

Real Time Changes in Monetary Policy

Marcelle Chauvet and Heather L. R. Tierney *

April 2009

Abstract

This paper investigates potential changes in monetary policy over the last decades using a nonparametric vector autoregression model. In the proposed model, the conditional mean and variance are time-dependent and estimated using a nonparametric local linear method, which allows for different forms of nonlinearity, conditional heteroskedasticity, and non-normality. Our results suggest that there have been gradual and abrupt changes in the variances of shocks, in the monetary transmission mechanism, and in the Fed's reaction function. The response of output was strongest during Volcker's disinflationary period and has since been slowly decreasing over time. There have been some abrupt changes in the response of inflation, especially in the early 1980s, but we can not conclude that it is weaker now than in previous periods. Finally, we find significant evidence that policy was passive during some parts of Burn's period, and active during Volcker's disinflationary period and Greenspan's period. However, we find that the uncovered behavior of the parameters is more complex than general conclusions suggest, since they display considerable nonlinearities over time. A particular appeal of the recursive estimation of the proposed VAR-ARCH is the detection of discrete local deviations as well as more gradual ones, without smoothing the timing or magnitude of the changes.

KEY WORDS: Monetary Policy, Taylor Rule, Local Estimation, Nonlinearity, Nonparametric, Structural Vector Autoregression, Autoregressive Conditional Heteroskedasticity.

JEL Classification Code: E40, E52, E58

* Contact author: Marcelle Chauvet, Department of Economics, University of California, Riverside, CA 92521-0247; email: chauvet@ucr.edu; phone: (951) 827-1587; fax: (951) 827-5685; Heather L.R. Tierney, School of Business and Economics, College of Charleston; 5 Liberty Street, Charleston, SC 29424, email: tierneyh@cofc.edu; phone: (843) 953-7070; fax: (843) 953-5697.

1. Introduction

A large body of research has focused on the examination of the increased stability in output and inflation since the mid 1980s. Some authors attribute it to changes in the real side of the economy, such as technological and financial innovations that may affect firms and consumers' behavior and make the economy less susceptible to shocks. Others consider that the culprit is a reduction in the size of shocks rather than in its propagation through the economy. In particular, the increased stability might be associated with changes in the impact of monetary shocks on the economy or with changes in the monetary transmission mechanism. However, there is no consensus on whether there have been changes and on their nature.

The investigation of these potential changes is important as it has implications on the temporary or permanent nature of this stability as the 2007-2009 financial crisis reminded us. If the main source is a reduction in the occurrence or size of shocks, economic stability will subside when confronted with larger ones, whereas it will be more lasting if associated with changes in the propagation of shocks. There has been recent evidence indicating the role of changes in monetary policy on real economic activity. Several papers find that the impact of monetary policy on inflation and output has reduced in the last decades.¹ For example, Boivin and Gianonni (2002, 2006) using a structural vector autoregression and a general equilibrium model, respectively, show a smaller impact of monetary policy shocks on the economy since the early 1980s. This finding has led to the investigation of whether monetary policy has become less powerful, perhaps because changes in financial innovations or other structural changes may have enabled the private sector to insulate themselves from the impact of fluctuations in interest rate. On the other hand, the evidence might reflect the opposite, that is, a more transparent and effective monetary policy conduct may have counteracted the impact of shocks to inflation and output (see e.g. Clarida, Galí, and Gertler 2000 and Boivin and Giannoni 2002).

This paper investigates potential changes in monetary policy over the last five decades using a nonparametric vector autoregression model. We use a recursive framework in which all model parameters are time dependent. In particular, the conditional mean and variance are nonparametrically estimated using a local linear estimator, which allows for any sort of nonlinearity in the relationship between variables as well as conditional heteroskedasticity or other sources of potential non-normality. The recursive estimation enables detection of some features that have been recently emphasized in the literature such as potential nonlinearities and nonstationarities in the dynamics of the VAR system. The framework is used to investigate the nature of changes in monetary policy shocks, in the Fed's reaction

¹ This is the overall conclusion of the Conference the special issue of the Federal Reserve Bank of New York's Economic Policy Review associated with the Conference Financial Innovation and Monetary Transmission, as compiled in Kuttner and Mosser (2002).

function, and in the monetary transmission mechanism to the economy. In particular, we can examine important questions such as whether there have been changes in the effectiveness of monetary policy or a reduction in the impact of monetary policy shocks on the economy.

Several recent papers have investigated potential changes in the monetary policy conduct over time. An influential recent paper is Clarida, Gali and Gertler (2000), which studies the Fed's reaction function in the last decades. They provide evidence that Taylor's rule was not met in the pre-Volcker period – that is, an increase in inflation was associated with a smaller increase in interest rate. However, they find evidence of abrupt changes related to Volcker's monetary policy in 1979, and from this date and on the Fed's reaction function has been consistent with Taylor's rule. Estimating dynamic general equilibrium models, Lubik and Schorfheide (2004) and Boivin and Gianonni (2006) also find that monetary policy has been more successful at stabilizing the economy and at ruling out undesired non-fundamental fluctuations since the early 1980s.

This finding has been contested regarding the nature of changes – whether they were abrupt or more gradual. In addition, some authors find that taking into account heteroskedasticity or the use of revised versus real time data might change the conclusions. Cogley and Sargent (2001) using a reduced VAR with drifting parameters find that changes in monetary policy have been more gradual. Sims (1999, 2001), and Sims and Zha (2006a) study changes in monetary policy via models with discrete breaks that capture potential switching policy and provide evidence that Cogley and Sargent's (2001) conclusions might be related to changes in the variance of shocks pre and post Volcker. Using a linear univariate framework, Stock's (2001) discussion of Cogley and Sargent's (2001) also points in this direction. However, Cogley and Sargent (2005), extending their previous work to include heteroskedasticity in the VAR shocks, find that there have been significant changes in the policy parameters separated from changes in variance. Orphanides (2001) raises problems in estimating the Fed's reaction function with revised data since this conceals how policymakers might have historically reacted to the information available to them at each point in time. Orphanides (2002, 2004) re-estimates Clarida, Gali and Gertler's (2000) model using real time data and finds that the Fed's reaction function is not very different pre and post Volcker, in contrast with Clarida et al. (2000).

Boivin (2006) proposes a model that combines some of the main issues discussed in the literature. In particular, a forward looking Taylor's rule with time varying parameters that takes into account heteroskedasticity in the policy shock is estimated using real time data. Boivin (2006) finds evidence of important changes in monetary conduct over time, although these changes are modeled as being gradual. Primiceri (2005) and Cogley and Sargent (2005) propose a time-varying parameter VAR that assumes gradual and continuous drifts in the estimated coefficients and second moments using revised data. The

findings are also that there have been systematic and non-systematic changes in monetary policy in the last decades.

The goal of this paper is to contribute to this debate by considering a flexible framework that can account for multifaceted dynamics in the evolution of monetary policy. Consistent with evidence from a large recent literature, a monetary model that ignores changes in parameters and heteroskedasticity may be misleading. In fact, models that allow for time varying parameters (TVP) with heteroskedasticity have become a popular tool to investigate changes in monetary policy. However, these models favor continuous drifting coefficients and variance over discrete breaks. If there is a discrete change the TVP model would be mis-specified, although it can still yield an insightful approximation (see Boivin 1999 for a discussion on this).² Koop, Leon-Gonzalez, and Strachan (2007) extend this approach to a more sophisticated time-varying VAR mixture innovation model that allows estimation of gradual or abrupt changes in parameters.

In the proposed nonparametric vector autoregressive conditional heteroskedasticity (VAR-ARCH) model, the conditional mean and variances are estimated by fitting a local linear regression function in the neighborhood of target values, with lower weights for observations farther away from them. Thus, the estimation takes into account information in the tail regions with potentially smaller bias near the boundary of support. In addition, potential nonlinearities, nonstationarities, and asymmetric behavior can be examined without the need for specifying a functional form or parametric densities for the conditional moments, which may provide useful insight for further parametric fitting. Finally, since the model is observation-driven and recursively estimated over real time samples, it leaves up to the data to determine the nature of variations and, as a result, the time series of the estimated coefficients and variances may change smoothly or sharply over time.

Structural VARs enable us to separate out systematic responses to changes in interest rates from exogenous monetary policy shocks. The errors from the interest rate equation are generally interpreted as monetary policy shocks and taken as a measure of policy changes. However, when VARs are estimated using ex-post revised data, the shocks could also reflect measurement error due to data revisions. As discussed in Bernanke and Mihov (1998) and Orphanides (2001), the presence of data revisions in the errors may bias the response of some variables in addition to compromising their interpretation as monetary policy shocks. The use of unrevised real time data, on the other hand, captures how policymakers have historically reacted to the information available to them at each point in time for

² In the context of stability in monetary VAR models, some find evidence supporting it such as Bernanke and Mihov (1998) and Christiano, Eichenbaum, and Evans (1999), while others find that the parameters of monetary VAR are unstable as in Bernanke, Gertler, and Watson (1997), Boivin (1999), and Boivin and Gianonni (2002), among several others. Boivin (1999) argue that the differences are due mainly to the small sample properties of the stability tests, and to the effect of the number of parameters on the power of these tests. He concludes that there is compelling evidence of abrupt changes in monetary VARs.

implementing policy, as they reflect uncertainties regarding the state of the economy and inflation, and the residuals represent monetary policy shocks as taken place in real time. Given the difference in results obtained by Orphanides (2002, 2004) using real time data and Clarida, Gali and Gertler's (2000) using revised data, this paper investigates changes in monetary policy in real time using unrevised data.

Our results suggest that there have been abrupt as well as gradual changes in the variances of the shocks, in the monetary transmission mechanism, and in the Fed's reaction function. First, the variance of the interest rate innovation displays a hump shape, with a sharp rise in the late 1970s and a gradual decrease since the mid 1980s. Second, our evidence confirms previous discussions regarding a recent reduced impact of monetary policy shocks on output, unemployment, and inflation compared to historical values. However, we find that the responses are not only time-varying, but also highly nonlinear with more oscillations until the mid 1980s and more stability in the second part. The negative response of output was strongest during Volcker's disinflationary period and has since been slowly decreasing over time. Although point estimates indicate that the impact of monetary shocks on inflation has been decreasing, we can not conclude that it is smaller now than in the early 1980s when taking into account the uncertainty surrounding these estimates. However, the strongest negative response of inflation took place in the mid 1970s. In addition, there have been some abrupt changes in the response of inflation, especially in the early 1980s. The large swings in these relationships unveil the fragility of split-sample comparisons – as it is commonly done in the literature.

Finally, we consider the evolution of the real time response of interest rate to inflation by recursively estimating a forward-looking Taylor rule as proposed in Clarida et al (2000), from which we derive a time-varying degree of activism. The movements broadly agree with the results of several authors, including Judd and Rudebusch (1998), Clarida, Gali, and Gertler (2000), Boivin (2006), and Cogley and Sargent (2001, 2005). In particular, taking into account uncertainty surrounding estimates, there is significant evidence that policy was passive from the mid to the late 1970s, but active during Volcker's disinflationary period and during the Greenspan period.

However, we find that the uncovered behavior of the parameters is more complex than general conclusions suggest, since they display considerable nonlinearities over time. Overall, we find that there were gradual and abrupt changes in the degree of activism and in the response of the economy to monetary policy shocks. A particular appeal of the recursive estimation is the detection of discrete local deviations as well as more gradual ones, without smoothing the timing or magnitude of the changes. Thus, the estimates can shed light on the nature and timing of these changes.

The structure of this paper is as follows: Section 2 presents the monetary structural VAR model. Section 3 discusses the estimation of the local linear nonparametric reduced-form and the structural VAR-

ARCH model. Section 4 interprets and compares the proposed VAR-ARCH model with existing literature. The empirical results are presented in Section 5, and Section 6 concludes.

2. Monetary Vector Autoregression Model³

Following Bernanke and Blinder (1992), assume that the “true” model of the economy is represented by the p^{th} -order structural vector autoregression:

$$\mathbf{Y}_t = \mathbf{b}^y + \sum_{i=0}^p \mathbf{E}_i \mathbf{Y}_{t-i} + \sum_{i=0}^p \mathbf{C}_i R_{t-i} + \mathbf{F}^y \mathbf{v}_t^y \quad (1)$$

$$R_t = b^R + \sum_{i=0}^p \mathbf{D}_i \mathbf{Y}_{t-i} + \sum_{i=0}^p G_i R_{t-i} + v_t^R \quad (2)$$

The system (1) and (2) allows both contemporaneous and p lagged values of each variable in any equation. Let \mathbf{Y}_t be an $[(n-1) \times 1]$ vector of nonpolicy macroeconomic variables, and R_t be a scalar representing the policy instrument. Equation (2) describes the response of the policy instrument to changes in the system while Equation (1) represents the response of the nonpolicy variables.⁴ The vector of structural disturbances \mathbf{v}_t^y is orthogonal and also mutually uncorrelated with the scalar exogenous policy shock v_t^R . The latter assumption implies that v_t^R is independent from contemporaneous macroeconomic conditions, which is part of the definition of an exogenous policy shock.

We consider two baseline models. The first is a more parsimonious VAR model with only inflation and growth rate of output in the non-policy vector \mathbf{Y}_t . The second is composed of two additional variables in the non-policy block: unemployment and commodity inflation. The inclusion of commodity inflation is commonly done in the literature to avoid the price puzzle – that monetary tightening tends to lead to an increase rather than a fall in the price level (see e.g. Bernanke and Blinder 1992, Sims 1992, Balke and Emery 1994, Sims and Zha 2006b, Christiano, Eichenbaum, and Evans 1996, Hanson 2004, Giordani 2004, Castelnuovo and Surico 2006, etc.).

The Federal funds rate is used as an indicator of policy stance, R_t . Although there have been changes in the Fed’s operating procedure in the last four decades, the Federal funds rate has been the main policy instrument in the U.S. over most of this time, as maintained by several authors including Bernanke and Blinder (1992), Bernanke and Mihov (1998), and a large literature that followed.

