CESifo WORKING PAPERS

8060
2020

January 2020

Nonlinear Business Cycle and Optimal Policy: A VSTAR Perspective

Vito Polito

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 8060*

Nonlinear Business Cycle and Optimal Policy: A VSTAR Perspective

Abstract

This paper studies optimal macroeconomic policy when nonlinearity in the business cycle is described by a vector smooth transition autoregression (VSTAR). A structural identification of the VSTAR that yields a low-dimension and certainty-equivalent nonlinear quadratic regulator (NLQR) problem is derived. Optimal rules are calculated by adapting from the engineering theory the approach of State Dependent Riccati Equation, which allows standard dynamic programming techniques to solve NLQR problems. The methodology is employed to study optimal conventional and quantitative easing (QE) monetary policy using a VSTAR model estimated on data for the United States during 1979-2018. The model allows for regime changes during periods of economic slack and when interest rates are near the zero lower bound. The results highlight the quantitative significance of nonlinearity in the analysis of optimal monetary policy and how the size, timing and composition of QE can influence macroeconomic dynamics.

JEL-Codes: C300, C600, E500.

Keywords: smooth transition models, nonlinear quadratic regulator, zero lower bound, quantitative easing, optimal monetary policy.

Vito Polito
University of Sheffield
9 Mapping Street
United Kingdom – S1 4DT
v.polito@sheffield.ac.uk

January 2020

1 Introduction

This paper studies optimal macroeconomic policy when the interaction between the economy and the policy instruments over the business cycle is described by a widely-used class of nonlinear autoregression models: vector smooth transition autoregression (VSTAR). VSTAR models have become increasingly popular in macroeconometrics given their ability to better describe data compared to linear models. In applied works VSTAR is often used for forecasting and to highlight history dependence in the response of macroeconomic variables to shocks.[1] While the econometric literature has long established methods for specification, estimation and structural analysis with VSTAR,[2] no attention has been devoted so far to the analysis of optimal policy.

The present paper provides the first attempt to fill this gap. The paper describes a methodology for calculating optimal policy rules when macroeconomic dynamics are represented by a VSTAR. The methodology is used to study the *optimal* coordination between conventional (interest rate) monetary policy and large scale asset purchase programs, frequently referred to as quantitative easing (QE), undertaken by the Fed over the past ten years. The macroeconomic dynamics resulting from the optimal policy serve as benchmark to evaluate those observed from the data, which reflect the *actual* policy undertaken by the Fed. This seems a pertinent application of the proposed methodology, given the extent of nonlinearity displayed by macroeconomic data of the United States over the last 40 years and, particularly, by the federal funds rate and the Fed's balance sheet since the (2007-2009) Great Recession.[3]

To highlight the methodological contributions of the paper it is useful to recap what is known from the literature on optimal policy analysis based on linear vector autoregression (hereafter VAR). When the macroeconomy is described by a reduced-form VAR at least two possible structural forms can be identified for the purpose of calculating an optimal policy rule (Sack, 2000; Stock and Watson, 2001). These structural models differ in terms of the timing assumption regarding the interaction between the economy and the policy instruments. Under A1, the economy responds with a lag to change in policy. Under A2, the economy responds within the same period in which policy is changed. If the objective function of the decision maker is quadratic, an optimal decision rule can be calculated as the solution to a linear quadratic regulator (LQR) problem using either of these two structural models to represent the constraints faced by the regulator. Crucially, decision rules under either A1 or A2 satisfy certainty equivalence, solving the LQR problem regardless of the particular sequence of disturbances. For this reason decision rules under A1 and A2 can be computed

[1] See, e.g., Weise (1999), Balke (2000), Galvão (2006), Auerback and Gorodnichenko (2012), Caggiano, Castelnuovo and Groshenny (2014), Caggiano et al. (2015), Galvão and Owyang (2018).

[2] See, e.g., Teräsvirta, Tjøstheim and Granger (2010), Hubrich and Teräsvirta (2013), Kilian and Lütkepohl (2017).

[3] Nonlinearity in the mean and volatility of macroeconomic and financial time series of the United States has long been documented, see, e.g., Stock and Watson (1996) and McConnell and Pérez-Quirós (2000), Ang and Bekaert (2002). Many empirical studies on the transmission mechanism of monetary policy over the business cycle in the United States give clear evidence in support of models that, like the VSTAR, allow for gradually evolving parameters and heteroscedastic shocks, including, among the others, Primiceri (2005); Canova and Gambetti (2009); Baumeister, Liu, and Mumtaz (2013).

using standard dynamic programming techniques.[4]

The first methodological contribution of the paper refers to the identification of a structural model suitable for optimal policy analysis from a reduced-form VSTAR. The paper shows that both A1 and A2 can still be used to identify two alternative nonlinear structural models from a reduced-form VSTAR. This time, however, differences between the two structural models go well beyond the timing of interaction among variables. It is found that under A1 the structural model forms a nonlinear quadratic regulator (NLQR) problem whose solution is high dimensional and no longer consistent with certainty equivalence. In contrast, A2 has the advantage of delivering a NLQR problem whose solution is low dimensional and compatible with certainty equivalence.

The second methodological contribution of the paper refers to the calculation of optimal decision rules that solve NLQR problems when the constraint is given by the nonlinear structural model identified from the VSTAR. The paper shows how to adapt from the engineering theory the so-called State Dependent Riccati Equation (SDRE) method. This consists of employing state dependent coefficient (SDC) factorization to transform the nonlinear structural model into an affine structure with SDC matrices. As a result, the NLQR problem becomes isomorphic to the LQR problem. Under A2, the problem is also low dimensional and certainty equivalent, and it can be solved with standard dynamic programming methods. The solution gives an optimal feedback rule with time-varying coefficients. The paper further shows how to combine this with the structural model to derive the VSTAR under the optimal rule for dynamic analysis.

It is important to highlight the usefulness of the SDRE method for applied economics. Policy analysis presently done in macroeconomics is still largely based on the well-known paradigm of the LQR (Ljungqvist and Sargent, 2018). This is applicable as long as the model of the economy is linear. Of course, any nonlinear model of the economy could be linearized and then analyzed with LQR. However, as well as providing solutions that are not globally valid, linearization assumes away three crucial features of macroeconomic data: asymmetries, threshold effects and large transitional changes. A VSTAR could well capture these features, but the LQR solution would disregard them, potentially resulting in incorrect and/or inefficient decision making. Alternatively, policy analysis using a nonlinear model could be carried out with numerical methods.[5] While these deliver solutions that are globally valid without disregarding nonlinearity, their main drawback is that they are computationally intensive. In contrast, the SDRE method for solving NLQR problems has the advantage of accounting for nonlinearity while being computationally simple and easy to understand given his similarity to the LQR.

The quantitative contribution of the paper consists of applying the methodology on a VSTAR estimated with maximum likelihood on monthly data for the United States during 1979-2018. The model includes four variables describing the economy and the financial system conditions: the inflation rate, the growth rate of industrial production, the unemployment rate and an indicator of credit risk. Monetary policy is described by three indicators: the federal funds rate and assets held by the Fed, distinguishing between Treasury securities and as-

[4] See, e.g., chapter 5 of Ljungqvist and Sargent (2018) for LQR solution under A1; chapter 10 of Chow (1976) for LQR solution under A2; Polito and Wickens (2012) for qualitative and quantitative comparison of the two solutions.

[5] See, e.g., Judd (1998), Miranda and Fackler (2002).

sets issued by the private sector. For this reason the analysis can account for the effects on aggregate quantities and prices of changes in both the size and composition of the Fed's balance sheet.[6] Nonlinearity in the mean and variance of the VSTAR can come through two sources. One is the economy, capturing changes in the coefficients of the model during periods of economic slack, as in Ramey and Zubairy (2018). The other is monetary policy, capturing changes in coefficients eventually occurring when the federal funds rate is near to the zero lower bound (ZLB). For the purpose of validation, it is shown that the estimated VSTAR provides (i) a good fit of the data, better than a number of alternative specifications, and (ii) a plausible description of the response of the policy instruments to shocks at different times of the United States's monetary history.

The estimated VSTAR is used to evaluate the macroeconomic effects of conventional and QE monetary policy undertaken by the Fed since 2008. To this end, macroeconomic dynamics obtained from the VSTAR once the monetary policy instruments are jointly optimized are used as benchmark. Three tasks are carried out: evaluating the gains from the joint optimization of monetary policy instruments; studying the economy and policy instruments responses to shocks once the optimal policy is implemented; undertaking a counterfactual simulation of macroeconomic dynamics under the optimal policy.

The gain from the joint optimization of the monetary policy instruments it is found to be large. According to the quantitative analysis, the average unemployment rate would have reduced by 1.7 percent had the federal funds rate and QE been coordinated optimally since 2008. However, this is not as large as the reduction achieved by the actual QE policy relative to a scenario of no-QE which is measured to be about 2.8 percent. The optimization gain is also found to be larger when measured over the Great Recession and its aftermath, rather than the whole 1979-2018. This suggests that the measured benefits from monetary policy coordination are larger during periods when nonlinearity is more significant. Further, the gain is recalculated using the VAR both over the full sample and the post-2008 period. In both cases the measured gains are higher than those from the VSTAR. This suggests that linear models can potentially overstate the benefits from monetary policy optimization.

Both actual and optimal QE policy display a significant degree of asymmetry and history dependence in their responses to demand and supply shocks. However, differences in the responses to shocks of QE under the optimal and the actual policy are significant only in the very short-run horizon (four to six months). The response of QE under the optimal policy shows two clear patterns. After a demand shock QE results in increase in the size of the balance sheet and shift in the portfolio mix of assets towards Treasury securities. After a supply shock QE still leads to increase in the size of the balance sheet but there is not shift in the portfolio mix, since holdings of both Treasuries and private securities increase.

[6] The theoretical literature on the transmission mechanisms of QE argues that central bank holdings of either government or private-sector bonds influence the economy through two separate channels. These are referred to as the portfolio-balancing channel for government assets and the credit channel for private assets, see, e.g. Joyce et al. (2012) or Kuttner (2018). In this literature asset purchases of government and private bonds are frequently referred to as quantitative and credit easing, respectively. As in Kuttner (2018), throughout this paper asset purchases are referred to as QE regardless of whether these entail government or private sector bonds.

The counterfactual simulation of the likely evolution of the economy and the policy instruments, had the federal funds rate and QE been coordinated optimally since 2008, shows that the observed overall increase in the Fed's balance sheet since 2008 was not far from that prescribed by the optimal policy. Differences between the actual and the optimal policy are larger during the first phase of QE, as the latter would have prescribed larger purchases of Treasuries. This would have further lowered the credit spread and increased inflation during 2008-2013. Finally, the observed duration of the ZLB period is found to be not too far from that otherwise prescribed by the optimal policy. This result is in contrast with the view that the ZLB should have terminated much earlier (around the beginning of 2011) had monetary policy during the Great Recession being conducted as predicted by a standard Taylor rule (Federal Reserve Bank of St. Louis, 2015).

Two potential caveats should be highlighted about these results. The first concerns to the specification of the Fed preferences once QE is an available monetary policy instrument. The economic theory offers little guidance on this. Dynamic stochastic general equilibrium (DSGE) models with QE feature, among the others, heterogeneous households. For this reason a micro-founded loss function is undefined in these models, since the aggregation of agents' preferences depends on arbitrary weights assigned to different households' type. Given the Fed's dual mandate of price stability and full employment, optimal policy is computed in this paper assuming a standard quadratic loss function in inflation and unemployment, following from Sims and Wu (2019b). For robustness, the likely effects of alternative specifications of the Fed preferences are appraised by varying the loss function parameters. The second caveat is that, in principle, counterfactual policy analysis with the VSTAR is subject to the Lucas (1976) critique. While the quantitative significance of this observation is debated, see Sims and Zha (2006), the latest consensus to guard against it is to consider policy changes that entail small deviations from the observed dynamics, see Rudebusch (2005) and Benati (2019). Following this, the loss function adopted in the quantitative analysis is expanded to include explicit penalties for changes in the policy instruments. This has the effect of inducing a certain degree of gradualism in the optimal policy reaction functions, while keeping the path of the macroeconomy under the optimal policy close to that observed in the data.[7]

The paper makes several contributions to the literature. The methodological part is related to the literature on nonlinear time series modelling with VSTAR, surveyed by Granger and Teräsvirta (1993), Teräsvirta, Tjøstheim and Granger (2010) and Hubrich and Teräsvirta (2013). This literature concentrates on the specification, estimation and structural analysis of VSTAR models. The present paper tackles a different task, using VSTAR for optimal policy analysis. The methodological part of the paper also contributes to the macroeconomic literature using dynamic programming to solve optimal decision making problems. As mentioned above, the predominant paradigm here is still the LQR, which

[7] Despite the Lucas critique, there are many prominent examples of counterfactual analyses based on reduced-from models in the macroeconomic literature of conventional monetary policy, see Sack (2000), Stock and Watson (2001), Sims and Zha (2006), and more recently of QE, see Lenza, Pill and Reichlin (2010), Chung et al (2012), Giannone et al (2012), Kapetanios et al (2012), Baumeister and Benati (2013), Dahlhaus, Hess and Reza (2018). Many of these works analyze counterfactual deviations of the policy instruments much larger than those considered in this paper.

applies to linear economic environments (Ljungqvist and Sargent, 2018). The SDRE method described in the paper suggests a way of extending the application of the same LQR techniques to nonlinear economic environments. The applied part of the paper contributes to the empirical literature on the macroeconomic effects of monetary policy. Within this a number of recent studies quantifies the effects of QE through either impulse response function (IRF) or counterfactual analysis using vector autoregressions or semi-structural models. Examples include Lenza, Pill and Reichlin (2010), Chung et al. (2012), Giannone et al. (2012), Kapetanios et al. (2012), Baumeister and Benati (2013), Gambacorta et al. (2014), Dahlhaus, Hess and Reza (2018). These works consider only the effects stemming from changes in the overall size of the balance sheet, typically against a counterfactual scenario of no-QE intervention. The present paper considers the impact of changes in both size and composition of the Fed's balance sheet, and evaluates these using as reference a different counterfactual scenario, the optimal policy. Finally, the paper is also related to the DSGE literature on the transmission channels of unconventional monetary policy, see Curdia and Woodford (2011), Gagnon et al. (2011), Woodford (2016), Harrison (2017), Reis (2017), providing instead an evaluation based on an a-theoretical model like the VSTAR. Within this DSGE literature, Guerrieri and Iacoviello (2017) and Sims and Wu (2019a,b) have recently emphasized the significance of macroeconomic asymmetries due to nonlinearity stemming from both the private and the policy sectors of the macroeconomy. The present paper shares with these works a similar emphasis on the dual sources of nonlinearity, but for the purpose of empirical analysis.

The paper is organized as follows. Section 2 sets a reduced-form VSTAR model that encompasses several specifications adopted in the applied macroeconomic literature. This is used to identify the two possible structural representations under A1 and A2 and to highlight their differences. Section 3 describes how the SDRE method is employed to solve a NLQR problem where the constraint faced by the regulator is given by the structural VSTAR model under A2. In this section it is also shown how to combine the optimal policy rule with the NLQR constraint to derive the VSTAR under control. Potential issues of stability and solutions are also discussed here. Section 4 presents the estimated VSTAR model, describing the data, the econometric methodology, the maximum likelihood results and the IRF analysis. Section 5 presents the results from the analysis of optimal policy. Section 6 concludes with a summary. Appendix A gives more detail on the identification of the structural representations from the VSTAR model. Appendix B describes the derivation of the solution to the NLQR problem. Appendix C gives more details on the data used for the quantitative analysis. Appendix D describes the algorithm for calculating the IRFs. Appendix E includes further robustness results from the optimal policy analysis.

2 Macroeconomic Model

The macroeconomic model is a VSTAR including variables describing the economy and the policy sectors. There are two sources of nonlinearity, one stemming from the economy, the other from the policy sector. Nonlinearity affects the coefficients of the mean and the variance structure of the VSTAR.

2.1 Reduced Form

Let $\mathbf{y}_t$ be a $n \times 1$ vector partitioned as $\mathbf{y}_t = \left[\begin{array}{cc} \mathbf{x}_t' & \mathbf{u}_t' \end{array} \right]'$, with $\mathbf{x}_t$ denoting a $p \times 1$ vector of variables describing the economy and $\mathbf{u}_t$ a $m \times 1$ vector of policy instruments. The dynamic of $\mathbf{y}_t$ for $t \geq 0$ is described by the nonlinear first-order stochastic process:

$$\begin{aligned} \mathbf{y}_{t+1} &= \left[1 - g_M\left(\mathbf{s}_u'\mathbf{y}_t\right)\right] \left\{ \begin{array}{c} \left[1 - l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\boldsymbol{\lambda}_1 + \boldsymbol{\Lambda}_1\left(L\right)\mathbf{y}_t\right] + \\ l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\boldsymbol{\lambda}_2 + \boldsymbol{\Lambda}_2\left(L\right)\mathbf{y}_t\right] \end{array} \right\} + \\ & g_M\left(\mathbf{s}_u'\mathbf{y}_t\right) \left\{ \begin{array}{c} \left[1 - l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\boldsymbol{\lambda}_3 + \boldsymbol{\Lambda}_3\left(L\right)\mathbf{y}_t\right] + \\ l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\boldsymbol{\lambda}_4 + \boldsymbol{\Lambda}_4\left(L\right)\mathbf{y}_t\right] \end{array} \right\} + \mathbf{v}_{t+1}, \end{aligned} \tag{1}$$

where $\mathbf{y}_0$ is given; $\boldsymbol{\lambda}_j$ are vectors of constant coefficients of dimensions $n \times 1$; L is the lag operator; $\boldsymbol{\Lambda}_j\left(L\right)$ denote q lags of the vector $\mathbf{y}_t$, i.e. $\boldsymbol{\Lambda}_j\left(L\right) = \sum_{k=1}^{q} \boldsymbol{\Lambda}_{jk} L^k$ with each $\boldsymbol{\Lambda}_{jk}$ being a $n \times n$ matrix of coefficients; $\mathbf{v}_t$ is a vector of reduced-form disturbances to $\mathbf{y}_t$, with $\mathbf{v}_t \sim N\left(\mathbf{0}, \boldsymbol{\Omega}_t\right)$, $j = 1, ..., 4$. The covariance matrix $\boldsymbol{\Omega}_t$ is given by:

$$\begin{aligned} \boldsymbol{\Omega}_t &= \left[1 - g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\left[1 - l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_1 + l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_2\right\} + \\ & g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\left[1 - l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_3 + l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_4\right\}, \end{aligned} \tag{2}$$

where $\boldsymbol{\Omega}_j$ are symmetric matrices of coefficients, $j = 1, ..., 4$.

The terms $g_i\left(\mathbf{s}_u'\mathbf{y}_t\right)$ and $l_i\left(\mathbf{s}_x'\mathbf{y}_t\right)$ denote continuos (transition) functions, bounded between zero and one, capturing the effect of nonlinearity in the transmission mechanism of shocks in the mean, $i = M$, and variance, $i = V$, of the VSTAR model in (1) and (2), with $\mathbf{s}_x$ and $\mathbf{s}_u$ being selection vectors identifying transition variables from $\mathbf{x}_t$ and $\mathbf{u}_t$ respectively. Transition variables can be either single, or linear combinations of, variables from the economy, $\mathbf{s}_x'\mathbf{y}_t$, and the policy vector, $\mathbf{s}_u'\mathbf{y}_t$. These determine the state in which $\mathbf{y}_t$ and $\boldsymbol{\Omega}_t$ are transiting during any given period t. In particular, $l_i\left(\mathbf{s}_x'\mathbf{y}_t\right)$ determines changes in the VSTAR coefficients due to changes in the state of the economy. These are nested within changes in the VSTAR coefficients due to variation of the policy stance, as captured by $g\left(\mathbf{s}_u'\mathbf{y}_t\right)$.

The transition functions in the mean and variance of the VSTAR allow for changes across regimes to be either gradual or abrupt. In this second case, the model tends to a threshold vector autoregression model. Setting all transition functions equal to zero yields a VAR model. The VSTAR also nests two alternative specifications that attribute nonlinearity to the mean or the variance alone. Using in (1) the restriction $\boldsymbol{\lambda}_j = \boldsymbol{\lambda}$ and $\boldsymbol{\Lambda}_j\left(L\right) = \boldsymbol{\Lambda}\left(L\right)$, $j = 1, ..., 4$, yields a VSTAR with nonlinear variance only. Alternatively, the restriction in (2) $\boldsymbol{\Omega}_j = \boldsymbol{\Omega}$, $j = 1, ..., 4$, yields a VSTAR with nonlinear mean only.

Equations (1) and (2) encompass the typical specification of the VSTAR model adopted in the applied macroeconomic literature.[8] This assumes that there is only one source of nonlinerity due to change in the economy that impacts at the same time on the mean and variance processes. Such a specification is obtained by restricting $g_i\left(\mathbf{s}_u'\mathbf{y}_t\right) = 0$, $i = M$,V, and $l_M\left(\mathbf{s}_x'\mathbf{y}_t\right) = l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)$ in (1) and (2). However, macroeconomic policy decisions are also subject to

[8] See, e.g., Weise (1999), Balke (2000), Galvão (2006), Auerback and Gorodnichenko (2012), Caggiano, Castelnuovo and Groshenny (2014), Caggiano et al (2015), Galvão and Owyang (2018).

nonlinearity. For example, conventional monetary policy action is constrained by the ZLB on the nominal rate of interest. At the same time, the size and composition of many central banks balance sheet display very different dynamics before and after the Great Recession. In this respect, the VSTAR model in (1)-(2) provides a natural framework to study monetary policy, as it can capture the asymmetries stemming from policy action.

The idea of considering simultaneously the effects of dual nonlinearity, from the economy and the policy sector, is not new. This is explored in the recent works of Guerrieri and Iacoviello (2017) and Sims and Wu (2019a,b) that use DSGE models where nonlinearity stems from the private sector, due to the presence of collateral constraints on borrowing, and from the policy sector, due to the nonnegativity constraint on the monetary policy rate. The VSTAR model specified in (1) and (2) share with this works the similar emphasis on the dual nonlinearity from the economy and policy sectors, with the purpose of evaluating its significance for empirical analysis.

2.2 Structural Form(s)

The VSTAR specified in (1)-(2) is a reduced-form model, therefore suitable for econometric estimation and dynamic analysis like forecasting. It can also be used to analyze the response of the economy to two types of policy intervention: unanticipated and anticipated changes in the policy instruments. The latter involves replacing the coefficients of the equations corresponding to the policy instruments, which describe the so-called actual policy rules, with either ad-hoc coefficients or those calculated from an optimal policy. As in a VAR, both types of policy analysis require identifying from the reduced-form model a structural form (Sack 2000, Stock and Watson, 2001).

Identification of a structural model to quantify the effects of surprise policy changes in a VSTAR is typically carried out through a full Cholesky factorization (necessary condition) of the reduced-form covariance matrix. Since $\mathbf{y}_t$ is a $n \times 1$ vector, the number of possible structural transformations is $n!$.

