

Economics 704a Lecture 3: Evidence for Monetary Non-Neutrality

Adam M. Guren

Boston University

Spring 2026

Empirical Evidence on Role of Money

- Big questions in monetary economics:
 - How is an economy with money different from an economy without money?
 - What effects do change in monetary policy have on real activity and inflation?
- Before we add money to canonical RBC model, turn to empirical evidence.
 - Will provide facts any successful model must match and motivate features we will add to model.
- Estimating effect of monetary policy is hard because *policy is endogenous* and *expectations affect equilibrium today*.
 - In fact, if monetary policy is successful and countercyclical, estimates of its effect would be biased towards zero.
 - Intuition: Monetary policy would move around but output would be stable.

Evidence for Nominal Rigidity

1. Vector Autoregressions
2. VAR Evidence for Non-Neutrality
3. Other Approaches
4. Natural Experiments

Introduction to VARs

- Before we get started, introduce a key econometric tool: Vector Autoregression or VAR.
 - Proposed by Sims (1980), who won the Nobel Prize for it.
 - Wanted a way to describe economic time series with minimal theoretical restrictions.
- This is a key tool in macro to summarize relationships between macroeconomic time series.
 - To motivate / test models.
 - Examine response to structural shocks.
 - Frequently used at central banks.

Stationarity and White Noise

- Cannot find regularities if things do not repeat themselves.
- Leads to concept of *stationarity*.
 - A time series $\{x_t\}$ is stationary if the mean, variance, and autocorrelation can be well approximated by sufficiently long time averages.
 - In other words, $\{x_t\}$ is covariance stationary if:

$$E \{x_t\} = \mu \quad \forall t \text{ and } E \{(x_t - \mu)(x_{t-k} - \mu)\} = g_k \quad \forall t, k$$

- Sometimes not a great assumption (e.g., economies in transition), but for post-war US GDP, it works.
 - Otherwise, detrend or difference.
- A *white noise process* has mean zero, a constant variance, and is serially uncorrelated.

Autoregressions

- An *autoregression* is a regression of a time series $\{x_t\}$ on lags of itself.
- Example: AR(1)

$$x_t = \beta_0 + \beta_1 x_{t-1} + \varepsilon_t$$

- Stationary and stable if $|\beta_1| < 1$.
 - Otherwise goes off to infinity and never mean reverts.
- Can estimate AR(p)

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_p x_{t-p} + \varepsilon_t$$

by OLS if $\{x_t\}$ is stationary and ε_t is a white noise process.

Vector Autoregression

- A *vector autoregression* is a generalization of an autoregression in which x_t is a *vector* of time series.
- Simple example we will use:

$$x_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}$$

- However can be of arbitrary size n .
- The *reduced-form* single-lag VAR (or VAR(1)) of x_t is then:

$$x_t = A_0 + A_1 x_{t-1} + e_t$$

where A_0 is an $n \times 1$ vector and A_1 is an $n \times n$ matrix.

Reduced-Form VAR: Estimation

- More generally, for a VAR of size n with p lags,

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \dots + A_p x_{t-p} + e_t$$

x_t , A_0 , and e_t are $n \times 1$ vectors and A_i are $n \times n$ matrices.

- There are thus $n + pn^2$ coefficients and $(n + 1)n/2$ in the variance-covariance matrix.
- The right hand side only contains predetermined variables of a stationary process, and the error terms are assumed to be serially uncorrelated with constant variance (can relax).
 - So can estimate each equation by OLS.
 - Application of seemingly unrelated regression.

Forecasting

- Can forecast using VAR:

$$E_t x_{t+1} = A_0 + A_1 x_t$$

$$E_t x_{t+2} = A_0 + A_1 E_t x_{t+1} = A_0 + A_1 [A_0 + A_1 x_t]$$

- Often-used diagnostic tool is the forecast error variance decomposition (FEVD).
 - Tells us proportion of variance of movements in $\{y_t\}$ or $\{z_t\}$ due to $e_{1,t}$ and $e_{2,t}$.
 - Like a partial R^2 of forecast error by forecast horizon.
 - See econometrics class for derivation.

