of Econometrics* **74**, 119–147.

Lan, H. and Meyer-Gohde, A. (2013*a*), Pruning in perturbation DSGE models - Guidance from nonlinear moving average approximations, Discussion papers sfb 649, Humboldt University.

Lan, H. and Meyer-Gohde, A. (2013*b*), 'Solving DSGE models with a nonlinear moving average', *Journal of Economic Dynamics and Control* **37**(12), 2643–2667.

Lan, H. and Meyer-Gohde, A. (2014), 'Solvability of perturbation solutions in DSGE models', *Journal of Economic Dynamics and Control* **45**(C), 366–388.

Lombardo, G. and Sutherland, A. (2007), 'Computing second-order accurate solutions for rational expectation models using linear solution methods', *Journal of Economic Dynamics and Control* **31**, 515–530.

Lombardo, G. and Uhlig, H. (2014), A theory of pruning, Working paper series number 1696, European Central Bank.

Matthes, C. and Barnichon, R. (2014), Measuring the non-linear effects of monetary policy, Working paper, CREi.

Mehra, R. and Prescott, E. C. (1985), 'The equity premium: A puzzle', *Journal of Monetary Economics* **15**, 145–161.

Mutschler, W. (2015), 'Identification of DSGE models - The effects of higher-order approximation and pruning', *Journal of Economic Dynamics and Control* **56**, 34–54.

Peralta-Alva, A. and Santos, M. S. (2012), Analysis of numerical errors, Working paper, Federal Reserve Bank of St. Louis.

Rudebusch, G. D. and Swanson, E. T. (2008), 'Examining the bond premium puzzle with a DSGE model', *Journal of Monetary Economics* **55**, 111–126.

Rudebusch, G. D. and Swanson, E. T. (2012), 'The bond premium in a DSGE model with long-run real and nominal risks', *American Economic Journal: Macroeconomics* **4**(1), 1–43.

Ruge-Murcia, F. (2012), 'Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles', *Journal of Economic Dynamics and Control* **35**, 914–938.

Ruge-Murcia, F. (2013), *Generalized method of moments estimation of DSGE models*, In Handbook of Research Methods and Applications in Empirical Macroeconomics, N. Hashimzade and M. Thornton (Eds.), Edward Elgar Publishing.

Ruge-Murcia, F. J. (2007), 'Methods to estimate dynamic stochastic general equilibrium models', *Journal of Economic Dynamics and Control* **31**, 2599–2636.

Schmitt-Grohé, S. and Uribe, M. (2004), 'Solving dynamic general equilibrium models using a second-order approximation to the policy function', *Journal of Economic Dynamics and Control* **28**, 755–775.

Schmitt-Grohe, S. and Uribe, M. (2007), Optimal inflation stabilization in a medium-scale macroeconomic model, *in* K. Schmidt-Hebbel and R. Mishkin, eds, 'Moneatry Policy Under Inflation Targeting', Central Bank of Chile, Santiago, Chile, pp. 125–186.

Smets, F. and Wouters, R. (2007), 'Shocks and frictions in US business cycles: A Bayesian DSGE approach', *American Economic Review* **97**(3), 586–606.

Smith, J. A. A. (1993), 'Estimating nonlinear time-series models using simulated vector autoregressions', *Journal of Applied Econometrics* **8, Supplement: Special Issue on Econometric Inference Using Simulation Techniques**, S63–S84.

Swanson, E. (2012), 'Risk aversion and the labor margin in dynamic equilibrium models', *American Economic Review* **102**(4), 1663–1691.

Swanson, E. (2013), Risk aversion, risk premia, and the labor margin with generalized recursive preferences, Working paper series, Federal Reserve Bank of San Francisco.

Swanson, E. (2015), A Macroeconomic Mmodel of Equities and Real, Nominal, and Defaultable Bonds, Technical report, University of California, Irvine.

Tallarini, T. D. (2000), 'Risk-sensitive real business cycles', *Journal of Monetary Economics* **45**, 507–532.

Weil, P. (1990), 'Nonexpected utility in macroeconomics', *Quarterly Journal of Economics* **105**(1), 29–42.

