ENDOGENOUS TIME VARIATION IN VECTOR AUTOREGRESSIONS

2021

BANCO DE ESPAÑA
Eurosistema

Documentos de Trabajo
N.º 2108

Danilo Leiva-Leon and Luis Uzeda

ENDOGENOUS TIME VARIATION IN VECTOR AUTOREGRESSIONS [*]

Danilo Leiva-Leon [**]

BANCO DE ESPAÑA

Luis Uzeda [***]

BANK OF CANADA

(*) For their comments and suggestions, we would like to thank Máximo Camacho, Fabio Canova, Efrem Castelnuovo, Todd Clark, Laurent Ferrara, Ana Galvao, Serdar Kabaca, Michele Lenza, John Maheu, James Morley, Haroon Mumtaz, Helmut Lütkepohl, Adrian Pagan, Mikkel Plagborg-Moller, Giorgio Primiceri, Rodrigo Sekkel, Benjamin Wong, Tomasz Wozniak, two anonymous referees and conference and seminar participants at the *6th International Association of Applied Econometrics Conference, 25th Computing Economics and Finance Conference, 27th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, 39th International Symposium on Forecasting, 2nd Workshop in Structural VAR Models, 3rd Forecasting at Central Banks Conference, 9th Workshop in Time Series Econometrics,* Bank of Canada, University of Sydney and University of Melbourne. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada, Banco de España or the Eurosystem.
(**) E-mail: danilo.leiva@bde.es.
(***) E-mail: luzedagarcia@bank-banque-canada.ca.

Documentos de Trabajo. N.º 2108

2021

ISSN: 1579-8666 (on line)

Abstract

We introduce a new class of time-varying parameter vector autoregressions (TVP-VARs) where the identified structural innovations are allowed to influence the dynamics of the coefficients in these models. An estimation algorithm and a parametrization conducive to model comparison are also provided. We apply our framework to the US economy. Scenario analysis suggests that, once accounting for the influence of structural shocks on the autoregressive coefficients, the effects of monetary policy on economic activity are larger and more persistent than in an otherwise standard TVP-VAR. Our results also indicate that cost-push shocks play a prominent role in understanding historical changes in inflation-gap persistence.

Keywords: TVP-VAR, state-space, endogeneity, bayesian, monetary policy.

JEL classification: C11, C32, E31, E52.

Resumen

Este artículo introduce una nueva clase de modelos de vectores autorregresivos con parámetros cambiantes en el tiempo (TVP-VAR). En los modelos propuestos, se permite que las innovaciones estructurales puedan influir en la dinámica de sus coeficientes. También se proporciona un algoritmo de estimación y una parametrización conducente a la comparación de modelos. Este nuevo marco econométrico es aplicado a la economía estadounidense. Un análisis de escenarios sugiere que, una vez que se tiene en cuenta la influencia de las innovaciones estructurales en los coeficientes autorregresivos, los efectos de la política monetaria sobre la actividad económica son mayores y más persistentes que en un TVP-VAR estándar. Resultados adicionales sugieren que las innovaciones que impulsan los costos desempeñan un papel destacado en los cambios históricos en la persistencia de la brecha de inflación.

Palabras clave: TVP-VAR, estado-espacio, endogeneidad, bayesiano, política monetaria.

Códigos JEL: C11, C32, E31, E52.

1 Introduction

Time-varying parameter vector autoregressions (TVP-VARs) are a well-established tool for empirical analysis of changes in the relationship between economic variables. In part, the appeal of these models stems from the fact that they can capture a wide range of economic dynamics while preserving a tractable structure inherited from fixed-coefficient VARs. Moreover, TVP-VARs can be regarded as a reduced-form representation of nonlinear environments adopted for policy design, such as dynamic stochastic general equilibrium (DSGE) models that exhibit parameter variation.[1]

Initial efforts to work with TVP-VARs date back to Doan et al. (1984), Sims (1993), Canova (1993) and Stock and Watson (1996). Nevertheless, these models have arguably become more popular after papers such as Cogley and Sargent (2005) and Primiceri (2005) applied them to investigate changes in the transmission mechanism of monetary policy. Since then, numerous other studies followed using TVP-VARs to tackle different issues. For instance, Mumtaz and Surico (2009) used a TVP-VAR to assess macro-finance instabilities in the relationship between the term structure of interest rates and the economy. Gali and Gambetti (2015) and Paul (2019) relied on the same econometric framework to focus on issues related to changes in the sensitivity of asset prices to monetary policy, while Baumeister and Peersman (2013) explored variations in the price elasticity of oil demand. TVP-VARs have also been associated with modeling changes in inflation dynamics, as in Clark and Terry (2010) and Bianchi and Civelli (2015).

A common feature in all these studies is that the innovations producing parameter variations are not identified. Consequently, the traditional TVP-VAR framework, albeit useful to model structural changes, lacks a formal strategy to shed light on why such changes may occur in the first place. Our main contribution is to propose a new class of TVP-VARs where parameter changes are explicitly associated with the structural innovations identified within these models.[2] Such innovations, as is well known, are the objects that commonly have an economic and causal interpretation in the context of VAR modeling. Therefore, our strategy provides a TVP-VAR framework that not only accounts for parameter changes, but is also informative on what drives such changes.

[1] For example, Cogley et al. (2015) show that DSGE models with learning can be recast as a reduced-form TVP-VAR.

[2] In keeping with the common jargon for VARs, we will adopt the terminologies 'structural innovations' and 'structural shocks' interchangeably throughout this paper.

Perhaps the paper that is most closely related to ours is Cogley and Sargent (2001). To the best of our knowledge, these authors proposed the first (and only, to this date) TVP-VAR that accommodates dependence between measurement errors and coefficient innovations. While the work in Cogley and Sargent (2001) certainly lays an important foundation to ours, the class of models developed in this paper is the first to speak directly to the role of structural shocks behind parameter variation in TVP-VARs. More specifically, Cogley and Sargent (2001) consider cross-covariances between the VAR *reduced form* errors and the drifting coefficients. Instead, we directly parameterize such coefficients in terms of the *identified shocks*. In doing so, we simplify both measurement and validation of how structural shocks affect coefficient variations.[3] Also, our approach introduces nuances, such as accounting for both contemporaneous and lagged effects of the identified shocks on the VAR coefficients. Sections 2 and 3 elaborate further on all these points.

Importantly, the framework we propose nests the case of TVP-VARs where the drifting coefficients and structural shocks are – as commonly assumed – orthogonal to each other. This is achieved by specifying the law of motion for the time-varying coefficients as a function of two distinct (and independent) elements: (i) a set of identified structural innovations; and (ii) a coefficient-specific error term. Keeping the latter is useful, as it allows us to apply formal statistical procedures to gauge the evidence (or lack thereof) in favor of our approach.[4] In particular, we adopt a Bayesian technique for verification of exclusion restrictions, namely, the Savage-Dickey Density Ratio method (see Verdinelli and Wasserman (1995)). Thus, a second contribution of this paper is the provision of a model parametrization that is conducive to testing the validity of the TVP-VARs proposed here.

A third contribution of this paper is the development of a Markov Chain Monte Carlo (MCMC) algorithm to estimate the class of TVP-VARs proposed in this study. Our MCMC sampler is efficient and builds upon previous work on precision sampling methods in Chan and Jeliazkov (2009). Estimation techniques are generalized to accommodate several ways the structural innovations can

[3]For example, with reduced form errors – which are convolutions of multiple structural shocks – it is less straightforward to tease out information about how a specific economic shock of interest contributes to the overall dynamics of the coefficients in the model. In addition, there are numerous cross-covariances between coefficient innovations and reduced-form errors in TVP-VARs. Numerosity further complicates testing and summarizing the relationship between structural shocks and the VAR coefficients if focusing on correlations in terms of the reduced form errors.

[4]Keeping a coefficient-specific error term is useful for other reasons as well. For example, it prevents changes in the coefficients from being solely driven by the structural shocks, which in turn could lead to excess comovement amongst the state variables (i.e. the coefficients). Also, such errors can help approximating shocks that are not accounted for in the VAR.

enter the drifting coefficient equations. More precisely, deciding whether structural innovations affect the VAR coefficients contemporaneously or with lags requires only adjusting the number of non-zero bands in a sparse matrix.

To differentiate our framework from the extant literature on TVP-VARs, hereafter, we refer to the class of models proposed in this paper as 'endogenous' TVP-VARs. Note that the term endogenous is applied here simply to reflect the explicit relationship between the structural innovations and the time-varying coefficients, which is absent in traditional TVP-VARs. In this sense, the latter can be perceived as being 'exogenous' TVP-VARs, i.e. where coefficient changes occur independently from the structural innovations.

We illustrate the usefulness of our framework with two substantial empirical applications. In the first one, we adopt a small scale TVP-VAR along the lines of Cogley et al. (2010) to study changes in the persistence of the transitory (or gap) component of inflation. Inflation-gap persistence is measured using the same statistical metric proposed by these authors. That is, a measure akin to an R^2 coefficient of determination and which is a function of the time-varying coefficients in the VAR. Given the direct link between these coefficients and the structural innovations in our models, we are able to evaluate the extent to which VAR shocks – identified with sign restrictions (see, e.g., Uhlig (2005)) – matter for changes in the persistence of the inflation-gap.[5] Overall, our results indicate that cost-push shocks played a salient role as a driver for inflation-gap movements over the past five decades.

In the second application, we conduct scenario analysis based on exogenous and endogenous TVP-VARs. In particular, we investigate the effects of alternative monetary policy decisions on economic activity during two periods: (i) the transition towards the (effective) zero lower bound of the policy rate in the early stages of the 2008 financial crisis; and (ii) the normalization process of interest rates that began in 2015. The idea behind these scenarios is to illustrate how differences between endogenous and exogenous TVP-VARs manifest themselves in the context of normative analysis that are policy relevant. Specifically, we compare the simulated impulse responses for inflation and unemployment that emerge from these two models in response to monetary policy

[5] Sign restrictions are a useful strategy to generate macroeconomic impulse responses that are consistent with economic theory. That said, selecting an identification strategy is, of course, arbitrary. If desired, alternative identification schemes such as short- and long-run restrictions, Cholesky, identification through heteroskedasticity or narrative-based approaches could be integrated into our framework. We leave such extensions for future work.

shocks. A key result from this exercise is that the additional channel for the propagation of policy shocks in the endogenous setting – i.e. the direct impact of shocks on the time-varying coefficients – generates a larger and more protracted effect of monetary policy on economic activity in both scenarios. Such a result receives strong support from the data in a model comparison exercise.

Finally, our paper is also related to a strand of the literature that focuses on endogenizing structural changes in macroeconometric models. For example, Kim et al. (2008) and Kang (2014) propose methods to endogenize structural changes in the context of univariate Markov-switching models. For VARs, attempts to endogenize structural changes have traditionally relied on piecewise-type models (e.g. regime-switching) where changes in the autoregressive parameters are dictated by some observable time series. Notable examples for these are threshold VARs in, e.g., Tsay (1998) and Galvão (2006), and smooth transition VARs in, e.g., Anderson and Vahid (1998) and Auerbach and Gorodnichenko (2012). More recently, Carriero et al. (2018) and Mumtaz and Theodoridis (2019) – also using observable time series – adopted Bayesian techniques to endogenize volatility changes in VARs.

The rest of the paper is organized as follows. Section 2 presents a general framework for endogenous TVP-VAR models. Section 3 discusses the estimation of the proposed framework. Section 4 investigates several sources of macroeconomic instability in the US economy through the lens of our models. Section 5 concludes.

2 Modeling Endogenous Parameter Instabilities

In this section we introduce a new class of TVP-VARs that allows for the intercept and autoregressive coefficients to be driven by identified structural shocks.[6] We show that the proposed framework not only nests the traditional approach to model parameter variations in VARs but also accommodates different alternatives to endogenize dynamics for the drifting coefficients. In particular, we put forward three alternatives for specifying endogenous TVP-VARs. The first one considers the case when structural shocks embedded in the VAR affect the coefficients contemporaneously. The second alternative focuses on the case when coefficients are influenced by past

[6] How shock identification is achieved is addressed in Section 4.

realizations of the structural shocks. Lastly, we consider a more general scenario when both contemporaneous and lagged structural innovations are allowed to influence parameter instabilities in the model.

2.1 Traditional TVP-VAR Models

Let $y_t = (y_{1,t}, y_{2,t}, ..., y_{N,t})'$ be a $N \times 1$ vector of economic variables whose dynamic relationships are governed by the vector of time-varying coefficients ϕ_t. A baseline representation for a TVP-VAR with p lags can thus be written as:

$$y_t = X_t \phi_t + A e_t, \tag{1}$$

$$\phi_t = \phi_{t-1} + V_t, \tag{2}$$

where $\phi_t = (\phi_{1,t}, \cdots, \phi_{N,t})'$, such that each element $\phi_{i,t}$ for $i = 1, ..., N$ is an $Np + 1 \times 1$ vector following – as is standard in TVP-VARs – random walk dynamics driven by the vector V_t.[7] The latter collects innovations that can be mutually correlated or independent. Lagged regressors and constant terms are collected in $X_t = I_N \otimes (1, y'_{t-1}, ..., y'_{t-p})$, where I_N is an N-dimensional identity matrix and the symbol $\otimes$ denotes the Kronecker product. The contemporaneous relationships amongst reduced-form errors are captured by the impact matrix A, while e_t represents the vector of – unit variance and mutually uncorrelated – structural shocks, i.e. reduced-form innovations are given by $u_t = A e_t$.[8]

This paper is concerned with the connection between e_t and V_t. In this regard, both processes are typically assumed to be jointly Gaussian and evolving independently at all leads and lags. Conceptually, the assumption of independence between e_t and V_t implies that changes in the transmission mechanism of economic shocks, encapsulated in ϕ_t, remain entirely driven by sources of information that are not identified within the VAR system. Therefore, albeit useful to infer changes in ϕ_t, the framework in (1)-(2) remains silent about why such changes may take place over time.

[7] While assuming random walk dynamics is the usual and parsimonious strategy to model drifting coefficients in TVP-VARs, some authors, as in Canova (1993), allow for a more general framework where the coefficients follow stationary autoregressive processes. Nevertheless, random walk coefficients provide enough flexibility to approximate the dynamics from quite distinct data-generating processes (see, e.g., Canova et al. (2015)).

[8] We discuss the computation of A in Section 4.1.

Unlike previous applications of TVP-VARs, we relax the independence assumption for e_t and V_t and propose methods to model and test the relationship between these two terms.

Before discussing our general framework, one comment, however, is in order. Thus far, we have deliberately abstracted from time variation in second-moment parameters. That is not to say that such a modeling feature is unimportant. In fact, heteroskedasticity plays an important role in TVP-VAR analysis as in Primiceri (2005), Sims and Zha (2006) and Canova and Gambetti (2009), to name a few. Hence, as a robustness check, in Section 4.5 we also consider endogenous TVP-VARs that allow for time-varying volatilities.[9] To keep notation clear, however, in what follows we discuss our methodology in the context of homoskedastic endogenous TVP-VARs.

2.2 A More General Framework

Our modeling strategy consists of letting ϕ_t in (1) and (2) to be partially explained by identified economic shocks. That is, the vector of innovations V_t is decomposed into two orthogonal sources of information:

$$V_t = g(e_t) + v_t, \tag{3}$$

where $g(e_t)$ denotes some function of the identified structural shocks and v_t is a vector collecting mutually uncorrelated coefficient-specific errors. The latter can be interpreted as unidentified shocks that may be associated with variables that are not accounted for in the VAR. Notably, while e_t contains N elements, v_t is a $(N^2p+N)\times 1$ vector, which makes identification of coefficient-specific errors intrinsically hard. Nonetheless, even if some sources of parameter instability cannot be identified, it remains important to measure what portion of changes in ϕ_t can be attributed to identifiable drivers.

There are two relevant issues that are raised under the formalization in (3). The first one regards the functional form of $g(\bullet)$. In principle, structural shocks could affect parameter stability in a linear or nonlinear fashion. Since TVP-VARs can approximate more complex nonlinear structures while preserving a simpler framework, it seems a bit incongruous to reverse such simplicity now by introducing nonlinearities into the state equations. Therefore, in this paper we develop endogenous

[9] The challenging task of jointly endogenizing first and second moments for TVP-VARs is, however, beyond the scope of this paper.

TVP-VARs where economic shocks affect the VAR coefficients linearly and leave nonlinearities for subsequent work.

The second issue corresponds to the timing in the relationship between V_t and $g(e_t)$, or equivalently, between e_t and ϕ_t. If structural shocks are associated with, say, policy makers' and agents' decisions, the pace at which such decisions might exert an influence on economic relations can vary. This raises the question whether e_t should affect ϕ_t contemporaneously or with lags. In what follows, we explore several options regarding this timing issue.[10]

Contemporaneous Innovations

The first case we consider is the case when structural shocks at time t can influence changes in the relationship between the variables in y_t within the same time period. Formally, we have:

$$\phi_t = \phi_{t-1} + H_{C,\lambda} e_t + v_t, \quad v_t \sim \mathcal{N}(0,\, \Omega_v), \tag{4}$$

where $\Omega_v = \texttt{diag}(\sigma^2_{v,1}, ..., \sigma^2_{v,N^2p+N})$ and $H_{C,\lambda} = \lambda_C' \otimes \iota$ denotes the Kronecker product between the $N \times 1$ vector $\lambda_C = (\lambda_{C,1}, ..., \lambda_{C,N})'$ and a $(N^2p+N) \times 1$ vector of ones, ι.

The vector λ_C – and its variants introduced below – represents a key piece of information for the remainder of this paper. It governs the sensitivity of ϕ_t to contemporaneous structural shocks. Also, as we will discuss in Section 3.2, statistical validation of our framework is organized around such a vector.

Next, note that the Kronecker structure in $H_{C,\lambda} = \lambda_C' \otimes \iota$ is introduced to allow for a one-to-one mapping between λ_C and e_t, i.e. each term in λ_C is associated with a specific structural shock in e_t and vice-versa. Hence, instead of reporting results for $N(N^2p+N)$ free parameters in λ_C – which would be the case had we assigned N different sensitivity parameters to each time-varying coefficient in ϕ_t – we focus on the overall effect of e_t on ϕ_t captured by N sensitivity parameters. Such an overidentification strategy mitigates parameter proliferation and greatly helps to summarize the relevance of structural shocks for changes in ϕ_t.

Also, since there are considerably fewer structural shocks than VAR coefficients, the expression in (4) describes a factor structure where commonalities across ϕ_t are absorbed by e_t. Strong

[10] For example, in the context of theoretical (or micro-founded) models, lags in the transmission of shocks to the economy can be reconciled with the existence of frictions, such as nominal rigidities (e.g., Smets and Wouters (2007)) and capital adjustment costs (e.g., Cooper and Haltiwanger (2006)).

evidence in favor of comovements in ϕ_t for TVP-VARs are documented, for example, in Cogley and Sargent (2005) via principal components analysis.

Lagged Innovations

The second case we introduce is when structural shocks can alter the relationship between the variables in y_t with a one-period lag. This could be the case, for instance, if the underlying economic processes embedded in y_t are substantially persistent such that the effect of shocks on ϕ_t might require some time to materialize. Accordingly, the state equation in (2) can be recast as:

$$\phi_t = \phi_{t-1} + H_{L,\lambda} e_{t-1} + v_t, \quad v_t \sim \mathcal{N}(0, \Omega_v). \tag{5}$$

As in the contemporaneous case, we set $\Omega_v = \texttt{diag}(\sigma^2_{v,1}, ..., \sigma^2_{v,N^2p+N})$ and $H_{L,\lambda} = \lambda_L' \otimes \iota$, where the vector λ_L contains the N parameters that govern the sensitivity of ϕ_t to the lagged structural shocks.

Contemporaneous and Lagged Innovations

The third case is the most general one since it encompasses the previous two possibilities. That is, we assume that time-varying coefficients might be potentially influenced by both contemporaneous and lagged structural shocks. Therefore, the dynamics for the VAR coefficients are be given by:

$$\phi_t = \phi_{t-1} + H_{C,\lambda} e_t + H_{L,\lambda} e_{t-1} + v_t, \quad v_t \sim \mathcal{N}(0, \Omega_v). \tag{6}$$

Note that this general specification nests all the four possible cases for modeling parameter variation described above: (i) exogenous; (ii) contemporaneous; (iii) lagged; and (iv) contemporaneous and lagged. In particular, depending on whether the corresponding matrices $H_{C,\lambda}$ and $H_{L,\lambda}$ are set to zero, the vector of innovations in (2) can be expressed in four different ways:

$$V_t = \begin{cases} \text{(I)} \;\; v_t & \text{, if } H_{C,\lambda} = 0,\ H_{L,\lambda} = 0, \\ \text{(II)} \;\; H_{C,\lambda} e_t + v_t & \text{, if } H_{C,\lambda} \neq 0,\ H_{L,\lambda} = 0, \\ \text{(III)} \;\; H_{L,\lambda} e_{t-1} + v_t & \text{, if } H_{C,\lambda} = 0,\ H_{L,\lambda} \neq 0, \\ \text{(IV)} \;\; H_{C,\lambda} e_t + H_{L,\lambda} e_{t-1} + v_t & \text{, if } H_{C,\lambda} \neq 0,\ H_{L,\lambda} \neq 0. \end{cases} \tag{7}$$

Given its generality, the model defined by equations (1) and (6) constitutes our benchmark specification in the empirical applications in Section 4.[11]

3 Estimation

In this section we outline the estimation techniques adopted to estimate the models discussed in Section 2. In particular, we propose an algorithm to estimate endogenous TVP-VARs that is based on simple Gibbs sampling steps.[12] The drifting coefficients (ϕ_t) are estimated using precision-based (or precision sampling) methods as in Chan and Jeliazkov (2009), instead of Kalman filter-based techniques.[13] Precision-based algorithms can be loosely viewed as a 'vectorized' version of the Kalman filter in that precision sampling operates directly on a representation of (1) and (6) where all variables are stacked over $t = 1, \cdots, T$. As a result, precision-based algorithms provide a setup that is arguably simpler than Kalman filter-based alternatives when conducting Bayesian estimation of state-space models. In particular, precision sampling methods do not require the recursive algebraic steps to derive filtering and smoothing equations, which are necessary when using the Kalman filter for Bayesian estimation of state-space models.[14].

Next, let $\mathbf{y} = (y_1, \ \cdots, \ y_T)'$ and the vector of innovations $\mathbf{e}$ and $\mathbf{v}$ being similarly defined, equations (1) and (6) can thus be expressed using the following stack representation:

$$\mathbf{y} = \widetilde{\mathbf{X}}\widetilde{\phi} + \mathbf{L_A}\mathbf{e}, \tag{8}$$

$$\widetilde{\phi} = \widetilde{\phi}_0 + \mathbf{L}_{\boldsymbol{\lambda}}\mathbf{e} + \mathbf{v}, \tag{9}$$

$$\begin{bmatrix} \mathbf{e} \\ \mathbf{v} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_{NT} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_v \end{bmatrix} \right), \tag{10}$$

[11] Nevertheless, for comparison, we also show selected results associated with more restrictive specifications.

[12] Evaluation of the mixing properties of our algorithm is available in the Online Appendix.

[13] As pointed out in McCausland et al. (2011), precision-based methods typically reduce computational complexity and expedite state simulation.

