On the Empirical (Ir)Relevance of the Zero Lower Bound Constraint*

Davide Debortoli
Universitat Pompeu Fabra, CREI
and Barcelona GSE

Jordi Galí
CREI, Universitat Pompeu Fabra
and Barcelona GSE

Luca Gambetti
Universitat Autonoma de Barcelona
and Barcelona GSE

January 2018

Abstract

We estimate a time-varying structural VAR that describes the dynamic responses of a number of U.S. macro variables to different identified shocks. We find no significant changes in the estimated responses over the period when the federal funds rate attained the zero lower bound (ZLB). This result is consistent with the hypothesis of "perfect substitutability" between conventional and unconventional monetary policies. Montecarlo simulations based on artificial time series generated from a standard New Keynesian model point to the validity of our empirical approach to detect the changes in equilibrium dynamics associated with ZLB episodes.

JEL Classification No.: E44, E52

Keywords: regime changes, liquidity trap, unconventional monetary policies, time-varying structural vector-autoregressive models.

*E-mails: `davide.debortoli@upf.edu`, `jgali@crei.cat`, and `luca.gambetti@uab.cat`. The authors aknowledge the financial support of the Spanish Ministry of Economy and Competitiveness, through grants ECO2014-55555-P (Debortoli), ECO2017-87827-P (Galí) and ECO2015-67602-P (Gambetti), and the support of the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563) and the Barcelona Graduate School of Economics.

1 Introduction

The magnitude of the global financial crisis and the subsequent recession and disinflation episode has forced many central banks to lower their policy rates down to their theoretical zero lower bound (ZLB). Given the impossibility of further reductions in the policy rate, central banks have increasingly relied on unconventional monetary policies (UMPs) i.e. policies that seek to substitute for changes in the short-term nominal rate –the instrument of monetary policy in normal times– when the latter attains the ZLB. Two prominent examples of unconventional policies adopted by several central banks in recent years include (i) forward guidance and (ii) large asset purchase programs. Both policies can be interpreted as attempts to stimulate economic activity through a variety of channels: lowering expectations of future policy rates, reducing the term and/or risk premia of longer term debt and, more generally, increasing the overall liquidity of the financial system.

In the face of these policy challenges, a growing literature has emerged that aims at evaluating empirically the effectiveness of different unconventional monetary policies, using a variety of approaches.[1] The present paper seeks to contribute to that literature. Our empirical approach involves the estimation of a structural vector autoregressive model with time-varying coefficients (TVC-SVAR) describing the dynamic responses of a number of U.S. macro variables to different shocks (both supply and demand), as well as the changes over time in those responses. Under the null hypothesis of "perfect substitutability" between conventional and unconventional monetary policies we should not find any significant change in the responses of real variables over the period during which the Federal Funds Rate (henceforth, the policy rate) attained its zero lower bound (from January 2009 through December 2015). This is indeed what we find, suggesting that UMPs have been highly effective at getting around the ZLB constraint.

[1] See below for a review of the literature.

We carry out our analysis using two alternative empirical models, which differ in terms of the variables they contain and the number of shocks they allow for. Our baseline model, similar to that in Galí and Gambetti (2009), uses data on work hours and labor productivity and allows for two shocks: "technology" and "demand." We compare the estimated responses to those shocks during non-ZLB and ZLB periods with those predicted by a standard New Keynesian model with a monetary authority following a nonlinear Taylor rule subject to the ZLB. Our findings are stark: while the theoretical model implies very large differences in the responses to shocks depending on whether the economy is under a binding ZLB or not, we cannot uncover any significant such differences in the U.S. economy, suggesting that UMPs have been highly effective at getting around the ZLB. In order to rule out that such a finding is not due to an inability of the empirical model to uncover such changes we estimate our TV-VAR using artificial time series generated by our theoretical model, and show that, contrary to what we see when using actual data, the estimated responses display large differences between ZLB and non-ZLB periods.

Our findings using the previous baseline model carry over to an extended empirical model which also includes data on long term interest rates and inflation in addition to hours and labor productivity, decomposing each variable into four shocks "technology," "demand," "monetary policy," and "other."

The remainder of the paper is organized as follows. Section 2 describes the baseline empirical model. Section 3 reports the estimated responses and their changes over time and compares them with the responses implied by a standard New Keynesian model. Section 4 describes the Montecarlo simulations. Section 5 reports evidence based on the extended model. Section 6 discusses the related literature. Section 7 concludes.

2 Empirical Model

The present section describes our baseline empirical model, which consists of a structural vector autoregression model with time-varying coefficients (TVC-SVAR), which we use to estimate the dynamic responses of hours, productivity and output to identified "technology" and "demand" shocks, as in Galí and Gambetti (2009).[2] Our identification approach, based on Galí (1999), assumes that only technology shocks may have a permanent effect on labor productivity. The main motivation for using a model with time-varying coefficients lies in our interest in assessing the extent to which the ZLB episode experienced by the U.S. economy over the period 2009:Q1-2015:Q4 led to a change in the way the U.S. economy responded to different shocks. In addition, the use of a TVC-SVAR provides a flexible specification which allows for other structural changes that the U.S. economy may have experienced over the full sample period.[3]

2.1 Model Specification

Let $\mathbf{x}_t$ be an n-dimensional time series vector. We assume that the vector admits the following TVC-VAR representation

$$\mathbf{x}_t = \mathbf{A}_{0,t} + \mathbf{A}_{1,t}\mathbf{x}_{t-1} + \mathbf{A}_{2,t}\mathbf{x}_{t-2} + ... + \mathbf{A}_{p,t}\mathbf{x}_{t-p} + \mathbf{u}_t \tag{1}$$

where $\mathbf{A}_{0,t}$ is a vector of time-varying intercepts, $\mathbf{A}_{i,t}$, for $i = 1, ..., p$ are matrices of time-varying coefficients, and $\mathbf{u}_t$ is a Gaussian white noise vector process with covariance matrix $\boldsymbol{\Sigma}_t$. The reduced form innovations are assumed to be a time-varying linear transformation of the underlying structural shocks $\boldsymbol{\varepsilon}_t$ given by

$$\mathbf{u}_t \equiv \mathbf{Q}_t\boldsymbol{\varepsilon}_t$$

[2] Though focusing on different variables, the specification of our reduced form time-varying VAR follows closely that in Primiceri (2005).