In contrast, identification of the structural model to quantify the effects of changes in the decision rules requires only a zero block transformation (sufficient condition) of the reduced-form model, to separate the policy vector from the economy vector. Therefore, there are (at least) two possible structural representations.[9] A detailed description of the derivation of these two representation is provided in Appendix A. To illustrate here, consider partitioning (1) to separate

[9] Given a zero block restriction, the economy and the policy vectors can be separated. Other structural representations could be obtained by further restricting the covariance matrix of the equations for the variables in the economy vector. These further restrictions are however not necessary when using a VAR (or a VSTAR) to study the effects of changes in the policy rules.

the economy from the policy instruments as:

$$\left[\begin{array}{c}\mathbf{x}_{t+1}\\ \mathbf{u}_{t+1}\end{array}\right] = \tag{3}$$

$$\left[1-g_M\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\begin{array}{l}\left[1-l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\begin{array}{c}\boldsymbol{\lambda}_{x1}\\ \boldsymbol{\lambda}_{u1}\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{x1}\left(L\right)\\ \boldsymbol{\Lambda}_{u1}\left(L\right)\end{array}\right]\mathbf{y}_t\\ +l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left\{\left[\begin{array}{c}\boldsymbol{\lambda}_{x2}\\ \boldsymbol{\lambda}_{u2}\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{x2}\left(L\right)\\ \boldsymbol{\Lambda}_{u2}\left(L\right)\end{array}\right]\mathbf{y}_t\right\}\end{array}\right\}+$$

$$g_M\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\begin{array}{l}\left[1-l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\begin{array}{c}\boldsymbol{\lambda}_{x3}\\ \boldsymbol{\lambda}_{u3}\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{x3}\left(L\right)\\ \boldsymbol{\Lambda}_{u3}\left(L\right)\end{array}\right]\mathbf{y}_t\\ +l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left\{\left[\begin{array}{c}\boldsymbol{\lambda}_{x4}\\ \boldsymbol{\lambda}_{u4}\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{x4}\left(L\right)\\ \boldsymbol{\Lambda}_{u4}\left(L\right)\end{array}\right]\mathbf{y}_t\right\}\end{array}\right\}+\left[\begin{array}{c}\mathbf{v}_{xt+1}\\ \mathbf{v}_{ut+1}\end{array}\right],$$

where $\boldsymbol{\Lambda}_{xj}\left(L\right)=\left[\begin{array}{cc}\boldsymbol{\Lambda}_{xxj}\left(L\right) & \boldsymbol{\Lambda}_{xuj}\left(L\right)\end{array}\right]$, $\boldsymbol{\Lambda}_{uj}\left(L\right)=\left[\begin{array}{cc}\boldsymbol{\Lambda}_{uxj}\left(L\right) & \boldsymbol{\Lambda}_{uuj}\left(L\right)\end{array}\right]$, $j=1,...,4$. The covariance matrix (2) is partitioned conformably as:

$$\boldsymbol{\Omega}_t=\left[\begin{array}{cc}\boldsymbol{\Omega}_{xxt} & \boldsymbol{\Omega}_{xut}\\ \boldsymbol{\Omega}_{uxt} & \boldsymbol{\Omega}_{uut}\end{array}\right] \tag{4}$$

$$=\left[1-g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\begin{array}{c}\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx1} & \boldsymbol{\Omega}_{xu1}\\ \boldsymbol{\Omega}_{ux1} & \boldsymbol{\Omega}_{uu1}\end{array}\right]\\ +l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx2} & \boldsymbol{\Omega}_{xu2}\\ \boldsymbol{\Omega}_{ux2} & \boldsymbol{\Omega}_{uu2}\end{array}\right]\end{array}\right\}+$$

$$g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\begin{array}{c}\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx3} & \boldsymbol{\Omega}_{xu3}\\ \boldsymbol{\Omega}_{ux3} & \boldsymbol{\Omega}_{uu3}\end{array}\right]\\ +l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx4} & \boldsymbol{\Omega}_{xu4}\\ \boldsymbol{\Omega}_{ux4} & \boldsymbol{\Omega}_{uu4}\end{array}\right]\end{array}\right\}.$$

Assume that the economy does not react within the same period of a change in policy, A1. This implies the restriction $\boldsymbol{\Omega}_{xut}=\mathbf{0}$ in (4) and yields the structural representation for the economy vector:

$$\begin{aligned}\mathbf{y}_{t+1} &= \left[1-g_M\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\begin{array}{c}\left[1-l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\boldsymbol{\phi}_1+\boldsymbol{\Phi}_1\left(L\right)\mathbf{y}_t+\boldsymbol{\Gamma}_1\mathbf{u}_t\right]\\ +l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left(\boldsymbol{\phi}_2+\boldsymbol{\Phi}_2\left(L\right)\mathbf{y}_t+\boldsymbol{\Gamma}_2\mathbf{u}_t\right)\end{array}\right\}+ \\ &\quad g_M\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\begin{array}{c}\left[1-l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\boldsymbol{\phi}_3+\boldsymbol{\Phi}_3\mathbf{y}_t+\boldsymbol{\Gamma}_3\left(L\right)\mathbf{u}_t\right]\\ +l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\boldsymbol{\phi}_4+\boldsymbol{\Phi}_4\left(L\right)\mathbf{y}_t+\boldsymbol{\Gamma}_4\mathbf{u}_t\right]\end{array}\right\}+\mathbf{e}_{t+1}, \end{aligned} \tag{5}$$

$$\boldsymbol{\Sigma}_t = \left[\begin{array}{cc}\boldsymbol{\Omega}_{xxt} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right], \tag{6}$$

where $\boldsymbol{\phi}_j=\left[\begin{array}{cc}\boldsymbol{\lambda}_{xj}' & \mathbf{0}'\end{array}\right]'$; $\boldsymbol{\Phi}_j\left(L\right)=\left[\begin{array}{cc}\boldsymbol{\Lambda}_{xxj}\left(L\right) & \boldsymbol{\Lambda}_{xuj}\left(L\right)-\boldsymbol{\Lambda}_{xu1j}\\ \mathbf{0} & \mathbf{0}\end{array}\right]$, with (i) $\boldsymbol{\Lambda}_{xxj}\left(L\right)$ and $\boldsymbol{\Lambda}_{xuj}\left(L\right)$ denoting partitions of the responses of the economy variables, $\boldsymbol{\Lambda}_{xj}\left(L\right)$, to their lagged values and to the policy instruments, respectively, and (ii) $\boldsymbol{\Lambda}_{xu1j}\left(L\right)$ denoting the responses of the economy variables to the first lag of the policy instruments; $\boldsymbol{\Gamma}_j=\left[\begin{array}{cc}\boldsymbol{\Lambda}_{xuj1}' & \mathbf{0}'\end{array}\right]'$; $j=1,...,4$ and $\mathbf{e}_t=\left[\begin{array}{cc}\mathbf{v}_{xt}' & \mathbf{0}'\end{array}\right]'\sim N\left(\mathbf{0},\boldsymbol{\Sigma}_t\right)$.

Conversely, assume that the economy reacts within the same period of a change in policy, A2. This implies the block restriction $\boldsymbol{\Omega}_{uxt}=\mathbf{0}$ in (4) yields

the structural representation for the economy vector:

$$\begin{aligned}\mathbf{y}_{t+1} &= \left[1-g_M\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\begin{array}{c}\left[1-l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\boldsymbol{\phi}_{1t}+\boldsymbol{\Phi}_{1t}\left(L\right)\mathbf{y}_t\right]\\ +l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\boldsymbol{\phi}_{2t}+\boldsymbol{\Phi}_{2t}\left(L\right)\mathbf{y}_t\right]\end{array}\right\}+ \qquad (7)\\ &\quad g_M\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\begin{array}{c}\left[1-l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\boldsymbol{\phi}_{3t}+\boldsymbol{\Phi}_{3t}\left(L\right)\mathbf{y}_t\right]\\ +l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\boldsymbol{\phi}_{4t}+\boldsymbol{\Phi}_{4t}\left(L\right)\mathbf{y}_t\right]\end{array}\right\}+\boldsymbol{\Gamma}_t\mathbf{u}_{t+1}+\mathbf{e}_{t+1},\\ \boldsymbol{\Sigma} &= \left[\begin{array}{cc}\boldsymbol{\sigma}_x^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{array}\right], \qquad (8)\\ \boldsymbol{\Gamma}_t &= \left[\begin{array}{cc}\mathbf{G}_{12t}' & \mathbf{I}\end{array}\right]', \mathbf{G}_{12t}=\boldsymbol{\Omega}_{xut}\boldsymbol{\Omega}_{uut}^{-1} \qquad (9)\end{aligned}$$

where $\boldsymbol{\phi}_{jt}=\left[\begin{array}{cc}\left(\boldsymbol{\lambda}_{xj}-\mathbf{G}_{12t}\boldsymbol{\lambda}_{uj}\right)' & \mathbf{0}'\end{array}\right]'$, $\boldsymbol{\Phi}_{jt}\left(L\right)=\left[\begin{array}{c}\boldsymbol{\Lambda}_{xj}\left(L\right)-\mathbf{G}_{12t}\boldsymbol{\Lambda}_{uj}\left(L\right)\\ \mathbf{0}\end{array}\right]$, $j=1,...,4$, $\mathbf{e}_t=\left[\begin{array}{cc}\left(\mathbf{v}_{xt}-\mathbf{G}_{12t}\mathbf{v}_{ut}\right)' & \mathbf{0}'\end{array}\right]'\sim N\left(\mathbf{0},\boldsymbol{\Sigma}\right)$, and $\boldsymbol{\sigma}_x^2$ denotes the variance of $\mathbf{v}_{xt}-\mathbf{G}_{12t}\mathbf{v}_{ut}$ (see Appendix A).

The nonlinear dynamic structure of $\mathbf{y}_t$ under both (5) and (7) describes a VSTAR with exogenous variables, since the policy vector $\mathbf{u}_t$ appears on the right side in both models as an exogenous variable.[10] Therefore, both (5) and (7) can be used to represent the nonlinear economy constraint faced by a regulator in charge of setting $\mathbf{u}_t$. When the objective of the regulator is approximated by a quadratic function, the resulting problem is referred to as the nonlinear quadratic regulator (NLQR) problem.

It is useful to point out the differences between (5)-(6) and (7)-(9). The structural model in (5)-(6) corresponds to the equations for the economy vector $\mathbf{x}_t$ in the reduced-form model (3). Under this representation, the economy variables respond with a lag to changes in policy. In contrast, the structural model (7)-(9) is obtained by extrapolating from the reduced-form covariance matrix (4) the contemporaneous response of the economy to policy and then incorporating this into the systematic part of $\mathbf{x}_t$. Thus the economy responds within the period to changes in policy. This difference in the timing of the two structural representation is similar to that obtained when applying A1 and A2 to a VAR.[11]

Aside from the timing of interaction among variables, there are however three further differences between (5)-(6) and (7)-(9) that are specific to the nonlinear nature of the VSTAR model in (1)-(2). The first refers to the type of nonlinearity featuring the response coefficient of the economy variables $\mathbf{x}_t$ to changes in the policy vector $\mathbf{u}_t$. In model (5) this reflects the nonlinearity in the mean of the VSTAR as it is given by $\boldsymbol{\Gamma}_j=\left[\begin{array}{cc}\boldsymbol{\Lambda}_{xuj1}' & \mathbf{0}'\end{array}\right]'$, $j=1,...,4$. In model (7), this is determined by the nonlinearity in the variance of the VSTAR as it is given by $\boldsymbol{\Gamma}_t$ in (9). The second difference between (5)-(6) and (7)-(9) refers to the coefficients $\boldsymbol{\phi}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Gamma}$ in the two structural models. In model (5) these are fixed, and it is their nonlinear combination that varies across states, reflecting the state-dependent nature of time variation in the coefficients of the VSTAR. In model (7), these coefficients are time varying and written as

[10] Note how the coefficients corresponding to the equations for the policy instruments in $\mathbf{u}_t$ are all zero in both (5) and (7).

[11] There is not agreement on the correct timing protocol among economists. Sack (2000) employs A1, whereas Stock and Watson (2001) argue in favour of A2. Starting from a reduced-form VAR, Polito and Wickens (2012) evaluate optimal monetary policy in the United States under these two alternative timing protocols and find that, quantitatively, differences in the solutions are generally not large.

$\boldsymbol{\phi}_{jt}$, $\boldsymbol{\Phi}_{jt}$ and $\boldsymbol{\Gamma}_{jt}$, $j = 1, ..., 4$, since they are re-calculated to incorporate time variation from the reduced-form covariance matrix. The third difference refers to the variance structure of the disturbances in the two representations. While this is not constant across observations and nonlinear in (6), disturbances are homoscedastic in the structural model (9).

Both (5)-(6) and (7)-(9) can be employed as nonlinear dynamic constraints of a NLQR problem. The structural model (5)-(6) has however two main drawbacks. First, the solution to the NLQR problem becomes highly dimensional. This is because the coefficients in (5) depend on the lagged value of the policy vector, the variable set by the regulator. Thus changes in $\mathbf{u}_t$ affect on impact the entire system through Γ_j and $g_M(\mathbf{s}_u'\mathbf{y}_t)$. Second, the solution would no longer satisfy the certainty equivalence principle, since changes in the lagged value of the policy vector would also affect the variance structure in (6). In contrast, the structural model in (7)-(9) yields a solution that has lower dimensionality compared to (5)-(6). This is because changes in the policy vector $\mathbf{u}_{t+1}$ affect the economy through Γ_j but not $g_M(\mathbf{s}_u'\mathbf{y}_t)$, which is pre-determined when policy choices are made. In addition, the structural model in (7)-(9) satisfy the certainty-equivalence principle, because the disturbances in (7) have constant variance given by (9). In other words, the representation in (7)-(9) brings a low-dimension and certainty-equivalent structural identification of the VSTAR model.

To further appreciate the significance of nonlinearity in the structural identification of VSTAR for the analysis of changes in the policy rules, consider the structural models obtained instead by applying A1 and A2 to a reduced-form VAR. Setting $g_i(\mathbf{s}_u'\mathbf{y}_t) = l_i(\mathbf{s}_x'\mathbf{y}_t) = 0$, $i = M, V$, in (1) and (2) yields the reduced-form VAR:

$$\begin{aligned} \mathbf{y}_{t+1} &= \boldsymbol{\lambda}_1 + \boldsymbol{\Lambda}_1(L)\mathbf{y}_t + \mathbf{v}_{t+1},\ \mathbf{v}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}_t) \\ \boldsymbol{\Omega}_{t+1} &= \boldsymbol{\Omega}_1. \end{aligned}$$

A1 and A2 yield the following structural representations:[12]

$$\begin{aligned} \text{A1} &: \mathbf{y}_{t+1} = \widetilde{\boldsymbol{\lambda}}_1 + \widetilde{\boldsymbol{\Lambda}}_1(L)\mathbf{y}_t + \widetilde{\Gamma}_1\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{v}}_{t+1},\ \widetilde{\mathbf{v}}_t \sim N\left(\mathbf{0}, \widetilde{\boldsymbol{\Omega}}\right) \\ \text{A2} &: \mathbf{y}_{t+1} = \widehat{\boldsymbol{\lambda}}_1 + \widehat{\boldsymbol{\Lambda}}_1(L)\mathbf{y}_t + \widehat{\Gamma}_1\widehat{\mathbf{u}}_{t+1} + \widehat{\mathbf{v}}_{t+1},\ \widehat{\mathbf{v}}_t \sim N\left(\mathbf{0}, \widehat{\boldsymbol{\Omega}}\right). \end{aligned}$$

The two structural representations from the VAR differ only in terms of the timing interaction between the economy and the policy instruments, being with a lag under A1 and contemporaneous under A2. The use of either of them as constraint for optimization makes no difference in terms of problem dimensionality due to the absence of the transition functions. At the same time, disturbances are homoscedastic in both models, implying that certainty-equivalent policy rules can be calculated from both.

Given the advantages highlighted above, the NLQR problem is specified and solved in the next section using the structural representation based on A2 in (7)-(9). Stock and Watson (2001) argue in favour of A2, on the ground that it is reasonable to expect the economy to react within the the same period of a change in policy whenever data are observed at monthly or lower frequency. This

[12] For the derivation of these two structural models from a VAR, see Polito and Wickens (2012).

point is further reinforced when the economy vector includes variables reflecting prices and financial market behaviour. The timing of A2 is also consistent with that of a typical New Keynesian model, since the IS and Phillips curves in this model result in quantities and prices that respond within the same period to changes in the policy rate.

3 NLQR Problem

The structural model of the economy identified from the VSTAR can be used to evaluate optimal macroeconomic policy. This is done by formulating a general NLQR problem in which the structural model represents the constraints faced by a quadratic regulator. The optimal policy is calcuated with the SDRE method and then combined with the structural model to derive the reduced-form VSTAR under the optimal policy. Potential issues of stability are also considered.

3.1 Specification

The problem of setting optimal policy is specified as a closed-loop regulator determining the sequence of policy instruments $\{\mathbf{u}_t\}_{t=1}^{\infty}$ that minimizes the quadratic loss function

$$V_0 = E_0 \sum_{t=0}^{\infty} \beta^t \left[(\mathbf{y}_t - \overline{\mathbf{y}})' \mathbf{Q} (\mathbf{y}_t - \overline{\mathbf{y}}) \right], \tag{10}$$

where E_0 denotes mathematical expectation conditional on information at time $t = 0$; $\beta \in (0, 1)$ is the discount factor; $\overline{\mathbf{y}}$ is a vector of targets; $\mathbf{Q}$ is a symmetric positive semidefinite matrix of coefficients. The regulator takes the initial value $\mathbf{y}_0$ as given and sets $\{\mathbf{u}_t\}_{t=1}^{\infty}$ subject to the constraint of the VSTARX model in (7)-(9).[13]

It is not possible to derive a closed-form for the solution of the NLQR problem of minimizing (10) subject to the nonlinear structure of (7)-(9). Economists typically proceed in two possible ways. One option is to linearize the nonlinear model (7) so that the NLQR problem effectively becomes an LQR and a closed-form solution can be calculated with standard dynamic programming tecniques (Ljungqvist and Sargent, 2018). The alternatively is to solve the problem numerically using the nonlinear model of the economy as it is or a higher-order approximation. Using linearization, would somehow be inconsistent with the main aim of this paper, i.e. studying the significance of macroeconomic nonlinearities for optimal policy analysis. The main drawback of numerical methods is that they can be computationally intensive. In the context of VSTAR, Luukkonen, Saikkonen and Teräsvirta (1988) show how to employ Taylor-series expansion of the transition function to derive a polynomial approximation of the VSTAR that can be used to test for nonlinearity in the data. The issue of using this approach for the purpose of solving a NLQR problem is that the economy constraint would remain still highly nonlinear, even when using a first-order approximation of the transition function.

[13] Finding the sequence of $\{\mathbf{u}_t\}_{t=1}^{\infty}$ that minimizes (10) is equivalent to calculate the solution to the optimal policy under commitment. The analysis can be easily extended to (i) compute solutions under discretion, where the regulator minimizes the loss function only in the current period, $(\mathbf{y}_t - \overline{\mathbf{y}})' \mathbf{Q} (\mathbf{y}_t - \overline{\mathbf{y}})$, and/or (ii) allow for time variation in the weighting matrix $\mathbf{Q}$.

The control engineering community has a long history in modelling and solving NLQR problems. Most methods are designed to deal with a generic nonlinear model, rather than being tailored to the specifics of a VSTAR. Among these, a method of increasing popularity is that of the State Dependent Riccati Equation (SDRE). The SDRE method extends the method of Riccati equation iteration used in LQR to the nonlinear case. In practice, this is done by approximating the NLQR problem into a sequence of one-period LQR problems that can be solved independently and with standard dynamic programming techniques. The SDRE method has the advantage of being numerically fairly simple in comparison to many other nonlinear optimization tecniques, as well as being closely related to the well-understood Riccati equation method used for LQR.[14] The next sections describe the implementation of the SDRE method with the VSTAR.

3.2 State Dependent Riccati Equation Method

The SDRE method applied to solve the NLQR problem of minimizing (10) subject to (7)-(9) includes two steps. The first consists of employing state dependent coefficients (SDCs) factorization to transform the nonlinear model (7) in an affine structure with SDC matrices. As a result, the NLQR problem can be written in a form that is isomorphic to the LQR problem and it can be solved with standard dynamic programming techniques. The second step consists of solving the NLQR problem iterating on the resulting SDREs. The solution yields feedback rules for the policy instruments with time-varying response coefficients.

3.2.1 State Dependent Coefficients Factorization

In the engineering control community SDC factorization is typically used for solving (i) continuous time, (ii) deterministic NLQR problems, (iii) using the method of variation. The method is adapted here for solving (i) discrete time, (ii) stochastic NLQR problems, (iii) using dynamic programming. To illustrate the idea of SDC factorization, consider the differential equation for an autonomous system, with fully observable state, nonlinear in the state and affine in the control: $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})\,\mathbf{u}(t)$, $\mathbf{x}(0) = \mathbf{x}_0$, where $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\mathbf{u} \in \mathbb{R}^m$ is the control vector, $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ and $\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^n$ are continuos functions, $\mathbf{f}(\mathbf{0}) = 0$, $\mathbf{g}(\mathbf{x}) \neq 0\ \forall \mathbf{x}$, and $t \in [0, \infty]$. The two nonlinear functions can be factorized as $\mathbf{f}(\mathbf{x}) = \mathbf{A}(\mathbf{x})\,\mathbf{x}$ and $\mathbf{g}(\mathbf{x}) = \mathbf{B}(\mathbf{x})$ where $\mathbf{A}(\mathbf{x})$ and $\mathbf{B}(\mathbf{x})$ are matrices of coefficients that depend on the state vector. Thus the nonlinear differential equation can be written in an affine form with SDC matrices: $\dot{\mathbf{x}} = \mathbf{A}(\mathbf{x})\,\mathbf{x} + \mathbf{B}(\mathbf{x})\,\mathbf{u}$. For this reason, early applications refer to the SDC factorization as *quasi linearization* (Pearson, 1962) or *apparent*

[14] Early work in the engineering literature on the SDRE method was done by Pearson (1962), Garrard, McClamroch and Clark (1967), Burghart (1969) and Wernly and Cook (1975). Cloutier et al. (1996a,b) and Mracek and Cloutier (1998) illustrate the theory and application of the SDRE in the context of affine nonlinear models, establishing conditions required to achieve optimality, stability and convergence. Beeler, Tran and Banks (2000a) compare SDRE with other methods for control of nonlinear models. Beeler, Tran and Banks (2000b) develop SDRE in the context of tracking control and state estimation. Beeler (2004) reviews the SDRE method and considers a number of possible variations to overcome numerical issues.

linearization (Wernli and Cook, 1975).[15]

Following this, the nonlinear system in equation (7) can be factorized in terms of SDC matrices and rewritten as

$$\mathbf{y}_{t+1} = \mathbf{c}(\mathbf{y}_t) + \mathbf{A}(\mathbf{y}_t)\,\mathbf{y}_t + \mathbf{B}(\mathbf{y}_t)\,\mathbf{u}_{t+1} + \mathbf{e}_{t+1} \tag{11}$$

for $t \geq 0$, using:

$$\begin{aligned}
\mathbf{c}(\mathbf{y}_t) &= [1 - g_M(\mathbf{s}_u'\mathbf{y}_t)]\{[1 - l_M(\mathbf{s}_x'\mathbf{y}_t)]\,\boldsymbol{\phi}_{1t} + l_M(\mathbf{s}_x'\mathbf{y}_t)\,\boldsymbol{\phi}_{2t}\} + \\
&\quad g_M(\mathbf{s}_u'\mathbf{y}_t)\{[1 - l_M(\mathbf{s}_x'\mathbf{y}_t)]\,\boldsymbol{\phi}_{3t} + l_M(\mathbf{s}_x'\mathbf{y}_t)\,\boldsymbol{\phi}_{4t}\}, \\
\mathbf{A}(\mathbf{y}_t) &= [1 - g_M(\mathbf{s}_u'\mathbf{y}_t)]\{[1 - l_M(\mathbf{s}_x'\mathbf{y}_t)]\,\boldsymbol{\Phi}_{1t}(L) + l_M(\mathbf{s}_x'\mathbf{y}_t)\,\boldsymbol{\Phi}_{2t}(L)\} + \\
&\quad g(\mathbf{s}_u'\mathbf{y}_t)\{[1 - l_M(\mathbf{s}_x'\mathbf{y}_t)]\,\boldsymbol{\Phi}_{3t}(L) + l_M(\mathbf{s}_x'\mathbf{y}_t)\,\boldsymbol{\Phi}_{4t}(L)\}, \\
\mathbf{B}(\mathbf{y_t}) &= \boldsymbol{\Gamma}_t,
\end{aligned}$$

with the upper part of $\boldsymbol{\Gamma}_t$, i.e. $\mathbf{G}_{12t}$, being given by (9).[16] Three results of the above SDC factorisation are worth highlighting. First, by construction the SDC factorization gives a representation of the model of the economy that is mathematically equivalent to the structural model in equation (7). Second, the SDC vectors $\mathbf{c}(\mathbf{y}_t)$ and the SDC matrices $\mathbf{A}(\mathbf{y}_t)$ are constructed adding through weighted coefficients across the four states, with weights given by $[1 - g_M(\mathbf{s}_u'\mathbf{y}_t)][1 - l_M(\mathbf{s}_x'\mathbf{y}_t)]$, $[1 - g_M(\mathbf{s}_u'\mathbf{y}_t)]\,l_M(\mathbf{s}_x'\mathbf{y}_t)$, $g_M(\mathbf{s}_u'\mathbf{y}_t)[1 - l_M(\mathbf{s}_x'\mathbf{y}_t)]$ and $g_M(\mathbf{s}_u'\mathbf{y}_t)\,l_M(\mathbf{s}_x'\mathbf{y}_t)$ for state $j = 1, 2, 3$ and 4, respectively. However, the SDC matrix $\mathbf{B}(\mathbf{y}_t)$ encapsulates the highly nonlinear structure of (9). Third, all SDC matrices in (11) depend on the previous period state vector. For this reason, these are unaffected in the current period by a change in policy.[17]

SDC factorizations are not new in macroeconomics. Guerrieri and Iacoviello (2017) use it to solve a DSGE model with dual nonlinearity from both the private and policy sector. Their solution takes the form of two SDC VAR models, depending on whether the ZLB constraint is binding or not.