[14] For a detailed comparison between Kalman filter- and precision-based algorithms, see McCausland et al. (2011)

where:

$$\widetilde{\phi} = \mathbf{L}\phi = \underbrace{\begin{bmatrix} I & 0 & \cdots & 0 \\ -I & I & & \\ 0 & -I & \ddots & \vdots \\ \vdots & & \ddots & \\ 0 & & \cdots & -I \quad I \end{bmatrix}}_{\mathbf{L}} \underbrace{\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \\ \phi_T \end{bmatrix}}_{\phi}, \quad \widetilde{\mathbf{X}} = \mathbf{X}\mathbf{L}^{-1} = \underbrace{\begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ & & & \\ 0 & 0 & \cdots & X_T \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} I & 0 & \cdots & 0 \\ I & I & & \\ I & I & \ddots & \vdots \\ \vdots & & \ddots & \\ I & I & \cdots & I \quad I \end{bmatrix}}_{\mathbf{L}^{-1}},$$

$$\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{e} = \underbrace{\begin{bmatrix} H_{C,\lambda} & 0 & \cdots & & 0 \\ H_{L,\lambda} & H_{C,\lambda} & \cdots & & 0 \\ 0 & H_{L,\lambda} & \ddots & & \\ \vdots & & \ddots & & \vdots \\ 0 & & \cdots & H_{L,\lambda} & H_{C,\lambda} \end{bmatrix}}_{\mathbf{L}_{\boldsymbol{\lambda}}} \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ \\ e_T \end{bmatrix}}_{\mathbf{e}}, \quad \mathbf{L}_{\mathbf{A}} = I_T \otimes A, \quad \boldsymbol{\Sigma}_v = I_T \otimes \Omega_v \text{ and}$$

$\widetilde{\phi}_0 = (\phi_0,\ 0, \cdots, 0)'$. ϕ_0 represents an $N^2p+N \times 1$ vector that collects initialization conditions for the VAR coefficients. We treat such conditions as additional parameters that are estimated under our MCMC algorithm. Each identity matrix (I) in $\mathbf{L}$ and $\mathbf{L}^{-1}$ is $N^2p+N \times N^2p+N$. All other elements in the matrices above are defined exactly as discussed in Section 2. Finally, note that the three endogenous TVP-VAR variants discussed in Section 2.2 can be obtained by simply applying the exclusion restrictions in (7) to the appropriate band in $\mathbf{L}_{\boldsymbol{\lambda}}$, i.e.

$$\mathbf{L}_{\boldsymbol{\lambda}} = \left\{ \underset{\text{(Contemporaneous + Lagged)}}{\begin{bmatrix} H_{C,\lambda} & 0 & \cdots & & 0 \\ H_{L,\lambda} & H_{C,\lambda} & \cdots & & 0 \\ 0 & H_{L,\lambda} & \ddots & & \\ \vdots & & \ddots & & \vdots \\ 0 & & \cdots & H_{L,\lambda} & H_{C,\lambda} \end{bmatrix}}, \underset{\text{(Contemporaneous)}}{\begin{bmatrix} H_{C,\lambda} & 0 & \cdots & 0 \\ 0 & H_{C,\lambda} & \cdots & 0 \\ 0 & 0 & \ddots & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & H_{C,\lambda} \end{bmatrix}}, \underset{\text{(Lagged)}}{\begin{bmatrix} 0 & 0 & \cdots & & 0 \\ H_{L,\lambda} & 0 & \cdots & & 0 \\ 0 & H_{L,\lambda} & \ddots & & \\ \vdots & & \ddots & & \vdots \\ 0 & & \cdots & H_{L,\lambda} & 0 \end{bmatrix}} \right\}.$$

3.1 Sampling Endogenous States

Now let $\boldsymbol{\theta} = \left\{ \mathbf{L}_{\mathbf{A}},\ \widetilde{\phi}_0,\ \mathbf{L}_{\boldsymbol{\lambda}},\ \boldsymbol{\Sigma}_v \right\}$ denote the set containing parameters for each of the models in Table 2. An MCMC sampler for the system in (8)-(10) can be summarized as a two-step algorithm where posterior draws are obtained by sequentially sampling from the following densities: (i) $f(\widetilde{\phi}|\mathbf{y},\boldsymbol{\theta})$; and (ii) $f(\boldsymbol{\theta}|\mathbf{y},\widetilde{\phi})$. In what follows, we focus on deriving an expression for the full conditional posterior, $f(\widetilde{\phi}|\mathbf{y},\boldsymbol{\theta})$, since it represents the main contribution of this paper on the estimation front. Sampling details for $\boldsymbol{\theta}$ and a discussion on the priors can be found in the Online Appendix.[15]

To sample $\widetilde{\phi}$ from $f(\widetilde{\phi}|\mathbf{y},\boldsymbol{\theta})$, one first needs to obtain an expression for the likelihood function, $\mathscr{L}\left(\widetilde{\phi},\boldsymbol{\theta}|\mathbf{y}\right) = f(\mathbf{y}|\widetilde{\phi},\boldsymbol{\theta})$, and the prior density, $f(\widetilde{\phi}|\boldsymbol{\theta})$, i.e. the marginal distribution of $\widetilde{\phi}$ unconditional on $\mathbf{y}$. The prior can be readily obtained from (9):

$$\widetilde{\phi}|\boldsymbol{\theta} \sim \mathscr{N}\left(\boldsymbol{\mu}_{\widetilde{\phi}},\ \boldsymbol{\Sigma}_{\widetilde{\phi}\widetilde{\phi}}\right), \tag{11}$$

where the first and second moments – conditional on $\boldsymbol{\theta}$ – in (11) are respectively given by:[16]

$$\boldsymbol{\mu}_{\widetilde{\phi}} = \mathbb{E}\left(\widetilde{\phi}_0 + \mathbf{L}_{\boldsymbol{\lambda}}\mathbf{e} + \mathbf{v}\right) = \widetilde{\phi}_0, \tag{12}$$

$$\boldsymbol{\Sigma}_{\widetilde{\phi}\widetilde{\phi}} = \mathbb{V}\mathrm{ar}\left(\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{e} + \mathbf{v}\right) = \mathbf{L}_{\boldsymbol{\lambda}}\mathbb{V}\mathrm{ar}(\mathbf{e})\mathbf{L}_{\boldsymbol{\lambda}}' + \mathbb{V}\mathrm{ar}(\mathbf{v}) = \mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\boldsymbol{\lambda}}' + \boldsymbol{\Sigma}_v. \tag{13}$$

Next, to derive an expression for the likelihood, we use a standard result for multivariate Normal distributions, namely the joint density $f(\mathbf{y},\widetilde{\phi}|\boldsymbol{\theta})$ can be factorized into a conditional and a marginal coordinate, i.e. $f(\mathbf{y},\widetilde{\phi}|\boldsymbol{\theta}) = f(\mathbf{y}|\widetilde{\phi},\boldsymbol{\theta})f(\widetilde{\phi}|\boldsymbol{\theta})$. By setting the marginal density to be the prior in (11), then the properties of the Normal distribution (see, e.g., Theorem 3.8 in Kroese and Chan (2014), chapter 3.6) ensure that the likelihood function $f(\mathbf{y}|\widetilde{\phi},\boldsymbol{\theta})$ is Gaussian such that:

$$\mathbf{y}|\widetilde{\phi},\boldsymbol{\theta} \sim \mathscr{N}\left(\boldsymbol{\mu}_{\mathbf{y}} + \boldsymbol{\Sigma}_{\widetilde{\phi}\mathbf{y}}'\boldsymbol{\Sigma}_{\widetilde{\phi}\widetilde{\phi}}^{-1}\left(\widetilde{\phi} - \boldsymbol{\mu}_{\widetilde{\phi}}\right),\ \boldsymbol{\Sigma}_{\mathbf{yy}} - \boldsymbol{\Sigma}_{\widetilde{\phi}\mathbf{y}}'\boldsymbol{\Sigma}_{\widetilde{\phi}\widetilde{\phi}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\phi}\mathbf{y}}\right). \tag{14}$$

The expression above introduces three new terms, $\boldsymbol{\mu}_{\mathbf{y}}$, $\boldsymbol{\Sigma}_{\mathbf{yy}}$ and $\boldsymbol{\Sigma}_{\widetilde{\phi}\mathbf{y}}$. The first two denote first and second moments, respectively, which – conditional on $\widetilde{\phi}$ and $\boldsymbol{\theta}$ – are obtained from the measurement equation in (8). Specifically, we have:

[15] It should be noted, however, that the class of priors adopted here are broadly in line with previous TVP-VARs studies, e.g., Cogley and Sargent (2005) and Primiceri (2005).

[16] To make notation less cumbersome – while explicitly mentioning what they are — we omit conditional factors in the expressions for expectation, variance and covariance operators.

$$\boldsymbol{\mu}_{\mathbf{y}} = \mathbb{E}\left(\widetilde{\mathbf{X}}\widetilde{\boldsymbol{\phi}} + \mathbf{L}_{\mathbf{A}}\mathbf{e}\right) = \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\phi}}, \tag{15}$$

$$\boldsymbol{\Sigma}_{\mathbf{yy}} = \mathbb{V}\mathrm{ar}\left(\mathbf{L}_{\mathbf{A}}\mathbf{e}\right) = \mathbf{L}_{\mathbf{A}}\mathbb{V}\mathrm{ar}(\mathbf{e})\mathbf{L}_{\mathbf{A}}{}' = \mathbf{L}_{\mathbf{A}}\mathbf{L}_{\mathbf{A}}{}'. \tag{16}$$

The cross-covariance term, $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\mathbf{y}}$, appears in (14) since the vector of structural shocks ($\mathbf{e}$) is a common driver to both $\mathbf{y}$ and $\widetilde{\boldsymbol{\phi}}$. Therefore, from (8) and (9) and recalling $\mathbf{e}$ and $\mathbf{v}$ are independent random vectors, we have:

$$\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\mathbf{y}} = \mathbb{C}\mathrm{ov}\left(\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{e},\ \mathbf{L}_{\mathbf{A}}\mathbf{e}\right) = \mathbf{L}_{\boldsymbol{\lambda}}\mathbb{C}\mathrm{ov}\left(\mathbf{e},\ \mathbf{e}\right)\mathbf{L}_{\mathbf{A}}{}' = \mathbf{L}_{\boldsymbol{\lambda}}\mathbb{V}\mathrm{ar}(\mathbf{e})\mathbf{L}_{\mathbf{A}}{}' = \mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\mathbf{A}}{}'. \tag{17}$$

Of course, in the absence of endogenous time variation, $\mathbf{L}_{\boldsymbol{\lambda}}$ is a null matrix and so is $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\mathbf{y}}$, which returns an expression for (14) that is consistent with the likelihood function for exogenous TVP-VARs.[17]

Finally, applying Bayes' rule to combine (11) and (14) and using the results in (12)-(13), (15)-(16) and (17) yields:

$$\begin{aligned} f(\widetilde{\boldsymbol{\phi}}|\mathbf{y},\boldsymbol{\theta}) &\propto f(\mathbf{y}|\widetilde{\boldsymbol{\phi}},\boldsymbol{\theta}) f(\widetilde{\boldsymbol{\phi}}|\boldsymbol{\theta}), \\ &\propto \exp\left[-\frac{\left(\mathbf{y}_* - \mathbf{B}\widetilde{\boldsymbol{\phi}}\right)' \mathbf{K}_{\mathbf{y}}{}^{-1}\left(\mathbf{y}_* - \mathbf{B}\widetilde{\boldsymbol{\phi}}\right) + \left(\widetilde{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}_0\right)' \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\widetilde{\boldsymbol{\phi}}}^{-1}\left(\widetilde{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}_0\right)}{2}\right], \\ &\propto \exp\left[-\frac{\widetilde{\boldsymbol{\phi}}'\left(\mathbf{B}'\mathbf{K}_{\mathbf{y}}{}^{-1}\mathbf{B} + \left(\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\boldsymbol{\lambda}}' + \boldsymbol{\Sigma}_v\right)^{-1}\right)\widetilde{\boldsymbol{\phi}} - 2\left(\mathbf{y}_*{}'\mathbf{K}_{\mathbf{y}}{}^{-1}\mathbf{B} + \widetilde{\boldsymbol{\phi}}_0'\left(\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\boldsymbol{\lambda}}' + \boldsymbol{\Sigma}_v\right)^{-1}\right)\widetilde{\boldsymbol{\phi}}}{2}\right], \end{aligned} \tag{18}$$

where we define:

$$\begin{aligned} \mathbf{y}_* &= \mathbf{y} + \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\mathbf{y}}'\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\widetilde{\boldsymbol{\phi}}}^{-1}\boldsymbol{\mu}_{\widetilde{\boldsymbol{\phi}}} = \mathbf{y} + \mathbf{L}_{\mathbf{A}}\mathbf{L}_{\boldsymbol{\lambda}}'\left(\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\boldsymbol{\lambda}}' + \boldsymbol{\Sigma}_v\right)^{-1}\widetilde{\boldsymbol{\phi}}_0, \\ \mathbf{B} &= \widetilde{\mathbf{X}} + \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\mathbf{y}}'\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\widetilde{\boldsymbol{\phi}}}^{-1} = \widetilde{\mathbf{X}} + \mathbf{L}_{\mathbf{A}}\mathbf{L}_{\boldsymbol{\lambda}}'\left(\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\boldsymbol{\lambda}}' + \boldsymbol{\Sigma}_v\right)^{-1}, \\ \mathbf{K}_{\mathbf{y}} &= \boldsymbol{\Sigma}_{\mathbf{yy}} - \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\mathbf{y}}'\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\widetilde{\boldsymbol{\phi}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\mathbf{y}} = \mathbf{L}_{\mathbf{A}}\mathbf{L}_{\mathbf{A}}{}' - \mathbf{L}_{\mathbf{A}}\mathbf{L}_{\boldsymbol{\lambda}}'\left(\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\boldsymbol{\lambda}}' + \boldsymbol{\Sigma}_v\right)^{-1}\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\mathbf{A}}{}'. \end{aligned}$$

The expression in (18) reveals a Gaussian kernel such that:

$$\widetilde{\boldsymbol{\phi}}|\mathbf{y},\boldsymbol{\theta} \sim \mathscr{N}\left(\overline{\mathbf{d}}_{\widetilde{\boldsymbol{\phi}}},\ \overline{\mathbf{D}}_{\widetilde{\boldsymbol{\phi}}}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{\widetilde{\boldsymbol{\phi}}} = \overline{\mathbf{D}}_{\widetilde{\boldsymbol{\phi}}}\left(\mathbf{B}'\mathbf{K}_{\mathbf{y}}{}^{-1}\mathbf{y}_* + \left(\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\boldsymbol{\lambda}}' + \boldsymbol{\Sigma}_v\right)^{-1}\widetilde{\boldsymbol{\phi}}_0\right), \\ \overline{\mathbf{D}}_{\widetilde{\boldsymbol{\phi}}} = \left(\mathbf{B}'\mathbf{K}_{\mathbf{y}}{}^{-1}\mathbf{B} + \left(\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{L}_{\boldsymbol{\lambda}}' + \boldsymbol{\Sigma}_v\right)^{-1}\right)^{-1}. \end{cases} \tag{19}$$

[17] A similar rationale applies for the prior variance-covariance matrix $\boldsymbol{\Sigma}_{\widetilde{\boldsymbol{\phi}}\widetilde{\boldsymbol{\phi}}}$ in (13).

To produce draws for $\widetilde{\phi}|\mathbf{y},\boldsymbol{\theta}$ one needs to construct $\overline{\mathbf{d}}_{\widetilde{\phi}}$ and $\overline{\mathbf{D}}_{\widetilde{\phi}}$. This can be done using the posterior simulation algorithm of Chan and Jeliazkov (2009). Draws for $\phi|\mathbf{y},\boldsymbol{\theta}$ can then be recovered by simply computing $\phi = \mathbf{L}^{-1}\widetilde{\phi}$.

3.2 A Parametrization for Model Comparison

Before moving on to the empirical applications, it is useful to illustrate a strategy to carry out model comparison between endogenous and exogenous TVP-VARs. Within a Bayesian framework a natural way to conduct such an exercise is by computing the posterior odds, i.e. the ratio of posterior model probabilities between two competing specifications. In our case this can be formulated as:

$$\overbrace{\frac{f(\text{Endog. TVP-VAR}|\mathbf{y})}{f(\text{Exog. TVP-VAR}|\mathbf{y})}}^{\texttt{Posterior Odds}} = \overbrace{\frac{f(\mathbf{y}|\text{Endog. TVP-VAR})}{f(\mathbf{y}|\text{Exog. TVP-VAR})}}^{\texttt{Bayes Factor}} \times \overbrace{\frac{f(\text{Endog. TVP-VAR})}{f(\text{Exog. TVP-VAR})}}^{\texttt{Prior Odds}}.$$

It is common to assume that two models are equally likely a priori, hence the posterior odds above simplify to computing the Bayes factor.

Now, recall from (7) that endogenous TVP-VARs nest their exogenous counterpart. As a result, a convenient Bayesian tool to compute the Bayes factor is to apply the Savage-Dickey Density Ratio method (Verdinelli and Wasserman (1995)). This requires evaluating the following expression:[18]

$$\text{Bayes factor} = \frac{f(\boldsymbol{\lambda}=0)}{f(\boldsymbol{\lambda}=0|\mathbf{y})}, \tag{20}$$

where $\boldsymbol{\lambda}$ is the vector collecting contemporaneous and lagged sensitivity parameters in Equation (6), $\boldsymbol{\lambda} = \left(\lambda_{C,1},\ \cdots,\ \lambda_{C,N},\ \lambda_{L,1},\ \cdots,\ \lambda_{L,N}\right)'$.

From (20), it is clear that computation of the Bayes factor requires evaluating the marginal prior and posterior for $\boldsymbol{\lambda}$ at the restriction $\boldsymbol{\lambda}=0$. As demonstrated in the Online Appendix, such an exercise is simple and involves only evaluating the ratio between two Gaussian densities. In particular, the posterior in the denominator can be readily estimated via Monte Carlo integration

[18] Loosely speaking, the Savage-Dickey Density Ratio approach can be seen as a Bayesian analog to Wald-type tests in the sense that both approaches denote suitable statistical methods to compare nested models.

using Gibbs draws from the conditional posterior $f(\boldsymbol{\lambda}|\mathbf{y},\widetilde{\phi},\boldsymbol{\theta}_{-\boldsymbol{\lambda}})$, where we use $\boldsymbol{\theta}_{-\boldsymbol{\lambda}}$ to denote any parameter in $\boldsymbol{\theta}$ except for $\boldsymbol{\lambda}$.

To obtain posterior draws for $\boldsymbol{\lambda}$, first, we need to express the state equation in (9) in terms of $\boldsymbol{\lambda}$. This is an important step, since expressions for the prior and conditional posterior in terms of $\mathbf{L}_{\boldsymbol{\lambda}}$ are not of a known form. Thus, by a simple change of variable $\mathbf{L}_{\boldsymbol{\lambda}}\mathbf{e} = \mathbf{L}_{\mathbf{e}}\boldsymbol{\lambda}$, we can rewrite (9) as:

$$\widetilde{\phi} = \widetilde{\phi}_0 + \mathbf{L}_{\mathbf{e}}\boldsymbol{\lambda} + \mathbf{v}, \tag{21}$$

where

$$\mathbf{L}_{\mathbf{e}}\boldsymbol{\lambda} = \left(\begin{bmatrix} e_1' & 0' \\ e_2' & e_2' \\ \vdots & \vdots \\ e_T & e_{T-1}' \end{bmatrix} \otimes \iota \right) \begin{bmatrix} \lambda_{C,1} \\ \vdots \\ \lambda_{C,N} \\ \lambda_{L,1} \\ \vdots \\ \lambda_{L,N} \end{bmatrix}.$$

That is, $\mathbf{L}_{\mathbf{e}}$ denotes the Kronecker product between a $T \times 2N$ matrix containing the contemporaneous and lagged structural shocks and a column vector of ones, ι, which – as defined in Section 2.2 – has the same row dimension as ϕ_t.

Fortunately, given the modular nature of MCMC algorithms, one can use the conditionality of $\boldsymbol{\lambda}$ on $\mathbf{y}$, $\widetilde{\phi}$ and $\mathbf{L}_{\mathbf{A}}$ in $f(\boldsymbol{\lambda}|\mathbf{y},\widetilde{\phi},\boldsymbol{\theta}_{-\boldsymbol{\lambda}})$, to back out the vector of structural shocks $\mathbf{e}$ from (8) and treat $\mathbf{L}_{\mathbf{e}}$ as a predetermined regressor in Equation (21).[19] This means that derivation of an expression for $f(\boldsymbol{\lambda}|\mathbf{y},\widetilde{\phi},\boldsymbol{\theta}_{-\boldsymbol{\lambda}})$ – a crucial step to compute (20) – can be based on the simple linear regression form in (21) for which standard results can be applied (see, e.g., Koop et al. (2007)).

4 Macroeconomic Instabilities: Measurement and Sources

We now illustrate the usefulness of the proposed endogenous TVP-VAR framework by means of two empirical applications for the US economy. In the first application, we focus on identifying the shocks that might have contributed to changes in the persistence of the inflation gap. In

[19] For simplicity, we set pre-sample values of e_t to zero.

the second, we conduct scenario analysis for inflation and unemployment under different monetary policy decisions about the trajectory of interest rates. In all these applications, we employ a small-scale structural TVP-VAR along the lines of Cogley and Sargent (2001) and Cogley et al. (2010).[20] In keeping with these studies, we set lag-length in the VAR to two and estimate the models using quarterly data on inflation (the annual percentage change in the seasonally adjusted GDP implicit price deflator), the unemployment rate (seasonally adjusted civilian unemployment rate, all workers over age sixteen), and the short-term nominal interest rate for the US (yield on 3-month Treasury bill rate). Our original sample runs from 1948Q2 to 2018Q2, where the first fifteen years are used to train the priors for initializing the time-varying coefficients. Therefore, our effective sample runs from 1963Q2 until 2018Q2. To circumvent issues related to interest rates being operationally restrained by the zero lower bound (ZLB), for the period between 2009Q1 and 2015Q4 we splice the series for the short-term nominal rate with the shadow rate measure proposed in Wu and Xia (2016), which allows for negative realizations of the policy rate.[21]

Following the description of the models in Section 2, we first present results for homoskedastic endogenous TVP-VARs. In Section 4.5 we provide robustness checks for models that allow for heteroskedasticity. Also, unless stated otherwise, the results we report below are based on our most general endogenous TVP-VAR, i.e. the one that allows for the intercept and autoregressive coefficients to be affected by both contemporaneous and lagged structural shocks. All models were estimated using 40000 draws of the Gibbs sampling algorithm proposed in Section 3 for which the first 10000 burn-in draws were discarded. For brevity, further details on estimation as well as additional results for the other forms of endogenous TVP-VARs are reported in the Online Appendix.

4.1 Identification

Since we are interested in labeling VAR shocks in a manner that is consistent with economic theory, we adopt a sign-restriction approach (e.g., Uhlig (2005)) to identify the structural innova-

[20] Section A2 of the Online Appendix investigates the scalability of the algorithm proposed in this study. In particular, an exercise based on simulated data shows that computation of endogenous TVP-VARs remains tractable, speed-wise, when including up to seven variables (i.e., N=7) in the system. This is the same number of variables used in prototypical medium-scale DSGE models, such as in Smets and Wouters (2007).

[21] The data for inflation, unemployment and the interest rate were obtained from the Federal Reserve Bank of St. Louis website. Data for the Wu-Xia shadow federal funds rate were obtained from the Federal Reserve Bank of Atlanta website.

tions. As pointed out in Canova (2011), sign restrictions provide a strategy to incorporate non-parametric (sign-related) restrictions from DSGE models into VARs.

In the applications that follow we identify three shocks, namely a cost-push, a monetary policy and a demand shock. While these shocks, of course, do not represent an exhaustive list, they emerge as natural candidates to shed light on phenomena related to inflation and unemployment dynamics. For example, prototypical characterizations of New Keynesian DSGE models, that are tailored to study real and nominal effects of monetary policy, rely on a similar set of shocks (see, e.g., Clarida et al. (1999) and Lubik and Schorfheide (2004)). Moreover, recall that by allowing for additional, albeit unidentified, coefficient-specific errors we do not rule out potential effects from other extraneous sources in our TVP-VARs.

In particular, we assume that a positive cost-push (or negative supply) shock lowers real activity by increasing unemployment while leading inflation and the policy rate to rise. In contrast, a positive demand shock decreases the unemployment rate, but is accompanied by an increase in inflation and the policy rate. Finally, a positive – i.e. contractionary – monetary policy shock implies an increase in the policy rate while lowering inflation and increasing the unemployment rate. These restrictions are identical to the ones adopted in Fry and Pagan (2011) and are summarized in Table 1. To obtain an impact matrix (A) that conforms with the sign restrictions just described above – akin to Baumeister and Peersman (2013) and Mumtaz and Zanetti (2015) – we use Algorithm 2 in Rubio-Ramirez et al. (2010) to generate a rotation matrix that spans the space of all possible permutation of the signs in Table 1.[22] A detailed description of how we compute A within our Gibbs sampling algorithm is presented in Section A1.1 of the Online Appendix.

[22] It should be noted, however, that allowing for a rotation matrix to achieve sign identification comes with an important caveat. As pointed out in Baumeister and Hamilton (2015), the inclusion of a rotation matrix into the estimation procedure implicitly implies a prior for structural objects, such as impulse responses, that is informative and depends on the number of variables in the model. One exception is the case of VARs with $N = 3$ (as assumed in this paper), where such an implicit prior turns out to be an uninformative one. In a similar vein, Bognanni (2018) discusses ordering dependency issues for the estimation of TVP-VARs and also illustrates implications to structural objects in the context of sign-identified VARs. Note, however, that when the impact matrix (A) is time *invariant* (as assumed in this paper) the author's novel parametrization for the law of motion of the autoregressive coefficients (see Equation 54 in the author's working paper version) is equivalent to the standard random walk case adopted in this paper. Also, motivated by Bognanni's critique, we re-estimated our baseline model under different variable orderings, which produced virtually identical results.

4.2 Evidence on Endogenous Time Variation

We begin by evaluating the statistical evidence for endogenizing the time-varying coefficients. As discussed in Sections 2.2 and 3.2, validation of our methodology can be carried out by investigating whether the vector of loadings ($\boldsymbol{\lambda}$) should be introduced in the model. Figure 1 shows the posterior density estimates for each loading. Notably, all densities exhibit most of their masses located away from zero. Such a result lends support to the idea that structural shocks can indeed affect the dynamics of the VAR coefficients.

Next, we rely on a more formal statistical procedure to test whether $\boldsymbol{\lambda} = 0$. In particular, we use the expression in (20) to compute the Bayes factor between endogenous and exogenous TVP-VARs. Table 2 reports the list of models used in such an exercise. Table 3 reports values for twice the natural logarithm of the Bayes factor ($2\log(\text{BF})$) for cases when, loosely speaking, the model 'under the null hypothesis' is an exogenous TVP-VAR and the model 'under the alternative' is one of the three endogenous TVP-VARs in Table 2.[23] All in all, Bayes factor analysis reinforces our findings from Figure 1, i.e. regardless whether coefficients are driven by contemporaneous, lagged or both types of structural innovations, evidence supporting endogenous TVP-VARs is substantial. Notably, all values for $2\log(\text{BF})$ are greater than ten which, following Kass and Raftery (1995), can be interpreted as strong evidence in favor of endogenizing the dynamics of the coefficients. In terms of posterior odds, such results indicate that – for the empirical applications in this paper – endogenous TVP-VARs are more likely than their exogenous counterpart by a factor greater than 150.[24]

It is important to note that statistical evidence in favor of our framework does not require identifying all shocks in the system and including them jointly in the law of motion of the VAR coefficients, i.e. the methodology supports partial identification. To illustrate this, we compute the Bayes factors between endogenous and exogenous TVP-VARs when coefficients in the former are a function of a *single* structural shock. To be clear, in this exercise we estimate endogenous TVP-VARs where just one shock is identified.[25] The values for $2\log(\text{BF})$ are reported in Table 4. Our

[23] Kass and Raftery (1995) suggest the use of $2\log(\text{BF})$ since such a metric is on the same scale as other familiar test statistics like the deviance information criteria and the likelihood ratio test.