Impulse Responses

- Theorem: VAR has a vector moving average representation.
- Example with one lag (and a zero constant term, to keep things simple):

$$\begin{aligned}x_t &= Ax_{t-1} + e_t \\&= A(Ax_{t-2} + e_{t-1}) + e_t \\&= e_t + Ae_{t-1} + A^2e_{t-2} + A^3e_{t-3} + \dots\end{aligned}$$

- The response of x_t to a one unit shock to $e_{1,t}$ in period t after n periods with no other e shocks is:

$$IR(n) = A^n \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Impulse Response Functions

$$IR(n) = A^n \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- This is called a reduced-form *impulse response function* to e_{1t} and is a convenient way to represent how shocks $\{e_t\}$ affect $\{x_t\}$.
 - Can plot graphically and create standard error bands.
- Intuitively, this is the difference between two processes $\{x_t\}$ that are made up of identical shocks $\{e_t\}$ except in period t , where an additional unit one shock is added to $e_{1,t}$.

Structural VARs

- Unfortunately, reduced form VARs are restrictive.
 1. No simultaneous causality.
 2. Shocks have to be uncorrelated and white noise.
 - Note: These are the same problem written different ways.
- Generalize to a *structural VAR*, which relaxes these.
- Tackle problem 1 first and allow for simultaneous causality:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt}$$

where $\varepsilon_{y,t}$ and $\varepsilon_{z,t}$ are independent white noise processes.

- Cannot directly estimate because y_t is correlated with $\varepsilon_{z,t}$ and vice-versa, violating exclusion restriction.

Reduced-Form Representation

- Write structural VAR as a matrix:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$
$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t$$

- Premultiply by B^{-1} to get reduced-form representation:

$$x_t = A_0 + A_1 x_{t-1} + e_t$$

where $A_0 = B^{-1}\Gamma_0$, $A_1 = B^{-1}\Gamma_1$, and $e_t = B^{-1}\varepsilon_t$.

- Note reduced form errors e_t are of form:

$$e_{1t} = (\varepsilon_{yt} - b_{12}\varepsilon_{zt}) / (1 - b_{12}b_{21})$$

- Stationary white noise, but *correlated with one another*.
- For IRFs and FEVDs, want responses to ε s not e s.

The Identification Problem

- *Cannot invert from reduced form to a unique structural VAR unless add restrictions.*
- Reduced form has 9 unknowns, six a s plus 3 terms of var-covar:

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t}$$

$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t}$$

- Structural form has 10 unknowns, 8 b s and γ s plus 2 terms of var-covar matrix (uncorrelated shocks):

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt}$$

- Fundamentally under-identified.
 - Intuitively, e_s depend on both ε_{yt} and ε_{zt} so cannot invert from e_s to ε_s .
 - Cannot distinguish which structural shock causes the reduced form errors. Mathematically, do not know rotation.

Recursive VARs and Identification

- Solution is recursive system:
 - Assume y_t has contemporaneous effect on z_t , but z_t has no contemporaneous effect on y_t , so $b_{12} = 0$.
 - Jargon: “order” y_t first, in the sense that it is “causally prior.”
- System is:

$$y_t = b_{10} + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt}$$

so

$$e_{1t} = \varepsilon_{yt} \text{ and } e_{2t} = \varepsilon_{zt} - b_{21}\varepsilon_{yt}$$

- Exactly identified because one parameter (b_{12}) is now a zero.
9 parameters in both structural VAR and reduced form.
- Intuition: Can now distinguish ε_{yt} and ε_{zt} shocks.
 - Only ε_{yt} shocks affect contemporaneous values of y_t .
 - e_{1t} attributed completely to ε_{yt} ; can invert es to get ε s.

Cholesky Decomposition

- Lower triangular assumption on the structural residuals is called a *Cholesky decomposition*.
 - Most common identification scheme for VAR.
- Generalize this to n variable and p lag VAR.
 - B is then an $n \times n$ matrix.
 - Exact identification requires $(n^2 - n) / 2$ restrictions between the regression residuals and structural innovations.
 - Cholesky does this by setting exactly $(n^2 - n) / 2$ values of the B matrix to zero.

