Drifts and Volatilities under Measurement Error: Assessing Monetary Policy Shocks over the Last Century

Pooyan Amir-Ahmadi Christian Matthes
Mu-Chun Wang *

December 7, 2015

Abstract

How much have the dynamics of U.S. time series changed over the last century? Has the evolution of the Federal Reserve as an institution over the 100 years altered the transmission of monetary policy shocks? To tackle these questions, we build a multivariate time series model with time-varying parameters and stochastic volatility that features measurement errors in observables. We find substantial changes in the structure of the economy. There is also large variation in the impact of monetary policy shocks, but the majority of this variation is driven by changes in exogenous volatility.

JEL Classification: C50, E31, N12
Keywords: Bayesian VAR, Time variation, Measurement error, U.S. monetary policy

*Affiliations: Goethe University Frankfurt (Amir-Ahmadi), Federal Reserve Bank of Richmond (Matthes), and University of Hamburg (Wang).

We would like to thank Luca Benati, Gabriel Fagan, Stefan Gerlach, Jochen Günther, Thomas Lubik, James Nason, Andrew Owens, Giorgio Primiceri, Ricardo Reis, Pierre Sarte, Frank Schorfheide, Felipe Schwartzman, Daniel Tracht, Alexander Wolman as well as three anonymous referees and seminar participants at CREATES (Aarhus), UPF, Banque de France, Kiel Institute for the World Economy, the Bundesbank, the FRBR-UVa Research Jamboree and the Central Bank of Ireland for helpful comments. Parts of this paper were written while Matthes was visiting the Bundesbank, whose hospitality is gratefully acknowledged.

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1 Introduction

We study over 100 years of US economic data on inflation, real output, short-term and long-term nominal interest rates as well as money growth through the lens of a time-varying parameter model to assess how the dynamics of the economy have changed and how the impact of monetary policy shocks has evolved over time.

Our sample covers two World Wars, the Great Depression, the Great Inflation, the Great Moderation, the recent financial crisis and the associated recession, technological revolutions and the founding of the Federal Reserve, so there is ample reason to believe that indeed the dynamics and co-movement of the variables we consider might have changed over time.

To tackle these questions, we have to confront the measurement issues inherent in historical macroeconomic data, as discussed by Romer (1989). Long-run historical time series are usually compiled using a variety of sources for different time periods, so measurement issues (i.e. changes in the quality of the data) are unavoidable. We combine a model of possibly mis-measured historical data with a time-varying parameter VAR with stochastic volatility along the lines of Primiceri (2005) and Cogley & Sargent (2005) to jointly assess the importance of measurement error and time variation in the dynamics of historical macroeconomic time series. Our model of measurement error builds on the measurement error models used by Cogley, Sargent & Surico (2015) and Cogley & Sargent (2014), who introduce measurement error in time series model of inflation, and Schorfheide, Song & Yaron (2014), who model measurement errors in consumption. Romer (1989) emphasized that the large volatility we see in U.S. GDP data before the end of WWII is substantially due to measurement error. To tackle the measurement issue Ritschl, Sarferaz & Uebele (2015) employ a time-varying dynamic factor model for long time series of US economic activity indicators and find similar results. Our approach confirms the findings in Romer (1989). To our knowledge, this is the first paper to explicitly introduce measurement errors in a VAR framework, let alone a VAR with time-varying parameters and stochastic volatility.

To gauge how much the dynamics of the US economy have changed during our sample period, we first calculate four different measures of time variation implied by our multivariate time-varying parameter model: variation in persistence, volatility, long-run averages, and co-movement. We find that along all these lines there is substantial variation, even after taking into account measurement error. The correlation structure between our variables of interest has changed dramatically over time.

One of the most pressing questions in macroeconomics is that of the effects of unanticipated changes in policy instruments, particularly for the case of monetary policy (Christiano, Eichenbaum & Evans (1999)). We want to analyze both how the impact of unexpected movements in monetary policy has evolved across monetary regimes as well as how much those unexpected movements have contributed to overall volatility in the economy. Defining a monetary policy shock for post-WWII data is straightforward: many economists tend to think of the Federal Reserve after WWII as choosing a path for the short-term interest rate. If we were to have a model (or equation) for the short-term nominal interest rate, we could define the monetary policy shock as the residual after properly accounting for movements in all variables deemed relevant for the setting of the short-term interest rate. The same would hold true if the Federal Reserve consistently used changes in money growth as its policy instrument. However, in our sample we are faced with the difficulty that there has been no consistent conduct of monetary policy. For example, the Federal Reserve targeted monetary aggregates as recently as the beginning of Paul Volcker's tenure (Hetzel (2008)). We thus aim to identify monetary policy shocks not as identified shocks associated with a certain equation or variable, but rather by their impact on the economy through the use of sign restrictions. These sign restrictions identify the set of impulse responses consistent with the sign restrictions.

We find that effects of an 'average' (one standard deviation) shock have changed substantially. These changes could be driven by changes in both the average size of a shock (changes in the standard deviation) and the dynamic responses to shocks. We disentangle these possible causes and find that the size of the innovation is the major driver of the changes in the effects of a monetary policy innovation.

Our work is related to the growing literature on time-varying VARs, most notably Cogley & Sargent (2002), who were the first to use this class of models, and Primiceri (2005), who first identified monetary policy shocks in this class of models. Our finding of surprising stability of the effects of a monetary policy shock (once we condition on the size of the shock) has precedents in the literature: using a recursive identification scheme, Primiceri (2005) finds that there is not much time variation in impulse responses to shocks of a given size in post-WWII data. Sims & Zha (2006) argue that most of the time variation in post-WWII US time series is driven by changes in the volatility of innovations. Canova & Gambetti (2009) argue that the transmission of monetary policy shocks has been relatively stable over the post-WWII period. Using a simple split sample analysis and not considering measurement error, Sims (1999) argues that the response to monetary policy shocks has not changed dramatically between the inter-war and post-war periods. In line with our findings, Amir-Ahmadi & Ritschl (2013) use a factor-augmented VAR for the interwar period and find effects of monetary policy shocks on real activity that are comparable with the effects in post-WWII data. In comparison to those papers, our paper combines a long-run historical perspective, a careful treatment of measurement errors and time variation in parameters and volatilities.

Analysis that focus only on a sample split around 1980 will have difficulty identifying structural changes at other points in time. Focusing only on pre- or post WWII data would not allow us to compare the monetary transmission mechanism for the Federal Reserve's entire history. Finally, our analysis shows that different statistics (such as correlations, impulse responses, or forecasts) change substantially at different points in time, making an a priori choice of subsamples hard to defend. We thus find that building one model for the entire sample that explicitly tackles the issues of time variation, stochastic volatility and measurement error seems the most straightforward way to answer the questions we are interested in.

2 The Model

We are interested in modeling the dynamics of the following vector of observables:

$$\tilde{y}_t = \begin{pmatrix} \Delta gdp_t & \pi_t & i^s_t & spread_t & \Delta money_t \end{pmatrix}' \tag{1}$$

where Δgdp_t is the one-year difference in the log of real output, π_t is the one-year inflation rate, i_t^s is a short-term nominal interest rate, $spread_t$ is the spread between a long-term nominal interest rate and our short-term nominal interest rate, and finally, $\Delta money_t$ is the one year difference in the log of a monetary aggregate. Our benchmark monetary aggregate is the monetary base.

Christina Romer's work (see, for example, Romer (1986) and Romer (1989)) has brought the measurement issues associated with historical macroeconomic data front and center. We tackle measurement issues by building a model of measured data that allows for relatively general stochastic processes for the measurement errors. We use over 100 years of data and thus have to confront the possibility of not only measurement error, but also changes in the measurement process (as suggested by the work of Christina Romer). To do this, we extend the framework of Cogley, Sargent & Surico (2015). They build a univariate model for inflation, allowing for an autocorrelated measurement error process whose parameters can change when the underlying data source changes. Because this framework allows for time variation in the parameters of the measurement error only at those known break dates, we can, just as Cogley, Sargent & Surico (2015), separately identify the changes in the measurement error process and the changes in the process of the underlying 'true' data. We model the measurement error processes as independent across variables, while allowing for autocorrelation in each of the errors. We assume that the observed data vector $\tilde{y}_t$ is a function of a measurement error vector m_t and the 'true' (unobserved) data y_t:[1]

$$\tilde{y}_t = y_t + M(L)m_t \tag{2}$$

$M(L)$ is a diagonal matrix where the i-th diagonal element $M_i(L)$ is a polynomial in the lag operator whose role we will describe in detail below. Our model for the 'true' unobserved data y_t follows Primiceri (2005)[2] and models y_t as a time-varying VAR:

$$y_t = c_t + \sum_{j=1}^{L} A_{j,t} y_{t-j} + e_t \tag{3}$$

where the intercepts c_t, the autoregressive matrices $A_{j,t}$, and the covariance matrix Ω_t of e_t are allowed to vary over time. We set the number of lags $L = 2$. To be able to parsimoniously describe the dynamics of our model, we define $X_t' \equiv I \otimes (1, y_{t-1}' ..., y_{t-L}')$ and rewrite (3) in the following state space form[3]:

$$y_t = X_t' \theta_t + e_t \tag{4}$$

$$\theta_t = \theta_{t-1} + u_t \tag{5}$$

The observation equation (4) is a more compact expression for (3). The state equation (5) describes the law of motion for the intercepts and autoregressive matrices. The covariance matrix of the innovations in equation (4) is modeled after Primiceri (2005):

$$e_t = \Lambda_t^{-1} \Sigma_t \varepsilon_t \tag{6}$$

Λ_t is a lower triangular matrix with ones on the main diagonal and representative non-fixed element λ_t^i. Σ_t is a diagonal matrix with represen-

[1]While Christina Romer's work has emphasized mismeasurement issues in output and unemployment, Cogley, Sargent & Surico (2015) and Cogley & Sargent (2014) have put their focus on inflation. It might be less clear why we also assume measurement error for interest rates. We do so because we combine different sources of interest rate data.

[2]The modeling assumptions we make for this part of the model are widely used in empirical macroeconomics. An overview of the methods used and assumptions made in this literature is given by Koop & Korobilis (2010).

[3]I denotes the identity matrix.

tative non-fixed element σ_t^j. The dynamics of the non-fixed elements of Λ_t and Σ_t are given by:

$$\lambda_t^i = \lambda_{t-1}^i + \zeta_t^i \tag{7}$$

$$\log \sigma_t^j = \log \sigma_{t-1}^j + \eta_t^j \tag{8}$$

To conclude the specification of the VAR for the true data, we need to make distributional assumptions on the innovations ε_t, u_t, η_t and ζ_t, where η_t and ζ_t are vectors of the corresponding scalar innovations in the elements of Σ_t and Λ_t. We assume that all these innovations are normally distributed with covariance matrix V, which we, following Primiceri (2005), restrict as follows:

$$V = Var\left[\begin{pmatrix} \varepsilon_t \\ u_t \\ \zeta_t \\ \eta_t \end{pmatrix}\right] = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix} \tag{9}$$

S is further restricted to be block diagonal, which simplifies inference. We estimate this model using the Gibbs sampling algorithm described in Del Negro & Primiceri (2013)[4] augmented with additional Gibbs steps for the drawing of the measurement errors and the parameters of the measurement error process. A summary of this algorithm can be found in the online appendix, where we also describe the priors, whose specification is standard in the literature.

Some variables that we use in our model are measured in year-over-year rates[5] (GDP growth, inflation and money growth). To capture the fact that it might in fact be the *levels* of these variables that are measured with error (along the lines of Schorfheide et al. (2014)), we introduce the

[4]We use 100000 draws.