[24] Such result is based under the assumption that both endogenous and exogenous TVP-VARs are equally likely a priori. For more details on how to interpret the Bayes factor, see Section 3.2 in Kass and Raftery (1995).

[25] In other words, we impose sign restrictions only to the column in the impact matrix (A) associated with a particular shock of interest, while leaving the remaining columns in A unrestricted (sign-wise).

results favor endogenous over exogenous time variation for each of the three shocks considered in this paper. Nonetheless, models that condition coefficient variations on a single lagged structural shock (plus a coefficient-specific error) do not receive support from the data. This result indicates that – at least for the applications in this paper – if one is interested in identifying just one shock and exploring how it may affect interdependencies in a VAR, then accounting for its contemporaneous seems more appropriate. Put differently, partial identification of shocks is supported within the context of our empirical applications. Although, this support seems to be contingent on the timing of shocks. Of course, for different applications this result may change.

4.3 Application 1: Inflation-Gap Persistence

With modern central banking being largely concerned with committing to an inflation target, sustained credibility of a monetary authority hinges (among other things) on: (i) its ability to infer whether deviations of headline inflation from some underlying target are long-lasting; and (ii) identifying the potential drivers associated with such deviations. In light of these, Cogley et al. (2010) proposed a two-step exercise to study historical changes in the persistence of the US inflation-gap that combined measurement from an exogenous TVP-VAR and structural assessment based on a New-Keynesian DSGE model. Following their analysis, our first empirical application seeks to study both the dynamics and underlying drivers of inflation-gap persistence through the lens of our framework – i.e. the endogenous TVP-VAR model – and investigate how our findings relate to those in Cogley et al. (2010). In addition, since our sample runs until 2018Q2 – whereas theirs ends in 2006Q4 – we also explore what insights, if any, our methodology may offer for the more recent history of the inflation gap.

4.3.1 Measuring Persistence

To assemble descriptive statistical evidence about potential changes in inflation-gap persistence, Cogley et al. (2010) propose a metric that defines persistence in terms of the predictive content past shocks carry to future variations in a time series of interest. The idea being that if a process is persistent, then the predictive content of past shocks should take long to die out. The authors show that such concept can be formulated analogously to a time-varying $R^2_{h,t}$ statistic for

$h-$step ahead forecasts. Specifically, one can measure variations in the persistence of a time series by verifying, at any given point in time, the speed at which $R^2_{h,t}$ converges to zero as the forecast horizon h extends. For the sake of brevity, technical details on how to construct $R^2_{h,t}$ are relegated to the Online Appendix. For now, it suffices to remember that $R^2_{h,t}$ is a metric bounded between zero and one and that what characterizes changes in inflation-gap persistence is the pace at which $R^2_{h,t}$ converges to zero as h increases. Therefore, a highly (weakly) persistent process at time t is one whose $R^2_{h,t}$ statistic slowly (quickly) converges to zero as h lengthens.

Figure 2 plots the evolution of $R^2_{h,t}$ from 1962Q1 until 2018Q2 for $h = 1,\ 4$ and 8. Our estimates (posterior medians) indicate that the inflation gap was considerably more persistent during the 1970s and early 1980s than during the remainder of the sample. For example, in some instances during the 1970s, the $R^2_{h,t}$ remains close to one even at $h = 8$. In contrast, after the early 1980s the decline in inflation-gap persistence becomes apparent with $R^2_{h,t}$ dropping more rapidly as h increases for virtually all points in time. All in all, these results are broadly in line with what Cogley et al. (2010) documented.

Moreover, the latter part of our sample – that is not covered in their study – provides an additional result. Since 2015, inflation-gap persistence seems to be gradually picking up again. In fact, results reported in the Online Appendix for other endogenous TVP-VAR variants document an even steeper upward trend in $R^2_{h,t}$ beginning a bit after 2012. Interestingly, the emergence of such a 'trend' approximately coincides with the Federal Reserve (Fed hereafter) making the official announcement of pursuing a 2% inflation target. It is also consistent with commonly used measures of US inflation, such as 'core inflation', being persistently below 2% since the end of the 2008 financial crisis.

4.3.2 Sources of Changes in Inflation-Gap Persistence

We now illustrate how our framework can be used to gain insights into the possible causes underlying historical changes in inflation-gap persistence. To this end, it is important to recognize that variations in $R^2_{h,t}$ in Figure 2 are directly related to movements in the VAR coefficients. As previously mentioned, such movements are a function of four different sources, namely three structural shocks (cost-push, monetary policy and demand) and unidentified coefficient-specific

errors.[26] Therefore, to further interpret the results on inflation-gap persistence, we propose a statistical measure along the lines of Kose et al. (2003) that captures the contribution of each structural shock to the overall variation in the drifting coefficients. More precisely, note from (6) that changes in each of the elements in ϕ_t can be re-expressed in first-difference form $(\Delta\phi_{j,t} = \phi_{j,t} - \phi_{j,t-1})$ as:

$$\Delta\phi_{j,t} = \sum_{i=1}^{N} \left(\lambda_{C,i} e_{i,t} + \lambda_{L,i} e_{i,t-1}\right) + v_{j,t}, \quad \text{for } j = 1, ..., M, \tag{22}$$

where M denotes the total number of drifting coefficients, i.e. $M = N^2 p + N$.[27]

We are interested in providing a simple and informative way to interpret how structural shocks contribute to coefficient variations and, consequently, to changes in inflation-gap persistence. To this end, first recall that we assume structural shocks have unit-variance and that all innovations in (22) are mutually and serially uncorrelated. Also, as discussed in Section 2.2, to mitigate parameter proliferation we assumed that the term $\sum_{i=1}^{N} \left(\lambda_{C,i} e_{i,t} + \lambda_{L,i} e_{i,t-1}\right)$ is equivalent across all M states. With these in mind, a measure to summarize total variation in the VAR coefficients can be obtained by averaging the variance of $\Delta\phi_{j,t}$ over $j = 1, ..., M$. Formally, from (22), we set:

$$\begin{aligned} \texttt{Total Coefficient Variation} &= \frac{1}{M} \sum_{j=1}^{M} \mathbb{V}\text{ar}\left(\Delta\phi_{j,t}\right), \\ &= \sum_{i=1}^{N} \left(\lambda_{C,i}^2 + \lambda_{L,i}^2\right) + \frac{1}{M} \sum_{j=1}^{M} \sigma_{v,j}^2. \end{aligned} \tag{23}$$

The metric above decomposes changes in the VAR coefficients into two distinguishable components.[28] The first summation in (23) denotes the portion of parameter variation that is attributed to the structural shocks. The second summation represents the residual, or idiosyncratic, contribution due to coefficient-specific errors. As a result, the share of total coefficient variation associated with a particular structural innovation and coefficient-specific errors can be respectively defined as:

$$\texttt{Share (Shock i)} = \frac{\lambda_{C,i}^2 + \lambda_{L,i}^2}{\sum_{i=1}^{N} \left(\lambda_{C,i}^2 + \lambda_{L,i}^2\right) + \frac{1}{M} \sum_{j=1}^{M} \sigma_{v,j}^2}, \tag{24}$$

[26] Notably, for heteroskedastic TVP-VARs, changes in $R_{h,t}^2$ can also be induced by changes in second-moment parameters. In Section 4.5 we show that our main results are virtually unchanged when allowing for multiple volatility breaks.

[27] As previously defined, N denotes the number of structural shocks and p stands for lag-length.

[28] Since we assume that the innovations on the right-hand-side of (22) are independently and identically distributed, $\mathbb{V}\text{ar}\left(\Delta\phi_{j,t}\right)$ can be interpreted as an unconditional moment. When allowing for breaks in the volatility, as we do in Section 4.5, $\mathbb{V}\text{ar}\left(\Delta\phi_{j,t}\right)$ then becomes the conditional variance for $\Delta\phi_{j,t}$.

for i = Cost-Push, Demand, Monetary Policy, and

$$\texttt{Share (Residual)} = \frac{\frac{1}{M}\sum_{j=1}^{M}\sigma_{v,j}^2}{\sum_{i=1}^{N}\left(\lambda_{C,i}^2+\lambda_{L,i}^2\right)+\frac{1}{M}\sum_{j=1}^{M}\sigma_{v,j}^2}. \tag{25}$$

Table 5 reports the contribution from both identified and unidentified innovations to the overall degree of variation in the VAR coefficients according to (24) and (25). Such contributions sum up to one, which facilitates the interpretation of our results. All in all, we find that cost-push and monetary policy shocks seem to be the main driving forces behind coefficient changes in our model and, consequently, behind the changes in inflation-gap persistence reported in Figure 2. In particular, cost-push shocks account for more than 50% of overall parameter variation, followed by monetary policy shocks that contribute with around 30%. The remaining 20% is distributed almost equally between demand shocks and the coefficient innovations that the model does not identify.[29] As in Section 4.3.1, our results are again conceptually in line with Cogley et al. (2010) who also find that variations in the persistence of the US inflation gap over the past five decades are most likely related to cost-push and monetary policy shocks.[30]

Next, we examine the timing in the relationship between shocks and macroeconomic instabilities. By appropriately excluding $\lambda_{C,i}^2$ or $\lambda_{L,i}^2$ in the numerator of (24), Table 5 separates contemporaneous from lagged contributions to total coefficient variation for each structural innovation. Some features are noteworthy. For example, the contemporaneous contribution from cost-push and demand shocks is larger than their lagged counterpart. In contrast, for monetary policy shocks, the lagged contribution is larger than the contemporaneous one. Even if all these differences are not substantial, they seem to suggest that non-policy shocks generated by the aggregate behaviour of agents can alter macroeconomic relationships more promptly than shocks generated by policy makers. Such a result reinforces the traditional view that the effects of monetary policy on the economy are felt with lags, as posited in the seminal work by Friedman (1961) and more recently revisited in Havranek and Rusnak (2013).

[29] The results in Table 5 denote posterior medians, and percentiles, of the sampling distribution for the statistics in (24) and (25). To be clear, at each MCMC iteration we compute and retain the values for (24) and (25), which in turn gives us a distribution for these two metrics.

[30] Again, we emphasize that while our findings on what drives changes in inflation-gap persistence are obtained from an endogenous TVP-VAR model, in Cogley et al. (2010) they stem from a DSGE model.

4.4 Application 2: Monetary Policy Counterfactuals

In this section we carry out scenario analysis to investigate quantitative and qualitative differences between endogenous and exogenous TVP-VARs. We focus on two relatively recent episodes of high relevance for policy makers: (i) the path of the policy rate towards the ZLB after the onset of the 2008 financial crisis; and (ii) the period of monetary policy normalization that started in 2015. In particular, our first scenario assesses how inflation and unemployment would have responded had the Fed pursued a less accommodative (or 'less dovish') sequence of interest rate cuts during the early stages of the crisis. The second scenario evaluates inflation and unemployment responses to the normalization process of monetary policy had such a process been conducted under a more aggressive (or 'more hawkish') stance.

4.4.1 A Simulated Impulse Response Approach to Construct Scenarios

Our scenario analysis is based on generating simulated impulse responses (SIRs) for the variables in y_t – inflation and unemployment, more specifically – following a given sequence of simulated policy surprises. Such responses are constructed for both exogenous and endogenous TVP-VARs. To this end, first note that Equation (1) – for a lag length of two – can be recast as:

$$y_t = \mu_t + \Phi_{1,t} y_{t-1} + \Phi_{2,t} y_{t-2} + A e_t, \tag{26}$$

where μ_t and $\Phi_{p,t}$ (for $p = 1,\ 2$) denote an $N \times 1$ vector and a $N \times N$ matrix that collects the drifting intercepts and slope coefficients in ϕ_t, respectively.

Therefore, SIRs can be obtained by applying the following recursions:

$$\Xi_{t+h} = \begin{cases} A s_{t+h} & \text{for } h = 0, \\ \Phi_{1,t+h} \Xi_{t+h-1} + s_{t+h} & \text{for } h = 1, \\ \sum_{p=1}^{2} \Phi_{p,t+h} \Xi_{t+h-p} + s_{t+h} & \text{for } h = 2, \cdots, H, \end{cases} \tag{27}$$

where s_{t+h} denotes an $N \times 1$ vector containing simulated structural shocks at $t+h$, for $h = 0,\ 1, ..., H$.

A few comments are in order. First, to isolate the effects of monetary policy surprises, we set all other shocks (i.e. demand and cost-push) in s_{t+h} to zero for $h = 0,\ 1, ..., H$, i.e.:[31]

[31] Coefficient-specific innovations (v_t) are also set to zero over the impulse response horizon.

$$s_{t+h} = \begin{bmatrix} 0 \\ 0 \\ e_{t+h}^{mp,\ j} \end{bmatrix} \text{ for } j \in \{base,\ counter\} \text{ and } h = 0,\ 1, ..., H, \tag{28}$$

where $e_{t+h}^{mp,\ base}$ and $e_{t+h}^{mp,\ counter}$ denote simulated monetary policy shocks for the baseline and counterfactual scenarios, respectively.[32] Therefore, the third column of the 3×3 matrix Ξ_{t+h} collects the responses of the variables in y_t to simulated policy surprises at $h = 0,\ 1, ..., H$.

Second, the recursions in (27) essentially follow a textbook-like approach to compute *orthogonalized* impulse responses as in, e.g., Lütkepohl (2005). In other words, the dynamic responses for the variables in y_t are a function of the reduced-form autoregressive coefficients $(\Phi_{p,t})$, the impact matrix (A) – which in our case incorporates the sign restrictions – and a structural shock of interest that is contemporaneously uncorrelated to the remaining shocks in the system.[33] One distinction, however, is that we allow s_{t+h} to enter (27) additively for $h \geq 1$. This follows from the fact that we are interested in how the system in (26) responds to a *sequence* of policy surprises rather than to just a one-off perturbation at $h = 0$.

Third, recall that in the case of endogenous TVP-VARs, $\Phi_{p,t+h}$ is a function of structural shocks of interest. Consequently, when applying the recursions in (27) to these models, the autoregressive coefficients will change accordingly over the impulse response horizon every time a policy shock takes a non-zero value in s_{t+h}. This contrasts with the exogenous case, where the autoregressive coefficients will remain constant over the entire response horizon, since structural shocks do not enter the law of motion of the VAR coefficients in such models.

We now turn to discussing how policy surprises are simulated in our scenarios. For the baseline case, $e_{t+h}^{mp,\ base}$ in (28) is defined as a share, α_{t+h}, of the *observed* change in the policy rate from time $t+h-1$ to time $t+h$. That is, $e_{t+h}^{mp,\ base} = \alpha_{t+h} \Delta r_{t+h}^{base}$, where r_{t+h}^{base} denotes the actual policy rate figure at $t+h$. Notably, α_{t+h} is introduced to accommodate the idea that a change in the policy rate embeds a surprise (or unanticipated) component. To allow for uncertainty associated with such a component, we treat α_{t+h} as a (continuous) random variable drawn from a uniform distribution

[32] Details on how we simulate $e_{t+h}^{mp,\ base}$ and $e_{t+h}^{mp,\ counter}$ are provided below.

[33] Another common approach to define impulse responses for nonlinear models is to apply *generalized* impulse responses, which entails computing the difference between two conditional means. We compute these type of responses in Section A5 of the Online Appendix to illustrate the transmission of all shocks in our system more broadly.

defined over the interval from 0 to 0.5. – i.e. $\alpha_{t+h} \sim U(0, 0.5)$. Moreover, for additional flexibility, a different α_{t+h} is drawn at each point of the impulse response horizon, $h = 0,\ 1, ..., H$. Similarly, in the counterfactual case, we set $e_{t+h}^{mp,\ counter} = \alpha_{t+h} \Delta r_{t+h}^{counter}$, except now $r_{t+h}^{counter}$ denotes some *hypothetical* value for the policy rate at $t+h$ that is calibrated to reflect a more "dovish" or "hawkish" policy stance depending on the scenario of interest. Thus, the mechanics to generate SIRs can be summarized as follows: for a given estimate of $\Phi_{p,t}$ (at time t and $p = 1,\ 2$) and A, we run the recursions in (27) 1000 times such that for each run a different sequence of policy shocks is simulated – as discussed above – and collected into s_{t+h} (as in (28)). At the end of this procedure, a distribution of SIRs is generated as a result of the randomness associated with α_{t+h}.[34]

Lastly, and admittedly, albeit relying on impulse responses, our strategy to construct scenarios is somewhat different from extant work based on (exogenous) TVP-VARs, such as Baumeister and Benati (2013). Indeed, a common strategy to compute scenarios is to apply techniques such as the ones discussed in Waggoner and Zha (1999), Hamilton and Herrera (2004) or Bańbura et al. (2015). In the context of endogenous TVP-VARs, however, implementing the methods discussed in these studies entails considerable computational complexity which, in our view, warrants future research.[35]Nonetheless, notwithstanding measurement caveats, the policy shocks we simulate are, by construction, proportionally related to the actual and hypothetical paths of the policy rate. Therefore, the results from our SIR analysis are, at a minimum, informative in a qualitative sense.

4.4.2 Reaching the Zero Lower Bound

There is an ongoing debate about the macroeconomic consequences of the ZLB. On the one hand, some works provide evidence that the ZLB had strong and detrimental effects on the US

[34]We use the posterior medians of $\Phi_{1,t}$, $\Phi_{2,t}$ and A for the recursions in (27). For the endogenous TVP-VAR case, we also use the posterior median for the elements in the vector of loadings (λ) that pre-multiplies the policy shock in the law of motion for the VAR coefficients.

[35]More formally, approaches to construct counterfactuals as in Waggoner and Zha (1999) and Hamilton and Herrera (2004), rely on solving a system of equations where a sequence of shocks, s, is a function of the VAR parameters, ϕ and A, and the assumed path of some variable of interest, z, i.e. $s = \mathscr{F}(\phi, A, z)$. For constant parameter VARs and exogenous TVP-VARs, there is a mapping from ϕ, A and z to s, i.e. $\{\phi, A, z\} \rightarrow s$, that renders the problem of solving for s more tractable. In contrast, for endogenous TVP-VARs, such a mapping is a more complex one. Note that when autoregressive coefficients are themselves a function of the (contemporaneous and lagged) structural shocks, we have a mapping between VAR parameters and shocks where the latter appear both in the domain and codomain of this relationship, i.e. $\{\phi(s), A, z\} \rightarrow s$. Consequently, solving for s would entail solving a (potentially) highly nonlinear system of equations given by $s = \mathscr{F}(\phi(s), A, z)$. Similarly, applying the recursive approach in Bańbura et al. (2015) would entail not only extending their methodology and parametrization to exogenous TVP-VARs but also adjusting their technique to the endogenous TVP-VAR case.

economy. From a theoretical perspective, Gust et al. (2017) argue that the ZLB has represented a significant constraint for policy makers, which exacerbated the Great Recession and inhibited the subsequent recovery. In particular, the authors find that the ZLB accounted for about 30% of the substantial economic contraction exhibited in 2009 and a potentially even larger fraction of the slow recovery. Likewise, on the empirical side, Hess et al. (2012) provide evidence that the ZLB importantly constrained the ability of conventional monetary policy to limit the depth and duration of the Great Recession. The authors do so by relying on several models, including a TVP-VAR with exogenous time variation.

On the other hand, Debortoli et al. (2019), also by employing a TVP-VAR model, find no significant changes in the response of a number of US macroeconomic variables to some structural shocks during ZLB period. They attribute this result to the hypothesis of 'perfect substitutability' between conventional and unconventional monetary policy during such a period. Similarly, Swanson (2018) argues that the ZLB has not been and still does not represent a significant constraint for the Fed. In fact, Baumeister and Benati (2013) show that unconventional monetary policy has had a significant effect on both output growth and inflation for the U.S. and U.K. economies.

Notably, the debate in the studies outlined above takes a more positive standpoint on the macroeconomic consequences of the ZLB. In this paper, we take a slightly different, rather normative, perspective about the ZLB and ask the following question: What would have happened with unemployment and inflation rates had the policy rate not reached the ZLB? To shed light on this question, we compare the estimated impact of a sequence of monetary policy shocks on unemployment and inflation under two scenarios.

The first scenario employs a sequence of shocks derived from what was actually observed during the financial crisis, namely the Fed lowered the policy rate until it reached its (effective) ZLB in 2008Q4. In particular, there were six rate cuts from 2007Q2 until 2008Q3. In contrast, the counterfactual scenario postulates a 'less dovish' stance. Instead of six, we assume the Fed implemented four rate cuts since 2007Q2, hence leaving the policy rate at its 2008Q2 level, 2.09%.[36] Figure 3 illustrates the path of the policy rate under these two cases.

[36] Accordingly, policy surprises – constructed as discussed in Section 4.4.1 – in (the third row of) s_{t+h} will take non-zero values for $h = 0,\ 1,\ 2,\ 3,\ 4$ and 5 and for $h = 0,\ 1,\ 2$ and 3 in the baseline and counterfactual cases, respectively.

Chart B of Figure 3 plots the simulated response of the unemployment rate obtained with the exogenous TVP-VAR model under the 'dovish' and 'less dovish' scenarios for monetary policy. The reported bands correspond to the uncertainty associated with the unanticipated component of policy rate changes (α_{t+h}), as discussed in Section 4.4.1. Both responses coincide until the five-quarter-ahead horizon, since the sequences of shocks in both scenarios are the same up to this point by construction. Henceforth, the actual and counterfactual interest rate paths start to diverge, thus leading to different dynamic responses for unemployment. More specifically, our estimates indicate that, ceteris paribus, by keeping rate cuts until the ZLB, the Fed contributed to reduce the unemployment rate by more than a percentage point than if it had halted rate cuts in 2008Q1.

Chart C of Figure 3 reports equivalent scenarios for unemployment, except now (simulated) responses are constructed using the endogenous TVP-VAR model. Under both scenarios – and following the same sequence of policy shocks as in Chart B – the unemployment rate exhibits larger and more persistent contractions than the ones obtained with the exogenous TVP-VAR. In particular, at the lowest point of the impulse response in the ZLB case, the impact of the policy shock on unemployment from the endogenous TVP-VAR is nearly a full percentage point stronger than what we observe for the exogenous TVP-VAR. Such results can be attributed to the self-exciting mechanism in the endogenous setting – supported by the data – that allows the structure of the macroeconometric model to be directly affected by policy shocks. In other words, not only monetary policy shocks can affect unemployment via the standard mechanisms in impulse response analysis, but they can also alter the relationship between economic variables – summarized by the VAR coefficients – along the response horizons.

Next, for inflation, our results in Charts D and E suggest that the differences between exogenous and endogenous TVP-VARs are less stark than those observed for unemployment scenarios. Such results are consistent with the analysis in Section 4.3, where a less persistent inflation gap – possibly reflecting better 'anchorage' of inflation expectations – might dampen substantial transitory fluctuations in inflation following monetary policy surprises.

4.4.3 Engaging in Monetary Policy Normalization

Six years after the end of the Great Recession, the Fed decided to engage in a path to raise interest rates, commonly referred to as the normalization of monetary policy. The 'Policy Normal-

ization Principles and Plans', presented in September 2014 by the FOMC established three goals: (i) begin increases in the short-term interest rate, constituting the end of the ZLB; (ii) reduce the size of the Fed's balance sheet so that monetary policy can work as it did before the Great Recession; and (iii) transform the Fed's asset holdings to a composition similar to those of pre-Great Recession times. In December 2015, the FOMC took the first step in implementing this plan.

Only a few years have passed since the beginning of the monetary policy normalization process, and yet another debate is starting to grow. This one regards whether the actions taken by the Fed in this normalization process have been adequate or not. On the one hand, Powell (2018) and Williams (2018) argued in recent speeches that the timing and magnitudes for rate hikes were appropriate. This is sustained by an unemployment rate below the FOMC's estimates of its long-run natural rate, and a much less responsive inflation to changes in resource utilization. On the other hand, Feldstein (2017) advocated in another speech that the Fed could reduce the risk of a financial correction by raising interest rates more quickly than it currently projects. Specifically, reaching a nominal rate of 4% by the end of 2019 or 2020, hence aiming for a real rate of 2% given an inflation target of 2%.

There is still scarce literature that focuses on the macroeconomic effects of monetary policy normalization in the US economy. To the best of our knowledge, this is the first study that empirically addresses the implications of alternative normalization paths by quantifying their associated effects on real activity and inflation. To this end, the scenario analysis in this section is based on the following question: what would be the implications to unemployment and inflation if the Fed had increased the policy rate at a stronger pace during the normalization process? Our approach to construct such scenarios is similar to the one discussed in Section 4.4.2. More precisely – constructing the unanticipated component embedded in interest rate changes as in Section 4.4.2 – we evaluate the effects of monetary policy shocks on unemployment and inflation since 2015Q4 under two alternative scenarios. The first one represents the actual normalization path taken by the Fed, which consisted of a 200 basis-point increase carried out over three years. In particular, starting in 2015Q4, the Fed raised rates by 25 basis-point increases with the last hike occurring in 2018Q4.[37]

[37] In 2019Q3 the Fed resumed rate cuts. Moreover, around the time we finished writing this paper the world economy was hit by a major economic shock associated with the coronavirus pandemic (COVID-19). As a result, the effective federal funds rate is currently back to its effective ZLB of 0.05%. For future work it could be interesting to apply our methodology to investigate macroeconomic instabilities associated with such a global shock.