³ As it is usual in the literature, bold letters denote vector and matrices.

⁴ We also estimate a forward looking version of the Taylor rule in Section 5.3.

The system (1)-(2) can be written in a more compact form. Let $\mathbf{Z}_t = [\mathbf{Y}_t \ R_t]'$ be a $nx1$ vector containing the values that the n variables \mathbf{Y}_t and R_t take at date t , $\mathbf{B}_i = \begin{bmatrix} \mathbf{E}_i & \mathbf{C}_i \\ \mathbf{D}_i & \mathbf{G}_i \end{bmatrix}$ be the $n \times n$ coefficient matrices with $i = 0, \dots, p$, $\mathbf{b} = [\mathbf{b}^y \ b^R]'$ be the $nx1$ vector of constants, and $\mathbf{v}_t = [\mathbf{v}_t^y \ \mathbf{v}_t^R]'$ be the $nx1$ vector of disturbances. The structural model can be written as:

$$\mathbf{B}_0 \mathbf{Z}_t = \mathbf{b} + \mathbf{B}_1 \mathbf{Z}_{t-1} + \dots + \mathbf{B}_p \mathbf{Z}_{t-p} + \mathbf{F} \mathbf{v}_t \quad \mathbf{v}_t \sim (0, \mathbf{\Omega}), \quad (3)$$

and the reduced form VAR is:

$$\mathbf{Z}_t = \mathbf{a} + \mathbf{A}_1 \mathbf{Z}_{t-1} + \dots + \mathbf{A}_p \mathbf{Z}_{t-p} + \mathbf{u}_t \quad \mathbf{u}_t \sim (0, \mathbf{\Theta}), \quad (4)$$

The parameters of the reduced-form VAR are a linear combination of structural parameters, which includes contemporaneous relations among the endogenous variables. The reduced form coefficients satisfy the following conditions: the vector of constants $\mathbf{a} = \mathbf{B}_0^{-1} \mathbf{b}$, and the matrices $\mathbf{A}_i = \mathbf{B}_0^{-1} \mathbf{B}_i$ for $i = 1, \dots, p$. The vector of innovations \mathbf{u}_t is modeled as an interdependent system of equations such that:

$$\mathbf{B}_0 \mathbf{u}_t = \mathbf{F} \mathbf{v}_t \quad (5)$$

$\begin{matrix} nxn & nx1 & nxn & nx1 \end{matrix}$

and $\mathbf{\Theta} = \mathbf{B}_0^{-1} \mathbf{\Gamma} \mathbf{B}_0^{-1'}$, for $\mathbf{\Gamma} = \mathbf{F} \mathbf{\Omega} \mathbf{F}'$. In order to identify the parameters, restrictions on the matrices $\mathbf{B}_i, \mathbf{F}, \mathbf{A}_i, \mathbf{\Omega}$ are necessary. In general, \mathbf{B}_0 is normalized to have ones on the leading diagonal. In addition, the unobservable structural shocks are assumed to be mutually uncorrelated, which is obtained by setting $\mathbf{\Omega} = \mathbf{I}_n$ or as a diagonal matrix. An alternative identification found in the literature is to set $\mathbf{F} = \mathbf{I}_n$, which implies that the vector \mathbf{u}_t is modeled as the interdependent system of linear equations $\mathbf{B}_0 \mathbf{u}_t = \mathbf{v}_t$.

Although orthogonality of the structural innovations isolates the dynamic impact of a shock, it is not sufficient to achieve identification. Under the conventional assumption that all structural disturbances are conditionally homoskedastic further assumptions need to be imposed. As an illustration, let \mathbf{P} be an orthogonal transformation matrix ($\mathbf{P} \mathbf{P}' = \mathbf{P}' \mathbf{P} = \mathbf{I}_n$). Then:

$$\mathbf{B}_0^* \mathbf{u}_t = \mathbf{F}^* \mathbf{v}_t \quad (6)$$

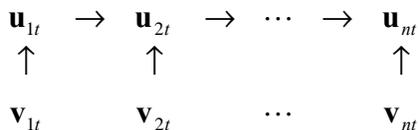
is observationally equivalent to (5) with $\mathbf{B}_0^* = \mathbf{P} \mathbf{B}_0, \mathbf{F}^* = \mathbf{P} \mathbf{F}$ with orthogonally rotated structural disturbances for any admissible transformation matrices. A necessary condition for identification is that at least $\frac{n(n-1)}{2}$ additional restrictions are imposed on the system.

We use two approaches to impose identifying restrictions on the model. First, following a large literature including Bernanke and Blinder (1992), Rotemberg and Woodford (1997), Bernanke and Mihov (1998), Boivin and Gianonni (2002), among several others, a minimum set of restriction that allows identification of the impact of policy shocks is to assume that they affect the macroeconomic variables with one lag, $C_o = 0$. That is, non-policy variables depend on contemporaneous and lagged values of \mathbf{Y}_t , but only on lagged values of R_t . This assumption implies a delayed response of the economy to policy shocks due to lags in the transmission mechanism. The policy variable, on the other hand, is assumed to respond to all contemporaneous and lagged variables in the system. That is, policymakers are assumed to have contemporaneous information about the nonpolicy variables. This is a plausible assumption with the use of unrevised real time data, which is the first information available to the Fed to implement policy on a current basis. Consistently with a broad literature we enter commodity inflation as it commonly appears in other VAR studies, as responding to contemporaneous information in the non-policy block given that it is set in forward-looking asset markets (see e.g. Eichenbaum 1992, Christiano and Eichenbaum 1992, and Bernanke and Mihov 1998, Hanson 2004, etc.). We find that this ordering is the one that results in a smaller price puzzle. However, changing this ordering does not significantly affect the other responses. The reader is referred to Section 5.4 for a discussion of these choices and of the robustness of the results to alternative specifications.

The remainder identification conditions for the non-policy block are formulated based on Swanson and Granger's (1997) method, which uses the notion of 'instantaneous causality' (see e.g. Lütkepohl 1990) to identify the causal structure for the innovations. Assume that the innovations of the VAR can be arranged as follows:

$$\mathbf{u}_{1t} = \mathbf{v}_{1t}; \mathbf{u}_{2t} = \gamma_2 \mathbf{u}_{1t} + \mathbf{v}_{2t}; \dots; \mathbf{u}_{nt} = \gamma_n \mathbf{u}_{n-1,t} + \mathbf{v}_{nt}.$$

This structure can be represented by a causal graph as:⁵



This graph means, for example, that \mathbf{u}_{2t} may be expressed as a function of \mathbf{u}_{1t} , \mathbf{u}_{3t} as a function of \mathbf{u}_{2t} , etc. Swanson and Granger (1997) show that this causal ordering implies that $E(\mathbf{u}_{ht} \mathbf{u}_{kt} | \mathbf{u}_{lt}) = 0$ for $h < l < k$, with $h < n$, and suggest testing the partial correlation between \mathbf{u}_{ht} and \mathbf{u}_{kt} conditional on \mathbf{u}_{lt} in order to recover the causal ordering empirically. For example, given $\mathbf{u}_{3t} = \gamma_3 \mathbf{u}_{2t} + \mathbf{v}_{3t}$ and

⁵ Note that the causal graph with the arrows pointed forward implies the same restriction on the conditional expectation as the arrows pointed backward, and, therefore, the direction of the graph is not identified.

$E(\mathbf{u}_{2t}\mathbf{v}_{3t})=0$, the condition $E(\mathbf{u}_{1t}\mathbf{u}_{3t} | \mathbf{u}_{2t})=0$ implies that given \mathbf{u}_{2t} the variable \mathbf{u}_{1t} does not help to predict \mathbf{u}_{3t} . Conversely, if $\gamma_3 \neq 0$, $E(\mathbf{u}_{2t}\mathbf{u}_{3t} | \mathbf{u}_{1t}) \neq 0$ and the variable \mathbf{u}_{2t} helps to predict (‘cause’) \mathbf{u}_{3t} .

We apply this method over recursive real time samples. Although the results change occasionally, for most part of the sample considered the ordering suggested is that inflation does not respond contemporaneously to unemployment and output, unemployment does not respond to inflation contemporaneously, but it helps predict output within the same period.⁶

Finally, we fix the sign of the factorization matrix $\mathbf{B}_0^{-1}\mathbf{F}$ so that the diagonal elements are all positive (see Christiano, Eichenbaum, and Evans 1999 for a discussion). This normalization insures that the policy shocks are identified as well as the responses of non-policy and policy variables to policy shocks. Summing up, the conditions for identification of the VAR with five variables implies that

$$\mathbf{B}_0\mathbf{u}_t = \mathbf{F}\mathbf{v}_t$$

$$\begin{pmatrix} 1 & -b_{12} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -b_{31} & -b_{32} & 1 & 0 & 0 \\ 0 & -b_{42} & -b_{43} & 1 & 0 \\ -b_{51} & -b_{52} & -b_{53} & -b_{54} & 1 \end{pmatrix} \begin{pmatrix} u_t^U \\ u_t^\pi \\ u_t^O \\ u_t^{PC} \\ u_t^R \end{pmatrix} = \begin{pmatrix} f_{11} & 0 & 0 & 0 & 0 \\ 0 & f_{22} & 0 & 0 & 0 \\ 0 & 0 & f_{33} & 0 & 0 \\ 0 & 0 & 0 & f_{44} & 0 \\ 0 & 0 & 0 & 0 & f_{55} \end{pmatrix} \begin{pmatrix} v_t^U \\ v_t^\pi \\ v_t^O \\ v_t^{PC} \\ v_t^R \end{pmatrix},$$

that is,

$$\begin{aligned} u_t^U &= b_{12}u_t^\pi + f_{11}v_t^U \\ u_t^\pi &= f_{22}v_t^\pi \\ u_t^O &= b_{31}u_t^U + b_{32}u_t^\pi + f_{33}v_t^O \\ u_t^{PC} &= b_{42}u_t^\pi + b_{43}u_t^O + f_{44}v_t^{PC} \\ u_t^R &= b_{51}u_t^U + b_{52}u_t^\pi + b_{53}u_t^O + b_{54}u_t^{PC} + f_{55}v_t^R \end{aligned}$$

The structural VAR is over-identified with one degree of freedom. The structural VAR with three variables follow the same identifying restrictions as above for the policy shock and non-policy variables output and inflation.

⁶ We allow for the possibility that inflation helps predict output contemporaneously given that the coefficient in the test was significant in about half of the recursive samples considered. That is, output may be an *unshielded collider* on the path of inflation, unemployment, and output (see Demiralp, Hoover, and Perez 2008). We also find that commodity prices do not respond contemporaneously to the unemployment rate for most of the recursive samples. We have considered the cases in which one or both of these coefficients are set to zero, as discussed in Section 5.4. The results are not sensitive to these restrictions.

3. Estimation

In a first step, we estimate a reduced form VAR using the nonparametric local linear method equation by equation as described in section 3.1.⁷ The coefficients estimated obtained with this method are then used as initial values in the structural VAR, which is estimated using the GMM method described in Section 3.2.

3.1 Reduced-Form VAR

Parametric Estimation

Let \mathbf{x}_t be an $[(np+1) \times 1]$ vector containing a constant and the p lags of each of the elements of \mathbf{Z}_t and \mathbf{A}' be a $[n \times (np+1)]$ matrix of coefficients:

$$\mathbf{x}_t \equiv \begin{bmatrix} 1 \\ \mathbf{Z}_{t-1} \\ \mathbf{Z}_{t-2} \\ \vdots \\ \mathbf{Z}_{t-p} \end{bmatrix} \quad \text{and} \quad \mathbf{A}' \equiv [\mathbf{a} \ \boldsymbol{\alpha}_1 \ \boldsymbol{\alpha}_2 \ \dots \ \boldsymbol{\alpha}_p].$$

The standard vector autoregressive system (4) can then be written as:

$$\mathbf{Z}_t = \mathbf{A}' \mathbf{x}_t + \mathbf{u}_t \quad (7)$$

Where \mathbf{u}_t is the vector of zero mean disturbances, which are independent of \mathbf{x}_t . The least squares estimators (OLS) of \mathbf{A} is:

$$\hat{\mathbf{A}}' = \left[\sum_{t=1}^T \mathbf{Z}_t \mathbf{x}_t' \right] \left[\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right]^{-1},$$

$[n \times (np+1)]$

From the regression of Z_{jt} on \mathbf{x}_t :

$$Z_{jt} = \boldsymbol{\alpha}_j \mathbf{x}_t + u_{jt} \quad (8)$$

We obtain the estimated coefficient vector:

$$\hat{\boldsymbol{\alpha}}_j' = \left[\sum_{t=1}^T Z_{jt} \mathbf{x}_t' \right] \left[\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \right]^{-1},$$

$[1 \times (np+1)]$

which corresponds to the j^{th} row of $\hat{\mathbf{A}}'$.

⁷ In the reduced-form only lagged values of the same endogenous variables appear on the right-hand side of each equation.

Nonparametric Estimation

Consider a more general version of Equation (7):

$$Z_{jt} = m_j(\mathbf{x}_t) + \varpi^{1/2}(\mathbf{x}_t)u_{jt} \quad (9)$$

where $m_j(\mathbf{x}_t) = E(Z_{jt} | \mathbf{x}_t)$, $u_{jt} \sim (0,1)$, and $\varpi_j(\mathbf{x}_t)$ is the conditional variance of the residuals based on information \mathbf{x}_t . Notice that since the conditioning information is composed of lagged values of Z_{jt} this model corresponds to an autoregressive conditional heteroskedastic autoregressive model (ARCH). The framework is very flexible and more general than standard ARCH models, as it admits non-normality. In addition, since it does not make assumptions on its functional form, it allows fitting of many alternative nonlinear forms.⁸

Parametric analysis assumes that the functional form linking m_j and ϖ_j to \mathbf{x}_t is known. However, if the function is mis-specified regarding the joint distribution or its form, the estimators will be biased and inconsistent. For example, the normality assumption implies symmetry, unimodality, and mesokurtosis.⁹ Given the possibility that these assumptions might not hold and the effects on the estimation, we consider nonparametric modeling, which does not depend on specific structures of the conditional expectation and variance.