[5]Specifically we calculate the year-over-year rates as $100 * ln(\frac{y_t}{y_{t-4}})$.

lag polynomials $M_i(L)$, which are the diagonal elements of the diagonal matrix $M(L)$. For variables that we assume are directly measured with error, we set $M_i(L) = 1$, so that the measurement error operates directly on this variable. If, instead, the measurement error for variable i is associated with the level of a variable that we include in *annual* growth rates in the VAR, we set $M_i(L) = 1 - L^4$. We use this specification for output growth, inflation and money growth in the benchmark model. The benchmark model uses $M_i(L) = 1$ for interest rates and the interest rate spread.[6]In the online appendix we study a specification of the model in which we set $M_i(L) = 1 \; \forall i$, in line with the model used for inflation by Cogley, Sargent & Surico (2015). We call that specification our growth rate specification.

Each element i of m_t follows an AR(1) process:

$$m_t^i = \rho_j^i m_{t-1}^i + \sigma_j^i \varepsilon_t^{m,i} \tag{10}$$

We allow the measurement error process for each variable to change whenever a data source changes over our sample.[7] The data source for each variable is indexed by j. The coefficients for each measurement error process can thus change at points in time that can be different for each variable. Just as Cogley, Sargent & Surico (2015), we assume that the latest source corresponds to data measured without error. This helps identify measurement error process parameters and is also (at least implicitly) the assumption underlying most of Christina Romer's work such as Romer (1989). The innovations $\varepsilon^{m,i}$ are Gaussian with mean 0 and

[6]Our approach can also easily accommodate other situations such as measurement error in levels when the observed data is in one period growth rates ($M_i(L) = 1 - L$) or when the measurement error is in levels, but we observe an N-period average ($M_i(L) = 1/N \sum_{n=1}^{N} L^{n-1}$).

[7]The exact timing of the break dates is described later. We also add one additional break date for GDP, but this is inconsequential for our results.

variance 1, and are independent over time and across i, as well as independent of all other innovations in the model.

The break points in the measurement error processes for the different variables are given in table 1. The dates are motivated by changes in the sources of the data, which are described in detail in the online appendix. For robustness, we have included an additional break data for real GDP growth in 1930 to account for possible measurement issues during the great depression. It turns out that adding this break date does not alter results in any meaningful way. After the last break date for each variable, that variable is assumed to be *measured without error*.

	Δgdp_t	π_t	i_t^s	$spread_t$	$\Delta money_t$
1^{st} break date	1930:Q1	1947:Q1	1920:Q1	1920:Q1	1918:Q1
2^{nd} break date	1947:Q1	–	–	–	1936:Q1
3^{rd} break date	–	–	–	–	1959:Q1

Table 1: Break dates (if applicable) for parameters in the measurement error processes for all variables

3 Data

In this section we briefly describe our data (plotted in Figure 9) and discuss the estimated measurement errors. Figure 9 focuses on the part of the sample where measurement error is present - a plot of the entire data set can be found in the online appendix. We use quarterly U.S. data covering the period from the first quarter 1876 to the second quarter of 2011. This time span is of specific interest as it covers the pre-Fed period as well as all chairmanships prior to Janet Yellen, which represent potentially different monetary policy regimes. Furthermore, the period covers 29 recessions, as measured by the NBER, of different duration and depth. In our application, we will use the first part of our constructed sample (up until 1913) to inform the prior for our VAR. The estimation of our time-varying VAR starts in 1914. The earlier part of our sample comes from Balke & Gordon (1986), whereas most of the post-WWII data is taken from the FRED database at the St. Louis Fed.

3.1 Measurement Errors in the Data

Figure 9 plots the observed data as well as the estimated 'true' data until 1960 (in percentages). To focus on the role of measurement error in the early part of the sample, we do not plot the more 'standard' part of the sample, which consists of standard US macroeconomic data after 1960. The online appendix shows a plot of the entire dataset. Along the lines of Romer (1989), GDP growth is identified by our estimation as having significant amounts of measurement error during some periods before WWII.

Overall, our estimated median 'true' GDP growth exhibits lower volatility than observed GDP growth up to 1947. Most notably, during the period from 1930 to 1947 observed GDP growth is almost four times more volatile than our estimated 'true' GDP growth, which explicitly takes measurement errors into account. While still clearly visible, the severity and magnitude of the Great Depression in our 'true' estimated GDP growth is substantially smaller and the corresponding expansion during WWII is estimated to be more shallow. It is worth noting that our estimation procedure does not just automatically smooth out movements in the observed series - the large movements in observed GDP before 1930 are estimated to be movements in the underlying true GDP growth process. Measurement error plays a substantially smaller role for all other variables.

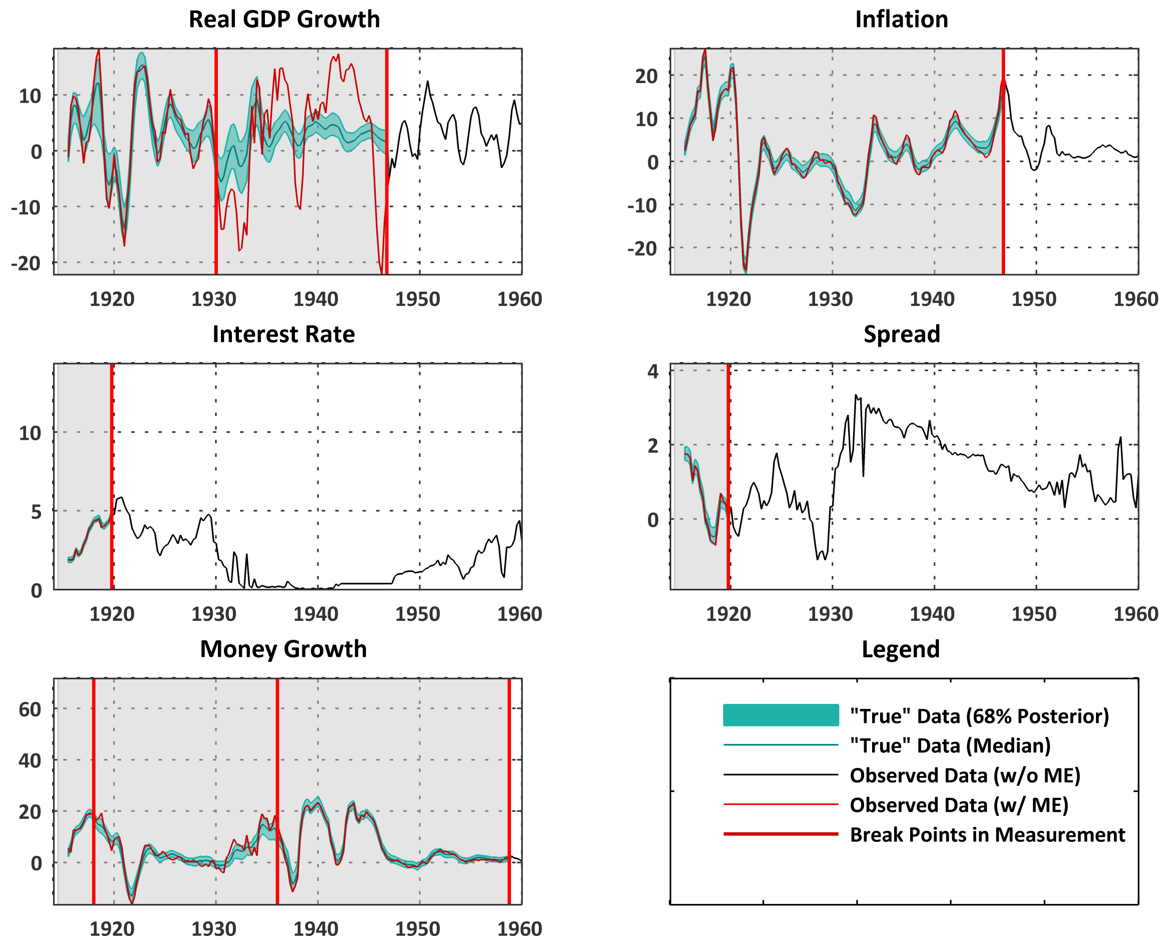

Figure 1: Data until 1960

To assess how important measurement error is during different periods, we focus on real GDP growth where we suspect measurement error is most important, as highlighted by Romer (1989). Table 2 shows the standard deviation of observed and estimated 'true' real GDP growth for the period before the end of WWII, both in absolute terms and relative to post-WWII volatility. For simplicity, we focus on the median of the estimated GDP growth series.

The standard deviation of both observed and estimated GDP growth from 1915 to 1946 is higher than the standard deviation of what we assume to be perfectly observed GDP growth after WWII (1947-2006). We find that measurement error drives a substantial fraction of the GDP growth volatility between 1930 and 1946 and estimated 'true' GDP growth is quite smooth. It is important to note that for the period $1930-1946$ we do *not* find 'true' GDP growth to be less volatile because observed GDP growth is less volatile - actually the standard deviation of observed GDP growth is higher. Rather, our model uses the VAR structure of the 'true' data (and thus the relation between the variables in our dataset) to conclude that 'true' GDP growth must have been smoother during that period. While we do find more volatile 'true' GDP growth before the end of WWII relative to the immediate post-war period, the differences relative to post-WWII data are substantially smaller than what one would expect from observed data - in particular, in spite of the Great Depression and WWII, true GDP growth between 1915 and 1946 was only 87 % more volatile than after WWII, a far cry from the 268% difference implied by observed data. These findings are qualitatively in line with Romer (1986), who argues that the postwar stabilization was substantially less significant than generally believed.

	Estimated	Observed	$\frac{\text{Estimated}}{\text{Post WWII}}$	$\frac{\text{Observed}}{\text{Post WWII}}$
1915-1946	4.8	9.6	1.9	3.7
1915-1929	6.4	7.7	2.5	2.9
1930-1946	2.9	11.0	1.1	4.2
1947-2006	2.6	2.6	1	1

Table 2: Standard deviation of observed and estimated real GDP growth for different periods

4 Results

In this section we describe how the dynamics of our estimated VAR model have changed over time. The goals of this section are two-fold: we are not only interested in these statistics in their own right, but also want to know if those changes will be reflected in changes in the impulse responses that we describe later.

4.1 Examining Stochastic Volatility

First, we study the estimated volatilities of the innovations hitting our model. We focus on the square roots of the diagonal elements of $\Omega_t = Var(e_t)$, which incorporate both time variation in Σ_t and Λ_t.