3.2.2 State Dependent Coefficients Solution

Following the SDC factorization, the closed-loop regulator problem consists of finding the sequence $\{\mathbf{u}_{t+1}\}_{t=0}^{\infty}$ that minimizes (10) subject to (11) given $\mathbf{y}_0$. For convenience, this is re-written below:

$$\begin{aligned}
\min_{\{\mathbf{u}_{t+1}\}_{t=0}^{\infty}} V_0 &= E_0 \sum_{t=0}^{\infty} \beta^t \left[(\mathbf{y}_t - \overline{\mathbf{y}})'\,\mathbf{Q}\,(\mathbf{y}_t - \overline{\mathbf{y}})\right] \\
s.t. &: \mathbf{y}_{t+1} = \mathbf{c}(\mathbf{y}_t) + \mathbf{A}(\mathbf{y}_t)\,\mathbf{y}_t + \mathbf{B}(\mathbf{y}_t)\,\mathbf{u}_{t+1} + \mathbf{e}_{t+1}, \\
\mathbf{e}_{t+1} &\sim N(\mathbf{0}, \boldsymbol{\Sigma}) \text{ and } \mathbf{y}_0 \text{ given.}
\end{aligned}$$

[15] Wernli and Cook (1975) also consider the SDC factorization of the more general model $\dot{x} = f(x, u, t)$. This is given by $f(x, u, t) = A(x, u, t)\,x + B(x, u, t)\,u$, where $A(x, u, t)$ and $B(x, u, t)$ are matrices of coefficients, both depending on x, u and t.

[16] It is worth highlighting that the SDC factorization is very general, and it can be applied to many other nonlinear structures, including threshold models, artificial neural networks and bilinear time series models.

[17] An issue often highlighted in the engineering literature on the SDRE method, see Cloutier et al. (1996a) or Beeler (2004), is that SDC factorizations are not unique, since there are many nonlinear forms that can be encapsulated into the SDC matrices in equation (11). This is true when the SDC factorization is used to approximate an unknown nonlinear model. However, this issue does not apply here, since the starting point of the analysis is the regime changing description of the economy through the known VSTAR model in (1)-(2).

This formulation of the NLQR problem is isomorphic to the LQR except for the objects $\mathbf{c}(\mathbf{y}_t)$, $\mathbf{A}(\mathbf{y}_t)$ and $\mathbf{B}(\mathbf{y}_t)$, whose elements are functions of the state vector $\mathbf{y}_t$. Given this, the SDRE method treats the matrices $\mathbf{c}(\mathbf{y}_t)$, $\mathbf{A}(\mathbf{y}_t)$ and $\mathbf{B}(\mathbf{y}_t)$ as fixed in any given period $t \geq 0$ and solves the NLQR problem in that period as a LQR.[18] The Bellman equation for any $t \geq 0$ can be written as

$$V(\mathbf{y}_t) = \min_{\mathbf{u}_{t+1}} \left[(\mathbf{y}_t - \overline{\mathbf{y}})' \mathbf{Q} (\mathbf{y}_t - \overline{\mathbf{y}}) + \beta E_t V(\mathbf{y}_{t+1}) \right].$$

The value function is guessed to be quadratic and involving undetermined coefficients: $V(\mathbf{y}_t) = \mathbf{y}_t' \mathbf{P}_t \mathbf{y}_t + 2\mathbf{y}_t' \mathbf{p}_t + p_t$, where $\mathbf{P}_t$, $\mathbf{p}_t$ and p_t are a positive semi-definite symmetric matrix, a vector and a positive scalar, respectively, whose coefficients are fixed in any given period $t \geq 0$, but vary over time. After replacing the postulated solution on the right side of the Bellman equation and using (11) to compute expectations, differentiation with respect to $\mathbf{u}_{t+1}$ yields the SDRE solution to the NLQR problem:

$$\mathbf{u}_{t+1} = \mathbf{k}_t + \mathbf{K}_t \mathbf{y}_t \tag{12}$$

$$\mathbf{k}_t = -\left[\mathbf{B}(\mathbf{y}_t)' \mathbf{P}_t \mathbf{B}(\mathbf{y}_t) \right]^{-1} \mathbf{B}(\mathbf{y}_t)' \left[\mathbf{P}_t \mathbf{c}(\mathbf{y}_t) + \mathbf{p}_t \right], \tag{13}$$

$$\mathbf{K}_t = -\left[\mathbf{B}(\mathbf{y}_t)' \mathbf{P}_t \mathbf{B}(\mathbf{y}_t) \right]^{-1} \mathbf{B}(\mathbf{y}_t)' \mathbf{P}_t \mathbf{A}(\mathbf{y}_t), \tag{14}$$

where:

$$\begin{aligned} \mathbf{P}_t = {} & \mathbf{Q} + \beta \mathbf{A}(\mathbf{y}_t)' \mathbf{P}_t \mathbf{A}(\mathbf{y}_t) - \\ & \beta \mathbf{A}(\mathbf{y}_t)' \mathbf{P}_t \mathbf{B}(\mathbf{y}_t) \left[\mathbf{B}(\mathbf{y}_t)' \mathbf{P}_t \mathbf{B}(\mathbf{y}_t) \right]^{-1} \mathbf{B}(\mathbf{y}_t)' \mathbf{P}_t \mathbf{A}(\mathbf{y}_t), \end{aligned} \tag{15}$$

$$\mathbf{p}_t = \left\{ \mathbf{I} - \beta \left[\mathbf{K}_t' \mathbf{B}(\mathbf{y}_t)' - \mathbf{A}(\mathbf{y}_t)' \right] \right\}^{-1} \left\{ \begin{array}{c} \beta \left[\mathbf{A}(\mathbf{y}_t)' + \mathbf{K}_t' \mathbf{B}(\mathbf{y}_t)' \right] \\ \times \mathbf{P}_t \mathbf{c}(\mathbf{y}_t) - \mathbf{Q} \overline{\mathbf{y}} \end{array} \right\}, \tag{16}$$

$$p_t = (1-\beta)^{-1} \left\{ \begin{array}{c} \overline{\mathbf{y}}' \mathbf{Q} \overline{\mathbf{y}} + \beta \mathbf{c}(\mathbf{y}_t)' \mathbf{P}_t \mathbf{c}(\mathbf{y}_t) + \beta \mathbf{k}_t' \mathbf{B}(\mathbf{y}_t)' \\ \left[\mathbf{P}_t \mathbf{c}(\mathbf{y}_t) + \mathbf{p}_t \right] + \beta tr \left[\mathbf{\Sigma}_{t+1} \mathbf{P}_t \right] + 2\beta \mathbf{c}(\mathbf{y}_t)' \mathbf{p}_t \end{array} \right\}. \tag{17}$$

The optimal feedback rule (12) is a linear function of the state vector with time-varying coefficients, since the Kalman gain $\mathbf{K}_t$ in (14) and the intercept term $\mathbf{k}_t$ in (13) are both determined by SDCs matrices. The feedback rule (12) is derived as follows in any given $t \geq 0$. First solve the Riccati equation (RE) in equation (15) to obtain $\mathbf{P}_t$. Since the coefficients of $\mathbf{A}(\mathbf{y}_t)$ and $\mathbf{B}(\mathbf{y}_t)$ are predetermined, the RE equation (15) can be solved, for example, iterating until convergence on $\mathbf{P}_t$. This solution is a function of $\mathbf{y}_t$ and for this reason the RE has to be solved conditional on the value of the state vector $\mathbf{y}_t$. Second, calculate $\mathbf{K}_t$ from equation (14). Third, compute $\mathbf{p}_t$ from (16). Fourth, use the calculated values for $\mathbf{P}_t$ and $\mathbf{p}_t$ to compute $\mathbf{k}_t$ from (13). The optimal feedback policy can then be calculated from equation (12). This gives a time-varying sequence of $\mathbf{u}_{t+1}$ for $t \geq 0$. Finally, compute the sequence of p_t from (17) for $t \geq 0$.

The algorithm described above illustrates the simplicity and computational advantages of the SDRE method. In any period $t \geq 0$, given $\mathbf{y}_t$, the objects $\mathbf{c}(\mathbf{y}_t)$, $\mathbf{A}(\mathbf{y}_t)$ and $\mathbf{B}(\mathbf{y}_t)$ can be regarded as fixed and the feedback rule coefficients in that period can be computed upon iteration of (15), which is effectively a standard algebraic Riccati equation. It is worth observing that in a LQR problem where the constraint of the regulator is a system with fixed coefficients (i.e.

[18] Appendix B derives the solution step-by-step.

$\mathbf{c}(\mathbf{y}_t) = \mathbf{c}$, $\mathbf{A}(\mathbf{y}_t) = \mathbf{A}$ and $\mathbf{B}(\mathbf{y}_t) = \mathbf{B}$), the value function parameters $\mathbf{P}$, $\mathbf{p}$, and p can be computed *offline* before the policy rule in (12) is implemented, since (15)-(17) depend only on the parameters of the objective function and the time-invariant coefficients $\mathbf{c}$, $\mathbf{A}$ and $\mathbf{B}$. In contrast, the derivation of the SDRE solution described above can only be computed *online,* since the parameters of the constraint in (11) depend on the state vector, which in turn changes over time due to the implementation of the feedback rule (12). The *online* solution works as follows. In period $t = 0$, equations (13)-(17) are solved given $\mathbf{y}_0$ and (12) is used to compute $\mathbf{u}_1$. For $t > 0$, two options are available, depending on whether the state vector is updated with simulated or observed $\mathbf{y}_t$'s. The first consists of replacing the optimal $\mathbf{u}_t$ into (11) to update the state and compute $\mathbf{y}_{t+1}$, which can then be used for the next stage of optimization to derive $\mathbf{u}_{t+1}$, and so on. The second option is only feasible for in-sample counterfactual simulation since it consists of updating in every period the state vector using its observed value. Clearly, the deviations between the observed and the optimal paths of $\mathbf{u}_t$ are smaller when the updating is based on the observed $\mathbf{y}_t$'s. Constraints on some of the policy instruments included in $\mathbf{u}_t$, like for example the nonnegativity of the federal funds rate in Section 5, are implemented in the online solution directly upon computation of the policy vector $\mathbf{u}_t$ in each $t \geq 0$.

3.3 Reduced Form

The closed-loop solution in equations (12)-(14) takes the form of a linear feedback rule with time-varying matrices of coefficients, $\mathbf{K}_t$ and $\mathbf{k}_t$. In any $t \geq 0$, these can be combined with the equations for the economy in SDC form in (11) as:

$$\begin{bmatrix} \mathbf{I} & -\mathbf{B}(\mathbf{y}_t) \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{u}_{t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{c}(\mathbf{y}_t) \\ \mathbf{k}_t \end{bmatrix} + \begin{bmatrix} \mathbf{A}(\mathbf{y}_t) \\ \mathbf{K}_t \end{bmatrix} \mathbf{y}_t + \begin{bmatrix} \mathbf{e}_{t+1} \\ \mathbf{0} \end{bmatrix}. \quad (18)$$

The system (18) can be written as a reduced-form model with time varying coefficients:

$$\mathbf{y}_{t+1} = \boldsymbol{\lambda}_t^* + \boldsymbol{\Lambda}_t^*(L)\,\mathbf{y}_t + \mathbf{e}_{t+1}^*, \quad (19)$$

where

$$\boldsymbol{\lambda}_t^* = \begin{bmatrix} \mathbf{c}(\mathbf{y}_t) + \mathbf{B}(\mathbf{y}_t)\,\mathbf{k}_t \\ \mathbf{k}_t \end{bmatrix}, \boldsymbol{\Lambda}_t^*(L) = \begin{bmatrix} \mathbf{A}(\mathbf{y}_t) + \mathbf{B}(\mathbf{y}_t)\,\mathbf{K}_t \\ \mathbf{K}_t \end{bmatrix}, \mathbf{e}_t^* = \begin{bmatrix} \mathbf{e}_t \\ \mathbf{0} \end{bmatrix}$$

and $\mathbf{e}_t \sim (\mathbf{0}, \boldsymbol{\Sigma})$, with covariance matrix $\boldsymbol{\Sigma}$ being defined in (9). Equation (19) is the VSTAR model under the optimal feedback rule, written in SDC form. This differs from the reduced-form VSTAR in (1) and (2) in three respects. First, the coefficients corresponding to the equations for the economy in the VSTAR in (1) and (2) are replaced with a linear combination of the corresponding coefficients from the structural model (7)-(9) and the policy rule coefficients (see the top parts of $\boldsymbol{\lambda}_t^*$ and $\boldsymbol{\Lambda}_t^*(L)$). Second, the coefficients corresponding to the equations for the policy vector in the VSTAR in (1) and (2) are replaced with those in (12). Third, the variance structure of the VSTAR in (19) is constant, while it is time-varying and state-dependent in the the original VSTAR in (1) and (2).

It is worth noting that the system (11) can be used to study the effects of any policy rule including those with ad-hoc coefficients. To this end it is sufficient to replace the coefficients $\mathbf{K}_t$ and $\mathbf{k}_t$ in (18) with any given values, say $\widetilde{\mathbf{K}}_t$ and $\widetilde{\mathbf{k}}_t$, and then derive the corresponding reduced-form system as in (19).

3.4 Stability

The VSTAR model under the optimal feedback rule in (19) describes a closed-loop system with time-varying parameters, since the feedback rule coefficients, $\mathbf{k}_t$ and $\mathbf{K}_t$, and the coefficients of the regulator constraint, $\mathbf{c}(\mathbf{y}_t)$, $\mathbf{A}(\mathbf{y}_t)$ and $\mathbf{B}(\mathbf{y}_t)$, are time varying. This system is asymptotically stable if $\mathbf{Q}$ is positive semidefinite and the pair $\{\mathbf{A}(\mathbf{y}_t), \mathbf{B}(\mathbf{y}_t)\}$ is stabilizable for all $\mathbf{y}_t$ over $t \geq 0$ (sufficient conditions). Stabilizability requires that $\mathbf{y}_t$ converges to a fixed point at $t \to \infty$, given any starting value $\mathbf{y}_0$. For any $t \geq 0$, assume the eigenvalues of $\mathbf{\Lambda}_t^*(L)$ to be distinct and consider the time-varying eigenvalue decomposition $\mathbf{\Lambda}_t^*(L) = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t^{-1}$, where each column of $\mathbf{D}_t$ is an eigenvector of $\mathbf{\Lambda}_t^*(L)$ and $\mathbf{R}_t$ is a diagonal matrix of eigenvalues of $\mathbf{\Lambda}_t^*(L)$. Ignoring the intercept and considering only its deterministic part, the closed-loop system (19) can be written as $\mathbf{y}_{t+1} = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t^{-1} \mathbf{y}_t$ in any $t \geq 0$. The solution to this difference equation in any given period $t \geq 0$ can be calculated using backward substitution, which yields $\mathbf{y}_j = \mathbf{D}_t \mathbf{R}_t^j \mathbf{D}_t^{-1} \mathbf{y}_0$.[19] This is stable in any given period $t \geq 0$ and for any initial condition $\mathbf{y}_0$ if and only if in that period the eigenvalues in the matrix $\mathbf{A}(\mathbf{y}_t) + \mathbf{B}(\mathbf{y}_t)\mathbf{K}_t$, i.e. the diagonal elements of $\mathbf{R}_t$, are all strictly less than unity in absolute value.

In general, as long as the matrix $\mathbf{Q}$ is positive semidefinite it is always possible for the control solution to stabilize the pair $\{\mathbf{A}(\mathbf{y}_t), \mathbf{B}(\mathbf{y}_t)\}$. Even if the open-loop system (i.e. the VSTAR) is highly unstable in a given period $t \geq 0$, for example dysplaying one or more explosive roots, then the closed-loop system is still stable. Convergence may show poor dynamics in the sense that the control may display an erratic initial path and large swings, thereby requiring the system longer time to settle. If the instruments vector is subject to constraints, these may be binding in some $t \geq 0$ therefore further deteriorating the performance of the control solution. Lewis, Vrabie and Syrmos (2012) suggest that asymptotic stability can be insured more efficiently by employing in any period in which the pair $\{\mathbf{A}(\mathbf{y}_t), \mathbf{B}(\mathbf{y}_t)\}$ is highly unstable another stabilizable pair, $\{\overline{\mathbf{A}}(\mathbf{y}_t), \overline{\mathbf{B}}(\mathbf{y}_t)\}$, whose roots are more stable than the original $\{\mathbf{A}(\mathbf{y}_t), \mathbf{B}(\mathbf{y}_t)\}$. To construct the stabilizable pair $\{\overline{\mathbf{A}}(\mathbf{y}_t), \overline{\mathbf{B}}(\mathbf{y}_t)\}$, one can use the time-invariant pair $\{\mathbf{A}, \mathbf{B}\}$ obtained from the time-invariant version of (1)-(2), i.e. the VAR. If this pair is stabilizable than it can be used as replacement for $\{\mathbf{A}(\mathbf{y}_t), \mathbf{B}(\mathbf{y}_t)\}$ in any period $t \geq 0$ when this is either highly unstable or not stabilizable at all.

Arguably, the choice of the stabilizable pair $\{\overline{\mathbf{A}}(\mathbf{y}_t), \overline{\mathbf{B}}(\mathbf{y}_t)\}$ is not unique and the time-invariant pair $\{\mathbf{A}, \mathbf{B}\}$ is not necessarely the most efficient. For example, the control algorithm could be executed one-period ahead only over a given time horizon, say $t \in (0, T)$ and several stable $\{\mathbf{A}(\mathbf{y}_t), \mathbf{B}(\mathbf{y}_t)\}$ pairs could be identified. Each of these, or some linear combination of them, could be used as stabilizable pair. The preferred pair may be chosen as that yielding less volatility. However, the time-invariant pair $\{\mathbf{A}, \mathbf{B}\}$ has three advantages compared to these alternatives. First, it can be computed offline before knowing the state of the economy and before computing the optimal control rule, as it only requires knowledge of the time invariant matrices $\mathbf{A}$ and $\mathbf{B}$. Second, it is faster to implement as it does not require any preliminary assessment of the many possible stable pairs. Third, the stabilizing pair $\{\mathbf{A}, \mathbf{B}\}$ exists as long as

[19] Note that this result is derived in any period $t \geq 0$ keeping $\mathbf{R}_t$ and $\mathbf{D}_t$ fixed and compounding over the upper script j, starting from a given $\mathbf{y}_0$.

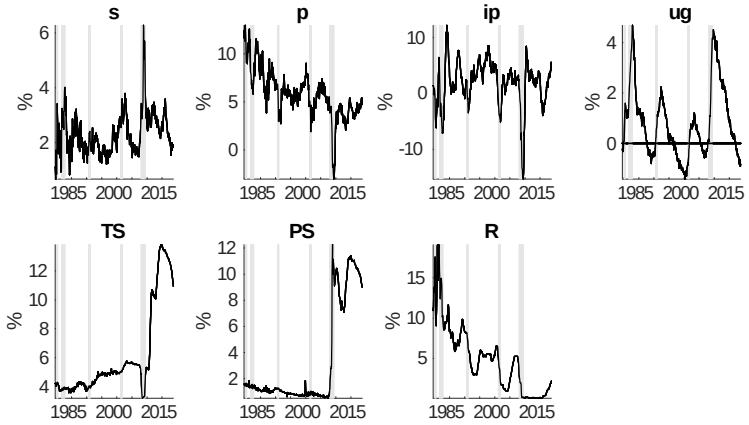

Figure 1: Aggregate monthly data, United States, 1979:8 to 2018:10. NBER recessions in grey.

a stable VAR can be inferred from the VSTAR, while there is not guarantee that a stable $\{\mathbf{A}(\mathbf{y}_t), \mathbf{B}(\mathbf{y}_t)\}$ pair can be identified over $t \in (0, T)$.

Given the above, stability of the VSTAR under control in the numerical analysis is monitored as follows. First, the numerical algorithm implementing the SDRE solution in (12)-(16) is applied. If the resulting loss is lower that that obtained from the simulation of the VSTAR, then the closed-loop solution is kept. Otherwise, the solution is recomputed using the stabilizing pair $\{\mathbf{A}, \mathbf{B}\}$ from the VAR in any period $t \geq 0$ in which the VSTAR under control in (19) is unstable. If this still delivers an higher loss, then the the solution is recomputed using the stabilizing pair $\{\mathbf{A}, \mathbf{B}\}$ from the VAR in any period $t \geq 0$ in which the open-loop pair $\{\mathbf{A}(\mathbf{y}_t), \mathbf{B}(\mathbf{y}_t)\}$ is unstable.

4 Estimated VSTAR

This section describes the data used for the empirical analysis and the steps undertaken to specify, estimate and validate the VSTAR model used for the subsequent analysis of the optimal monetary policy.

4.1 Data

The VSTAR model is estimated using aggregate monthly data for the United States from 1979:8 to 2018:10.[20] The economy vector includes four variables: the credit risk, measured as the spread between the Baa corporate rate and the 10-year treasury rate (s); the inflation rate (p), measured as the annual change in the personal consumption expenditures deflator; the growth rate of industrial production (ip); and the difference between the actual and the natural rate of unemployment, the unemployment gap (ug). The policy vector includes

[20] Details on the sources of the raw data and how these are transformed to obtain the time series used for the quantitative analysis are given in Appendix C.

three variables: Treasury securities held by the Fed as a proportion to GDP (TS); other private securities held by the Fed as a proportion to GDP (PS); and the federal fund rate (R). The selected measure of credit risk is used as a proxy for financial conditions in the macroeconomic system.[21] Inflation and the unemployment gap are used as indicators for the Fed targets for price stability and maximum employment, respectively, as in Federal Reserve Bank of St. Louis (2015). The sum of TS and PS equals total assets held in the Fed's balance sheet, as a proportion to GDP. Thus the specified VSTAR includes conventional and QE monetary policy tools, and can account for variation in both the size of the Fed's balance sheet and the composition of its assets portfolio.

Figure 1 plots the data, highlighting in grey periods of economic recessions measured by the NBER.[22] Several clear irregularities are visible over the sample period: the large spikes in the credit risk indicator, inflation, industrial production and unemployment gap data; the asymmetric dynamics of the unemployment gap above and below zero; the large transitional changes in the QE variables, TS and PS, around the Great Recession; and the flat path of the federal fund rate while at the ZLB. Most of the applied macroeconomic literature uses VSTAR to capture asymmetries in the business cycle of the economy.[23] The asymmetric dynamics of the unemployment gap also capture this. The clear irregularities observed in the dynamics of the policy instruments further motivate the use of a nonlinear model like the VSTAR in (1) and (2) that can account for nonlinearity in both the economy and policy vectors.

