

Endogenous Time Variation in Vector Autoregressions

by Danilo Leiva-Leon¹ and Luis Uzeda²

¹Banco de España

danilo.leiva@bde.es

²Canadian Economic Analysis Department
Bank of Canada, Ottawa, Ontario, Canada K1A 0G9

luzedagarcia@bank-banque-canada.ca



Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

Acknowledgements

For their comments and suggestions, we would like to thank Maximo Camacho, Fabio Canova, Efrem Castelnuovo, Daniel de Munnik, 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 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.

Abstract

We introduce a new class of time-varying parameter vector autoregressions (TVP-VARs) where the identified structural innovations are allowed to influence — contemporaneously and with a lag — the dynamics of the intercept and autoregressive 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 the effects of monetary policy on economic activity are larger and more persistent in the proposed models than in an otherwise standard TVP-VAR. Our results also indicate that cost-push shocks play an important role in understanding historical changes in inflation persistence.

Topics: Econometric and statistical methods; Transmission of monetary policy; Inflation and prices

JEL codes: C11; C32; E31; E52

Résumé

Nous présentons une nouvelle classe de modèles vectoriels autorégressifs (VAR) à paramètres variables dans le temps dans laquelle les innovations structurelles désignées explicitement peuvent influencer – simultanément et en décalage – sur la dynamique de la constante et des coefficients autorégressifs. Nous fournissons également un algorithme d'estimation et une paramétrisation facilitant la comparaison des modèles. Nous appliquons ensuite ce cadre à l'économie américaine. Selon notre analyse de scénarios, les effets de la politique monétaire sur l'activité économique seraient plus importants et plus persistants dans les modèles proposés que dans un modèle VAR à paramètres variables dans le temps autrement standard. Nos résultats montrent aussi que les chocs d'inflation par les coûts aideraient grandement à expliquer les variations de la persistance de l'inflation au fil du temps.

Sujets : Méthodes économétriques et statistiques; Transmission de la politique monétaire; Inflation et prix

Codes JEL : C11; C32; E31; E52

Non-Technical Summary

Macroeconometric models subject to parameter instabilities have been successfully used to assess the effectiveness of policies in a constantly changing economic environment. Interestingly, understanding the drivers underlying such parameter instabilities is something that has remained largely unaddressed to date. We provide an empirical framework that helps to shed light on the contribution of economic shocks responsible for changes in the relationship between key economic variables.

We propose a new class of vector autoregression models with time-varying parameters, where the underlying structural shocks are allowed to affect parameter changes and, consequently, the relationship between the variables embedded in the model. An estimation method to provide robust inference for this new class of models is also provided along with procedures to determine the validity of the proposed framework.

Our methodology is applied to study macroeconomic instabilities in the US economy. We show that cost-push shocks have been a prominent driver behind transitory fluctuations in the US inflation rate. Moreover, after implementing a couple of scenario analyses, we demonstrate that the effects of monetary policy shocks on the US economy tend to be amplified when allowing for the possibility that policy actions can directly impact the parameters of the model in question.

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, Litterman, and Sims \(1984\)](#), [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

¹For example, [Cogley, Matthes, and Sbordone \(2015\)](#) show that DSGE models with learning can be recast as a reduced-form TVP-VAR.

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

strategy provides a TVP-VAR framework that not only accounts for parameter changes, but is also informative on what drives such changes.

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 more traditional TVP-VARs. This is achieved by specifying the law of motion for the time-varying coefficients as a function of two distinct 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

³For 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.

⁴Keeping 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.

(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 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, Primiceri, and Sargent \(2010\)](#) 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

⁵Sign 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.

(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 dynamic responses for inflation and unemployment that emerge from these two models in response to monetary policy 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, Piger, and Startz \(2008\)](#) 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, Clark, and Marcellino \(2018\)](#) 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

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

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 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 $y_t = (y_{1,t}, y_{2,t}, \dots, y_{N,t})'$ 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 $\phi_t = (\phi_{1,t}, \dots, \phi_{N,t})'$, such that each element $\phi_{i,t}$ for $i = 1, \dots, N$ 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 $X_t = I_N \otimes (1, y'_{t-1}, \dots, y'_{t-p})$, 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

⁷While 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, Ferroni, and Matthes \(2015\)](#)).

⁸We discuss the structure of A in Section 4.

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. 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 \(2006\)](#), [Koop, Leon-Gonzalez, and Strachan \(2009\)](#) 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 second-moment breaks in both measurement and state equations.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

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

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 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.¹¹ Formally, we have:

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

where $\Omega_v = \text{diag}(\sigma_{v,1}^2, \dots, \sigma_{v,N^2p+N}^2)$ and $H_{C,\lambda} = \lambda_C' \otimes \iota$ denotes the Kronecker product between the $N \times 1$ vector $\lambda_C = (\lambda_{C,1}, \dots, \lambda_{C,N})'$ 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

¹⁰For 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 \(2007\)](#)), time-to-build effects (e.g., [Kydland and Prescott \(1982\)](#)), and capital adjustment costs (e.g., [Cooper and Haltiwanger \(2006\)](#)).

¹¹This might be the case, for example, in financial applications where variables usually tend to experience abrupt changes rather than gradual ones. For example, uncertainty shocks due to unexpected political outcomes could almost instantaneously influence the relationship between financial and macroeconomic variables, such as stock exchange indexes and expectations of agents about the future path of policy rates.

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 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:

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

As in the contemporaneous case, we set $\Omega_v = \text{diag}(\sigma_{v,1}^2, \dots, \sigma_{v,N^2p+N}^2)$ 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 given by:

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

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:

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

Given its generality, the model defined by equations (1) and (6) constitutes our benchmark specification in the empirical applications in Section 4. Nevertheless, we also show selected results associated with more restrictive specifications for comparison purposes.

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 \(2009\)](#) and [Chan \(2013\)](#), 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

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

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

stacked over $t = 1, \dots, T$. 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 $\mathbf{y} = (y_1, \dots, y_T)'$ 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:

$$\mathbf{y} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\phi}} + \mathbf{L}_A\mathbf{e}, \quad (8)$$

$$\tilde{\boldsymbol{\phi}} = \tilde{\boldsymbol{\phi}}_0 + \mathbf{L}_\lambda\mathbf{e} + \mathbf{v}, \quad (9)$$

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

where:

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

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

¹⁴For a textbook treatment of the Kalman filter, the reader is referred to [Durbin and Koopman \(2012\)](#). For a detailed comparison between Kalman filter- and precision-based algorithms, see [McCausland, Miller, and Pelletier \(2011\)](#).

$\tilde{\phi}_0 = (\phi_0, 0, \dots, 0)'$. 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}_λ , i.e.

$$\mathbf{L}_\lambda = \left\{ \begin{array}{c} \text{(Contemporaneous + Lagged)} \\ \left[\begin{array}{cccc} H_{C,\lambda} & 0 & \cdots & 0 \\ H_{L,\lambda} & H_{C,\lambda} & \cdots & 0 \\ 0 & H_{L,\lambda} & \ddots & \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & H_{L,\lambda} & H_{C,\lambda} \end{array} \right], \end{array} \right. \left\{ \begin{array}{c} \text{(Contemporaneous)} \\ \left[\begin{array}{cccc} H_{C,\lambda} & 0 & \cdots & 0 \\ 0 & H_{C,\lambda} & \cdots & 0 \\ 0 & 0 & \ddots & \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & H_{C,\lambda} & \end{array} \right], \end{array} \right\} \left\{ \begin{array}{c} \text{(Lagged)} \\ \left[\begin{array}{cccc} 0 & 0 & \cdots & 0 \\ H_{L,\lambda} & 0 & \cdots & 0 \\ 0 & H_{L,\lambda} & \ddots & \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & H_{L,\lambda} & 0 \end{array} \right] \end{array} \right\}.$$

3.1 Sampling Endogenous States

Now let $\theta = \{\mathbf{L}_A, \tilde{\phi}_0, \mathbf{L}_\lambda, \Sigma_v\}$ 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) $f(\tilde{\phi}|\mathbf{y}, \theta)$; and (ii) $f(\theta|\mathbf{y}, \tilde{\phi})$. In what follows, we focus on deriving an expression for the full conditional posterior, $f(\tilde{\phi}|\mathbf{y}, \theta)$, since it represents the main contribution of this paper on the estimation front. Sampling details for θ and a discussion on the priors can be found in the Online Appendix.¹⁵

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

$$\tilde{\phi}|\theta \sim \mathcal{N}(\mu_{\tilde{\phi}}, \Sigma_{\tilde{\phi}\tilde{\phi}}), \quad (11)$$

¹⁵It 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).

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

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

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\phi}}\tilde{\boldsymbol{\phi}}} = \text{Var}(\mathbf{L}_\lambda \mathbf{e} + \mathbf{v}) = \mathbf{L}_\lambda \text{Var}(\mathbf{e}) \mathbf{L}_\lambda' + \text{Var}(\mathbf{v}) = \mathbf{L}_\lambda \mathbf{L}_\lambda' + \boldsymbol{\Sigma}_v. \quad (13)$$

Next, to derive an expression for the likelihood, we use a standard result for multivariate Normal distributions, namely the joint density $f(\mathbf{y}, \tilde{\boldsymbol{\phi}}|\boldsymbol{\theta})$ can be factorized into a conditional and a marginal coordinate, i.e. $f(\mathbf{y}, \tilde{\boldsymbol{\phi}}|\boldsymbol{\theta}) = f(\mathbf{y}|\tilde{\boldsymbol{\phi}}, \boldsymbol{\theta})f(\tilde{\boldsymbol{\phi}}|\boldsymbol{\theta})$. 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 $f(\mathbf{y}|\tilde{\boldsymbol{\phi}}, \boldsymbol{\theta})$ is Gaussian such that:

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

The expression above introduces three new terms, $\boldsymbol{\mu}_y$, $\boldsymbol{\Sigma}_{yy}$ and $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\phi}}y}$. The first two denote first and second moments, respectively, which – conditional on $\tilde{\boldsymbol{\phi}}$ and $\boldsymbol{\theta}$ – are obtained from the measurement equation in (8).¹⁷ Specifically, we have:

$$\boldsymbol{\mu}_y = \mathbb{E} \left(\tilde{\mathbf{X}}\tilde{\boldsymbol{\phi}} + \mathbf{L}_A \mathbf{e} \right) = \tilde{\mathbf{X}}\tilde{\boldsymbol{\phi}}, \quad (15)$$

$$\boldsymbol{\Sigma}_{yy} = \text{Var}(\mathbf{L}_A \mathbf{e}) = \mathbf{L}_A \text{Var}(\mathbf{e}) \mathbf{L}_A' = \mathbf{L}_A \mathbf{L}_A'. \quad (16)$$

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

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\phi}}y} = \text{Cov}(\mathbf{L}_\lambda \mathbf{e}, \mathbf{L}_A \mathbf{e}) = \mathbf{L}_\lambda \text{Cov}(\mathbf{e}, \mathbf{e}) \mathbf{L}_A' = \mathbf{L}_\lambda \text{Var}(\mathbf{e}) \mathbf{L}_A' = \mathbf{L}_\lambda \mathbf{L}_A'. \quad (17)$$

Of course, in the absence of endogenous time variation, \mathbf{L}_λ is a null matrix and so is $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\phi}}y}$,

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

¹⁷Strictly speaking, \mathbf{y} in (14) is also conditional on \mathbf{X} (and so is $\tilde{\boldsymbol{\phi}}$ in the full conditional posterior $f(\tilde{\boldsymbol{\phi}}|\mathbf{y}, \boldsymbol{\theta})$). However, we follow the common practice of omitting regressors when presenting the conditional factors for random variables and random vectors in a regression setting.

which returns an expression for (14) that is consistent with the likelihood function for exogenous TVP-VARs.¹⁸

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

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

where we define:

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

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

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

To produce draws for $\tilde{\phi}|\mathbf{y}, \boldsymbol{\theta}$ one needs to construct $\bar{\mathbf{d}}_{\tilde{\phi}}$ and $\bar{\mathbf{D}}_{\tilde{\phi}}$. 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 $\phi = \mathbf{L}^{-1}\tilde{\phi}$.