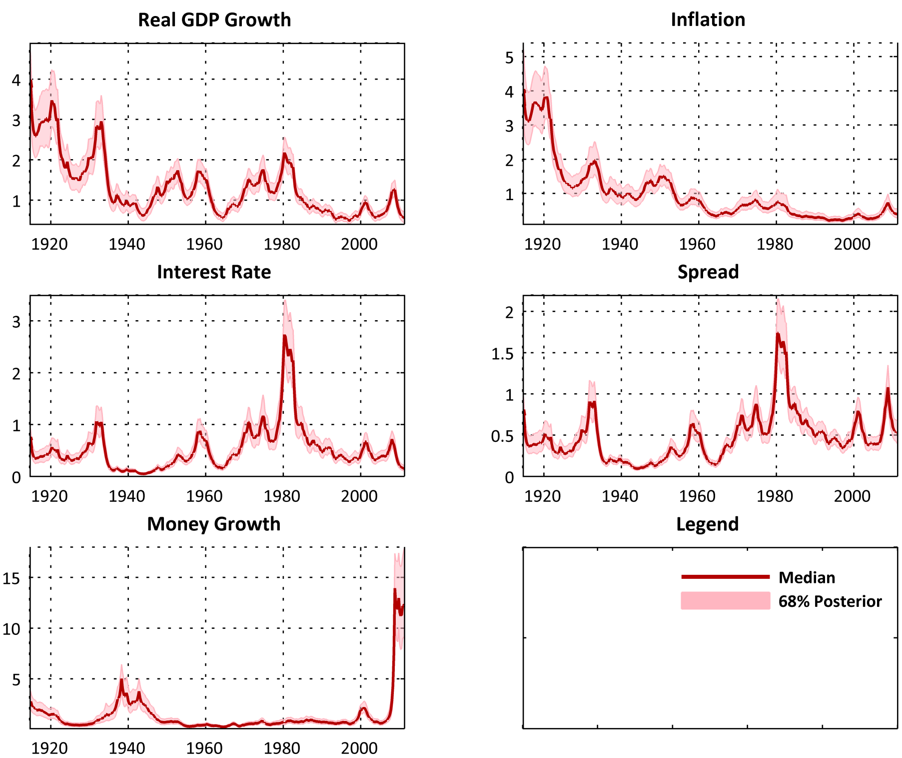

Figure 2: Volatility of reduced form residuals

Figure 2 plots the median as well as 16th and 86th percentile bands of these time-varying standard deviations of the reduced-form residuals. The residuals associated with real GDP growth, inflation and money growth are substantially more volatile during the first part of our sample, whereas the residuals of short-term interest rates and the spread are more volatile after World War II, in particular around the Volcker disinflation of 1980. The residual in money growth, on the other hand, does not have a substantially larger volatility around the Volcker disinflation. One important take-away from this exercise is that relative to the decrease in volatility of real GDP and inflation after the Great Depression and World War II, the so-called "Great Moderation" is almost invisible in the estimated volatilities because the overall level of volatility is so much higher in the earlier part of the sample.[8]

The volatility of the reduced form error in the equation for the spread shows a discrete jump in 1980. Interestingly, while average volatility in that error has come down in the 1980s and 1990s, the levels remain elevated relative to pre-1980 values. Since the reduced form residuals are the one-quarter ahead forecast errors, our model implies that one-quarter ahead forecasts of the slope of the yield curve have thus become less precise since 1980, an interesting hypothesis for future work.

4.2 Time t Approximations to Moments of Forecasts

To analyze the estimated time variation further, we ask what the first and second moments of our observables would be if the dynamics of the observables were governed by parameter estimates that are fixed at the level estimated at one particular time t[9]. Since we do not impose stationarity on our VAR, we cannot compute the unconditional moments under the assumption that the time t parameter estimates do not change in the future. Instead, we compute time t moments for different forecast horizons, which do not require the (smoothed) time t estimates of the companion form matrix of the VAR having all eigenvalues (except for the

[8]The "Great Moderation" refers to decreases of volatility in observables, not necessarily residuals, but it seems natural to expect part of this decrease to be reflected in residuals with smaller variance.

[9]Cogley & Sargent (2005) have used this approach.

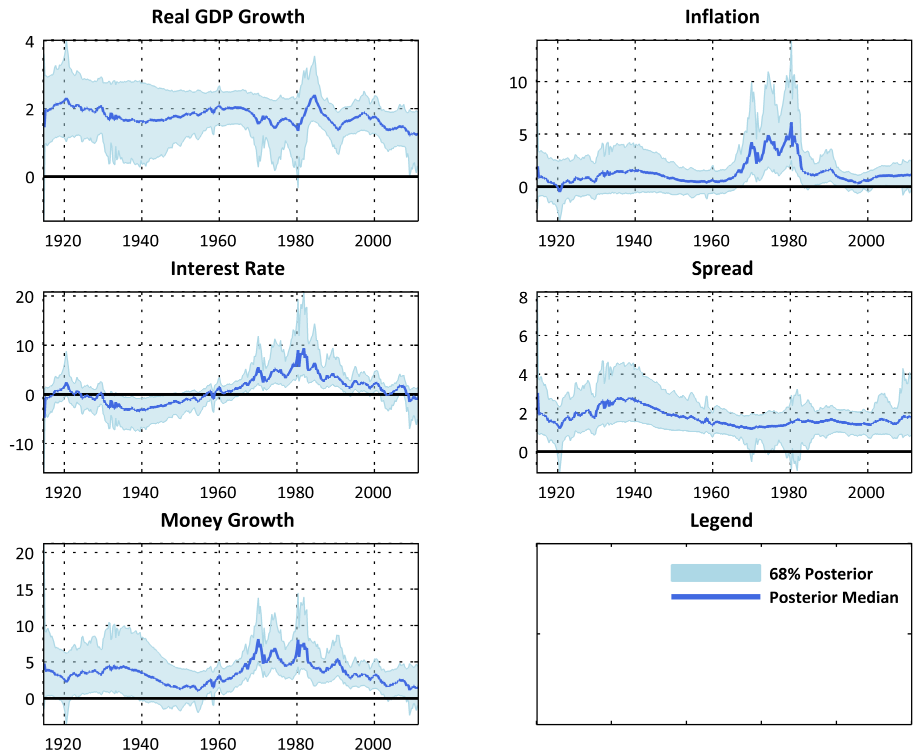

Figure 3: Evolving forecast means: 20 years ahead

eigenvalue associated with the intercept) being less than 1 in absolute value. Following the suggestion of Sims (2001) and the work by Cogley & Sargent (2005), we calculate summary statistics that help us understand how the economy has changed over time using *all* available information (since we use smoothed estimates).

Figure 3 plots the posterior medians and 16th and 86th percentile bands of the evolution of these forecast means at the 20-years-ahead horizon. A substantial part of the time variation is actually in the uncertainty surrounding the forecast means rather than in the median, which does not move too much for long periods of time for the observables we consider.

The period from 1920 to 1940 (which encompasses the Great Depression) is represented in Figure 3 as a time of substantial uncertainty surrounding long-run values, but it is (maybe surprisingly) not associated with substantial movement in the median of the forecasts. Our model thus attributes a substantial part of the Great Depression to temporary changes in volatilities. Benati & Lubik (2014) have a similar finding using inventory and sales data.[10]

The 1970s instead are viewed by our model as a time in which the long-run outlook was quite bleak in terms of GDP growth and inflation.

The Volcker disinflation around 1980 is seen as a major structural break in our model. Average forecasted inflation dropped dramatically, average forecasted output growth increased by 1 percent in annual terms, and the uncertainty surrounding these long-run-forecasts shrank. The recent financial crisis does not dramatically manifest itself in these long-run averages.

We use the h-step ahead forecast variance $Var_t[y_{t+h}]$ [11] to construct time t approximations of the forecast correlations between our observables, which are depicted in Figure 4. For the sake of brevity, we focus on the 5 year horizon for these plots. Other horizons are qualitatively similar. There is substantial time variation in these correlations. The error bands are in general quite wide. A substantial number of these correlations feature large movements in the 1970s and then a structural break at the time of the Volcker disinflation. Starting with the output growth/inflation correlation, we see that inflation and output growth are substantially negatively correlated when looking at the median correlation, but that this correlation is significantly different from 0 only in the 1940s. A similar pattern can be observed for the output growth/interest rate relationship. The 68 percent error bands for the output growth/spread

[10] The estimates are based on all sample information. Out-of-sample forecasts using only information up to that time period would presumably look quite different.

[11] Lütkepohl (2010) describes how to construct this variance.

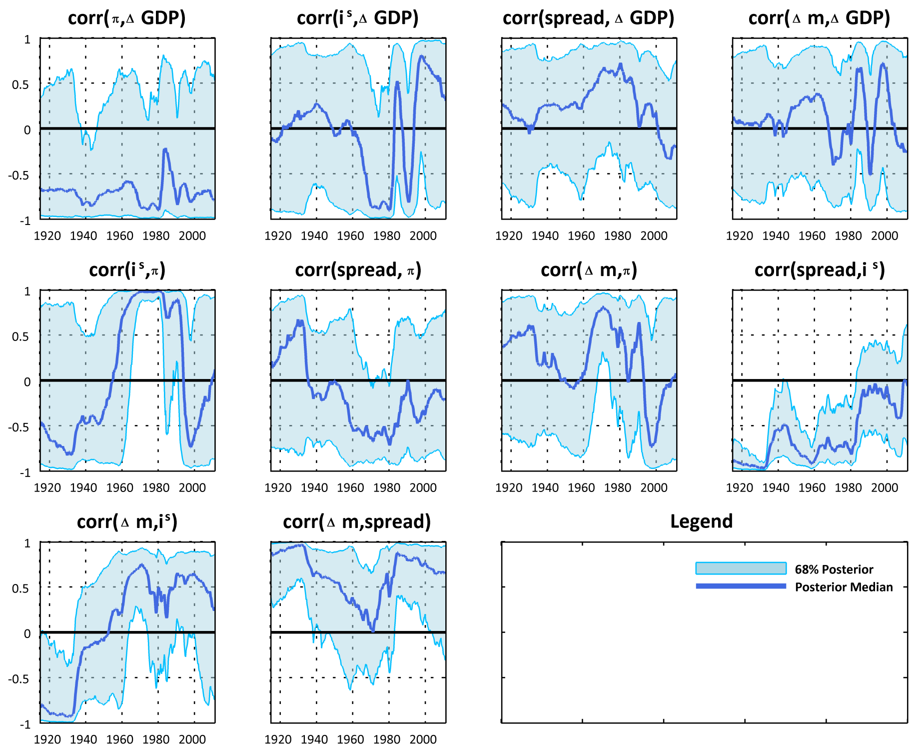

Figure 4: Forecast correlations: 5 years ahead

correlation contain 0 for the entire sample. Note though that the median correlation decreases substantially in magnitude after 1980. The mid-1980s have been identified before as a point in time after which yield curve information does not carry much information for forecasting output growth.[12] Inflation and interest rates have not been significantly correlated until 1960 (the "Gibson Paradox" studied by Cogley, Surico & Sargent (2012)). The correlation then grew throughout the 1960s and was close to 1 during the 1970s. Revisiting the by now common theme, the correlation falls dramatically with the disinflation of the early 1980s[13].

[12] Wheelock & Wohar (2009) state that "Several studies find that the spread has been less useful for forecasting output growth since the mid-1980s, at least for the United States."

[13] To see why a 0 correlation between inflation and the nominal interest rate may be surprising, remember the Fisher equation in its approximate linear form:

$$i_t^s = r_t + E_t \pi_{t+1} \tag{11}$$

If we think about the real interest rate r_t being roughly constant in the distant future then this equation tells us that in the long run short-term interest rates and inflation should move one-for-one. In our model we can not subtract our inflation measure from

A possible explanation for the disappearance of a significant correlation could be that during periods of low correlation between inflation and nominal interest rates inflation expectations are ‘well-anchored’ in that they do not move much in response to movements in variables at the time when the forecast is made. Inspecting our long-run forecast of inflation, we do indeed see little movement in forecasted inflation during times of low correlation between forecasted inflation and forecasted short-term interest rates.

Inflation and money growth are not significantly correlated before the mid-1960s, when the correlation becomes positive. The strength of this correlation disappears immediately with the beginning of the Volcker chairmanship and the associated disinflation. This again points to a positive relationship at high levels of inflation, but not at the substantially lower levels we have observed since the 1980s.

We see a substantially negative relationship between the short-term interest rate and the spread before 1980. This correlation has since become much closer to 0, meaning forecasted long-run movements in the short rate do not feed (linearly) into the slope of the yield curve. This might have implications for monetary policy - policymakers hope to influence long-term interest rates by moving the short-term interest rate. These correlations, however, are not conditional on specific shocks hitting the economy.