The sample covers three key periods of the monetary economic history of the United States. The first is from the appointment of Paul Volcker as Fed Chairman in August 1979 until around the beginning of Great Recession in September 2008. Over this period the main instrument of monetary policy is federal funds rate. The size of the Fed's balance sheet remaines relatively constant, being on average about 6 percent of GDP. The composition of the assets portfolio gradually changes, with the share of TS increasing from about 70 to about 90 percent of the total portfolio by mid 2007. The second period covers the Great Recession and the post-crisis slow recovery, ending just before the Fed announcement of the normalization program, in December 2015. Throughout this phase the federal fund rate is near zero (below 25 base points) and the Fed's portfolio of assets increases in size, reaching about 25 percent of GDP, and changes in composition, as PS cover on average about half of the portfolio due to the large purchases of mortgage-backed assets.[24] The third phase of the United States monetary policy covered by the sample period includes the early stage of the so-called normalization program, between January 2016 (when the federal funds rate rises above 25 base points, 0.34 percent, for the first time after

[21] Galvão and Owyang (2018) use a factor VSTAR model to derive a financial conditions factor that is highly correlated with macroeconomic and credit risk variables of the United States economy. The indicator s is the measure of credit risk that display the highest correlation with their financial conditions factor, see Table 2 in Galvão and Owyang (2018). A popular alternative in the QE literature to proxy credit risk is the VIX. In contrast to those required to construct s, time series of the VIX are only available from the 1990s onward.

[22] According to the NBER, over the sample period the United States economy was in recession five times: between January-July 1980, July 1981-November 1982, July 1990-March 1991, March-November 2001 and December 2007-June 2009.

[23] See, e.g., Auerback and Gorodnichenko (2012), Caggiano, Castelnuovo and Groshenny (2014), Caggiano et al (2015), Galvão and Owyang (2018).

[24] A detailed description of the different QE programs undertaken by the Fed since 2008 can be found in Kuttner (2018).

seven years) until the end of the sample. By October 2018, the federal funds rate has reached 2.4 percent, the size of the balance sheet has declined to about 20% of GDP, while the assets portfolio composition has remained diversified with TS averaging about 54.5 percent of total assets. This latest phase of monetary policy in the United States is therefore characterized by a low federal funds rate, yet away from the ZLB, and a size of the Fed's balance sheet larger than in the decades preceding the Great Recession. Whether this is a transitional period or the beginning of a new policy configuration is early to establish. Nevertheless this feature of the data will be reflected in the structure of the transmission mechanism estimated through the VSTAR and in the analysis of optimal policy.

It is important to highlight at this stage a possible limitation the chosen breakdown for the balance sheet variables, since this does not account for the maturity structure of the assets held by the Fed. In principle, this could be included in the VSTAR, at least to a first-order approximation, by separating short- and long-term holdings of TS. This would however require a large increase in the number of coefficients of the VSTAR, making the estimation untractable. Thus the optimal policy analysis can only consider whether in any given period TS and PS holdings should increase, decrease or remain unchanged, The analysis is however silent on whether the maturity structure of assets within these two cathegories should also change. This should be kept in mind while interpreting the results in Section 5. Nevertheless, the distinction between TS and PS is still important, both in theory and practice. As noted in the Introduction, this distinction is highlighted in the economic theory on the transmission mechanism of QE, where changes of TS and PS affect aggregate demand through the portfolio-balance and the credit channel, respectively. According to Kuttner (2018), TS and PS are also key tools of QE policy in the United States which is focused on the asset side of the balance sheet.

4.2 Specification

Variables are ordered as $\mathbf{y}_t' = \begin{bmatrix} s_t & \pi_t & ip_t & ug_t & TS_t & PS_t & R_t \end{bmatrix}$. As in Dahlhaus, Hess and Reza (2018) and Galvão and Owyang (2018), the VSTAR lag lenght is equal to one, since a larger number of lags would make the estimation of the VSTAR parameters unfeasible.

Estimation requires the specification of the transition functions $l_i(\mathbf{s}_x'\mathbf{y}_t)$ and $g_i(\mathbf{s}_u'\mathbf{y}_t)$, $i = M, V$. The function $l_i(\mathbf{s}_x'\mathbf{y}_t)$ is set to capture nonlinearity due to the amount of slack in the economy, measured using the unemployment gap. Thus $\mathbf{s}_x' = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ and $l_i(\mathbf{s}_x'\mathbf{y}_t) = l_i(ug_t)$. This is similar to Ramey and Zubairy (2018), except that in their work the natural unemployment rate is constant at 6.5 percent, while it is time-varying in this paper following the definition of unemployment gap adopted by Federal Reserve Bank of St. Louis (2015). The unemployment gap is not the only possible measure of economic slack. Auerbach and Gorodnichenko's (2012), for example, employ the moving average of GDP growth. Using this measure would however significantly cut the sample size, since monthly observations of real GDP are only available from the early 1990s.[25] The function $g_i(\mathbf{s}_u'\mathbf{y}_t)$ is set to capture changes in the VSTAR

[25] As explained in Appendix C, an estimate of the monthly GDP is still needed in order to scale the QE variables before the 1990s. To this end, a series of monthly nominal GDP is interpolated from available quarterly data. In principle, this could be added to the available data on monthly GDP and then employed as transition variable. This however would affect

parameters depending on whether or not the economy is near to the ZLB. Thus $\mathbf{s}_u' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ and $g_i(\mathbf{s}_u'\mathbf{y}_t) = g_i(R_t)$.

Transition across states due to nonlinearity in the economy is modelled using a logistic function $l_i(ug_t) : (-\infty, +\infty) \rightarrow [0,1]$ described as

$$l_i(ug_t) = \{1 + \exp[-\gamma_{li}(ug_t - c_i)]\}^{-1}, \tag{20}$$

where $\gamma_{li} > 0$ denotes the speed of adjustment across states for the parameters of $\mathbf{y}_t$, $i = M$, and $\mathbf{\Omega}_t$, $i = V$; c_i indicates the threshold parameter. Since the federal funds rate cannot be negative, transition across states due to nonlinearity in the policy vector is instead modelled with the incomplete gamma function $g_i(R_t) : (0, +\infty) \rightarrow [0,1]$ described as:

$$g_i(R_t) = \frac{1}{\Gamma(\gamma_{gi})} \int_0^{R_t} e^{-t} t^{\gamma_{gi}-1} dt, \tag{21}$$

where $\Gamma(\gamma_{gi})$ is the gamma function with shape parameter γ_{gi} for $\mathbf{y}_t$, $i = M$, and $\mathbf{\Omega}_t$, $i = V$; and the implied scale parameter has been set equal to one. Logistic functions are widely employed in the applied macroeconomic literature to capture nonlinearity stemming from the economy, see, e.g., Auerback and Gorodnichenko (2012), Caggiano et al. (2015) and Galvão and Owyang (2018). The incomplete gamma function is used to capture nonlinearity driven by a variable that is nonnegative, being for example volatility, as in Lanne and Saikkonen (2005), or the nominal rate of interest, as in Hurn et al. (2018). To discipline the economic interpretation of the estimated paremeters, the thresholds in (20) are set equal to zero, i.e. $c_i = 0$, $i = M, V$. This means that the nonlinearity stemming from the economy vector depends on whether the actual rate of unemployment is above or below the natural rate of unemployment or, in line with Ramey and Zubairy (2018), equivalently the amount of slackness in the economy.

To impose the positive definiteness restriction on the covariance matrix while simultaneously reducing the dimension of the parameter vector, the estimation uses the covariance matrix decomposition based on the BEKK model of Engle and Kroner (1995). This sets $\mathbf{\Omega}_j = \widetilde{\mathbf{\Omega}}_j' \widetilde{\mathbf{\Omega}}_j$ in (2), where each $\widetilde{\mathbf{\Omega}}_j$ is a lower triangular matrix, $j = 1, ..., 4$.

As well as estimating the unrestricted model (1)-(2), five alternative specifications are considered for comparison, each being nested in the unrestricted VSTAR: (i) VSTARC, which constraints the transition parameters in the logistic and gamma function for the mean and variance to be the same, i.e. $\gamma_{lM} = \gamma_{lV}$ and $\gamma_{gM} = \gamma_{gV}$; (ii) VSTARV, which allows for nonlinearity only in the variance structure but keeps a constant mean; (iii) VSTARM, with allows for nonlinearty in the mean but has contant variance; (iv) VSTARP, which restricts each equation for the QE instruments to follow an AR(1) when the federal funds rate is at the ZLB, and the equation for the the federal funds rate to follow an AR(1) during the ZLB period; (v) VAR, with constant mean and variance. The VSTARC specification imposes restrictions on the transition variables similar to those used by Auerback and Gorodnichenko (2012) and Galvão and Owyang (2018). The VSTARP specification restricts the dynamics so that QE is exogenous when

the estimate of nonlinearity by any possible approximation error from the interpolation.

conventional monetary policy is active, and viceversa, as assumed by Hurn et al. (2018) and Sims and Wu (2019a,b). The VSTARM and VSTARV specifications restrict nonlinearity to the mean and variance, respectively. The VAR gives a linear benchmark against which to compare any nonlinear specification.

4.3 Estimation

Each model is estimated with full information maximum likelihood using standard iterative algorithms. Assuming a normal distribution for the disturbance terms, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{\Omega}_t)$, the log-likelihood function of $\mathbf{y}_t$ at observation $t \in (0, T)$, for any of the VSTAR models described above with parameter vector Π is $\ln l_t(\Pi) = -\frac{n}{2}\ln 2\pi - 0.5\ln|\mathbf{\Omega}_t| - 0.5\mathbf{v}_t'\mathbf{\Omega}_t^{-1}\mathbf{v}_t$. The full information maximum likelihood estimator Π_U for a sample T conditioning on the first $q = 1$ observations is calculated by maximizing the conditional log-likelihood function $\ln L_T(\Pi) = [2(T-q)]^{-1}\sum_{t=q+1}^{T}\ln l_t(\Pi)$.[26]

Table 1 ranks each estimated model in terms of its log-likelihood, reported in the second column. The following three columns report the number of parameters (NoP) estimated for each model; the likelihood ratio (LR) test comparing the log-likelihood of each restricted model against the unrestriced VSTAR; and the critical value of the chi-square distribution at 5 percent. The last three columns report the Akaike (AIC), the Hannan and Quinn (HIC) and the Schwarz (SIC) information criteria, which penalize the log-likelihood of each model taking into account its dimensionality.

Overall, the results in Table 1 validate the use of the specified VSTAR, as this provides the best fit of the data along most criteria. All five restricted specifications are rejected against the unrestricted VSTAR under the LR test. The VSTAR is also selected as the preferred specification under the AIC and the HIC, while it ranks below the VSTARP only according to the SIC. It is interesting to observe that all specifications allowing for nonlinearity in the variance (the first four in Table 1) have a log-likelihood higher than the two homoscedastic specifications (VSTARM and VAR). This points towards the relative importance of allowing for nonlinearity in the volatility, as also highlighted in Polito and Spencer (2016). Perhaps unsurprisingly given the data dynamics observed in Figure 1, the VAR provides in general the worst fit under almost any diagnostic considered in Table 1.

The top panels in Figure 2 illustrates the evolution of the transition probabilities determining variation in the coefficients in the mean (top-left) and the variance (top-right) of the estimated VSTAR. The black lines show the probability of the economy being in a slack state, $l_i(ug_t) \simeq 1$, or not being in a slack state, $l_i(ug_t) \simeq 0$. The red lines those of policy being away from the

[26] The VSTAR is first estimated using as starting values the coefficients from the OLS VAR. All other restricted models are estimated after imposing the required restrictions on the coefficients estimated from the VSTAR and then using these as starting values. Restricting the transition functions to zero returns the OLS VAR coefficients. The VSTAR coefficients are re-estimated after normalizing the transtion variables ug_t and R_t to have zero mean and unit variance, but this does not lead to significant differences in the parameter estimates. The log-likelihood function is maximized using the quasi Newtown method, the default option in MATLAB fminunc.

Model	lnL_T	NoP	LR	$\chi^2_{NoR}(5\%)$	AIC	HIC	SIC
VSTAR	0.0089	340			1.43	2.61	4.43
VSTARP	-0.1395	302	139.16	51	1.57	2.63	4.26
VSTARC	-0.1704	338	29.04	5.99	1.78	2.95	4.77
VSTARV	-0.8025	161	592.94	209.04	2.30	2.87	3.74
VSTARM	-1.6298	224	1528.72	1368.90	4.22	5.01	6.22
VAR	-2.2982	77	2155.68	1402.93	4.92	5.19	5.61

Notes: $LR = -2\left(T-q\right)\left[lnL_T\left(\Pi_R\right) - lnL_T\left(\Pi_U\right)\right]$; $AIC = -2lnL_T\left(\Pi\right) + \frac{2k}{T-q}$; $HIC = -2lnL_T\left(\Pi\right) + 2k\ln\left[\frac{\ln(T-q)}{T-q}\right]$; $SIC = -2lnL_T\left(\Pi\right) + k\frac{\ln(T-q)}{T-q}$; $T = 471$; $q = 1$; $k = NoP$.

Table 1: Maximum likelihood estimates of the unrestricted VSTAR model and five alternative restricted specifications.

ZLB, $g_i\left(R_t\right) \simeq 1$, or near to the ZLB, $g_i\left(R_t\right) \simeq 0$. NBER recession periods are also highlighted in grey for reference. With regard to nonlinearity in the mean (top left panel), four sharp shifts in the VSTAR coefficients can be observed, all linked to the economy vector and occurring around (NBER) recessions. After 2001, nonlinearity stemming from the policy instruments becomes significant too, in particular during 2009-2015, when the federal funds rate is close to the ZLB. With regard to volatility (top right panel), changes in the coefficients due to nonlinearity in the economy vector also occur around recessions. In contrast, coefficient changes driven by R_t increase in frequency over the whole sample period, reflecting the downward trend in the federal funds rate observable in Figure 1.

The bottom panels in Figure 2 report scatter diagrams of $l_M\left(ug_t\right)$ vs. $l_V\left(ug_t\right)$ (bottom left) and $g_M\left(R_t\right)$ vs. $g_V\left(R_t\right)$ (bottom right). These display the correlation between coefficient changes in the mean and variance. It can be clearly observed that changes in coefficients linked to the rate of unemployment in the mean and variance of the VSTAR tend to comove, with transition probability oscillatting above and below 0.5. Coefficient changes linked to the federal funds rate in the mean and variance of the VSTAR tend to comove too, though gradually shifting over the whole sample period in line with the downward trend of the federal fund rate.

4.4 Impulse Response Functions

IRF analysis is carried out to verify whether the estimated VSTAR can deliver dynamics of the policy instruments in response to economic shocks consistent with the economic theory. If this is the case, then the estimated VSTAR can be regarded as a reliable starting point for the subsequent analysis of optimal policy.

The transmission mechanism of shocks in the VSTAR depends on the magnitude and size of the shocks, as well as the state of the economy when a shock occurs. The IRFs are nonlinear, since they allow for regime changing in the mean and variance over the simulation horizon and the ZLB constraint is enforced on by restricting the IRFs of the simulated federal funds rate to be nonnegative. The IRFs are also structural because they are computed from the simulation of

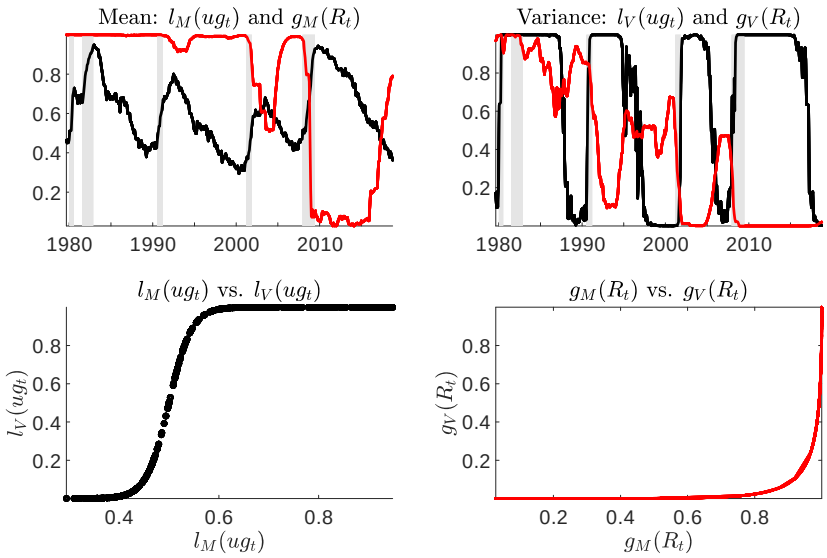

Figure 2: Transition probabilities estimated from the VSTAR.

the structural rather than the reduced-form VSTAR.[27]

Existing analyses of IRFs with VSTAR are based on the algorithm for calculating generalized IRFs of Koop, Pesaran and Potter (1996).[28] This identifies shocks through the Cholesky factorization of the VSTAR covariance matrix $\mathbf{\Omega}_t$ in each first period when a shock hits the economic system. While in small scale VAR the Cholesky decomposition yields structural shocks that can be easily reconciled with an economic interpretation, this is not necessarily the case in larger (and nonlinear) models like the VSTAR in (1)-(2). For this reason, identification for IRF analysis is carried out using restrictions.[29][30]

The sign restrictions are set to identify a demand and a supply shock. The demand shock is such that both the spread and the unemployment gap increase while inflation and industrial production decrease at the one month horizon.[31] The supply shock is such that inflation and unemployment increase while in-

[27] For these reasons these are nonlinear structural IRFs, following from Kilian and Lütkepohl (2017).

[28] A detailed account of how this algorithm is implemented in the VSTAR is given by Hubrich and Teräsvirta (2013).

[29] Appendix D describe the algorithm used for computing the IRFs. In essence, in each instance when the algorithm requires the computation of the Cholesky factor, this is replaced with a set of modified Cholesky factors obtained by post-multiplying the original one by an orthogonal matrix, which is generated from the QR decomposition of a matrix of random elements of the same size of the covariance matrix in the VSTAR.

[30] Bruns and Piffer (2017) also identify shocks in a VSTAR using sign restrictions. However, in their VSTAR the transition function applies directly to the triangular decomposition of the covariance matrix rather than the covariance matrix as a whole. In other words, Bruns and Piffer estimate the structural VSTAR directly, rather than recovering it from the unrestricted model as in the present paper and in most of the VSTAR literature.

[31] The sign restrictions on p, ip, and ug are standard in the identification of demand shocks. The sign on the spread is based on the observation that the sample correlation coefficients of the spread with p, ip, and ug are -0.543, -0.607 and 0.468, respectively.

dustrial production falls at the one month horizon. To evaluate the different responses during periods when either conventional monetary policy or QE is active, unconditional IRFs are computed using as starting values for $\mathbf{y}_t$ the averages over two sub-samples of equal size, following from Kilian and Lütkepohl (2017). The first is from 1994:6 to 2001:5. This is referred to as the normal times, capturing a period of the United States economic history when the federal funds rate is well above the ZLB constraint.[32] The second set of unconditional IRFs uses as starting values observations for the ZLB period, from 2009:1-2015:12.[33]

Figures 3 plots the IRFs of the VSTAR policy instruments plus the total size of the Fed's balance sheet (TS+PS) to demand and supply shocks over a 24 months horizon. In each panel, black lines denote the IRFs during normal times, red lines the IRFs for the ZLB period, and shaded areas around the IRFs denote two-standard deviations confidence bands.

The response of the policy instruments to a demand shock in the top panels of Figure 3 reflect the expected form of expansionary monetary policy intervention at different times. During normal times a demand shock leads to a reduction of the federal funds rate, while QE shows little dynamics. During the ZLB the monetary expansion occurs through QE. It is remarkable how the path of the IRFs for the two QE variables reproduces changes in the Fed assets' portfolio similar to those observed after QE1-QE2. These entail an initial expansion through the increase of PS holding, subsequently supported by a large increase of the TS holding.[34] These changes in the composition of the Fed's balance sheet during QE1 and QE2, consisting of a large initial purchase of PS followed by TS, are also visible from the dynamic of TS and PS around the Great Recession in Figure 1.

The IRFs to the supply shock in the bottom panels of Figure 3 display two interesting features. As for the demand shock, these show how different policy instruments operate at different time periods (active conventional monetary policy and passive QE during normal, passive conventional monetary policy and active QE at the ZLB). Differently from the demand shock, the monetary policy response to a supply shock is contractionary during normal times while it is expansionary during the ZLB period. This reflects the fact that in the data large changes in the Fed's balance sheet are associated with the slow real recovery in the aftermath of the Great Recession, which is captured by the estimated VSTAR.

Overall, the IRF analysis suggests that the estimated VSTAR can well replicates key features of the typical conduct of monetary policy in the United States during normal times and during the ZLB years.

[32] Note how the transition function $g_M(R_t)$ in the top-left panel of Figure 2 is close to one for most of the 1990s, which suggests that the policy rate is far from the ZLB over this period of time.

[33] Table 6 at the end of Appendix D reports all starting values used for calculating the IRFs.

[34] QE1 lasted 17 months. The Fed initially purchased private securities, consisting of $500 billion in agency mortgage-backed securities (MBS) and $100 billion in agency debt from Fannie Mae, Freddie Mac and Federal Home Loan Banks. Later the Fed purchased an additional $750 billion of MBS, $100 billion in agency debt, and $300 billion in long-term Treasuries. QE2 lasted for 8 months. During this phase, the Fed purchased a total of $767 billion of long-term Treasuries. See Kuttner (2018), for further details.

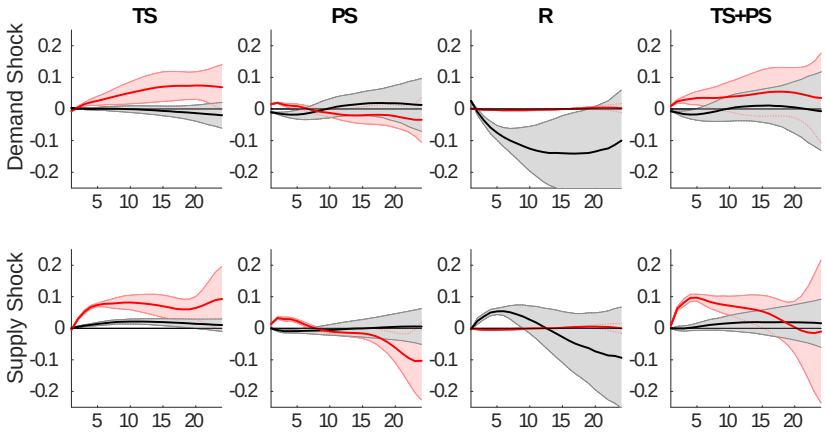

Figure 3: Unconditional IRFs of the policy instruments to demand and supply shocks during normal times (black) and the ZLB period (red). Shaded areas are two standard deviations confidence bands.

5 Optimal VSTAR

Quantitative analyses of optimal macroeconomic policy typically accomplish three tasks: evaluating the gains from the optimization of policy; studying the economy response to shocks once the optimal policy is implemented; undertaking counterfactual simulations to compare the actual macroeconomic dynamics against the optimal ones. Each of these tasks is carried out here after applying the methodology for computing optimal policy in Sections 2 and 3 starting from the estimated VSTAR described in Section 4. Once optimal rules are computed, macroeconomic dynamics can be simulated using the VSTAR under the optimal policy in (19). This is used to derive a benchmark against which to compare the dynamics observed from the data, which are a reflection of the actual monetary policy undertaken.

The benchmark dynamics obtained under the optimal policy are not unique. These crucially depend on the assumptions on the optimization protocol followed by the Fed (whether commitment or discretion); the constraints faced by Fed; the preferences of the Fed in terms of objective function and targets; and the number of available instruments. The analysis here considers solutions under commitment for joint optimization (optimal coordination) of both the federal funds rate and the QE instruments. Thus the economy vector is specified as $\mathbf{x}_t' = \begin{bmatrix} s_t & \pi_t & ip_t & ug_t \end{bmatrix}$ and the policy vector as $\mathbf{u}_t = \begin{bmatrix} TS_t & PS_t & R_t \end{bmatrix}$. As highlighted in section 2.2, the coefficients from the estimated VSTAR are used to recover the structural coefficients for the nonpolicy variables $\mathbf{x}_t$ under the timing assumption A2. Therefore, the structural model for the nonpolicy variables forming the constraint faced by the Fed is equivalent to that in (7)-(9).