We extend the nonparametric local linear least squares method (LLLS) to estimate the reduced form vector autoregressive model based on the work of Fan (1993), Fan and Gijbels (1996), Ruppert and Wand (1994), and Fan and Yao (1998), among others. The local linear and polynomial methods have been widely used in nonparametric regression during recent years due to its attractive mathematical efficiency, bias reduction, and adaptation of edge effects. It is also well documented that the local linear method is superior in theory and applications among most nonparametric smoothing methods (see Fan and Gijbels 1996).

The basic idea is that the conditional mean and variance functions in each equation of the VAR are estimated by locally averaging the values Z_{jt} corresponding to \mathbf{x}_t by fitting a local linear vector autoregression function in the neighborhood of the target \mathbf{x} , within an interval specified by the vector of variable window widths, \mathbf{h}_j , and the weighting process determined by the multivariate Gaussian kernel

⁸ Some of the related literature is Tierney (2005), which proposes a bivariate VAR to study the Fisher Effect using the nonparametric local linear method and Silverman's (1986) rule of thumb bandwidth. Bossaerts, Hardle, Haftern (1995) propose a univariate nonparametric conditional autoregressive heteroskedastic nonlinear model (CHARN), which is estimated using local linear methods. Bachmeier, Leelahanon, and Li (2007) propose a nonparametric monetary VAR, considering homoskedastic errors. The model is estimated using Nadaraya-Watson's kernel estimator (Nadaraya 1965, Watson 1964). Finally, Härdle, Tsybakov, and Yang (1998) propose a vector conditional heteroskedastic autoregressive nonlinear models (CHARN) and apply it to estimate volatility matrices in foreign exchange markets using the fixed rule-of-thumb bandwidth proposed by Silverman's (1986).

⁹ Section 5.4 shows that these restrictions do not hold for most equations over the recursive samples.

density. The shape of the empirical kernel is not consequential as long as it places more weight on the observations closest to the conditioning observation and increasingly less weight as the distance between the measuring and conditioning observations increases. On the other hand, the choice of the window width is very important in nonparametric estimation, as it determines how many of the \mathbf{x}_t around \mathbf{x} are to be used in forming the average – or the local sample size. The optimal amount of smoothing minimizes the trade-off between bias and variance (i.e. the mean squared error).

We extend to a multivariate variable setting the bandwidth for local polynomial regressions proposed by Fan and Gijbels (1995). This bandwidth selector combines the notions of the plug-in and the cross-validation method (see Rupert and Wand 1994 for a discussion). Fan and Gijbels (1995) show this to be a consistent estimator of the asymptotic optimal bandwidth. In addition, it is less variable and more efficient than the cross-validation bandwidth rule (Kalyanaramany 2009). The estimation procedure is described in the Appendix.

3.2 Structural VAR

Estimation

The structural VAR is:

$$\mathbf{B}_0(\mathbf{x}_t)\mathbf{Z}_t = \mathbf{m}^s(\mathbf{x}_t) + \mathbf{v}_t \quad (10)$$

where $\mathbf{m}^s(\mathbf{x}_t) = E(\mathbf{B}_0(\mathbf{x}_t)\mathbf{Z}_t | \mathbf{x}_t)$ and $\mathbf{v}_t \sim (0, \mathbf{\Omega}(\mathbf{x}_t))$. The structural parameters are estimated using the Generalized Method of Moments (GMM). The assumption that the structural shocks are mutually uncorrelated implies the moment conditions $E(v_{it}v_{jt}) = 0$ for $i < j$ subject to the restrictions from equation (5). The estimation procedure involves two steps. First, the reduced form parameters $\hat{\boldsymbol{\theta}}_j(\mathbf{x}_t) = \{\hat{\boldsymbol{\alpha}}_j(\mathbf{x}_t), \hat{\boldsymbol{\omega}}_j(\mathbf{x}_t)\}$ are estimated using the nonparametric method described in the previous section. The GMM structural parameters are then obtained as a solution to the minimization problem:

$$\hat{\boldsymbol{\lambda}} = \arg \min_{\{\boldsymbol{\lambda}\}} \left\{ \left[\sum_{t=p+1}^T M_{v_t}(\boldsymbol{\lambda})' \right] \mathbf{Q}(\mathbf{x}_t) \left[\sum_{t=p+1}^T M_{v_t}(\boldsymbol{\lambda}) \right] \right\} \quad (11)$$

subject to the identification restrictions imposed by Equation (5), where $\boldsymbol{\lambda}$ is the vector of structural parameters containing the elements of the matrices $\mathbf{\Omega}(\mathbf{x}_t)$ and $\mathbf{B}_0(\mathbf{x}_t)$, $M_{v_t}(\boldsymbol{\lambda}) = (v_{it}v_{jt})_{i < j}$, and $\mathbf{Q}(\mathbf{x}_t)$ is the weighting matrix:

$$\mathbf{Q}(\mathbf{x}_t) = \sum_{t=p+1}^T E \left[M_{v_t}(\boldsymbol{\lambda}) M_{v_t}(\boldsymbol{\lambda})' \right]^{-1}.$$

Allowing for time dependent conditional volatilities imposes overidentification conditions in the structural VAR (see e.g. Normandin, L. Phaneuf 2004). The conditional heteroskedasticity structure of the system is:

$$\mathbf{B}_0(\mathbf{x}_t)\mathbf{\Theta}(\mathbf{x}_t)\mathbf{B}_0(\mathbf{x}_t)' = \mathbf{\Gamma}(\mathbf{x}_t)$$

With linearly independent conditional variances of the structural disturbances, the system (5) is statistically identified since $\mathbf{B}_0(\mathbf{x}_t)^{-1}$ is unique under orthogonal transformations so that $\mathbf{P} = \mathbf{I}$ and $\mathbf{P} = \mathbf{I}^{1/2}$ are the only admissible transformations preserving the orthogonality of the rotated structural innovations in Equation (6) since $\mathbf{\Gamma}^*(\mathbf{x}_t) = \mathbf{P}\mathbf{\Gamma}(\mathbf{x}_t)\mathbf{P}'$ is diagonal. Given the restrictions imposed on the system as discussed in Section 2, the assumption of independent conditional variances is not necessary for identification. However, a simpler correlation structure facilitates interpretation of the source of uncertainties at the structural level.

4. Model Interpretation and Comparison with Literature

Recently, some authors have proposed parametric monetary VARs in which both the coefficients and the covariance of the structural shocks are time-varying, such as Cogley and Sargent (2005), Primiceri (2005), and Koop, Leon-Gonzalez and Strachan (2007). These papers consider TVP with drifting coefficients and multivariate stochastic volatility, which are estimated using Bayesian methods. Intuitively, the TVP estimator at time t applies decreasing weights to observations far away from t , while the standard OLS method applies equal weights to all observations. In these papers identification is obtained through a triangular factorization of the reduced form covariance matrix. In Cogley and Sargent (2005) and Primiceri (2005), the innovations of the random coefficients and variances are modeled as normal distributions and assumed to be mutually uncorrelated. The main difference between Cogley and Sargent (2005) and Primiceri (2005) is that the latter assumes that the simultaneous relations among variables are also time variant. In our notation, this is modeled in Primiceri (2005) as allowing the elements of the matrix \mathbf{B}_0 to follow random walks. However, they attain some difference in results as it is discussed in the next section.¹⁰

The proposed framework shares some similarities and differences with these models. First, as explained in Section 3.1, the reduced-form VAR is estimated using the nonparametric local least squares method, which fits a hyperplane within an interval as specified by the window width of the kernel density function. The observations of \mathbf{x}_t that lie far from \mathbf{x} will get low weight in the determination of the

¹⁰ As in Primiceri (2005), Koop, Leon-Gonzalez, and Strachan's (2007) model also allows the elements of the matrix \mathbf{B}_0 to be time-varying. However, Koop, Leon-Gonzalez and Strachan's TVP with mixture innovation model does not impose as many restrictions in the parameters and allows for the estimation of breaks in parameters as well as gradual changes.

conditional mean and variance for the reduced form and, the structural parameters. Intuitively, the estimation procedure implies that the estimators are obtained as local averaging values of the dependent variable that are close in terms of the values taken on by the regressors. Since the regressors are lagged values of the dependent variables, the estimators are also time-dependent. Thus, the model proposed in this paper is a multivariate nonlinear version of ARCH models in which the variance is a function of the squares of past observations.¹¹ Although our VAR-ARCH model allow for time dependent coefficients and variance, it does not assume a functional form for the conditional variance or parametric distributions for the processes followed by the coefficients and errors. These features are carried over in the estimation of the structural VAR.

In stochastic volatility models the conditional variance depends on some unobserved components rather than on past observations, and the disturbances in the system are assumed to follow parametric distributions. A disadvantage of stochastic volatility models is that it can not move away from one lag dependence, given that it is based on a random walk, whereas this is not a restriction for ARCH models.¹² In ARCH models — and in our model more specifically, the conditional mean and variance of the parameters are not assumed to be unrelated from one another. Finally, as in Primiceri (2005) and Koop, Leon-Gonzalez and Strachan (2007), our model also allows the innovation of structural shocks to have a time-varying effect on the variables in the system, given that the matrix of structural coefficients $\mathbf{B}_0(\mathbf{x}_t)$ is time dependent.

There is also a recent literature that studies changes in monetary policy via models with discrete breaks governed by Markov chains that capture potential switching policy, such as Sims (1999, 2001), and Sims and Zha (2006a). These models differ from Cogley and Sargent (2005) and Primiceri (2005), as the latter captures gradual and continuous time variation in the coefficients and variances of the shocks. In fact, if breaks are present in the data, the TVP model is mis-specified. In this case, the weighting implied by these models generates smooth estimates of the discrete change, with the parameters beginning to change before the actual breakpoint. On the other hand, Koop, Leon-Gonzalez, and Strachan's (2007) method considers gradual and/or abrupt changes in parameters.

Finally, our identification scheme of structural shocks is not obtained by using triangular factorization, but by utilizing overidentified restrictions followed by a large literature on monetary VAR (e.g Bernanke and Blinder 1992, Rotenberg and Woodford 1997, Bernanke and Mihov 1998, Boivin and Gianonni 2002, etc.), and using Swanson and Granger (1997) 'instantaneous' causality test.

¹¹ Notice that the equations in the reduced-form nonparametric VAR are estimated one-by-one.

¹² In fact, as stated in Cogley and Sargent (2005), their choice of the stochastic volatility specification as an alternative for ARCH models was made based on convenience, since it simplifies the Bayesian algorithm by allowing to impose the restriction that the innovations to the parameters and to the variances are uncorrelated. Shephard (1996) discusses the differences between the observation-driven ARCH models and the parameter-driven stochastic volatility models.

Our model is recursively estimated over real time samples, and the resulting time series of the resulting coefficients and variances could potentially change smoothly or abruptly over time. Since the recursive VAR-ARCH estimation proposed is observation-driven, rather than parameter-driven as in stochastic volatility models, by allowing for both types of changes leaves it up to the data to determine whether the time variation has been abrupt or gradual.

The goal of this paper is to contribute to the debate in the literature regarding changes in monetary policy by considering a flexible specification that accounts for potential nonlinearity, heteroskedasticity, gradual and abrupt changes, and non-normality in one integrated empirical framework, without imposing functional forms on the relations between the parameters and the observations. In addition, by allowing time-dependent coefficients (including $\mathbf{B}_0(\mathbf{x}_t)$) and covariance, the framework allows analysis of changes in the size of shocks and on their effect on the variables in the system. Finally, the empirical results may suggest a parametric structure of the functions independent of specific model assumptions.

5. Empirical Results¹³

The Data

We use real-time data for the analysis. These are the original and unrevised data available to economic agents at any given date in the past as opposed to the commonly used revised data currently available from government statistical agencies. The real time series were painstakingly compiled by the Federal Reserve Bank of Philadelphia. Details on how they were constructed can be found in Croushore and Stark (2001).

The series used in the two baseline VARs are the federal funds rate, the unemployment rate, and the annualized log first difference of the commodity price index, the real GDP, and of the GDP deflator. The last two variables have been substantially revised over the period considered, including major definitional changes as well as correction of discrepancies caused by lags in the availability of primary data. We use the collected real time realizations, or *vintages*, of these time series and of the unemployment rate as they would have appeared at the end of each quarter from the fourth quarter of 1965 to the first quarter of 2007.¹⁴ For each vintage the sample collected begins in the second quarter of 1959 and ends with the most recent data available for that vintage. Since the federal funds rate and the commodity price index are not revised we use the currently available series.¹⁵

¹³ Since the real time analysis involves recursive estimation for each vintage, each point in the Figures 1, 5, 6, 8, and 9 corresponds to estimates from applying the local linear estimation to a particular vintage.

¹⁴ See Orphanides and Van Norden (2005) for a discussion of the reliability of real time estimates of alternative ways to detrend output in real time.

¹⁵ The federal funds rate is obtained from the St. Louis F.R.E.D. and commodity price index from the Commodity Research Bureau (CRB).