¹⁸A similar rationale applies for the prior variance-covariance matrix $\boldsymbol{\Sigma}_{\tilde{\phi}\tilde{\phi}}$ in (13).

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:

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

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:¹⁹

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

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

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 using Gibbs draws from the conditional posterior $f(\boldsymbol{\lambda}|\mathbf{y}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\theta}_{-\boldsymbol{\lambda}})$, 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

¹⁹Loosely 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.

posterior in terms of \mathbf{L}_λ are not of a known form. Thus, by a simple change of variable $\mathbf{L}_\lambda \mathbf{e} = \mathbf{L}_e \boldsymbol{\lambda}$, we can rewrite (9) as:

$$\tilde{\boldsymbol{\phi}} = \tilde{\boldsymbol{\phi}}_0 + \mathbf{L}_e \boldsymbol{\lambda} + \mathbf{v}, \quad (21)$$

where

$$\mathbf{L}_e \boldsymbol{\lambda} = \left(\left(\begin{bmatrix} e'_1 & 0' \\ e'_2 & e'_2 \\ \vdots & \vdots \\ e'_T & e'_{T-1} \end{bmatrix} \otimes \iota \right) \begin{bmatrix} \lambda_{C,1} \\ \vdots \\ \lambda_{C,N} \\ \lambda_{L,1} \\ \vdots \\ \lambda_{L,N} \end{bmatrix} \right).$$

That is, \mathbf{L}_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} , $\tilde{\boldsymbol{\phi}}$ and \mathbf{L}_A in $f(\boldsymbol{\lambda}|\mathbf{y}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\theta}_{-\lambda})$, to back out the vector of structural shocks \mathbf{e} from (8) and treat \mathbf{L}_e as a predetermined regressor in Equation (21).²⁰ This means that derivation of an expression for $f(\boldsymbol{\lambda}|\mathbf{y}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\theta}_{-\lambda})$ – 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, Poirier, and Tobias (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 the second, we conduct scenario analysis for inflation and unemployment under different monetary policy decisions about the trajectory of interest rates.

²⁰For simplicity, we set pre-sample values of e_t to zero.

In all these applications, we employ a small-scale structural TVP-VAR along the lines of [Cogley and Sargent \(2001\)](#) and [Cogley, Primiceri, and Sargent \(2010\)](#). 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 1947Q1 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 1962Q1 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 innovations. As pointed out in [Canova \(2011\)](#), sign restrictions provide a strategy to incorporate non-parametric (sign-related) restrictions from DSGE models into VARs.

²¹The 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](#).

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, Gali, and Gertler \(1999\)](#) 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 the desired sign restrictions, we need to redefine the reduced-form errors in [Equation \(1\)](#), such that $u_t = APe_t$ (instead of $u_t = Ae_t$), where P is a rotation matrix that spans the space of orthogonal decompositions for u_t associated with any possible permutation of the signs in [Table 1](#). In other words, the matrix P can generate structural shocks whose contemporaneous effects on inflation, unemployment and the policy rate conform with the signs in [Table 1](#). Also, since P is a matrix from the orthogonal group, we have $PP' = I$, hence leaving the unconditional covariance matrix for u_t unchanged, i.e. $\mathbb{E}(u_t u_t') = \mathbb{E}(APe_t e_t' P' A') = AP \mathbb{E}(e_t e_t') P' A' = APP' A' = AA'$.

To obtain P we use the algorithm proposed by [Rubio-Ramirez, Waggoner, and Zha \(2010\)](#). In practice, we retain only the draws for P from their algorithm that satisfy the restrictions in [Table 1](#). Accordingly, let $\tilde{e}_t = Pe_t$ denote the sign-identified structural

innovations. The model in (1) and (6) can then be recast as:

$$y_t = X_t\phi_t + A\tilde{e}_t, \quad (22)$$

$$\phi_t = \phi_{t-1} + H_{C,\lambda}\tilde{e}_t + H_{L,\lambda}\tilde{e}_{t-1} + v_t. \quad (23)$$

The state simulation strategy discussed in Section 3.1 is thus applied to the endogenous TVP-VAR representation in (22)-(23).

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 (λ) 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 $\lambda = 0$. 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 ($21\log(\text{BF})$) 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 2.²² 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 $21\log(\text{BF})$ 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

²²Kass and Raftery (1995) suggest the use of $21\log(\text{BF})$ since such a metric is on the same scale as other familiar test statistics like the deviance information criteria and the likelihood ratio test.

a factor greater than 150.²³

We also compute the Bayes factors between endogenous and exogenous TVP-VARs when coefficients in the former are a function of a single structural shock. The values for $2\log(\text{BF})$ are reported in Table 4. Again, our results favor endogenous over exogenous time variation, although 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 how a particular shock might affect interdependencies in a VAR, then allowing for the contemporaneous rather than the lagged effect of such a shock seems more appropriate.

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, Primiceri, and Sargent \(2010\)](#) 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 verify if a modeling strategy that relies on single framework – i.e., endogenous TVP-VARs – is able to deliver results on inflation-gap persistence that are, at least at a high level, comparable to the findings in [Cogley, Primiceri, and Sargent \(2010\)](#). 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, Primiceri, and Sargent \(2010\)](#) propose a metric that defines persistence

²³Such 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 \(1995\)](#).

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 $R_{h,t}^2$ 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 $R_{h,t}^2$ converges to zero as the forecast horizon h extends. For the sake of brevity, technical details on how to construct $R_{h,t}^2$ are relegated to the Online Appendix. For now, it suffices to remember that $R_{h,t}^2$ is a metric bounded between zero and one and that what characterizes changes in inflation-gap persistence is the pace at which $R_{h,t}^2$ converges to zero as h increases. Therefore, a highly (weakly) persistent process at time t is one whose $R_{h,t}^2$ statistic slowly (quickly) converges to zero as h lengthens.

Figure 2 plots the evolution of $R_{h,t}^2$ 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 $R_{h,t}^2$ remains close to one even at $h = 8$. In contrast, after the early 1980s the decline in inflation-gap persistence becomes apparent with $R_{h,t}^2$ dropping more rapidly as h increases for virtually all points in time. All in all, these results are broadly in line with what [Cogley, Primiceri, and Sargent \(2010\)](#) 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 $R_{h,t}^2$ 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 current narratives on inflation being persistently below its target since the end of the 2008 financial crisis (see, e.g., [Andolfatto and Spewak \(2019\)](#)).

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 $R_{h,t}^2$ in Figure 2 are directly related to movements in the VAR coefficients.²⁴ 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.²⁵ Therefore, to further interpret the results on inflation-gap persistence, we propose a statistical measure along the lines of [Kose, Otrok, and Whiteman \(2003\)](#) that captures the contribution of each structural shock to the overall variation in the drifting coefficients. More precisely, note from (23) 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:

$$\Delta\phi_{j,t} = \sum_{i=1}^N (\lambda_{C,i}\tilde{\epsilon}_{i,t} + \lambda_{L,i}\tilde{\epsilon}_{i,t-1}) + v_{j,t}, \quad \text{for } j = 1, \dots, M, \quad (24)$$

where M denotes the total number of drifting coefficients, i.e. $M = N^2p + N$.²⁶

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 (24) are mutually and serially uncorrelated. Also, as discussed in Section 2.2, to mitigate parameter proliferation we assumed that the term $\sum_{i=1}^N (\lambda_{C,i}\tilde{\epsilon}_{i,t} + \lambda_{L,i}\tilde{\epsilon}_{i,t-1})$ 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

²⁴Again, details on the construction of $R_{h,t}^2$ can be found in the Online Appendix or in [Cogley, Primiceri, and Sargent \(2010\)](#).

²⁵Notably, 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.

²⁶As previously defined, N denotes the number of structural shocks and p stands for lag-length.

variance of $\Delta\phi_{j,t}$ over $j = 1, \dots, M$. Formally, from (24), we set:

$$\begin{aligned} \text{Total Coefficient Variation} &= \frac{1}{M} \sum_{j=1}^M \text{Var}(\Delta\phi_{j,t}), \\ &= \sum_{i=1}^N (\lambda_{C,i}^2 + \lambda_{L,i}^2) + \frac{1}{M} \sum_{j=1}^M \sigma_{v,j}^2. \end{aligned} \quad (25)$$

The metric above decomposes changes in the VAR coefficients into two distinguishable components.²⁷ The first summation in (25) 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:

$$\text{Share (Shock } i) = \frac{\lambda_{C,i}^2 + \lambda_{L,i}^2}{\sum_{i=1}^N (\lambda_{C,i}^2 + \lambda_{L,i}^2) + \frac{1}{M} \sum_{j=1}^M \sigma_{v,j}^2}, \quad (26)$$

for $i = \text{Cost-Push, Demand, Monetary Policy, and}$

$$\text{Share (Residual)} = \frac{\frac{1}{M} \sum_{j=1}^M \sigma_{v,j}^2}{\sum_{i=1}^N (\lambda_{C,i}^2 + \lambda_{L,i}^2) + \frac{1}{M} \sum_{j=1}^M \sigma_{v,j}^2}. \quad (27)$$

Figure 3 plots the contribution from both identified and unidentified innovations to the overall degree of variation in the VAR coefficients according to (26) and (27). It is straightforward to see that 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

²⁷Since we assume that the innovations on the right-hand-side of (24) are independently and identically distributed, $\text{Var}(\Delta\phi_{j,t})$ can be interpreted as an unconditional moment. When allowing for breaks in the volatility, as we do in Section 4.5, $\text{Var}(\Delta\phi_{j,t})$ then becomes the conditional variance for $\Delta\phi_{j,t}$.

between demand shocks and the coefficient innovations that the model does not identify.²⁸ As in Section 4.3.1, our results are again conceptually in line with Cogley, Primiceri, and Sargent (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.

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 (26), Figure 4 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).

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

²⁸The results in Figure 3 denote medians of the sampling distribution for the statistics in (26) and (27). To be clear, at each MCMC iteration we compute and retain the values for (26) and (27), which in turn gives us a distribution for these two metrics. Results in Figure 3 are thus medians based on such distributions.

hawkish’) stance.

4.4.1 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 economy. From a theoretical perspective, [Gust, Herbst, Lopez-Salido, and Smith \(2017\)](#) 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, Laforte, Reifschneider, and J. \(2012\)](#) 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, Gali, and Gambetti \(2019\)](#), 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.

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 captures 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%. Figure 5 illustrates the path of the policy rate under these two cases.