Error Correlation Version

- Assume instead there is no simultaneous causality but $e_t = C\varepsilon_t$ where ε_t are the independent structural shocks and C is an $n \times n$ matrix:

$$x_t = A_0 + A_1x_{t-1} + C\varepsilon_t$$

- Equivalent to reduced form VAR

$$C^{-1}x_t = C^{-1}A_0 + C^{-1}A_1x_{t-1} + \varepsilon_t$$

- Same as before. Recursive VAR if C^{-1} is lower triangular.
- Same problem – cannot tell apart shocks.
- Now direct relationship between e s and ε s instead of relationship arising through simultaneous causality.
- Alternate interpretation of Cholesky: Assume ε_{yt} affects both y_t and z_t but ε_{zt} has no contemporaneous effect on y_t .

Cholesky Decomposition: Key Assumptions

- Cholesky is a **STRONG** assumption.
 - No reverse-causality.
 - No omitted variables correlated with “lower ordered” variables and “higher ordered” variables.
 - Strong exclusion restriction.
- “Ordering” sounds innocuous. It’s not.
 - $n!$ possible orderings!

Stock and Watson: VAR Criticisms in Practice

1. What really are the VAR “shocks?”

- Concern: Shocks reflect factors omitted from model. If correlated with included variables, then OVB.
- In practice: imagine you order policy last and thus statistically model effect of variables on policy.
 - Assuming regression captures all channels through which policy responds to developments in economy.
 - Omitted channels may lead to correlation between policy and outcomes.
- Ex: “Price puzzle” of why inflation rises with contractionary monetary shock.
 - One answer is Fed is forward looking and rises rates when it (correctly) anticipated inflation.
 - VAR omits variables that predict this inflation.

2. Parameter instability.

3. Timing assumptions do not reflect real-time data availability, causing misspecification.

Other SVAR Identification Schemes

- VARs criticized for being too reduced form.
- *Structural VAR (SVAR) approach* uses economic theory rather than Cholesky decomposition to invert reduced form VAR to structural VAR (that is, to recover structural innovations from reduced form residuals).
 - Must impose $(n^2 - n) / 2$ restrictions.
- Examples:
 - Gali (1999) splits Solow residual into tech and non-tech shocks by assuming that only tech shocks affect long-run productivity.
 - Sign restrictions (Uhlig).
 - Assuming cross-sectional regression holds (e.g., Beraja, Hurst, and Ospina, 2016).

Local Projections

- Jorda (2005) proposes more robust estimate of impulse response by local projection.
- Assume you have a shock S_t and an outcome Y_t . Idea is to directly and non-parametrically estimate the impulse response:

$$Y_{t+h} = \beta_0 + \beta_h S_t + \gamma X_t + \varepsilon_t^h$$

- Then the β_h s for $h = 0, \dots, T$ is the estimated T -period impulse response.
- Controls X_t often include lags of outcome and shock to control for autocorrelation and anticipatory effects.
- Local projections increasingly popular.
 - Less extrapolation of model at long horizons. More robust to misspecification, which compounds over time.
 - But much wider standard errors. Bias-variance tradeoff between LP and SVAR (Plagborg-Møller and Wolf, 2021; Li, Plagborg-Møller, and Wolf, 2024).
 - See Plagborg-Møller et al. Macro Annual for best practices.

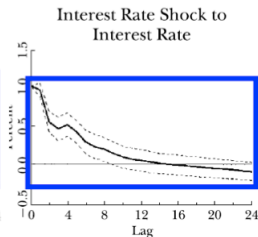
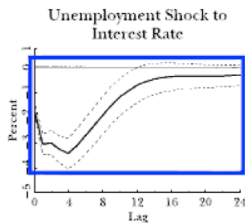
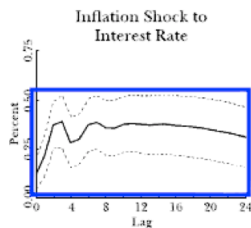
VAR Evidence on Non-Neutrality

- Apply VARs to study the effects of monetary shocks.
- Key challenge is endogeneity.
 - Changes in monetary policy occur for good reasons.
 - Error term ε_t correlated with outcome:

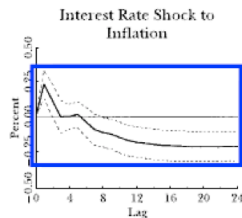
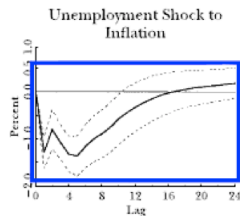
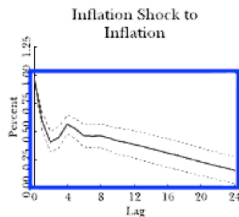
$$\Delta y_t = \alpha + \beta \Delta i_t + \varepsilon_t$$

- Start with simple VAR from Stock and Watson (2001).
 - 3 variables: inflation, unemployment, and Federal Funds interest rate.
 - Order π_t , u_t , R_t in recursive VAR.
 - π_t affects u_t and R_t contemporaneously but not vice-versa.
 - u_t affects R_t contemporaneously but not vice-versa.
 - See paper for data description (FEVD, Granger Causality tests).