Table 1: Estimation Results

The reported estimates are from the second step in GMM using the optimal weighting matrix with 10 lags in the Newey-West estimator. For $\mathcal{M}_{FB}^{RRA}$ and $\mathcal{M}_{FB,Taylor}^{RRA}$, the value of the RRA is restricted to 5 and not estimated. For these two models, preliminary results show that $\omega = 1.00$, which is also imposed for $\mathcal{M}_{FB}^{RRA}$ and $\mathcal{M}_{FB,Taylor}^{RRA}$.

	No feedback from long bonds	With feedback from long bonds			
	$\mathcal{M}_0$	$\mathcal{M}_{FB}$	$\mathcal{M}_{FB}^{RRA}$	$\mathcal{M}_{FB,Taylor}$	$\mathcal{M}_{FB,Taylor}^{RRA}$
β	0.9995	0.9971	0.9971	0.9972	0.9968
	(0.0001)	(0.0006)	(0.0005)	(0.0007)	(0.0004)
b	0.6720	0.7073	0.7046	0.7149	0.6620
	(0.0526)	(0.0374)	(0.0513)	(0.0372)	(0.0285)
h_{ss}	0.3427	0.3395	0.3391	0.3392	0.3385
	(0.0011)	(0.0010)	(0.0006)	(0.0006)	(0.0014)
ϕ_2	0.9757	0.6446	0.7064	0.5667	0.5926
	(0.2668)	(0.1365)	(0.2585)	(0.1478)	(0.1029)
RRA	615.7	23.07	5	13.1617	5
	(26.96)	(30.88)		(7.6267)	
κ	5.3986	9.7446	7.0361	9.1347	8.7656
	(0.8263)	(0.9127)	(0.9092)	(1.1303)	(0.8479)
α	0.8101	0.8002	0.8560	0.7978	0.8055
	(0.0066)	(0.0089)	(0.0073)	(0.0101)	(0.0089)
ρ_r	0.6491	0.8582	0.7626	0.8680	0.8362
	(0.0288)	(0.0245)	(0.0387)	(0.0296)	(0.0279)
β_π	1.2668	2.1708	3.3417	2.2720	3.0149
	(0.1512)	(0.2880)	(0.2409)	(0.1838)	(0.3846)
β_y	0.0315	0.2294	0.2442	0.2286	0.1809
	(0.0257)	(0.0624)	(0.0344)	(0.0326)	(0.0221)
$\mu_{\Upsilon,ss}$	1.0012	1.0011	1.0008	1.0013	1.0007
	(0.0011)	(0.0013)	(0.0010)	(0.0013)	(0.0011)
$\mu_{z,ss}$	1.0052	1.0053	1.0054	1.0053	1.0053
	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0006)
ρ_a	0.7450	0.7847	0.7733	0.7843	0.7059
	(0.0557)	(0.0229)	(0.0211)	(0.0274)	(0.0237)
ρ_G	0.8033	0.8147	0.9588	0.8229	0.8705
	(0.0950)	(0.0532)	(0.0214)	(0.0618)	(0.0556)
g_{ss}/y_{ss}	0.2062	0.2071	0.2060	0.2083	0.2209
	(0.0029)	(0.0029)	(0.0032)	(0.0035)	(0.0068)
σ_a	0.0161	0.0126	0.0178	0.0126	0.0168
	(0.0020)	(0.0018)	(0.0013)	(0.0021)	(0.0015)
σ_G	0.0422	0.0524	0.0249	0.0517	0.0373
	(0.0122)	(0.0109)	(0.0036)	(0.0119)	(0.0127)
σ_d	0.0131	0.0087	0.0089	0.0078	0.0068
	(0.0020)	(0.0015)	(0.0019)	(0.0018)	(0.0010)
π_{ss}	1.0121	1.0116	1.0094	1.0124	1.0166
	(0.0006)	(0.0004)	(0.0005)	(0.0011)	(0.0013)
ω	–	**0.9104**	**1.00**	0.9915	1.00
		(0.2301)		(0.0681)	
β_{xhr}	–	–		**−0.0690**	**−0.5190**
				(0.0858)	(0.0940)
Memo					
IES	0.053	0.063	0.058	0.067	0.095
u_{ss}	−2.273	−1.670	−1.774	−1.561	−1.565
ϕ_3	−1466.0	−70.89	−13.36	−43.54	−14.12

Table 2: Model Fit

All variables are expressed in annualized terms, except for $\log(g_t/y_t)$ and $\log h_t$.