[3] These may include the change in the cyclical behavior of productivity emphasized in Galí and Gambetti (2009), as well as the change in monetary policy starting with Paul Volcker's tenure at the Fed uncovered in Clarida, Galí and Gertler (2000).

where $\mathbb{E}\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\} = I$ and $\mathbb{E}\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-k}'\} = 0$ for all t and $k = 1, 2, 3, ...$ It follows that $\mathbf{Q}_t\mathbf{Q}_t' = \boldsymbol{\Sigma}_t$.[4]

Let $\boldsymbol{\theta}_t = vec(\mathbf{A}_t')$ where $\mathbf{A}_t = [\mathbf{A}_{0,t}, \mathbf{A}_{1,t}..., \mathbf{A}_{p,t}]$ and $vec(\cdot)$ is the column stacking operator. The vector $\boldsymbol{\theta}_t$ is assumed to evolve according to the following equation:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \tag{2}$$

where $\boldsymbol{\omega}_t$ is Gaussian white noise vector process with covariance matrix $\boldsymbol{\Omega}$.

Time variation of $\boldsymbol{\Sigma}_t$ is modeled in the standard way. Let $\boldsymbol{\Sigma}_t = (\mathbf{F}_t^{-1})\mathbf{D}_t(\mathbf{F}_t^{-1})'$, where $\mathbf{F}_t^{-1}$ is lower triangular, with ones on the main diagonal, and $\mathbf{D}_t$ a diagonal matrix. The vector containing the diagonal elements of $\mathbf{D}_t^{1/2}$, denoted by $\boldsymbol{\sigma}_t$, is assumed to evolve according to the process

$$\log \boldsymbol{\sigma}_t = \log \boldsymbol{\sigma}_{t-1} + \boldsymbol{\zeta}_t. \tag{3}$$

Moreover let $\boldsymbol{\phi}_{i,t}$ denote the column vector with the non-zero elements of the $(i+1)$-th row of $\mathbf{F}_t$. We assume

$$\boldsymbol{\phi}_{i,t} = \boldsymbol{\phi}_{i,t-1} + \boldsymbol{\nu}_{i,t} \tag{4}$$

where $\boldsymbol{\zeta}_t$ and $\boldsymbol{\nu}_{i,t}$ are Gaussian white noise vector processes with zero mean and (constant) covariance matrices $\boldsymbol{\Xi}$ and $\boldsymbol{\Psi}_i$, respectively. We further assume that $\boldsymbol{\nu}_{i,t}$ is independent of $\boldsymbol{\nu}_{j,t}$, for all $j \neq i$, and that $\boldsymbol{\omega}_t$, $\boldsymbol{\varepsilon}_t$, $\boldsymbol{\zeta}_t$ and $\boldsymbol{\nu}_{i,t}$ (for all i) are mutually independent. Estimation is carried out as in Del Negro and Primiceri (2013).[5]

Reduced form impulse response functions can be derived from the local moving average (MA) representation of the model. First let us consider the companion form representation of eq. (1):

$$\tilde{\mathbf{x}}_t = \tilde{\mu}_t + \tilde{\mathbf{A}}_t\tilde{\mathbf{x}}_{t-1} + \tilde{\mathbf{u}}_t$$

[4] The assumption that $\mathbb{E}\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\} = I$ is without loss of generality, since the scale of the coefficients of the matrix $\mathbf{Q}_t$ is unrestricted.

[5] We refer the reader to Galí and Gambetti (2016) for details.

where $\tilde{\mathbf{x}}_t \equiv [\mathbf{x}'_t, \mathbf{x}'_{t-1}, ..., \mathbf{x}'_{t-p+1}]'$, $\tilde{\mathbf{u}}_t \equiv [\mathbf{u}'_t, 0, ..., 0]'$, $\tilde{\mu}_t \equiv [\mathbf{A}'_{0,t}, 0, ..., 0]'$ and $\tilde{\mathbf{A}}_t \equiv \begin{pmatrix} & \mathbf{A}_t \\ I_{n(p-1)} & \mathbf{0}_{n(p-1),n} \end{pmatrix}$ is the corresponding companion matrix. The (local) time-varying reduced form MA representation of the model is given by

$$\mathbf{x}_t = \mathbf{b}_t + \sum_{j=0}^{\infty} \mathbf{B}_{t,j} \mathbf{u}_{t-j}$$

where $\mathbf{b}_t$ is a vector of time-varying means and $\mathbf{B}_{t,j}$, for $j = 0, 1, 2, \ldots$, is given by the first n rows and n columns of the matrix $\tilde{\mathbf{A}}_t$ raised to the power of j.

2.2 Identification

In our baseline bivariate VAR we identify the technology shock following Galí (1999). In particular, denoting with y_t (log) output and with n_t (log) hours worked, we consider a VAR for the vector $\mathbf{x}_t \equiv [\Delta(y_t - n_t), n_t]'$. The technology shock is defined to be the only shock in vector $\boldsymbol{\varepsilon}_t$ to have a long run effect on (log) labor productivity $y_t - n_t$.

The identification is implemented as follows. Let $\mathbf{S}_t$ be the Cholesky factor of $\mathbf{B}_t(1)\mathbf{\Sigma}_t\mathbf{B}_t(1)'$ where $\mathbf{B}_t(1) = \sum_{j=0}^{\infty} \mathbf{B}_{t,j}$. Then let $\mathbf{Q}_t = \mathbf{B}_t(1)^{-1}\mathbf{S}_t$. The dynamic responses of the variables in $\mathbf{x}_t$ to the two structural shocks hitting the economy at time t at an horizon of j periods ahead are given by

$$\mathbf{C}_{t,j} = \mathbf{B}_{t,j}\mathbf{Q}_t \tag{5}$$

for $j = 0, 1, 2, ...$ The first shock is the technology shock. We refer to the second shock as a "demand" shock.

In the four-variable VAR we identify 4 shocks: technology, demand, monetary policy and temporary supply shock. We use a mix of long run and sign restrictions. The restrictions used are the following: (i) the technology shock is the only shock driving labor productivity in the long run; (ii) a positive demand shock has a positive effect on prices, GDP and the long-term interest rate at a one-year horizon; (iii) an expansionary monetary policy shock has a positive effect on prices, GDP and a negative effect on the long-term interest rate at a one-year horizon; (iv) a positive supply shock has a negative effect on prices and a positive effect on GDP at a one-year horizon.

The identification is implemented as follows. Let

$$\mathbf{H}_t^n = \begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mathbf{H}_t^{n-1} \end{pmatrix}$$

where $\mathbf{0}$ is n-dimensional column vector of zeros, $\mathbf{H}_t^n$ and $\mathbf{H}_t^{n-1}$ are orthogonal matrices of dimension $n \times n$ and $n-1 \times n-1$ respectively. To impose the restrictions we follow the standard algorithm of Rubio-Ramirez, Waggoner and Zha (2010). We draw $\mathbf{H}_t^{n-1}$ using the QR decomposition and compute the potential structural impulse response functions as in (5) with $\mathbf{Q}_t = \mathbf{B}_t(1)^{-1}\mathbf{S}_t\mathbf{H}_t^n$. We retain the draw if the sign restrictions are satisfied. We collect a total of 500 draws at each point in time. With no loss of generality we order the shocks as follows: technology, demand, monetary policy and supply.