It is important to note that there were no rate hikes in 2016Q1, 2016Q2, 2016Q3 and 2017Q3. In the second scenario, we assume that the Fed incurred a stronger normalization path increasing the policy rate by 200 basis points in two years instead of three. In this case, the Fed would have raised the policy rate by 25 basis-point increases at every quarter starting in 2015Q4 with the last hike happening in 2017Q3. Therefore, the counterfactual constitutes a 'more hawkish' scenario for policy normalization.

Charts B and C of Figure 4 plot the simulated responses for unemployment obtained with the exogenous and endogenous TVP-VAR models, respectively. In both cases, unemployment rises sooner under the faster normalization scenario. Nevertheless, when assuming endogenous time variation, the increases in the unemployment rate are greater and take longer to die out than what is obtained from the exogenous TVP-VAR. Again, the amplification and persistence effects in the endogenous setting can be attributed to the additional mechanism in such a model that allows for changes in the relationship between macroeconomic variables as monetary policy surprises materialize.

Charts D and E of Figure 4 show the response of inflation to monetary policy shocks under exogenous and endogenous time variation, respectively. As in the scenarios for Section 4.4.2, the differences between endogenous and exogenous TVP-VARs are less pronounced than those obtained for unemployment. Nevertheless, it is worth noting that the inflation overshooting that occurs around horizon 10 in the exogenous case is more substantial than that observed for the endogenous TVP-VAR. This suggests that the latter model better mitigates price-puzzle related issues that may occur over the scenario profile.[38]

4.5 Robustness Checks

We examined the robustness of our proposed framework along several dimensions, namely allowing for heteroskedasticity, conducting prior sensitivity analysis and applying a different identification strategy. Overall, our main findings carry over to all these checks. In the interest of space, below we provide a brief description of these checks. More details can be found in Section A4 of the Online Appendix. In particular, we: (i) allow for two different methods to account for changes

[38] Note that sign restrictions are imposed only on the contemporaneous responses, i.e. at $h = 0$. Therefore, the price-puzzle could still be present at further horizons, i.e. when $h > 0$.

in the volatility of reduced-form errors, i.e., fixing the break dates (according to Fed chairmanship periods) and stochastic volatility;[39] (ii) conduct a model comparison exercise under alternative settings on the priors for the loading vector ($\boldsymbol{\lambda}$) and second-moment parameters;[40] and (iii) adopt an identification approach based on short- and long-run restrictions to illustrate the implications an alternative identification strategy may have on the proposed class of models.[41]

5 Conclusion

TVP-VARs have been widely used to infer instabilities amongst economic variables. In the context of such models, parametric instabilities are commonly specified as drifting coefficients. Interestingly, formal econometric procedures to infer why such coefficients may drift is something that has remained largely unaddressed to date. In this paper, we developed a new class of TVP-VARs that shows that coefficient changes can, in great part, be explained by the structural shocks identified within such models, hence yielding a self-exciting, or endogenous, dynamic system.

We applied the proposed endogenous TVP-VAR framework to study macroeconomic instabilities in the US economy. In our modeling strategy, allowing for endogenous changes in the VAR coefficients provided structural inference that is comparable to traditional (or exogenous) TVP-VARs, while offering valuable information regarding the underlying sources of time variation. In particular, we showed that cost-push shocks have been a prominent driver behind transitory fluctuations in the US inflation rate. Moreover, scenario analysis indicated that the effects of monetary policy shocks on unemployment tend to be more persistent and amplified when allowing for the possibility that policy actions can impact coefficients in a reduced form setting. This result is very much in line with the spirit of the Lucas critique (Lucas (1976)).

Finally, future research could leverage the class of models proposed in this paper to investigate the role of economic shocks as drivers of parameter changes in a more micro-founded sense (e.g. changes in elasticities, discount rates, policy and pricing parameters). Such a type of analysis

[39] Section A3 in the Online Appendix details how to augment our MCMC algorithm when allowing for stochastic volatility.

[40] We also conduct model comparison while relaxing the diagonal assumption for the covariance matrix associated with the coefficient-specific errors (Ω_ν).

[41] That is, instead of *set identification* of the parameters in the impact matrix (as in the case of sign restrictions), we investigate whether *point identification* of such parameters confirms the results obtained with sign restrictions.

could explore, for instance, the functional relationship between a DSGE model with time-varying parameters and its (approximate) reduced form representation as a TVP-VAR. Another useful extension could be to develop TVP-VARs where endogenous time variation affects first and second moment parameters jointly.

References

Anderson, H. M. and Vahid, F. (1998). Testing multiple equation systems for common nonlinear components. *Journal of Econometrics*, 84(1):1–36.

Auerbach, A. J. and Gorodnichenko, Y. (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy*, 4(2):1–27.

Bańbura, M., Giannone, D., and Lenza, M. (2015). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of forecasting*, 31(3):739–756.

Baumeister, C. and Benati, L. (2013). Unconventional monetary policy and the great recession: Estimating the macroeconomic effects of a spread compression at the zero lower bound. *International Journal of Central Banking*, 9(2):165–212.

Baumeister, C. and Hamilton, J. (2015). Sign restrictions, structural vector autoregressions, and useful prior information. *Econometrica*, 83(5):1963–1999.

Baumeister, C. and Peersman, G. (2013). Time-varying effects of oil supply shocks on the us economy. *American Economic Journal: Macroeconomics*, 5(4):1–28.

Bianchi, F. and Civelli, A. (2015). Globalization and inflation: Evidence from a time-varying var. *Review of Economic Dynamics*, 18(2):406–433.

Bognanni, M. (2018). A class of time-varying parameter structural vars for inference under exact or set identification. *Federal Reserve Bank of Cleveland, Working Paper*, 18-11.

Canova, F. (1993). Modelling and forecasting exchange rates with a bayesian time-varying coefficient model. *Journal of Economic Dynamics & Control*, 17:233–261.

Canova, F. (2011). *Methods for Applied Macroeconomic Research*. Princeton University Press.

Canova, F., Ferroni, F., and Matthes, C. (2015). Approximating time varying structural models with time invariant structures. *Banque de France Working Paper*, (578).

Canova, F. and Gambetti, L. (2009). Structural changes in the us economy: Is there a role for monetary policy? *Journal of Economic Dynamics & Control*, 33:477–490.

Carriero, A., Clark, T. E., and Marcellino, M. G. (2018). Endogenous uncertainty. *FRB of Cleveland Working Paper*, (18-05).

Chan, J. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Clarida, R., Gali, J., and Gertler, M. (1999). The science of monetary policy: a new keynesian perspective. *Journal of economic literature*, 37(4):1661–1707.

Clark, T. and Terry, S. (2010). Time variation in the inflation passthrough of energy prices. *Journal of Money, Credit and Banking*, 42(7):1419–1433.

Cogley, T., Matthes, C., and Sbordone, A. M. (2015). Optimized taylor rules for disinflation when agents are learning. *Journal of Monetary Economics*, 72:131–147.

Cogley, T., Primiceri, G., and Sargent, T. (2010). Inflation-gap persistence in the us. *American Economic Journal: Macroeconomics*, 2(1):43–69.

Cogley, T. and Sargent, T. (2001). Evolving post-world war ii u.s. inflation dynamics. *NBER Macroeconomics Annual*, pages 331–373.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post wwii us. *Review of Economic Dynamics*, 8(2):262–302.

Cooper, R. W. and Haltiwanger, J. C. (2006). On the nature of capital adjustment costs. *The Review of Economic Studies*, 73(3):611–633.

Debortoli, D., Gali, J., and Gambetti, L. (2019). On the empirical (ir)relevance of the zero lower bound constraint. In *NBER Macroeconomics Annual*, page Forthcoming. MIT press.

Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, 3(1):1–100.

Feldstein, M. (2017). Normalizing monetary policy. Remarks by Martin Feldstein at the Cato Institute's 35th Annual Monetary Conference.

Friedman, M. (1961). The lag in effect of monetary policy. *Journal of Political Economy*, 69(5):447–466.

Fry, R. and Pagan, A. (2011). Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature*, 49(4):938–60.

Gali, J. and Gambetti, L. (2015). The effects of monetary policy on stock market bubbles: some evidence. *American Economic Journal: Macroeconomics*, 7(1):233–257.

Galvão, A. B. C. (2006). Structural break threshold vars for predicting us recessions using the spread. *Journal of Applied Econometrics*, 21(4):463–487.

Gust, C., Herbst, E., Lopez-Salido, D., and Smith, M. (2017). The empirical implicatons of the interest-rate lower bound. *American Economic Review*, 107(7):1971–2006.

Hamilton, J. D. and Herrera, A. M. (2004). Comment: oil shocks and aggregate macroeconomic behavior: the role of monetary policy. *Journal of Money, credit and Banking*, pages 265–286.

Havranek, T. and Rusnak, M. (2013). Transmission lags of monetary policy: A meta-analysis. *International Journal of Central Banking*, 9(4):39–76.

Hess, C., Laforte, J., Reifschneider, D., and Williams, J. (2012). Have we underestimated the likelihood and severity of zero lower bound events? *Journal of Money, Credit and Banking*, 44(1):47–82.

Kang, K. (2014). Estimation of state-space models with endogenous markov regime-switching parameters. *Econometrics Journal*, 17:56–82.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kim, C., Piger, J., and Startz, R. (2008). Estimation of markov regime-switching regression models with endogenous switching. *Journal of Econometrics*, 143:263–273.

Koop, G., Poirier, D. J., and Tobias, J. L. (2007). *Bayesian Econometric Methods*. Cambridge University Press.

Kose, M. A., Otrok, C., and Whiteman, C. (2003). International business cycles: World, region and country-specific factors. *American Economic Review*, 93(4):1216–1239.

Kroese, D. P. and Chan, J. C. (2014). *Statistical Modeling and Computation*. Springer.

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

Lucas, R. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

McCausland, W. J., Miller, S., and Pelletier, D. (2011). Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics & Data Analysis*, 55(1):199–212.

Mumtaz, H. and Surico, P. (2009). Time-varying yield curve dynamics and monetary policy. *Journal of Applied Econometics*, 24:895–913.

Mumtaz, H. and Theodoridis, K. (2019). Dynamic effects of monatery policy shocks on macroeconomic volatility. *Journal of Monetary Economics*, Forthcoming.

Mumtaz, H. and Zanetti, F. (2015). Labor market dynamics: a time-varying analysis. *Oxford Bulletin of Economics and Statistics*, 77(3):319–338.

Paul, P. (2019). The time-varying effect of monetary policy on asset prices. *Review of Economics and Statistics*, Forthcoming.

Powell, J. (2018). Monetary policy in a changing economy. Remarks by Chairman Jerom H. Powell at "Changing Market Structure and Implications for Monetary Policy," a symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72:821–852.

Rubio-Ramirez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, 77(2):665–696.

Sims, C. (1993). A nine-variable probabilistic macroeconomic forecasting model. *Business Cycles, Indicators and Forecasting, NBER Studies in Business Cycles*, 18:179–214.

Sims, C. and Zha, T. (2006). Were there regime switches in the u.s. monetary policy? *American Economic Review*, 96(1):54–81.

Smets, F. and Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 97(3):586–606.

Stock, J. and Watson, M. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Swanson, E. (2018). The federal reserve is not very constrained by the lower bound on nominal interest rates. *Brooking Papers on Economic Activity*, Fall.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*, 93(443):1188–1202.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

Verdinelli, I. and Wasserman, L. (1995). Computing bayes factors using a generalization of the savage-dickey density ratio. *Journal of the American Statistical Association*, 90(430):614–618.

Waggoner, D. and Zha, T. (1999). Conditional forecasts in dynamic multivariate models. *The Review of Economics and Statistics*, 81(4):639–651.

Williams, J. (2018). Moving towards 'normal' u.s. monetary policy. Remarks by John C. Williams at the Joint Bank Indonesia-Federal Reserve Bank of New York Central Banking Forum, Nusa Dua, Indonesia.

Wu, J. and Xia, F. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291.

Tables and Figures

Table 1: Summary of sign restrictions

	Cost-push shock	Monetary policy shock	Demand shock
Inflation	+	-	+
Unemployment	+	+	-
Interest Rate	+	+	+

Note: + (-) means a positive (negative) response of the variable in the row to a positive realization of the shock in the column.

Table 2: List of models

Identifier	Model Features
Model I	Exogenous TVP-VAR Parameter changes are driven by unidentified coefficient-specific innovations
Model II	Endogenous Contemporaneous TVP-VAR Parameter changes are driven by contemporaneous identified shocks and unidentified coefficient-specific innovations
Model III	Endogenous Lagged TVP-VAR Parameter changes are driven by lagged identified shocks and unidentified coefficient-specific innovations
Model IV	Endogenous Contemporaneous and Lagged TVP-VAR Parameter changes are driven by contemporaneous + lagged identified shocks and unidentified coefficient-specific innovations.

Table 3: Model comparison results between endogenous and exogenous TVP-VARs based on all identified shocks jointly affecting the VAR coefficients in the endogenous case, i.e. results based on full identification of shocks

	Model II	Model III	Model IV
$\boldsymbol{\lambda}$	19.81	19.83	48.84

Note: Model comparison results are based on computing $2\log(Bayes\ factor)$ between TVP-VARs with and without endogenous time variation. Values greater than ten should be interpreted as very strong evidence in favor of endogenous TVP-VARs. See Kass and Raftery (1995) for details on using the Bayes factor as a metric for model comparison. Bayes factor estimates are based on Equation (20). For Model IV $\boldsymbol{\lambda} = (\lambda_{C,1},\ \lambda_{C,2},\ \lambda_{C,3},\ \lambda_{L,1},\ \lambda_{L,2},\ \lambda_{L,3})'$. For all the others, $\boldsymbol{\lambda} = (\lambda_{j,1},\ \lambda_{j,2},\ \lambda_{j,3})'$ for $j = C,\ L$. Exogenous TVP-VARs assume $\boldsymbol{\lambda} = \mathbf{0}$.

Table 4: Model comparison results between endogenous and exogenous TVP-VARs based on a single identified shock affecting the VAR coefficients in the endogenous case, i.e. results based on partial identification of shocks

	Model II	Model III	Model IV
λ_1 (cost-push shock)	433.42	-17.74	37.77
λ_2 (demand shock)	150.91	-18.00	31.26
λ_3 (monetary policy shock)	19.62	-18.26	38.13

Note: Entries are based on computing $2\log(Bayes\ factor)$ between TVP-VARs with and without endogenous time variation. Values greater than ten should be interpreted as very strong evidence in favor of endogenous TVP-VARs. See Kass and Raftery (1995) for details on using the Bayes factor as a metric for model comparison. Bayes factor estimates are based on Equation (20). For Model IV $\lambda_i = (\lambda_{C,i}\ \lambda_{L,i})'$, for $i = 1,2,3$. For all the others $\lambda_i = \lambda_{j,i}$ for $j = C,\ L$ and $i = 1,2,3$. Exogenous TVP-VARs assume $\lambda_i = 0$ for $i = 1,2,3$.

Table 5: Contributions of structural shocks to parameter instability

Total contribution (%)			
Source of parameter instability	*Median*	*16th perc.*	*84th perc.*
Cost-push shock	53.16%	47.13%	58.37%
Demand shock	11.22%	8.07%	15.19%
Monetary Policy shock	28.42%	24.07%	33.20%
Residual	7.21%	7.05%	9.35%
Contemporaneous and lagged contributions			
Source of parameter instability	*Median*	*16th perc.*	*84th perc.*
Contemporaneous cost-push shock	27.11%	23.06%	31.17%
Lagged cost-push shock	26.08%	22.31%	29.45%
Contemporaneous demand shock	9.11%	6.27%	12.21%
Lagged demand shock	2.30%	1.13%	4.02%
Contemporaneous monetary policy shock	12.09%	9.13%	15.40%
Lagged monetary policy shock	16.11%	14.02%	19.41%

Note: Results are obtained using Equations (24) and (25). The bottom part of the table reports the contributions in the upper part of the table when separated into contemporaneous and lagged shocks.

Figure 1: Posterior density estimates for the loadings associated with each structural shock driving the VAR coefficients under Model IV described in Table 2

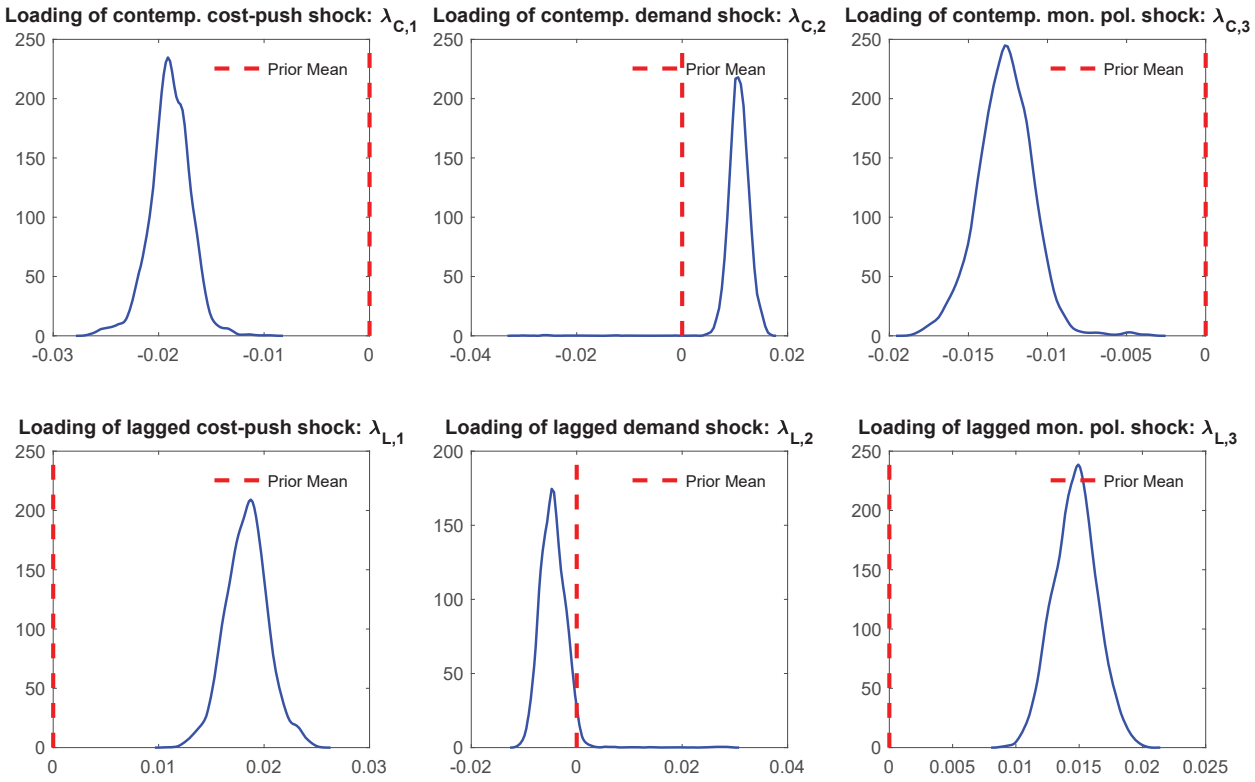
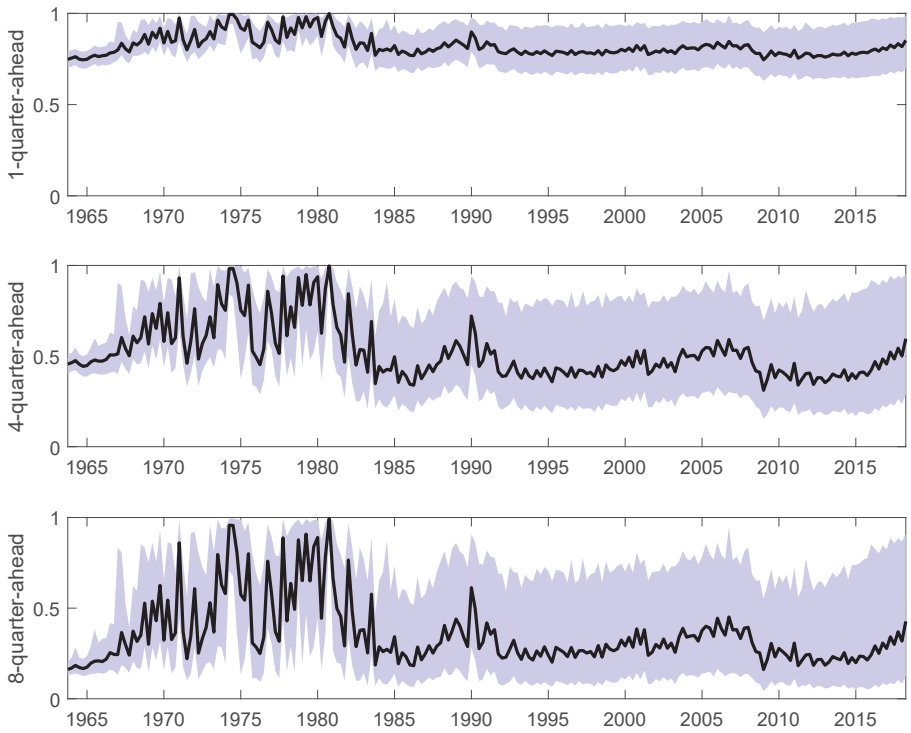


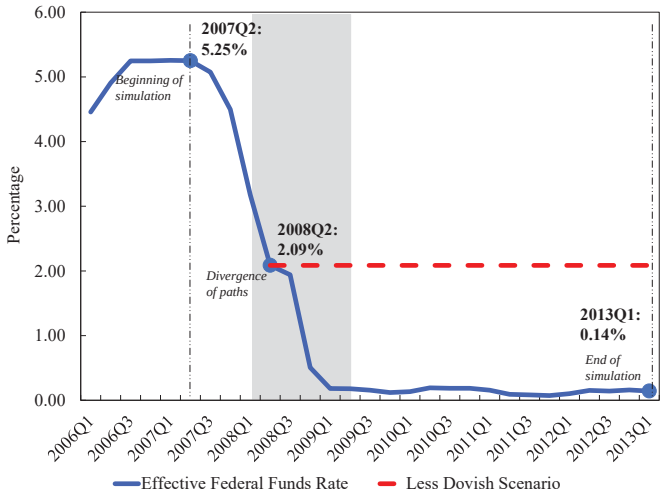
Figure 2: Inflation-gap persistence based on $R^2_{h,t}$ statistics



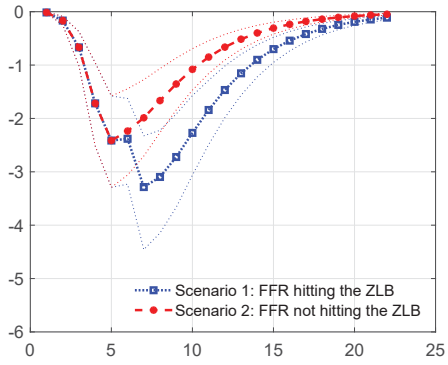
Note: The figure plots the time-varying measure of inflation-gap persistence, $R^2_{h,t}$, for the one-, four- and eight-quarter-ahead forecasting horizons. $R^2_{h,t}$ is constructed using Equation 12 in Cogley et al. (2010). Solid black lines denote posterior medians. Shaded areas represent the 16th and 84th percentiles of the corresponding posterior densities.

Figure 3: Counterfactual scenario for monetary policy in 2007: the trajectory towards the ZLB

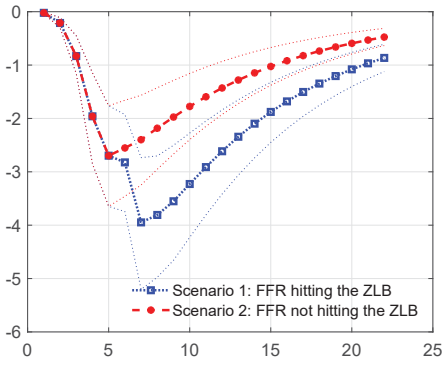
(a) Alternative Paths of Monetary Policy Stance



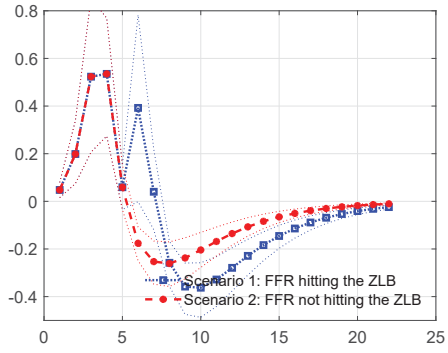
(b) Simulated Response of Unemp.: Exogenous



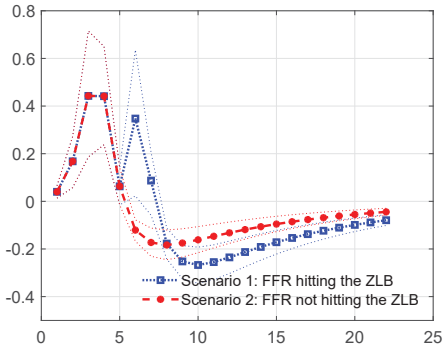
(c) Simulated Response of Unemp.: Endogenous



(d) Simulated Response of Inflation: Exogenous



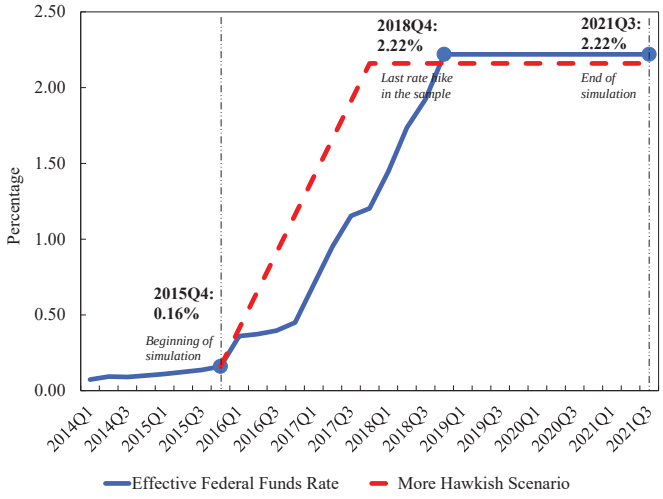
(e) Simulated Response of Inflation: Endogenous



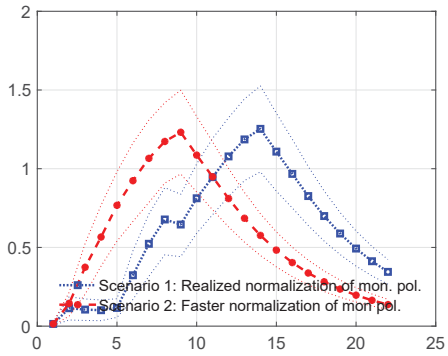
Note: Chart (a) plots the actual path of the federal funds rate (FFR) (blue line) and the alternative (non-ZLB) path of the policy rate (red line). Charts (b) and (c) plot the simulated response of the unemployment rate for both scenarios generated from an exogenous and endogenous TVP-VAR, respectively. Charts (d) and (e) plot the simulated response of inflation for both scenarios generated from an exogenous and endogenous TVP-VAR, respectively. Thin dotted lines represent the 16th and 84th percentile for the distribution of simulated responses in each scenario, as discussed in Section 4.4.1.