The local nonparametric estimation of the VAR applied recursively to the increasing real time samples yields a set of parameters for the 1965:4 vintage, based on the observations from 1959:1 to 1965:3 after taking into account the lags p ; a second set for the 1966:1 vintage and so on until the 2007:1 vintage. We, thus, obtain times series of the coefficients, variances, and impulse response functions resulting from each recursive estimation. The reader is referred to Section 5.4 for a discussion of the robustness of the results to alternative specifications.¹⁶

5.1 Evolution of Shocks

It has been documented in the literature that the variance of innovations to monetary policy has decreased substantially across sub samples. Figure 1a shows the recursively estimated real time variance of the identified monetary policy shocks and 90% bootstrap confidence bands.¹⁷ The results confirm previous findings of time varying variance, with a large difference in the estimates before and after 1979. In particular, the variance of interest rate shocks rises considerably in the mid 1970s and more than double between 1979 and 1982. This last feature is associated with Volcker's regime, in which the Fed's operation policy changed to target monetary aggregates. We find that after reaching a peak in the early 1980s, the variance of the shocks to interest rates decreases slowly but steadily until the end of the sample. Although the recent values are a lot smaller than the ones in the early 1980s, they are higher than the ones prevailing in the 1960s and early 1970s. Cogley and Sargent (2005) also find a similar pattern with a gradual step increase in variance in the pre-1979 period, but the variances are found to fall more slowly in our framework in the second half of the sample. Primiceri (2005), on the other hand, finds a sharper rise and decrease in the variance right around 1979-1982.

One of the possible reasons for the difference in the results is the estimation over recursive samples and the ARCH modeling approach in which the error variance is associated with the conditional mean of the squared data. That is, changes in interest rate variability depend on lags of the level of the interest rate itself. In fact, the variance of the shocks follows the same humped shape as the level of interest rates, with a sharp rise in the late 1970s and a gradual decrease since mid 1980s.

¹⁶ As discussed in Section 5.4, the results obtained from the two VARs are very similar, and we report in the empirical section the results from the VAR that includes more information.

¹⁷ Bootstrap confidence bands for this section and the subsequent ones are constructed based on the method proposed by Franke et al (2003, 2004) for nonparametric autoregressive function with conditional heteroskedasticity. The basic idea of the nonparametric bootstrap is to mimic the structure of the data generating process in equation (9) (Section 3.1) by resamples from the empirical distribution for calculating confidence bands for the parameters of the model. The number of Monte Carlo replications for the bootstrap is chosen to be equal to 399. The computational burden of implementing the nonparametric approach with a data-driven method is nontrivial.

5.2 Responses to Monetary Policy Shocks

We study potential time varying changes in the impact of monetary shocks by examining the real time response of output, inflation, and unemployment in different dates in the sample. We choose 1975:1, 1980:4 to 1981:3, 1996:1, and the last period of the sample, 2006:4. These dates are selected to reflect different economic conditions, and to represent the governance of Burns, Volcker, Greenspan, and Bernanke.

We find results that are broadly consistent with existing literature. As Bernanke and Mihov (1998), Clarida, Gali, and Gertler (2000), Cogley and Sargent (2001, 2005), Lubik and Schorfheide (2004), Boivin and Gianonni (2002, 2006), Boivin (2006), and Kuttner and Mosser (2002), among several others, we find a reduction in the effect of monetary policy shocks.

This is also found in Koop, Leon-Gonzalez, and Strachan (2007), especially with respect to the response of inflation, and less so for the impact on unemployment. Although the confidence bands are large and contain zero individually, they conclude that this does not imply that jointly there are no differences in impulse responses over time. Primiceri's (2005) results are qualitatively similar, but the latter concludes instead that coefficients do not vary much over time.

Figures 2a, 3a, and 4a show the impulse response function of unemployment, output, and inflation to an unexpected increase in the Fed funds rate in these dates.¹⁸ The other graphs in these figures plot the pairwise differences between responses for some dates along with 90% bootstrap confidence bands. Monetary policy shocks have a strong and lasting negative effect on unemployment during the Volcker period, and the uncertainty around these values increases over the horizon considered. The response of unemployment is also large in the Burns period, but it reaches a peak after five quarters. In the Greenspan period the response of unemployment is substantially smaller than in the previous ones.

The negative response of output follows a similar pattern, being strongest in the Volcker period, especially between 1980:3 and 1981:3, and weakest in the more recent periods in 1996:1 and in 2006:4. The pairwise differences across dates are substantial and significant when taking into account the 90% bootstrap confidence bands.¹⁹ The response of inflation to monetary policy shocks shows an even greater time variation. The largest negative response occurs in the 1975:1, in the Burns period, followed by milder but still negative impact in the Greenspan and the Bernanke periods (1996:1 and 2006:4). On the other hand, inflation shows a positive response in the Volcker period between 1980:4 and 1981:3, even though the VAR considered includes commodity inflation.²⁰

¹⁸ The response of unemployment between 1980:4 and 1981:3 is similar, so we only plot one of these subsamples.

¹⁹ There is a small but insignificant positive increase in output in the second period for some dates, which is also found by several authors and is named 'output puzzle' (see e.g. Boivin and Gianonni 2002 or Giordani 2004).

²⁰ Sims (1992), Christiano, Eichenbaum, and Evans (1996) and Sims and Zha (2006b) show that when one modifies such shock measures by including current and lagged values of commodity prices the price puzzle often decreases.

A broader picture of the changes can be seen in the evolution of the impulse response functions over recursive real time samples. As an illustration, Figure 5 plots the time series of the real time response of output to monetary policy shocks in the third quarter, which is the period with strongest response as found in Figure 3a. The response gradually increases (in absolute value) from the beginning of the sample until it reaches a peak in 1981. In fact, the largest responses are found between 1980 and 1982. From then on the response is much more stable and decreases gradually until the end of the sample to values smaller than in the 1970s and 1980s.

Figure 6 shows the evolution of the response of inflation to monetary policy shocks in horizons three and four. The third quarter captures the dynamics of the price puzzle. Some authors have shown that the price puzzle is sensitive to the sample used, and that it is stronger for the period pre-1980 than for post-1980.²¹ We find that a significant price puzzle is present in only a small part of the sample, between 1977 and 1981, with a sharp decrease in the positive response of inflation in the first quarter of 1981. This suggests that analyses based on comparison of sub-samples can be quite misleading depending on the dates chosen.²²

The dynamics of the response in quarter four plotted in Figure 6 indicates that inflation had the strongest negative response to monetary shocks between 1974 and 1976 (see Figure 4a). The point estimates of the response become positive during the Volcker disinflationary period, and turn negative again from the mid-1980s on. However, the uncertainty surrounding the estimates as shown by the confidence bands only allows concluding that there was large negative response in the mid 1970s compared to other times, and also a significant milder negative response since the mid 1990s and on.

In summary, the evidence confirms previous discussions regarding a recent reduced impact of monetary policy shocks on output, unemployment, and inflation compared to historical values. However, we find that the responses are not only time-varying, but also highly nonlinear with more oscillations in the first part of the sample and more stability in the second part. In addition, there have been some abrupt changes in the responses of inflation to policy shocks.

5.3 Real Time Changes in Systematic Monetary Policy

We consider the evolution of the real time response of interest rate to inflation and unemployment by estimating a forward-looking Taylor rule based on Clarida, Gali, and Gertler (2000):

$$i_t = \beta_0 + \beta_1 E_t \bar{\pi}_{t,t+\zeta_\pi} + \beta_2 E_t \bar{u}_{t,t+\zeta_u} + \beta_3 i_{t-1} + \eta_t, \quad (12)$$

²¹ See, for example, Balke and Emery (1994), Giordani (2004), Hanson (2004), Castelnuovo and Surico (2006), etc.

²² A counterpart *parametric* VAR (see Section 5.4 for details on the model) estimated for sub-samples also shows responses that are quite different from one quarter to the next in 1981.

where i_t is the short term interest rate, $\bar{\pi}_{t,t+\zeta_\pi}$ and $\bar{u}_{t,t+\zeta_u}$ are average inflation and average unemployment from t to $t+\zeta_\pi$, and from t to $t+\zeta_u$, respectively. The Taylor principle implies that the long run interest rate response to inflation is:

$$\beta^{LR} = \frac{\beta_1}{1 - \beta_3}.$$

This parameter is generally interpreted in the literature as a measure of the degree of activism of the systematic monetary policy. If $\beta^{LR} \geq 1$, then the policy rule is said to be activist, that is, there is a more than proportional increase in the nominal interest rate when inflation rises, leading to an increase in the real interest rate. In a passive policy, the increase in the nominal interest rate is less than proportional, which can be measured by $\beta^{LR} < 1$.

Boivin (2006), Primiceri (2005), and Cogley and Sargent (2005) also consider a time-varying degree of activism, and we compare our results to theirs. We follow Cogley and Sargent (2005) in estimating the parameters of the policy rule as projection coefficients from the VAR in two steps. First, the forecasts of average inflation and unemployment are projected onto a set of instruments. The instruments are the forecasts $E_{t-1}\bar{\pi}_{t,t+\zeta_\pi}$ and $E_{t-1}\bar{u}_{t,t+\zeta_u}$ for $E_t\bar{\pi}_{t,t+\zeta_\pi}$ and $E_t\bar{u}_{t,t+\zeta_u}$, respectively, which are obtained from estimating over recursive real time samples the identified VAR.²³ In a second step, Equation (12) is estimated with these fitted values, using the nonparametric local least squares method described in Section 3.1. Thus, we do not impose a parametric density for estimating the Taylor rule.

Figure 7 shows the time varying real time estimates of β^{LR} . The movements broadly agree with the results of several authors, including Judd and Rudebusch (1998), Clarida, Gali, and Gertler (2000), Boivin (2006), and Cogley and Sargent (2001, 2005), which find that the policy rule was passive during the Burns period, and activist during Volcker's disinflation period and during Greenspan's chairmanship. However, we find that the uncovered behavior of the parameter is more complex than these general conclusions suggest. The degree of activism displays sharp movements and nonlinearities over time. In particular, the early 1970s are characterized by strong activist policy, which lasted until around the oil shock and recession in 1973-1974. In 1975 the response to inflation falls sharply below one. Policy remains passive until around 1978-1979. During the Volcker disinflation period the degree of activism increases above one and it remains active throughout the rest of the sample. However, notice that

²³ Similarly to Cogley and Sargent (2005), we project future inflation and unemployment on a constant and two lags of each. In addition we assume that $\zeta_\pi = 4$ and $\zeta_u = 2$, consistent with the transmission mechanism of changes in interest rate and the response of inflation and unemployment. We have also considered other short term horizons such as $\zeta_\pi = 3$ and $\zeta_u = 3$, as well as other combinations, and the results were not very sensitive to these choices.

β^{LR} falls somewhat in the mid 1980s – but to values still above one, and rises again during the Greenspan period stabilizing around values well above one.

The magnitude of the policy response and its changes over time needs to be evaluated taking into account the statistical uncertainty around these estimates. In fact, there are times in which we can not conclude that the level of the response is below or above one. On the other hand, some changes in the degree of activism are significant when considering the 90% confidence bands, such as a passive policy from the mid to the late 1970s, and an active policy during Volcker’s disinflationary period and during Greenspan’s period.

Compared to the literature, our results are closer to the TVP version of the Taylor rule proposed by Boivin (2006), particularly the specification that allows for three regimes for the variance.²⁴ Using this framework, Boivin (2006) finds variations in the dynamics of the degree of activism at around the same time as ours. Cogley and Sargent’s (2005) results concur with ours in their middle sample, but displays large oscillations in the beginning and end of the sample.²⁵ On the other hand, Primiceri (2005) finds that although policy is more reactive to inflation in the last two decades compared to the 1960s and 1970s, the long run response to inflation remains above one in the entire sample. Primiceri (2005) attributes this difference to the fact that the results are based on smoothed estimates that use full sample information. However, Cogley and Sargent (2005) use a similar framework and estimation methods and find large swings in β^{LR} , with values reaching below one from the mid to the late 1970s. The difference in timing and magnitude of the changes in the long run response obtained by these papers might be a product of the parametric assumptions made in these models. If the assumptions do not hold, the parameters might be biased and inconsistent. A particular appeal of recursively estimating the proposed model allows detection of discrete local deviations as well as more gradual ones, without smoothing the timing or magnitude of the changes.

5.4 Robustness

Comparison of Results with Alternative Specifications

Results from VAR models are known to be sensitive to their specification. Our standard specification delivers sensible response functions, some of the results being broadly consistent with existing literature. For instance, Bernanke and Mihov (1998), Clarida, Gali, and Gertler (2000), Barth and Ramey (2001), Cogley and Sargent (2001, 2005), Boivin and Gianoni (2002, 2006), Boivin (2006), among others report a similar reduction in the effect of policy shocks on inflation, output (or unemployment). In particular,

²⁴ The middle one is from the fourth quarter of 1979 to the fourth quarter of 1982.

²⁵ Their estimates obtained in the mid-sample are the ones surrounded by least uncertainty in their methodology.

Bernanke and Mihov (1998) use a more sophisticated representation of the Fed's operating procedure while Barth and Ramey (2001) use long-run restrictions and attain similar findings.

We have also estimated two structural VARs that include three or five variables as discussed in Section 2. The inclusion or exclusion of information in our VAR does not affect the main conclusions obtained. In fact, the results are strikingly similar. One difference is that, as expected, the exclusion of commodity inflation accentuates the price puzzle – it is significant for a much longer period.²⁶

Next we address the question on whether the results are sensitive to the VAR ordering. Although, this assumption is more appropriate when using revised data, we have alternatively estimated the model assuming that policymakers do not observe contemporaneously the non-policy variables, which is motivated by lags in the collection of data (i.e. ordering the federal funds rate first). We find the results qualitatively equivalent in both cases. There are only minor changes in the size of the innovation variance of the federal funds rate but with the same broad pattern of evolution over time. In addition, we find a larger but insignificant output puzzle when the policy variable is ordered first. This is consistent with the comparison of these orderings in Christiano, Eichenbaum, and Evans (1999). Finally, changing the order of the non-policy variables has insignificant effects on the results.

In summary, the main empirical features that we aim to explain, namely the time-varying changes and reduced effect of monetary shocks on output and inflation, the pattern of changes in variances over time, and time variations in the degree of activism are corroborated by different specifications and identifying assumptions.