The responses plotted in the figures below are the average estimated responses over two sample periods of equal length $T = 28$ quarters: the ZLB period (2009Q1-2015Q4) and the seven-year period preceding it (2002Q1-2008Q4; henceforth, the pre-ZLB period). We have taken the latter period as a benchmark for comparison due to its proximity to the ZLB period, and given the fact that over a longer sample period other factors unrelated to the ZLB may have caused structural changes in the economy that could be reflected in the estimated impulse reponses.[6] For each sample period, we compute the average impulse responses at horizon j as

$$\bar{\mathbf{C}}_j = T^{-1} \sum_t \mathbf{C}_{t,j}.$$

2.3 Data

We use quarterly U.S. data drawn from the FRED database, spanning the period 1948Q1 to 2015Q4. The variables considered are: real output per hour of all persons (nonfarm

[6] See, e.g. Galí and Gambetti (2009) for evidence of the changes associated with the outset of the Great Moderation.

business sector) (OPHNFB); hours of all persons (nonfarm business sector) (HOANBS); civilian noninstitutional population (CNP16OV); GDP deflator (GDPDEF); 10-Year Treasury Constant Maturity Rate (GS10). In the bivariate VAR the specification includes the growth rate of labor productivity (OPHNFB) and the log of per-capita hours (HOANBS over CNP16OV). The four-variable VAR extends the bivariate model by including also the growth rate of the GDP deflator (GDPDEF) and the 10 year bond yeld (GS10).

3 Evidence from Bivariate VAR

Figure 1 displays the *average* impulse responses of labor productivity, hours and GDP (all expressed in logs) to a (one standard deviation) shock to technology (left panel) and demand (right panel), the ZLB period (solid blue) and the pre-ZLB period (dashed red). The figure also displays the 68 percent confidence band corresponding to the impulse responses for the ZLB period.

Our main finding, captured in Figure 1, is easy to summarize: there are no significant differences between the ZLB and pre-ZLB sample periods in the responses of labor productivity, hours, and output to either technology or demand shocks. In fact, the point estimates suggest that, if anything, the response of hours and output to a one standard deviation demand shock during the ZLB period seems somewhat dampened relative to the earlier sample.

The previous evidence may be given alternative interpretations. A first interpretation, and the one we end up favoring, is that the adoption and implementation of unconventional monetary policies, mostly in the form of QE programs and forward guidance, have proved to be extremely effective, fully compensating for the Fed's inability (or unwillingness, partly) to lower the policy rate below the zero threshold. Under that interpretation, the deep slump experienced by the U.S. economy during 2008-2009 would be a consequence of unusually large shocks, but would not have been made worse by a binding ZLB.

Under a second interpretation, however, the previous findings just reflect the inability of our empirical approach, based on TVC-SVAR, to capture any potential changes in the economy's response to shocks resulting from the ZLB constraint. In the next section we assess the power of our empirical method by means of a Montecarlo simulation.

3.1 Montecarlo Simulations

In this section we report the findings from the application of the TVC-SVAR approach used above to the time series generated by a basic New Keynesian model subject to a ZLB constraint.

In particular, we consider a standard business cycle model with a representative household, and a continuum of identical monopolistically competitive firms, each facing (quadratic) price adjustment costs as in Rotemberg (1982).[7] Economic fluctuations are driven by two exogenous driving forces: technology shocks and a discount rate shifter, the latter being the source of aggregate demand changes.

Consistenly with the identification assumption in the previous section, we assume that technology follows a random walk process (in logs), given by $a_t = a_{t-1} + \epsilon_t^a$, where $\epsilon_t^a \sim N(0, \sigma_a)$. The discount rate shifter is modeled as the sum of two componenents, i.e. $\gamma_t = z_t + \rho_t$. The term z_t represents "recurrent" demand shocks over the business cycle, and follows an *AR(1)* process, i.e. $z_t = \rho_z z_{t-1} + \epsilon_t^z$ with $0 < \rho_z < 1$ and $\epsilon_t^a \sim N(0, \sigma_z)$. The term ρ_t captures instead an "unusual" (and large) discount rate shift, and is assumed to follow a two-state Markov chain, i.e. $\rho_t \in \{\rho_L, \rho\}$ where $\rho_L < 0 < \rho$, with transition matrix $\mathbf{M}_\rho$. This specification is convenient to analyze the effects of supply and demand shocks both in normal times (when $\rho_t = \rho$) and at the ZLB (when $\rho_t = \rho_L$).

[7] In particular, we assume that the representative household's utility function is $U_t \equiv \log(C_t) - (N^\phi)/(1+\phi)$ and the production function is $Y_t = \exp(a_t) N_t^{1-\alpha}$.

The equilibrium conditions of this economy can be summarized by means of the following five equations:

$$\frac{1}{1+r_t} = \mathbb{E}_t \frac{\Lambda_{t,t+1}}{\Pi_{t+1}} \tag{6}$$

$$\Pi_t (\Pi_t - 1) = \mathbb{E}_t \Lambda_{t,t+1} \frac{Y_{t+1}}{Y_t} (\Pi_{t+1} - 1) + \frac{\epsilon}{\theta} \left(MC_t - \frac{\epsilon - 1}{\epsilon} \right) \tag{7}$$

$$MC_t = \frac{1}{1-\alpha} \exp \left(-\frac{1+\phi(1-\alpha)}{1-\alpha} a_t \right) C_t Y_t^{\phi + \frac{\alpha}{1-\alpha}} \tag{8}$$

$$C_t = \Delta_t^p Y_t \tag{9}$$

$$i_t = \max \left[\rho + \phi_\pi \log(\Pi_t) + \phi_y \log(Y_t / Y_{t-1}), 0 \right]. \tag{10}$$

Equation (6) is the household's intertemporal optimality condition —a consumption Euler equation— decribing the relationship between the nominal interest rate i_t, expected inflation Π_{t+1}, and the stochastic discount factor $\Lambda_{t,t+1} \equiv \exp(-\gamma_{t+1})(C_t/C_{t+1})$, where C_t denotes household's consumption and γ_{t+1} is the discount rate shifter. Equation (7) is a New-Keynesian Phillips curve, representing the firms' optimal pricing behavior, and describes the relationship between inflation (current and future) and real marginal costs MC_t, and where ϵ and θ denotes the (Dixit-Stiglitz) elasticity of substitution between goods varieties and the price adjustment cost parameter, respectively. The evolution of the marginal cost, which can be obtained combining the household's labor supply and the production function, is described by equation (8), where Y_t denotes aggregate output, $1 - \alpha$ is the labor income share, and ϕ is the (inverse) Frisch elasticity of substitution. Equation (9) is the market clearing condition, where $\Delta_t^p \equiv \left[1 - \frac{\theta}{2} (\Pi_t - 1)^2\right]$ denotes the price adjustment costs. Finally, equation (10) is the central bank's interest rate rule, subject to a ZLB constraint.