To construct the scenarios described above, we proceed as follows. First, we estimate the two versions of the TVP-VAR models, exogenous and endogenous, with data from 1962Q1 until 2007Q1. Then, at 2007Q2 we compute impulse responses for unemployment and inflation using two sequences of monetary policy shocks that – in line with the description of the scenarios above – differ only after the fifth-quarter-ahead horizon.²⁹

Two features deserve comment. First, the meetings of the Federal Open Market Committee (FOMC) typically occur eight times a year, hence our scenarios correspond to quarterly approximations of the dates when decisions about the policy rate were actually made. Second, changes in the policy rate embed both an anticipated and an unanticipated component. Therefore, to obtain a path for the unanticipated component, i.e. our sequence of policy shocks, we take a simple approach and assume that monetary policy surprises during the crisis accounted for 25% of each actual change in the policy rate.³⁰

Admittedly, fixing a predetermined share provides us with only a crude approximation for the unanticipated component embedded in interest rate changes. Nevertheless, such an approach can still be useful for scenario analysis. For example, given the heightened state of concerns about liquidity and contagion risks that dominated the US banking sector around 2007Q2 (and in the ensuing years), it is plausible to posit that market participants largely anticipated back then that the Fed would shift towards a more accommodative stance. In this sense, setting the unanticipated share of interest rate changes to 25% is just a simple strategy to incorporate this idea that monetary policy easing was somewhat predictable in the lead up to (and during the early stages of) the financial crisis. Moreover, any mismeasurement this approximation approach could bring is equally applied to both scenarios and models. Therefore, comparison of the impulse responses across scenarios and models is still informative in a qualitative sense.

Chart B of Figure 5 plots the response of the unemployment rate obtained with the exogenous TVP-VAR model under the ‘dovish’ and ‘less dovish’ scenarios for monetary

²⁹A relatively similar approach was adopted in [Leeper and Zha \(2003\)](#), to evaluate linear projections conditional on hypothetical paths of monetary policy.

³⁰The impact matrix for both endogenous and exogenous TVP-VARs is estimated using the same sign restrictions discussed Section 4.1

policy. Both responses coincide until the five-quarter-ahead horizon, which is unsurprising, since the sequences of shocks in both scenarios are the same up to this point. 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 5 reports equivalent scenarios for unemployment, except now dynamic 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 macroeconomic 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.

4.4.2 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 Normalization 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 4.4.1](#). More precisely – constructing the unanticipated component embedded in interest rate changes as in [Section 4.4.1](#) – 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 2018Q4.³¹ 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 6 plot the dynamic 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 6 show the response of inflation to monetary policy shocks under exogenous and endogenous time variation, respectively. As in the scenarios for Section 4.4.1, 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.

4.5 Accounting for Changes in the Volatility

In addition to changes in the intercept and autoregressive coefficients, TVP-VARs also often accommodate changes in the variance of the error terms. Therefore, even though our methodological contribution is about endogenizing first-moment parameters, a natural

³¹In 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 – as national accounts data become available following COVID-19 – it could be interesting to apply our methodology to investigate macroeconomic instabilities associated with such a global shock.

question is whether our previous empirical results are robust to heteroskedasticity.

To tackle this issue, we relax the assumption of homoskedasticity and re-conduct the analysis in Sections 4.2, 4.3 and 4.4 while allowing for: (i) breaks in the covariance matrix for the reduced form innovations; and (ii) breaks in the conditional variance of the drifting coefficients. For the latter, we introduce breaks in the loading coefficients associated with each (unit-variance) structural shock and in the variance of the coefficient-specific errors.

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.³² Notably, the last break date 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.

Certainly, another way to incorporate heteroskedasticity would be to follow [Primiceri \(2005\)](#) and let the conditional variances and covariances in the system to evolve as drifting coefficients. Since the focus in this paper is on endogenizing first-moment parameters, we leave such extensions – which would considerably increase the computational cost of estimation – for future work. Note, however, that our approach to model second-moment changes is not too distinct from [Primiceri \(2005\)](#), at least conceptually. More precisely, like [Primiceri \(2005\)](#), we allow all parameters in the covariance matrix in the measurement equation to change. We just do so in a more parsimonious and, admittedly, less flexible fashion. Also, the similarity between some of our results and [Cogley, Primiceri, and Sargent \(2010\)](#) – who use an approach to model second-moments that is much closer to [Primiceri \(2005\)](#) – suggests that our strategy to accommodate heteroskedasticity seems adequate for the purpose of this paper.

In summary, the main insights for homoskedastic endogenous-TVP-VARs carry over to

³²For details, see the discussion in Section 5 of their paper.

their heteroskedastic counterparts. Therefore, in the interest of brevity, we report most of our robustness checks in the Online Appendix. Nonetheless, a couple of noteworthy results emerged when controlling for volatility breaks. Figure 7 plots – for the four volatility regimes discussed above – the share of total coefficient variation that is due to the identified and unidentified innovations in the endogenous TVP-VAR setting.³³ As in Section 4.3.2, cost-push and monetary policy shocks remain the most important drivers behind coefficient variations. Nevertheless, the analysis here presents two nuances.

First, the share of coefficient variation due to cost-push shocks became considerably higher (approximately 60%) from 1980 onwards relative to its value in the 1970s (around 40%). Such a result aligns well with recent empirical evidence based on micro-level data – as in De Loecker and Eeckhout (2017) – that suggests firms in the US have been experiencing an increase in markups since the early 1980s.³⁴ Second, Chart B shows that monetary policy shocks contributed less to parameter changes during the 1979-1987 period. This possibly reflects the three-year period from 1979 until 1982 during the Volcker administration when the Fed temporarily targeted monetary aggregates instead of the federal funds rate. Recall that the latter is the policy instrument used in our TVP-VARs.

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.

³³To be clear, each chart in Figure 7 is computed using expressions (26) and (27) presented in Section 4.3.2, except now differences across charts reflect, as mentioned previously, the fact that we allow for breaks in the loading coefficients associated with each structural shock and in the variance of the coefficient-specific residuals.

³⁴Berry, Gaynor, and Morton (2019) argue that candidate drivers for such expanding markups include a rise in fixed costs, network effects, monopsony effects in labor markets, an increase in rent-seeking behavior and globalization effects.

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 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.

Tables and Figures

Table 1: Summary of sign restrictions

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

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

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

	Model II	Model III	Model IV
λ	19.81	19.83	48.84

Note: Model comparison results are based on computing $2\log(\text{Bayes factor})$ 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 $\lambda = (\lambda_{C,1}, \lambda_{C,2}, \lambda_{C,3}, \lambda_{L,1}, \lambda_{L,2}, \lambda_{L,3})'$. For all the others, $\lambda = (\lambda_{j,1}, \lambda_{j,2}, \lambda_{j,3})'$ for $j = C, L$. Exogenous TVP-VARs assume $\lambda = \mathbf{0}$.

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

	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 $2 \log(\text{Bayes factor})$ 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 $i = 1, 2, 3$.

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

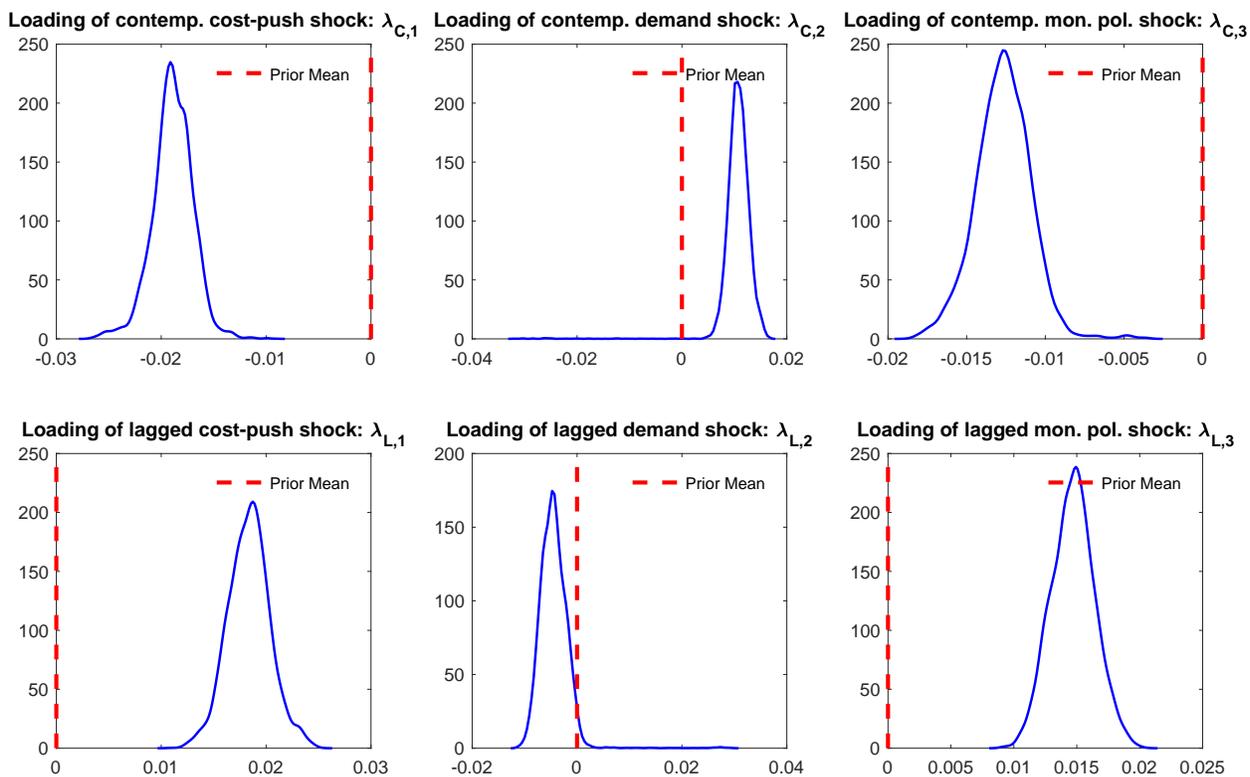
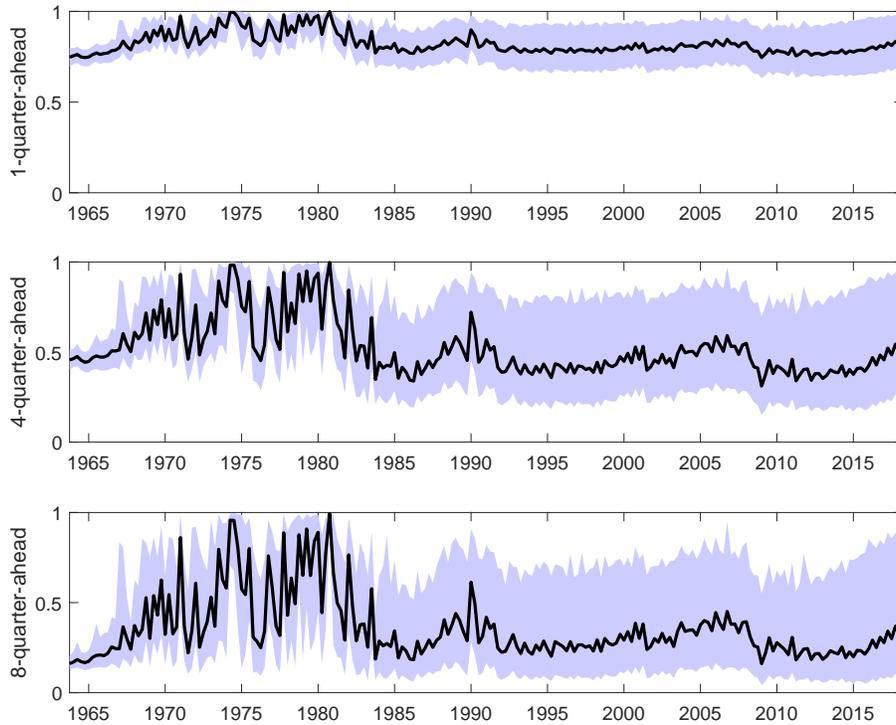
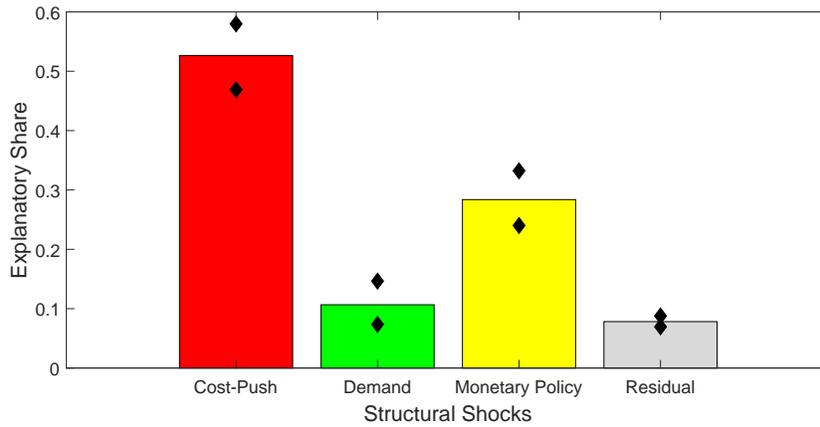