---

our measure of the short-term nominal interest rate to get a measure of the ex-post real rate because our short-term interest rate is a three month (annualized) interest rate, whereas we use an annual inflation measure. In terms of long-run forecasts, the difference between an annual interest rate and an annualized 3 month interest rate for a safe asset like we consider should be small. Also, we plot the correlation between inflation and the nominal interest rate 20 years in the future, whereas the Fisher equation would call for the correlation between the nominal interest rate in 20 years and the inflation rate in 20 years and one quarter. Given our long forecast horizon this seems inconsequential.

Finally, the correlation between money growth and the spread has moved into positive territory after 1980. Only at the very end of the sample do these correlations move toward 0 again. Taken at face value, this implies that from 1980 to the early 2000s money growth could have been useful in predicting movements of the yield curve.

The correlations described in this section share common themes: substantial changes around 1980 and correlations that are larger in absolute value when some of the time series themselves are relatively large (such as the inflation/GDP growth correlation). This points to substantial non-linearity in reduced-form Phillips curve relationships, for example.

It is worth pointing out that even though 1980 is identified as a break point in many correlations, there are other break points, many of them in the early 1960s. We interpret this as evidence that a simple split sample analysis using pre- and post-Volcker data is bound to miss interesting aspects of time variation in the economy. Furthermore, many analyses using post-WWII US data only use data starting in the late 1950s or early 1960s (Primiceri (2005) and Sims & Zha (2006) are two prime examples) and will thus not be able to detect those changes.

4.3 Impulse Responses to a Monetary Shock Over the Last Century

Having characterized the substantial changes in the reduced form dynamics of US time series over the last century, we now turn to the question of the effects of monetary policy on the economy and changes of those effects over time.

This section will first describe the impact of a one standard-deviation monetary policy shock on the economy. Then we will examine the relative importance of fluctuations in volatility when compared to changes in the dynamic response to shocks.

The following assumption summarizes our sign restrictions:

Assumption 1: A *monetary policy impulse vector* at time t is an impulse vector a_t, so that the impulse responses to a_t of the price level, the level of output and money growth are not positive and the impulse responses for the short term interest rate are not negative, all at horizons $k = 0, \ldots, K$.[14]

Our benchmark specification does not restrict the impulse responses of the spread. Setting $K = 2$, we impose the sign restrictions at each point in time for the specified contemporaneous responses and for the first and second quarters.[15] This is in line with Uhlig (2005), who uses five months in a monthly model, and Benati (2010), who imposes the restrictions on impact and for the two following quarters. In contrast to the benchmark case in Uhlig (2005), we do restrict the response of output not to react positively following a contractionary monetary policy shock. Most theoretical macroeconomic models feature meaningful output responses to monetary policy shocks, a feature that we use to guide our identification restrictions (see Canova & Paustian (2011) for an introduction to this approach). The candidate time t impulse vector a_t is given by

$$a_t = \Lambda_{t|T}^{-1} \Sigma_{t|T} \alpha_t \tag{12}$$

where α_t is a column vector of conformable size drawn from the unit sphere of norm 1, which we vary across draws to capture the uncertainty

[14] Imposing the restrictions on inflation and output growth, instead of the price level and the level of output, leads to quantitatively very similar results.

[15] Our results are robust to restricting the response of output growth and inflation instead of the level of output and the price level.

implied by the sign restrictions. To compute one draw of the impulse response vector, we simulate data from our model under two scenarios: one in which all random innovations are drawn from their estimated distributions, and another where all innovations are drawn from their estimated distribution except for one time period, where we impose a monetary shock of a given size.

This approach builds on Uhlig (2005), Faust (1998), Canova & Nicolo (2002), and Canova & Gambetti (2009). Additional details regarding implementation and normalization are provided in the online appendix. The impulse responses we show follow Canova & Gambetti (2009) and Benati & Mumtaz (2007) and take into account *all sources of uncertainty* in our model, including the uncertainty associated with the rotation α_t (which is not point-identified) and the uncertainty associated with parameters changing in the future - we calculate generalized impulse responses along the lines of Koop, Pesaran & Potter (1996).

Using sign restrictions to identify a monetary policy frees us from having to make an assumption about what variable is used as the monetary policy instrument. With these sign restrictions, we aim to capture the effects of an unanticipated monetary shock for the different policy regimes in place throughout our sample. Before the Great Depression, the Fed adhered to the Real Bills doctrine. While this represented a policy regime that a-priori could be viewed as featuring a different *systematic* part of monetary policy, there does not seem to be ample reason to think that the sign restrictions we use were not valid during that period. Even though the Gold Standard was in place during that time, Bernanke (2013) identifies it as not binding. After the Great Depression the Federal Reserve was substantially influenced by the Treasury until the 1951

Fed-Treasury accord. This is another episode where the systematic part of monetary policy could be different from other periods, but we do not have substantial reason a-priori to doubt that unexpected effects of monetary policy followed our identifying assumptions during that period. Nonetheless, we can not rule out that our identification restrictions are too strong in that a monetary policy shock might not have the effects we ascribe to it during certain periods. To give an example, Lubik & Schorfheide (2003) show that in a New Keynesian model under indeterminacy, some (but not all) equilibria display dynamics where inflation rises with a positive (i.e. contractionary) monetary policy shock. Incorporating this kind of information, while at the same time not giving up on our reasonable identifying assumptions for most of the sample, would force us to make substantially stronger assumptions such as when exactly those dynamics are in place and what exactly the identification assumptions in those periods are. We feel that the associated costs outweigh the benefits given that our assumptions seem standard for most of our sample.

4.3.1 A Historical Assessment of Consequences of Monetary Policy Shocks

In a model with time-varying parameters, we could compute impulse response paths for each variable at each point in the sample. To keep this overwhelming amount of information manageable, we organize our discussion by dividing our sample into 7 time periods that together span our entire sample and each stand for a succinct time period in the Federal Reserve's history. We borrow these time periods from the Federal Reserve itself, which used the same time-line to summarize its history during the celebration of its centennial.[16] For each of those 7 periods we calculate generalized impulse response functions that take into account all sources of uncertainty in our model. We generate impulse response paths to one standard deviation shocks for every time period t and all posterior draws and then compute the average response for each of the seven periods, which can be seen in figure (5). Since the standard deviations of shocks in our model change over time, the responses we show in this section are best interpreted as the response to an average sized shock during each time period, where the average size of the shock can change across time periods. The black vertical line displays the horizon until which the sign restrictions are imposed. Even though we impose sign restrictions on the level of output and the price level, we plot the impulse responses in terms of inflation and output growth. First focusing on possible policy instruments of the Federal Reserve, we see that the median responses of the interest rate and the money base vary substantially over regimes, both in their median responses and in the variability around those responses. The variability of the money growth response has decreased monotonically over the different time periods for each response horizon, whereas the response of the nominal rate increased in variability before it started decreasing. The median response of inflation decreases until 1982, but then increases again. The median response of the nominal rate has decreased somewhat over time for each horizon, but those changes are small compared to the variability implied by the error bands. The median response of money growth has decreased more substantially over time. Our main takeaway here is that the impact of monetary policy shocks on policy instruments of the Fed has changed

[16] http://www.federalreservehistory.org/Events

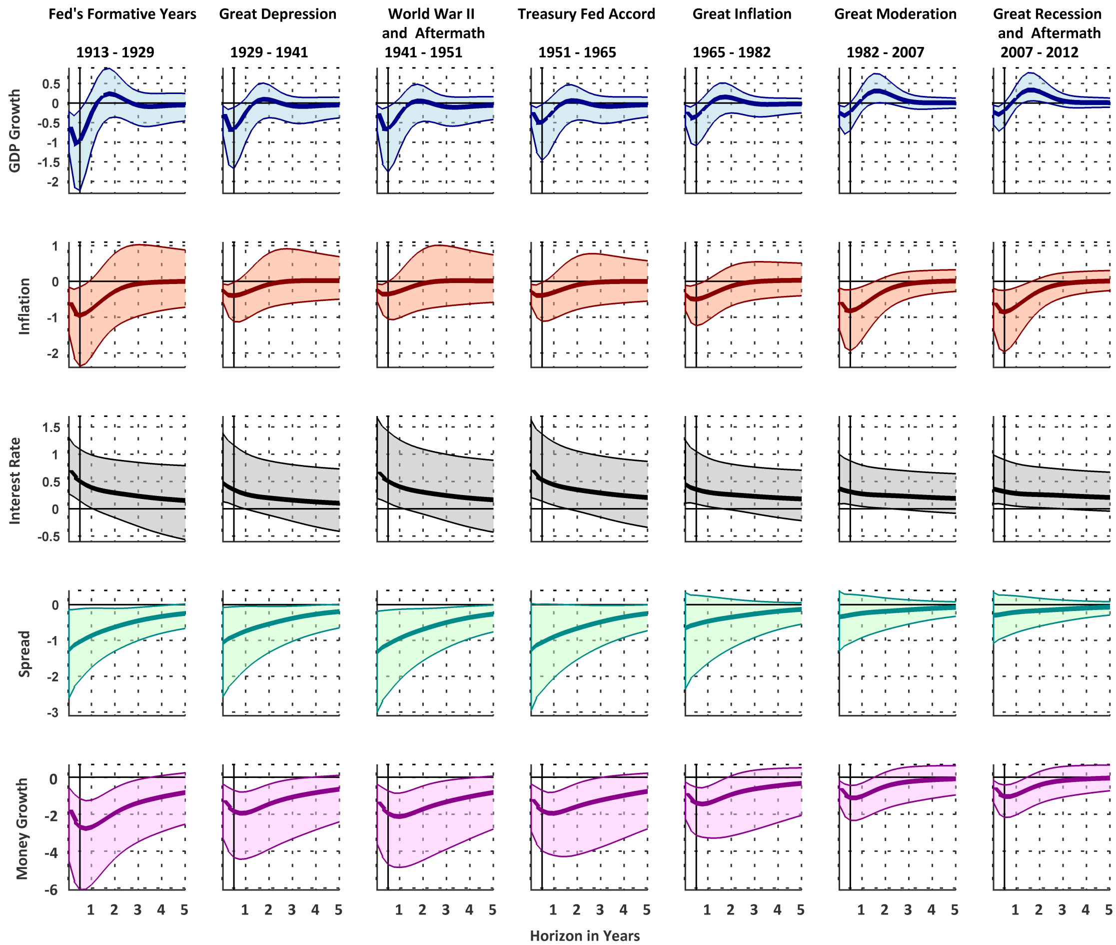

Figure 5: One standard deviation impulse responses across time. The solid bold line is the median response, the shaded area represents the 68 % posterior probability bands centered at the median.

substantially over time, especially when it comes to the uncertainty surrounding the impact.

Before the Great Inflation period, monetary policy shocks had a significantly negative effect on the spread. While the median impact of monetary policy shocks on the spread has remained negative throughout, the magnitude of the response has decreased over time for every horizon and the error bands include 0 for every horizon starting with the Great Inflation period.

The response of GDP growth over time becomes more muted as well as substantially less uncertain. This response is remarkably stable between

1929 and 1951 (or, to almost the same extent, until 1965). The median response starts to become smaller in magnitude and less uncertain after the Fed Treasury Accord that gave the Fed substantially independence.[17] This pattern is more pronounced with our alternative specification of measurement error (see the online appendix for the corresponding figures). However, it is important not to lose sight of the big picture: the patterns of impulse responses to monetary policy shocks have remained surprisingly stable throughout our sample.