	Data	No feedback from long bonds $\mathcal{M}_0$	With feedback from long bonds $\mathcal{M}_{FB}$	$\mathcal{M}_{FB}^{RRA}$	$\mathcal{M}_{FB,Taylor}$	$\mathcal{M}_{FB,Taylor}^{RRA}$
Means						
$\Delta c_t \times 100$	2.439	2.350	2.378	2.337	2.424	2.271
$\Delta i_t \times 100$	3.105	2.847	2.817	2.650	2.943	2.535
$\pi_t \times 100$	3.757	3.404	3.323	3.329	3.323	3.471
$r_t \times 100$	5.605	5.567	5.495	5.471	5.482	5.596
$r_{t,40} \times 100$	6.993	6.924	6.919	6.808	6.841	6.982
$xhr_{t,40} \times 100$	1.724	2.090	1.492	1.362	1.386	1.422
$\log(g_t/y_t)$	−1.575	−1.578	−1.576	−1.576	−1.576	−1.577
$\log h_t$	−1.084	−1.083	−1.083	−1.083	−1.083	−1.083
Stds (in pct)						
Δc_t	2.685	2.701	2.668	2.633	2.687	2.714
Δi_t	8.914	8.687	8.938	8.747	8.873	8.889
π_t	2.481	2.669	2.709	2.509	2.709	2.640
r_t	2.701	2.520	2.547	2.450	2.510	2.572
$r_{t,40}$	2.401	2.282	2.057	2.171	2.052	2.193
$xhr_{t,40}$	22.978	12.930	12.683	9.167	12.322	7.977
$\log g_t/y_t$	8.546	8.264	9.300	10.449	9.438	9.663
$\log h_t$	1.676	2.396	1.892	2.432	1.869	2.124
Auto-correlations						
$corr(\Delta c_t, \Delta c_{t-1})$	0.254	0.238	0.336	0.357	0.351	0.315
$corr(\Delta i_t, \Delta i_{t-1})$	0.506	0.355	0.132	0.171	0.138	0.185
$corr(\pi_t, \pi_{t-1})$	0.859	0.824	0.878	0.932	0.865	0.849
$corr(r_t, r_{t-1})$	0.942	0.966	0.989	0.972	0.988	0.976
$corr(r_{t,40}, r_{t-1,40})$	0.963	0.989	0.988	0.994	0.988	0.996
$corr(xhr_{t,40}, xhr_{t-1,40})$	−0.024	−0.006	−0.005	0.003	−0.007	0.010
$corr(\log g_t/y_t, \log g_{t-1}/y_{t-1})$	0.9922	0.888	0.859	0.972	0.865	0.932
$corr(\log h_t, \log h_{t-1})$	0.792	0.543	0.549	0.611	0.535	0.477

Table 2: Model Fit (continued)

		No feedback from long bonds	With feedback from long bonds			
	Data	$\mathcal{M}_0$	$\mathcal{M}_{FB}$	$\mathcal{M}_{FB}^{RRA}$	$\mathcal{M}_{FB,Taylor}$	$\mathcal{M}_{FB,Taylor}^{RRA}$
$corr\left(\Delta c_t, \Delta i_t\right)$	0.594	0.518	0.522	0.543	0.528	0.598
$corr\left(\Delta c_t, \pi_t\right)$	−0.362	−0.313	−0.304	−0.229	−0.314	−0.285
$corr\left(\Delta c_t, r_t\right)$	−0.278	−0.212	−0.189	−0.218	−0.194	−0.173
$corr\left(\Delta c_t, r_{t,40}\right)$	−0.178	−0.111	−0.135	−0.092	−0.141	−0.085
$corr\left(\Delta c_t, xhr_{t,40}\right)$	0.271	0.495	0.360	0.484	0.348	0.554
$corr\left(\Delta i_t, \pi_t\right)$	−0.242	−0.452	−0.337	−0.237	−0.341	−0.374
$corr\left(\Delta i_t, r_t\right)$	−0.265	−0.151	−0.104	−0.123	−0.099	−0.094
$corr\left(\Delta i_t, r_{t,40}\right)$	−0.153	−0.057	−0.050	−0.038	−0.048	−0.032
$corr\left(\Delta i_t, xhr_{t,40}\right)$	0.021	0.706	0.254	0.691	0.254	0.831
$corr\left(\pi_t, r_t\right)$	0.628	0.938	0.841	0.906	0.832	0.805
$corr\left(\pi_t, r_{t,40}\right)$	0.479	0.822	0.890	0.932	0.876	0.879
$corr\left(\pi_t, xhr_{t,40}\right)$	−0.249	−0.379	−0.190	−0.223	−0.182	−0.296
$corr\left(r_t, r_{t,40}\right)$	0.861	0.847	0.830	0.805	0.832	0.789
$corr\left(r_t, xhr_{t,40}\right)$	−0.233	−0.150	−0.067	−0.161	−0.0596	−0.180
$corr\left(r_{t,40}, xhr_{t,40}\right)$	−0.121	−0.053	−0.111	−0.055	−0.106	−0.036