Figure 2: Inflation-gap persistence based on $R_{h,t}^2$ statistics



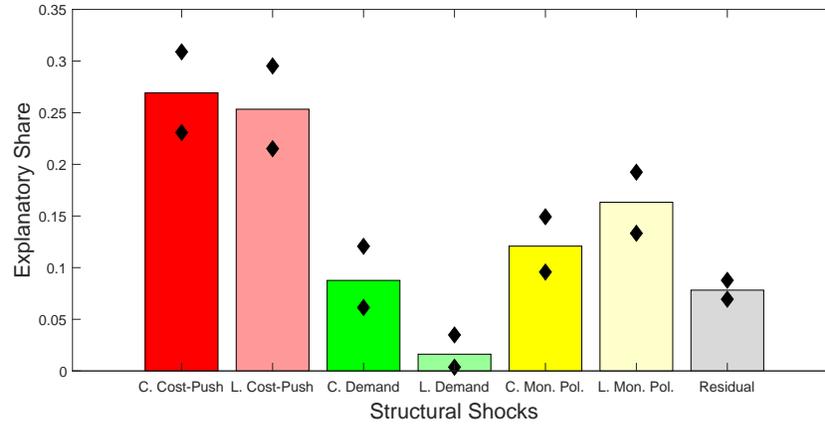
Note: The figure plots the time-varying measure of inflation-gap persistence, $R_{h,t}^2$, for the one-, four- and eight-quarter-ahead forecasting horizons. $R_{h,t}^2$ is constructed using Equation 12 in [Cogley, Primiceri, and Sargent \(2010\)](#).

Figure 3: Contribution of structural shocks to parameter instability



Note: The figure shows the contributions of the structural shocks (cost-push, demand and monetary policy), along with the contribution of the exogenous shocks (residual), to the overall changes in the drifting VAR coefficients. Results are based on Equations (26) and (27). Black dots indicate the 16th and 84th percentiles for posterior densities associated with such contributions.

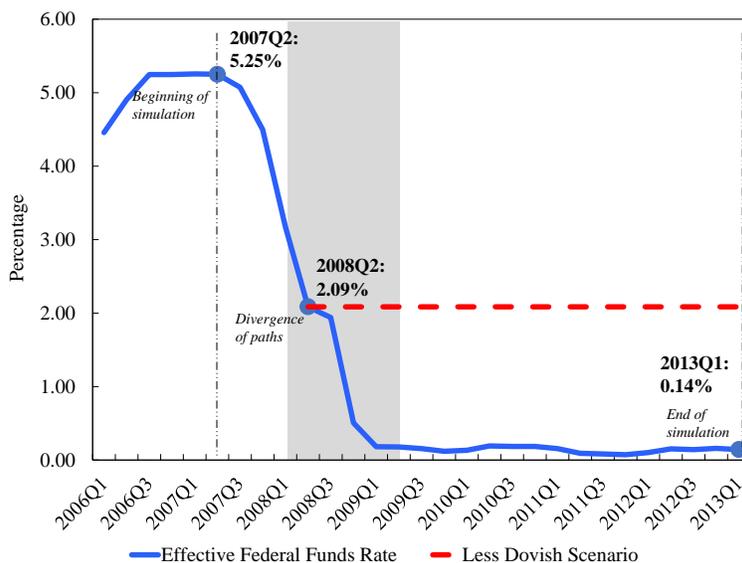
Figure 4: Contribution of contemporaneous and lagged structural shocks to parameter instability



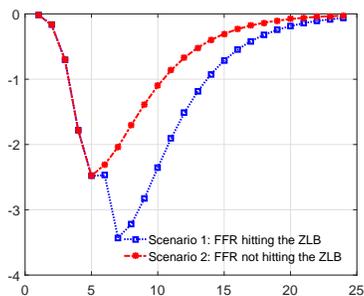
Note: The figure shows the contributions contemporaneous (C.) and lagged (L.) structural shocks (cost-push, demand and monetary policy), along with the contribution of exogenous shocks (residual), to the overall changes in the drifting VAR coefficients. Results are based on Equations (26) and (27). Black dots indicate the 16th and 84th percentiles for posterior densities associated with such contributions.

Figure 5: Counterfactual scenario for monetary policy in 2007: the trajectory towards the zero lower bound

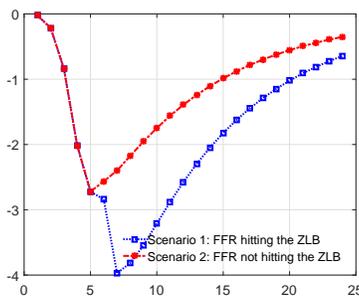
(a) Alternative Paths of Monetary Policy Stance



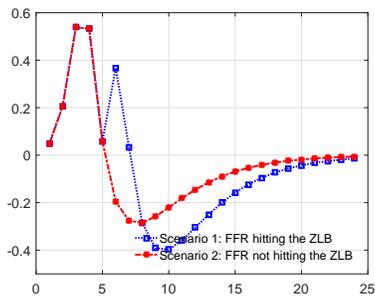
(b) Response of Unemployment: Exogenous



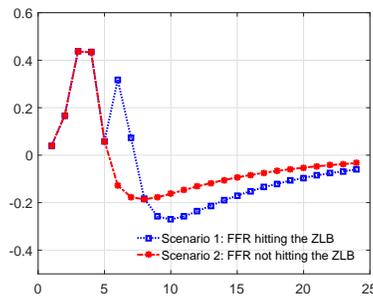
(c) Response of Unemployment: Endogenous



(d) Response of Inflation: Exogenous



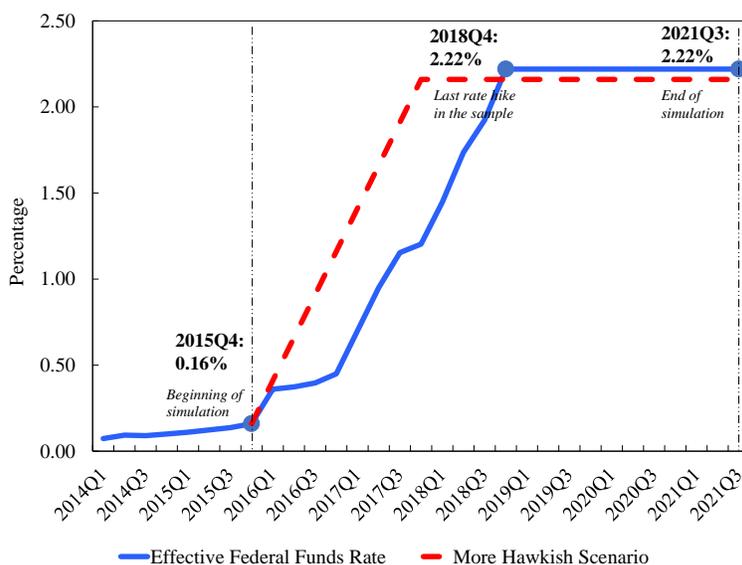
(e) Response of Inflation: Endogenous



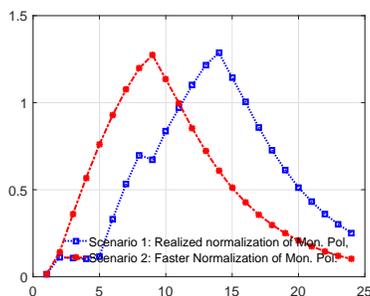
Note: Chart (a) plots the actual path of the policy rate (blue line) and the alternative non-ZLB path of the policy rate (red line). Charts (b) and (c) plot the response of unemployment rate, for both scenarios, obtained with a TVP-VAR with exogenous and endogenous time-varying coefficients, respectively. Charts (d) and (e) plot the response of inflation, for both scenarios, obtained with a TVP-VAR with exogenous and endogenous time-varying coefficients, respectively.

Figure 6: Counterfactual scenario for monetary policy in 2015: interest rate normalization

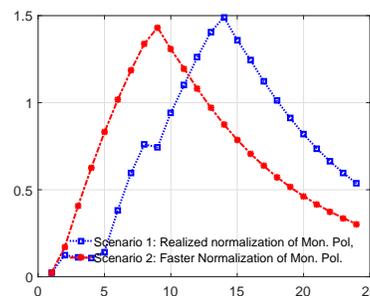
(a) Alternative Paths of Monetary Policy Stance



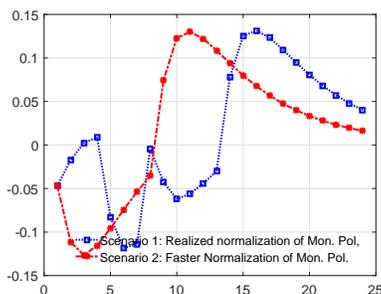
(b) Response of Unemployment: Exogenous



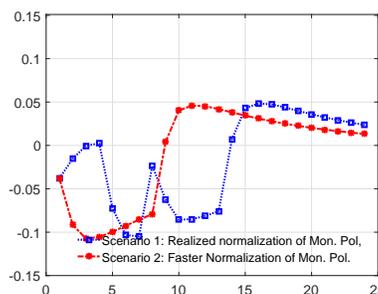
(c) Response of Unemployment: Endogenous