Given that we set-identify the responses to a monetary policy shock (that is, even if we knew the reduced form parameters with certainty, we could still not exactly pin down the impulse responses), readers might be interested in the amount of uncertainty coming only from the partial identification of the impulse response function. Following Moon, Granziera & Schorfheide (2013) we report in figure (6) the full set of identified impulse responses conditional on the posterior mean estimates of all reduced form parameters (so that reduced form parameter uncertainty plays no role when computing these error bands) along with the standard 90% posterior error bands that take into account parameter uncertainty and the partial identification of the impulse responses. As noted by Moon et al. (2013), either of these bands could be wider than the other. Next, we turn to isolating the driving force behind these changes - are the changes in the volatility of monetary policy shocks the reason behind the more muted real effects or did the transmission of monetary policy shocks of a given size change?

[17]The Fed-Treasury Accord removed the Fed's obligation to maintain a low-interest rate peg on government securities.

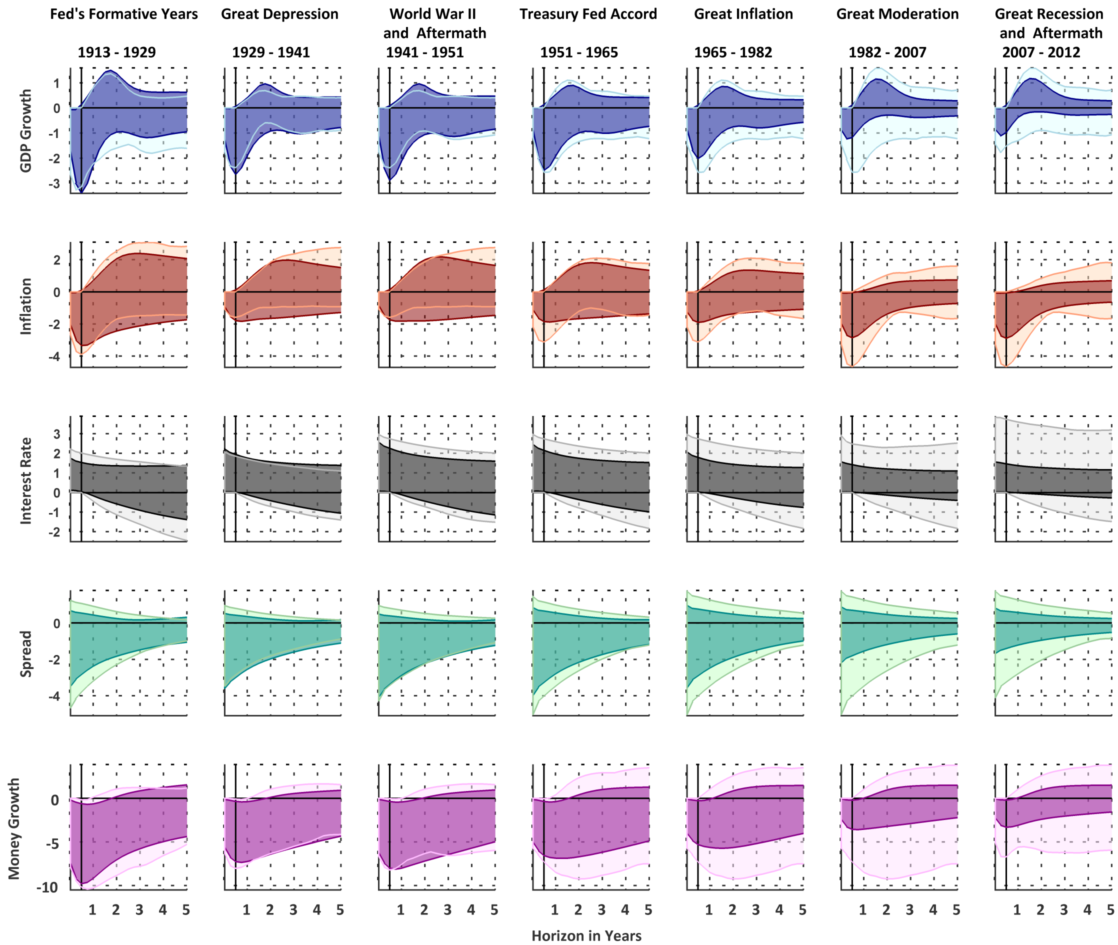

Figure 6: Full identified set at the posterior mean of one standard deviation impulse responses (lighter shaded area) along with the 90% posterior bands (darker shaded area).

4.3.2 Characterizing the Evolution of the Monetary Transmission Mechanism

So far, we have documented both changes in contemporaneous volatilities of forecast errors associated with our VAR as well as changes in parameters governing the dynamic responses in our model. We now disentangle the effects of changes in contemporaneous volatility and changes in response parameters. This allows us to identify changes in the monetary policy *transmission* rather than changes in the effects of monetary policy that are driven by changes in the volatilities of exogenous disturbances.

To do so, we want to isolate the effects of changes in Σ_t on the impulse responses. To do so we follow Canova & Gambetti (2009), who normalize their sign-restriction-based impulse responses by fixing the contemporaneous effect on the nominal interest rate. In figure 7 we consider a 25 basis point increase in the nominal rate (but any other normalization would just rescale the impulse responses we show). This does not mean that we think the nominal rate was the policy instrument throughout our sample - we just normalize the impact of the monetary policy shock through time. As mentioned before, our sign restriction approach allows us to not take a stand on the nature of the policy instrument. For the sake of brevity, we report the median response as well as the 68 % posterior bands, but omit the identified set of impulse responses. Again focusing first on possible policy instruments, we see only small changes in the median response for money growth and only mild reductions in uncertainty. It now becomes clear that the nominal rate response has become more persistent over time, a fact that had been masked by the changes in the impact of a monetary policy shock on the nominal rate on impact and the uncertainty surrounding that impact effect.

After 1982, the median response of inflation has become larger across horizons, and the probability of a strong negative response of inflation to a 25 basis point increase in the nominal rate has increased substantially. The response of the spread to a monetary policy shock shows the same pattern as in the previous section - the median response is negative throughout, but the error bands include 0 for every horizon starting with the Great Inflation period. Monetary policy in the latter part of the sample thus shifts the entire yield curve without changing the slope. The impact of monetary policy on long-term rates has increased over time, consistent

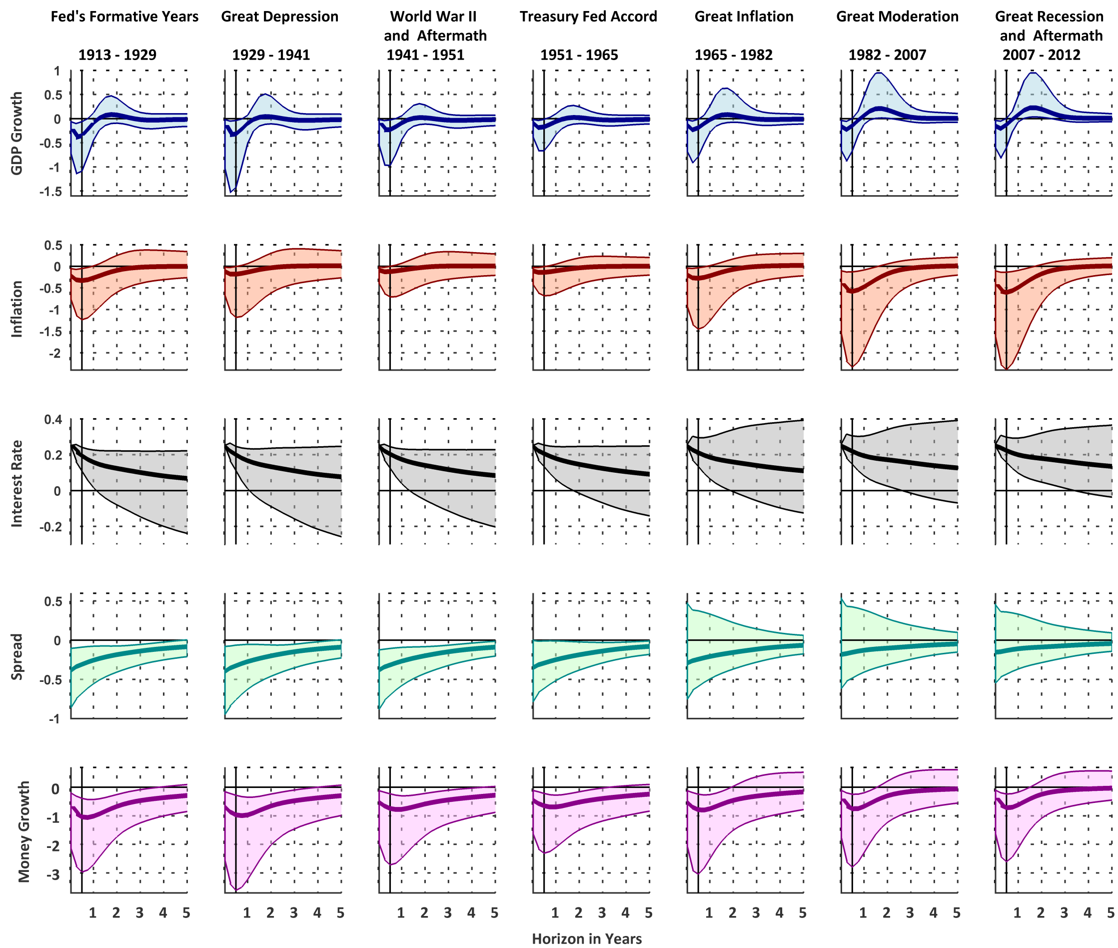

Figure 7: Impulse responses normalized on nominal rate impact across time.The solid bold line is the median response, the shaded area represents the 68 % posterior probability bands centered at the median.

with a narrative that attributes more credibility to the Fed in the latter part of the sample.

The median impact of monetary policy shocks for a given size on real GDP growth have decreased over time (similar to what we found for the one standard deviation shock), but this decrease is small compared to the width of the error bands at any point in time. For real GDP growth, we do not see a substantial difference in the response before and after Volcker. One feature that we do find for both shocks of fixed size and one standard deviation shocks, is that the downside risk associated with a contractionary monetary policy shock have become smaller over time -

the 16th percentile error band for the GDP growth response has become smaller in magnitude over time for all horizons.

While the reaction of policy instruments to a monetary policy shock of a given size has changed over time, we do find more similarities than differences in the responses of real GDP growth and inflation across the different time periods we consider. While there is substantial time variation in reduced-form moments, as we documented in previous sections, our identification scheme for monetary policy shocks implicitly attributes a substantial part of those changes to changes not associated with monetary policy shocks.

5 Conclusion

To study changes in the dynamics of the U.S. economy over the last century, we enrich a time-varying parameter VAR model along the lines of Primiceri (2005) to allow for possibly mismeasured data. We find substantial evidence of measurement error before and during WWII (particularly in GDP, in line with Romer (1989)), time variation in volatilities of reduced form innovations and substantial time variation in the correlations between the macroeconomic variables we consider. In particular, the early 1980s were a time period that our model associates with substantial shifts in the structure of the economy.

Changes in the responses to a monetary policy shock are clearly present, but once we condition on the size of the shock (i.e. the initial impact on short-term nominal interest rates) the most noteworthy finding is that of surprising stability over the past century.

References

Amir-Ahmadi, P. & Ritschl, A. (2013), 'Depression econometrics: A favar model of monetary policy during the great depression', *mimeo* .

Balke, N. & Gordon, R. J. (1986), Appendix B Historical Data, *in* 'The American Business Cycle: Continuity and Change', NBER Chapters, National Bureau of Economic Research, Inc, pp. 781–850.

Baumeister, C. & Benati, L. (2013), 'Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound', *International Journal of Central Banking* **9**(2), 165–212.

Baumeister, C. & Peersman, G. (2013), 'Time-Varying Effects of Oil Supply Shocks on the US Economy', *American Economic Journal: Macroeconomics* **5**(4), 1–28.