The preferences of the Fed are specified in terms of minimization of a discounted intertemporal quadratic function that includes two main components. The first penalizes deviations of the inflation rate and the unemployment gap form their respective targets. This is consistent with the Fed dual mandate of

price stability and maximum employment. An alternative would be to use a quadratic loss function derived from a fully micro-founded DSGE model of the economy. As noted by Sims and Wu (2019b), DSGE models require heterogeneous agents for QE to be effective in equilibrium. For this reason, derivation of an aggregate objective function from a DSGE model would not be unique, since this would necessarily rely on ad-hoc weights attached to different agents. The second component of the loss function penalizes changes in the monetary policy instruments. This allows to control for different degrees of gradualism in the use of each policy instrument, mimicking for example the gradual movements of the federal funds rate observed in several periods over 1979-2018, as well as the sluggish evolution of the QE variables before the Great Recession. At the same time, this second component ensures that the path of the economy and policy instruments under the optimal rules does not deviate too much from that observed in the data, therefore containing the quantitative importance of the Lucas (1976) critique on the results, see Rudebusch (2005) and Benati (2019). The extent to which this limited deviation is achieved in the present analysis is highlighted in section 5.2.[35]

Given the above, the Fed's objective function is specified as:

$$V_0 = E_0 \left(1-\beta\right) \sum_{t=0}^{\infty} \beta^t \left[\begin{array}{c} \upsilon_p \left(p_t - \overline{p}\right)^2 + \upsilon_{ug} \left(ug_t - \overline{ug}\right)^2 + \\ \upsilon_{\Delta TS} \Delta TS_t^2 + \upsilon_{\Delta PS} \Delta PS_t^2 + \upsilon_{\Delta R} \Delta R_t^2 \end{array} \right], \qquad (22)$$

where $\overline{p}$ and $\overline{ug}$ denote inflation and unemployment gap targets, respectively; υ_z, $z = p$, ug, ΔTS, ΔPS and ΔR, denote weights attached to the stabilization of inflation, unemployment gap and the three policy instruments; Δ is the lag operator. Thus, the first two terms on the right side of (22) capture the Fed's preferences for inflation and unemployment stabilization, the next three terms control the degree of gradualism in changes of the policy instruments. The objective function (22) can be interpreted as capturing the Fed's desire for macroeconomic (prices and unemployment) stabilization while keeping to a minimum changes of policy instruments relative to previous period.

The results of the optimal policy analysis also depend on the values assigned to the six parameters included in the objective function (22): β, $\overline{p}$, $\overline{ug}$, υ_p, υ_{ug}, $\upsilon_{\Delta TS}$, $\upsilon_{\Delta PS}$, and $\upsilon_{\Delta R}$. The discount factor β is set equal to 0.996, corresponding to an annual rate of interest equal to the sample average for R_t of 4.83 percent. The results are not significantly affected by other reasonable values of the discount factor. The unemployment target is set as $\overline{ug} = 0$, since the unemployment gap is measured as the deviationof the actual rate of unemployment from the natural rate, and the Fed's mandate includes full employment (Federal Reserve Bank of St. Louis, 2015). The specification of the remaining parameters is less clear cut as it is reasonable to expect that the Fed preferences in terms of inflation target and weights on policy instruments may have changed during 1979-2018. For this reason, to establish a baseline specification the inflation target is set equal to the sample average over the simulation horizon and all policy weights are given equal value, υ_p=υ_u=$\upsilon_{\Delta TS}$=$\upsilon_{\Delta PS}$=$\upsilon_{\Delta R} = 1$. When

[35] An alternative approach to limit the quantitative significance of the Lucas critique is that used by Sack (2000). This consists of calibrating the inflation target and the relative weight between inflation and unemployment to minimize the in-sample distance between the actual and the optimal the paths of the policy instruments. This approach however relies on the assumption that the Fed's preferences have not changed between 1979-2018.

necessary, the robustness of the quantitative results that follow is evaluated against other reasonable choices of these parameters.

5.1 Gain from Optimization

What would have been the gain in terms of macroeconomic stabilization if QE policy had been optimally coordinated with interest rate policy? The answer to this question is first quantified by considering macroeconomic dynamics from December 2007, 11 months before the Fed announcement of QE1, to October 2017, the date from which the Fed's balance sheet is allowed to shrink gradually as existing assets mature.[36] The gain from monetary policy optimization is evaluated by comparing two scenarios. The first is referred to as the *actual policy*, since this is based on the observed dynamics of the economy and policy instruments over the considered subperiod. The second scenario is the *optimal policy*. This is constructed assuming that the Fed chooses the optimal combination of conventional and QE policy in each month and this optimal policy is implemented given the true state of the economy in each month, as in Sack (2000).[37] The economy response to the optimal policy in any given month is computed including the actual shocks occurring in that month, using the VSTAR in (19). In addition, the inflation target in the loss function (22) is set as $\overline{p} = 3.38$, corresponding to the average inflation rate from December 2007 onward. As a further reference, a *no-policy* scenario is also simulated.[38] This is based on the implied economy and policy dynamics obtained if no-change in QE had been implemented since December 2007.

Following from Dennis and Söderström (2006), two measures of the optimization gain are computed. The first is the (percentage) change in the loss due to implementation of a new policy, V', relative to the loss measured from another policy, V, i.e. $G = 100 \times [1 - V'/V]$. The second is the unemployment-equivalent compensation. To clarify this, consider the unemployment gap term in (22). This can be written as $\upsilon_{ug} ug_t^2 = \upsilon_{ug} (u_t - \overline{u_t})^2$, where u_t and $\overline{u}_t$ denote the actual and the natural rate of unemployment, respectively. The unemployment-equivalent compensation is defined as the permanent deviation $\widehat{u} \neq 0$ of the actual unemployment rate from the natural rate that results in a change of the loss by $(1-\beta) \sum_{t=0}^{\infty} \beta^t \upsilon_{ug} \widehat{u}^2$ such that V_0 equals V'. it follows that $\widehat{u} = \sqrt{(V - V')/\upsilon_{ug}}$. Thus G is a direct measure of macroeconomic stabilization differential between two policies, whereas $\widehat{u}$ quantifies the gain from optimization in terms of implied permanent reduction in the rate of unemployment.

Table 2 presents the results. Columns 2 to 6 report the volatilities of the five variables included in (22) relative to their respective targets; column 7 reports the implied losses; whereas the last two columns report the calculated

[36] The start date is chosen to account for possible aggregate responses in the economy before the announcement of the first QE program, due to the private sector's anticipation of future policy intervention.

[37] There are several possible extensions of the present analysis, including computation of the gains upon the optimization of the federal funds rate alone and/or QE alone.

[38] Counterfactual analyses based on a no-policy scenario are frequent in the recent literature on the macroeconomic effects of QE, see, e.g., Lenza, Pill and Reichlin (2010), Chung et al (2012), Giannone et al (2012), Kapetanios et al (2012), Baumeister and Benati (2013), Dahlhaus, Hess and Reza (2018).

measures of the optimization gain. For robustness, six different specifications of the weights in (22) are considered. The first (I) is the baseline which gives equal weight to each item in (22). The remaining five specifications evaluate the effects of lower weight to the stabilization of either inflation (II), unemployment (III), all policy instruments (IV), the two QE instruments (V), or the federal funds rate (VI). For each set of policy weights the loss is computed as the undiscounted weighted average of the five terms in (22).[39] For each set of policy weights the stabilization gains and unemployment-equivalent compensations are measured first to compare the effects monetary policy optimization relative to the actual policy undertaken by the Fed, and then to compare the actual policy relative to the no-QE scenario.

Two main results emerge from Table 2. The first is that policy optimization would have significantly increased the stabilization gain to the United States economy during 2008-2017. On average across different weight specifications, optimal policy would have reduced macroeconomic volatility by about 28.5 percent relative to the actual policy. In terms of unemployment-equivalent compensation, policy optimization would have resulted in a average reduction of the unemployment rate of about 1.7 percent. The second result emerging from Table 2 is that the gains from switching from the actual to the optimal policy are not as large as those achieved by the actual implementation of QE relative to the no-QE policy scenario. On average across the different weight specifications, macroeconomic volatility and the unemployment rate would have been about 43.6 and 2.8 percent higher, respectively, had QE not been implemented during 2008-2017.

Looking across the different specifications of weights, it can be observed that gains from either the actual or the optimal policy increase when the Fed targets inflation volatility more aggressively than unemployment volatility (compare case II with either I or III). This is because inflation is relative more volatile than unemployment during 2008-2017, in turn implying that assigning a higher weight to this variable brings larger gains. Reducing the weight attached to the QE policy instruments in the objective function produces higher stabilization gains (compare case I with cases IV and V). This is because the further reduction in the volatility of inflation and unemployment more than compensates the increase in the volatility of the instruments. Changing the weight on the federal funds rate makes little difference, since this is the least volatile variable (compare I and VI).

Two further robustness exercises are carried out. The first evaluates the gains from the optimal coordination between interest rate and QE policy over the whole sample period 1979-2018. To this end, the inflation target in (10) is set to be equal to the sample average, i.e. $\overline{p} = 5.84$. The second robustness exercise evaluates how the measured stabilization gains change once these are calculated using the VAR, either post-2008 or over the full sample. Table 3 reports the averages across policy weight specifications of the stabilization gains and unemployment-equivalent compensations obtained from these robustness checks, together with those from the results in Table 2 for comparison.[40]

Two clear results are visible from Table 3. First, the measured gains from

[39] For β close to 1, the discounted quadratic loss is approximatively equal to the weighted average of the deviation from target of the variables included in it, see Bertsekas (2012).

[40] The data calculated from these simulations for each set of policy weights are reported in Appendix E.

optimization reduce once evaluated over the full sample compared to the post-2008 period. This holds whether they are calculated using the VSTAR or the VAR. This suggests that the benefit from the joint optimization of the federal funds rate and the QE instruments are higher during the post Great Recession periods relative to the so-called normal times. The second clear result visible from Table 3 is that optimization gains are systematically overstated when using the VAR. This points towards the importance of accounting for nonlinearity in the quantitative analysis of monetary policy, highlighting the difference it makes using a VAR rather than a non linear model like the VSTAR.

5.2 Impulse Response Functions

How the economy and the policy instruments respond to shocks once QE and the federal funds rate policies are optimally coordinated? To answer this question, the IRFs are calculated from the VSTAR under the optimal policy in (19) assuming that the economy and the policy instruments ($\mathbf{y}_t$) start from two different positions.[41] The first is the ZLB period, in which only QE is the active policy instrument. The second is the post-ZLB period, in which both the federal funds rate and QE are available.[42] In both cases, the preferences of the Fed are set as under the baseline calibration. The inflation target is however recalculated as the corresponding sub-sample average, being 3.05 percent during the ZLB period and 4.30 for the post-ZLB period.[43]

Figures 4 and 5 plots the IRFs from the VSTAR under the optimal policy rules (blue) to a demand and a supply shock, respectively, during the ZLB period. For comparison, in both figures the corresponding responses from the estimated VSTAR, which reflect the dynamics under the actual Fed monetary policy are also reported (red).[44] The IRFs in Figure 4 show that when the federal funds rate is constrained at the ZLB, the Fed responds under the optimal policy to a demand shock by increasing the size of the balance sheet more sharply than under the actual policy, at least in the short run (first 4 to 6 months). At the same time there is significant portfolio rebalancing towards Treasury securities, since TS holdings increase while PS holdings decrease over the 24 months horizon. These changes in size and composition of the Fed portfolio have the effects of dampening the impact of the demand shock on inflation and unemployment. All dynamics under the optimal policy rules are highly nonlinear, in particular those for industrial production, which appears to decrease in the first 5 months following the shock despite the expansion in the size of the balance sheet, before increasing sharply over the subsequent 2 months and falling afterwards. The main distinguishing feature of the IRFs to a supply shock in Figure 5 is that,

[41] As descrided in Appendix D, the IRFs under the optimal policy are computed using the same algortihm for the IRFs in Section 4.4 except that simulation is now based on the VSTAR in (19).

[42] For the ZLB period, the starting position is the average of $\mathbf{y}_t$ from 2009:1-2015:12. For the post ZLB period, is the average of the 2016:1-2018:8 period. See Table 6 in Appendix D.

[43] The IRFs to shocks during normal times are not reported here for reason of space since these are known. The federal funds rate responds to shocks more aggressively under the optimal monetary policy than under the actual response. See for examples Sack (2000), Polito and Wickens (2012), Polito and Spencer (2015).

[44] The responses for the policy instruments under the actual policy rules are the same as those in Figures 3.

	Volatilities					Loss	Stabilization	
	$(p_t - \overline{p})^2$	ug_t^2	ΔTS_t^2	ΔPS_t^2	ΔR_t^2	V	G	$\widehat{u}$
Policy	**Baseline**: $\upsilon_p = \upsilon_{ug} = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	3.63	5.55	0.02	0.05	0.38	9.64		
Optimal	1.42	5.04	0.04	0.08	0.44	7.02	27.14*	1.62*
(No-QE)	11.46	5.68	0.02	0.00	0.00	17.17	43.87**	2.74**
	Weights II: $\upsilon_p = 0.5$, $\upsilon_{ug} = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	3.63	5.55	0.02	0.05	0.38	7.82		
Optimal	1.62	5.05	0.03	0.08	0.42	6.39	18.31*	1.20*
(No-QE)	11.46	5.68	0.02	0.00	0.00	11.43	31.60**	1.90**
	Weights III: $\upsilon_{ug} = 0.5$, $\upsilon_p = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	3.63	5.55	0.02	0.05	0.38	6.86		
Optimal	1.41	5.23	0.05	0.07	0.43	4.57	33.35*	2.14*
(No-QE)	11.46	5.68	0.02	0.00	0.00	11.43	52.11**	3.86**
	Weights IV: $\upsilon_{ug} = \upsilon_p = 1$, $\upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 0.5$							
Actual	3.63	5.55	0.02	0.05	0.38	9.41		
Optimal	1.25	4.76	0.06	0.12	0.47	6.35	32.53*	1.75**
(No-QE)	11.46	5.68	0.02	0.00	0.00	17.15	45.16**	2.78*
	Weights V: $\upsilon_{ug} = \upsilon_p = \upsilon_{\Delta R} = 1$, $\upsilon_{\Delta TS} = \upsilon_{\Delta PS} = 0.5$							
Actual	3.63	5.55	0.02	0.05	0.38	9.42		
Optimal	1.34	4.76	0.05	0.12	0.42	6.42	31.85*	1.73*
(No-QE)	11.46	5.68	0.02	0.00	0.00	17.17	45.13**	2.78**
	Weights VI: $\upsilon_{ug} = \upsilon_p = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = 1$, $\upsilon_{\Delta R} = 0.5$							
Actual	3.63	5.55	0.02	0.05	0.38	9.62		
Optimal	1.33	5.05	0.05	0.08	0.47	6.95	27.74*	1.63*
(No-QE)	11.46	5.68	0.02	0.00	0.00	17.15	43.90**	2.74**
	Average (across weights)							
Actual	3.63	5.55	0.02	0.05	0.38	8.79		
Optimal	1.40	4.98	0.05	0.09	0.44	6.28	28.49*	1.68*
(No-QE)	11.46	5.68	0.02	0.00	0.00	15.73	43.63**	2.80**

Notes: Loss is weighted sum of volatilities; $G = 100 \times [1 - V'/V]$; $\widehat{u} = \sqrt{(V - V')/\upsilon_{ug}}$; '*' indicates that G and $\widehat{u}$ are comparing actual and optimal policy; '**' indicates that G and $\widehat{u}$ are comparing actual and no-QE policy.

Table 2: Gains from monetary policy optimization computed from the VSTAR during 2008-2017.

	Post-2008 (2017:12-2016:10)		Full sample (1979:8-2018:10)	
	VSTAR	VAR	VSTAR	VAR
G	28.49	44.10	19.59	31.34
$\widehat{u}$	1.68	2.05	1.39	1.76

Table 3: Average gains from monetary policy optimization computed from the VSTAR and the VAR for either the post-2008 or the 1979-2018 period.

while the Fed's balance sheet still increases more than under the actual policy, both TS and PS increase over the 24 moths horizon.

Figures 6 and 7 show the IRFs under the actual (red) and optimal (blue) policy to demand and supply shocks, respectively, during the post-ZLB period. Figure 6 shows that the federal funds rate falls and the Fed's balance sheet expands in response to a demand shock under the actual policy. Remarkably, the optimal policy response comes mainly in the form of a sharp rebalancing of the Fed portfolio towards TS, with little action displayed by the the federal fund rate and the size of the Fed portfolio. The IRFs of the policy instruments under the optimal policy show less variability than under the actual policy, since both inflation and unemployment return to targets way more quickly than under the actual policy. The responses to a supply shock under the actual policy in Figure 7 show a reduction in the federal funds rate and a slight increase in the Fed's balance sheet. In contrast, under the optimal policy, there is a large increase in the Fed's balance sheet due to the simultaneous rise in both TS and PS holding.

In summary, QE under the optimal policy results in increase of the size of the balance sheet and significant portfolio rebalancing toward TS in response to a demand shock. QE also results in increase of the size of the balance sheet in response to a supply shock, but no portfolio rebalancing. These patterns hold regardless of whether the federal funds rate is at the ZLB or (just) above.

There is another feature of the results in Figures 4 - 7 that is worth pointing out. Above it was mentioned that the latest consensus to guard against the quantitative relevance of the Lucas critique is that of considering only policy changes that result in modest deviation of macroeconomic dynamics from the observed ones. Rudebusch (2005) interprets 'modest' as not being statistically detectable. Benati (2019) considers deviations smaller than 1 percent five years after the beginning of the policy intervention. The baseline calibration used here, which penalizes changes in the policy instruments equally, achieves an outcome in the spirit of the modest changes proposed in this literature. This is because most of the IRFs calculated under the optimal policy lie within the confidence intervals of the actual IRFs by the end of the two-years horizon.

5.3 Counterfactual Monetary Policy

What would have been the dynamics of the economy and the policy instruments had the federal funds rate and QE been coordinated optimally since 2008? To answer this question, the VSTAR under the optimal policy is used for a counterfactual experiment. This consists of calculating the dynamic evolution of the economy and policy instruments from 2008 onward under the assumption that the federal funds rate and QE had been set optimally over that period of time. The resulting dynamics are used as reference against which to evaluate the observed dynamics for the economy and policy instruments, which reflect the actual monetary policy conducted over the same period of time. The simulation is carried out taking as initial position the economy and the policy vector in December 2007, and using the baseline specification of the Fed preferences.

Given the uncertainty surrounding the effective inflation rate targeted by the Fed over the past ten years, the counterfactual simulation is carried out for three alternative values of the inflation target: the sample average, $\overline{\pi} = 5.8$, $\overline{\pi} = 4$ and $\overline{\pi} = 2$. These do not reflect any specific view on the actual preferences of

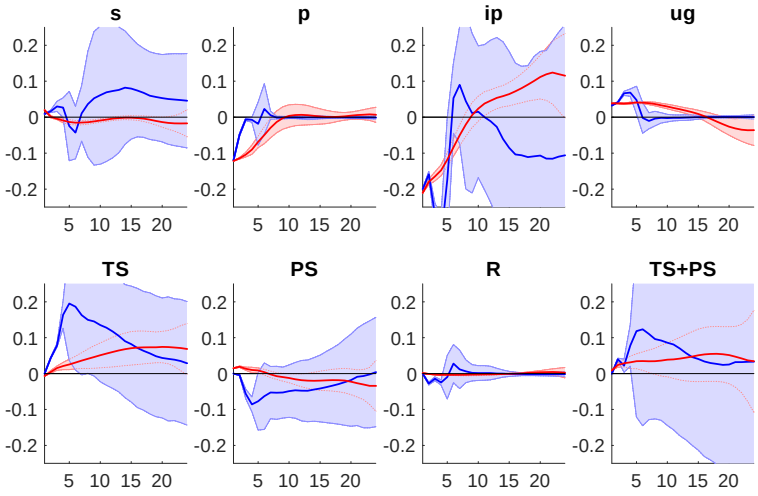

Figure 4: Unconditional IRFs to a demand shock during the ZLB period under the actual (red) and the optimal (blue) policy rules. Shaded areas are two standard deviations confidence bands.

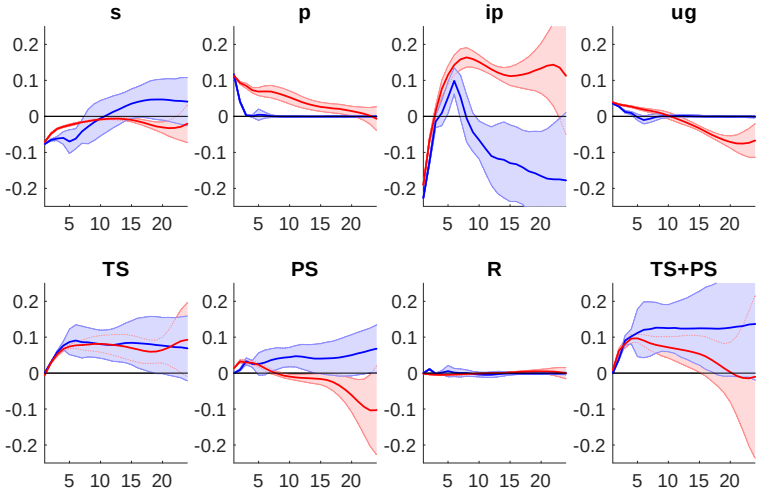

Figure 5: Unconditional IRFs to a supply shock during the ZLB period under the actual (red) and the optimal (blue) policy rules. Shaded areas are two standard deviations confidence bands.

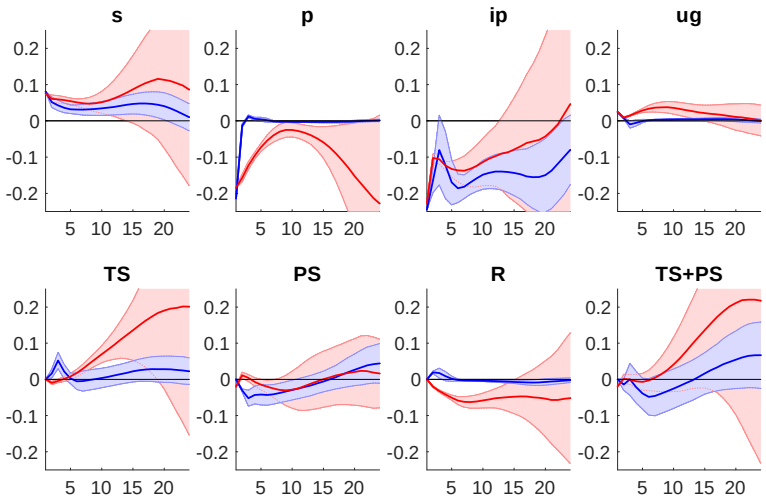

Figure 6: Unconditional IRFs to a demand shock during the post ZLB period under the actual (red) and the optimal (blue) policy rules. Shaded areas are two standard deviations confidence bands.