(d) Response of Inflation: Exogenous

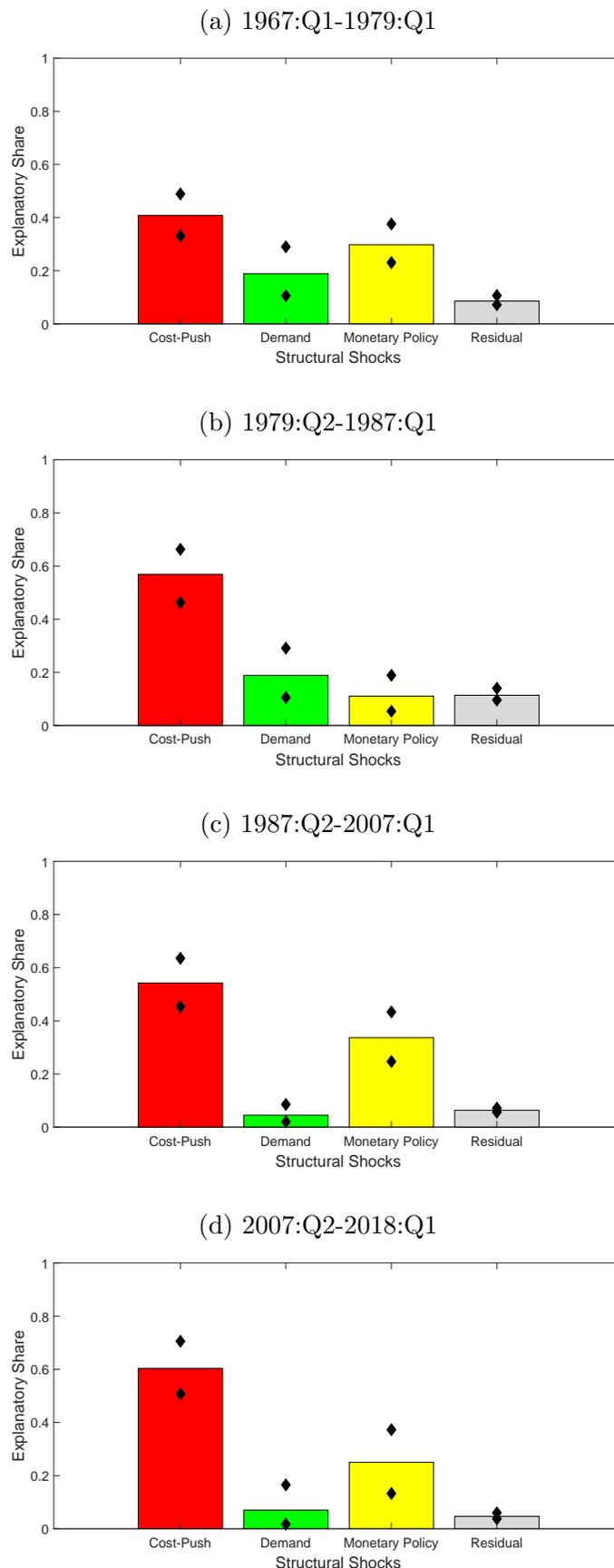


(e) Response of Inflation: Endogenous



Note: Chart (a) plots the actual path of the policy rate normalization (blue line) and the path of a faster normalization of the policy rate (red line). Charts (b) and (c) plot the response of unemployment rate, for both scenarios, obtained with a TVP-VAR with exogenous and endogenous time-varying coefficients, respectively. Charts (d) and (e) plot the response of inflation, for both scenarios, obtained with a TVP-VAR with exogenous and endogenous time-varying coefficients, respectively.

Figure 7: Contribution of structural shocks to parameter instability (volatility breaks)



Note: The figure shows the contributions of the structural shocks, along with the contribution of the exogenous shocks, to the overall changes in the drifting VAR coefficients. Results are based on Equations (26) and (27). Black dots indicate the 16th and 84th percentiles for posterior densities associated with such contributions.

References

- ANDERSON, H. M., AND F. VAHID (1998): “Testing multiple equation systems for common nonlinear components,” *Journal of Econometrics*, 84(1), 1–36. [1](#)
- ANDOLFATTO, D., AND A. SPEWAK (2019): “Understanding Lowflation,” *Federal Reserve of St. Louis Review*, 101(1), 1–26. [4.3.1](#)
- AUERBACH, A. J., AND Y. GORODNICHENKO (2012): “Measuring the output responses to fiscal policy,” *American Economic Journal: Economic Policy*, 4(2), 1–27. [1](#)
- BAUMEISTER, C., AND G. PEERSMAN (2013): “Time-varying effects of oil supply shocks on the US economy,” *American Economic Journal: Macroeconomics*, 5(4), 1–28. [1](#)
- BERRY, S., M. GAYNOR, AND F. MORTON (2019): “Do increasing markups matter? Lessons from empirical industrial organization,” *Journal of Economic Perspectives*, 33(3), 44–68. [34](#)
- BIANCHI, F., AND A. CIVELLI (2015): “Globalization and inflation: Evidence from a time-varying VAR,” *Review of Economic Dynamics*, 18(2), 406–433. [1](#)
- CANOVA, F. (1993): “Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model,” *Journal of Economic Dynamics & Control*, 17, 233–261. [1](#), [7](#)
- (2011): *Methods for Applied Macroeconomic Research*. Princeton University Press. [4.1](#)
- CANOVA, F., F. FERRONI, AND C. MATTHES (2015): “Approximating time varying structural models with time invariant structures,” *Banque de France Working Paper*, (578). [7](#)
- CANOVA, F., AND L. GAMBETTI (2009): “Structural changes in the US economy: Is there a role for monetary policy?,” *Journal of Economic Dynamics & Control*, 33, 477–490. [2.1](#)
- CARRIERO, A., T. E. CLARK, AND M. G. MARCELLINO (2018): “Endogenous Uncertainty,” *FRB of Cleveland Working Paper*, (18-05). [1](#)
- CHAN, J. C. (2013): “Moving average stochastic volatility models with application to inflation forecast,” *Journal of Econometrics*, 176(2), 162–172. [3](#)
- CHAN, J. C., AND I. JELIAZKOV (2009): “Efficient simulation and integrated likelihood estimation in state space models,” *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2), 101–120. [1](#), [3](#), [3.1](#)
- CLARIDA, R., J. GALI, AND M. GERTLER (1999): “The science of monetary policy: a new Keynesian perspective,” *Journal of economic literature*, 37(4), 1661–1707. [4.1](#)

- CLARK, T., AND S. TERRY (2010): “Time variation in the inflation passthrough of energy prices,” *Journal of Money, Credit and Banking*, 42(7), 1419–1433. [1](#)
- COGLEY, T., C. MATTHES, AND A. M. SBORDONE (2015): “Optimized Taylor rules for disinflation when agents are learning,” *Journal of Monetary Economics*, 72, 131–147. [1](#)
- COGLEY, T., G. PRIMICERI, AND T. SARGENT (2010): “Inflation-gap persistence in the US,” *American Economic Journal: Macroeconomics*, 2(1), 43–69. [1](#), [4](#), [4.3](#), [4.3.1](#), [24](#), [4.3.2](#), [4.5](#), [2](#)
- COGLEY, T., AND T. SARGENT (2001): “Evolving Post-World War II U.S. Inflation Dynamics,” *NBER Macroeconomics Annual*, pp. 331–373. [1](#), [4](#)
- COGLEY, T., AND T. J. SARGENT (2005): “Drifts and volatilities: Monetary policies and outcomes in the post WWII US,” *Review of Economic Dynamics*, 8(2), 262–302. [1](#), [2.2](#), [15](#)
- COOPER, R. W., AND J. C. HALTIWANGER (2006): “On the nature of capital adjustment costs,” *The Review of Economic Studies*, 73(3), 611–633. [10](#)
- DE LOECKER, J., AND J. EECKHOUT (2017): “The rise of market power and the macroeconomic implications,” *NBER Working Paper*, 23687. [4.5](#)
- DEBORTOLI, D., J. GALI, AND L. GAMBETTI (2019): “On the empirical (ir)relevance of the zero lower bound constraint,” in *NBER Macroeconomics Annual*, p. Forthcoming. MIT press. [4.4.1](#)
- DOAN, T., R. LITTERMAN, AND C. SIMS (1984): “Forecasting and conditional projection using realistic prior distributions,” *Econometric reviews*, 3(1), 1–100. [1](#)
- DURBIN, J., AND S. J. KOOPMAN (2012): *Time Series Analysis by State Space Models*. Oxford University Press. [14](#)
- FELDSTEIN, M. (2017): “Normalizing monetary policy,” Remarks by Martin Feldstein at the Cato Institute’s 35th Annual Monetary Conference. [4.4.2](#)
- FRIEDMAN, M. (1961): “The lag in effect of monetary policy,” *Journal of Political Economy*, 69(5), 447–466. [4.3.2](#)
- FRY, R., AND A. PAGAN (2011): “Sign restrictions in structural vector autoregressions: A critical review,” *Journal of Economic Literature*, 49(4), 938–60. [4.1](#)
- GALI, J., AND L. GAMBETTI (2015): “The effects of monetary policy on stock market bubbles: some evidence,” *American Economic Journal: Macroeconomics*, 7(1), 233–257. [1](#)

- GALVÃO, A. B. C. (2006): “Structural break threshold VARs for predicting US recessions using the spread,” *Journal of Applied Econometrics*, 21(4), 463–487. [1](#)
- GUST, C., E. HERBST, D. LOPEZ-SALIDO, AND M. SMITH (2017): “The empirical implications of the interest-rate lower bound,” *American Economic Review*, 107(7), 1971–2006. [4.4.1](#)
- HAVRANEK, T., AND M. RUSNAK (2013): “Transmission lags of monetary policy: A Meta-Analysis,” *International Journal of Central Banking*, 9(4), 39–76. [4.3.2](#)
- HESS, C., J. LAFORTE, D. REIFSCHNEIDER, AND W. J. (2012): “Have we underestimated the likelihood and severity of zero lower bound events?,” *Journal of Money, Credit and Banking*, 44(1), 47–82. [4.4.1](#)
- KANG, K. (2014): “Estimation of state-space models with endogenous Markov regime-switching parameters,” *Econometrics Journal*, 17, 56–82. [1](#)
- KASS, R. E., AND A. E. RAFTERY (1995): “Bayes factors,” *Journal of the American Statistical Association*, 90(430), 773–795. [4.2](#), [22](#), [23](#), [3](#), [4](#)
- KIM, C., J. PIGER, AND R. STARTZ (2008): “Estimation of Markov regime-switching regression models with endogenous switching,” *Journal of Econometrics*, 143, 263–273. [1](#)
- KOOP, G., R. LEON-GONZALEZ, AND R. STRACHAN (2009): “On the evolution of the monetary policy transmission mechanism,” *Journal of Economic Dynamics & Control*, 33, 997–1017. [2.1](#)
- KOOP, G., D. J. POIRIER, AND J. L. TOBIAS (2007): *Bayesian Econometric Methods*. Cambridge University Press. [3.2](#)
- KOSE, M. A., C. OTROK, AND C. WHITEMAN (2003): “International Business Cycles: World, Region and Country-Specific Factors,” *American Economic Review*, 93(4), 1216–1239. [4.3.2](#)
- KROESE, D. P., AND J. C. CHAN (2014): *Statistical Modeling and Computation*. Springer. [3.1](#)
- KYDLAND, F. E., AND E. C. PRESCOTT (1982): “Time to build and aggregate fluctuations,” *Econometrica: Journal of the Econometric Society*, pp. 1345–1370. [10](#)
- LEEPER, E., AND T. ZHA (2003): “Modest Policy Interventions,” *Journal of Monetary Economics*, 50(8), 1673–1700. [29](#)
- LUBIK, T. A., AND F. SCHORFHEIDE (2004): “Testing for indeterminacy: An application to US monetary policy,” *American Economic Review*, 94(1), 190–217. [4.1](#)
- LUCAS, R. (1976): “Econometric policy evaluation: A critique,” *Carnegie-Rochester Conference Series on Public Policy*, 1, 19–46. [5](#)