Benati, L. (2010), Evolving Phillips trade-off, Working Paper Series 1176, European Central Bank.

Benati, L. & Lubik, T. (2014), 'Sales, inventories, and real interest rates: A century of stylized facts', *Journal of Applied Econometrics* p. forthcoming.

Benati, L. & Mumtaz, H. (2007), U.S. evolving macroeconomic dynamics: a structural investigation, Working Paper Series 0746, European Central Bank.

Bernanke, B. S. (2013), 'A Century of US Central Banking: Goals, Frameworks, Accountability', *Journal of Economic Perspectives* **27**(4), 3–16.

Bianchi, F. & Ilut, C. (2013), Monetary/fiscal policy mix and agents' beliefs, Technical report.

Canova, F. & Gambetti, L. (2009), 'Structural changes in the US economy: Is there a role for monetary policy?', *Journal of Economic Dynamics and Control* **33**(2), 477–490.

Canova, F. & Nicolo, G. D. (2002), 'Monetary disturbances matter for business fluctuations in the G-7', *Journal of Monetary Economics* **49**(6), 1131–1159.

Canova, F. & Paustian, M. (2011), 'Business cycle measurement with some theory', *Journal of Monetary Economics* **58**(4), 345–361.

Carter, C. K. & Kohn, R. (1994), 'On gibbs sampling for state space models', *Biometrika* **81**(3), 541–553.

Christiano, L. J., Eichenbaum, M. & Evans, C. L. (1999), Monetary policy shocks: What have we learned and to what end?, *in* J. B. Taylor & M. Woodford, eds, 'Handbook of Macroeconomics', Vol. 1 of *Handbook of Macroeconomics*, Elsevier, chapter 2, pp. 65–148.

Cogley, T., Matthes, C. & Sbordone, A. (2015), 'Optimized taylor rules for disinflation when agents are learning', *The Journal of Monetary Economics, forthcoming* .

Cogley, T. & Sargent, T. J. (2002), Evolving Post-World War II U.S. Inflation Dynamics, *in* 'NBER Macroeconomics Annual 2001, Volume 16', NBER Chapters, National Bureau of Economic Research, Inc, pp. 331–388.

Cogley, T. & Sargent, T. J. (2005), 'Drift and volatilities: Monetary policies and outcomes in the post WWII U.S.', *Review of Economic Dynamics* **8**(2), 262–302.

Cogley, T. & Sargent, T. J. (2014), 'Measuring price-level uncertainty and instability in the U.S., 1850-2012', *Review of Economics & Statistics, forthcoming* .

Cogley, T., Sargent, T. J. & Surico, P. (2015), 'Price-level uncertainty and instability in the United Kingdom', *Journal of Economic Dynamics and Control* **52**(C), 1–16.

Cogley, T., Surico, P. & Sargent, T. J. (2012), The return of the Gibson paradox, Working paper, NYU and LBS.

Del Negro, M. & Primiceri, G. (2013), Time-varying structural vector autoregressions and monetary policy: a corrigendum, Staff Reports 619, Federal Reserve Bank of New York.

Faust, J. (1998), 'The robustness of identified VAR conclusions about money', *Carnegie-Rochester Conference Series in Public Policy* **49**, 207–244.

Hetzel, R. L. (2008), *The Monetary Policy of the Federal Reserve - A History*, Cambridge Universit Press.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **65**(3), 361–93.

Koop, G. & Korobilis, D. (2010), Bayesian multivariate time series methods for empirical macroeconomics, Working paper, University of Strathclyde.

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

Lubik, T. A. & Schorfheide, F. (2003), 'Computing sunspot equilibria in linear rational expectations models', *Journal of Economic Dynamics and Control* **28**(2), 273–285.

Lütkepohl, H. (2010), *New Introduction to Multiple Time Series Analysis*, Springer Verlag.

Moon, H. R., Granziera, E. & Schorfheide, F. (2013), 'Inference for vars identified with sign restrictions', *mimeo* .

Primiceri, G. (2005), 'Time varying structural vector autoregressions and monetary policy', *Review of Economic Studies* **72**(3), 821–852.

Ritschl, A., Sarferaz, S. & Uebele, M. (2015), 'The u.s. business cycle, 1867-2006: A dynamic factor approach', *Review of Economics & Statistics, forthcoming* .

Romer, C. D. (1986), 'New estimates of prewar gross national product and unemployment', *The Journal of Economic History* **46**(02), 341–352.

Romer, C. D. (1989), 'The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869-1908', *Journal of Political Economy* **97**(1), 1–37.

Rubio-Ramirez, J. F., Waggoner, D. F. & Zha, T. (2010), 'Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference', *Review of Economic Studies* **77**(2), 665–696.

Schorfheide, F., Song, D. & Yaron, A. (2014), Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach, Technical report.

Sims, C. (1999), 'The role of interest rate policy in the generation and propagation of business cycles: What has changed since the 30's', *Proceedings 1998 Boston FRB Annual Res. Conf., Federal Reserve Bank of Boston* pp. 121–160.

Sims, C. A. (2001), 'Comment on Sargent and Cogley's "Evolving US Postwar Inflation Dynamics"', *NBER Macroeconomics Annual Volume 16* .

Sims, C. A. & Zha, T. (2006), 'Were there regime switches in U.S. monetary policy?', *American Economic Review* **96**(1), 54–81.

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

Wheelock, D. C. & Wohar, M. E. (2009), 'Can the term spread predict output growth and recessions? a survey of the literature', *Federal Reserve Bank of St. Louis Review* .

Appendix

A Data

A.1 Annual Output Growth

Our output growth series is obtained by splicing two different real output series covering different time spans. We use real GNP series as constructed by Balke & Gordon (1986) from the first quarter of 1876 to the fourth quarter of 1946 (1930:Q1 and 1947:Q1 are the break dates for the parameters of the measurement error process associated with this series). Starting in 1947:Q1, we use the real GDP series provided by the St. Louis Fed FRED database covering the first quarter of 1947 to the second quarter of 2011. The spliced series are transformed in logs and then we take year-on-year differences.

A.2 Annual Inflation Rate

The corresponding annual inflation rate is also based on the combination of two different series on the output deflator. Again the first part comes from Balke & Gordon (1986) covering the period 1876Q1-1946Q4. Just as with output growth, the break date for the parameters of the measurement error process is 1947:Q1. The second part of the series comes from the St. Louis Fed FRED database covering the time span 1948Q1-2011Q2. Again we transform the data into year-on-year growth rates.

A.3 Short-Term Interest Rate

The short term interest rate plays the role of a potential direct or indirect monetary policy instrument for at least a substantial part of the time span we analyze. There is no single series on shorter interest rates at quarterly frequency for the full sample, which requires constructing a series based on several data sources reflecting short-term borrowing conditions. From 1920Q1-2011Q2 we use data on the 90-day T-Bill rate from the secondary market. Prior to that we backcast the series including as regressors data on call money rates and commercial paper rates. These two series show a strong contemporaneous correlation with the T-Bill rate when all series are available. The series used for backcasting and our target short term interest rate series are all available at monthly frequency. Specifically, we regress 90-day T-Bill rate on call money rates and commercial paper rates based on a sample running from February 1920 to April 1934[18]. Combining the resulting coefficients with our regressors we can backcast our target series back to the first quarter of 1876. This way we interpolate backward the missing observations for the 90-day T-Bill rate. We thus avoid using the six-month short term interest rate, which would lead to a maturity mismatch combining the three-month and six-month rates. Furthermore, we prefer the shorter maturity rate as a potential monetary policy instrument. We use annualized interest rates throughout. The break point for the parameters of the measurement error process is 1920:Q1.

A.4 Long-Term Interest Rate

As for the term spread, we employ the difference between a constructed measure of the long-term interest rate and the short-term interest rate described in the previous section. The lack of a consistent long-term government benchmark interest rate requires the combination and backcasting of three indicators. From 1920Q1-2011Q2 we use data on the

[18]Some experimentation with alternative windows for the backcasting exercise lead to essentially same results.

10-year government bond yields at constant maturities. Prior to that, we backcast the series including as regressors data on railroad bond yields (high grade) and a railroad bond yields index. Again, there is a strong contemporaneous correlation between the series we use to approximate the long-term interest rate and that interest rate itself when all series are available.

We regress 10-year government bond yields at constant maturities on railroad bond yields (high grade) and railroad bond yields index based on a sample running from February 1920 to April 1934. Combining the resulting coefficients with our regressors we can backcast our target series back to the first quarter in 1876. The long-term interest rate is expressed in annual terms. Just as with short-term rates, the break point for the parameters of the measurement error process here is 1920:Q1.

A.5 Annual Base Money Growth

The monetary base measure we use to represent a direct or indirect monetary policy instrument is obtained by combining by two series. The first part of the sample from 1876Q1-1958Q4 comes from Balke & Gordon (1986) and the second part from the FRED database covering 1959Q1-2011Q2. Since the Balke & Gordon (1986) data uses different sources itself, we allow for further breaks in 1918 and 1935 in addition to the break point in 1958.[19]

[19] While the measurement of the money base has undergone multiple changes over the years, our reading of St. Louis Fed documentation on this subject helped us identify these possibly major break points. The St Louis Fed uses similar sources for the first part of the sample as do Balke & Gordon (1986). Links to this documentation are:

1. `http://research.stlouisfed.org/publications/review/03/09/Anderson.pdf`
2. `http://research.stlouisfed.org/aggreg/newbase.html`

B Prior Choice

We choose priors in a way to stay as close as possible to the previous literature, while taking into account our larger sample and the addition of measurement error processes. We use data from 1876:Q1 to 1913:Q4 to initialize the priors for the VAR for y_t by using a fixed coefficient VAR, similarly to Primiceri (2005).

The priors for the measurement errors and the associated parameters are set similarly to Cogley, Sargent & Surico (2015). The priors are the same for each data source, but vary across variables to take into account the different volatilities of each variable.

We use independent normal-inverse gamma priors for each set of measurement error process coefficients. As we change the measurement error process for inflation and output growth, we keep the structure of the prior (i.e. the distributional assumptions), but modify some of the parameters of the priors for the measurement error processes to take into account how the measurement error enters the measurement equations for inflation and output growth. The prior for the AR coefficients for the measurement errors is independent across variables and break dates. It is Gaussian with mean 0 and prior standard deviation 0.45, which we keep for both sets of specifications, our benchmark specification and the specification with $M_i(L) = 1 \forall i$. The prior for the variance of the innovation of the AR processes is inverse-gamma and independent across variables and break dates. The mode of the inverse-gamma distribution is set to a fraction of the standard deviation of each variable during the training sample (the prior scale parameters are set to 2). Cogley, Sargent & Surico (2015) use 50 % of the training sample standard deviation for the prior mode for their model of inflation. For our growth rate specification (which is the specification Cogley, Sargent & Surico (2015) used for their model of inflation), we use the same value for our inflation measurement error as well as for the interest rate and spread series, but found that the standard deviation for real GDP growth and money growth during the training sample is so high that mechanically using the same value as Cogley, Sargent & Surico (2015) for those series resulted in somewhat implausible estimates. Instead of a scaling factor of 50 % of the training standard deviation, we in that case thus a scaling factor of 15 % of the training sample deviation for GDP growth and money growth. Since we use priors for σ_j^i with an infinite variance (Cogley, Sargent & Surico (2015) also use a prior with infinite variances for their corresponding parameter), this change does not restrict the posterior to assign only a minor role to measurement error.