Table 3: Model Specification Test

The objective function in step 1, denoted Q^{step1}, is computed with the weighting matrix $\mathbf{W}_T = diag\left(\hat{\mathbf{S}}_{mean}^{-1}\right)$, where $\hat{\mathbf{S}}_{mean}$ denotes the variance of the sample moments as computed by the Newey-West estimator using 10 lags. The objective function in step 2, denoted Q^{step2}, is computed using the optimal weighting matrix with 10 lags in the Newey-West estimator. The P-value is for the J-test for model misspecification based on the objective function in step 2.

	No feedback from long bonds	With feedback from long bonds			
	$\mathcal{M}_0$	$\mathcal{M}_{FB}$	$\mathcal{M}_{FB}^{RRA}$	$\mathcal{M}_{FB,Taylor}$	$\mathcal{M}_{FB,Taylor}^{RRA}$
Objective function: Q^{step1}	16.929	14.546	16.333	14.546	16.286
Objective function: Q^{step2}	0.0887	0.0860	0.0835	0.0841	0.0874
Number of moments	39	39	39	39	39
Number of parameters	19	20	19	21	20
P-value	0.685	0.658	0.796	0.618	0.700

Table 4: Decomposing the 10-year term premium

Moments for the 10-year term premium $TP_{t,40}$ are reported in annualized basis points, whereas moments for the remaining variables are at a quarterly frequency and unscaled. Moments for the quantity of risk cannot be computed directly by the perturbation method (because $\mathbb{V}_t\left(M_{t,t+1}\right)$ and, hence, the market price of risk are zero in the steady state), and we therefore compute these moments from simulated sample paths of 1,000,000 observations for $TP_{t,40}$ and the market price of risk.

	$TP_{t,40}$	$\mathbb{V}_t\left(M_{t,t+1}\right)$	Market price of risk	Quantity of risk
Means ($\pi_{ss} = \hat{\pi}_{ss}^{GMM}$)				
$\mathcal{M}_0$	145.20	0.0321	0.0327	0.2881
$\mathcal{M}_{FB}$	150.19	8.12×10^{-4}	8.29×10^{-4}	3.3701
$\mathcal{M}_{FB}^{RRA}$	139.82	2.46×10^{-4}	2.51×10^{-4}	13.8306
$\mathcal{M}_{FB,Taylor}$	142.96	3.08×10^{-4}	3.14×10^{-4}	11.0778
$\mathcal{M}_{FB,Taylor}^{RRA}$	142.23	2.00×10^{-4}	2.05×10^{-4}	21.9432
Means ($\pi_{ss} = 1.00$)				
$\mathcal{M}_0$	32.19	0.0030	0.0030	0.2702
$\mathcal{M}_{FB}$	42.72	2.33×10^{-4}	2.35×10^{-4}	4.5682
$\mathcal{M}_{FB}^{RRA}$	88.01	1.25×10^{-4}	1.26×10^{-4}	17.5490
$\mathcal{M}_{FB,Taylor}$	52.42	1.04×10^{-4}	1.05×10^{-4}	12.6119
$\mathcal{M}_{FB,Taylor}^{RRA}$	93.26	7.17×10^{-5}	7.23×10^{-5}	32.4587
Stds ($\pi_{ss} = \hat{\pi}_{ss}^{GMM}$)				
$\mathcal{M}_0$	115.88	0.0503	0.0515	137.86
$\mathcal{M}_{FB}$	114.37	8.63×10^{-4}	8.85×10^{-4}	790.08
$\mathcal{M}_{FB}^{RRA}$	93.39	1.58×10^{-4}	1.62×10^{-4}	593.49
$\mathcal{M}_{FB,Taylor}$	110.06	3.04×10^{-4}	3.12×10^{-4}	1076.32
$\mathcal{M}_{FB,Taylor}^{RRA}$	131.56	1.92×10^{-4}	1.98×10^{-4}	4084.97
Stds ($\pi_{ss} = 1.00$)				
$\mathcal{M}_0$	1.42	3.39×10^{-4}	3.55×10^{-4}	0.0219
$\mathcal{M}_{FB}$	0.88	2.04×10^{-5}	2.13×10^{-5}	0.3379
$\mathcal{M}_{FB}^{RRA}$	10.11	1.16×10^{-5}	1.20×10^{-5}	1.6744
$\mathcal{M}_{FB,Taylor}$	1.12	9.47×10^{-6}	9.83×10^{-6}	1.1751
$\mathcal{M}_{FB,Taylor}^{RRA}$	5.31	4.93×10^{-6}	5.10×10^{-6}	3.6225