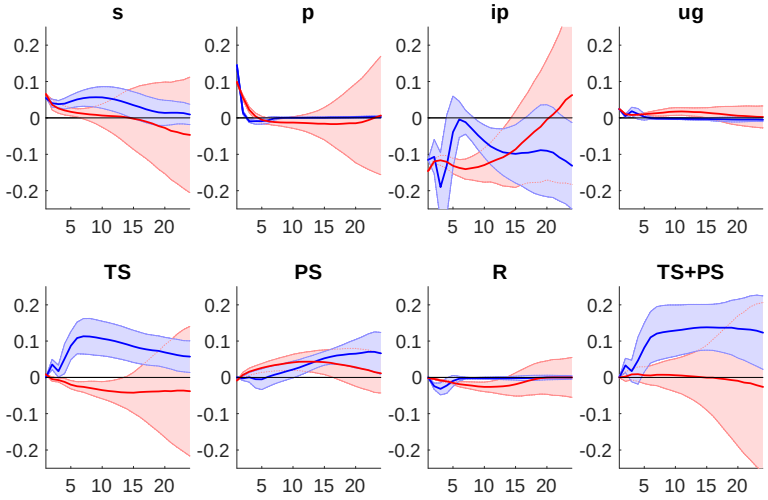

Figure 7: Unconditional IRFs to a supply shock during the post ZLB period under the actual (red) and the optimal (blue) policy rules. Shaded areas are two standard deviations confidence bands.

the Fed, but are selected for gauging indication on the sensitivity of the results to the choice of the inflation target.[45] It is important to clarify the role of the inflation target in these simulations. All other things equal, the optimal policy would require expansionary (contractionary) QE whenever actual inflation is below (above) target. As the unemployment gap is positive during most of the post-2008 period, the lower is actual inflation relative to the target, the stronger is the response of monetary policy to unemployment stabilization over the counterfactual simulation.

Figure 8 plots the dynamics of the economy and the policy instruments from these counterfactual simulations. In each subpanel, black lines denote the actual dynamics observed from the data; whereas blue, red and green lines denote the dynamics simulated from the optimal VSTAR under the three alternative inflation targets of 5.8, 4 and 2 percent, respectively. The simulation is carried out assuming that in each period the Fed implements the optimal policy given the observed state of the economy.

Three main results can be observed from Figure 8. First, the dynamics of the economy are fairly similar across the three counterfactual simulations. Compared to the observed dynamics the main differences are for nominal and financial variables, as credit risk would have been lower during 2008-2013 and there would have not been deflation during 2009 under the optimal policy regardless of the choice of the inflation target. The dynamics of industrial production and the unemployment gap are remarkably similar to those observed from the data, under any of the three counterfactual simulations.

The second result observable from Figure 8 concerns the path of the federal funds rate, which appears close to that observed from the data under any of the counterfactual simulations. To shed further light on this, Table 4 gives a breakdown of the duration of the ZLB period under the actual and optimal policies implied by the counterfactual.[46] In the data, the ZLB period lasts 84 months, from December 2008 to September 2015, with the liftoff starting in October 2015. The duration of the ZLB period from the counterfactual simulations depends on the calibration of the inflation target. With the highest inflation target of 5.8 percent, the ZLB period would have been slightly shorter, 80 months November 2008 to March 2015, and the liftoff would have started in April 2015. With the 2 percent inflation target the ZLB period would have lasted longer, 101 months from November 2008 to April 2017, and the liftoff would have only started in May 2017. The 4 percent inflation target replicates the actual duration of the ZLB period more closely. To put these results into context, it is worth highlighting that the durations of the ZLB period according to this optimal policy analysis are considerably longer than those predicted by a standard Taylor rule (without any nonnegative boundary on the policy rate). According to the calculations made in Federal Reserve Bank of St. Louis (2015), the ZLB should have terminated much earlier (around the beginning of 2011) had monetary policy during the Great Recession being conducted as predicted

[45] Ball (2014) argues in favour of a 4 percent inflation target. The 2 percent inflation target was officially announced by the Fed as a long-run objective in January 2012. Mumtaz and Theodoridis (2019) find that the implicit inflation target of the Fed has been on average close to 2 percent between 2008 and 2016, picking above 5 percent around 2010. The average inflation rate from 2008 onward sits between the 2 and 4 percent targets, being 3.36 percent.

[46] For this purpose, the federal funds rate is deemed to be at the ZLB in any period in which is equal to or below 0.25 percent.

by a standard Taylor rule.

The third result observable from Figure 8 concerns the dynamics of the Fed's balance sheet. These appear quite close the observed ones under any of the three counterfactual scenarios. The most notable exception is the period 2009-2013 period, when the size of the simulated Fed's balance sheet is larger than the actual one. This is due to the more rapid increase in TS purchases under any of the counterfactual relative to the observed purchases. Further the normalization phase, i.e. the gradual reduction in the Fed assets holding (see Federal Reserve Bank of St. Louis, 2015), would have started around December 2015, regardless of the choice of the inflation target.

The counterfactual simulations for the QE instruments presented in Figure 8 can be used to provide direct evaluations of either specific segments of the QE phase or specific asset purchase programs undertaken by the Fed since November 2008. In particular, the Fed undertook three large-scale asset purchases, commonly known as QE1, QE2, and QE3; and the Maturity Extension Program (MEP), also known as the 'Operation Twist'.[47] The upper part of Table 5 reports how TS holdings, PS holdings and the total size of the Fed's balance sheet changed during each of these programs. The remaining part of the Table reports the corresponding changes under each of the three counterfactual simulation. Under the actual QE policy, the total increase in the balance sheet was about 17.96 percent of GDP. The size of the change of the Fed's balance sheet as a percentage of GDP prescribed under the optimal policy is close to the actual one, ranging between 17 and 18.25 percent depending on the inflation target. Comparing across programs, it can be noted that according to this analysis, QE1 should have been almost twice larger, regardless of the inflation target. Consequently, this would have resulted in QE2 and QE3 programs of smaller size under the optimal policies compared to the actual. It would have also resulted in a reduction of the size of the balance sheet (relative to GDP) during the MEP under the optimal policy about four times larger compared to the actual reduction observed in the data.

Overall, according to this analysis the scale of total QE intervention observed in the data is close to that prescribed by the optimal QE policy under conventional calibration of the Fed preferences. Rather than the overall size, the main differences between actual and optimal QE are in terms of their timing and composition, since the optimal QE would have entailed earlier increases in the Fed's holding of TS compared to the actual policy. According to this analysis, the main outcomes of these changes would have been a lower spread and higher inflation in the five years after the Great Recession. None of the counterfactual simulation gives evidence in favour of an earlier liftoff.

6 Conclusion

Macroeconomic data display many nonlinear features (asymmetries, thresholds and large swings) that often make linear models inadequate for reliable quantitative analysis. VSTAR is a popular tool employed in econometrics and applied

[47] The MEP had the the objective to put downward pressure on longer-term interest rates. It consisted of purchases of long-term Treasuries, accompanied by the sale of the same quantity of short term securities, leaving the overall size of the balance sheet unchanged.

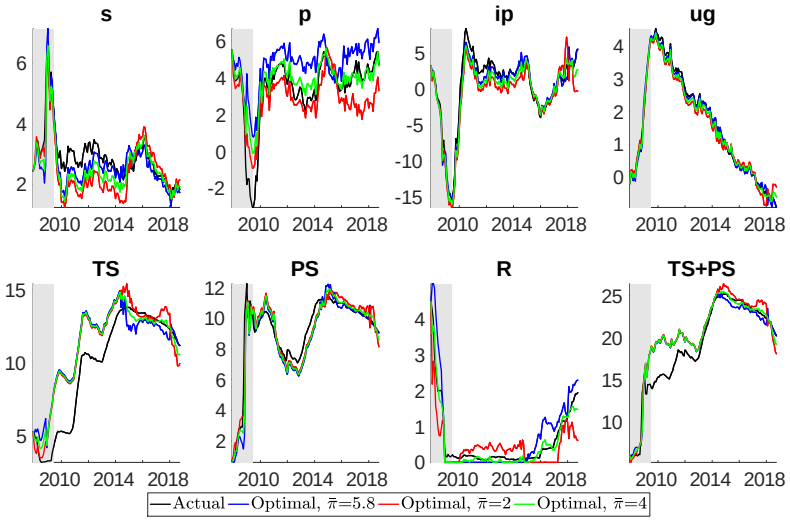

Figure 8: Optimal conventional and QE monetary policy in the United States from 2007:12 to 2018:10.

	Start	End	Duration (months)
	Actual Policy		
	Dec 08	Dec 15	85
$\bar{\pi}$(%)	Optimal Policy		
5.8	Nov 08	Jun 15	80
4	Nov 08	Dec 15	86
2	Nov 08	Mar 17	101

Table 4: Duration of the ZLB period under the actual and optimal policy.

	QE1	QE2	MEP	QE3	Total
			Actual Policy		
TS	2.02	5.10	-0.48	3.69	
PS	5.35	-1.75	-0.08	4.12	
Total	7.37	3.35	-0.56	7.81	17.96
			Optimal Policy		
			$\overline{\pi} = 5.8$		
TS	3.34	3.89	-1.38	-0.01	
PS	9.11	-2.92	-0.57	5.54	
Total	12.45	0.97	-1.95	5.53	17.00
			$\overline{\pi} = 4$		
TS	4.08	3.87	-1.36	1.45	
PS	8.20	-2.93	-0.60	4.89	
Total	12.28	0.94	-1.96	6.34	17.60
			$\overline{\pi} = 2$		
TS	4.87	3.85	-1.34	3.03	
PS	7.21	-2.94	-0.63	4.19	
Total	12.09	0.91	-1.97	7.22	18.25

Table 5: Change in the Fed's asset portfolio in percentage of GDP under actual and optimal policy over different QE phases.

macroeconomics for inference, structural analysis and forecasting when data are nonlinear. This paper uses VSTAR for a different task, the analysis of optimal policy.

It is shown that at least two possible nonlinear structural representations can be identified from a reduced-form VSTAR for calculating an optimal policy rule. One of these makes the analysis particularly tractable, since it preserves certainty equivalence and ensures low dimensionality of the optimization problem.

This structural representation from the VSTAR can then be used as the constraint faced by a nonlinear regulator in charge of setting the path of policy instruments that minimizes a given quadratic objective. This NLQR problem is solved by adapting from the engineering theory the SDRE method. This consists of employing SDC factorization to transform the nonlinear structural representation from the VSTAR into a VAR with SDC matrices, which makes the NLQR problem isomorphic the LQR. Consequently, a nonlinear optimization problem can be solved with standard dynamic programming techniques.

The advantage of the SDRE method relative to linearization is that it does not neglect the potential effects of nonlinearity on optimal decisions. At the same time the SDRE method has the advantage compared to numerical methods of being simple and computationally tractable, since it is still based on the iteration upon convergence of the well-understood Riccati Equation.

The methodology is illustrated to study the optimal coordination between conventional interest rate monetary policy and QE in the United States and evaluate its effects on aggregate quantities and prices along three important dimensions: the gains from optimization; the responses to shocks; and a post-2008 counterfactual analysis of the dynamic of the United States economy when conventional monetary policy and QE are set optimally.

The empirical results show that the gains from the joint optimization of (conventional and QE) monetary policy instruments can be large though not has those achieved by the actual QE policy relative to a scenario of no-QE. The optimization gain is found to be larger when measured over the Great Recession and its aftermath, suggesting that the benefits from monetary policy coordination are more substantial during periods when nonlinearity is more significant. The optimization gain is also found to be larger when the analysis is repeated using a VAR, which suggests that linear models can potentially overstate the benefits from monetary policy optimization.

With regard to their responses to shocks, both actual and optimal QE policy display a significant degree of asymmetry and history dependence. However, differences between the optimal and the actual policy are significant only in the very short-run horizon (four to six months). The response of QE under the optimal policy shows two clear patterns. After a demand shock, the size of the balance sheet increases and the portfolio mix of assets shifts towards Treasury securities. After a supply shock, the size of the balance sheet increases but there is not shift in the portfolio mix.

The counterfactual simulation shows that the observed overall increase in the Fed's balance sheet since 2008 was not far from that prescribed by the optimal policy. Differences between the actual and the optimal policy are larger during the first phase of QE, as the latter would have prescribed larger purchases of Treasuries. The main macroeconomic outcome of this earlier expansion of the balance sheet would have been a lower credit spread and higher rate of inflation in during 2008-2013. Finally, in contrast to the predictions of a Taylor-type monetary policy rule, the observed duration of the ZLB period is found to be in line with that otherwise prescribed by the optimal policy.

Overall, these results highlights not only the feasibility of optimal policy analysis with VSTAR but also the many new interesting dimensions of monetary policy that can be explored.

7 Acknowledgment

I am grateful for helpful discussions, suggestions and comments on early drafts of this paper to Juan Paez-Farrell, Paulo Monteiro, Michele Piffer, Peter Spencer, Timo Teräsvirta, Paul Trodden, Mike Wickens. I am grateful for comments received by seminar and conference participants at the 4th International Workshop on Financial Markets and Nonlinear Dynamics, University of Brunnel, University of Sheffield.

8 References

Ang A, Bekaert G. 2002. Regime Switches in Interest Rates. *Journal of Business and Economic Statistics*, 20, 2, 163-182.

Auerbach AJ, Gorodnichenko Y. 2012. Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, 4, 2, 1-27.

Balke SN. 2000. Credit and Economic Activity: Credit Regimes and Nonlinear Propagation of Shocks. *The Review of Economics and Statistics*, 82, 344-349.

Ball LM. 2014. The Case for a Long-Run Inflation Target of Four Percent. *IMF Working Papers 14/92*, International Monetary Fund.

Baumeister C, Liu P, Mumtaz H. 2013. Changes in the Effects of Monetary Policy on Disaggregate Price Dynamics. *Journal of Economic Dynamics and Control,* 37, 3, 543-560.

Beeler SC. 2004. State-Dependent Riccati Equation Regulation of Systems with State and Control Nonlinearities. Tech. Rep. NASA/CR-2004-213245, NIA 2004-08, NASA Langley Research Center, National Institute of Aerospace, Hampton, VA 23681.

Beeler SC, Tran HT, Banks HT. 2000a. Feedback Control Methodologies for Nonlinear Systems. *Journal of Optimization Theory and Applications*, 107, 1-33.

Beeler SC, Tran HT, Banks HT. 2000b. State Estimation and Tracking Control of Nonlinear Dynamical Systems. CRSC Technical Report CSRC-TR00-19, N.C. State University.

Benati L. 2019. Would a Modest Policy intervention Have Prevented the U.S. Housing Bubble? Mimeo.

Benati L. 2015. The Long-Run Phillips Curve: A Structural VAR Investigation. *Journal of Monetary Economics*, 76, 15-28.

Benati L, Surico P. 2009. VAR Analysis and the Great moderation. *American Economic Review*, 99, 4, 1636-52.

Bernanke B, Mihov I. 1998. Measuring monetary policy. *Quarterly Journal of Economics,* 113, 869-902.

Bertsekas DP. 2012. *Dynamic Programming and Optimal Control*, Vol. 2. Nashua, NH: Athena Sci. 4th eds.

Burghart JA. 1969. A Technique for Suboptimal Control of Nonlinear Systems. *IEEE Transactions on Automatic Control*, 14, 530-533.

Canova F, Gambetti L. 2009. Structural Changes in the US Economy: Is There a Role for Monetary Policy? *Journal of Economic Dynamics and Control* 33, 2, 477-490.

Caggiano G, Castelnuovo E, Colombo V, Nodari G. 2015. Estimating Fiscal Multipliers: News from a Non-Linear World. *The Economic Journal*,125, 746–776.

Caggiano G, Castelnuovo E, Groshenny N. 2014. Uncertainty Shocks and Unemployment Dynamics in U.S. Recessions. *Journal of Monetary Economics*, 67, 78-92.

Chung H, Laforte J-P, Reifschneider D, Williams JC. 2012. Have We Underestimated the Likelihood and Severity of Zero Lower Bound Events? *Journal of Money, Credit and Banking*, 44, 47-82.

Cloutier JR, D'Souza CN, Mracek CP. 1996a. Nonlinear Regulation and Nonlinear H_∞ Control Via the State-Dependent Riccati Equation Technique: Part1. Theory. Proceedings of the First International Conference on Nonlinear Problems in Aviation and Aerospace, Daytona Beach, FL.

Cloutier JR, D'Souza CN, Mracek CP. 1996a. Nonlinear Regulation and Nonlinear H_∞ Control Via the State-Dependent Riccati Equation Technique: Part2. Examples. Proceedings of the First International Conference on Nonlinear Problems in Aviation and Aerospace, Daytona Beach, FL.

Curdia V, Woodford M. 2011. The central-bank balance sheet as an instrument of policy. *Journal of Monetary Economics*, 58, 54-79.

Dahlhaus T, Hess K, Reza A. 2018. International Transmission Channels of U.S. Quantitative Easing: Evidence from Canada. *Journal of Money, Credit and Banking*, 50(2-3), 545-563.

Dennis R, Söderström ULF. 2006. How Important Is Precommitment for Monetary Policy? *Journal of Money, Credit and Banking*, 38, 4, 847-872.

Engle RF, Kroner KF. 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11, 122-150.

Federal Reserve Bank of St. Louis. 2015. The Road to Normal: New Directions in Monetary Policy. Annual Report.

Gagnon J, Raskin M, Remache J, Sack B. 2011. The financial market effects of the Federal Reserve's large-scale asset purchases. *International Journal of Central Banking*, 7, 10, 3-43.

Galvão AB. 2006. Structural Break Threshold VARs for Predicting US Recessions Using the Spread. *Journal of Applied Econometrics*, 21, 463-487.

Galvão AB, Owyang MT. 2018. Financial Stress Regimes and the Macroeconomy. *Journal of Money, Credit and Banking*, 50, 7, 1479-1505.

Garrard WL, McClamroch NH, Clark LG. 1967. An Approach to Sub-Optimal Feedback Control of Nonlinear System. *International Journal of Control*, 5, 425-435.

Giannone D, Lenza M, Pill H, Reichlin L. 2012. The ECB and The Interbank Market. *The Economic Journal*, 122, F467-F486.

Granger C, Teräsvirta T. 1993. *Modelling Non-Linear Economic Relationships*. Oxford University Press.

Guerrieri L, Iacoviello M. 2017. Collateral Constraints and Macroeconomic Asymmetries. *Journal of Monetary Economics*, 90, 28-49.

Harrison R. 2017. Optimal quantitative easing. *Bank of England working papers 678*, Bank of England.

Hubrich K, Teräsvirta T. 2013. Thresholds and Smooth Transitions in Vector Autoregressive Models. in *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims*, Fomby TB, Kilian L, Murphy A (ed.), 273-326.

Hurn S, Johnson N, Silvennoinen A, Teräsvirta T. 2018. Transition from the Taylor rule to the zero lower bound. Mimeo.

Judd KL. 1988. *Numerical Methods in Economics*. Cambridge, MA: The MIT Press.

Joyce M, Miles D, Scott A, Vayanos D. 2012. Quantitative Easing and Unconventional Monetary Policy - An Introduction. The Economic Journal, 122, 271-288.

Kapetanios G, Mumtaz H, Stevens I, Theodoridis K. 2012. Assessing the Economy-Wide Effects of Quantitative Easing. *Economic Journal*, 122, 316-47.

Kilian L, Lütkepohl H. 2017. *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Koop G, Pesaran MH, Potter SM. 1996. Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics*, 74, 1, 119-147.

Kuttner KN. 2018. Outside the Box: Unconventional Monetary Policy in the Great Recession and Beyond, *Journal of Economic Perspectives*, 32, 4, 121-146.

Lanne M, Saikkonen P. 2005. Nonlinear GARCH models for highly persistent volatility. *Econometrics Journal,* 8, 251-276.

Lenza M, Pill H, Reichlin L. 2010. Monetary Policy in Exceptional Times. *Economic Policy*, 25, 295-339.

Lewis FL, Vrabie D, Syrmos VL. 2012. *Optimal Control*, Third Ed., John Wiley & Sons, Inc.

Ljungqvist L, Sargent TJ. 2018. *Recursive Macroeconomic Theory*, fourth edition. The MIT Press, Cambridge, Massachussets.

Lucas R. 1976. Econometric policy evaluation: a critique. In *The Phillips Curve and Labor Markets*, Brunner K, Meltzer A (eds), Carnegie–Rochester Conference Series on Public Policy 1. American Elsevier, New York, 19-46.

Luukkonen R, Saikkonen P, Teräsvirta T. 1988. Testing Linearity Against Smooth Transition Autoregressive Models. *Biometrika*, 75, 3, 491-499.

Martin V, Hurn S, Harris D. 2013. Econometric Modelling with Time Series. Cambridge University Press.

Miranda MJ, Fackler PL. 2002. *Applied Computational Economics and Finance*. The MIT Press, Cambridge, Massachussets.

Mracek CP, Cloutier JR. 1988. Control Design for the Nonlinear Benchmark Problem Via the State-Dependent Riccati Equation Method. *International Journal of Robust and Nonlinear Control*, 8, 401-433.

Mumtaz H, Theodoridis K. 2019. The Federal Reserve Inflation Target and Macroeconomic Dynamics. A SVAR Analysis. Mimeo.

Pearson JD. 1962. Approximation Methods in Optimal Control. *Journal of Electronics and Control*, 13, 453-465.

Polito V, Spencer P. 2015. Optimal Control of Heteroscedastic Macroeconomic Models. *Journal of Applied Econometrics*, 31, 1430-1444.

Polito V, Wickens MR. 2012. Optimal Monetary Policy using an Unrestricted VAR. *Journal of Applied Econometrics*, 27, 4, 525-553.

Primiceri GE. 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies,* 72, 3, 821-852.

Reis R. 2017. QE in the future: the central bank's balance sheet in a fiscal crisis. *IMF Economic Review,* 65, 1, 71-112.

Ramey VA, Zubairy S. 2018. Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data. *Journal of Political Economy*, 126, 2, 850-901.

Rudebusch GD. 2005. Assessing the Lucas Critique in Monetary Policy Models. *Journal of Money, Credit, and Banking*, 37, 2, 245-272.

Sack B. 2000. Does the Fed Act Gradually? A VAR Analysis. *Journal of Monetary Economics*, 46, 1, 229-256.

Sims C, Zha T. 2006a. Does monetary policy generate recessions? *Macroeconomic Dynamics,* 10, 231-272.

Sims C, Zha T. 2006b. Were There Regime Switches in U.S. Monetary Policy? *American Economic Review*, 96, 1, 54-81.

Sims E, Wu JC. 2019a. Evaluating Central Banks' Tool Kit: Past, Present, and Future. *NBER Working Paper No. 26040*. NBER.

Sims E, Wu JC. 2019b. The Four Equations New Keynesian Model. *NBER Working Paper No. 26067*. NBER.

Stock JH, Watson MW. 2001. Vector Autoregression. *Journal of Economic Perspectives*, 15, 4, 101-115.

Stock JH, Watson MW. 1996. Evidence of Structural Instability in Macroeconomic Time Series Relations. *Journal of Business and Economic Statistics,* 14, 1, 11-30.

Teräsvirta T, Tjøstheim D. Granger CWJ. 2010. Modelling Nonlinear Economic Timeseries. Oxford University Press Inc, New York.

Weise CL. 1999. The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression. *Journal of Money, Credit and Banking*, 31, 1, 85-108.

Wernli A, Cook G. 1975. Suboptimal Control for the Nonlinear Quadratic Regulator Problem. *Automatica,* 11, 75-84.

Woodford M. 2016. Quantitative Easing and Financial Stability. *NBER Working Paper No. 22285,* NBER.

A Structural Representations

Consider the reduced-form residuals in (4). These can can be written as

$$\left[\begin{array}{c}\mathbf{v}_{xt}\\ \mathbf{v}_{ut}\end{array}\right]=\left[\begin{array}{cc}\mathbf{I} & \mathbf{G}_{12t}\\ \mathbf{G}_{21t} & \mathbf{I}\end{array}\right]\left[\begin{array}{c}\boldsymbol{\epsilon}_{xt}\\ \boldsymbol{\epsilon}_{ut}\end{array}\right]$$

where $\left[\begin{array}{cc}\boldsymbol{\epsilon}_{xt}' & \boldsymbol{\epsilon}_{ut}'\end{array}\right]'$ is a vector of uncorrelated error terms, with $\boldsymbol{\epsilon}_{xt}\sim i.i.d.\left(\mathbf{0},\boldsymbol{\sigma}_x^2\right)$ and $\boldsymbol{\epsilon}_{ut}\sim i.i.d.\left(\mathbf{0},\boldsymbol{\sigma}_u^2\right)$. Thus $\mathbf{v}_{xt}=\boldsymbol{\epsilon}_{xt}+\mathbf{G}_{12t}\boldsymbol{\epsilon}_{ut}$ and $\mathbf{v}_{ut}=\mathbf{G}_{21t}\boldsymbol{\epsilon}_{xt}+\boldsymbol{\epsilon}_{ut}$.