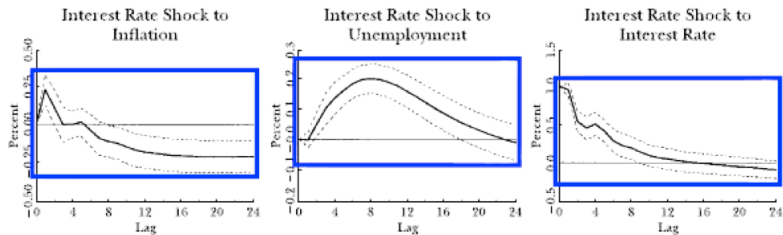
IRFs: Taylor Rule



IRFs: Phillips Curve

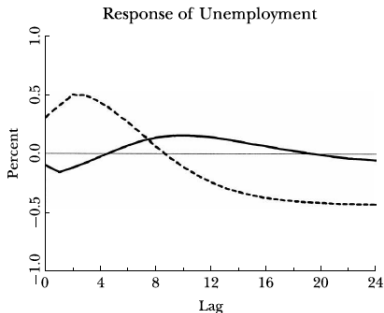
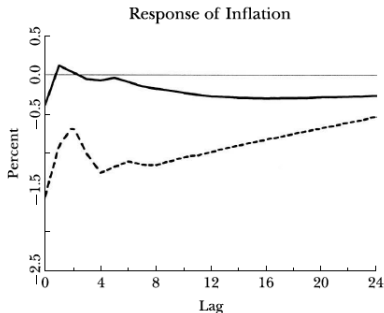


IRFs: Monetary Non-Neutrality



IRFs: Two Structural Approaches

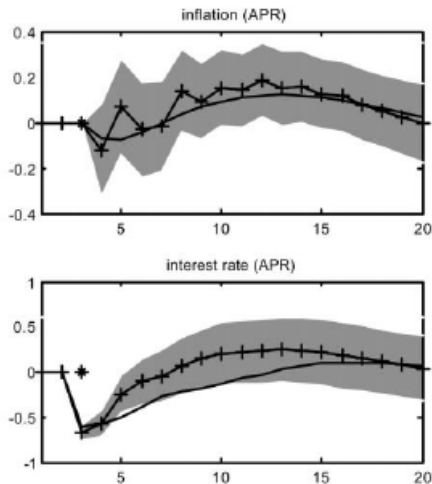
- Stock and Watson then use a structural VAR in which they impose a Taylor rule for identification rather than recursive VAR.
 - In solid: backward-looking Taylor rule.
 - In dashed: forward-looking Taylor rule.
 - Structural assumptions (and hence ordering) not innocuous.



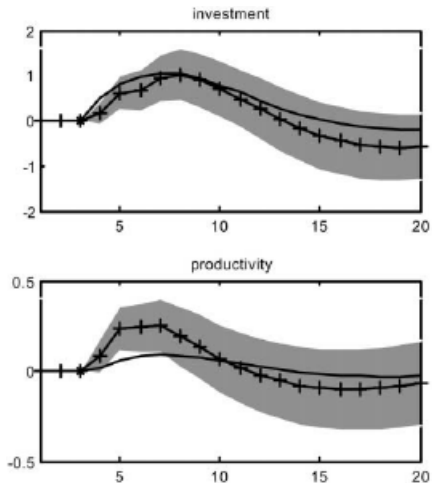
Christiano, Eichenbaum, and Evans (2005)

- Paper does a lot. For now just focus on VAR evidence.
- Run an 9-variable VAR. Ordering:
 1. Real GDP
 2. Real Consumption
 3. GDP Deflator
 4. Real Investment
 5. Real Wage
 6. Labor Productivity
 7. Federal Funds Rate
 8. Real Profits
 9. M2 Growth
- Economic conditions can affect monetary policy, but monetary policy only affects economic conditions with a lag.
 - Trying to get around endogeneity of monetary policy by statistically modeling it. But still Stock-Watson concerns.