- MCCAUSLAND, W. J., S. MILLER, AND D. PELLETIER (2011): “Simulation smoothing for state-space models: A computational efficiency analysis,” *Computational Statistics & Data Analysis*, 55(1), 199–212. [13](#), [14](#)
- MUMTAZ, H., AND P. SURICO (2009): “Time-varying yield curve dynamics and monetary policy,” *Journal of Applied Econometrics*, 24, 895–913. [1](#)
- MUMTAZ, H., AND K. THEODORIDIS (2019): “Dynamic effects of monetary policy shocks on macroeconomic volatility,” *Journal of Monetary Economics*, Forthcoming. [1](#)
- PAUL, P. (2019): “The time-varying effect of monetary policy on asset prices,” *Review of Economics and Statistics*, Forthcoming. [1](#)
- 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. [4.4.2](#)
- PRIMICERI, G. (2005): “Time Varying Structural Vector Autoregressions and Monetary Policy,” *Review of Economic Studies*, 72, 821–852. [1](#), [2.1](#), [15](#), [4.5](#)
- RUBIO-RAMIREZ, J. F., D. F. WAGGONER, AND T. ZHA (2010): “Structural vector autoregressions: Theory of identification and algorithms for inference,” *The Review of Economic Studies*, 77(2), 665–696. [4.1](#)
- SIMS, C. (1993): “A Nine-Variable Probabilistic Macroeconomic Forecasting Model,” *Business Cycles, Indicators and Forecasting, NBER Studies in Business Cycles*, 18, 179–214. [1](#)
- SIMS, C., AND T. ZHA (2006): “Were there regime switches in the U.S. monetary policy?,” *American Economic Review*, 96(1), 54–81. [2.1](#), [4.5](#)
- SMETS, F., AND R. WOUTERS (2007): “Shocks and frictions in US business cycles: A Bayesian DSGE approach,” *American Economic Review*, 97(3), 586–606. [10](#)
- STOCK, J., AND M. WATSON (1996): “Evidence on Structural Instability in Macroeconomic Time Series Relations,” *Journal of Business & Economic Statistics*, 14(1), 11–30. [1](#)
- SWANSON, E. (2018): “The Federal Reserve is not very constrained by the lower bound on nominal interest rates,” *Brookings Papers on Economic Activity*, Fall. [4.4.1](#)
- TSAY, R. S. (1998): “Testing and modeling multivariate threshold models,” *Journal of the American Statistical Association*, 93(443), 1188–1202. [1](#)

- 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. [1](#), [4.1](#)
- VERDINELLI, I., AND L. WASSERMAN (1995): “Computing Bayes factors using a generalization of the Savage-Dickey density ratio,” *Journal of the American Statistical Association*, 90(430), 614–618. [1](#), [3.2](#)
- 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. [4.4.2](#)
- WU, J., AND F. XIA (2016): “Measuring the macroeconomic impact of monetary policy at the zero lower bound,” *Journal of Money, Credit and Banking*, 48(2-3), 253–291. [4](#)

Online Appendix - Not for Publication

Appendix A: 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 with one minor difference. We use the notation $\tilde{\mathbf{e}}$ to denote the $NT \times 1$ vector of sign-identified shocks, i.e. pre-multiplied by some given rotation matrix, as discussed in Section 4.1, i.e.:

$$\mathbf{y} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\phi}} + \mathbf{L}_A\tilde{\mathbf{e}}, \quad (1)$$

$$\tilde{\boldsymbol{\phi}} = \tilde{\boldsymbol{\phi}}_0 + \mathbf{L}_\lambda\tilde{\mathbf{e}} + \mathbf{v}, \quad (2)$$

$$\begin{bmatrix} \tilde{\mathbf{e}} \\ \mathbf{v} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_{NT} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_v \end{bmatrix} \right), \quad (3)$$

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.

Priors

We assume standard independent priors for the four blocks of model parameters, namely $\boldsymbol{\lambda}$, $\boldsymbol{\phi}_0$, $\Omega_u = A'A$ and $\Omega_v = \text{diag}(\sigma_{v,1}^2, \dots, \sigma_{v,M}^2)$, where, recall, M denotes the number of drifting coefficients. More specifically, we have:

$$\boldsymbol{\lambda} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\lambda),$$

$$\boldsymbol{\phi}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{\phi_0}, \boldsymbol{\Sigma}_{\phi_0}),$$

$$\Omega_u \sim \mathcal{IW}(\nu_u, \mathbf{V}_u),$$

$$\sigma_{v,i}^2 \sim \mathcal{IG}(\nu_{v,i}, S_{v,i}) \text{ for } i = 1, \dots, M.$$

We set $\Sigma_{\lambda} = \text{diag}(10^{-3}, \dots, 10^{-3})$, $\Sigma_{\phi_0} = \text{diag}(10^{-5}, \dots, 10^{-5})$ and μ_{ϕ_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 $i = 1, \dots, M$. Overall, this class of priors is standard in TVP-VAR analysis. In particular, the prior density for λ 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 λ could be prior – rather than data – induced.

Posterior Sampler

Let $\mathbf{s} = \{\phi, \lambda, \phi_0, \Omega_u = A'A, \Omega_v\}$ denote the set of states and parameters for the system in (1)-(3), where notation \mathbf{s}_{-j} represents all elements in \mathbf{s} except for j . An MCMC algorithm for the endogenous TVP-VARs introduced in Section 2 entails sequentially sampling from the following five conditional posterior distributions:

- (1) $f(\phi|\mathbf{y}, \mathbf{s}_{-\phi})$,
- (2) $f(\lambda|\mathbf{y}, \mathbf{s}_{-\lambda})$,
- (3) $f(\phi_0|\mathbf{y}, \mathbf{s}_{-\phi_0})$,
- (4) $f(\Omega_u|\mathbf{y}, \mathbf{s}_{-\phi})$,
- (5) $f(\Omega_v|\mathbf{y}, \mathbf{s}_{-\Omega_v})$.

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

- *Sampling λ*

Recall from Section 3.2 that we can express the state equation in (2) as:

$$\tilde{\phi} = \tilde{\phi}_0 + \mathbf{L}_{\tilde{\epsilon}}\lambda + \mathbf{v}, \tag{4}$$

where:

$$\mathbf{L}_{\tilde{\mathbf{e}}} = \begin{pmatrix} \begin{bmatrix} \tilde{e}'_1 & 0' \\ \tilde{e}'_2 & \tilde{e}'_2 \\ \vdots & \vdots \\ \tilde{e}'_T & \tilde{e}'_{T-1} \end{bmatrix} \end{pmatrix}.$$

Since – when sampling $\boldsymbol{\lambda} - \tilde{\boldsymbol{\phi}}$ and A are given, we can back out $\tilde{\mathbf{e}}$ from (1) by simply computing $\tilde{\mathbf{e}} = \mathbf{L}_{\mathbf{A}}^{-1}(\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\phi}})$. As a result, by virtue of the modular nature of MCMC algorithms, $\mathbf{L}_{\tilde{\mathbf{e}}}$ can be treated as a matrix of predetermined regressors. Given that, $\mathbf{v} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_v)$, then by standard regression results, we have:

$$\boldsymbol{\lambda} | \mathbf{y}, \mathbf{s}_{-\boldsymbol{\lambda}} \sim \mathcal{N}(\bar{\mathbf{d}}_{\boldsymbol{\lambda}}, \bar{\mathbf{D}}_{\boldsymbol{\lambda}}), \text{ where } \begin{cases} \bar{\mathbf{d}}_{\boldsymbol{\lambda}} = \bar{\mathbf{D}}_{\boldsymbol{\lambda}} \mathbf{L}'_{\tilde{\mathbf{e}}} \boldsymbol{\Sigma}_v^{-1} (\tilde{\boldsymbol{\phi}} - \tilde{\boldsymbol{\phi}}_0), \\ \bar{\mathbf{D}}_{\boldsymbol{\lambda}} = (\mathbf{L}'_{\tilde{\mathbf{e}}} \boldsymbol{\Sigma}_v^{-1} \mathbf{L}_{\tilde{\mathbf{e}}} + \boldsymbol{\Sigma}_{\boldsymbol{\lambda}}^{-1})^{-1}. \end{cases} \quad (5)$$

- *Sampling $\boldsymbol{\phi}_0$*

Let $\mathbf{L}_0 = \iota_0 \otimes I_M$, where (again) M denotes the number of drifting coefficients and $\iota_0 = (1, 0, \dots, 0)'$ is a $T \times 1$ vector. Thus, (4) can be recast as:

$$\tilde{\boldsymbol{\phi}} = \mathbf{L}_0 \boldsymbol{\phi}_0 + \mathbf{L}_{\tilde{\mathbf{e}}} \boldsymbol{\lambda} + \mathbf{v}. \quad (6)$$

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

$$\boldsymbol{\phi}_0 | \mathbf{y}, \mathbf{s}_{-\boldsymbol{\phi}_0} \sim \mathcal{N}(\bar{\mathbf{d}}_{\boldsymbol{\phi}_0}, \bar{\mathbf{D}}_{\boldsymbol{\phi}_0}), \text{ where } \begin{cases} \bar{\mathbf{d}}_{\boldsymbol{\phi}_0} = \bar{\mathbf{D}}_{\boldsymbol{\phi}_0} (\mathbf{L}'_0 \boldsymbol{\Sigma}_v^{-1} (\tilde{\boldsymbol{\phi}} - \mathbf{L}_{\tilde{\mathbf{e}}} \boldsymbol{\lambda}) + \boldsymbol{\Sigma}_{\boldsymbol{\phi}_0}^{-1} \boldsymbol{\mu}_{\boldsymbol{\phi}_0}), \\ \bar{\mathbf{D}}_{\boldsymbol{\phi}_0} = (\mathbf{L}'_0 \boldsymbol{\Sigma}_v^{-1} \mathbf{L}_0 + \boldsymbol{\Sigma}_{\boldsymbol{\phi}_0}^{-1})^{-1}. \end{cases} \quad (7)$$

- *Sampling Ω_u*

Note that by substituting (4) into (1), we can rewrite the measurement equation as follows:

$$\mathbf{y} = \tilde{\mathbf{X}} \tilde{\boldsymbol{\phi}}_0 + \tilde{\mathbf{X}} \tilde{\boldsymbol{\eta}} + \mathbf{u}, \quad (8)$$

where $\mathbf{u} = \mathbf{L}_A \tilde{\mathbf{e}} \sim \mathcal{N}(\mathbf{0}, \mathbf{L}_A \mathbf{L}'_A)$, such that $\mathbf{L}_A \mathbf{L}'_A = \text{diag}(\Omega_u, \dots, \Omega_u)$, and $\tilde{\boldsymbol{\eta}} = \mathbf{L}_{\tilde{\mathbf{e}}} \boldsymbol{\lambda} + \mathbf{v} = \tilde{\boldsymbol{\phi}} - \tilde{\boldsymbol{\phi}}_0$. Since the latter term ($\tilde{\boldsymbol{\phi}} - \tilde{\boldsymbol{\phi}}_0$) is known when sampling Ω_u , we can (again) apply the modularity of MCMC algorithms to treat $\tilde{\boldsymbol{\eta}}$ as a predetermined regressor in (8). Therefore, given we elicit an inverse-Wishart prior for Ω_u , standard regression results yield:

$$\Omega_u | \mathbf{y}, \mathbf{s}_{-\Omega_u} \sim \mathcal{IW}(\bar{\nu}_u, \bar{\mathbf{V}}_u), \text{ where } \begin{cases} \bar{\nu}_u = T + \nu_u, \\ \bar{\mathbf{V}}_u = \sum_{t=1}^T u_t u'_t + \mathbf{V}_u, \end{cases} \quad (9)$$

where u_t is a $N \times 1$ element in $\mathbf{u} = (u_1, u_2, \dots, u_T)$. Given an MCMC draw for Ω_u , we back out A via Cholesky factorization of the former, i.e. $\Omega_u = AA'$.¹

- *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:

$$\sigma_{v,i}^2 | \mathbf{y}, \mathbf{s}_{-\sigma_{v,i}} \sim \mathcal{IG}(\bar{\nu}_{v,i}, \bar{S}_{v,i}), \text{ where } \begin{cases} \bar{\nu}_{v,i} = \frac{T}{2} + \nu_{v,i}, \\ \bar{S}_{v,i} = \frac{\sum_{t=1}^T v_{t,i}^2}{2} + S_{v,i} \quad \text{for } i = 1, \dots, M. \end{cases} \quad (10)$$

- *Posterior Simulation for Exogenous TVP-VARs and Models with Volatility Breaks*

Our MCMC algorithm for exogenous TVP-VARs follows very similar steps to the ones described above. The only differences regard step 1 (as discussed in Section 3.1) and step 2. In particular, the latter is not required when estimating exogenous TVP-VARs since $\boldsymbol{\lambda}$ is not present in such models. As for (endogenous) models with second-moment breaks, the steps are exactly the same, except that steps 2, 4 and 5 need to be conducted for each of the four subsamples that result from the three volatility breaks discussed in Section 4.5.