The equilibrium of the economy is given by the sequence $\{C_t, Y_t, R_t, MC_t, \Pi_t\}_{t=0}^{\infty}$ satisfying equations (6)-(10), given the exogenous processes $\{a_t, \gamma_t\}_{t=0}^{\infty}$ and initial output Y_{-1}. Similarly to Fernández-Villaverde et. al. (2015), we solve the model using a global projection method to accurately account for the effect that uncertainty about the ZLB has on the economic decisions of households and firms.[8]

We adopt a quarterly calibration of the model. The discount rate ρ_t switches between a normal value of $\rho = 0.01$ and a low value $\rho_L = -0.004$, where in each period the probability of remaining in the normal regime is 99.6 percent, while the probability of remaining in a low-rate regime is 96.4 percent. These values imply that ZLB episodes occurs on average once every 270 quarters, and each episode lasts on average 28 quarters, which is consistent with the U.S. postwar experience. Furthermore, we set the parameter $\rho_z = 0.8$ and the standard deviations $\sigma_z = \sigma_a = 0.1$ percent. Under this calibration demand and technology shocks equally contribute to output fluctuations in normal times. The remaining parameters are set to standard values.[9]

Figure 2 displays the impulse responses of (log) labor productivity, (log) hours, and (log) output, to a technology and a discount rate shock under the two regimes, non-binding ZLB (dashed red) and binding ZLB (solid blue). Note that when the ZLB is binding the three variables experience a much larger response to the demand shock than under a nonbinding ZLB, due to the lack of a stabilizing monetary policy response. In response to a technology shock, however, the differences appear to be smaller, and in the case of labor productivity and hours, the responses are more muted under a binding ZLB regime. In both cases, however, the response under the nonbinding ZLB regime is closer to the efficient one (not shown).

Figure 3 shows the estimated responses obtained by applying the TVC-SVAR method described above to artificial data generated by our baseline New Keynesian model. More specifically, we generate 200 random samples for the time series of (log) labor productivity and (log) hours. Each sample has a length of 270 quarters, and we "force" the

[8]In particular, we approximate the model policy functions with Chebyshev polynomials (or splines), using a collocation method on a discrete grid for the three state variables (a_t, γ_t, Y_{t-1}).

[9]In particular, the (inverse) Frisch elasticity $\phi = 1$, the labor income share $1 - \alpha = 2/3$, the elasticity of substitution between good varieties $\epsilon = 6$, and the price adjustment cost parameter $\theta = 117.64$, which imply the same slope of the Phillips curve as in model with sticky prices a la Calvo and an average price duration of 4 quarters. The interest rate rule parameters are set to $\phi_\pi = 1.5$ and $\phi_y = 0.125$.

zero lower bound to be binding for the last 28 quarters, as in the actual data used in Section 3.[10]

Note that in the case of demand shocks, the average estimated responses corresponding to the ZLB period have a substantially larger size than those corresponding to the pre-ZLB period. The same is true for output –though not for labor productivity and hours– in response to a technology shock. The previous pattern of responses is fully consistent, at least in a qualitative sense, with the predictions of the New Keynesian model described above, thus pointing to the ability of the TVC-SVAR method to uncover potential differences in the economy's responses to shocks in an environment in which the ZLB is binding in a way that is not offset by any other policy intervention (e.g. unconventional monetary policies).[11]

Accordingly, the absence of any significant differences between the pre-ZLB and ZLB periods when we implement the same empirical approach to actual U.S. data suggests that the ZLB may have been binding over the 2009Q1-2005Q4 period only in a nominal sense, without altering the normal response of aggregate variables to shocks, possibly as a result of the unconventional monetary policies implemented by the Fed during that episode.

4 Evidence from a Larger VAR

In this section we report the evidence based on a four-variable TVC-SVAR that includes data on inflation and long-term nominal yields, in addition to (log) labor productivity and (log) hours. As discussed in Section 2.2, we identify four shocks –labeled as technology, demand, monetary policy and supply– using a combination of long-run and sign

[10]In particular, we assume that $\rho_t = 0.01$ for the first 242 quarters, while $\rho_t = -0.004$ for the last 28 quarters, and check that the ZLB is binding only over the latter period.

[11]This result is also consistent with the findings of Gust et. al. (2017), who perform a counterfactual excercise in an richer New Keynesian model estimated using actual U.S. data, and conclude that the ZLB constraint amplified the recession and slowed down the recovery.

restrictions.

Figure 4 displays the average estimated impulse responses of several macro variables to the four shocks, for both the pre-ZLB period (dashed red) and the ZLB period (solid blue). The differences in the estimated responses between the two periods are very small, for all variables and shocks. Most importantly, note that the lack of a significant gap between the responses in the two periods carries over to the nominal and real long-term yields, which are aguably more relevant in the determination of aggregate demand than the very short term nominal rates that were subject to a binding ZLB constraint. That finding is consistent with the hypothesis that by means of unconventional policies like forward guidance and quantitative easing the Fed managed to steer long terms rates in response to shocks as in normal times.

A potential explanation for the absence of a significant gap in the estimated responses across periods might point to the inability of the TVC-SVAR method to detect differences across monetary policy regimes of the order found in actual data. In order to evaluate that possibility we compare the average estimated responses during the ZLB period reported earlier to the average estimated responses corresponding to a period of equal length at the end of the pre-Volcker era. The choice of the latter period for the purposes of comparison is motivated by independent evidence suggesting that it was characterized by a monetary policy rule remarkably different from that adopted in the Volcker and post-Volcker eras.[12] Figure 5 displays the average impulse responses for the two periods considered, pointing to substantial differences between them, which helps discard the hypothesis of little power of our method.[13]

Finally, we have analyzed the contribution of demand shocks to fluctuations in GDP growth. In theory, as illustrated in section 3.1, when the ZLB is binding demand shocks

[12]See, e.g., Clarida et al. (2000)

[13]In particular, note that the estimates detect the differences in the response of labor productivity to demand shocks emphasized in Galí and Gambetti (2009), as well as the differences in the response of hours uncovered in Galí, López-Salido and Vallés (2003).

should become a more important source of business cycle fluctuations than under a nonbinding ZLB, due to the lack of a stabilizing monetary policy response. On the contrary, we find that the relative contribution of demand shocks declined significantly in the 1980s —possibly because of the adoption of better monetary policies— but did not increase during the ZLB period. This result provides further evidence consistent with the "perfect substitutability" hypothesis between conventional and unconventional monetary policies.[14]

5 Related Literature

Our paper contributes to a growing literature that aims at assessing the effectiveness of unconventional monetary policies through alternative approaches.