For our benchmark specification, we keep the prior on the AR coefficients, but reduce the prior mode for the innovation in the AR process. We do this because the level specification automatically introduces additional volatility (this is easiest to see if we think of iid measurement errors in levels - then the composite measurement error is the difference of two iid measurement errors and thus has twice the volatility of the original measurement error). We thus set the prior mode of the innovation volatilities in the measurement error processes to capture half of the standard deviation that the corresponding prior in the growth rate specification captured. We keep the scale parameters the same across the two specifications. Summarizing the priors for the measurement error process, we have

$$\rho_j^i \sim N(0, .45^2) \tag{13}$$

$$\sigma_j^i \sim IG(scaling_i * \hat{\sigma}^2_{i,train}, 2) \tag{14}$$

where $\hat{\sigma}^2_{i,train}$ is the estimate of the variance of observed variable i from the training sample and scaling is set as described above. We use (somewhat non-standard) notation for the inverse gamma where the first argument gives the prior mode and the second argument the scale parameter.[20]

An important prior for time-varying parameter VARs is the prior for Q, the covariance matrix of the residuals that enter the law of motion for θ. We assume that Q, which governs the amount of time variation in the VAR coefficients, follows an inverse Wishart distribution with the following parameters:

$$Q \sim IW(\kappa_Q^2 * 152 * V(\theta_{OLS}), 152) \tag{15}$$

where the prior degrees of freedom is set to 152, which is the length of our training sample, $V(\theta_{OLS})$ is the variance of the OLS estimator of the VAR coefficients in our training sample and $\kappa_Q = 0.01$ is the tuning parameter to parameterize the prior belief about the amount of time variation. Primiceri (2005) uses exactly the same approach to set his prior. Choosing the same approach allows us to keep our results comparable to his. The other priors are also set according to Primiceri (2005), adjusting for the larger size of our vector of observables. In contrast to Cogley & Sargent (2005), we do not impose the prior that the companion matrix of our VAR only has eigenvalues smaller than 1 in absolute value. We do this to be able to study if there is significant variation in the probability of this local non-stationarity.

[20]The scale parameter of an inverse gamma distribution is one of the two parameters commonly used for this family of distributions. Importantly, when we talk about the scale parameter, we do not mean $scaling_i$.

C Sources of Volatility

Volatility in time series models can be traced back to two sources: the innovations (or unpredictable components) that influence the time series of interest and the systematic response to those innovations. To make this point, consider a univariate AR(1) model with Gaussian innovations:

$$z_t = \rho z_{t-1} + w_t,\ w_t \sim N(0, \sigma_w^2) \tag{16}$$

Then the j-step ahead conditional variance is given by

$$Var_t(z_{t+j}) = \sigma_w^2 \sum_{k=1}^{j} \rho^{2(j-k)} \tag{17}$$

We can see that the volatility of this process is fully characterized by the autoregressive coefficient and the variance of the innovation. The next two sections present a similar characterization for our time-varying VAR. The objects corresponding to ρ in the multivariate context are the $\mathbf{A}_t$ matrices, which are high dimensional. To study dynamics, we can focus on the eigenvalues, but even those are large in number (given that they vary over time). The section below therefore focuses on the largest eigenvalue in absolute value. This object does not fully characterize the effects of time variation in persistence on volatility, but it does give an idea about whether or not our estimated model features (locally) unstable dynamics, which in turn will have an effect on volatility.

C.1 Are There Explosive Dynamics in U.S. Time Series?

We study the probability of matrix $\mathbf{A}_t$ having eigenvalues larger than 1 in our sample by checking the draws of $\mathbf{A}_t$ that are generated by our Gibbs sampler. We can do this because, as mentioned before, we do not follow Cogley & Sargent (2005) and impose conditions on the eigenvalues of the companion matrix of our VAR. The right panel of Figure 8 shows this probability, whereas the left panel shows draws from the posterior path of the maximum absolute eigenvalue as well as the median and 68 percent posterior probability bands.

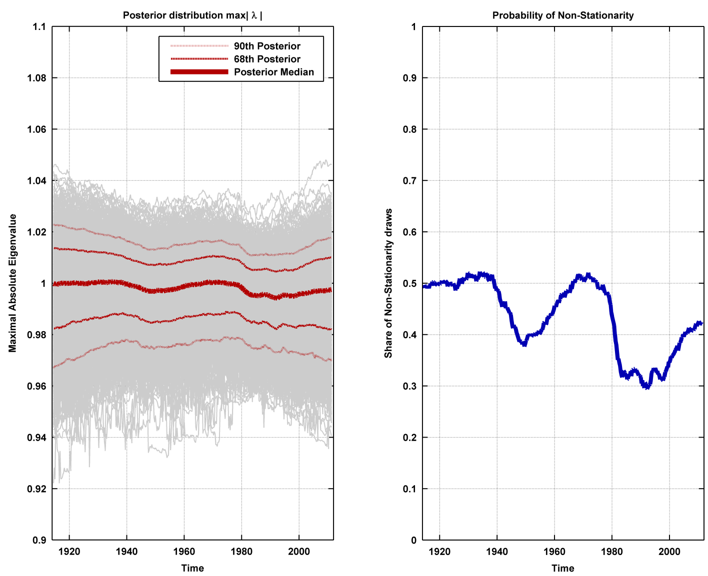

Figure 8: Explosive behavior

The average level of the probability until the 1940s is quite high, reaching over 0.5. The probability drops almost 20 percentage points at the end of WWII. It rises again until the end of the 1970s. The second big decrease in this probability following the Volcker disinflation could be interpreted in terms of a structural model in which agents have to learn about the true data-generating process (DGP): Cogley, Matthes & Sbordone (2015) show that times in which beliefs of private agents are far away from the DGP can lead to explosive dynamics, whereas the probability of explosive eigenvalues falls as beliefs move closer to the true DGP. An alternative structural model that can give temporarily explosive dynamics is given by Bianchi & Ilut (2013). To wit, we find two large changes in the probability of local non-stationarity.

Despite the fact that high probability of explosiveness can be found in various periods in the history, the left panel of Figure 8 shows that the absolute value of those eigenvalues larger than 1 is only slightly larger than 1. This means that even if there are eigenvalues larger than 1, it takes a long time for the economy to become noticeably unstable. Concerning the kind of stationarity restrictions used by Cogley & Sargent (2005), there is a substantial posterior probability of having explosive eigenvalues, making estimation algorithms with this restriction slow to converge. At the same time, the restriction itself is not far from being met for large parts of post-WWII data in the sense that the estimated eigenvalues are not far from 1[21].

D The Growth Rate Specification for Measurement Error

In this section we study a model with $M_i(L) = 1\ \forall i$. The differences in results between this approach and our benchmark are small. For the sake of brevity, we focus on the estimated 'true' data as well as impulse responses. The first figure shows the estimated 'true' data (with the median in bold red), and 68% posterior bands. The break points for the measurement errors are the same as for our benchmark, with the exception that we do not have a break for GDP growth in 1930. Adding this break here does not change the results. After the last break point the estimated 'true' data coincides with the observed data. We see in figure

[21] Given that we also use pre-WWII data, this approach would be harder to defend for our application.

9 that the estimated 'true'data is very similar to that obtained using our benchmark specification. The major difference between this specification and our benchmark is that this specification attributes the GDP growth downturns in the second half of the 1930s and after WWII to movements in actual data, whereas our benchmark mainly sees this as measurement error.

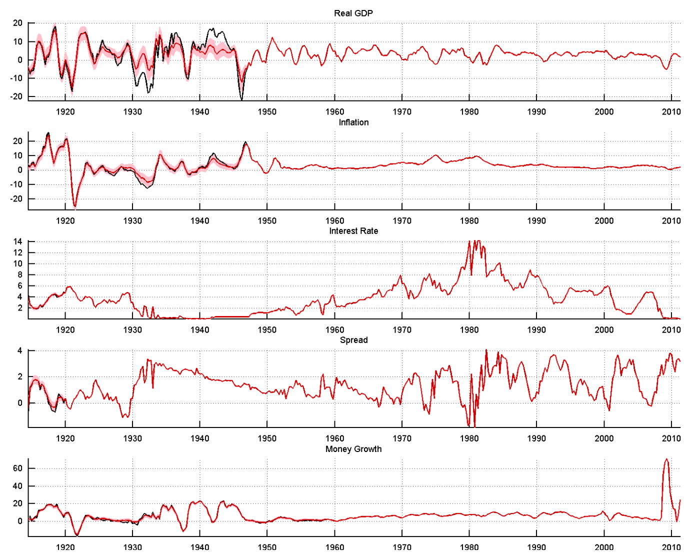

Figure 9: Data

Concerning the impulse responses, the patterns are very similar to our benchmark results. Figures 10 and 11 shoes the impulse responses. Once thing we do find in this specification, at least in the one standard deviation shock case, is a change in the behavior of the GDP growth impulse response with the Fed Treasury Accord, in line with some of the reduced form changes that we found both in the benchmark case and with this specification.

Table 1 shows the volatility of the estimated and observed GDP growth series for different periods. We see the same patterns as for our benchmark specification. The reduction in the volatility of the estimated 'true' series starting in 1930 is still present, albeit less pronounced.

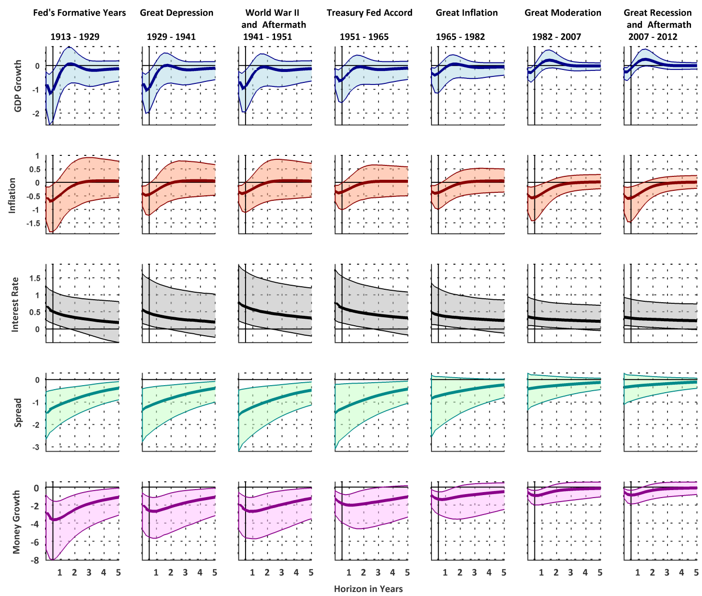

Figure 10: Impulse responses for one standard deviation shock

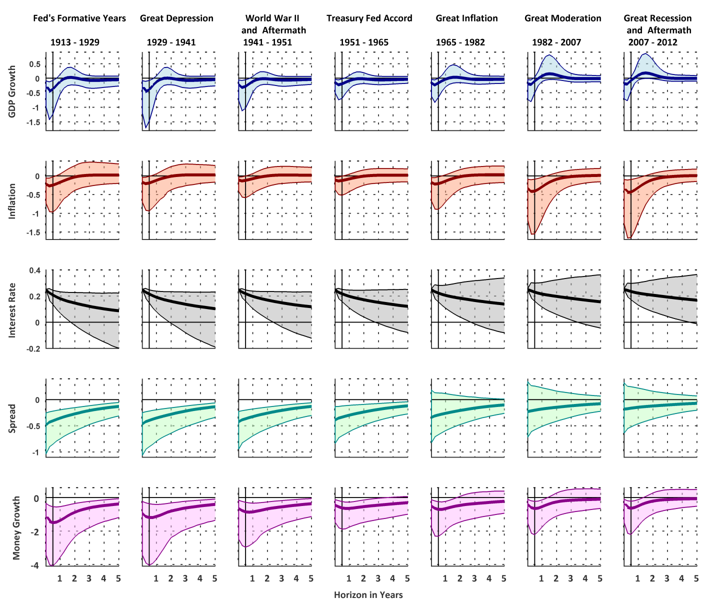

Figure 11: Impulse responses for 25 basis point nominal interest rate shock

E Estimation Algorithm

We use a Gibbs-Sampler to approximate the posterior distribution by generating $100,000$ draws. The exact implementation for time-varying parameters and stochastic volatilities follows Primiceri (2005) including

	Estimated	Observed	$\frac{\text{Estimated}}{\text{Post WWII}}$	$\frac{\text{Observed}}{\text{Post WWII}}$
1915-1946	6.1	9.6	2.4	3.7
1915-1929	6.7	7.7	2.6	2.9
1930-1946	5.5	11.0	2.1	4.2
1947-2006	2.6	2.6	1	1

Table 3: Standard deviation of observed and estimated real GDP growth for different periods calculated using the alternative specification of the measurement error process.

the corrigendum of Del Negro and Primiceri (2013). In addition, we propose a multivariate generalization of Cogley, Sargent & Surico (2015) to simulate the posterior distribution for measurement error process parameters and unobserved 'true' data.