We have also estimated recursively a parametric counterpart of the structural VARs with three and five variables and the same identification restrictions as our benchmark specifications. In order to capture heteroskedasticity we estimate the model assuming three different regimes from the variance. The middle one corresponds to the high volatile period between 1979:4 and 1984:1. Based on the variance plotted in Figure 1a, this division captures major heteroskedastic episodes in the monetary policy shocks.²⁷

Comparing the results, we find analogous temporal evolution in the recursive coefficients and variances of shocks obtained with the parametric VAR. Although the patterns are similar, the parametric VAR yields estimates that substantially underestimate or overestimate the nonparametric estimates. As an illustration, Figure 8 compares the dynamics of the variance of interest shocks. The parametric VAR underestimates the variance of interest shocks for most of the sample, except for the period between 1981

²⁶ For sake of brevity we do not report the results of the VAR with three variables here. The full range of results for this specification is reported in a previous version of this paper and is available from the authors on request. We have also estimated this version of the VAR considering changes instead of the level of interest rates as in Gali (1992).

²⁷ The parametric structural VARs were estimated using maximum likelihood by the method of scoring.

and 1986, when the variance is overestimated.²⁸ If the normality or functional form assumptions on the parametric specification do not hold, the coefficients of the parametric VAR may be biased.

We examine normality of the structural innovations of the parametric structural VAR using several tests.²⁹ Since the innovations are assumed to be serially uncorrelated, we first apply multivariate extensions of the traditional Jarque-Bera test and Kolmogorov–Smirnov (KS) tests.³⁰ The tests are applied to several dates in the sample. For all sub-periods examined we reject the joint normality of innovations with a p-value near zero, even for periods that exclude the highly volatile Volcker’s disinflation period, and for the more recent period. The tests also reject normality equation by equation for all sub-periods considered for at least three equations. Finally, normality is also rejected for the unemployment equation in all sub-samples examined.

We further investigate the normality assumption using Fan, Li, and Min’s (2006) extension of Zheng’s (2000) nonparametric conditional normality test. The statistic critical values are obtained following the bootstrapping procedure suggested by Fan et al (2006) to better approximate their finite-sample null distribution. The test is applied to several sub-periods and to the full sample and we find that it overwhelmingly rejects the assumption of normality at any conventional significance level.

Finally we compare the results of recursively estimating the proposed monetary VAR using only data available in the 2007:1 vintage. This comparison can shed light on differences in findings due to the use of real time versus revised data. As an illustration, we show the difference in the response of output to the identified monetary shock.³¹ Figure 9 plots the evolution of the response of output in period 3 from the nonparametric VAR estimated over real time recursive samples (same time series as in Figure 5) and the counterpart estimated series obtained using revised data (i.e. each point is the result of estimating the VAR for increasing samples from 1959:1 to 2006:4 using the same dataset as of 2007:1). Although the response shows the same general temporal pattern, the revised version underestimates the negative response of output to interest rate until the 2000s. This is especially accentuated during the Volcker disinflationary period. In addition, the decrease in the impact of monetary policy shock on output in the later part of the sample is much more pronounced when using real time data.

The estimation using revised data also overestimates or underestimates the other parameters of the model during some periods, concealing the impact of shocks that took place in real time. The findings support Orphanides (2001, 2002, 2004), which argue that evidence based on revised data does not

²⁸ The response of the economy also shows different timing and intensity to interest rate shocks estimated from the parametric VAR.

²⁹ We estimate the parametric model with lags $p=1, 2, 4$.

³⁰ Bai and Ng (2005) propose a method to test for normality on time series data with weakly dependence and apply it to some of the individual series used in our structural VAR. They find that the normality assumption is rejected for unemployment rate and inflation, but not for output growth.

³¹ Notice that the series for the federal funds rates it is not affected by definitional changes or any other revisions.

accurately depicts the information that was actually available to the Fed at the time decisions were made. We find that the recursive real time shocks are significantly different from the shocks based on revised data, which distort assessment of their impact on the economy in real time. The residuals obtained from the real time recursive estimation better reflect monetary policy shocks as taken place in real time.

We are investigating in an on-going project the implications of using different sets that carry out current or forward looking information on the parameter of the models. In particular, counterfactual experiments are used to reflect the Fed's information set in real time but with a transmission mechanism based on full information.

6. Conclusion

This paper investigates potential changes in monetary policy shocks, in the transmission mechanism, and in the Fed's reaction function in the last 50 years using a VAR-ARCH model. The structural monetary VAR-ARCH is estimated using a framework in which all model parameters are time-dependent. The method is applied to real time unrevised data and takes into account the possibility of abrupt or gradual changes in the parameters.

We find substantial evidence of time-varying changes and nonlinearities in the variance of shocks, in the impact of monetary policy shocks on the economy, and on the Fed's reaction function. There is considerably evidence of non-systematic and systematic changes in monetary policy, which have been both gradual and abrupt, with highly nonlinear responses from one period to the next at times. These nonlinearities in the relationships suggest that caution should be applied when analyzing changes through comparison of sub-samples. Overall, there has been a reduced negative impact of monetary policy in the economy. However, we can not conclude that inflation has been recently less responsive to monetary policy. In fact, point estimates indicate that inflation was the least responsive to policy shocks in the early 1980s, with an abrupt change in the response in 1981. This evidence suggests that the way instability is modeled is very important to uncover properly the nature of the changes in monetary policy during the 1970s and 1980s. Finally, we find that policy was significantly more active in Volcker's disinflationary period and in the Greenspan period than in some (but not all) parts of the Burns period.

The evidence from the evolution of changes in the variance of shocks, in the response of the economy, and in the Fed's response indicates that the dynamics are complex and highly nonlinear, and several of the changes took place simultaneously. In addition, the uncertainty surrounding estimates precludes precise conclusions regarding the source of changes.

References

- Bachmeier, L., S. Leelahanon, and Q. Li (2007), "Money Growth and Inflation in the United States," *Macroeconomic Dynamics*, 11, 113-127.
- Bai, J., and S. Ng (2005), "Tests for Skewness, Kurtosis, and Normality for Time Series Data," *Journal of Business & Economic Statistics*, 23, 49-60.
- Balke, N., and K.M. Emery (1994), "Understanding the Price Puzzle," *Federal Reserve Bank of Dallas Economic Review*, 4, 15-26.
- Barth, M. J., and V.A. Ramey (2001), "The Cost Channel of Monetary Transmission," *NBER Macroeconomics Annual*, 199-239.
- Bernanke, B. S., and A.S. Blinder (1992), "The Federal Funds Rate and the Channels of Monetary Transmission," *American Economic Review*, 82, 901-921.
- Bernanke, B.S., M. Gertler, and M.W. Watson (1997), "Systematic Monetary Policy and the Effects of Oil Price Shocks." *Brookings Papers on Economic Activity*, 1997, 91-142.
- Bernanke, B. S., and I. Mihov (1998), "Measuring Monetary Policy," *Quarterly Journal of Economics* 113, 869-902.
- Boivin, J. (1999), "Revisiting the Evidence on the Stability of Monetary VARs." Working Paper, Columbia University, Graduate School of Business.
- Boivin, J. (2006), "Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real Time Data," *Journal of Money, Credit and Banking*, 38, 1149-1173.
- Boivin, J., and M. Gianonni (2002), "Assessing Changes in the Monetary Transmission Mechanism: A VAR Approach," *FRBNY Economic Policy Review*, 8, 97-111.
- Boivin, J., and M. Gianonni (2006), "Has Monetary Policy Become More Effective?" *Review of Economics and Statistics*, 88, 445-462.
- Bossaerts, P., C. Hafner, and W. Härdle (1996), "Exchange Rates Have Surprising Volatility," in *Time Series Analysis*, eds. P.M. Robinson and M. Rosenblatt, New York: Springer-Verlag.
- Cai, Z. (2007), "Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors," *Journal of Econometrics*, 136, 163-188.
- Castelnuovo, E., and P. Surico (2004), "Model Uncertainty, Optimal Monetary Policy and the Preferences of the Fed," *Scottish Journal of Political Economy*, 51, 105-126.
- Christiano, L.J. and M. Eichenbaum (1992), "Identification and the Liquidity Effect of a Monetary Policy Shock," in *Political Economy, Growth and Business Cycles*, ed. by A. Cukierman, Z. Hercowitz, and L. Leiderman, Cambridge and London: MIT Press, 335- 370.
- Christiano, L.J., Eichenbaum, M. and Evans, C. (1996), "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds," *Review of Economics and Statistics*.
- Christiano, L. J., M. Eichenbaum, and C. Evans (1999), "Monetary Policy Shocks: What Have We Learned and to What End?" in *Handbook of Macroeconomics*, eds. J. B. Taylor and M. J. Woodford, Amsterdam: Elsevier, 65-148.
- Clarida, R., J. Galí, and M. Gertler (2000). "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115, 147-180.
- Cogley, T., and T.J. Sargent (2001), "Evolving Post World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual*, 16, 331-373.

- Cogley, T., and T. J. Sargent (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Dynamics*, 8, 262-302.
- Croushore, D., and T. Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105, 111-130.
- Eichenbaum, M. (1992), "Comment on Interpreting The Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review*, 36, 5, 1001-1011.
- Fan, J. (1993), Local Linear Regression Smoothers and their Minimax Efficiency. *Annals of Statistics*, 21, 196-216.
- Fan, J., and I. Gijbels (1995), "Data-Driven Selection in Polynomial Fitting: Variable Bandwidth and Spatial Adaptation," *Journal of the Royal Statistical Society- Series B*, 57, 371-394.
- Fan, J., and I. Gijbels (1996), "Local Polynomial Modeling and Its Applications," *Monographs on Statistics and Applied Probability* 66, London: Chapman and Hall.
- Fan, J., Q. Li, and I. Min (2006), "A Nonparametric Bootstrap Test of Conditional Distributions," *Journal of Nonparametric Statistics*, 2, 389-403.
- Fan, J., and Q. Yao (1998), "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika*, 85, 645-660.
- Franke, J., Kreiss, J., Mammen, E. and Neumann, M. (2003), "Properties of the Nonparametric Autoregressive Bootstrap," *Journal of Time Series Analysis*, 23, 555-585.
- Franke, J., M.N. Neumann, and J.P. Stockis (2004), "Bootstrapping Nonparametric Estimators of the Volatility Function," *Journal of Econometrics*, 118, 189-218.
- Giordani, P. (2004), "An Alternative Explanation of the Price Puzzle", *Journal of Monetary Economics*, 51, 1271-1296.
- Hall, P., and R. J. Carroll (1989), "Variance Function Estimation in Regression: The Effect of Estimating the Mean," *Journal of the Royal Statistical Society, Series B*, 51, 3-14.
- Hanson, M. S. (2004), "The 'Price Puzzle' Reconsidered," *Journal of Monetary Economics*, 51, 1385-1413.
- Härdle, W., and A. Tsybakov (1997), "Local Polynomial Estimators of the Volatility Function in Nonparametric Autoregression," *Journal of Econometrics*, 81, 223-242.
- Härdle, W., A. Tsybakov, and L. Yang (1998), "Nonparametric Vector Autoregression." *Journal of Statistical Planning and Inference*, 68, 221-245.
- Hastie, T., and Loader, C. (1993), "Local Regression: Automatic Kernel Carpentry (with discussion)," *Statistical Science*, 8, 120-143.
- Hoover, K., S. Demiralp, and S. Perez (2008), "A Bootstrap Method for Identifying and Evaluating a Structural Vector Autoregression," forthcoming in the *Oxford Bulletin of Economics and Statistics*.
- Judd, J. P., and G.D. Rudebusch (1998) "Taylor's Rule and the Fed: 1970-1997." *Federal Reserve Bank of San Francisco Economic Review* 3, 1-16.
- Kalyanaraman, K. "Bandwidth Choice for Regression Functionals with Application to Average Treatment," mimeo, Harvard University, January 2009.
- Koop, G., R. Leon-Gonzalez, and R. Strachan (2007), "On the Evolution of Monetary Policy," mimeo, University of Strathclyde.
- Kuttner, K. N., and P.C. Mosser (2002), "The Monetary Transmission Mechanism: Some Answers and Further Questions," *FRBNY Economic Policy Review*, 8, 15-26.

- Lubik, T., and F. Schorfheide (2004), "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 94, 190-217.
- Lütkepohl, H. (1990), "Testing for Causation Between two Variables in Higher Dimensional VAR Models," paper presented at the World Congress of the Econometric Society, Barcelona.
- Masry, E. (1996a), "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17, 571-599.
- Masry, E. (1996b), "Multivariate Regression Estimation: Local Polynomial Fitting for Time Series," *Stochastic Processes and Their Applications*, 65, 81-101.
- Nadaraya, E. A. (1965), "On Nonparametric Estimation of Density Functions and regressions Curves," *Theory of Probability and its Application*, 10, 186-190.
- Neumann, M. (1994), "Fully Data-Driven Nonparametric Variance Estimators," *Statistics*, 25, 189-212.
- M. Normandin, L. Phaneuf (2004), "Monetary policy shocks: Testing Identification Conditions under Time-Varying Conditional Volatility," *Journal of Monetary Economics*, 51, 1217-1243.
- Orphanides, A. (2001), "Monetary Policy Rules Based on Real-Time Data," *American Economic Review*, 91, 964-985.
- Orphanides, A. (2002), "Monetary Policy Rules and the Great Inflation," *American Economic Review*, 92, 115-121.
- Orphanides, A. (2004), "Monetary Policy Rules, Macroeconomic Stability and Inflation: A View from the Trenches," *Journal of Money, Credit and Banking*, 36, 151-175.
- Orphanides, A. and S. Norden (2005), "The Unreliability of Output Gap Estimates in Real Time," *Review of Economics and Statistics*, 84, 569-583.
- Parzen, E. (1962), "On Estimation of a Probability Density and Mode," *Annals of Mathematical Statistics*, 33, 1065-1076.
- Primiceri, G. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72, 821-852.
- Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics*, 27, 642-669.
- Rotemberg, J. J., and M. Woodford (1997), "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macroeconomics Annual*, 297-346.
- Ruppert, D., and M.P. Wand (1994), "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22, 1346-1370.
- Ruppert, D., Wand, M., Holst, U., and O. Hössjer (1997), "Local Polynomial Variance Function Estimation," *Technometrics*, 39, 262-273.
- Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility", in *Time Series Models in Econometrics, Finance and Other Fields*, eds. D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen, London: Chapman and Hall, 1-67.
- Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.
- Sims, C. A. (1992), "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review*, 36, 975-1000.
- Sims, C. A. (1999), "Drift and Breaks in Monetary Policy," Unpublished Manuscript, Princeton University, Department of Economics.