Swanson and Williams (2014) estimate the time-varying sensitivity of yields to macroeconomic announcements using high-frequency data, and conclude that long-term yields were essentially unconstrained throughout 2008 to 2012, and short-term yields seemed to be constrained only by late 2011. D'Amico and King (2013) use security-level data on Treasury prices and quantities to document a "local supply" effect along the yield curve during the large-scale aseet purchase (LSAP) interventions starting in 2009, documenting both a substantial response of yields to changes in supplies outstanding of a given maturity ("stock effect") as well as to the purchases themselves when they occurred ("flow effect"). The segmentation of the Treasury market suggested by their evidence would make it possible for QE programs to help stabilize the economy in the face of a binding ZLB constraint, thus overcoming the "irrelevance" result that emerges in frictionless settings.[15] Krishnamurthy and Vissing-Jorgensen (2011) and Hamilton and Wu

[14]In particular, we find that on impact the relative contribution of demand shocks to the forecast error of GDP growth exceeded 35% in the 1970s, and remained below 20% throughout the 2000s, including during the ZLB period. A similar result also holds for other forecast horizons, as shown in the online appendix.

[15]See, e.g., Eggertsson and Woodford (2003).

(2012) provide related evidence of relative supply effects on the yield curve.

In a more recent paper, D'Amico and King (2017) use a VAR with sign restrictions on survey forecasts and uncover strong and persistent effects on inflation and output of forward guidance policies, i.e. of policy interventions that rely on anticipated changes in future short-term rates. Swanson (2017) provides evidence pointing to large effects on the yield curve, stock prices and exchange rates of both forward guidance and LSAPs during the 2009-2015 ZLB period. Those effects are shown to be comparable in magnitude to the effects of conventional policies in the pre-ZLB period.

Our paper is closely related in spirit to the recent work of Wu and Xia (2016) and Wu and Zhang (2017). Wu and Xia (2016) propose a shadow policy rate indicator as a measure of the monetary policy stance that can also apply to ZLB periods. They find that the shadow rate has an impact on the economy during the ZLB period similar to the Federal Funds rate in the pre-ZLB period, and that the efforts by the Fed to stimulate the economy appear to have succeeded in maintaining the same level of economic activity as if that Fed had followed a Taylor rule without the ZLB. Wu and Zhang (2017) study a New-Keynesian model with a shadow rate which captures unconventional monetary policy at the ZLB, and conclude that a binding ZLB constraint does not alter the responses of aggregate variables to supply and demand shocks relative to periods with non-binding ZLB. Finally, Gust et. al. (2017) perform a counterfactual exercise using estimates from a non-linear DSGE model, and conclude that the zero lower bound was a significant constraint on monetary policy that exacerbated the recession and slowed down the recovery. Our simulations of a New Keynesian model in Section 3.1 are consistent with their findings, even though we use a much simpler model. Of course, our interest in this paper lies not so much in the predictions of a theoretical model regarding the importance of the ZLB constraint, but in the empirical evidence as to its effective relevance, given the potential effectiveness of unconventional policies.

6 Concluding Comments

The present paper contributes to a growing literature that aims to evaluate empirically the effectiveness of unconventional monetary policies during the recent ZLB episode. Our approach had made use of a structural vector autoregressive model with time-varying coefficients (TVC-SVAR) to describe the dynamic responses of a number of U.S. macro variables to different shocks (both supply and demand), as well as the changes over time in those responses. We find that those responses did not experience any major change during the ZLB period. That evidence is at odds with the predictions of a standard New Keynesian model where the central bank follows a nonlinear Taylor rule (subject to the ZLB) and in which no unconventional policies are implemented when the ZLB is binding (or in which those policies are neutral).

We interpret that finding as being consistent with (though not a proof of) the hypothesis that the unconventional monetary policies implemented during the ZLB years may have succeeded, at least to some extent, at getting around the constraints imposed by the ZLB on conventional monetary policy.

References

Clarida, Richard, Jordi Galí, and Mark Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, vol. 105, issue 1, 147-180.

D'Amico, Stefania and Thomas B. King (2013): "Flow and Stock Effects of Large Scale Treasury Purchases: Evidence on the Importance of Local Supply," *Journal of Financial Economics* 108(2), 425-448.

D'Amico, Stefania and Thomas B. King (2017): "What does anticipated Monetary Policy Do?" unpublished manuscript.

Del Negro, Marco and Giorgio Primiceri (2013): "Time-Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum" Federal Reserve Bank of New York Staff Report No. 619.

Eggertsson, Gauti, and Michael Woodford (2003): "The Zero Bound on Interest Rates and Optimal Monetary Policy," *Brookings Papers on Economic Activity*, vol. 1, 139-211.

Fernández-Villaverde, Jesús, Grey Gordon, Pablo Guerrón-Quintana, and Juan F. Rubio-Ramírez (2015): "Nonlinear adventures at the zero lower bound," *Journal of Economic Dynamics and Control*, 57(C), 182-204.

Galí, Jordi (1999): "Technology, Employment and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?," *American Economic Review* 89, 249-271.

Galí, Jordi, J. David López-Salido, and Javier Vallés (2003): "Technology Shocks and Monetary Policy: Assessing the Fed's Performance," *Journal of Monetary Economics*, vol. 50, no. 4., 723-743.

Galí, Jordi and Luca Gambetti (2009): "On the Sources of the Great Moderation," *American Economic Journal: Macroeconomics,* 1 (1), 26-57.

Gust, Christopher, Edward Herbst, David López-Salido and Matthew E. Smith

(2017): "The Empirical Implications of the Interest-Rate Lower Bound," *American Economic Review,* 107(7), 1971-2006.

Hamilton, James and Jing Cynthia Wu (2012): "The Effectiveness of alternative Monetary Policy Tools in Zero Lower Bound Environments," *Journal of Money, Credit and Banking* 44 (s1), 3-46.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen (2011): "The Effects of Quantitative Easing on Interest Rates," *Brookings Papers on Economic Activity* 43, 215-287.

Primiceri, Giorgio E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, pp. 821-852.