¹Recall from the discussion in Section 4.1 that the rotation matrix (P) applied to achieve sign identification does not affect the distribution of the reduced form innovations. In particular, we have: $\Omega_u = \mathbb{E}(u_t u'_t) = \mathbb{E}(A P e_t e'_t P' A') = A P \mathbb{E}(e_t e'_t) P' A' = A P P' A' = A A'$.

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:

$$\hat{f}(\boldsymbol{\lambda} = \mathbf{0}|\mathbf{y}) = \frac{1}{R} \sum_{r=1}^R f(\boldsymbol{\lambda} = \mathbf{0}|\mathbf{y}, \boldsymbol{\phi}^{(r)}, \boldsymbol{\phi}_0^{(r)}, \Sigma_u^{(r)}, \Sigma_v^{(r)}),$$

where $(\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}_0^{(1)}, \Sigma_u^{(1)}, \Sigma_v^{(1)}), \dots, (\boldsymbol{\phi}^{(R)}, \boldsymbol{\phi}_0^{(R)}, \Sigma_u^{(R)}, \Sigma_v^{(R)})$ are (post-burn-in) MCMC draws. Note that the density inside the summation above can be evaluated exactly using the results in (5).

Inefficiency Factors

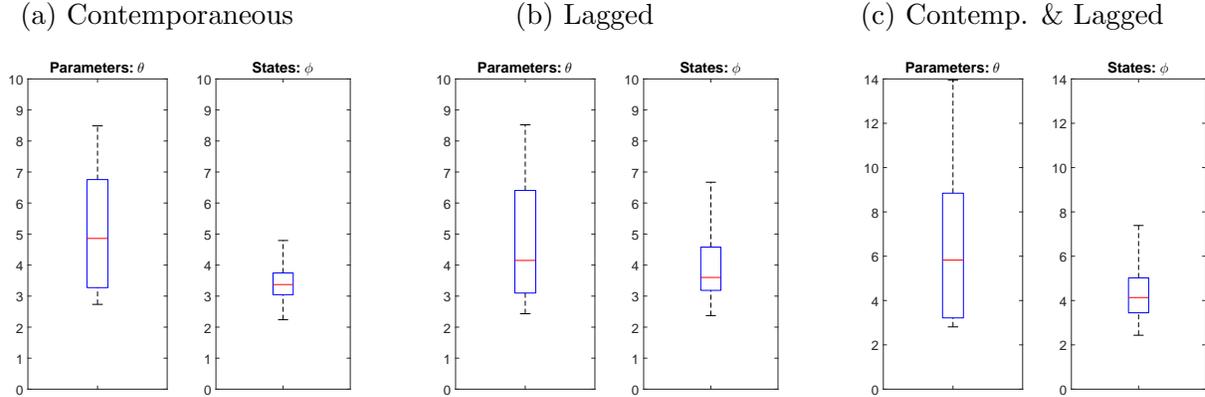
We report inefficiency factors of the posterior draws for all states and parameters using a common metric (see, e.g., Chib (2001)) given by:

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

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.² Figure 1 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 1 demonstrate that our posterior sampler exhibits good mixing properties.

²Another 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.

Figure 1: Inefficiency factors



Note: The figure shows the boxplots associated with the inefficiency factors corresponding to the posterior draws of parameters θ and states, time-varying coefficients, ϕ 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.

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 ϕ 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

³In other words, to the extent that large shocks induce big variations in policy and agents’ decisions – as

said, below we show that such stability related concerns are not prevalent for our empirical applications.

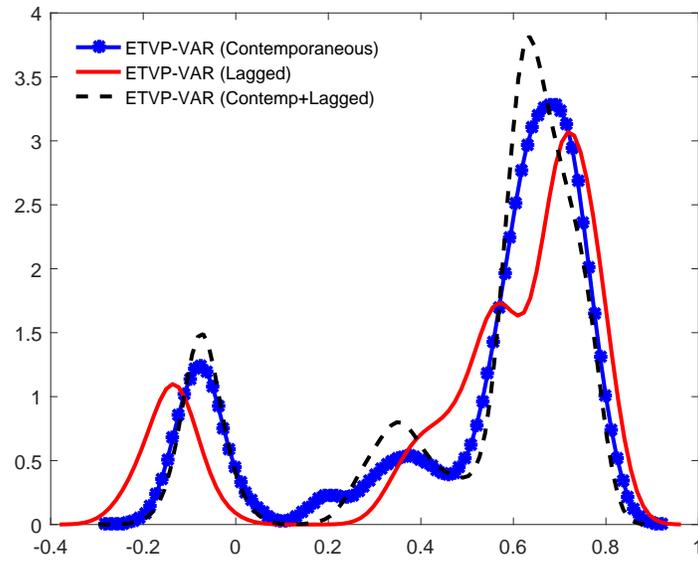
Certainly, one alternative to address stability issues could be to adopt a single-move MCMC sampler. In this case, instead of checking for the stability of $N \times p \times T$ eigenvalues at once, the problem would be reduced to checking $N \times p$ eigenvalues T times (i.e. evaluate 18 eigenvalues 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 2 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 2 corresponds to the mean obtained from 30000 post burn-in draws. Overall, results in Figure 2 reinforce that, while the MCMC algorithm might occasionally produce non-stable draws, such draws are not frequent in our estimation exercise.

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.

⁴The 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 2: Distribution of eigenvalues associated with the autoregressive coefficients for endogenous TVP-VARs



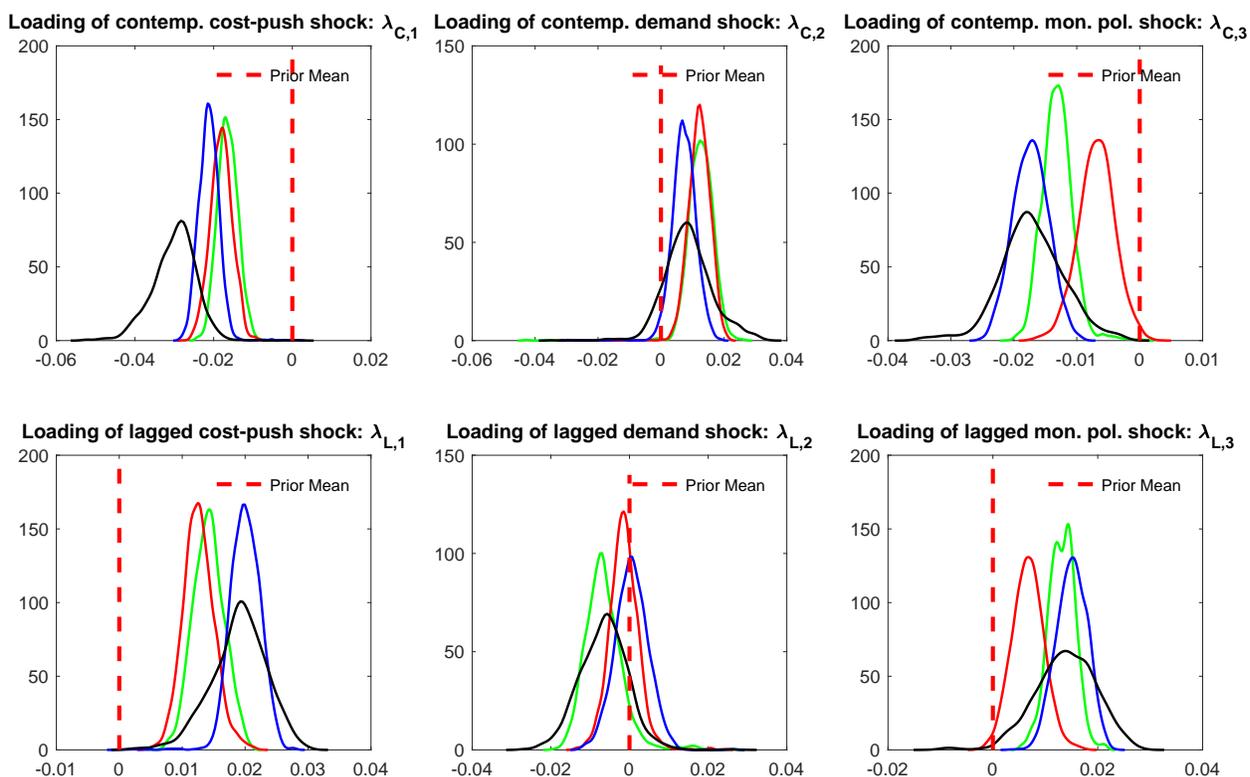
Note: The figure plots the kernel densities associated with the posterior means of all the eigenvalues for the three versions of endogenous TVP-VAR models, that is, with coefficients depending on contemporaneous, lagged, and contemporaneous and lagged structural innovations.

Appendix B: Robustness Checks for Heteroskedasticity

In this section, we present results associated to VAR specifications where the covariance matrix of the reduced form innovations and the conditional variance of the drifting coefficients are assumed to exhibit breaks at specific time periods. The break dates are set to 1979Q1, 1987Q2 and 2007Q2, which leads to four volatility regimes, that is, 1967Q1-1979Q1, 1979Q2-1987Q1, 1987Q2-2007Q2 and 2007Q2-2018Q2.

Evidence of Endogenous Time Variation

Figure 3: Posterior densities of the effect of structural shocks to time-varying coefficients (breaks in 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 underlying model allows for breaks in volatility. Green, red, blue and black lines correspond to regimes 1, 2, 3 and 4, respectively. The $2\log(\text{BF})$ 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.