- Sims, C. A. (2001), "Stability and Instability in US Monetary Policy Behavior," Unpublished Manuscript, Princeton University, Department of Economics.
- Sims, C. A., and Zha, T. (2006a), "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, 96, 54-81.
- Sims, C. A., and Zha, T. (2006b), "Does Monetary Policy Generate Recessions?" *Macroeconomic Dynamics*, 10, 2, 231-272.
- Stock, J. H. (2001), "Discussion of Cogley and Sargent 'Evolving Post World War II U.S. Inflation Dynamics'," *NBER Macroeconomics Annual*, 16, 379-387.
- Swanson, N. R., and , C.W.J. Granger (1997), "Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions," *Journal of the American Statistical Association*, 92, 357-367.
- Tierney, H. L. R. (2005), "Essays on Inflation and Nominal Interest Rates," *Ph.D. Dissertation*, UC Riverside.
- Watson, G. S. (1964), "Smooth Regression Analysis," *Sankhya Series A*, 26, 359-372.
- Yao, Q. and H. Tong (1994), "Quantifying the Influence of Initial Values on Nonlinear Prediction." *Journal of the Royal Statistical Society. Series B (Methodological)*, 56, 4, 701-725
- Zheng, J.X. (2000), "A Consistent Test of Conditional Parametric Distributions," *Econometric Theory*, 16, 667-691.

Figure 1 – Real Time Innovation Variance of Interest Rate, Output, Inflation, and Unemployment, and 90% Bootstrap Confidence Bands

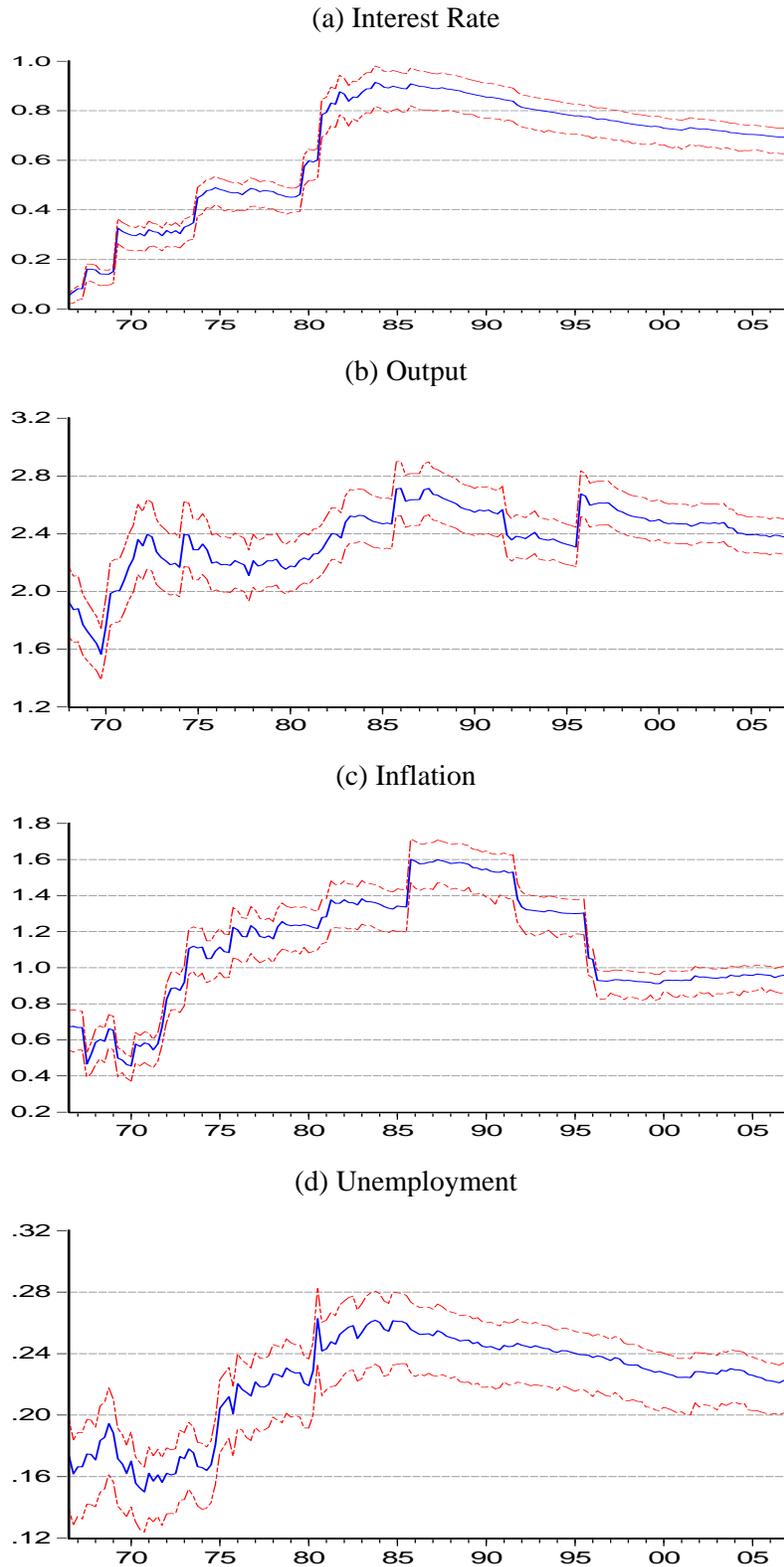


Figure 2 – Real Time Response of Unemployment to Shocks to Interest Rate in Different Dates and Pairwise Difference between Periods with 90% Bootstrap Confidence Bands

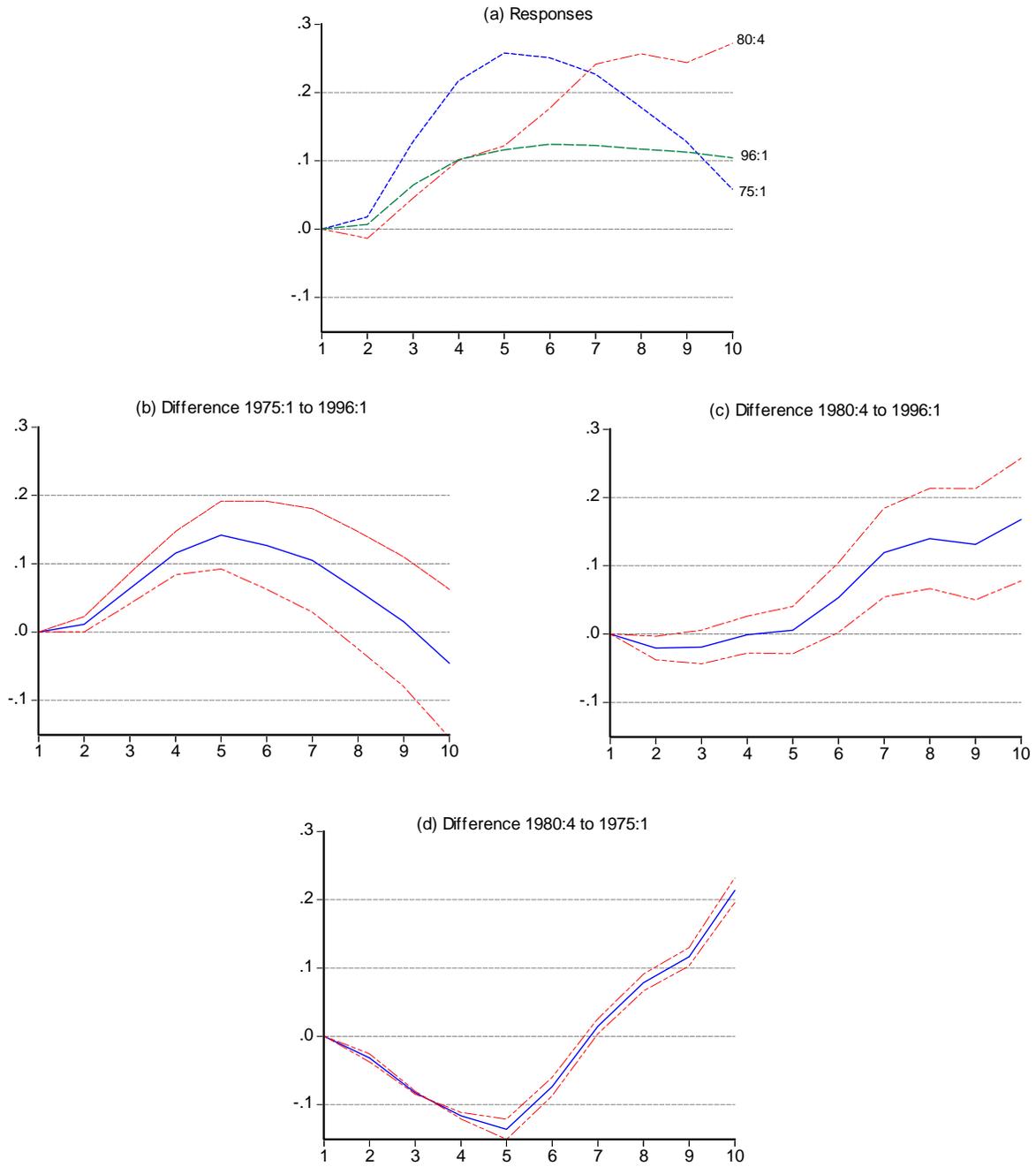


Figure 3 – Real Time Response of Output to Shocks to Interest Rate in Different Dates and Pairwise Difference between Periods with 90% Bootstrap Confidence Bands

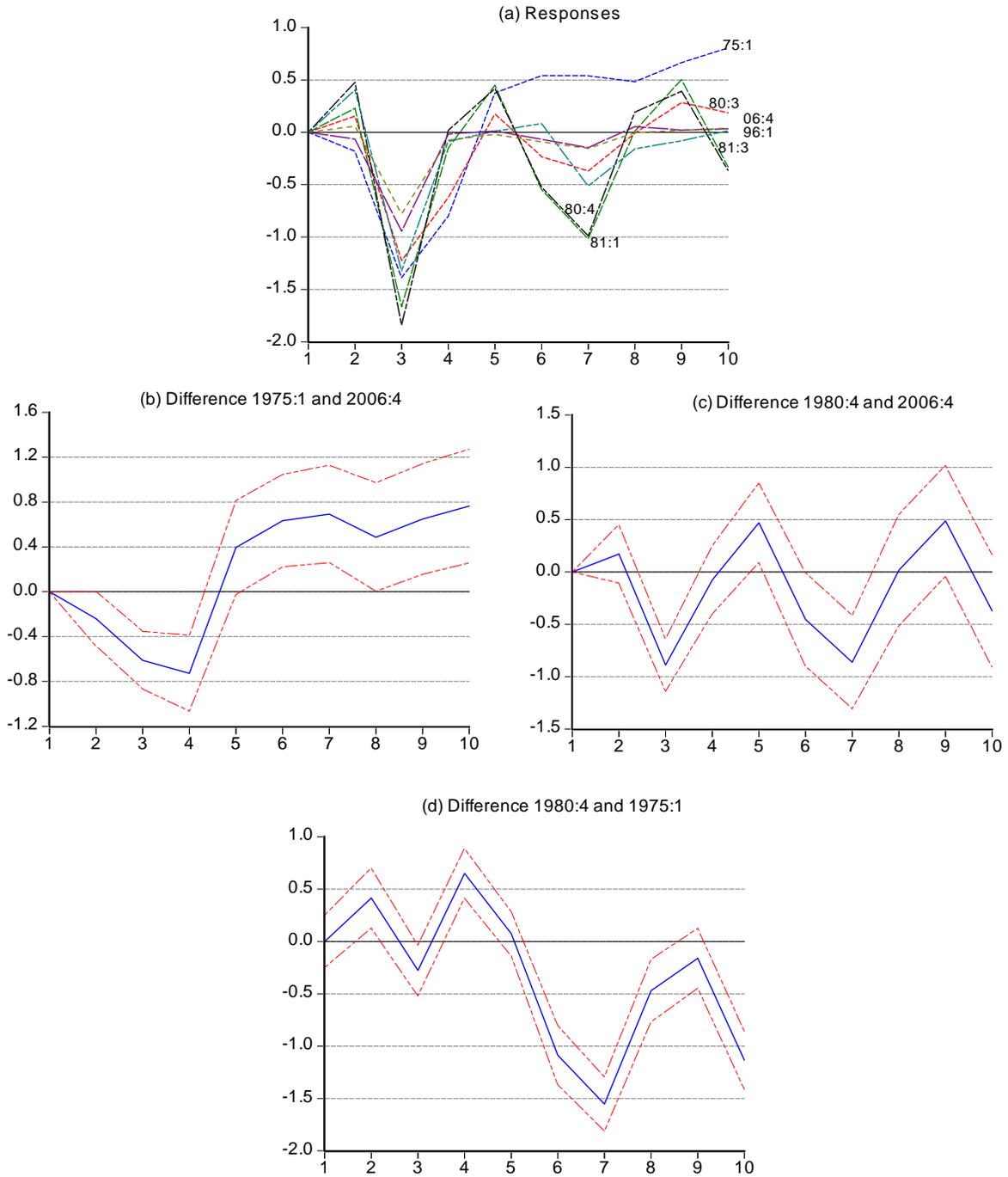


Figure 4 – Real Time Response of Inflation to Shocks to Interest Rate in Different Dates and Pairwise Difference between Periods with 90% Bootstrap Confidence Bands