Figure 1: Simulated sample path

The capital stock, the price dispersion index, and the inflation rate are expressed in deviation from the deterministic steady state, whereas consumption growth is de-meaned. Unless stated otherwise, all parameters are from $\mathcal{M}_0$. All variables are expressed at a quarterly level.

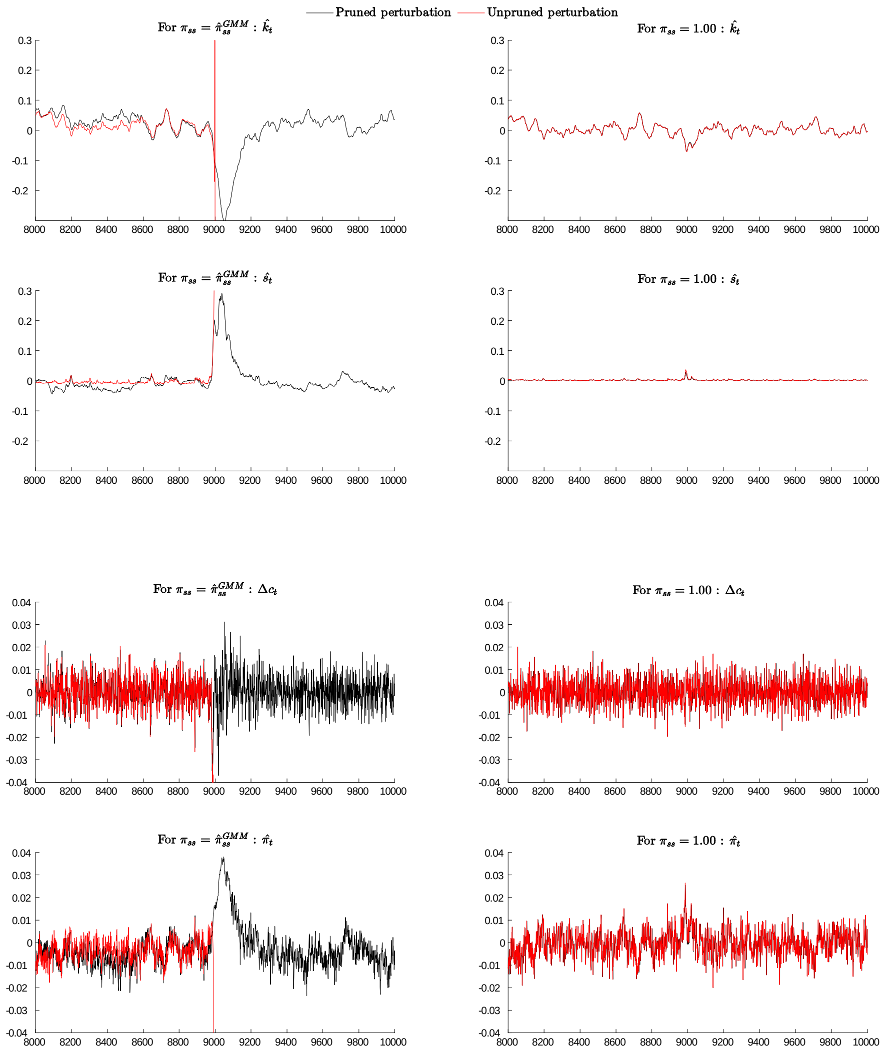

Figure 2: GIRFs: Technology shock

GIRFs following a positive one-standard-deviation shock to technology. The GIRFs are computed at the unconditional mean of the states using the estimated parameters for $\mathcal{M}^{RRA}_{FB,Taylor}$. All GIRFs are expressed in deviation from the steady state, except for excess holding period return and term premium, which are expressed in annualized basis points from their unconditional means.

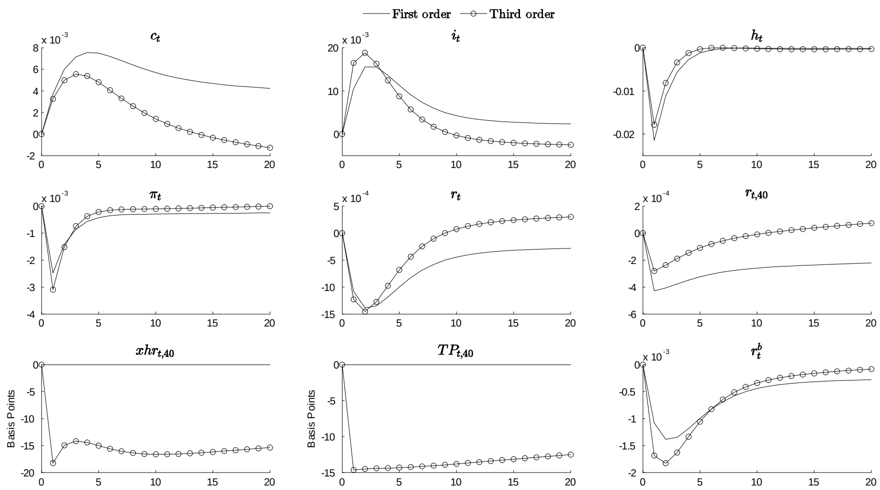

Figure 3: GIRFs: Government shock