Rubio-Ramirez, J., D. Waggoner and T. Zha (2010):"Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *The Review of Economic Studies*, 77: 665-696.

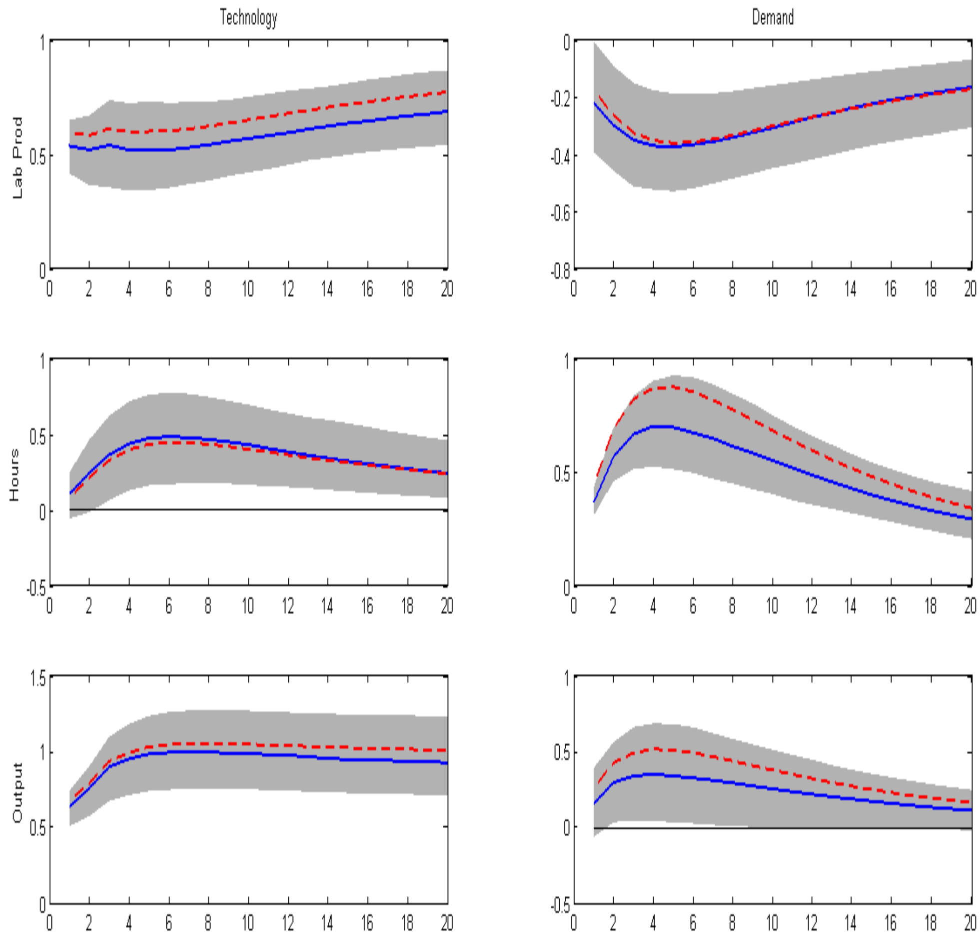
Swanson, Eric T., and John C. Williams (2014): “Measuring the Effect of the Zero Lower Bound on Medium- and Longer-Term Interest Rates,” *American Economic Review*, 104 (10), 3154–3185.

Wu, Jing Cynthia and Fan Dora Xia (2016): *Journal of Money Credit and Banking* 48(2-3), 253-291.

Wu, Jing Cynthia and Zhang (2017): “A Shadow Rate New Keynesian Model,” NBER working paper 22856.

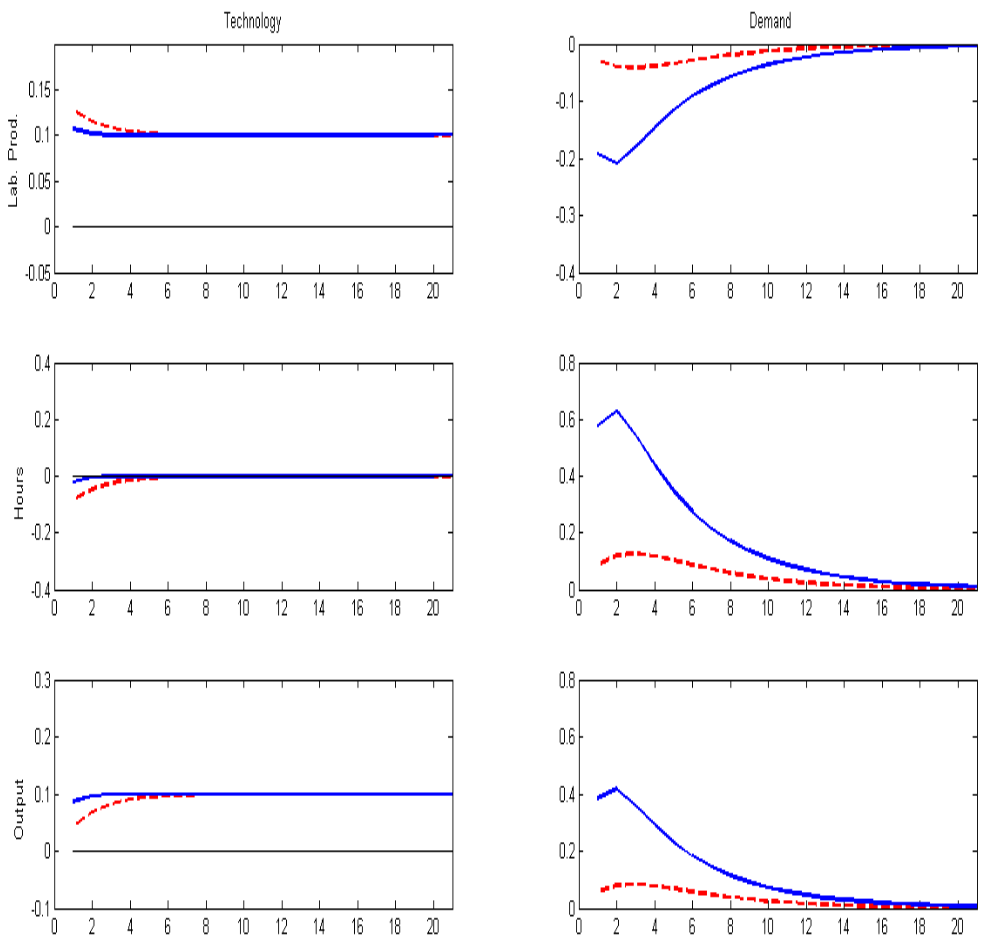
Figures

Figure 1: Impulse Response Functions of the Bivariate VAR: pre-ZLB vs. ZLB period