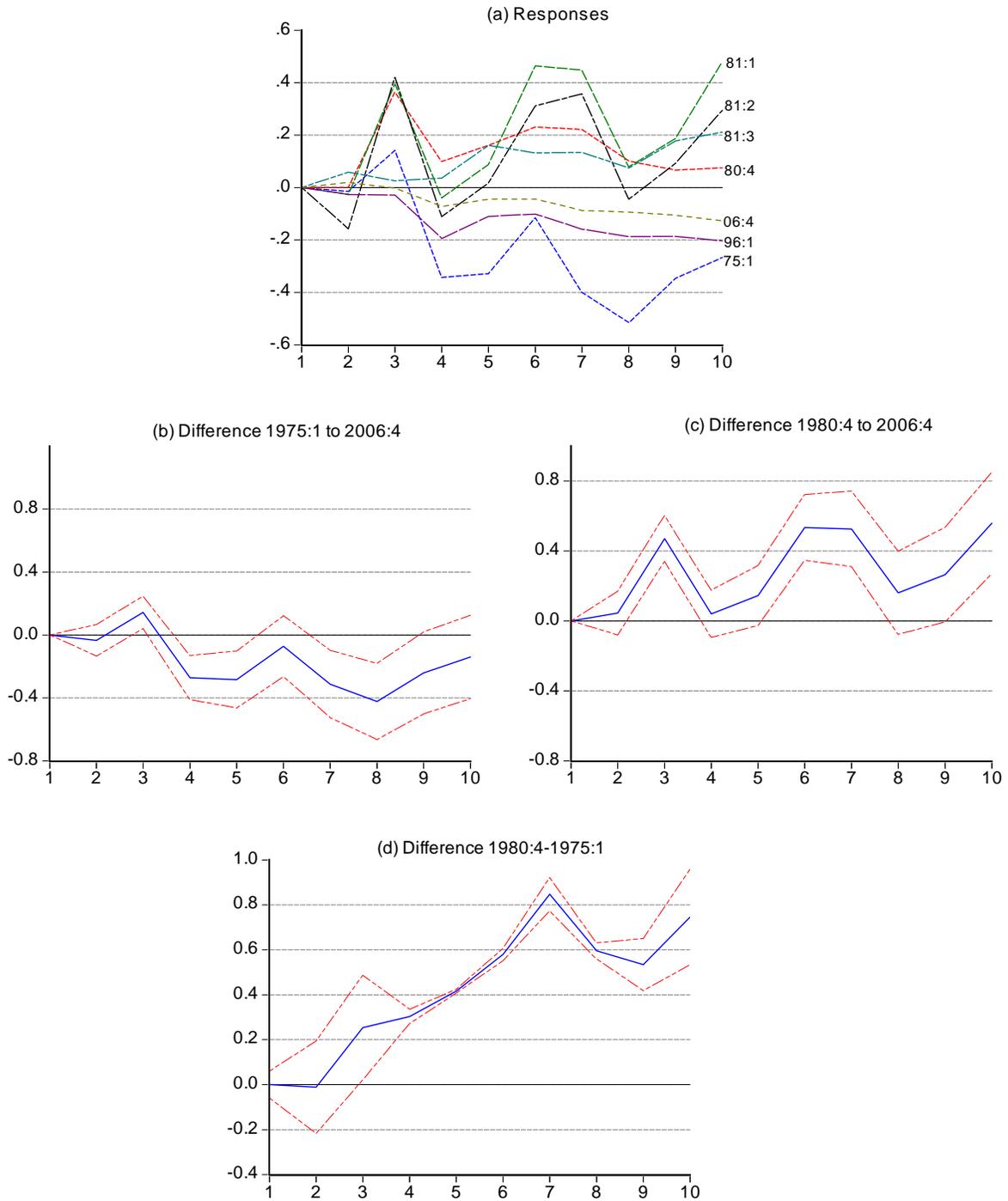


Figure 5 – Real Time Response of Output to Shocks to Interest Rate in Period 3 and 90% Bootstrap Confidence Bands

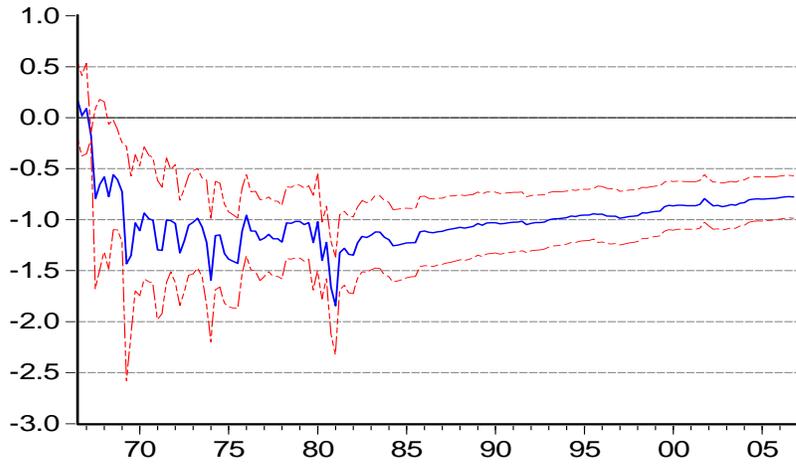


Figure 6 – Real Time Response of Inflation to Shocks to Interest Rate in Periods 3 and 4 and 90% Bootstrap Confidence Bands

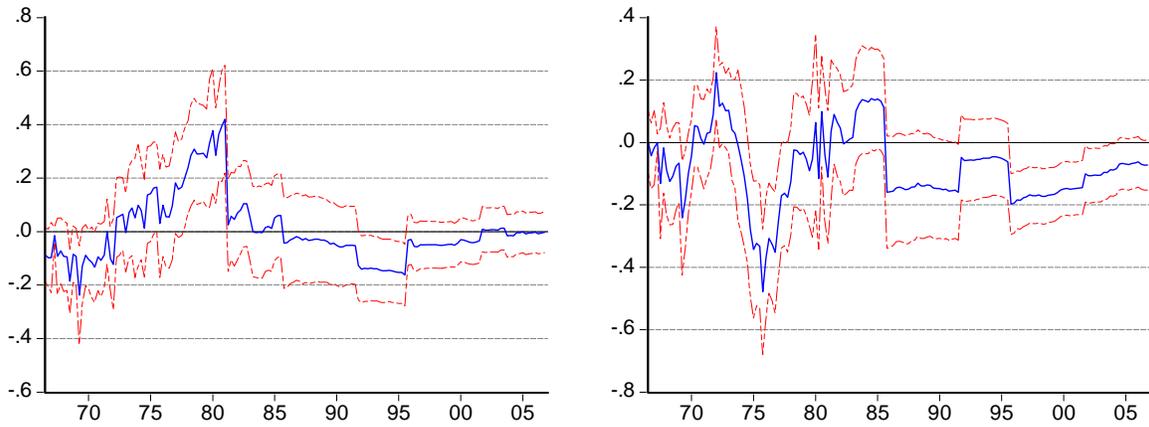


Figure 7 – Long Run Response of Interest Rates to Inflation – Estimates of Activism Coefficient and 90% Bootstrap Confidence Bands

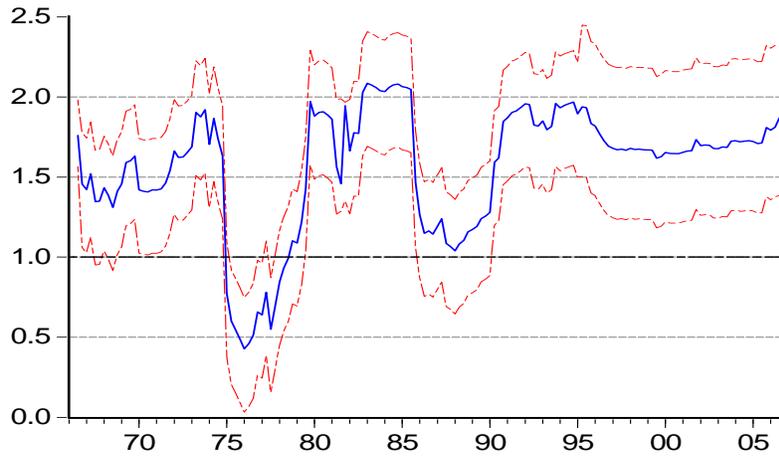


Figure 8 – Evolution of the Variance of Shocks to Interest Rate from Parametric (---) and Nonparametric (—) VAR

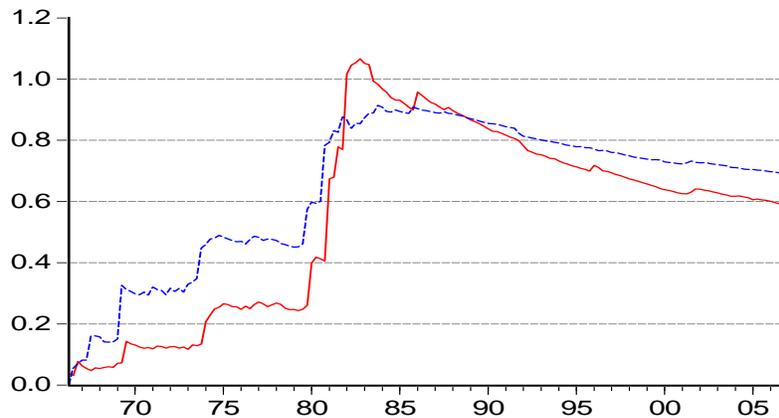
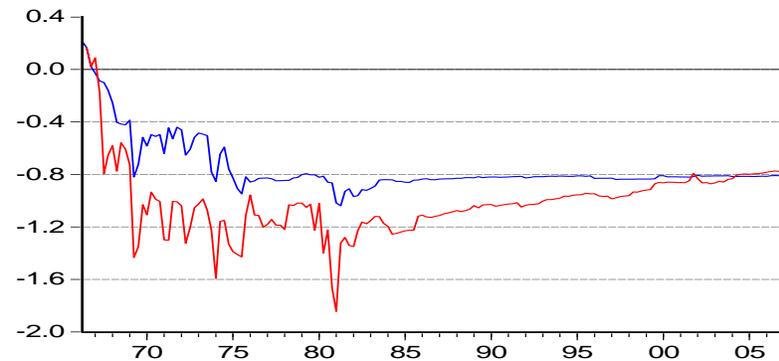


Figure 9 – Recursively Estimated Response of Output to Monetary Shocks of the Nonparametric VAR for 2001:7 Vintage, (—), for All Vintages in Real Time (---)



APPENDIX

Local Linear Non-Parametric Estimation

1. Conditional Expectation

Estimation

The conditional expectation $m_j(\mathbf{x}_t) = E(Z_{jt} | \mathbf{x}_t)$ from equation (9) (Section 3.1) can be approximated locally by:

$$m_j(\mathbf{x}_t) \approx m_j(\mathbf{x}) + m_j'(\mathbf{x})(\mathbf{x}_t - \mathbf{x})$$

for \mathbf{x}_t in the neighborhood of \mathbf{x} , obtained as a weighted average of the regressors using a multivariate

Gaussian kernel density, $K_j\left(\frac{\mathbf{x}_t - \mathbf{x}}{\mathbf{h}_j}\right)$. Let x_{ts} denote the s^{th} component of \mathbf{x}_t for $s = 1, \dots, (np + 1)$ and

$$k\left(\frac{x_{ts} - x_s}{h_{js}}\right) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x_{ts} - x_s}{h_{js}}\right)^2\right\} \quad s = 1, \dots, np + 1$$

be the univariate Rosenblatt-Parzen kernel

density (Rosenblatt 1956, Parzen 1962). The multivariate kernel density is:

$$K_j\left(\frac{\mathbf{x}_t - \mathbf{x}}{\mathbf{h}_j}\right) = k\left(\frac{x_{t1} - x_1}{h_{j1}}\right) \times \dots \times k\left(\frac{x_{t(np+1)} - x_{(np+1)}}{h_{j(np+1)}}\right), \quad (\text{A1})$$

where \mathbf{h}_j is a $(np + 1) \times 1$ vector of variable bandwidths. The method fits a local least squares hyperplane within an interval as specified by the window width of the kernel density function. For values of \mathbf{x}_t that lie

far from \mathbf{x} , for a given vector of window width \mathbf{h}_j , $K_j\left(\frac{\mathbf{x}_t - \mathbf{x}}{\mathbf{h}_j}\right)$ will be small. Hence, these observations

get low weight in determining $\hat{m}_j(\mathbf{x}_t)$.