Figure 4: Counterfactual scenario for monetary policy in 2015: interest rate normalization

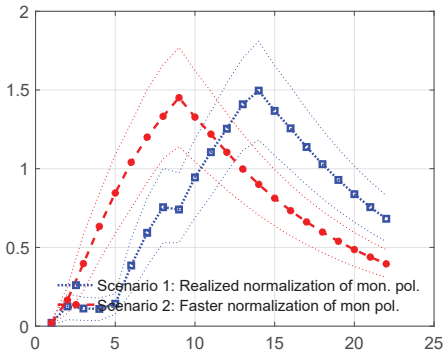
(a) Alternative Paths of Monetary Policy Stance



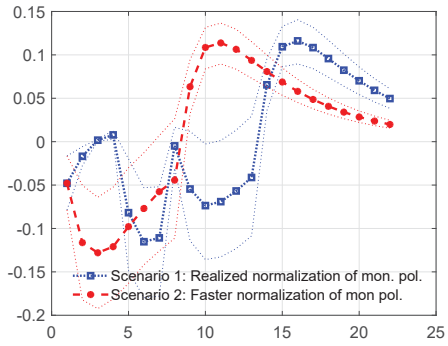
(b) Simulated Response of Unemp.: Exogenous



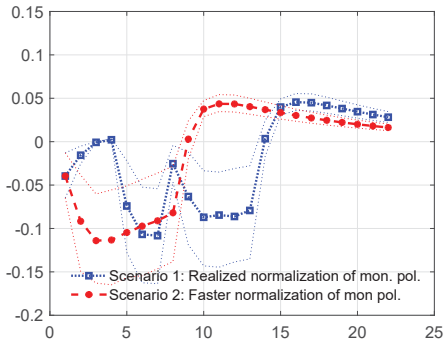
(c) Simulated Response of Unemp.: Endogenous



(d) Simulated Response of Inflation: Exogenous



(e) Simulated Response of Inflation: Endogenous



Note: Chart (a) plots the actual path of the policy rate normalization (blue line) and the faster normalization path of the policy rate (red line). Charts (b) and (c) plot the simulated response of the unemployment rate for both scenarios generated from an exogenous and endogenous TVP-VAR, respectively. Charts (d) and (e) plot the simulated response of inflation for both scenarios generated from an exogenous and endogenous TVP-VAR, respectively. Thin dotted lines represent the 16th and 84th percentile for the distribution of simulated responses in each scenario, as discussed in Section 4.4.1.

Online Appendix to *Endogenous Time Variation in Vector Autoregressions* *

Danilo Leiva-León† Banco de España

Luis Uzeda‡ Bank of Canada

Contents

*The views expressed are those of the authors. No responsibility for them should be attributed to the Bank of Canada, Banco de España or the Eurosystem.

†Danilo Leiva-León: Banco de España, Calle Alcalá, 48, 28014, Madrid, Spain Email: danilo.leiva@bde.es

‡Luis Uzeda: Bank of Canada, 234 Wellington Ave W, Ottawa, ON, K1A 0H9, Canada; Email: luzedagarcia@bank-banque-canada.ca

A1 Estimation Details

This appendix provides details on the MCMC algorithm used for estimation of endogenous TVP-VARs. It also elaborates on the computation of the Bayes factor and reports: (i) results for the mixing of the posterior sampler; and (ii) diagnostics for the stability of the VAR coefficients.

For convenience, we repeat below the stacked representation in Section 3.2 of the paper i.e.:

$$\mathbf{y} = \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\phi}} + \mathbf{L}_{\mathbf{A}}\mathbf{e}, \tag{A1}$$

$$\widetilde{\boldsymbol{\phi}} = \widetilde{\boldsymbol{\phi}}_{\mathbf{0}} + \mathbf{L}_{\boldsymbol{\lambda}}\mathbf{e} + \mathbf{v}, \tag{A2}$$

$$\begin{bmatrix} \mathbf{e} \\ \mathbf{v} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_{NT} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\boldsymbol{v}} \end{bmatrix} \right), \tag{A3}$$

where the reader is referred to Sections 3.1 and 3.2 of the paper for the exact structures of each vector and matrix above.

A1.1 Posterior Sampling

Priors

We assume standard independent priors for the four blocks of model parameters, namely $\boldsymbol{\lambda}$, $\boldsymbol{\phi_0}$, $\Omega_u = A'A$ and $\Omega_v = \texttt{diag}\left(\sigma_{v,1}^2, \cdots, \sigma_{v,M}^2\right)$, where, recall, M denotes the number of drifting coefficients. More specifically, we have:

$$\begin{aligned}
&\boldsymbol{\lambda} \sim \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}\right), \\
&\boldsymbol{\phi_0} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\phi_0}},\ \boldsymbol{\Sigma}_{\boldsymbol{\phi_0}}\right), \\
&\Omega_u \sim \mathcal{IW}\left(\nu_u,\ \mathbf{V}_u\right), \\
&\sigma_{v,i}^2 \sim \mathcal{IG}\left(\nu_{v,i},\ S_{v,i}\right) \text{ for } i = 1, ..., M.
\end{aligned}$$

We set $\boldsymbol{\Sigma}_{\boldsymbol{\lambda}} = \texttt{diag}(10^{-3}, ..., 10^{-3})$, $\boldsymbol{\Sigma}_{\boldsymbol{\phi_0}} = \texttt{diag}(10^{-5}, ..., 10^{-5})$ and $\boldsymbol{\mu}_{\boldsymbol{\phi_0}}$ – akin to Primiceri (2005) – denotes the OLS estimate from a fixed coefficient VAR estimated on the training sample (1947Q1-1961Q4). Moreover, we assume $\nu_u = 30$, $\mathbf{V}_u = 200I_N$; and $\nu_{v,i} = 60$ and $S_{v,i} = 0.01$ for $i = 1, ..., M$. Overall, this class of priors is standard in TVP-VAR analysis. In particular, the prior density for $\boldsymbol{\lambda}$ is tightly parameterized around zero to mitigate the possibility of endogenous time variation being artificially manufactured by very loose priors. The latter could force our posterior sampler to visit uninteresting regions of the parameter space where large values for $\boldsymbol{\lambda}$ could be prior – rather than data – induced.

Gibbs Sampling Steps

Let $\mathbf{s} = \{\boldsymbol{\phi},\ \boldsymbol{\lambda},\ \boldsymbol{\phi_0},\ A,\ \Omega_v\}$ denote the set of states and parameters for the system in (A1)-(A3), where notation $\mathbf{s}_{-j}$ represents all elements in $\mathbf{s}$ except for j. An MCMC algo-

rithm for the endogenous TVP-VARs introduced in Section 2 entails sequentially sampling from five conditional posterior distributions as summarized in the box below:

Summary of the Gibbs Sampling Algorithm

`Step-1` *Draw the vector of VAR coefficients (stacked over* $t = 1, ..., T$*),* $\boldsymbol{\phi}$*, from* $f(\boldsymbol{\phi}|\mathbf{y}, \mathbf{s}_{-\boldsymbol{\phi}})$,

`Step-2` *Draw the loadings,* $\boldsymbol{\lambda}$*, from* $f(\boldsymbol{\lambda}|\mathbf{y}, \mathbf{s}_{-\boldsymbol{\lambda}})$,

`Step-3` *Draw the coefficients' initial conditions,* $\boldsymbol{\phi_0}$*, from* $f(\boldsymbol{\phi_0}|\mathbf{y}, \mathbf{s}_{-\boldsymbol{\phi_0}})$,

`Step-4` *Drawing the impact matrix* A :

- `Step-4.1` *Draw the covariance matrix of the reduced-form innovations,* Ω_u*, from* $f(\Omega_u|\mathbf{y}, \mathbf{s}_{-\Omega_u})$,
- `Step-4.2` *Compute* $\texttt{Chol}(\Omega_u) = \tilde{A}$*, such that* $\Omega_u = \tilde{A}\tilde{A}'$,
- `Step-4.3` *Generate a rotation matrix* P *using Algorithm 2 in Rubio-Ramirez et al.* (2010),
- `Step-4.4` *Construct* $A = \tilde{A}P$ *and check if the columns in* A *satisfy the required sign restrictions. If they do, keep* A *and proceed to the next step. Otherwise, return to* `Step-4.3` *and keep iterating until obtaining a matrix* P *that delivers a sign-conforming impact matrix* A,

`Step-5` *Draw the covariance matrix,* Ω_v*, of coefficient-specific errors by sampling each* $\sigma^2_{v,i}$*, in* $\Omega_v = \texttt{diag}\left(\sigma^2_{v,1}, \cdots, \sigma^2_{v,M}\right)$ *from* $f(\sigma^2_{v,i}|\mathbf{y}, \mathbf{s}_{-\sigma^2_{v,i}})$ *for* $i = 1, ..., M$.

Step 1 above was discussed in Section 3.2. In what follows, we provide details for the other steps.

- *Sampling* $\boldsymbol{\lambda}$

Recall from Section 3.2 that we can express the state equation in (A2) as:

$$\widetilde{\boldsymbol{\phi}} = \widetilde{\boldsymbol{\phi}}_{\mathbf{0}} + \mathbf{L_e}\boldsymbol{\lambda} + \mathbf{v}, \tag{A4}$$

where:

$$\mathbf{L_e} = \left(\begin{bmatrix} e_1' & 0' \\ e_2' & e_2' \\ \vdots & \vdots \\ e_T & e_{T-1}' \end{bmatrix} \right).$$

Since – when sampling $\boldsymbol{\lambda}$ – $\widetilde{\boldsymbol{\phi}}$ and A are given, we can back out $\mathbf{e}$ from (A1) by simply computing $\mathbf{e} = \mathbf{L_A}^{-1}(\mathbf{y} - \widetilde{\mathbf{X}}\widetilde{\boldsymbol{\phi}})$. As a result, by virtue of the modular nature of MCMC

algorithms, $\mathbf{L_e}$ can be treated as a matrix of predetermined regressors. Given that, $\mathbf{v} \sim \mathcal{N}(0,\ \mathbf{\Sigma_v})$, then by standard regression results, we have:

$$\boldsymbol{\lambda}|\mathbf{y}, \mathbf{s}_{-\boldsymbol{\lambda}} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\lambda},\ \overline{\mathbf{D}}_{\lambda}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{\boldsymbol{\lambda}} = \overline{\mathbf{D}}_{\boldsymbol{\lambda}} \mathbf{L}_{\mathbf{e}}' \mathbf{\Sigma}_v^{-1} (\widetilde{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}_{\mathbf{0}}), \\ \overline{\mathbf{D}}_{\boldsymbol{\lambda}} = \left(\mathbf{L}_{\mathbf{e}}' \mathbf{\Sigma}_v^{-1} \mathbf{L}_{\mathbf{e}} + \mathbf{\Sigma}_{\boldsymbol{\lambda}}^{-1}\right)^{-1}. \end{cases} \tag{A5}$$

• *Sampling* $\boldsymbol{\phi_0}$

Let $\mathbf{L_0} = \iota_0 \otimes I_M$, where (again) M denotes the number of drifting coefficients and $\iota_0 = (1, 0, \cdots, 0)'$ is a $T \times 1$ vector. Thus, (A4) can be recast as:

$$\widetilde{\boldsymbol{\phi}} = \mathbf{L_0}\boldsymbol{\phi_0} + \mathbf{L_e}\boldsymbol{\lambda} + \mathbf{v}. \tag{A6}$$

Next, just like in the discussion for sampling $\boldsymbol{\lambda}$, standard regression results yield:

$$\boldsymbol{\phi_0}|\mathbf{y}, \mathbf{s}_{-\boldsymbol{\phi_0}} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\boldsymbol{\phi_0}},\ \overline{\mathbf{D}}_{\boldsymbol{\phi_0}}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{\boldsymbol{\phi_0}} = \overline{\mathbf{D}}_{\boldsymbol{\phi_0}} (\mathbf{L}_0' \mathbf{\Sigma}_v^{-1} (\widetilde{\boldsymbol{\phi}} - \mathbf{L}_{\widetilde{\mathbf{e}}} \boldsymbol{\lambda}) + \mathbf{\Sigma}_{\boldsymbol{\phi_0}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\phi_0}}), \\ \overline{\mathbf{D}}_{\boldsymbol{\phi_0}} = \left(\mathbf{L}_0' \mathbf{\Sigma}_{\eta}^{-1} \mathbf{L}_0 + \mathbf{\Sigma}_{\boldsymbol{\phi_0}}^{-1}\right)^{-1}. \end{cases} \tag{A7}$$

• *Sampling A*

We begin by reproducing the measurement equation in (A1):

$$\mathbf{y} = \widetilde{\mathbf{X}} \widetilde{\boldsymbol{\phi}} + \mathbf{L_A} \mathbf{e}, \tag{A8}$$

such that the vector of reduced-form errors is given by $\mathbf{L_A e} = \mathbf{u} \sim \mathcal{N}(\mathbf{0},\ \mathtt{diag}(\Omega_u, \cdots, \Omega_u))$.[1]

Our approach for estimating A follows the same principles as in Uhlig (2005) and Rubio-Ramirez et al. (2010), who also address Bayesian estimation of sign-restricted VARs. Specifically, these authors show that given a posterior draw for Ω_u, one can rotate its Cholesky decomposition, $\tilde{A}$ to recover an impact matrix that complies with pre-specified sign restrictions. This strategy relies on introducing an auxiliary algorithm into the posterior sampling strategy in order to generate a rotation matrix, P, which combined with $\tilde{A}$ yields sign-conforming values for the impact matrix $A = \tilde{A}P$.

Obtaining a draw for Ω_u, is straightforward. Since $\widetilde{\boldsymbol{\phi}}$ is given when sampling Ω_u, we can apply the modular nature of MCMC algorithms to treat the drifting coefficients as predetermined terms in (A8). Standard regression results thus yield:

$$\Omega_u|\mathbf{y}, \mathbf{s}_{-\boldsymbol{\Omega_u}} \sim \mathcal{IW}\left(\overline{\nu}_u,\ \overline{\mathbf{V}}_u\right), \text{ where } \begin{cases} \overline{\nu}_u = T + \nu_u, \\ \overline{\mathbf{V}}_u = \sum_{t=1}^{T} u_t u_t' + \mathbf{V}_u, \end{cases} \tag{A9}$$

[1] Recall from Section 3 that $\mathbf{L_A} = \mathtt{diag}(A, \cdots, A)$. Thus $\mathbf{u} \sim \mathcal{N}(\mathbf{0},\ \mathtt{diag}(\Omega_u, \cdots, \Omega_u))$ can be equivalently represented as $\mathbf{u} \sim \mathcal{N}(\mathbf{0},\ \mathtt{diag}(AA', \cdots, AA'))$.

where u_t for $t = 1, ..., T$ is a $N \times 1$ vector-element in $\mathbf{u} = (u_1', \cdots, u_T')'$.

Given a posterior draw for Ω_u, we compute its Cholesky factorization which returns $\tilde{A}$:

$$\tilde{A} = \begin{bmatrix} \alpha_{1,1} & 0 & 0 \\ \alpha_{2,1} & \alpha_{2,2} & 0 \\ \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} \end{bmatrix}. \tag{A10}$$

To generate P we apply Algorithm 2 in Rubio-Ramirez et al. (2010), which can be summarized as follows:[2]

We draw an $N \times N$ matrix $J \sim \mathcal{N}(0, I)$ and compute its QR decomposition. The latter yields $J = PD$, where P is the desired rotation matrix and D is a diagonal matrix whose non-zero elements are normalized to be positive. Next, we compute $A = \tilde{A}P$ and verify if the columns in A satisfy the restrictions in Table 1 of the paper. If they do not, we redraw J and compute its QR decomposition. We repeat this procedure until generating a rotation matrix P that conforms with the predetermined sign restrictions.

- *Sampling* Ω_v

Since Ω_v is diagonal, standard methods ensure that we can sample each individual variance parameter in Ω_v from an inverse-Gamma density. Formally, we have:

$$\sigma_{v,i}^2 | \mathbf{y}, \mathbf{s}_{-\boldsymbol{\sigma 2_{v,i}}} \sim \mathcal{IG}\left(\overline{\nu}_{v,i},\ \overline{S}_{v,i}\right), \text{ where } \begin{cases} \overline{\nu}_{v,i} = \frac{T}{2} + \nu_{v,i}, \\ \overline{S}_{v,i} = \frac{\sum_{t=1}^{T} v_{t,i}^2}{2} + S_{v,i} \quad \text{for } i = 1, ..., M. \end{cases} \tag{A11}$$

A1.2 Computing the Bayes Factor

Recall from Section 3.2 that computation of the Bayes factor entailed evaluating the prior – $f(\boldsymbol{\lambda})$ – and the posterior – $f(\boldsymbol{\lambda}|\mathbf{y})$ – at the restriction $\boldsymbol{\lambda} = \mathbf{0}$. The prior can be evaluated exactly, while an estimator for the posterior can be obtained using the following Monte Carlo average:

$$\hat{f}(\boldsymbol{\lambda} = \mathbf{0}|\mathbf{y}) = \frac{1}{R} \sum_{r=1}^{R} f(\boldsymbol{\lambda} = \mathbf{0}|\mathbf{y}, \boldsymbol{\phi}^{(r)}, \boldsymbol{\phi}_{\mathbf{0}}^{(r)}, \Sigma_u^{(r)}, \Sigma_v^{(r)}),$$

where $(\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}_{\mathbf{0}}^{(1)}, \Sigma_u^{(1)}, \Sigma_v^{(1)}), ..., (\boldsymbol{\phi}^{(R)}, \boldsymbol{\phi}_{\mathbf{0}}^{(R)}, \Sigma_u^{(R)}, \Sigma_v^{(R)})$ are (post-burn-in) MCMC draws. Note that the density inside the summation above can be evaluated exactly using the results in (A5).

[2] Note that since P is a rotation matrix, and consequently $PP' = I$, appending such a matrix to our Gibbs sampling procedure does not alter first and second moments for the vector of the reduced-form errors $\mathbf{u}$.

A1.3 Stability Diagnostics

For multi-move samplers – such as ours – imposing inequality restrictions to ensure all time-varying autoregressive coefficients comply with a stable system can complicate estimation considerably. This is because it has to be assumed that the MCMC algorithm can only proceed to the next sampling step if and only if all the autoregressive coefficients at all points in time simultaneously comply with stability conditions. In practice, this typically leads sampling algorithms to get 'stuck' ad infinitum before moving to the next draw.

For the applications in this paper, strict adherence to stability conditions implies simultaneously evaluating 1314 eigenvalues (i.e. $N \times p \times T$ eigenvalues) at each MCMC step for $\boldsymbol{\phi}$ and confirming that not a single eigenvalue is greater than one in absolute value. Intuitively, by virtue of law of large numbers, the probability that at least one eigenvalue would violate such condition is virtually one. Moreover, in a context where autoregressive coefficients are allowed to be influenced by structural innovations, achieving stability at all points in time can be quite challenging. For instance, during periods of high macroeconomic or financial instability, some structural shocks may exhibit extremely large magnitudes, hence behaving as outliers and potentially temporarily distorting the stability of the VAR system.[3] That said, below we show that such stability related concerns are not prevalent for our empirical applications.

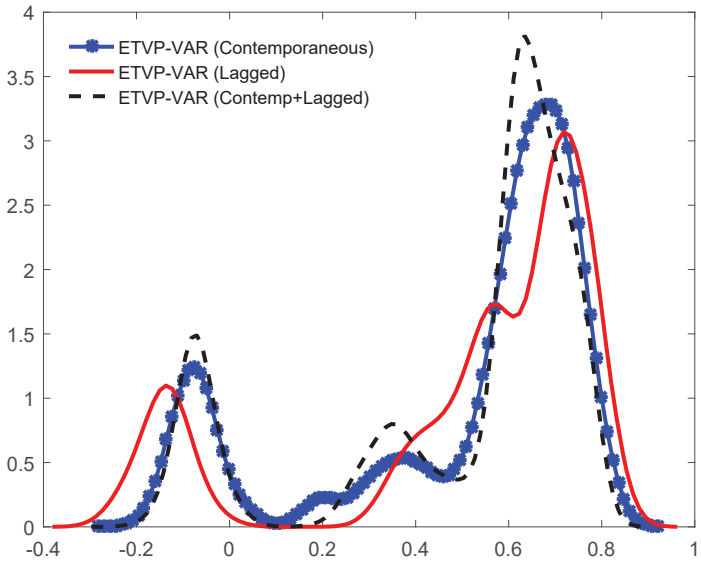
Certainly, one alternative to address stability issues could be to adopt a single-move MCMC sampler. In this case, instead of checking for the stability of $N \times p \times T$ eigenvalues at once, the problem would be reduced to checking $N \times p$ eigenvalues T times (i.e. evaluate 18 eigenvalues 73 times). However, single-move MCMC samplers are not problem-free either. In particular, such a type of algorithms typically fare quite poorly in terms of mixing properties.[4]

Due to the reasons just described, results reported in this paper did not involve reject-sampling procedures to ensure stability of all time-varying coefficients. Nonetheless, we still view it as important to check whether explosive draws are indeed a considerable issue in the context of our empirical applications. To do so, Figure A1 plots the distribution of all 1314 eigenvalues for the three types of endogenous TVP-VARs discussed in the paper. To be precise, each of the 1314 eigenvalues in Figure A1 corresponds to the mean obtained from 30000 post burn-in draws. Overall, results in Figure A1 reinforce that, while the MCMC algorithm might occasionally produce non-stable draws, such draws are not frequent in our estimation exercise.

[3] In other words, to the extent that large shocks induce big variations in policy and agents' decisions – as is typically the case during recessions – it is not implausible to conceive that economic dynamics can exhibit 'explosive-like' behavior at times. It could then be argued that letting structural shocks temporarily push coefficients to the unit-root region of the parameter space – if such a situation is indeed favored by the data – can provide a useful and more realistic indication of economic dynamics during particular episodes.

[4] The reader is referred to Koop and Potter (2011) for a detailed discussion on single versus multi-move samplers in the context of (exogenous) TVP-VARs.

Figure A1: Distribution of eigenvalues associated with the autoregressive coefficients for endogenous TVP-VARs



Note: The figure plots the kernel densities associated with the posterior means of all the eigenvalues for the three versions of endogenous TVP-VAR models, that is, with coefficients depending on contemporaneous, lagged, and contemporaneous and lagged structural innovations.

A2 Computational Performance

A2.1 Inefficiency Factors

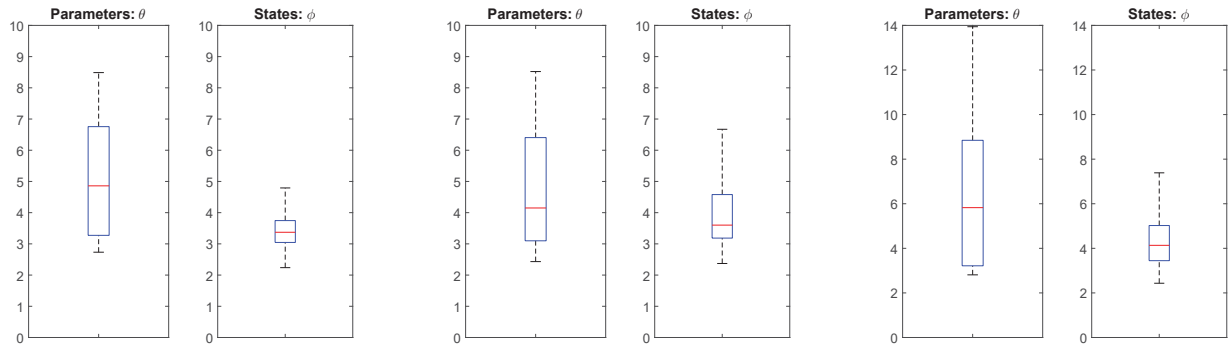
We report inefficiency factors of the posterior draws for all states and parameters using a common metric (see, e.g., Chib (2001)) given by:

$$1 + 2\sum_{j=1}^{J} \rho_j,$$

where ρ_j is the sample autocorrelation at lag j through lag J. In our empirical application we set J to be large enough until autocorrelation tapers off. In an ideal setting where MCMC draws are virtually independent draws, inefficiency factors should be one. As a rule of thumb, inefficiency factors around twenty are typically interpreted as an indication of fast mixing.[5] Figure A2 reports boxplots to summarize inefficiency factor results. The middle line denotes the median inefficiency factor. Lower and upper lines respectively represent the 25 and 75 percentiles, while whiskers extend to the maximum and minimum inefficiency factors. All in all, results in Figure A2 demonstrate that our posterior sampler exhibits good mixing properties.