A.1 Assumption A1

Set $\mathbf{G}_{12t}=\mathbf{0}$, so that $\mathbf{v}_{xt}=\boldsymbol{\epsilon}_{xt}$. Under this restriction (4) becomes

$$\begin{aligned}\boldsymbol{\Omega}_t &= \left[\begin{array}{cc}\boldsymbol{\Omega}_{xxt} & \mathbf{0}\\ \boldsymbol{\Omega}_{uxt} & \boldsymbol{\Omega}_{uut}\end{array}\right]\\ &= \left[1-g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\begin{array}{c}\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx1} & \mathbf{0}\\ \boldsymbol{\Omega}_{ux1} & \boldsymbol{\Omega}_{uu1}\end{array}\right]\\ +l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx2} & \mathbf{0}\\ \boldsymbol{\Omega}_{ux2} & \boldsymbol{\Omega}_{uu2}\end{array}\right]\end{array}\right\}+\\ &\quad g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\begin{array}{c}\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx3} & \mathbf{0}\\ \boldsymbol{\Omega}_{ux3} & \boldsymbol{\Omega}_{uu3}\end{array}\right]\\ +l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx4} & \mathbf{0}\\ \boldsymbol{\Omega}_{ux4} & \boldsymbol{\Omega}_{uu4}\end{array}\right]\end{array}\right\}.\end{aligned}$$

Thus

$$\begin{aligned}\boldsymbol{\Omega}_{xxt} &= \left[1-g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_{xx1}+l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_{xx2}\right\}+\\ &\quad g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_{xx3}+l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_{xx4}\right\}\\ &= \boldsymbol{\epsilon}_{xt}^2,\\ \boldsymbol{\Omega}_{uxt} &= \left[1-g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_{ux1}+l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_{ux2}\right\}+\\ &\quad g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_{ux3}+l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_{ux4}\right\}\\ &= \mathbf{G}_{21t}\boldsymbol{\epsilon}_{xt}^2,\end{aligned}$$

which implies

$$\begin{aligned}\boldsymbol{\Omega}_{uxt} &= \mathbf{G}_{21t}\boldsymbol{\Omega}_{xxt}\\ \mathbf{G}_{21t} &= \boldsymbol{\Omega}_{uxt}\boldsymbol{\Omega}_{xxt}^{-1}\\ &= \left\{\begin{array}{c}\left[1-g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_{ux1}+l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_{ux2}\right\}+\\ g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_{ux3}+l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_{ux4}\right\}\end{array}\right\}\times\\ &\quad \left\{\begin{array}{c}\left[1-g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_{xx1}+l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_{xx2}\right\}+\\ g_V\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\left[1-l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\boldsymbol{\Omega}_{xx3}+l_V\left(\mathbf{s}_x'\mathbf{y}_t\right)\boldsymbol{\Omega}_{xx4}\right\}\end{array}\right\}^{-1}.\end{aligned}$$

The matrix

$$\mathbf{H}_t^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix}.$$

can then be used to map the reduced-form model into a structural one. To this end, pre-multiply both sides of (3) by the right side of $\mathbf{H}_t^{-1}$ to obtain:

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{u}_{t+1} \end{bmatrix} =$$
$$[1 - g_M(\mathbf{s}_u'\mathbf{y}_t)] \left\{ \begin{array}{c} [1 - l_M(\mathbf{s}_x'\mathbf{y}_t)] \left\{ \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_{x1} \\ \boldsymbol{\lambda}_{u1} \end{bmatrix} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_{x1}(L) \\ \boldsymbol{\Lambda}_{u1}(L) \end{bmatrix} \mathbf{y}_t \right\} \\ + l_M(\mathbf{s}_x'\mathbf{y}_t) \left\{ \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_{x2} \\ \boldsymbol{\lambda}_{u2} \end{bmatrix} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_{x2}(L) \\ \boldsymbol{\Lambda}_{u2}(L) \end{bmatrix} \mathbf{y}_t \right\} \end{array} \right\} +$$
$$g_M(\mathbf{s}_u'\mathbf{y}_t) \left\{ \begin{array}{c} [1 - l_M(\mathbf{s}_x'\mathbf{y}_t)] \left\{ \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_{x3} \\ \boldsymbol{\lambda}_{u3} \end{bmatrix} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_{x3}(L) \\ \boldsymbol{\Lambda}_{u3}(L) \end{bmatrix} \mathbf{y}_t \right\} \\ + l_M(\mathbf{s}_x'\mathbf{y}_t) \left\{ \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_{x4} \\ \boldsymbol{\lambda}_{u4} \end{bmatrix} + \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_{x4}(L) \\ \boldsymbol{\Lambda}_{u4}(L) \end{bmatrix} \mathbf{y}_t \right\} \end{array} \right\} +$$
$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{G}_{21t} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{xt+1} \\ \mathbf{v}_{ut+1} \end{bmatrix}.$$

This is solved as

$$\begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{u}_{t+1} - \mathbf{G}_{12t}\mathbf{x}_{t+1} \end{bmatrix} =$$
$$[1 - g_M(\mathbf{s}_u'\mathbf{y}_t)] \left\{ \begin{array}{c} [1 - l_M(\mathbf{s}_x'\mathbf{y}_t)] \left\{ \begin{bmatrix} \boldsymbol{\lambda}_{x1} \\ \boldsymbol{\lambda}_{u1} - \mathbf{G}_{12t}\boldsymbol{\lambda}_{x1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}_{x1}(L) \\ \boldsymbol{\Lambda}_{u1}(L) - \mathbf{G}_{12t}\boldsymbol{\Lambda}_{x1}(L) \end{bmatrix} \mathbf{y}_t \right\} + \\ + l_M(\mathbf{s}_x'\mathbf{y}_t) \left\{ \begin{bmatrix} \boldsymbol{\lambda}_{x2} \\ \boldsymbol{\lambda}_{u2} - \mathbf{G}_{12t}\boldsymbol{\lambda}_{x2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}_{x2}(L) \\ \boldsymbol{\Lambda}_{u2}(L) - \mathbf{G}_{12t}\boldsymbol{\Lambda}_{x2}(L) \end{bmatrix} \mathbf{y}_t \right\} \end{array} \right\}$$
$$+ g_M(\mathbf{s}_u'\mathbf{y}_t) \left\{ \begin{array}{c} [1 - l_M(\mathbf{s}_x'\mathbf{y}_t)] \left\{ \begin{bmatrix} \boldsymbol{\lambda}_{x3} \\ \boldsymbol{\lambda}_{u3} - \mathbf{G}_{12t}\boldsymbol{\lambda}_{x3} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}_{x3}(L) \\ \boldsymbol{\Lambda}_{u3}(L) - \mathbf{G}_{12t}\boldsymbol{\Lambda}_{x3}(L) \end{bmatrix} \mathbf{y}_t \right\} + \\ + l_M(\mathbf{s}_x'\mathbf{y}_t) \left\{ \begin{bmatrix} \boldsymbol{\lambda}_{x4} \\ \boldsymbol{\lambda}_{u4} - \mathbf{G}_{12t}\boldsymbol{\lambda}_{x4} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}_{x4}(L) \\ \boldsymbol{\Lambda}_{u4}(L) - \mathbf{G}_{12t}\boldsymbol{\Lambda}_{x4}(L) \end{bmatrix} \mathbf{y}_t \right\} \end{array} \right\} +$$
$$\begin{bmatrix} \mathbf{v}_{xt+1} \\ \mathbf{v}_{ut+1} - \mathbf{G}_{12t}\mathbf{v}_{xt+1} \end{bmatrix}.$$

After moving the term $\mathbf{G}_{12t}\mathbf{x}_{t+1}$ on the right side and multiplying through, the equations for the economy vector can be written in state-space form as

$$\begin{aligned}
&\left[\begin{array}{c}\mathbf{x}_{t+1}\\ \mathbf{u}_{t+1}\end{array}\right] = \\
&\left\{\begin{array}{l}
[1-g_M(\mathbf{s}_u'\mathbf{y}_t)]\left\{\begin{array}{l}[1-l_M(\mathbf{s}_x'\mathbf{y}_t)]\left\{\left[\begin{array}{c}\boldsymbol{\lambda}_{x1}\\ \mathbf{0}\end{array}\right]+\left[\begin{array}{cc}\boldsymbol{\Lambda}_{xx1}(L) & \boldsymbol{\Lambda}_{xu1}(L)-\boldsymbol{\Lambda}_{xu11}\\ \mathbf{0} & \mathbf{0}\end{array}\right]\left[\begin{array}{c}\mathbf{x}_t\\ \mathbf{u}_t\end{array}\right]\right\}+\\ \qquad\qquad l_M(\mathbf{s}_x'\mathbf{y}_t)\left\{\left[\begin{array}{c}\boldsymbol{\lambda}_{x2}\\ \mathbf{0}\end{array}\right]+\left[\begin{array}{cc}\boldsymbol{\Lambda}_{xx2}(L) & \boldsymbol{\Lambda}_{xu2}(L)-\boldsymbol{\Lambda}_{xu21}\\ \mathbf{0} & \mathbf{0}\end{array}\right]\left[\begin{array}{c}\mathbf{x}_t\\ \mathbf{u}_t\end{array}\right]\right\}\end{array}\right\}+\\
+\quad g_M(\mathbf{s}_u'\mathbf{y}_t)\left\{\begin{array}{l}[1-l_M(\mathbf{s}_x'\mathbf{y}_t)]\left\{\left[\begin{array}{c}\boldsymbol{\lambda}_{x3}\\ \mathbf{0}\end{array}\right]+\left[\begin{array}{cc}\boldsymbol{\Lambda}_{xx3}(L) & \boldsymbol{\Lambda}_{xu3}(L)-\boldsymbol{\Lambda}_{xu31}\\ \mathbf{0} & \mathbf{0}\end{array}\right]\left[\begin{array}{c}\mathbf{x}_t\\ \mathbf{u}_t\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{xu31}\\ \mathbf{0}\end{array}\right]\mathbf{u}_t\right\}+\\ \qquad\qquad l_M(\mathbf{s}_x'\mathbf{y}_t)\left\{\left[\begin{array}{c}\boldsymbol{\lambda}_{x4}\\ \mathbf{0}\end{array}\right]+\left[\begin{array}{cc}\boldsymbol{\Lambda}_{xx4}(L) & \boldsymbol{\Lambda}_{xu4}(L)-\boldsymbol{\Lambda}_{xu41}\\ \mathbf{0} & \mathbf{0}\end{array}\right]\left[\begin{array}{c}\mathbf{x}_t\\ \mathbf{u}_t\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{xu41}\\ \mathbf{0}\end{array}\right]\mathbf{u}_t\right\}\end{array}\right\}+\\
\left[\begin{array}{c}\mathbf{v}_{xt+1}\\ \mathbf{0}\end{array}\right].
\end{aligned}$$

In the above, note that $\boldsymbol{\Lambda}_{xuj}(L)=\sum_{k=1}^{q}\boldsymbol{\Lambda}_{xujk}L^k$, $j=1,...,4$. Thus the term $\boldsymbol{\Lambda}_{xuj1}$ is the matrix of coefficients pertinent to $\mathbf{u}_t$ in each state. The state-space representation above for the economy is equivalently written as in (5)-(6).

A.2 Assumption A2

Set $\mathbf{G}_{21t}=\mathbf{0}$, which implies $\mathbf{v}_{ut}=\boldsymbol{\epsilon}_{ut}$. Under this restriction (3) becomes

$$\begin{aligned}
\boldsymbol{\Omega}_t &= \left[\begin{array}{cc}\boldsymbol{\Omega}_{xxt} & \boldsymbol{\Omega}_{xut}\\ \mathbf{0} & \boldsymbol{\Omega}_{uut}\end{array}\right]\\
&= [1-g_V(\mathbf{s}_u'\mathbf{y}_t)]\left\{\begin{array}{c}[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx1} & \boldsymbol{\Omega}_{xu1}\\ \mathbf{0} & \boldsymbol{\Omega}_{uu1}\end{array}\right]\\ +l_V(\mathbf{s}_x'\mathbf{y}_t)\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx2} & \boldsymbol{\Omega}_{xu2}\\ \mathbf{0} & \boldsymbol{\Omega}_{uu2}\end{array}\right]\end{array}\right\}+\\
&\quad g_V(\mathbf{s}_u'\mathbf{y}_t)\left\{\begin{array}{c}[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx3} & \boldsymbol{\Omega}_{xu3}\\ \mathbf{0} & \boldsymbol{\Omega}_{uu3}\end{array}\right]\\ +l_V(\mathbf{s}_x'\mathbf{y}_t)\left[\begin{array}{cc}\boldsymbol{\Omega}_{xx4} & \boldsymbol{\Omega}_{xu4}\\ \mathbf{0} & \boldsymbol{\Omega}_{uu4}\end{array}\right]\end{array}\right\}.
\end{aligned}$$

It follows that:

$$\begin{aligned}
&\boldsymbol{\Omega}_{xut}=[1-g_V(\mathbf{s}_u'\mathbf{y}_t)]\{[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\boldsymbol{\Omega}_{xu1}+l_V(\mathbf{s}_x'\mathbf{y}_t)\boldsymbol{\Omega}_{xu2}\}+\\
&g_V(\mathbf{s}_u'\mathbf{y}_t)\{[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\boldsymbol{\Omega}_{xu3}+l_V(\mathbf{s}_x'\mathbf{y}_t)\boldsymbol{\Omega}_{xu4}\}=\mathbf{G}_{12t}\boldsymbol{\epsilon}_{ut}^2,\\
&\boldsymbol{\Omega}_{uut}=[1-g_V(\mathbf{s}_u'\mathbf{y}_t)]\{[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\boldsymbol{\Omega}_{uu1}+l_V(\mathbf{s}_x'\mathbf{y}_t)\boldsymbol{\Omega}_{uu2}\}+\\
&g_V(\mathbf{s}_u'\mathbf{y}_t)\{[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\boldsymbol{\Omega}_{uu3}+l_V(\mathbf{s}_x'\mathbf{y}_t)\boldsymbol{\Omega}_{uu4}\}=\boldsymbol{\epsilon}_{ut}^2,
\end{aligned}$$

which yields:

$$\begin{aligned}\boldsymbol{\Omega}_{xut} &= \mathbf{G}_{12t}\boldsymbol{\Omega}_{uut}\\ \mathbf{G}_{12t} &= \boldsymbol{\Omega}_{xut}\boldsymbol{\Omega}_{uut}^{-1}\end{aligned}$$

with

$$\begin{aligned}\boldsymbol{\Omega}_{xut} &= [1-g_V(\mathbf{s}_u'\mathbf{y}_t)]\{[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\boldsymbol{\Omega}_{xu1}+l_V(\mathbf{s}_x'\mathbf{y}_t)\boldsymbol{\Omega}_{xu2}\}+\\ &\quad g_V(\mathbf{s}_u'\mathbf{y}_t)\{[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\boldsymbol{\Omega}_{xu3}+l_V(\mathbf{s}_x'\mathbf{y}_t)\boldsymbol{\Omega}_{xu4}\}\\ \boldsymbol{\Omega}_{uut} &= [1-g_V(\mathbf{s}_u'\mathbf{y}_t)]\{[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\boldsymbol{\Omega}_{uu1}+l_V(\mathbf{s}_x'\mathbf{y}_t)\boldsymbol{\Omega}_{uu2}\}+\\ &\quad g_V(\mathbf{s}_u'\mathbf{y}_t)\{[1-l_V(\mathbf{s}_x'\mathbf{y}_t)]\boldsymbol{\Omega}_{uu3}+l_V(\mathbf{s}_x'\mathbf{y}_t)\boldsymbol{\Omega}_{uu4}\}.\end{aligned}$$

The solution for $\mathbf{G}_{12t}$ can be used to construct matrix

$$\mathbf{H}_t^{-1}=\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix},$$

which can then be used to map the reduced-form model into a structural one. To this end, pre-multiply both sides of (3) by the right side of $\mathbf{H}_t^{-1}$ to obtain:

$$\begin{aligned}&\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{t+1}\\ \mathbf{u}_{t+1}\end{bmatrix}=\\ &[1-g_M(\mathbf{s}_u'\mathbf{y}_t)]\left\{\begin{array}{c}[1-l_M(\mathbf{s}_x'\mathbf{y}_t)]\left\{\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}_{x1}\\ \boldsymbol{\lambda}_{u1}\end{bmatrix}+\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_{x1}(L)\\ \boldsymbol{\Lambda}_{u1}(L)\end{bmatrix}\mathbf{y}_t\right\}\\ +l_M(\mathbf{s}_x'\mathbf{y}_t)\left\{\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}_{x2}\\ \boldsymbol{\lambda}_{u2}\end{bmatrix}+\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_{x2}(L)\\ \boldsymbol{\Lambda}_{u2}(L)\end{bmatrix}\mathbf{y}_t\right\}\end{array}\right\}\\ &+g_M(\mathbf{s}_u'\mathbf{y}_t)\left\{\begin{array}{c}[1-l_M(\mathbf{s}_x'\mathbf{y}_t)]\left\{\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}_{x3}\\ \boldsymbol{\lambda}_{u3}\end{bmatrix}+\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_{x3}(L)\\ \boldsymbol{\Lambda}_{u3}(L)\end{bmatrix}\mathbf{y}_t\right\}\\ +l_M(\mathbf{s}_x'\mathbf{y}_t)\left\{\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}_{x4}\\ \boldsymbol{\lambda}_{u4}\end{bmatrix}+\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda}_{x4}(L)\\ \boldsymbol{\Lambda}_{u4}(L)\end{bmatrix}\mathbf{y}_t\right\}\end{array}\right\}+\\ &\begin{bmatrix}\mathbf{I} & -\mathbf{G}_{12t}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{v}_{xt+1}\\ \mathbf{v}_{ut+1}\end{bmatrix}.\end{aligned}$$

This is solved as

$$\begin{aligned}&\begin{bmatrix}\mathbf{x}_{t+1}-\mathbf{G}_{12t}\mathbf{u}_{t+1}\\ \mathbf{u}_{t+1}\end{bmatrix}=\\ &[1-g_M(\mathbf{s}_u'\mathbf{y}_t)]\left\{\begin{array}{c}[1-l_M(\mathbf{s}_x'\mathbf{y}_t)]\left\{\begin{bmatrix}\boldsymbol{\lambda}_{x1}-\mathbf{G}_{12t}\boldsymbol{\lambda}_{u1}\\ \boldsymbol{\lambda}_{u1}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\Lambda}_{x1}(L)-\mathbf{G}_{12t}\boldsymbol{\Lambda}_{u1}(L)\\ \boldsymbol{\Lambda}_{u1}(L)\end{bmatrix}\mathbf{y}_t\right\}+\\ l_M(\mathbf{s}_x'\mathbf{y}_t)\left\{\begin{bmatrix}\boldsymbol{\lambda}_{x2}-\mathbf{G}_{12t}\boldsymbol{\lambda}_{u2}\\ \boldsymbol{\lambda}_{u2}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\Lambda}_{x2}(L)-\mathbf{G}_{12t}\boldsymbol{\Lambda}_{u2}(L)\\ \boldsymbol{\Lambda}_{u2}(L)\end{bmatrix}\mathbf{y}_t\right\}\end{array}\right\}\\ &+g_M(\mathbf{s}_u'\mathbf{y}_t)\left\{\begin{array}{c}[1-l_M(\mathbf{s}_x'\mathbf{y}_t)]\left\{\begin{bmatrix}\boldsymbol{\lambda}_{x3}-\mathbf{G}_{12t}\boldsymbol{\lambda}_{u3}\\ \boldsymbol{\lambda}_{u3}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\Lambda}_{x3}(L)-\mathbf{G}_{12t}\boldsymbol{\Lambda}_{u3}(L)\\ \boldsymbol{\Lambda}_{u3}(L)\end{bmatrix}\mathbf{y}_t\right\}+\\ l_M(\mathbf{s}_x'\mathbf{y}_t)\begin{bmatrix}\boldsymbol{\lambda}_{x4}-\mathbf{G}_{12t}\boldsymbol{\lambda}_{u4}\\ \boldsymbol{\lambda}_{u4}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\Lambda}_{x4}(L)-\mathbf{G}_{12t}\boldsymbol{\Lambda}_{u4}(L)\\ \boldsymbol{\Lambda}_{u4}(L)\end{bmatrix}\mathbf{y}_t\end{array}\right\}+\\ &\begin{bmatrix}\mathbf{v}_{xt+1}-\mathbf{G}_{12t}\mathbf{v}_{ut+1}\\ \mathbf{v}_{ut+1}\end{bmatrix}.\end{aligned}$$

The equations for the economy vector can be written in state-space form as

$$\left[\begin{array}{c}\mathbf{x}_{t+1}\\ \mathbf{u}_{t+1}\end{array}\right]=$$
$$\left[1-g_M\left(\mathbf{s}_u'\mathbf{y}_t\right)\right]\left\{\begin{array}{c}\left[1-l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left\{\left[\begin{array}{c}\boldsymbol{\lambda}_{x1}-\mathbf{G}_{12t}\boldsymbol{\lambda}_{u1}\\ \mathbf{0}\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{x1}\left(L\right)-\mathbf{G}_{12t}\boldsymbol{\Lambda}_{u1}\left(L\right)\\ \mathbf{0}\end{array}\right]\mathbf{y}_t\right\}+\\ l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left\{\left[\begin{array}{c}\boldsymbol{\lambda}_{x2}-\mathbf{G}_{12t}\boldsymbol{\lambda}_{u2}\\ \mathbf{0}\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{x2}\left(L\right)-\mathbf{G}_{12t}\boldsymbol{\Lambda}_{u2}\left(L\right)\\ \mathbf{0}\end{array}\right]\mathbf{y}_t\right\}\end{array}\right\}$$
$$+g_M\left(\mathbf{s}_u'\mathbf{y}_t\right)\left\{\begin{array}{c}\left[1-l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\right]\left\{\left[\begin{array}{c}\boldsymbol{\lambda}_{x3}-\mathbf{G}_{12t}\boldsymbol{\lambda}_{u3}\\ \mathbf{0}\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{x3}\left(L\right)-\mathbf{G}_{12t}\boldsymbol{\Lambda}_{u3}\left(L\right)\\ \mathbf{0}\end{array}\right]\mathbf{y}_t\right\}+\\ l_M\left(\mathbf{s}_x'\mathbf{y}_t\right)\left[\begin{array}{c}\boldsymbol{\lambda}_{x4}-\mathbf{G}_{12t}\boldsymbol{\lambda}_{u4}\\ \mathbf{0}\end{array}\right]+\left[\begin{array}{c}\boldsymbol{\Lambda}_{x4}\left(L\right)-\mathbf{G}_{12t}\boldsymbol{\Lambda}_{u4}\left(L\right)\\ \mathbf{0}\end{array}\right]\mathbf{y}_t\end{array}\right\}+$$
$$+\left[\begin{array}{c}\mathbf{G}_{12t}\\ \mathbf{I}\end{array}\right]\mathbf{u}_{t+1}+\left[\begin{array}{c}\mathbf{v}_{xt+1}-\mathbf{G}_{12t}\mathbf{v}_{ut+1}\\ \mathbf{0}\end{array}\right],$$

or equivalently as in (7)-(9).