GIRFs following a positive one-standard-deviation shock to government spending. The GIRFs are computed at the unconditional mean of the states using the estimated parameters for $\mathcal{M}^{RRA}_{FB,Taylor}$. All GIRFs are expressed in deviation from the steady state, except for excess holding period return and term premium, which are expressed in annualized basis points from their unconditional means.

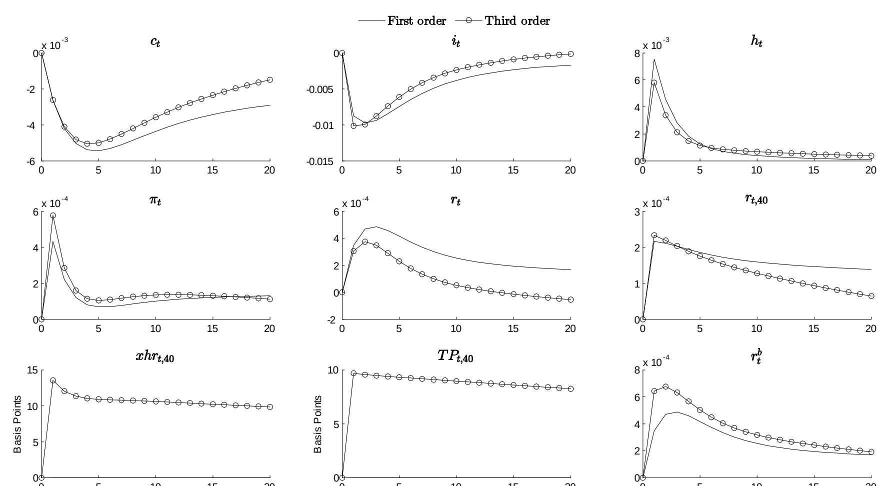

Figure 4: GIRFs: Preference shock

GIRFs following a positive one-standard-deviation shock to preferences. The GIRFs are computed at the unconditional mean of the states using the estimated parameters for $\mathcal{M}^{RRA}_{FB,Taylor}$. All GIRFs are expressed in deviation from the steady state, except for excess holding period return and term premium, which are expressed in annualized basis points from their unconditional means.