Figure A2: Inefficiency factors

(a) Contemporaneous (b) Lagged (c) Contemp. & Lagged



Note: The figure shows the boxplots associated with the inefficiency factors corresponding to the posterior draws of parameters $\boldsymbol{\theta}$ and states, time-varying coefficients, $\boldsymbol{\phi}$ of the model. Charts A, B and C show the results associated to TVP-VAR with coefficients driven by contemporaneous, lagged, and contemporaneous and lagged structural innovations, respectively.

A2.2 Scalability

As is well known, MCMC-based estimation of TVP-VARs becomes considerably challenging as model dimension increases. Therefore, in this section we examine how scalable our pro-

[5] Another way to interpret the inefficiency factor adopted here is to think that an inefficiency factor of 100 means that approximately 10000 posterior draws are required to convey the same information as 100 independent draws.

posed estimation algorithm is. To this end, we estimate the three endogenous TVP-VAR variants discussed in the main text – i.e. Models II, III and IV in Table 2 of the paper – using simulated data and setting $N = 3$, $N = 5$ and $N = 7$ while fixing sample size $T = 160$ quarters, i.e. 40 years.[6]

Computation started to become prohibitive whilst running Model IV – i.e. the more complex one – with $N = 7$. In particular, it took approximately 78 hours to run 15000 MCMC iterations. For all other cases, estimation was relatively tractable with estimation ranging from 30 minutes to 14 hours. Notably, the operation in our algorithm that entails considerable computational complexity is calculating the high-dimensional $TM \times TM$ posterior covariance matrix $\overline{\mathbf{D}}_{\tilde{\phi}}$ in Equation (19) of the manuscript.[7]

More precisely, as pointed out in Van Loan and Golub (1983) (see page 156), operations such as matrix inversion and Cholesky factorization – which are required to construct $\overline{\mathbf{D}}_{\tilde{\phi}}$ (see Chan and Jeliazkov (2009)) – have computational complexity equivalent to $\mathcal{O}((TM)^3)$ in the case of dense matrices, such as $\overline{\mathbf{D}}_{\tilde{\phi}}$ for endogenous TVP-VARs. In other words, as N increases – and, consequently, M – the number of algebraic operations that are necessary to generate $\overline{\mathbf{D}}_{\tilde{\phi}}$ increases at an exponential rate (with power 3). In addition, $\overline{\mathbf{D}}_{\tilde{\phi}}$ is yet more dense for Model IV since $\mathbf{L}_{\lambda}$ – which enters the construction of $\overline{\mathbf{D}}_{\tilde{\phi}}$ – has more non-zero elements than in the other two types of endogenous TVP-VARs we consider (see page 13 of the revised manuscript). This reduced sparsity explains the additional computing cost to generate $\overline{\mathbf{D}}_{\tilde{\phi}}$ for Model IV.easonable upper bound when running our algorithm without resorting to alternatives such as parallel computing, which we discuss next.

As an alternative to circumvent the aforementioned obstacles related to scalability, we developed a routine that allows potential users of our approach to parallelize estimation using Matlab's `parfor` function. Importantly, since MCMC is a procedure that entails path dependence across draws, our parallel-computing strategy preserves such a feature. Specifically, when executing the `parfor` loop, each "worker" in the parallel pool cycles through *all* the full-conditional posteriors in the sampling algorithm.[8] Hence, the benefit of this approach stems from having multiple "workers" running simultaneously, where each one is running on a reduced number of Gibbs iterations.

It is worth noting that parallelizing estimation is not cost-free either. In particular, each "worker" in the parallel pool is running a complete estimation procedure – i.e. cycling through all Gibss steps – relying on less CPU cores. As a result, the benefits of parallelizing become more salient when model dimension increases considerably. In such cases, producing a large number of MCMC draws based on a single "worker" (i.e. not parallelizing), even though exploring all CPU cores, becomes prohibitive time-wise due to the high level of

[6] Notably, $N = 7$ is the same number of variables used in prototypical medium-scale DSGE models, such as Smets and Wouters (2007) and Slobodyan and Wouters (2012).

[7] Recall that M denotes the number of drifting coefficients in the model, which increases exponentially as a function of N, i.e. $M = N^2p + N$, where p denotes lag length.

[8] To be clear, we are not parallelizing particular blocks of the MCMC algorithm, which could be problematic.

computational complexity. In contrast, generating a large number of draws by aggregating them from each "worker" in the pool running a reduced number of MCMC iterations – albeit each relying on less CPU cores – expedites estimation. Table A1 below illustrates the above mentioned trade-offs. In particular, while computing time is more or less comparable for the parallel and non-parallel procedures when $N = 3$ and $N = 5$, for $N = 7$ parallelization reduces the computation time by more than 60%.

Lastly, parallelization – along the lines proposed here – becomes more efficient the better the mixing properties of an MCMC algorithm are. In other words, good convergence properties suggest one can increase the number of "workers" in the parallel pool whilst reducing further the number of MCMC iterations assigned to each one (hence, speeding up computation). For the exercise in this section, we generated 15000 draws by assigning 3000 MCMC iterations to each of the five "workers" in the parallel pool. This seems reasonable based on the good mixing properties of our algorithm reported in Section A2.1 of the Online Appendix (see Figure A2).

Table A1: Computing time (in hours) for running 15000 MCMC iterations based on simulated data for a sample size of T=160

Identifier	Non-parallelized estimation		
	N=3	N=5	N=7
Model II	0.53	3.16	14.61
Model III	0.54	3.21	14.86
Model IV	1.17	10.78	78.10
	Parallelized estimation		
	N=3	N=5	N=7
Model II	0.61	3.37	5.41
Model III	0.62	3.76	5.50
Model IV	1.31	26.36	36.08

Note: N denotes the number of dependent variables in the TVP-VAR. Models II, III and IV denote the endogenous TVP-VAR versions where structural shocks affect the drifting coefficients contemporaneously, with a lag, and both contemporaneously and with a lag, respectively. See Table 2 in the paper for details. Parallelized estimation was conducted using Matlab's `parfor` function with five "workers" in the parallel pool, each running 3000 MCMC iterations. Parallel and non-parallel routines were implemented on a desktop with an Intel Xeon E5-2690 v2 @3.00 GHz processor.

A3 Incorporating Stochastic Volatility

In this section, we provide a detailed description of the MCMC steps required to estimate our proposed framework when augmented to accommodate stochastic volatility. As discussed in Section 4.5 of the paper, allowing for stochastic volatility is part of a broader list of robustness checks.

We begin by showing the augmented state-space representation of our baseline (endogenous) TVP-VAR framework due to the introduction of new latent variables into the model, $h_{t,i}$, for $i = 1, 2,\ 3$ that capture changes in the volatility of the reduced-form residuals. In particular, assuming a random-walk law of motion for $h_t = [h_{t,1}\ h_{t,2}\ h_{t,3}]'$ – as in Primiceri (2005), Stock and Watson (2007), Stock and Watson (2016), among others – we have:

$$y_t = X_t\phi_t + \tilde{A}\Lambda_t P e_t, \tag{A12}$$

$$\phi_t = \phi_{t-1} + H_{C,\lambda}e_t + H_{L,\lambda}e_{t-1} + v_t, \tag{A13}$$

$$h_t = h_{t-1} + \vartheta_t, \tag{A14}$$

$$\begin{bmatrix} e_t \\ v_t \\ \vartheta_t \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} I_N & 0 & 0 \\ 0 & \Omega_v & 0 \\ 0 & 0 & \Omega_\vartheta \end{bmatrix}\right). \tag{A15}$$

Note from Equation (A12) that to combine stochastic volatility with sign restrictions, we follow Baumeister and Peersman (2013) and Mumtaz and Zanetti (2015) and define the impact matrix as:

$$A_t = \tilde{A}\Lambda_t P, \tag{A16}$$

where

$$\tilde{A} = \begin{bmatrix} 1 & 0 & 0 \\ \alpha_{2,1} & 1 & 0 \\ \alpha_{3,1} & \alpha_{3,2} & 1 \end{bmatrix}, \quad \Lambda_t = \begin{bmatrix} \exp\left(\frac{h_{1,t}}{2}\right) & 0 & 0 \\ 0 & \exp\left(\frac{h_{2,t}}{2}\right) & 0 \\ 0 & 0 & \exp\left(\frac{h_{3,t}}{2}\right) \end{bmatrix} \tag{A17}$$

and P is a rotation matrix that ensures sign restrictions are satisfied.[9] In other words, computation of A_t entails sampling the log-volatilities (h_t), the lower triangular elements in $\tilde{A}$ and generating P. Estimation of h_t is conducted using the auxiliary mixture sampling approach of Omori et al. (2007) combined with precision sampling methods, as in Chan and Jeliazkov (2009). The lower triangular elements in $\tilde{A}$ are estimated using the equation-by-equation approach as in, e.g., Cogley and Sargent (2005) and Carriero et al. (2018). P is obtained using, again, Algorithm 2 in Rubio-Ramirez et al. (2010).

[9] Baumeister and Peersman (2013) and Mumtaz and Zanetti (2015) also allow for $\tilde{A}$ to be time-varying. For parsimony, we preserve the assumption that $\tilde{A}$ is time-invariant as in Cogley and Sargent (2005).

MCMC Steps

Steps 1, 2, 3 and 5 summarized in Box 1 in Section A1.1 remain unchanged when allowing for stochastic volatility. Step 4, however, needs to be modified since the conditional variances of the reduced-form errors are now state variables. In what follows, we describe in detail the derivation of the full conditional posteriors for $\mathbf{h} = (h_1, ..., h_T)$ and for (the parameters in) $\tilde{A}$. We also show two additional steps: (i) sampling Ω_ϑ, i.e. the covariance matrix for the innovations driving h_t; (ii) and sampling h_0, i.e. the initialization condition for the (log-) volatility states.

- *Sampling* $\mathbf{h} = (h_1, ..., h_T)'$

To sample the vector of log-volatilities $\mathbf{h}$, we combine the auxiliary mixture sampler approach of Omori et al. (2007) with the precision sampling techniques described in Chan and Jeliazkov (2009). To this end, note first that we can reexpress equations (A12) and (A14) as:

$$\tilde{A}^{-1}(y_t - X_t\phi_t) = \Lambda_t\tilde{e}_t, \tag{A18}$$

where $\tilde{e}_t = Pe_t$. Since P is a matrix of the orthogonal group (and, therefore, $PP' = I$), both e_t and $\tilde{e}_t$ follow the exact same distribution, i.e. $\tilde{e}_t \sim \mathcal{N}(0,\ I_N)$.

Next, squaring and subsequently taking natural logarithms of each element in both sides of (A18) leads to the following linear state space representation for h_t:

$$\tilde{y}_t^* = h_t + \tilde{e}_{t,\chi}, \tag{A19}$$

$$h_t = h_{t-1} + \vartheta_t. \tag{A20}$$

The system above – albeit linear – is no longer Gaussian. More precisely, each entry in $\tilde{e}_{t,\chi}$ follows a log chi-square distribution with one degree of freedom. We return to this point below, but first recall from our discussion in Section 3 of the paper that precision sampling techniques operate on a representation of systems of equations where each variable is stacked over $t = 1, ..., T$. Therefore, Equations (A19) and (A20) are re-casted as:

$$\widetilde{\mathbf{y}}^* = \mathbf{h} + \tilde{\mathbf{e}}_{\mathbf{t},\chi}, \tag{A21}$$

$$\mathbf{L}_h\mathbf{h} = \mathbf{h}_0 + \boldsymbol{\vartheta}, \tag{A22}$$

where:

$$\widetilde{\mathbf{y}}^* = (\tilde{y}_1^*, \cdots, \tilde{y}_T^*)',\ \mathbf{L}_h\mathbf{h} = \underbrace{\begin{bmatrix} I & 0 & \cdots & 0 \\ -I & I & & \\ 0 & -I & \ddots & \vdots \\ \vdots & & \ddots & \\ 0 & & \cdots -I & I \end{bmatrix}}_{\mathbf{L}_h} \underbrace{\begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ \\ h_T \end{bmatrix}}_{\mathbf{h}} \text{ and } \mathbf{h}_0 = (h_0, 0, \cdots, 0)'.$$

The vectors collecting the innovations $\tilde{\mathbf{e}}_{\mathbf{t},\chi}$ and $\boldsymbol{\vartheta}$ are similarly defined.

To bring the state space representation in (A21) and (A22) back to Gaussian form, we follow Omori et al. (2007) who proposed approximating the log chi-square distribution for $\tilde{\mathbf{e}}_{\mathbf{t},\chi}$ as a mixture of ten Normal densities.[10] Formally, let $\widetilde{\mathbf{e}}^*$ denote such mixture approximation, i.e.:

$$\tilde{\mathbf{e}}_{\mathbf{t},\chi} \approx \widetilde{\mathbf{e}}^* \sim p_1 \mathcal{N}\left(\boldsymbol{\alpha}_1,\ \boldsymbol{\Sigma}_1\right) + \cdots + p_{10} \mathcal{N}\left(\boldsymbol{\alpha}_{10},\ \boldsymbol{\Sigma}_{10}\right),$$

where $\boldsymbol{\alpha}_s$, $\boldsymbol{\Sigma}_s$ and the component-density probabilities p_s for $s = 1, \cdots, 10$ are all predetermined, and their values are given in Table 1 of Omori et al. (2007). Consequently, conditional on a given component-density $\mathcal{N}\left(\boldsymbol{\alpha}_s,\ \boldsymbol{\Sigma}_s\right)$, (A21)-(A22) can be expressed in linear Gaussian form as:

$$\widetilde{\mathbf{y}}^* = \mathbf{h} + \boldsymbol{\alpha}_s + \widetilde{\mathbf{e}}_s^*, \tag{A23}$$

$$\mathbf{L}_h \mathbf{h} = \mathbf{h}_0 + \boldsymbol{\vartheta}, \tag{A24}$$

$$\begin{bmatrix} \widetilde{\mathbf{e}}_s^* \\ \boldsymbol{\vartheta} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\vartheta \end{bmatrix} \right). \tag{A25}$$

The parametrization above requires sampling the vector of component-density indicators, $\mathbf{s} = (s_1, \cdots, s_T)$, from its full conditional posterior $f(\mathbf{s}|\widetilde{\mathbf{y}}^*, \mathbf{z})$ and subsequently sample $\mathbf{h}$ from $f(\mathbf{h}|\widetilde{\mathbf{y}}^*, \mathbf{z}_{-\mathbf{h}}, \mathbf{s})$.[11] Each element in $\mathbf{s} = (s_1, \cdots, s_T)$ can be drawn independently from a multinomial distribution parameterized by the full conditional posterior probability:

$$\Pr\left(s_t = i|\widetilde{y}_t^*, z_t\right) = \frac{\psi\left(h_t + \alpha_i,\ \sigma_i^2\right) p_i}{\sum_{j=1}^{10} \psi\left(h_t + \alpha_j,\ \sigma_j^2\right) p_j} \quad \text{for } i = 1, \cdots, 10,$$

where $\psi\left(h_t + \alpha_s,\ \sigma_s^2\right)$ denotes a Gaussian density evaluated at mean $h_t + \alpha_s$ and variance σ_s^2. Again, α_s and σ_s^2 values are given in Table 1 in Omori et al. (2007) and h_t denotes posterior draws obtained from (A26) as presented below.

Given $\Pr\left(s_t = i|\widetilde{y}_{\tau,t}^*, z_t\right)$ posterior draws for s_t can then be generated via the inverse transform method for each $t = 1, \cdots, T$ as follows:[12]

(a) Generate $\omega_t \sim Uniform(0,\ 1)$,
(b) Find the smallest $i \in \{1, 2, \cdots, 10\}$ that satisfies $\sum_{j=1}^{i} \Pr\left(s_t = j|\widetilde{y}_t^*, z_t\right) \geq \omega_t$,
(c) Return $s_t|\widetilde{y}_t^*, z_t = i$.

[10] Their approach extends the seven-component auxiliary mixture sampling from Kim et al. (1998).
[11] Such ordering is consistent with the discussion in Del Negro and Primiceri (2015).
[12] See algorithm 3.2 in Kroese et al. (2013) for a more detailed discussion of the inverse transform method for discrete random variables.

Given $\mathbf{s}$, standard regression results (see Koop et al. (2007)) can applied to the system in (A23)-(A25) to derive the conditional posterior distribution for $\mathbf{h}$:

$$\mathbf{h}|\widetilde{\mathbf{y}}^*, \mathbf{z}, \mathbf{s} \sim \mathcal{N}\left(\overline{\mathbf{d}}_h,\ \overline{\mathbf{D}}_h\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_h = \overline{\mathbf{D}}_h \left(\boldsymbol{\Sigma}_s^{-1}\widetilde{\mathbf{y}}_\alpha^* + \mathbf{L}_h'\boldsymbol{\Sigma}_\vartheta^{-1}\mathbf{h}_0\right), \\ \overline{\mathbf{D}}_h = \left(\boldsymbol{\Sigma}_s^{-1} + \mathbf{L}_h'\boldsymbol{\Sigma}_\vartheta^{-1}\mathbf{L}_h\right)^{-1}, \end{cases} \tag{A26}$$

where $\widetilde{\mathbf{y}}_\alpha^* = \widetilde{\mathbf{y}}^* - \mathbf{h} - \boldsymbol{\alpha}_s$. Draws from the density above are obtained using the precision sampler of Chan and Jeliazkov (2009).

• *Sampling* $\tilde{A}$

To sample $\tilde{A}$ we follow the same strategy as in Cogley and Sargent (2005), who explore the lower triangular structure of $\tilde{A}$ to represent the relationship between reduced-form errors (u_t) and orthogonalized shocks (e_t), based on $u_t = \tilde{A}\Lambda_t P e_t$, as a system of unrelated regressions, i.e.:

$$\tilde{A}\Lambda_t P e_t = \begin{bmatrix} u_{t,1} \\ u_{t_2} \\ u_{t,3} \end{bmatrix} = \begin{bmatrix} \tilde{u}_{t,1} \\ \alpha_{2,1}\tilde{u}_{t,1} + \tilde{u}_{t,2} \\ \alpha_{3,1}\tilde{u}_{t,1} + \alpha_{3,2}\tilde{u}_{t,2} + \tilde{u}_{t,3} \end{bmatrix}, s.t.\ \tilde{u}_{t,i} \sim \mathcal{N}\left(0,\ \exp(h_{t,i})\right) \text{ for } i = 1,\ 2,\ 3.$$

Conditional on the data, parameters and states, $\tilde{u}_{t,1}$ and u_{t_2} are known and can thus be used as a regressor and regressand, respectively, in the regression $u_{t_2} = \alpha_{2,1}\tilde{u}_{t,1} + \tilde{u}_{t,2}$. Once an estimate for $\alpha_{2,1}$ is obtained, the same rationale applies to $u_{t,3} = \alpha_{3,1}\tilde{u}_{t,1} + \alpha_{3,2}\tilde{u}_{t,2} + \tilde{u}_{t,3}$.

Therefore, starting from the second equation, we can express each equation in the VAR system as:

$$y_{t,i} = x_{t,i}\phi_j + \sum_{j=2}^{i} \alpha_{i,j-1}\tilde{u}_{t,j-1} + \tilde{u}_{t,i} \text{ for } i = 2,\ 3. \tag{A27}$$

Stacking both sides of the expression above over $t = 1, ..., T$ yields:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\phi}_i + \mathbf{X}_{\tilde{u}_i}\alpha_i + \tilde{\mathbf{u}}_i,\ s.t.\ \tilde{\mathbf{u}}_i \sim \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_{u_i}\right) \text{ for } i = 2,\ 3, \tag{A28}$$

where $\boldsymbol{\Sigma}_{u_i} = \texttt{diag}\left(\exp(h_{1,i}), \cdots, \exp(h_{T,i})\right)$. Combining (A28) with a Gaussian prior $\alpha_i \sim \mathcal{N}\left(\widehat{\alpha}_i,\ \boldsymbol{\Sigma}_{\alpha_i}\right)$ yields:[13]

$$\alpha_i|\mathbf{y}, \mathbf{z}_{-\alpha_i} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\alpha_i},\ \overline{\mathbf{D}}_{\alpha_i}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{\alpha_i} = \overline{\mathbf{D}}_{z_0}\left(\mathbf{X}_{u_i}'\boldsymbol{\Sigma}_{u_i}^{-1}\mathbf{X}_{u_i}(\mathbf{y}_i - \mathbf{X}_i\widetilde{\boldsymbol{\phi}}) + \boldsymbol{\Sigma}_{\alpha_i}\widehat{\alpha}_i\right), \\ \overline{\mathbf{D}}_{\alpha_i} = \left(\mathbf{X}_{u_i}'\boldsymbol{\Sigma}_{u_i}^{-1}\mathbf{X}_{u_i} + \boldsymbol{\Sigma}_{\alpha_i}^{-1}\right)^{-1} \quad \text{for } i = 2,\ 3. \end{cases} \tag{A29}$$

[13] We calibrate prior hyperparameters as $\widehat{\alpha}_i = 0_{i-1\times 1}$ and $\boldsymbol{\Sigma}_{\alpha_i} = 10I_{i-1}$ for $i = 2,\ 3$.

• *Sampling* Ω_h

We assume $\Omega_h = \texttt{diag}\left(\sigma_{h_1}^2, \cdots, \sigma_{h_N}^2\right)$ with each element following an Inverse Gamma prior $\sigma_{h_i}^2 \sim \mathcal{IG}\left(\nu_{h_i},\ S_{h_i}\right)$.[14] Hence, each $\sigma_{h_i}^2$ is sampled from the following density:

$$\sigma_{h_i}^2|\mathbf{y}, \mathbf{z}_{-\sigma_{h_i}^2} \sim \mathcal{IG}\left(\overline{\nu}_{h_i},\ \overline{S}_{h_i}\right), \text{ where } \begin{cases} \overline{\nu}_{h_i} = \frac{T}{2} + \nu_{h_i}, \\ \overline{S}_{h_i} = \frac{\sum_{t=1}^{T} (\vartheta_{t,i})^2}{2} + S_{h_i} \quad \text{for } i = 1, \cdots N, \end{cases} \tag{A30}$$

where $\vartheta_{t,i}$ denotes each element in the vector of innovations ϑ_t driving the log-volatilities in (A14).

• *Sampling* h_0

Let $\mathbf{L}_{h_0} = I_N \otimes \iota_0$ where $\iota_0 = (1, 0, \cdots, 0)'$. Then by a simple change of variable, $\mathbf{h}_0 = \mathbf{L}_{h_0} h_0$, Equation (A24) can be expressed as:

$$\mathbf{L}_h \mathbf{h} = \mathbf{L}_{h_0} h_0 + \boldsymbol{\vartheta}. \tag{A31}$$

Combining (A31) with the Gaussian prior $h_0 \sim \mathcal{N}\left(\widehat{h_0}, \mathbf{\Sigma}_{h_0}\right)$ yields:[15]

$$\mathrm{h}_0|\mathbf{y}, \mathbf{z}_{-\mathrm{h}_0} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\mathrm{h}_0},\ \overline{\mathbf{D}}_{\mathrm{h}_0}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{\mathrm{h}_0} = \overline{\mathbf{D}}_{\mathrm{h}_0} \left(\mathbf{L}_{h_0}' \mathbf{\Sigma}_{\vartheta}^{-1} \mathbf{L}_{h_0} \mathbf{h} + \mathbf{\Sigma}_{h_0} \widehat{\mathrm{h}_0}\right), \\ \overline{\mathbf{D}}_{\mathrm{h}_0} = \left(\mathbf{L}_{h_0}' \mathbf{\Sigma}_{\vartheta}^{-1} \mathbf{L}_{h_0} + \mathbf{\Sigma}_{h_0}^{-1}\right)^{-1}. \end{cases} \tag{A32}$$

[14] We calibrate prior hyperparameters as $\nu_{h_i} = 10^5$ and $S_{h_i} = 2.9 \times 10^3$ for $i = 1,\ 2,\ 3$.

[15] We calibrate prior hyperparameters as $\widehat{h_0} = (-3.2\ \ -3.2\ \ -3.2)'$ and $\mathbf{\Sigma}_{h_0} = 10^{-3} I_3$.

A4 Robustness Checks

In this section, we report results for three types of robustness checks: (i) allowing for heteroskedasticity; (ii) prior sensitivity analysis; and (iii) exploring a different identification strategy. In short, key results documented in the main text carry over to all these checks. Also, unless stated otherwise, all results in this section are obtained estimating Model IV (see Table 1 in the paper), i.e. the most flexible endogenous TVP-VAR variant.

A4.1 Allowing for Heteroscedasticity

We take two approaches to incorporate (conditional) heteroskedasticity, namely fixing the break dates and allowing for stochastic volatility. In the subsequent analysis and focus on three results discussed in the main text: (i) statistical evidence for endogenous TVP-VARs; (ii) gauging the contribution of each structural shock as a driver of parameter instability; and (iii) changes in the inflation-gap persistence. Regarding the latter, as discussed in Section 4.3, we follow Cogley et al. (2010) and compute variations in inflation-gap persistence using the time-varying $R^2_{h,t}$ coefficient of determination proposed in their paper. Formally, we have:

$$R^2_{h,t} \approx 1 - \frac{s'_\pi \left[\sum_{i=0}^{h-1} B_t^i F_t B_t^{i'}\right] s_\pi}{s'_\pi \left[\sum_{i=0}^{\infty} B_t^i F_t B_t^{i'}\right] s_\pi}, \tag{A33}$$

where B_t collects the autoregressive coefficients at time t from a TVP-VAR casted in companion form and F_t denotes the conditional variance-covariance matrix of the system (also in companion form). s_π is a vector that collects the coefficients in the VAR equation for the inflation rate.