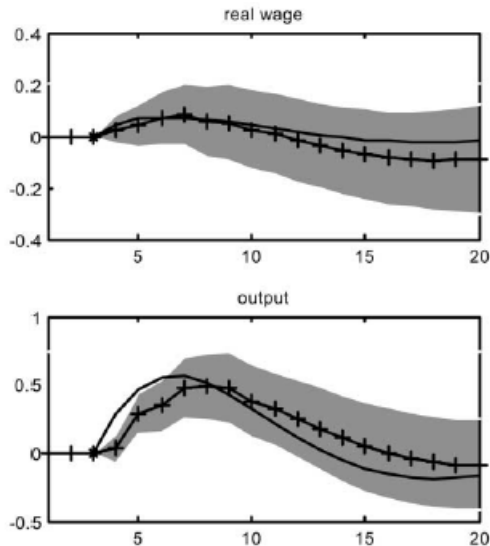
CEE: IRF To Money Shock



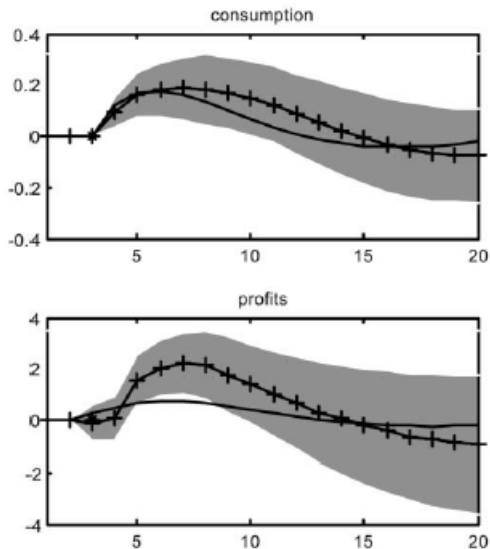
CEE: IRF To Money Shock



CEE: IRF To Money Shock



CEE: IRF To Money Shock



Christiano, Eichenbaum and Evans (2005) Summary

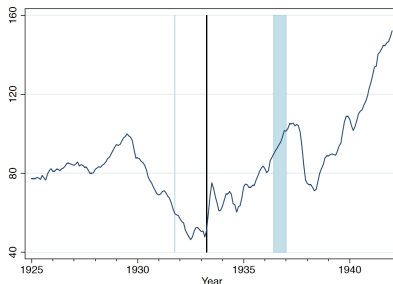
1. Hump-shaped response of output, consumption and investment, peaking at $1\frac{1}{2}$ years and returning to trend after 3.
 2. Hump-shaped response of inflation, peaking after two years.
 3. Interest rate falls for one year.
 4. Real profits, wages, and labor prod rise.
 5. Growth rate of money rises immediately.
- Phillips curve and Taylor rule as in Stock and Watson still hold.
 - Consistent with significant *monetary non-neutrality*
⇒ money affects real outcomes.

Other Approaches

- We may not like recursive VAR approach to identifying monetary shocks.
 - Four other approaches that try to deal with causality more directly of which I want you to be aware before moving to natural experiments.
1. Large Shocks
 2. Discontinuity-Based Approach
 3. Narrative Approach
 4. High Frequency Approach
- Good summary: Section 4 of Nakamura and Steinsson (2018) "Identification in Macroeconomics."

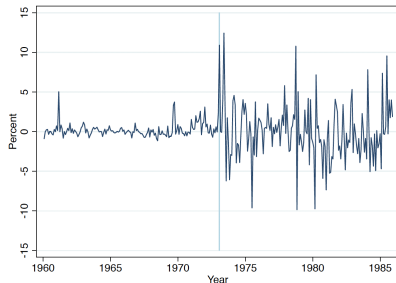
Large Shocks: Friedman and Schwartz (1963)

- Friedman and Schwartz (1963) famously argue that Fed made Great Depression worse.
- Focus on policy actions that are “of major magnitude,” not caused by other developments, sharp results that they compare to science experiment.
 - But others have questioned since.