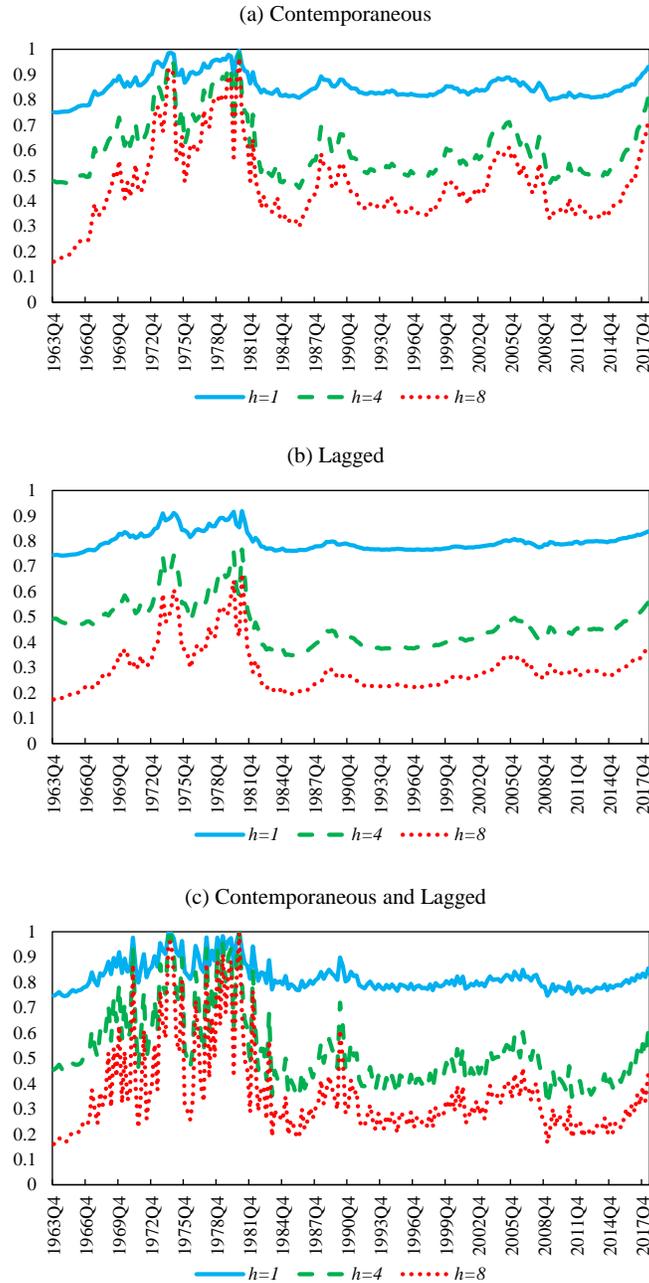
Application 1: Inflation-Gap Persistence

Following [Cogley, Primiceri, and Sargent \(2010\)](#), we compute a measure of inflation-gap persistence based on short- and medium-term predictability. Since the parameters of the model are assumed to be time-varying, the trend inflation is defined by its unconditional mean. Specifically, the employed measure of persistence, $R_{h,t}^2$, computes, at time t , the fraction of total variance of inflation gap h periods ahead, g_{t+h} , that can be attributed to shocks inherited from the past, relative to those shocks that will occur in the future. Since future shocks account for the forecast error, this measure can be expressed as one minus the ratio of the conditional variance to the unconditional variance, as follows,

$$R_{h,t}^2 \approx 1 - \frac{\mathbf{s}_\pi' \left[\sum_{i=0}^{h-1} \mathbf{B}_t^i \text{Var}(\boldsymbol{\xi}_{t+1}) \mathbf{B}_t^{i'} \right] \mathbf{s}_\pi}{\mathbf{s}_\pi' \left[\sum_{i=0}^{\infty} \mathbf{B}_t^i \text{Var}(\boldsymbol{\xi}_{t+1}) \mathbf{B}_t^{i'} \right] \mathbf{s}_\pi}, \quad (11)$$

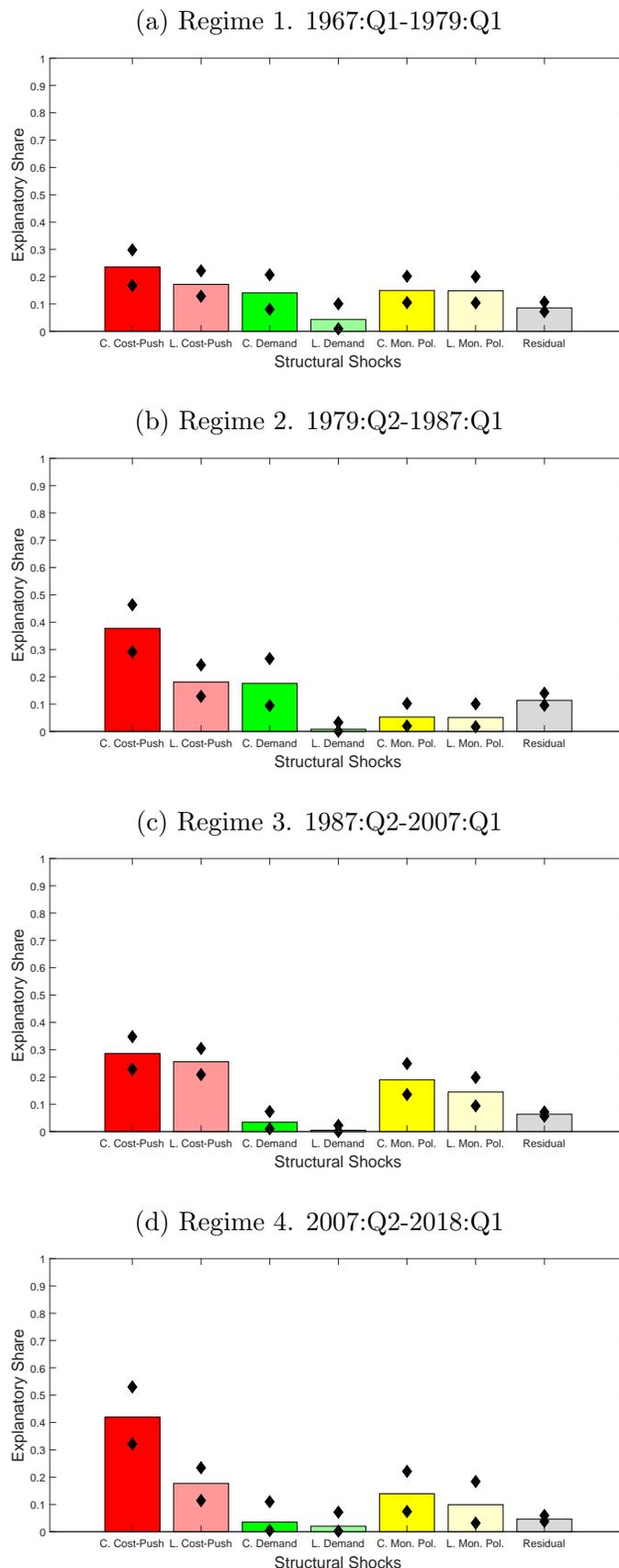
where \mathbf{B}_t and $\boldsymbol{\xi}_t$ correspond to the autoregressive coefficients matrix and reduced form innovations vector of the VAR in companion form, respectively, and \mathbf{s}_π denotes a vector that helps to select the coefficients associated with the equation of inflation rate. Notice that we perform two types of exercises, where the covariance matrix of $\boldsymbol{\xi}_t$ can be either time-invariant or subject to the breaks, as defined above. The measure, $R_{h,t}^2$, represents a fraction, and therefore, is bounded between zero and one. For small or medium $h \geq 1$, a small (large) fraction indicates weak (strong) persistence of inflation gap.

Figure 4: Inflation-gap persistence based on $R_{h,t}^2$ statistics (breaks in volatility)



Note: The figure plots the time-varying measure of inflation-gap persistence, $R_{h,t}^2$, for the one-, four- and eight-quarter-ahead forecasting horizons. $R_{h,t}^2$ is constructed using Equation 12 in [Cogley, Primiceri, and Sargent \(2010\)](#). 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.

Figure 5: Contribution of contemporaneous and lagged structural shocks to parameter instability (breaks in the volatility)

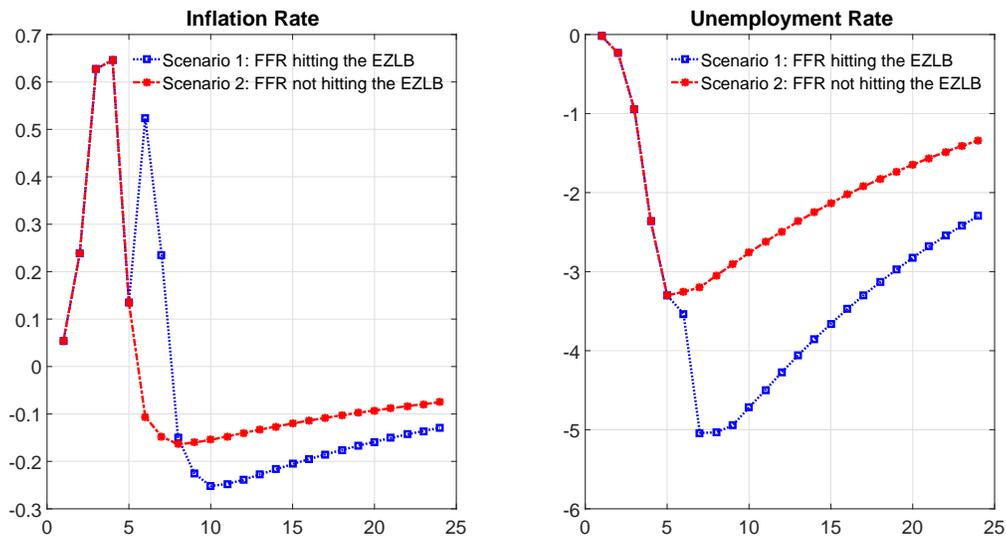


Note: The figure shows the contributions contemporaneous (C.) and lagged (L.) structural shocks, along with the contribution of exogenous shocks, to the overall changes in the drifting VAR coefficients. The contributions are based on equations (26) and (27), and normalized so that they sum up to one. Black dots indicate the 16th and 84th percentiles of the corresponding posterior densities.

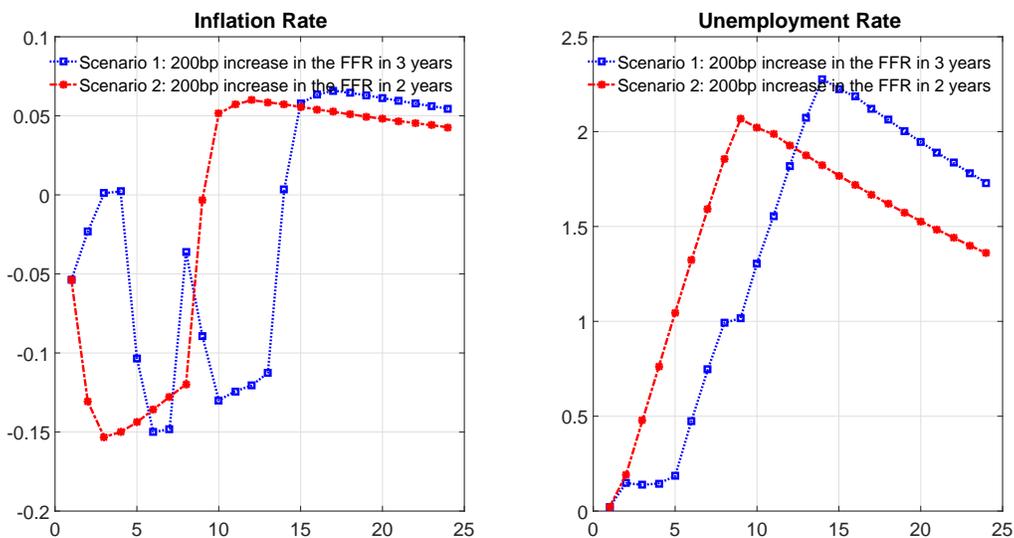
Application 2: Monetary Policy Counterfactuals

Figure 6: Responses of unemployment and inflation to monetary policy shocks (breaks in the volatility)

(a) The trajectory towards the zero lower bound in 2007



(b) The interest rate normalization in 2015



Note: Chart (a) plots the responses of unemployment rate and inflation rate under the scenarios associated with the zero lower bound period. Chart (b) plots the responses of unemployment rate and inflation rate under the scenarios associated with the normalization of the monetary policy.

References

- CHIB, S. (2001): “Markov chain Monte Carlo methods: computation and inference,” *Handbook of Econometrics*, 5, 3569–3649.
- COGLEY, T., G. PRIMICERI, AND T. SARGENT (2010): “Inflation-gap persistence in the US,” *American Economic Journal: Macroeconomics*, 2(1), 43–69.
- KOOP, G., AND S. POTTER (2011): “Time varying VARs with inequality restrictions,” *Journal of Economic Dynamics and Control*, 35(7), 1126–1138.
- PRIMICERI, G. (2005): “Time Varying Structural Vector Autoregressions and Monetary Policy,” *Review of Economic Studies*, 72, 821–852.