B Solution of the NLQR problem

The objective of the NLQR is to set the sequence of $\{\mathbf{u}_t\}_{t=1}^{\infty}$ that minimizes (10) subject to (11), taking $\mathbf{y}_0$ as given. In any period $t\geq 0$ the guessed value function for that period $V\left(\mathbf{y}_t\right)=\mathbf{y}_t'\mathbf{P}_t\mathbf{y}_t+2\mathbf{y}'\mathbf{p}_t+p_t$ can be replaced into the Bellman equation and the system (11) can be used to form expectations. This yields:

$$V\left(\mathbf{y}_t\right)=\min_{\mathbf{u}_{t+1}}\left\{\begin{array}{c}\left(\mathbf{y}_t-\overline{\mathbf{y}}\right)'\mathbf{Q}\left(\mathbf{y}_t-\overline{\mathbf{y}}\right)+\\ \beta\left[\begin{array}{c}\mathbf{c}\left(\mathbf{y}_t\right)+\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t\\ +\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{u}_{t+1}\end{array}\right]'\mathbf{P}_t\left[\begin{array}{c}\mathbf{c}\left(\mathbf{y}_t\right)+\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t\\ +\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{u}_{t+1}\end{array}\right]+\\ \beta E_t\mathbf{e}_{t+1}'\mathbf{P}_t\mathbf{e}_{t+1}\\ 2\beta\left[\mathbf{c}\left(\mathbf{y}_t\right)+\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t+\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{u}_{t+1}\right]'\mathbf{p}_t+\beta p_t\end{array}\right\}.$$

Multiplying through gives

$$V\left(\mathbf{y}_t\right)=\min_{\mathbf{u}_{t+1}}\left[\begin{array}{c}\mathbf{y}_t'\mathbf{Q}\mathbf{y}_t+\overline{\mathbf{y}}'\mathbf{Q}\overline{\mathbf{y}}-2\overline{\mathbf{y}}'\mathbf{Q}\mathbf{y}_t+\\ \beta\mathbf{c}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right)+\beta\mathbf{y}_t'\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t+\\ \beta\mathbf{u}_{t+1}'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{u}_{t+1}+2\beta\mathbf{y}_t'\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right)+\\ 2\beta\mathbf{u}_{t+1}'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right)+2\beta\mathbf{u}_{t+1}'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t+\\ \beta tr\left(\boldsymbol{\Sigma}\mathbf{P}_t\right)+2\beta\mathbf{c}\left(\mathbf{y}_t\right)'\mathbf{p}_t+2\beta\mathbf{x}_t'\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{p}_t+\\ 2\beta\mathbf{u}_{t+1}'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{p}_t+\beta p_t\end{array}\right].$$

Differentiation of the above w.r.t. $\mathbf{u}_{t+1}$ gives

$$\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{u}_{t+1}+\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right)+\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t+\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{p}_t=\mathbf{0},$$

which yields the solution

$$\mathbf{u}_{t+1}=-\left[\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\right]^{-1}\left\{\mathbf{B}\left(\mathbf{y}_t\right)'\left[\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right)+\mathbf{p}_t\right]+\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t\right\}.$$

The above is then rewritten in terms of the feedback rule (12)-(14). Replacing the above solution into the Bellman equation and multiplying through yields:

$$\begin{array}{l} \mathbf{y}_t'\mathbf{P}_t\mathbf{y}_t + 2\mathbf{y}_t'\mathbf{p}_t + p_t \\ = \\ \left[\begin{array}{c} \mathbf{y}_t'\mathbf{Q}\mathbf{y}_t + \overline{\mathbf{y}}'\mathbf{Q}\overline{\mathbf{y}} - 2\overline{\mathbf{y}}'\mathbf{Q}\mathbf{y}_t + \beta\mathbf{c}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \beta\mathbf{y}_t'\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t \\ +\beta\mathbf{k}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{k}_t + \beta\mathbf{y}_t'\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{K}_t\mathbf{y}_t + \\ 2\beta\mathbf{y}_t'\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{k}_t + 2\beta\mathbf{y}_t'\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + 2\beta\mathbf{k}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) \\ +2\beta\mathbf{y}_t'\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + 2\beta\mathbf{k}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t + \\ 2\beta\mathbf{y}_t'\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right)\mathbf{y}_t + \beta tr\left[\mathbf{\Sigma}\mathbf{P}_t\right] + 2\beta\mathbf{c}\left(\mathbf{y}_t\right)'\mathbf{p}_t + \\ 2\beta\mathbf{y}_t'\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{p}_t + 2\beta\mathbf{k}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{p}_t + 2\beta\mathbf{y}_t'\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{p}_t + \beta p_t \end{array}\right]. \end{array}$$

Equating the quadratic terms gives:

$$\begin{array}{rcl} \mathbf{P}_t & = & \mathbf{Q} + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right) + \beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{K}_t + \\ & & 2\beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right). \end{array}$$

Using (14), if follows that $\beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{K}_t = -\beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{A}\left(\mathbf{y}_t\right)$ and the above simplifies as (15). Equating the linear terms gives:

$$\begin{array}{rcl} \mathbf{p}_t & = & -\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{k}_t + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \\ & & \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{k}_t + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{p}_t + \beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{p}_t. \end{array}$$

Using $\beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{k}_t = -\beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\left[\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \mathbf{p}_t\right]$ the above becomes

$$\begin{array}{rcl} \mathbf{p}_t & = & -\mathbf{Q}\overline{\mathbf{y}} - \beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\left[\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \mathbf{p}\right] + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \\ & & \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{k}_t + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{p}_t + \beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{p}_t, \\ & = & -\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{k}_t + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{p}_t. \end{array}$$

Using (13) to replace $\mathbf{k}_t$ gives

$$\begin{array}{rcl} \mathbf{p}_t & = & -\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) - \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\left[\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\right]^{-1}\mathbf{B}\left(\mathbf{y}_t\right)'\left[\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \mathbf{p}_t\right] \\ & & +\beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{p}_t, \\ & = & -\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) - \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\left[\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\right]^{-1}\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) \\ & & -\beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\left[\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\right]^{-1}\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{p}_t + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{p}_t. \end{array}$$

Using (14) yields:

$$\mathbf{p}_t = -\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \beta\mathbf{K}_t\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{p}_t + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{p}_t,$$

which simplifies as

$$\mathbf{p}_t\left[\mathbf{I} - \beta\mathbf{K}_t'\mathbf{B}\left(\mathbf{y}_t\right)' - \beta\mathbf{A}\left(\mathbf{y}_t\right)'\right] = -\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{A}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \beta\mathbf{K}_t\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right),$$

that is solved as (16). Finally, combining the constant terms gives:

$$\begin{array}{rcl} p_t & = & \overline{\mathbf{y}}'\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{c}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \beta\mathbf{k}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{B}\left(\mathbf{y}_t\right)\mathbf{k}_t + 2\beta\mathbf{k}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{P}_t\mathbf{c}\left(\mathbf{y}_t\right) + \\ & & \beta tr\left[\mathbf{\Sigma}_{t+1}\mathbf{P}_t\right] + 2\beta\mathbf{c}\left(\mathbf{y}_t\right)'\mathbf{p}_t + 2\beta\mathbf{k}_t'\mathbf{B}\left(\mathbf{y}_t\right)'\mathbf{p}_t + \beta p_t, \end{array}$$

so that

$$\begin{aligned} p_t(1-\beta) &= \overline{\mathbf{y}}'\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{c}(\mathbf{y}_t)'\mathbf{P}_t\mathbf{c}(\mathbf{y}_t) + \beta\mathbf{k}_t'\mathbf{B}(\mathbf{y}_t)'\mathbf{P}_t\mathbf{B}(\mathbf{y}_t)\mathbf{k}_t + \\ &\quad 2\beta\mathbf{k}_t'\mathbf{B}(\mathbf{y}_t)'\mathbf{P}_t\mathbf{c}(\mathbf{y}_t) + \beta tr[\mathbf{\Sigma}_{t+1}\mathbf{P}_t] + 2\beta\mathbf{c}(\mathbf{y}_t)'\mathbf{p}_t + 2\beta\mathbf{k}_t'\mathbf{B}(\mathbf{y}_t)'\mathbf{p}_t. \end{aligned}$$

Using $\beta\mathbf{k}_t'\mathbf{B}(\mathbf{y}_t)'\mathbf{P}_t\mathbf{B}(\mathbf{y}_t)\mathbf{k}_t = -\beta\mathbf{k}_t'\mathbf{B}(\mathbf{y}_t)'[\mathbf{P}_t\mathbf{c}(\mathbf{y}_t) + \mathbf{p}_t]$ gives:

$$\begin{aligned} p_t(1-\beta) &= \overline{\mathbf{y}}'\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{c}(\mathbf{y}_t)'\mathbf{P}_t\mathbf{c}(\mathbf{y}_t) - \beta\mathbf{k}_t'\mathbf{B}(\mathbf{y}_t)'[\mathbf{P}_t\mathbf{c}(\mathbf{y}_t) + \mathbf{p}_t] + \\ &\quad 2\beta\mathbf{k}_t'\mathbf{B}(\mathbf{y}_t)'[\mathbf{P}_t\mathbf{c}(\mathbf{y}_t) + \mathbf{p}_t] + \beta tr[\mathbf{\Sigma}_{t+1}\mathbf{P}_t] + 2\beta\mathbf{c}(\mathbf{y}_t)'\mathbf{p}_t, \end{aligned}$$

which simplifies as

$$\begin{aligned} p_t(1-\beta) &= \overline{\mathbf{y}}'\mathbf{Q}\overline{\mathbf{y}} + \beta\mathbf{c}(\mathbf{y}_t)'\mathbf{P}_t\mathbf{c}(\mathbf{y}_t) + \beta\mathbf{k}_t'\mathbf{B}(\mathbf{y}_t)'[\mathbf{P}_t\mathbf{c}(\mathbf{y}_t) + \mathbf{p}_t] + \\ &\quad \beta tr[\mathbf{\Sigma}_{t+1}\mathbf{P}_t] + 2\beta\mathbf{c}(\mathbf{y}_t)'\mathbf{p}_t \end{aligned}$$

and then rewritten as (17).

C Data

Credit risk

- Data sources:
  - Moody's, Moody's Seasoned Baa Corporate Bond Yield [BAA], retrieved from FRED, Federal Reserve Bank of St. Louis;
    https://fred.stlouisfed.org/series/BAA, January 27, 2019.
  - Board of Governors of the Federal Reserve System (US), 10-Year Treasury Constant Maturity Rate [DGS10], retrieved from FRED, Federal Reserve Bank of St. Louis;
    https://fred.stlouisfed.org/series/DGS10, January 27, 2019.
- The credit risk indicator, s, is calculated as the difference between BAA and DGS10.

Inflation

- Data source:
  - U.S. Bureau of Economic Analysis, Personal Consumption Expenditures [PCE], retrieved from FRED, Federal Reserve Bank of St. Louis;
    https://fred.stlouisfed.org/series/PCE, January 27, 2019.
- The inflation rate p is the percent change from year ago of PCE, monthly, seasonally adjusted annual rate.

Industrial Production

- Data source:

  - Board of Governors of the Federal Reserve System (US), Industrial Production Index [INDPRO], retrieved from FRED, Federal Reserve Bank of St. Louis;
    https://fred.stlouisfed.org/series/INDPRO, January 29, 2019.

- The growth rate of industrial production *ip* is the percent change from year ago of INDPRO, monthly, seasonally adjusted.

Unemployment gap

- Data sources:

  - U.S. Bureau of Labor Statistics, Unemployment Rate [UNRATE], retrieved from FRED, Federal Reserve Bank of St. Louis;
    https://fred.stlouisfed.org/series/UNRATE, January 27, 2019.
  - U.S. Congressional Budget Office, Natural Rate of Unemployment (Short-Term) [NROUST], retrieved from FRED, Federal Reserve Bank of St. Louis;
    https://fred.stlouisfed.org/series/NROUST, January 27, 2019.

- The series UNRATE is in percent, monthly, seasonally adjusted. NROUST is in percent, quarterly, not seasonally adjusted. NROUST is converted in a monthly series using the MATLAB function *interp1*. The unemployment gap *ug* is calculated as the difference between UNRATE and the monthly NROUST.

Federal funds rate

- Data source:

  - Board of Governors of the Federal Reserve System (US), Effective Federal Funds Rate [FEDFUNDS], retrieved from FRED, Federal Reserve Bank of St. Louis;
  - https://fred.stlouisfed.org/series/FEDFUNDS, January 27, 2019.

- The federal funds rate R is the monthly FEDFUNDS.

Fed's Balance Sheet

- Data sources (Total Assets held by the Fed and Total U.S. government securities securities held by the Fed (all data refer to the end of month, are in millions of dollars and not seasonally adjusted):

  - From January 2003 to January 2019, Board of Governors of the Federal Reserve System (US), Assets: Total Assets: Total Assets (Less Eliminations From Consolidation): Wednesday Level [WALCL], retrieved from FRED, Federal Reserve Bank of St. Louis;
    https://fred.stlouisfed.org/series/WALCL, January 27, 2019.
  - From June 1996 to December 2002, manually copied from the Consolidated Statement of Condition of All Federal Reserve Banks in the Fed releases of Factors Affecting Reserve Balances - H.4.1;
    https://www.federalreserve.gov/releases/h41/;

  - From May 1975 to January 1978, using data downloaded from: https://fraser.stlouisfed.org/title/83.

- Data sources (monthly series of nominal GDP):

  - (i) Quarterly data on nominal GDP taken from U.S. Bureau of Economic Analysis, Gross Domestic Product [GDP], retrieved from FRED, Federal Reserve Bank of St. Louis;
    https://fred.stlouisfed.org/series/GDP, January 27, 2019.
    This is monthly interporlated between 1979 and 1992 using the MATLAB function *interp1*;
  - (ii) Monthly data nominal GDP between January 1992 and October 2018 using US Monthly GDP (MGDP) Index from Macroeconomic Advisers by IHS Markit, downloaded from:
    https://ihsmarkit.com/products/us-monthly-gdp-index.html.

- All assets time series are scaled by nominal GDP. TS is the Total U.S. government securities held by the Fed in percent of GDP. PS is the difference between Total Assets held by the Fed in percent of GDP and TS.

D Impulse Response Functions

The algorithm for computing nonlinear structural IRFs with sign restrictions conditional on information at time t includes the following 9 steps.

1. Compute N^Q matrices of random elements, each having the same size of the covariance matrix $\mathbf{\Omega}_t$ in equation (2). Compute the N^Q orthogonal matrices $\mathbf{Q}^{nq}$, $nq = 1, ..., N^Q$ each obtained from the QR decomposition of one of these random matrices.

2. Set the sequence of lagged data up to period $t-1$ to define the history F_{t-1} at date t.

3. Generate a baseline sequence of structural shocks at date t for each variable in $\mathbf{y}_t$ over the time horizon $h = 0, ..., H$. Then generate a sequence of perturbated shocks, which is equal to the baseline except for the shock of interest that is set equal to the value in the baseline plus a prespecified increase, δ, denoting the magnitude of this shock.

4. Given F_{t-1} and the baseline sequence of shocks in step 3, generate N^Q new paths of realizations of $\mathbf{y}_{t+h}$, $h = 0, ..., H$, by recursively updating the VSTAR model (1)-(2) conditional on F_{t-1}, after pre-multiplying the Cholesky decomposition of the covariance matrix (2) by the orthogonal matrices $\mathbf{Q}^{nq}$, $nq = 1, ..., N^Q$.

5. Given F_{t-1} and the perturbated sequence of shocks in step 3, generate N^Q new paths of realizations of $\mathbf{y}_{t+h}$, $h = 0, ..., H$, by recursively updating the VSTAR model (1)-(2) conditional on F_{t-1}, after pre-multiplying the Cholesky decomposition of the covariance matrix (2) by the orthogonal matrices $\mathbf{Q}^{nq}$, $nq = 1, ..., N^Q$.

Times	Start	End	s	p	ip	ug	TS	PS	R
Normal	Jun-94	May-01	1.91	6.11	4.80	-0.48	4.93	0.79	5.44
Pre-ZLB	Jan-02	Dec-08	2.48	5.06	1.30	0.33	5.28	1.28	2.75
ZLB	Jan-09	Dec-15	3.00	3.05	0.63	2.50	9.81	9.57	0.13
Post-ZLB	Jan-16	Oct-18	2.30	4.30	1.05	-0.22	12.5	10.18	1.01

Table 6: Histories used for computing IRFs during normal times, pre-ZLB period, ZLB period and post-ZLB period.

6. Subtract each path for $\mathbf{y}_{t+h}$ in step 4 from the corresponding path for $\mathbf{y}_{t+h}$ in step 5, $h = 0, ..., H$. This gives N^Q estimates of the IRFs conditional on F_{t-1}.

7. Set aside the IRFs in step 6 that satisfy the required sign restrictions.

8. Since the IRFs in step 7 depends on the particular random draw for the structural shocks in 3, repeat steps 3 to 7 N^R times and compute the median of the resulting IRF estimates. By the law of large numbers, this median converges to the conditional IRFs of $\mathbf{y}_{t+h}$ at horizon $h = 0, 1, ..., H$, to a given shock conditional on F_{t-1}.

9. The unconditional IRFs of $\mathbf{y}_{t+h}$ at horizon $h = 0, 1, ..., H$, can be computed by conditioning on the average of the subset of all histories of interest. Alternatively, unconditional IRFs of $\mathbf{y}_{t+h}$ at horizon $h = 0, 1, ..., H$, can also be computed by repeating steps 2-8 over many histories F_{t-1}, each of which is randomly drawn with replacement from the original data, and then averaging the values of the resulting conditional IRFs. This second procedure is however more time consuming and computatinally intensive than the first.

The IRFs calculated in the paper are based on $H = 24$, $N^Q = N^Q = 1000$. In the data, the spread is negatively correlated with inflation and industrial production, while it is positively correlated with the unemployment gap. Thus, $\delta = 1$ and in simulating the effect of a (negative) demand shock the first element in each sequence of perturbated shocks in step 3 is set equal to the baseline plus delta. Effetively, this is equivalent to consider increase in the credit spread as a result of a demand shock. Confidence bands are computed using two standard deviations of the median IRF in step 8. The calculation of the IRF under the optimal policy uses the same algorithm described above except that the N^Q new paths of realizations of $\mathbf{y}_{t+h}$, $h = 0, ..., H$, are obtained by recursively updating the VSTAR model in (19) rather than the VSTAR in (1)-(2). The unconditional IRFs are computed using subsample averages reported in Table 6.

E Gain from Optimization: Robustness Results

	Volatilities					Loss	Stabilization	
	$(p_t - \overline{p})^2$	ug_t^2	ΔTS_t^2	ΔPS_t^2	ΔR_t^2	V	G	$\widehat{u}$
Policy	**Baseline**: $\upsilon_p = \upsilon_{ug} = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	6.36	2.66	0.32	0.02	0.11	9.46		
Optimal	2.23	2.46	0.37	0.10	0.20	5.37	43.24	2.02
	Weights II: $\upsilon_p = 0.5$, $\upsilon_{ug} = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	6.36	2.66	0.32	0.02	0.11	6.28		
Optimal	2.66	2.44	0.33	0.07	0.16	4.33	31.08	1.40
	Weights III: $\upsilon_{ug} = 0.5$, $\upsilon_p = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	6.36	2.66	0.32	0.02	0.11	8.13		
Optimal	2.19	2.58	0.38	0.09	0.20	4.15	49.00	2.82
	Weights IV: $\upsilon_{ug} = \upsilon_p = 1$, $\upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 0.5$							
Actual	6.36	2.66	0.32	0.02	0.11	9.23		
Optimal	1.93	2.33	0.46	0.15	0.27	4.70	49.10	2.13
	Weights V: $\upsilon_{ug} = \upsilon_p = \upsilon_{\Delta R} = 1$, $\upsilon_{\Delta TS} = \upsilon_{\Delta PS} = 0.5$							
Actual	6.36	2.66	0.32	0.02	0.11	9.39		
Optimal	2.06	2.32	0.37	0.16	0.25	4.96	47.20	2.11
	Weights VI: $\upsilon_{ug} = \upsilon_p = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = 1$, $\upsilon_{\Delta R} = 0.5$							
Actual	6.36	2.66	0.32	0.02	0.11	9.30		
Optimal	2.10	2.47	0.47	0.09	0.22	5.11	45.00	2.05

Table 7: Gains from monetary policy optimization computed from the VSTAR: 1979-2018

	Volatilities					Loss	Stabilization	
	$(p_t - \overline{p})^2$	ug_t^2	ΔTS_t^2	ΔPS_t^2	ΔR_t^2	V	G	$\widehat{u}$
Policy	**Baseline**: $\upsilon_p = \upsilon_{ug} = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	3.63	5.55	0.02	0.05	0.38	9.64		
Optimal	2.06	5.49	0.03	0.06	0.26	7.89	18.07	1.32
	Weights II: $\upsilon_p = 0.5$, $\upsilon_{ug} = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	3.63	5.55	0.02	0.05	0.38	7.82		
Optimal	2.46	5.48	0.02	0.05	0.25	7.03	10.15	0.89
	Weights III: $\upsilon_{ug} = 0.5$, $\upsilon_p = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	3.63	5.55	0.02	0.05	0.38	6.86		
Optimal	2.06	5.53	0.02	0.06	0.26	5.17	24.58	1.84
	Weights IV: $\upsilon_{ug} = \upsilon_p = 1$, $\upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 0.5$							
Actual	3.63	5.55	0.02	0.05	0.38	9.41		
Optimal	1.60	5.41	0.03	0.09	0.30	7.22	23.22	1.48
	Weights V: $\upsilon_{ug} = \upsilon_p = \upsilon_{\Delta R} = 1$, $\upsilon_{\Delta TS} = \upsilon_{\Delta PS} = 0.5$							
Actual	3.63	5.55	0.02	0.05	0.38	9.42		
Optimal	1.63	5.36	0.02	0.09	0.30	7.21	23.47	1.49
	Weights VI: $\upsilon_{ug} = \upsilon_p = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = 1$, $\upsilon_{\Delta R} = 0.5$							
Actual	3.63	5.55	0.02	0.05	0.38	9.62		
Optimal	2.02	5.54	0.04	0.06	0.26	7.89	18.06	1.32

Table 8: Gains from monetary policy optimization computed from the VAR: 2008-2017

	Volatilities					Loss	Stabilization	
	$(p_t - \overline{p})^2$	ug_t^2	ΔTS_t^2	ΔPS_t^2	ΔR_t^2	V	G	$\widehat{u}$
Policy	**Baseline**: $\upsilon_p = \upsilon_{ug} = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	6.36	2.66	0.32	0.02	0.11	9.46		
Optimal	3.54	2.61	0.26	0.05	0.20	6.67	29.46	1.67
	Weights II: $\upsilon_p = 0.5$, $\upsilon_{ug} = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	6.36	2.66	0.32	0.02	0.11	6.28		
Optimal	4.26	2.61	0.25	0.03	0.14	5.16	17.78	1.06
	Weights III: $\upsilon_{ug} = 0.5$, $\upsilon_p = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 1$							
Actual	6.36	2.66	0.32	0.02	0.11	8.13		
Optimal	3.55	2.64	0.26	0.05	0.19	5.37	33.89	2.35
	Weights IV: $\upsilon_{ug} = \upsilon_p = 1$, $\upsilon_{\Delta TS} = \upsilon_{\Delta PS} = \upsilon_{\Delta R} = 0.5$							
Actual	6.36	2.66	0.32	0.02	0.11	9.23		
Optimal	2.75	2.58	0.27	0.10	0.32	5.68	38.48	1.89
	Weights V: $\upsilon_{ug} = \upsilon_p = \upsilon_{\Delta R} = 1$, $\upsilon_{\Delta TS} = \upsilon_{\Delta PS} = 0.5$							
Actual	6.36	2.66	0.32	0.02	0.11	9.39		
Optimal	2.85	2.55	0.26	0.11	0.33	5.89	37.35	1.87
	Weights VI: $\upsilon_{ug} = \upsilon_p = \upsilon_{\Delta TS} = \upsilon_{\Delta PS} = 1$, $\upsilon_{\Delta R} = 0.5$							
Actual	6.36	2.66	0.32	0.02	0.11	9.30		
Optimal	3.38	2.64	0.29	0.05	0.19	6.41	31.07	1.70

Table 9: Gains from monetary policy optimization computed from the VAR: 1979-2018