Next, recall that $R^2_{h,t}$ is a metric bounded between zero and one and that what characterizes changes in inflation-gap persistence is the pace at which $R^2_{h,t}$ converges to zero as the h-step-ahead horizon increases. Therefore, a highly (weakly) persistent process at time t is one whose $R^2_{h,t}$ statistic slowly (quickly) converges to zero as h lengthens.

• Fixed Breaks

Break dates are set to 1979Q1, 1987Q2 and 2007Q2, which leads to four 'volatility regimes' that broadly coincide with particular Fed chairmanship periods. That is, 1967Q1-1979Q1 and 1979Q2-1987Q1 can be regarded as the pre-Volcker (or Burns) and Volcker regimes respectively; 1987Q2-2007Q2 captures the Greenspan chairmanship; and finally 2007Q2-2018Q2 denotes the Bernanke-Yellen period. Such an approach is somewhat in line with Sims and Zha (2006), who – applying a VAR with regime-switching mean and variance – provided evidence that 'volatility regimes' in the US between 1959 and 2003 can be more or less characterized in terms of Fed chairmanships.[16] Notably, the last break date

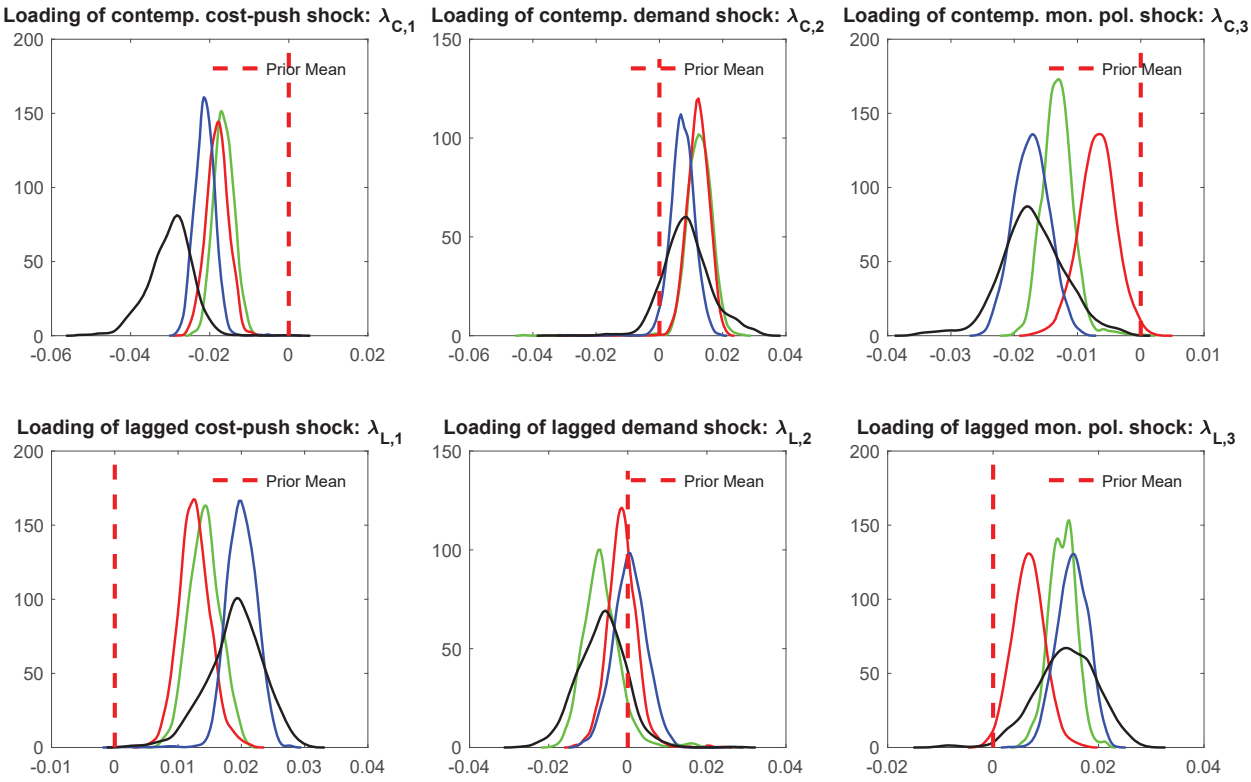
[16] For details, see the discussion in Section 5 of their paper.

is introduced to accommodate the possibility of a volatility change in the beginning of the Great Recession, in addition to capturing the early stages of Ben Bernanke's chairmanship.

Specifically, we relax the assumption of homoskedasticity while allowing for: (i) breaks in the covariance matrix for the reduced-form errors; and (ii) breaks in the conditional variance of the drifting coefficients. For the latter, we also introduce breaks in the loading coefficients associated with each (unit-variance) structural shock and in the variance of the coefficient-specific errors. It is worth noting that estimation of our models when allowing for fixed (second-moment) breaks follows the same steps summarized (in the box) presented in Section A1.1. The only adjustment required is that steps 2, 4 and 5 need to be conducted for each of the four subsamples that result from the three volatility breaks discussed above.

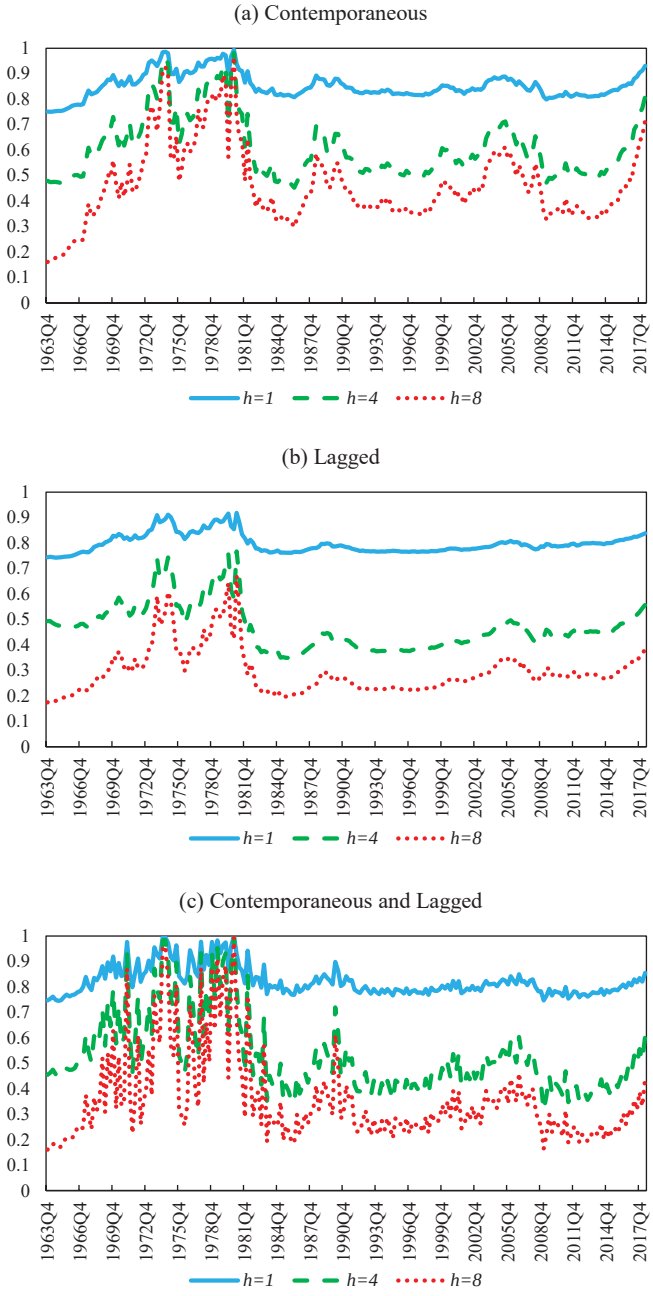
Results for our first robustness check are reported below in Figures A3 and A4 as well as in Table A2.

Figure A3: Posterior density estimates for the loadings associated with each structural shock driving the VAR coefficients under Model IV as described in Table 2 in the paper (fixed breaks)



Note: Charts in the top and bottom rows plot the posterior densities of the loadings associated with the contemporaneous and lagged structural shocks, respectively, that affect parameters of Model IV, described in Table 1 in the paper. Shock 1, Shock 2 and Shock 3 correspond to cost-push, demand and monetary policy shocks, respectively. In addition to volatility breaks in the reduced-form error and coefficient-specific error covariance matrices, the underlying model also allows for three fixed breaks (i.e. four regimes) in the loadings. Green, red, blue and black lines correspond to regimes 1, 2, 3 and 4, respectively. The `2log(BF)` between endogenous and exogenous TVP-VARs is 151.63, suggesting that evidence for endogenous time variation in the VAR coefficients is even stronger when allowing for second-moment breaks. See Kass and Raftery (1995) for details on using the Bayes factor as a metric for model comparison.

Figure A4: Inflation-gap persistence based on $R^2_{h,t}$ statistics (fixed breaks)



Note: The figure plots the time-varying measure of inflation-gap persistence, $R^2_{h,t}$, for the one-, four- and eight-quarter-ahead forecasting horizons. $R^2_{h,t}$ is constructed using Equation (A33). Estimates above denote posterior medians. Charts A, B and C show the results associated with endogenous TVP-VARs with coefficients driven by contemporaneous, lagged, and contemporaneous and lagged structural innovations, respectively.

Table A2: Contributions of structural shocks to parameter instability (fixed breaks)

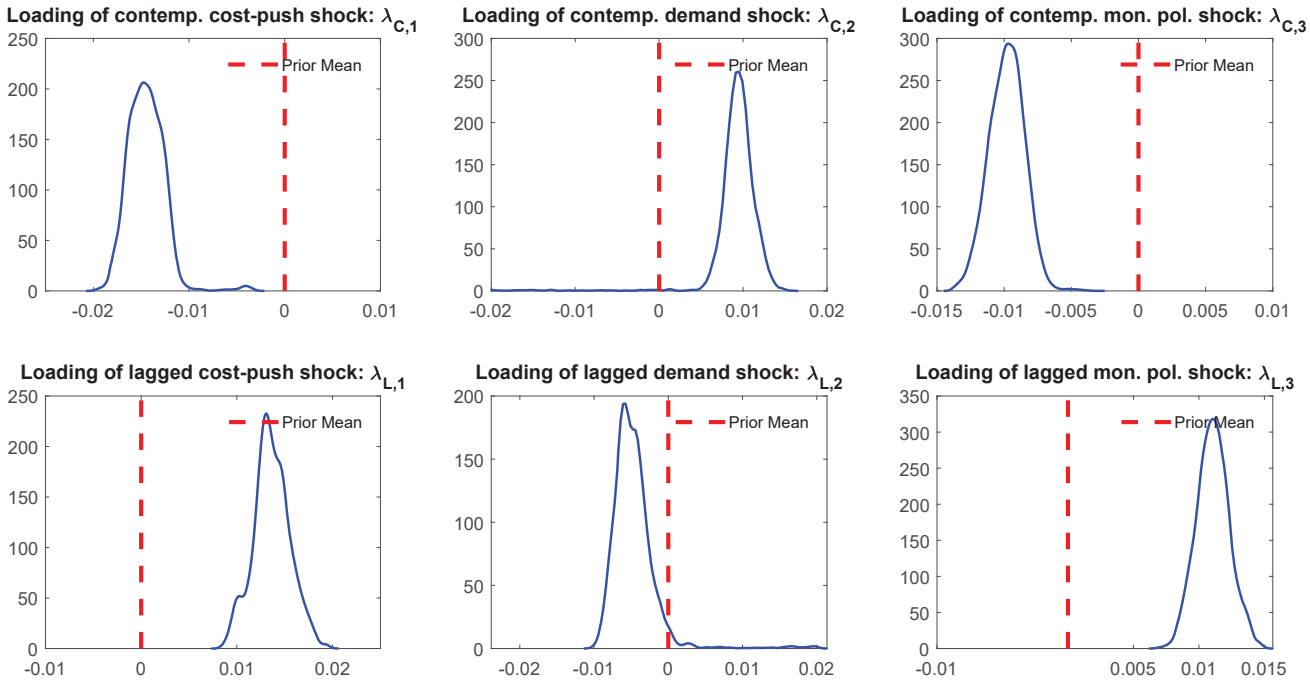
Total contribution (%)			
Source of parameter instability	*Median*	*16th perc.*	*84th perc.*
Regime 1: 1967:Q1-1979:Q1			
Cost-push shock	44.12%	35.38%	52.21%
Demand shock	16.22%	9.12%	25.31%
Monetary policy shock	30.41%	22.12%	39.05%
Residual	9.25%	8.17%	11.15%
Regime 2: 1979:Q2-1987:Q1			
Cost-push shock	58.10%	47.40%	67.31%
Demand shock	19.26%	11.31%	28.10%
Monetary policy shock	11.43%	6.27%	19.18%
Residual	11.21%	10.31%	14.29%
Regime 3: 1987:Q2-2007:Q1			
Cost-push shock	55.13%	46.22%	63.31%
Demand shock	4.27%	2.19%	9.01%
Monetary policy shock	33.41%	26.36%	42.04%
Residual	7.19%	5.21%	7.12%
Regime 4: 2007:Q2-2018:Q2			
Cost-push shock	61.26%	49.14%	70.02%
Demand shock	5.04%	1.43%	14.22%
Monetary policy shock	27.32%	16.15%	40.12%
Residual	6.38%	4.26%	6.10%

Note: The table reports the contributions of the structural shocks to the overall changes in the drifting VAR coefficients. Results are based on Equations (24) and (25) of the paper. In addition to volatility breaks in the reduced-form error and coefficient-specific error covariance matrices, the underlying model also allows for three fixed breaks (i.e. four regimes) in the loadings.

- **Stochastic Volatility**

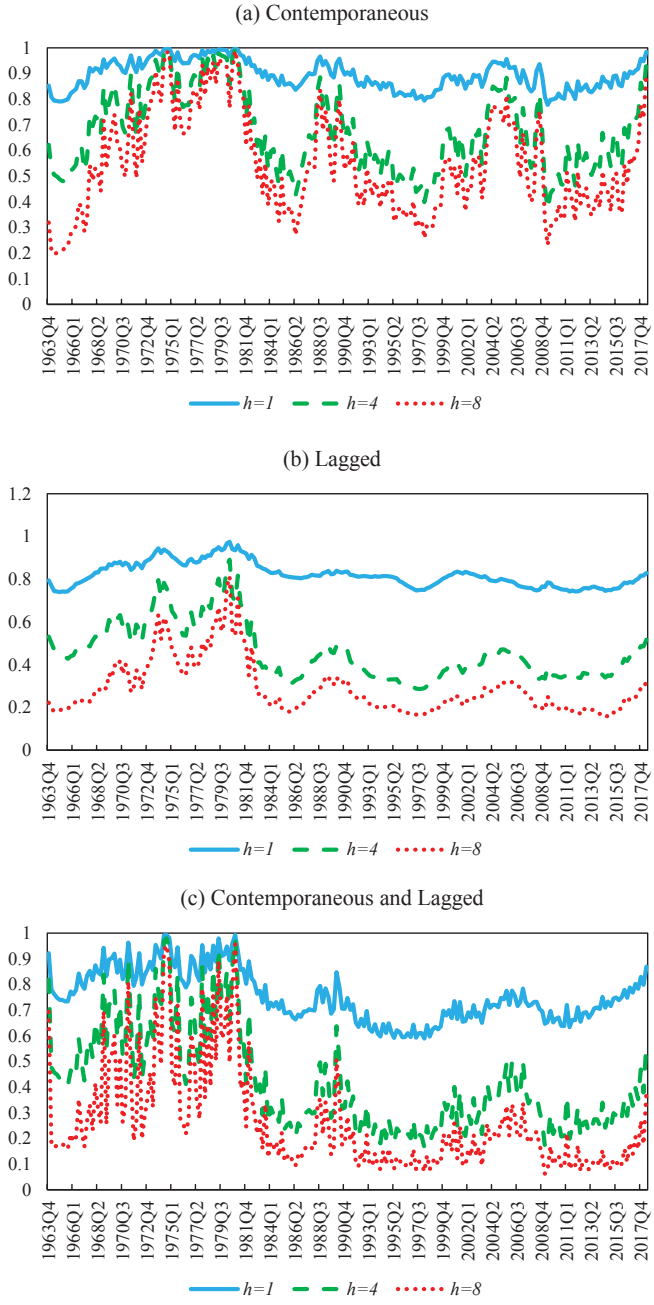
We now report robustness-check results when modeling the conditional variance of the reduced-form errors as a stochastic volatility process. Such results are shown in Figures A5 and A6 as well as in Table A3. Stochastic volatility variants of our proposed framework are estimated applying the modifications to our baseline algorithm, as discussed in Section A3.

Figure A5: Posterior density estimates for the loadings associated with each structural shock driving the VAR coefficients under Model IV described in Table 2 in the paper (stochastic volatility)



Note: Charts in the top and bottom rows plot the posterior densities of the loadings associated with the contemporaneous and lagged structural shocks, respectively, that affect parameters of Model IV, described in Table 1 in the paper. Shock 1, Shock 2 and Shock 3 correspond to cost-push, demand and monetary policy shocks, respectively. The `2log(BF)` between endogenous and exogenous TVP-VARs is 63.01, suggesting that evidence for endogenous time variation in the VAR coefficients remains very strong when allowing for stochastic volatility. See Kass and Raftery (1995) for details on using the Bayes factor as a metric for model comparison.

Figure A6: Inflation-gap persistence based on $R^2_{h,t}$ statistics (stochastic volatility)



Note: The figure plots the time-varying measure of inflation-gap persistence, $R^2_{h,t}$, for the one-, four- and eight-quarter-ahead forecasting horizons. $R^2_{h,t}$ is constructed using Equation (A33). The estimates are constructed as the medians of the corresponding posterior densities. Charts A, B and C show the results associated with endogenous TVP-VARs with stochastic volatility and with coefficients driven by contemporaneous, lagged, and contemporaneous and lagged structural innovations, respectively.

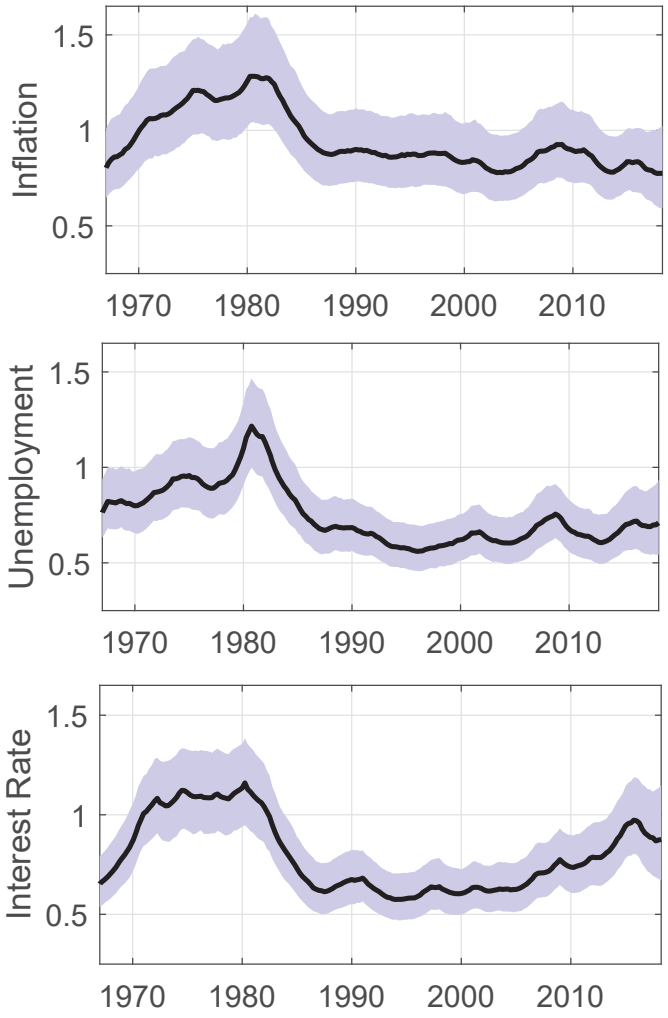
Table A3: Contributions of structural shocks to parameter instability (stochastic volatility)

Total contribution(%)			
Source of parameter instability	*Median*	*16th perc.*	*84th perc.*
Cost-push shock	47.17%	41.15%	53.21%
Demand shock	14.39%	10.21%	19.34%
Monetary policy shock	26.03%	22.32%	30.40%
Residual	12.41%	10.07%	15.21%

Note: The table reports the contributions of the structural shocks to the overall changes in the drifting VAR coefficients. Results are based on Equations (24) and (25) of the paper.

Lastly, we report the evolution of the volatility associated with the innovations in each VAR equation. All in all, results in Figure A7 reinforce the idea that the dynamics for the volatility of key macroeconomic variables can be roughly characterized by two regimes. That is, a pre- and post-mid 1980s volatility regime, each corresponding to a high and low volatility period, respectively. Such a result, however, should not be interpret as a claim that stochastic volatility is an unimportant feature. Rather, it indicates that for small-scale TVP-VARs similar to ours, parsimonious approaches to model volatility, such as fixing the break dates or regime-switching (as in, e.g., Sims and Zha (2006)) are also useful alternatives.

Figure A7: Time-varying volatilities



Note: Solid black lines denote posterior medians for $\sigma_{i,t} = \exp(h_{i,t}/2)$ for $i = 1, 2, 3$ and $t = 1, ..., T$ in the inflation, unemployment and interest rate equations. Shaded areas represent the corresponding 16th and 84th (posterior) percentiles.

A4.2 Prior Sensitivity Analysis

We now consider three types of prior sensitivity checks. First, we conduct additional model comparison exercises between our baseline endogenous TVP-VAR and its exogenous counterpart by allowing for several degrees of prior informativeness associated with the vector of loadings $\boldsymbol{\lambda}$. Specifically, we experiment with $\boldsymbol{\lambda} \sim \mathcal{N}(\mathbf{0},\ 0.1I)$, $\boldsymbol{\lambda} \sim \mathcal{N}(\mathbf{0},\ I)$ and $\boldsymbol{\lambda} \sim \mathcal{N}(\mathbf{0},\ 100I)$. As discussed in Section 3.2 of the paper, $\boldsymbol{\lambda}$ is a key parameter for model comparison, since statistical validation of our framework – via the Savage-Dickey density ratio method – is based on such a parameter.

Second, we examine whether statistical evidence and our results on the contribution of identified shocks to parameter instability would hold under more uninformative priors for second-moment parameters in our baseline model. To this end – and in keeping with Gelman

et al. (2013) (see pages 583-584) – we set to lower values: (i) the shape hyperparameter of the Inverse-Gamma prior for each element in Ω_v; and (ii) the degrees of freedom hyperparameter for the Inverse-Wishart prior associated with covariance matrix of reduced-form errors (Ω_u). More precisely, we set $\nu_{v,i} = 30$ (instead of 60) for $i = 1, ..., M$; and $\mathbf{V}_u = 15$ (instead of 30).

Finally, we also verify whether evidence in favor of our approach holds when relaxing the assumption that the covariance matrix for the coefficient-specific errors (i.e., Ω_v) is diagonal. Consequently, instead of assuming an Inverse-Gamma prior for each diagonal element in Ω_v, we elicit an Inverse-Wishart. In particular, we follow Primiceri (2005) and set $\Omega_v \sim \mathcal{IW}\left(40,\ 0.01^2 \times 40 \times \widehat{V}_{OLS}\right)$, where $\widehat{V}_{OLS}$ is the variance of ordinary least squares estimator for the coefficients of a time-invariant VAR that is fitted to a training sample (1948Q2 to 1962Q2).

Overall, the results reported in the main text are robust to all prior sensitivity checks discussed above. In particular, model comparison results between endogenous and exogenous TVP-VARs are very much in line with the "Barlett's Paradox" (see Bartlett (1957)). In other words, when comparing nested models, evidence in favor of the more complex one (in our case, endogenous TVP-VARs) weakens the flatter the prior for $\boldsymbol{\lambda}$ becomes. This is manifested by the decline in the Bayes factor as the prior for $\boldsymbol{\lambda}$ gradually becomes more uninformative from Checks 1 through 3 in Table A4. Nevertheless, we stress that we still gather strong evidence in favor of endogenous TVP-VARs even for considerably flat priors for $\boldsymbol{\lambda}$.[17] Also, the results in Table A5, based on more uninformative priors for second moments, further confirm our findings in the original manuscript on the importance of identified shocks to coefficient variations.

[17] By "strong evidence" we refer to the recommendations for interpreting Bayes factors as in, e.g., Kass and Raftery (1995), page 777.