The conditional expectation obtained from the local linear nonparametric method has important sampling properties as it adapts to a variety of random distributions of \mathbf{x}_t (Fan 1993).³² This is one of the best known methods for boundary adjustment since its bias is independent on the density. Hence, it potentially has smaller bias for regression estimation near the boundary of support. In addition, the best local linear estimator is the most efficient in an asymptotic minimax sense (efficiency in rates and in

³² A comparable approach is the nearest neighborhood estimator, in which the window width considers the distance of \mathbf{x}_t from the k^{th} (positive integers) nearest neighbor among the remaining $t-1$ data points. The results are generally comparable to local linear estimation, except for the known effects that nearest neighbor estimates appear rougher than local linear estimates and suffer more from the sparseness of data at the boundaries.

constant factor, Fan 1993) with good finite sampling and density adaptation properties (see Stone 1977, Cleveland 1979, Fan 1992, 1993, Fan and Gijbels 1992, 1996).³³

The local linear method is based on the minimization of the following equation:

$$\min_{\{\Phi_j(x)\}} (\mathbf{Z}_j - \mathbf{X}\Phi_j)' \mathbf{K}^*(\Psi_j) (\mathbf{Z}_j - \mathbf{X}\Phi_j) \quad (\text{A2})$$

where $\Phi_j(\mathbf{x}) = (E(Z_{jt} | \mathbf{x}_t = \mathbf{x}), (\mathbf{D}_j(\mathbf{x}))')'$ is a $((2+np) \times 1)$ vector containing the estimators of

$\alpha_j \mathbf{x}_t$ and of their derivative, $\mathbf{D}_j(\mathbf{x}) = \frac{\partial E(Z_{jt} | \mathbf{x}_t = \mathbf{x})}{\partial \mathbf{x}}$, \mathbf{Z}_j is the $T \times 1$ vector with Z_{jt} in the t^{th} row, \mathbf{X}

is the $T \times (2+np)$ design matrix and $\mathbf{K}^*(\Psi_j)$ with $\Psi_j = \begin{pmatrix} \mathbf{x}_t - \mathbf{x} \\ \mathbf{h}_{jt} \end{pmatrix}$ is a $T \times T$ diagonal matrix with t^{th}

diagonal $K_j \left(\frac{\mathbf{x}_t - \mathbf{x}}{\mathbf{h}_{jt}} \right)$. The solution to this problem is:

$$\hat{\Phi}_j(\mathbf{x}) = (\mathbf{X}' \mathbf{K}^*(\Psi_j) \mathbf{X})^{-1} \mathbf{X}' \mathbf{K}^*(\Psi_j) \mathbf{Z}_j,$$

which is equivalent to the standard generalized least squares (GLS). The conditional first and second moments are:

$$\begin{aligned} E(\hat{\Phi}_j | \mathbf{x}_1, \dots, \mathbf{x}_T) &= (\mathbf{X}' \mathbf{K}^*(\Psi_j) \mathbf{X})^{-1} \mathbf{X}' \mathbf{K}^*(\Psi_j) \mathbf{m} \\ &= \Phi_j + (\mathbf{X}' \mathbf{K}^*(\Psi_j) \mathbf{X})^{-1} \mathbf{X}' \mathbf{K}^*(\Psi_j) \mathbf{r} \end{aligned} \quad (\text{A3})$$

$$\text{Var}(\hat{\Phi}_j | \mathbf{x}_1, \dots, \mathbf{x}_T) = (\mathbf{X}' \mathbf{K}^*(\Psi_j) \mathbf{X})^{-1} (\mathbf{X}' \mathbf{W} \mathbf{X}) (\mathbf{X}' \mathbf{K}^*(\Psi_j) \mathbf{X})^{-1} \quad (\text{A4})$$

where $\mathbf{m} = \{m_j(\mathbf{x}_1), \dots, m_j(\mathbf{x}_T)\}'$, $\mathbf{r} = \mathbf{m} - \mathbf{X}\Phi_j$ be the residual of the local linear regression, and

$\mathbf{W} = \text{diag} \left[K_j^2 \left(\frac{\mathbf{x}_t - \mathbf{x}}{\mathbf{h}_j} \right) \omega_j^*(\mathbf{x}_t) \right]$, with $\omega_j^*(\mathbf{x}_t)$ as a pilot estimator of the conditional variance of the

residuals $\omega_{jt} = \{Z_{jt} - m_j(\mathbf{x}_t)\}^2$.

Method to Obtain the Variable Bandwidths

The choice of the window width is very important in nonparametric estimation. The amount of local information used to construct the average is controlled by the window width \mathbf{h}_j , which determines how many of the \mathbf{x}_t around \mathbf{x} are to be used in forming the local sample size. The estimates are consistent

³³ See, in addition, Masry (1996a, 1996b) for a discussion of uniform almost sure convergence rate and the pointwise asymptotic normality result of the multivariate local linear method.

when allowing the amount of local averaging to become more informative as the sample size increases ($T\mathbf{h}_j \rightarrow \infty$ as $T \rightarrow \infty$), while reducing the neighborhood in which the averaging is formed ($\mathbf{h}_j \rightarrow 0$). When the value of \mathbf{h}_j is small, $\hat{E}(Z_{jt} | \mathbf{x}_t = \mathbf{x})$ is the best local estimator. When $\mathbf{h}_j = 0$ we obtain $\hat{E}(Z_{jt} | \mathbf{x}_t = \mathbf{x}) = Z_{jt}$ whereas as $\mathbf{h}_j \rightarrow \infty$ $\hat{E}(Z_{jt} | \mathbf{x}_t = \mathbf{x}) = \frac{1}{T} \sum_{t=1}^T Z_{jt}$. When the data are over-smoothed (large \mathbf{h}_j), the bias term is large and the variance is small, whereas when the data are under-smoothed the opposite holds. The optimal amount of smoothing minimizes the trade-off between bias and variance (i.e. the mean squared error). Since \mathbf{h}_j tends to zero as T increases, all observations not close to \mathbf{x} are effectively deleted as the sample expands.³⁴

We use the bandwidth for local polynomial regressions proposed by Fan and Gijbels (1995), which estimates the local mean squared errors (MSE). This bandwidth selector combines the notions of the plug-in and the cross-validation method (see Rupert and Wand 1994 for a discussion). In the first step, a pilot statistic called Residual Squares Criterion (RSC) is obtained. The RSC is used to estimate the bias and variance in Equations (A3) and (A4), which are then used to estimate the mean squared error. This statistic is the building block to the bandwidth selection procedure. In the second step, we find a refined bandwidth that minimizes the estimated integrated MSE and use it to fit the local linear regression.

In the first step, the pilot statistic is obtained from:

$$RSC(\mathbf{x}; \mathbf{h}_j) = \hat{\omega}_j^*(\mathbf{x})\{1 + (q+1)V\}, \quad (\text{A5})$$

where $q=1$ for our local linear regression (the order of the polynomial), V is the first diagonal element of the matrix $\mathbf{G}_t^{-1}\mathbf{G}_t^*\mathbf{G}_t^{-1}$, with $\mathbf{G}_t = \mathbf{X}'\mathbf{K}^*(\Psi_j)\mathbf{X}$, $\mathbf{G}_t^* = \mathbf{X}'\mathbf{K}^{*2}(\Psi_j)\mathbf{X}$, and:

$$\hat{\omega}_j^*(\mathbf{x}) = \frac{1}{\text{tr}\{\mathbf{K}^*(\Psi_j) - (\mathbf{X}'\mathbf{K}^*(\Psi_j)\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}^{*2}(\Psi_j)\mathbf{X}\}} \sum_{t=1}^T (Z_{jt} - \hat{Z}_{jt})^2 K_j\left(\frac{\mathbf{x}_t - \mathbf{x}}{\mathbf{h}_j}\right)$$

is the normalized weighted residual sum squares. From Equation (A4), $\text{Var}(\hat{\Phi}_j | \mathbf{x}_1, \dots, \mathbf{x}_T) \approx \hat{\omega}_j^*(\mathbf{x})\mathbf{G}_t^{-1}\mathbf{G}_t^*\mathbf{G}_t^{-1}$.

The bias and variance in Equations (A3) and (A4) are not directly observed as they depend on \mathbf{r} and \mathbf{W} . Fan and Gijbels (1995) show that the bias can be approximated by $(\mathbf{X}'\mathbf{K}^*(\Psi_j)\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}^*(\Psi_j)\tau_j$, where for a given observation t :

$$\Phi_{jq+1}(\mathbf{x}_t - \mathbf{x})^{q+1} + \dots + \Phi_{jq+a}(\mathbf{x}_t - \mathbf{x})^{q+a}$$

are the t^{th} elements of τ_j . The approximate bias and variance to Equations (A3) and (A4) are:

³⁴ Notice that the local linear estimator is evaluated using all data points for each \mathbf{x}_t $t = 1, \dots, T$ for each x_k $k = 1, \dots, K$, where T is the effective sample size used to estimate the VAR after taking into account the lags and K is the number of local points. Thus, the number of evaluations is of order TK for each estimation over the recursive real time samples.

$$\text{Bias} \quad \mathbf{G}_t^{-1} \begin{pmatrix} \Phi_{jq+1} G_{q+1} + \dots + \Phi_{jq+a} G_{q+a} \\ \vdots \\ \Phi_{jq+1} G_{2q+1} + \dots + \Phi_{jq+a} G_{2q+a} \end{pmatrix} \quad (\text{A6})$$

$$\text{Variance} \quad (\mathbf{X}'\mathbf{K}^*(\Psi_j)\mathbf{X})^{-1}(\mathbf{X}'\mathbf{K}^{*2}(\Psi_j)\mathbf{X})(\mathbf{X}'\mathbf{K}^*(\Psi_j)\mathbf{X})^{-1}\hat{\omega}_j^*(x) \quad (\text{A7})$$

Thus, the bias and the variance can be estimated using the estimated parameters $\hat{\Phi}_{jq+1}, \dots, \hat{\Phi}_{jq+a}$ and $\hat{\omega}_j^*(x)$, respectively, which are obtained from fitting a $(q+a)^{\text{th}}$ order polynomial regression estimated with the pilot bandwidth RSC. For our local linear estimation case $q=1$. The MSE of $\hat{\Phi}_{jv}$ (for $v = 0, \dots, q$) is estimated as:

$$M\hat{S}E_{j,p,v}(\mathbf{x}; \mathbf{h}) = \text{bias} + \text{variance} \quad (\text{A8})$$

where the bias is given by the $(v+1)^{\text{th}}$ element of the estimated bias vector (A6), while the variance corresponds to the $(v+1)^{\text{th}}$ diagonal element of the estimated variance matrix (A7). We use this statistic as the building block to the bandwidth selection procedure. Fan and Gijbels (1995) show this to be a consistent estimator of the asymptotic optimal bandwidth. In addition, it is less variable and more efficient than the cross-validation bandwidth rule.

The procedure to estimate $m_j^v(\mathbf{x}_t)$ is as follows. We first fit a polynomial of order $q+a$ (choosing $a=2$ and $q=1$) and use the integrated RSC, $IRSC(h) = \int_{[c,d]} RSC(Z_j; h) dZ_j$ to select the optimal bandwidth for estimating Φ_{jq+1} .³⁵ We obtain estimates $\hat{\Phi}_{jq+1}$, $\hat{\Phi}_{jq+2}$, and $\hat{\sigma}_j^2(x)$. We then find a refined vector bandwidth that minimizes the estimated integrated MSE and use it to fit the local linear vector autoregression ($q=1$):

$$\hat{\mathbf{h}}_{j,q,v}^* = \arg \min_{\mathbf{h}_j} \left\{ \int_{[c,d]} M\hat{S}E_{q,v}(Z_j; \mathbf{h}_j) dZ_j \right\} \quad (\text{A9})$$

In practice, the integrals are implemented as averages over grid points as the estimated curves are evaluated in $\mathbf{x}_r, r = 1, \dots, R_{\text{grid}}$. The grid points are geometric: $\mathbf{h}_j = \mathbf{C}^r \mathbf{h}_{j\min}$, where $\mathbf{h}_{j\min}$ are the first grid points and \mathbf{C} is the grid span. Thus, to minimize the integrated MSE (IMSE) over the interval $[\mathbf{h}_{j\min}, \mathbf{h}_{j\max}]$, we start from $\mathbf{h}_{j\min}$, increase \mathbf{h}_j by the factor \mathbf{C} , and calculate IMSE at these grid points. We stop when the IMSE increases consecutively a couple of times or when $\mathbf{h}_j > \mathbf{h}_{j\max}$. We then choose (A9) as the grid points with the smallest computed IMSE.

³⁵ Fan and Gijbels (1995) show that setting $a = 2$ is computational efficient and not far from \sqrt{T} consistent for a local linear fit.

2. Conditional Variance

Estimation

The conditional variance is estimated using the residual-based method proposed by Fan and Yao (1998), which follows from the procedure above.³⁶ Let $E(\omega_{jt} | \mathbf{x}_t = x) = \overline{\omega}_j(x)$ be the conditional variance of the residuals from $Z_{jt} - m_j(\mathbf{x}_t)$. The local linear estimator of $\overline{\omega}_j(x)$ is $\hat{c}_j = \hat{\overline{\omega}}_j(x)$, which is obtained based on the minimization of the following equation:

$$(\hat{c}_j, \hat{d}_j) = \arg \min_{\{c, d\}} \sum_{t=1}^T \{\hat{\omega}_{jt} - c_j - d_j(\mathbf{x}_t - \mathbf{x})\}^2 K_j \left(\frac{\mathbf{x}_t - \mathbf{x}}{\mathbf{h}_{2j}} \right) \quad (\text{A10})$$

The local linear estimator for the variance also has the nice properties of the local linear conditional mean estimator. In particular, it is also efficient in an asymptotic minimax sense (Fan 1993) and adaptive to the conditional distribution of \mathbf{x}_t (Fan and Gijbels, 1992, Ruppert and Wand, 1994, Hastie and Loader, 1993).

Method to Obtain the Variable Bandwidths

The algorithm to obtain the conditional variance is: a) estimate the local linear regression by minimizing (A2) with bandwidth $\mathbf{h}_j = \mathbf{h}(\mathbf{x}_t; Z_{jt})$ and obtain $\hat{m}_j(\mathbf{x}_t)$; b) compute the squared residuals $\hat{\omega}_{jt} = \{Z_{jt} - \hat{m}_j(\mathbf{x}_t)\}^2$; c) find the conditional variance $\hat{c}_j = \hat{\overline{\omega}}_j(x)$ as the *arg min* in (A10) by applying the bandwidth $\mathbf{h}_{2j} = \mathbf{h}(\mathbf{x}_t; \hat{\omega}_{jt})$.

³⁶ Some of the literature on nonparametric estimation of variance is Yao and Tong (1994), Hall and Carroll (1989), Neumann (1994), Härdle and Tsybakov (1997), Ruppert et al. (1997), and Fan and Yao (1998), among several others.