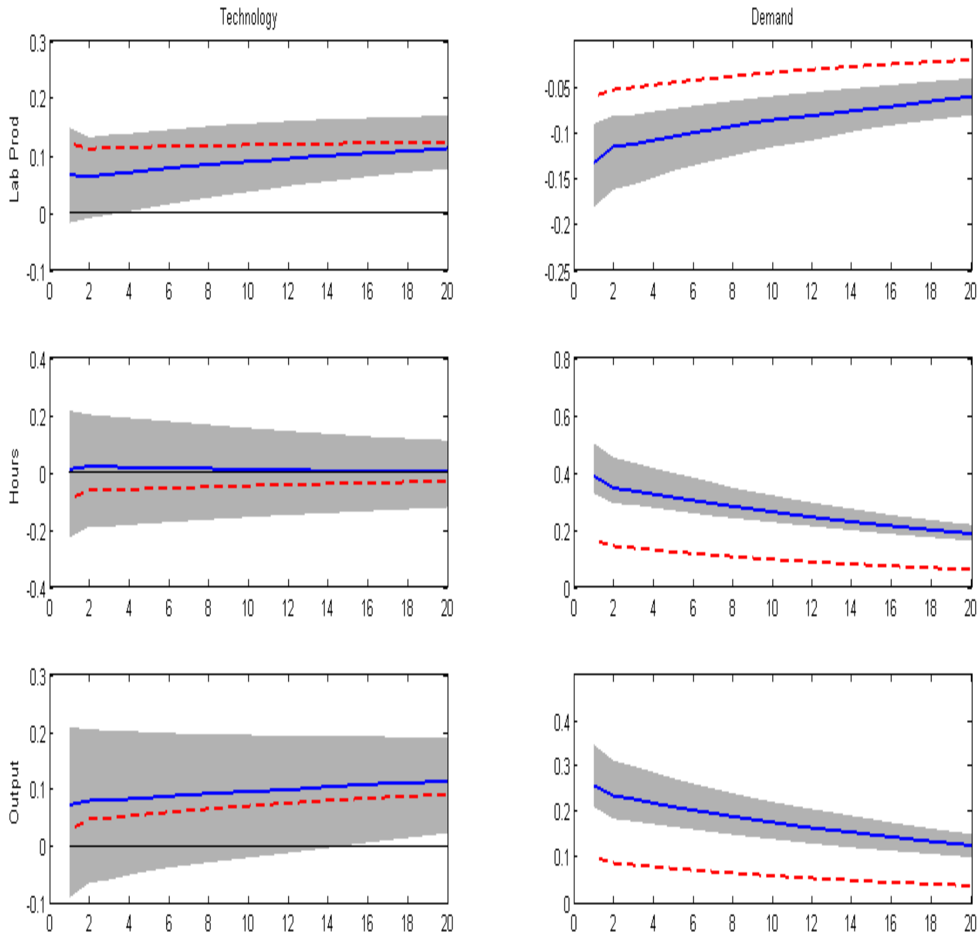
Note: The figure shows the impulse responses to technology (left column) and demand (right column) shocks. The blue line is the mean response and the grey area represents the 68% confidence bands of the average impulse response functions estimated over the zero lower bound period, 2009:Q1-2015:Q4. The dashed red line is the mean response of the average impulse response functions over the period 2002:Q1-2008:Q4.

Figure 2: Impulse Response Functions of the Theoretical Model: pre-ZLB vs. ZLB period



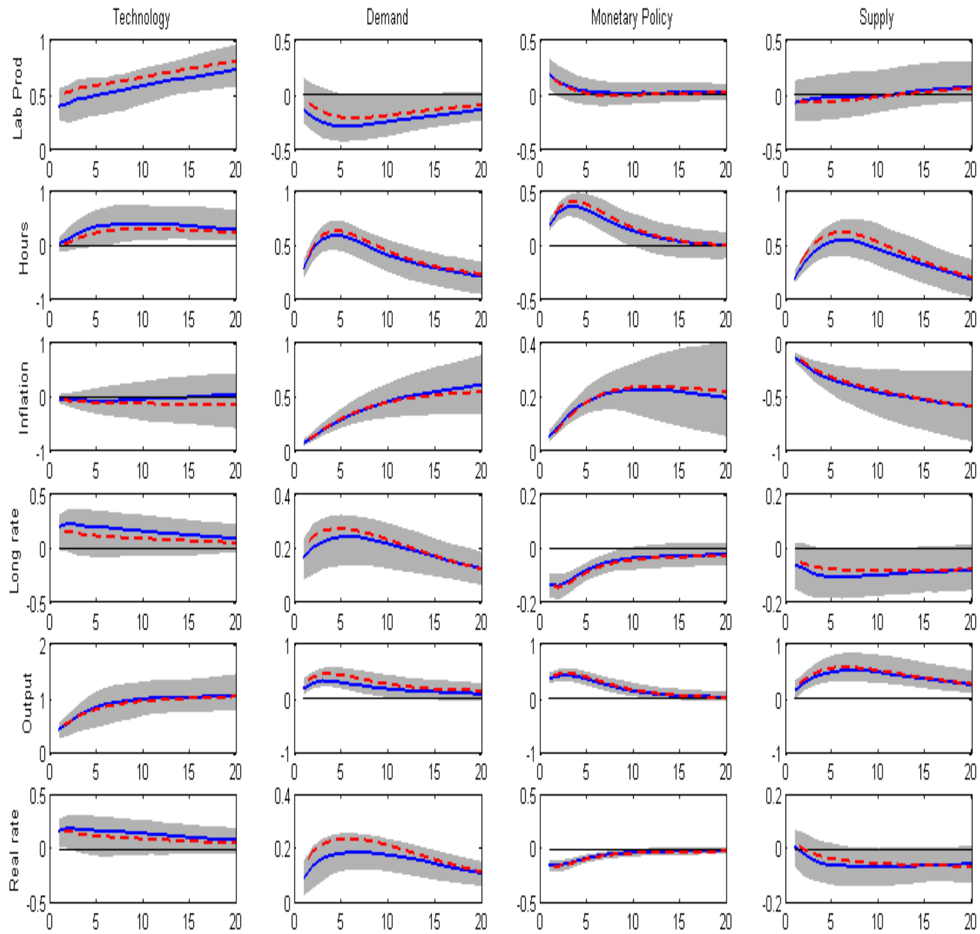
Note: The figure shows the impulse responses to technology (left column) and demand (right column) shocks. The blue line is the response under a binding zero lower bound, the red dashed line is the response under a non-binding zero lower bound.