Table A4: Model comparison between endogenous and exogenous TVP-VARs under different prior assumptions

Check 1	Check 2	Check 3
35.03	21.07	7.40
Check 4	Check 5	Check 6
63.67	62.13	60.61

Note: Values above denote $2\log(Bayes\ factor)$ between TVP-VARs with and without endogenous time variation. Comparison is based on the most flexible endogenous TVP-VAR variant (i.e. Model IV described in Table 2 of the paper) and its exogenous counterpart. Estimates greater than six and greater than ten should be interpreted as strong and very strong evidence in favor of endogenous TVP-VARs, respectively. See Kass and Raftery (1995) for details on using the Bayes factor as a metric for model comparison. Checks 1 through 3 denote different degrees of prior tightness for the loadings ($\boldsymbol{\lambda}$) in the drifting coefficients equations, with Checks 1, 2 and 3 allowing $\boldsymbol{\lambda}$ to range from tightly, moderately and loosely parameterized around zero, respectively. Checks 4 and 5 set prior hyperparameters for Ω_u and Ω_v, respectively, in order to make second-moment priors more uninformative. Check 6 relaxes the diagonal assumption for Ω_v and assumes the latter is a full covariance matrix.

Table A5: Contributions of structural shocks to parameter instability under more uninformative priors for second moments

Total contribution (%)			
Source of parameter instability	*Median*	*16th perc.*	*84th perc.*
Cost-push shock	55.22%	50.13%	61.04%
Demand shock	9.06%	5.87%	13.07%
Monetary policy shock	27.79%	22.53%	32.83%
Residual	6.99%	6.10%	8.10%

Note: The values above are computed using Equations (24) and (25) of the paper.

A4.3 An Alternative Identification Strategy

In this section, we examine whether a different identification scheme provides similar conclusions to the ones obtained with sign restrictions. Therefore, in what follows, instead of relying on set-identified shocks (such as in the case of sign restrictions), we adopt a more traditional approach based on point identification of shocks via short- and long-run restrictions.

To be clear, we follow Rubio-Ramirez et al. (2010) who – in keeping with the notion of long-run neutrality of monetary and demand shocks – identify monetary policy and demand shocks as having no long-run effects on real activity. In contrast, supply shocks are allowed to have unrestricted long-run effects on all variables in the model. Moreover, monetary policy shocks are assumed to have no contemporaneous impact on real activity, while the

short-run effects of both demand and supply shocks on all variables are left unrestricted.[18] These restrictions are summarized below in Table A6.

Table A6: Summary of short- and long-run restrictions

	Supply shock	Demand shock	Monetary policy shock
Inflation ($t = 0$)	*	*	*
Unemployment ($t = 0$)	*	*	0
Interest Rate ($t = 0$)	*	*	*
Inflation ($t = \infty$)	*	*	*
Unemployment ($t = \infty$)	*	0	0
Interest Rate ($t = \infty$)	*	*	*

Note: * and 0 respectively denote an unrestricted and muted response of the variable in the row to a positive realization of the shock in the column. The top three rows summarize the short-run restrictions ($t = 0$), i.e. how a variable respond to a shock upon impact. The bottom three rows summarize the long-run restrictions ($t = \infty$), i.e. the effect of a shock on a variable in the long horizon. For more details, see Section 6.1 in Rubio-Ramirez et al. (2010).

We also note that the short- and long-run identification approach we adopt provides shocks which are conceptually related to the ones we identify with sign restrictions. More precisely, both identification strategies are, ultimately, predicated on a supply and demand dichotomy. This is useful, as it makes results obtained from these two methods readily comparable. For example, we can assess whether the importance of supply shocks – manifested as cost-push shocks in the case of sign restrictions – is a specific result to sign-identified shocks or it may suggest a more general finding.

Overall, our main results for sign-identified shocks carry over to the case of shocks identified with short-run and long-run restrictions. For example, Figure A8 shows that the posterior densities for each loading (associated with each structural shock driving the VAR coefficients) exhibit most of their masses located away from zero. In particular, the value for twice the natural logarithm of the Bayes factor ($2\log(\text{BF})$) used for model comparison between our baseline endogenous TVP-VAR and its exogenous counterpart is 62.70. Taken together, these results reinforce that statistical evidence in favor of our framework is robust to the choice of identification.[19]

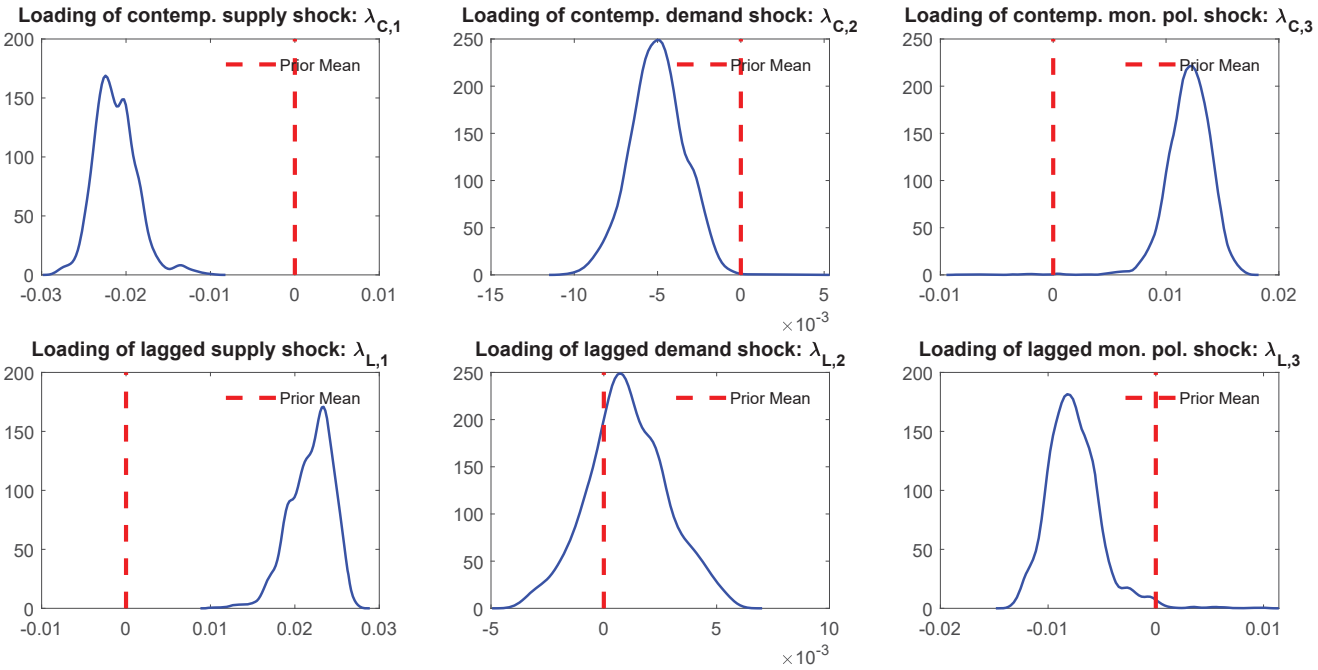
In addition, we also provide results on the contribution of identified shocks to parameter instability. Our results are comparable to the ones obtained in the case of sign-identified shocks. In particular, Table A7 shows that, when identifying supply shocks more broadly (i.e. not just as a cost-push shock), such shocks remain the main driver underlying parameter

[18] The identification approach described above is very much in line with the seminal work by Blanchard and Quah (1989), except that Rubio-Ramirez et al. (2010) also allow for identification of a monetary policy shock. For more details, see pages 683-686 in their paper.

[19] See Kass and Raftery (1995) for details on using the Bayes factor as a metric for model comparison.

variations in our model. In fact, the contribution of supply shocks as drivers of macroeconomic instability becomes even more pronounced, i.e. 73% instead of 53%, as observed for supply (cost-push) shocks identified with signs restrictions.

Figure A8: Posterior density estimates for the loadings associated with each structural shock driving the VAR coefficients under Model IV described in Table 2 of the paper (identification via short-run and long-run restrictions)



Note: The `2log(BF)` between endogenous and exogenous TVP-VARs is 62.70, suggesting very strong evidence in favor of endogenous time variation in the VAR coefficients when shocks are identified using short-run and long-run restrictions as illustrated in Rubio-Ramirez et al. (2010). See Kass and Raftery (1995) for details on using the Bayes factor as a metric for model comparison.

Table A7: Contributions of structural shocks to parameter instability (identification via short-run and long-run restrictions)

Total contribution (%)			
Source of parameter instability	*Median*	*16th perc.*	*84th perc.*
Supply shock	73.36%	67.23%	78.00%
Demand shock	2.47%	1.02%	3.91%
Monetary policy shock	16.24%	11.82%	21.34%
Residual	7.90%	6.86%	9.44%

Note: This table reports the contributions of the structural shocks to the overall changes in the drifting VAR coefficients when shocks are identified using short-run and long-run restrictions as illustrated in Rubio-Ramirez et al. (2010). Contributions are computed using Equations (24) and (25) of the paper.

A5 Generalized Impulse Responses

TVP-VARs have been widely been used to study changes in the transmission of aggregate shocks. Therefore, in this section we report time-varying impulse responses for all shocks considered in our empirical applications. To do so, we apply a common approach to define impulse responses for nonlinear models, namely generalized impulse response functions (GIRFs). In particular, to construct the GIRFs, we apply the algorithm discussed in Koop et al. (1996), which we summarize below.

GIRFs are defined as the difference between the conditional expectations of the variables in the model with and without the influence of a shock of size δ, that is,

$$GIRF_{t+h} = E[y_{t+h}|e_t = \delta, \psi_t] - E[y_{t+h}|e_t = 0, \psi_t],$$

where h denotes the horizon of the response and ψ_t denotes the set of information available up to time t. To compute the $GIRF_{t+h}$ we follow the next steps:

Step 1. Randomly draw from the Gibbs sample output one state of the economy, $\phi_t^{(j)}$, and a set of the model parameters, $A^{(j)}$, $\lambda_C^{(j)}$, $\lambda_L^{(j)}$, and $,\Omega_v^{(j)}$.[20]

Step 2. Generate a sequence of the structural shocks, $\xi_t^{(j)} = \left\{e_{t-1}^{(j)}, e_t^{(j)}, e_{t+1}^{(j)}, ..., e_{t+h}^{(j)}\right\}$, from a $N(0,1)$.

Step 3. Given $\lambda_C^{(j)}$, $\lambda_L^{(j)}$, and $,\Omega_v^{(j)}$, use the law of motion defined in Equation (6), of the main text, to generate a sequence of the time-varying coefficients, $\Phi_t^{(j)} = \left\{\phi_t^{(j)}, \phi_{t+1}^{(j)}, ..., \phi_{t+h}^{(j)}\right\}$.

Step 4. Given $\xi_t^{(j)}$ and $A^{(j)}$, generate a sequence of the reduced form innovations $U_t^{(j)} = \{u_t^{(j)}, u_{t+1}^{(j)}, ..., u_{t+h}^{(j)}\}$.

Step 5. Given $\Phi_t^{(j)}$ and $U_t^{(j)}$, compute the evolution of the variables y_t for the next h periods conditional on a shock of size δ, $E[y_{t+h}|e_t = \delta, \psi_t]$, and conditional on a shock of size 0, $E[y_{t+h}|e_t = 0, \psi_t]$. The difference between these two quantities defines our impulse response.

Step 6. Repeat steps 1 to 5 for $j = 1, ..., J$. Then, compute the mean of those J iterations, which constitutes one draw of the $GIRF_{t+h}$. We set $J = 100$.

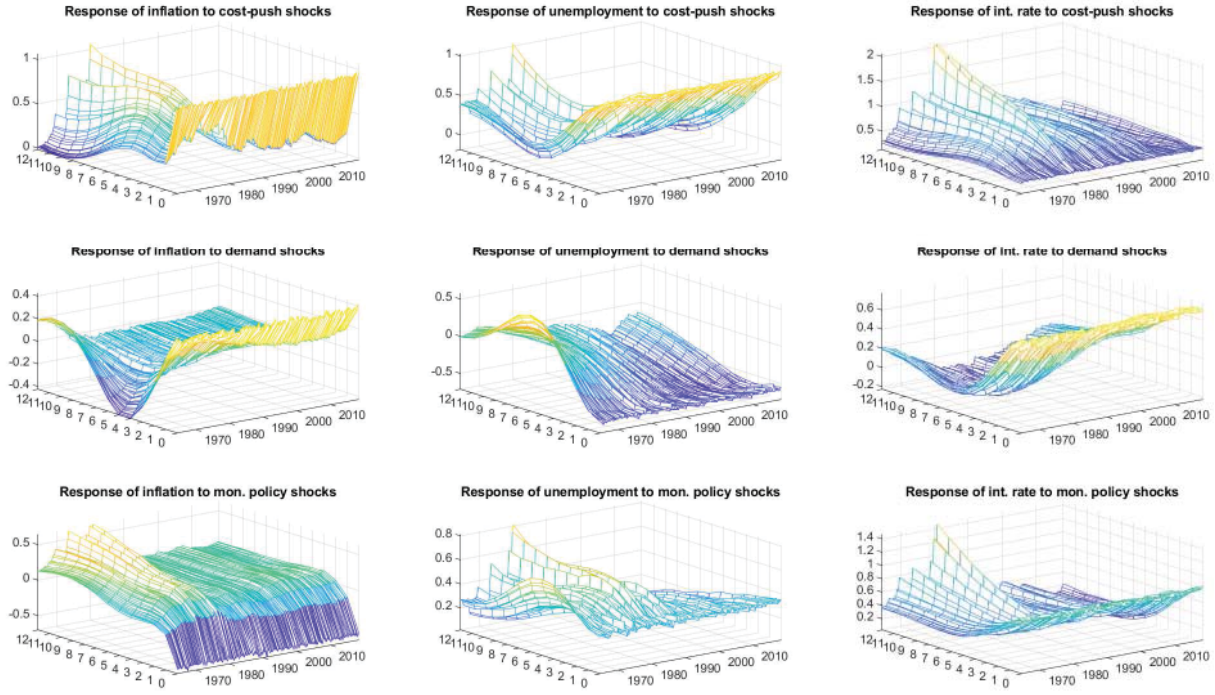
For each time period, we generate 500 draws, from which the median and corresponding quantiles can be used as estimates and credible sets, respectively.

Results for all GIRFs based on the procedure described above are shown in Figure (A9).

[20] Note that any selected draw of the impact multiplier matrix, $A^{(j)}$, already satisfies the sign restrictions used to identify the structural shocks.

Figure A9: Generalized impulse response functions for the Endogenous TVP-VAR model

(a) Responses of variables across time and horizons



(b) Four quarters ahead responses of variables

Note: In each chart, solid blue lines show the posterior median of the generalized impulse response distributions, while dashed red lines indicate the percentiles 16th and 84th of the corresponding posterior dentisty.

References

Bartlett, M. S. (1957). A comment on D.V. Lindley's statistical paradox. *Biometrika*, 44(3-4):533–534.

Baumeister, C. and Peersman, G. (2013). Time-varying effects of oil supply shocks on the us economy. *American Economic Journal: Macroeconomics*, 5(4):1–28.

Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, 19:655–73.

Carriero, A., Clark, T. E., and Marcellino, M. (2018). Measuring uncertainty and its impact on the economy. *Review of Economics and Statistics*, 100(5):799–815.

Chan, J. C. and Jeliazkov, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120.

Chib, S. (2001). Markov chain monte carlo methods: computation and inference. *Handbook of Econometrics*, 5:3569–3649.

Cogley, T., Primiceri, G., and Sargent, T. (2010). Inflation-gap persistence in the us. *American Economic Journal: Macroeconomics*, 2(1):43–69.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post wwii us. *Review of Economic Dynamics*, 8(2):262–302.

Del Negro, M. and Primiceri, G. E. (2015). Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum. *The review of economic studies*, 82(4):1342–1345.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kim, S., Shepard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65(3):361–393.

Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, 74(1):119–147.

Koop, G., Poirier, D. J., and Tobias, J. L. (2007). *Bayesian Econometric Methods*. Cambridge University Press.

Koop, G. and Potter, S. (2011). Time varying vars with inequality restrictions. *Journal of Economic Dynamics and Control*, 35(7):1126–1138.

Kroese, D. P., Taimre, T., and Botev, Z. I. (2013). *Handbook of Monte Carlo Methods*, volume 706. John Wiley & Sons.

Mumtaz, H. and Zanetti, F. (2015). Labor market dynamics: a time-varying analysis. *Oxford Bulletin of Economics and Statistics*, 77(3):319–338.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449.

Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72:821–852.

Rubio-Ramirez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *The Review of Economic Studies*, 77(2):665–696.

Sims, C. and Zha, T. (2006). Were there regime switches in the u.s. monetary policy? *American Economic Review*, 96(1):54–81.

Slobodyan, S. and Wouters, R. (2012). Learning in a medium-scale dsge model with expectations based on small forecasting models. *American Economic Journal: Macroeconomics*, 4(2):65–101.

Smets, F. and Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 97(3):586–606.

Stock, J. H. and Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, 39:3–33.

Stock, J. H. and Watson, M. W. (2016). Core inflation and trend inflation. *Review of Economics and Statistics*, 98(4):770–784.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

Van Loan, C. F. and Golub, G. H. (1983). *Matrix computations*. Johns Hopkins University Press Baltimore.

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

2001 JAVIER ANDRÉS, PABLO BURRIEL and WENYI SHEN: Debt sustainability and fiscal space in a heterogeneous Monetary Union: normal times vs the zero lower bound.

2002 JUAN S. MORA-SANGUINETTI and RICARDO PÉREZ-VALLS: ¿Cómo afecta la complejidad de la regulación a la demografía empresarial? Evidencia para España.

2003 ALEJANDRO BUESA, FRANCISCO JAVIER POBLACIÓN GARCÍA and JAVIER TARANCÓN: Measuring the procyclicality of impairment accounting regimes: a comparison between IFRS 9 and US GAAP.

2004 HENRIQUE S. BASSO and JUAN F. JIMENO: From secular stagnation to robocalypse? Implications of demographic and technological changes.

2005 LEONARDO GAMBACORTA, SERGIO MAYORDOMO and JOSÉ MARÍA SERENA: Dollar borrowing, firm-characteristics, and FX-hedged funding opportunities.

2006 IRMA ALONSO ÁLVAREZ, VIRGINIA DI NINO and FABRIZIO VENDITTI: Strategic interactions and price dynamics in the global oil market.

2007 JORGE E. GALÁN: The benefits are at the tail: uncovering the impact of macroprudential policy on growth-at-risk.

2008 SVEN BLANK, MATHIAS HOFFMANN and MORITZ A. ROTH: Foreign direct investment and the equity home bias puzzle.

2009 AYMAN EL DAHRAWY SÁNCHEZ-ALBORNOZ and JACOPO TIMINI: Trade agreements and Latin American trade (creation and diversion) and welfare.

2010 ALFREDO GARCÍA-HIERNAUX, MARÍA T. GONZÁLEZ-PÉREZ and DAVID E. GUERRERO: Eurozone prices: a tale of convergence and divergence.

2011 ÁNGEL IVÁN MORENO BERNAL and CARLOS GONZÁLEZ PEDRAZ: Sentiment analysis of the Spanish Financial Stability Report. (There is a Spanish version of this edition with the same number).

2012 MARIAM CAMARERO, MARÍA DOLORES GADEA-RIVAS, ANA GÓMEZ-LOSCOS and CECILIO TAMARIT: External imbalances and recoveries.

2013 JESÚS FERNÁNDEZ-VILLAVERDE, SAMUEL HURTADO and GALO NUÑO: Financial frictions and the wealth distribution.

2014 RODRIGO BARBONE GONZALEZ, DMITRY KHAMETSHIN, JOSÉ-LUIS PEYDRÓ and ANDREA POLO: Hedger of last resort: evidence from Brazilian FX interventions, local credit, and global financial cycles.

2015 DANILO LEIVA-LEON, GABRIEL PEREZ-QUIROS and EYNO ROTS: Real-time weakness of the global economy: a first assessment of the coronavirus crisis.

2016 JAVIER ANDRÉS, ÓSCAR ARCE, JESÚS FERNÁNDEZ-VILLAVERDE and SAMUEL HURTADO: Deciphering the macroeconomic effects of internal devaluations in a monetary union.

2017 JACOPO TIMINI, NICOLA CORTINOVIS and FERNANDO LÓPEZ VICENTE: The heterogeneous effects of trade agreements with labor provisions.

2018 EDDIE GERBA and DANILO LEIVA-LEON: Macro-financial interactions in a changing world.

2019 JAIME MARTÍNEZ-MARTÍN and ELENA RUSTICELLI: Keeping track of global trade in real time.

2020 VICTORIA IVASHINA, LUC LAEVEN and ENRIQUE MORAL-BENITO: Loan types and the bank lending channel.

2021 SERGIO MAYORDOMO, NICOLA PAVANINI and EMANUELE TARANTINO: The impact of alternative forms of bank consolidation on credit supply and financial stability.

2022 ALEX ARMAND, PEDRO CARNEIRO, FEDERICO TAGLIATI and YIMING XIA: Can subsidized employment tackle long-term unemployment? Experimental evidence from North Macedonia.

2023 JACOPO TIMINI and FRANCESCA VIANI: A highway across the Atlantic? Trade and welfare effects of the EU-Mercosur agreement.

2024 CORINNA GHIRELLI, JAVIER J. PÉREZ and ALBERTO URTASUN: Economic policy uncertainty in Latin America: measurement using Spanish newspapers and economic spillovers.

2025 MAR DELGADO-TÉLLEZ, ESTHER GORDO, IVÁN KATARYNIUK and JAVIER J. PÉREZ: The decline in public investment: “social dominance” or too-rigid fiscal rules?

2026 ELVIRA PRADES-ILLANES and PATROCINIO TELLO-CASAS: Spanish regions in Global Value Chains: How important? How different?

2027 PABLO AGUILAR, CORINNA GHIRELLI, MATÍAS PACCE and ALBERTO URTASUN: Can news help measure economic sentiment? An application in COVID-19 times.

2028 EDUARDO GUTIÉRREZ, ENRIQUE MORAL-BENITO, DANIEL OTO-PERALÍAS and ROBERTO RAMOS: The spatial distribution of population in Spain: an anomaly in European perspective.

2029 PABLO BURRIEL, CRISTINA CHECHERITA-WESTPHAL, PASCAL JACQUINOT, MATTHIAS SCHÖN and NIKOLAI STÄHLER: Economic consequences of high public debt: evidence from three large scale DSGE models.

2030 BEATRIZ GONZÁLEZ: Macroeconomics, Firm Dynamics and IPOs.

2031 BRINDUSA ANGHEL, NÚRIA RODRÍGUEZ-PLANAS and ANNA SANZ-DE-GALDEANO: Gender Equality and the Math Gender Gap.

2032 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Machine learning in credit risk: measuring the dilemma between prediction and supervisory cost.

2033 PILAR GARCÍA-PEREA, AITOR LACUESTA and PAU ROLDAN-BLANCO: Raising Markups to Survive: Small Spanish Firms during the Great Recession.

2034 MÁXIMO CAMACHO, MATÍAS PACCE and GABRIEL PÉREZ-QUIRÓS: Spillover Effects in International Business Cycles.

2035 ÁNGEL IVÁN MORENO and TERESA CAMINERO: Application of text mining to the analysis of climate-related disclosures.

2036 EFFROSYNI ADAMOPOULOU and ERNESTO VILLANUEVA: Wage determination and the bite of collective contracts in Italy and Spain: evidence from the metal working industry.

2037 MIKEL BEDAYO, GABRIEL JIMÉNEZ, JOSÉ-LUIS PEYDRÓ and RAQUEL VEGAS: Screening and Loan Origination Time: Lending Standards, Loan Defaults and Bank Failures.

2038 BRINDUSA ANGHEL, PILAR CUADRADO and FEDERICO TAGLIATI: Why cognitive test scores of Spanish adults are so low? The role of schooling and socioeconomic background

2039 CHRISTOPH ALBERT, ANDREA CAGGESE and BEATRIZ GONZÁLEZ: The Short- and Long-run Employment Impact of COVID-19 through the Effects of Real and Financial Shocks on New Firms.

2040 GABRIEL JIMÉNEZ, DAVID MARTÍNEZ-MIERA and JOSÉ-LUIS PEYDRÓ: Who Truly Bears (Bank) Taxes? Evidence from Only Shifting Statutory Incidence.

2041 FELIX HOLUB, LAURA HOSPIDO and ULRICH J. WAGNER: Urban air pollution and sick leaves: evidence from social security data.

2042 NÉLIDA DÍAZ SOBRINO, CORINNA GHIRELLI, SAMUEL HURTADO, JAVIER J. PÉREZ and ALBERTO URTASUN: The narrative about the economy as a shadow forecast: an analysis using Banco de España quarterly reports.

2043 NEZIH GUNER, JAVIER LÓPEZ-SEGOVIA and ROBERTO RAMOS: Reforming the individual income tax in Spain.

2101 DARÍO SERRANO-PUENTE: Optimal progressivity of personal income tax: a general equilibrium evaluation for Spain.

2102 SANDRA GARCÍA-URIBE, HANNES MUELLER and CARLOS SANZ: Economic uncertainty and divisive politics: evidence from the *Dos Españas*.

2103 IVÁN KATARYNIUK, VÍCTOR MORA-BAJÉN and JAVIER J. PÉREZ: EMU deepening and sovereign debt spreads: using political space to achieve policy space.

2104 DARÍO SERRANO-PUENTE: Are we moving towards an energy-efficient low-carbon economy? An input-output LMDI decomposition of CO_2 emissions for Spain and the EU28.

2105 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Understanding the performance of machine learning models to predict credit default: a novel approach for supervisory evaluation.

2106 JAVIER ANDRÉS, ÓSCAR ARCE and PABLO BURRIEL: Market polarization and the Phillips curve.

2107 JUAN de LUCIO and JUAN S. MORA-SANGUINETTI: New dimensions of regulatory complexity and their economic cost. An analysis using text mining.

2108 DANILO LEIVA-LEON and LUIS UZEDA: Endogenous time variation in vector autoregressions.