Let $\tilde{y}^T$ be the observed noisy data vector, $S^T = (y^T, m^T)$ be the vector of the unobserved data and associated measurement errors with and Θ^T be the collection of all parameters of the time-varying VAR with stochastic volatilities. Note that conditional on the (partly unobserved) 'true' data, the steps we borrow from Primiceri (2005) do not need to be altered: knowledge of the measurement error or the parameters of the measurement error process are irrelevant for those steps in the Gibbs sampler. The algorithm proceeds as follows[22]

1. Draw Σ^T from $p(\Sigma^T|y^T, \theta^T, \Lambda^T, V, s^T)$. This step requires us to generate draws from a nonlinear state space system. We use the approach by Kim, Shephard & Chib (1998) to approximate draws from the desired distribution. For a correct posterior sampling of the stochastic volatilities we follow the corrigendum in Del Negro and Primiceri (2013) and the modified steps therein.

2. Draw θ^T from $p(\theta^T|y^T, \Lambda^T, \Sigma^T, V)$. Conditional on all other param-

[22] A superscript T denotes a sample of the relevant variable from $t = 1$ to T.

eter blocks equations (4) and (5) from the main text form a linear Gaussian state space system. This step can be carried out using the simulation smoother detailed in Carter & Kohn (1994).

3. Draw Λ^T from $p(\Lambda^T|y^T, \theta^T, \Sigma^T, V)$. Again we draw these covariance states based on the simulation smoother of the previous step, exploiting our assumption that the covariance matrix of the innovations in the law of motion for the λ coefficients is block diagonal. This assumption follows Primiceri (2005), where further details on this step can be found.

4. Draw V from $p(V|\Sigma^T y^T, \theta^T, \Lambda^T)$. Given our distributional assumptions, this conditional posterior of the time-invariant variances follows an inverse-Wishart distribution, which we can easily sample from.

5. Draw S^T from $p(S^T|\Theta^T, \rho_m, \sigma_m^2, \tilde{y}^T)$

   Conditional on $\Theta^T, \rho_m, \sigma_m^2$ and $\tilde{y}^T$, equation (2) from the main text and and the set of equations (10) from the main text form a linear-Gaussian state-space system where the conditional posterior of S^T is also Gaussian and can be simulated using the Carter & Kohn (1994) sampler. The initialization of the Kalman filter is given by

$$S_0 = \begin{pmatrix} y_0 \\ m_0 \end{pmatrix} = \begin{pmatrix} \bar{y}_{train} \\ 0 \end{pmatrix}$$

   where $\bar{y}_{train}$ is the mean of the training sample of the observed data vector $\tilde{y}^T$. The initial MSE matrix P_0 is specified as

$$P_0 = \begin{pmatrix} \hat{\sigma}^2_{1,train} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 \\ \vdots & 0 & \hat{\sigma}^2_{M,train} & 0 & 0 & \vdots \\ 0 & 0 & 0 & \sigma^2_{1,\text{mode}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sigma^2_{M,\text{mode}} \end{pmatrix}$$

where $\hat{\sigma}^2_{i,train}$ is the unbiased variance estimate of the observed variable i from the training sample and $\sigma^2_{i,\text{mode}}$ is the prior mode of the variance of the measurement error i. We describe the state space system used to draw S in more detail in a separate section below.

6. Draw ρ_m and σ^2_m from $p(\rho_m, \sigma^2_m | S^T, \Theta^T, \tilde{y}^T)$

   Since all measurement error innovations are independent, the only relevant conditional information set is m^T. Conditioned on m^T and using the independent normal-inverse gamma prior for each of the measurement error process independently, the conditional posterior $p(\rho_m|\sigma^2_m, m^T)$ is normal and the conditional posterior $p(\sigma^2_m|\rho_m, m^T)$ is inverse gamma, which can be sampled using two Gibbs steps.

7. Draw s^T, the sequence of indicators for the mixture of normals needed for the Kim et al. (1998) stochastic volatility algorithm.

F Algorithm to Draw Generalized Impulse Responses

Here we describe the Monte Carlo procedure for the identification of the evolving generalized impulse response functions to contractionary monetary policy shocks employed via pure sign restrictions as briefly outlined in the main text. The exposition draws mostly on the procedure described in Benati & Mumtaz (2007), Baumeister & Benati (2013) and Baumeister & Peersman (2013), who build on Koop et al. (1996).

We compute the candidate generalized impulse responses as the difference between the conditional expectations with and without a specific value of the exogenous shock ϵ at time t

$$irf_{cand,t+k} = E[X_{t+k} \mid \epsilon_t, \omega_t] - E[X_{t+k} \mid \omega_t]$$

where X_{t+k} contains the forecasts of the endogenous variables at horizon k, ω_t represents the current information set that captures the entire history up to that point in time, and ϵ_t is the current disturbance term. At each point in time, the information set upon which we condition the forecasts contains the actual values of the lagged endogenous variables and a random draw of the model parameters and hyperparameters. To calculate the conditional expectations, we randomly draw from the Gibbs sampler output at a given time t the time-varying coefficients, the variance covariance matrix and the hyperparameters. We employ the transition laws and stochastically simulate the future paths of the coefficient vector and the components of the variance covariance matrix for up to 20 quarters into the future. By projecting the evolution of the system we account for all potential three sources of uncertainty from the corresponding innovations in the system. To obtain the time t structural impact matrix, $B_{0,t}$ we first obtain a rotation matrix Q following (Rubio-Ramirez, Waggoner & Zha (2010)) and combine it with the lower triangular Cholesky factor of $\Omega_{t|T}$ resulting in $B_{0,t} = \Lambda_{t|T}^{-1}\Sigma_{t|T}Q'$. Given this contemporaneous impact matrix, we compute the reduced-form innovations based on the relationship $e_t = B_{0,t}\epsilon_t$. From the set of candidate impulse responses derived in this way, only those satisfying our sign restriction are used to compute the impulse responses. All others are discarded. Based on these impulse responses, we calculate the statistics of interest. In particular, the minimum and maximum responses at each horizon are used to estimate the full identified set.

This procedure is computationally cumbersome and quite time consuming. We calculate the generalized impulse response functions at each point in time $t = 1, \ldots, 366$, given a random selection of 500 states of the economy explicitly taking into account possible future uncertainty in the structure of the economy along the horizon considered. For each of those random draws we calculate 50 candidate impulse response functions resulting in a total of 9.15 million candidates. The procedure described here (not including the Gibbs sampler to estimate the model in the reduced form model in the first place) takes for a given specification on an AMD Opteron(tm) Processor 6172, 2.10 GHz (8 processors), 16GB RAM and a 64-bit operating system about 7 days to run.

G Drawing Measurement Errors and 'True' Data

This section describes the state space system used to generate draws of the measurement errors and the unobserved 'true' data. For notational convenience, we focus here on our benchmark specification, but generalizing this section to the case with general $M_i(L)$ is straightforward.

Let $S_t = (\mathbf{Y}_t, \mathbf{m}_t^1, \mathbf{m}_t^2, \mathbf{m}_t^3, \mathbf{m}_t^4, \mathbf{m}_t^5)'$ be the vector of unobserved data and measurement errors where $\mathbf{m}_t^i = (m_t^i, m_{t-1}^i, m_{t-2}^i, m_{t-3}^i, m_{t-4}^i)'$ for $i = 1, 2, 5$ where the index i denotes the measurement error associated with the VAR variable i in the ordering[23]. The joint state space representation for

[23]We could in principle include higher order lags of measurement errors for the in-

observed noisy data vector $\tilde{y}_t$ is defined as (we write the dynamics of $\mathbf{Y}_t$ as a VAR(1) - it is the companion form of our original VAR):

$$\begin{aligned} \tilde{y}_t &= JS_t \\ S_t &= \delta_t + F_t S_{t-1} + \omega_t \end{aligned}$$

where the state law of motion is given by the VAR and the measurement error structure:

$$\begin{bmatrix} \mathbf{Y}_t \\ \mathbf{m}_t^1 \\ \mathbf{m}_t^2 \\ m_t^3 \\ m_t^4 \\ \mathbf{m}_t^5 \end{bmatrix} = \begin{bmatrix} \mu_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{A}_t & 0 & & \cdots & & 0 \\ 0 & \rho_m^1 & 0 & & & \\ & 0 & \rho_m^2 & 0 & & \vdots \\ \vdots & & 0 & \rho_m^3 & 0 & \\ & & & 0 & \rho_m^4 & 0 \\ 0 & & \cdots & & 0 & \rho_m^5 \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-1} \\ \mathbf{m}_{t-1}^1 \\ \mathbf{m}_{t-1}^2 \\ m_{t-1}^3 \\ m_{t-1}^4 \\ \mathbf{m}_{t-1}^5 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_t \\ \varepsilon_t^{m,1} \\ \varepsilon_t^{m,2} \\ \varepsilon_t^{m,3} \\ \varepsilon_t^{m,4} \\ \varepsilon_t^{m,5} \end{bmatrix}$$

where

$$\rho_m^i = \begin{bmatrix} \rho_m^i & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

and $\varepsilon_t^{m,i} = (\varepsilon_t^{m,i}, 0, ..., 0)'$ for $i = 1, 2, 5$.

The covariance matrix of ω_t is given by

$$Var(\omega_t) = \begin{bmatrix} Var(\mathbf{e}_t) & 0 & & \cdots & & 0 \\ 0 & Var(\varepsilon_t^{m,1}) & 0 & & & \\ & 0 & Var(\varepsilon_t^{m,2}) & 0 & & \vdots \\ \vdots & & 0 & Var(\varepsilon_t^{m,3}) & 0 & \\ & & & 0 & Var(\varepsilon_t^{m,4}) & 0 \\ 0 & & \cdots & & 0 & Var(\varepsilon_t^{m,5}) \end{bmatrix}.$$

terest rate and term spread for notational convenience. But in our application both variables are actually measured in levels.

Finally, the selection matrix J is specified such that

$$\begin{bmatrix} \Delta gdp_t \\ \pi_t \\ i_t^s \\ spread_t \\ \Delta money_t \end{bmatrix} = \begin{bmatrix} y_{1,t} + m_t^1 - m_{t-4}^1 \\ y_{2,t} + m_t^2 - m_{t-4}^2 \\ y_{3,t} + m_t^3 \\ y_{4,t} + m_t^4 \\ y_{5,t} + m_t^5 - m_{t-4}^5 \end{bmatrix} = J \times S_t$$