Figure 3: Impulse Response Functions of the Bivariate VAR using Artificial Data: pre-ZLB vs. ZLB period



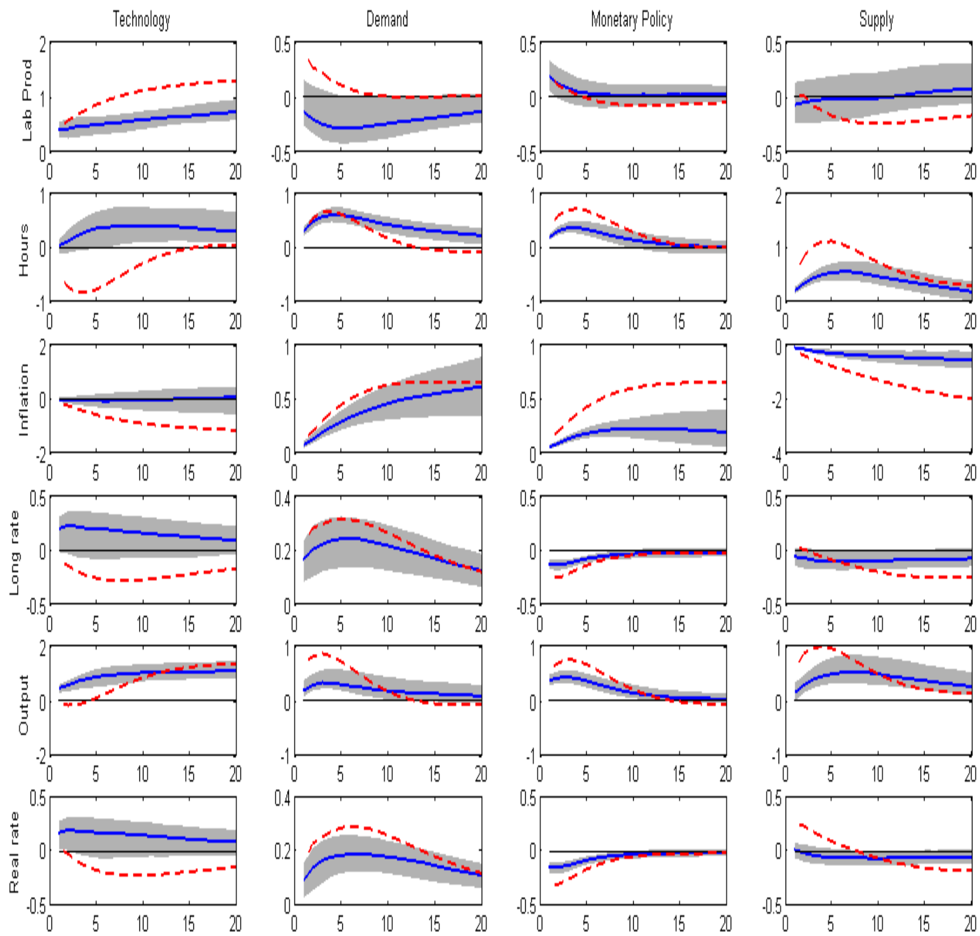
Note: The figure shows the impulse responses to technology (left column) and demand (right column) shocks. The blue line is the mean response and the grey area represents the 68% confidence bands of the average impulse response functions estimated over the zero lower bound period, 2009:Q1-2015:Q4. The dashed red line is the mean response of the average impulse response functions over the period 2002:Q1-2008:Q4.

Figure 4: Impulse Responses of the 4-variable VAR: pre-ZLB vs. ZLB period



Note: The figure shows the impulse responses to four shocks: technology (column 1), demand (column 2), monetary policy (column 3) and supply (column 4). The blue line is the mean response and the grey area represents the 68% confidence bands of the average impulse response functions estimated over the zero lower bound period, 2009:Q1-2015:Q4. The dashed red line is the mean response of the average impulse response functions over the period 2002:Q1-2008:Q4.

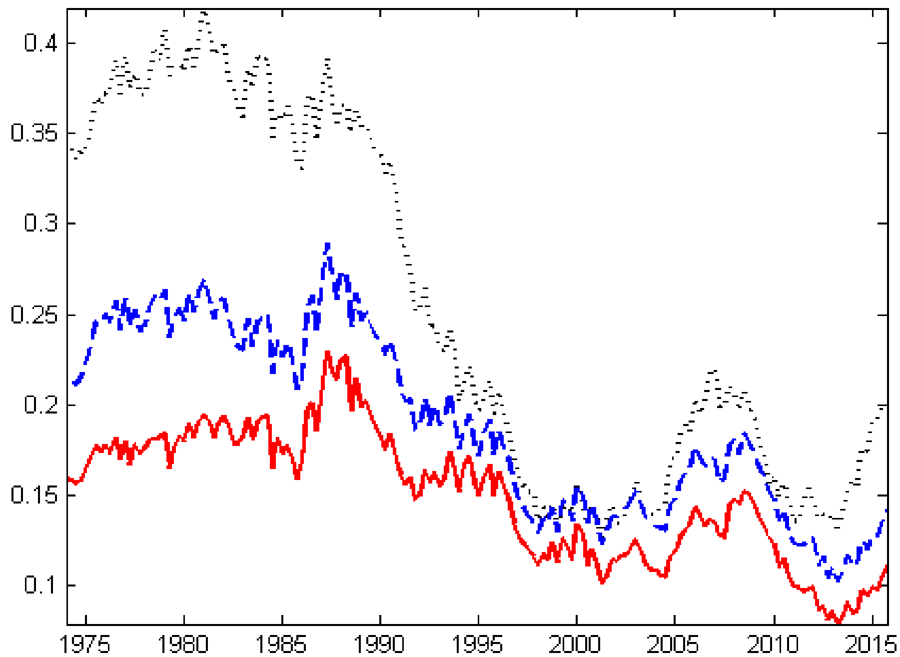
Figure 5: Impulse Responses of the 4-variable VAR: pre-Volcker vs. ZLB period



Note: The figure shows the impulse responses to four shocks: technology (column 1), demand (column 2), monetary policy (column 3) and supply (column 4). The blue line is the mean response and the grey area represents the 68% confidence bands of the average impulse response functions estimated over the zero lower bound period, 2009:Q1-2015:Q4. The dashed red line is the mean response of the average impulse response functions over the pre-Volcker period 1974:Q1-1979:Q2.

Online Appendix

Figure A-1: Contribution of Demand Shocks to the Volatility of GDP growth



Note: The figure shows the proportion of the variance of GDP growth explained by the demand shock in the 4-variable VAR at various horizons: on impact (dotted black line), after 8 quarters (dashed blue line), and after 16 quarters (solid red line).