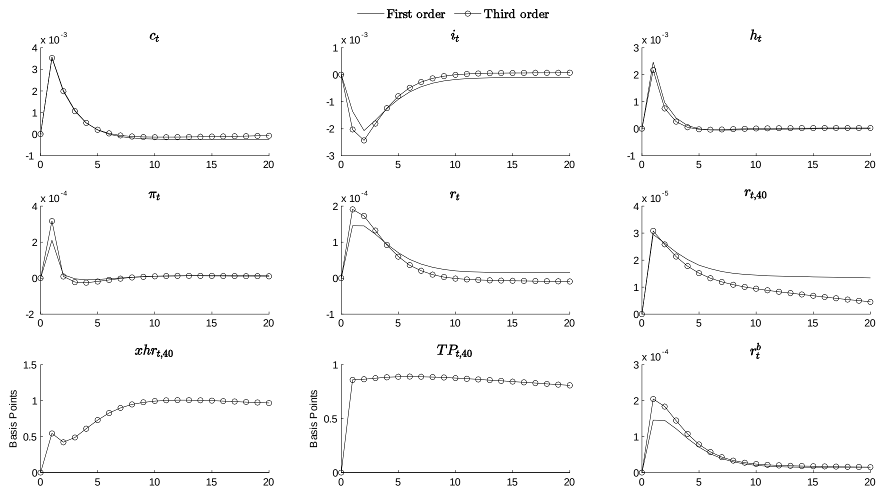

Figure 5: Conditional GIRFs: Expansions vs. Recessions

GIRFs following a positive one-standard-deviation shock to technology using the estimated parameters for $\mathcal{M}^{RRA}_{FB,Taylor}$. The state values representing recessions are defined from episodes in a simulated sample path with detrended negative output in the current and the previous two periods; otherwise, the economy is defined to be in expansion. The GIRFs are computed as the average across 500 draws from expansions and recessions. All GIRFs are expressed in deviation from the steady state, except for excess holding period return and term premium, which are expressed in annualized basis points from their unconditional means.

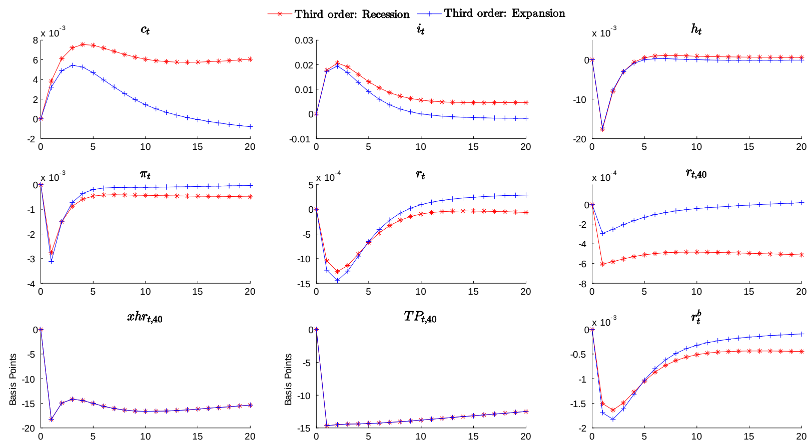

Figure 6: Conditional GIRFs: High vs. low inflation

GIRFs following a positive one-standard-deviation shock to technology using the estimated parameters for $\mathcal{M}^{RRA}_{FB,Taylor}$. The state values representing high inflation are defined from episodes with inflation larger than one standard deviation of inflation in a simulated sample path; otherwise, the economy is defined to be in a low inflation regime. The GIRFs are computed as the average across 500 draws from regimes of high and low inflation. All GIRFs are expressed in deviation from the steady state, except for excess holding period return and term premium, which are expressed in annualized basis points from their unconditional means.