Breakdown of Bretton Woods: Mussa (1986)

- In 1973 Bretton Woods fixed exchange rate system breaks.
 - Discontinuous and purely monetary change.
 - If money is neutral, should not affect real variables like real exchange rates.
- Monthly change in real Mark-Dollar exchange rate:



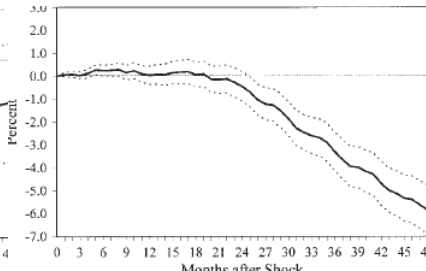
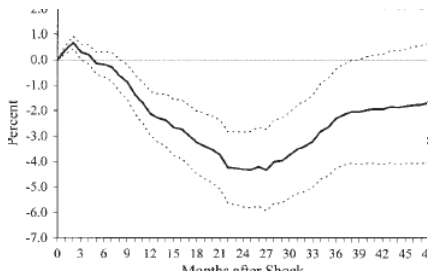
- But Itskhoki and Mukhin (2022) argue this is not evidence of monetary non-neutrality but instead financial frictions in foreign exchange markets.

Narrative Approach

- Narrative approach of Romer and Romer (1989) updated by Romer and Romer (2023).
 - Identify exogenous monetary shocks by using historical record.
 - Go through meeting transcripts, historical material, etc. to find a change in monetary policy unrelated to state of the economy e.g. a change in the Fed's preferences.
- Examples:
 - In December 1988, change view of what level of inflation is acceptable and raise rates.
 - January 1972 think unemployment settled too high and lower.
 - No monetary shocks 1988-2016! Only one expansionary shock in the entire series!
- Impulse responses to these “exogenous” policy dates show non-neutrality.

Narrative Approach

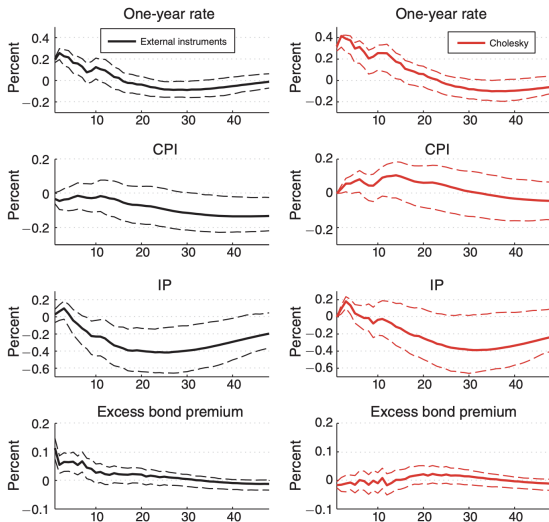
- More quantitative method in Romer and Romer (2004):
 - Determine intended FFR at meeting based on Fed's internal staff "Greenbook" forecast.
 - Regression controlling for level and change in forecasts of output, inflation, and unemployment.
 - Difference from FFR agreed upon at meeting to obtain shock.
- IRFs show non-neutrality:



High-Frequency Approach

- Shock series based on response of Fed funds futures in short window around Fed announcements.
 - Captures verbal communication in addition to rate.
 - Gürkaynak-Sack-Swanson (2005): Forward rates at long horizons affected by MP contrary to standard models.
 - Would like to look at responses of macro variables to show non-neutrality, but those are not at daily frequency!
- Two solutions to get at non-neutrality:
 - Gertler and Karadi (2015) time aggregate high frequency shocks and use as external instruments in VAR with low-frequency outcomes (e.g. inflation and output).
 - Nakamura and Steinsson (2018) compare responses of high frequency nominal and inflation-indexed Treasuries to separate real interest rate response and inflation expectations.
 - Issue with high frequency approach: Shocks are very small, lack of precision in local projection without VAR structure.

High-Frequency Approach: Gertler-Karadi (2015)



High-Frequency Approach Finds Non-Neutrality

- Gertler-Karadi (2015): Non neutrality, no price puzzle, credit spreads respond to monetary policy.
- Nakamura-Steinsson (2018)
 - Monetary shocks have large and persistent effects on real interest rates.
 - Monetary shocks have small effects on expected inflation at short horizons (< 1 year) and grows to a large effect over 2-3 years (hump shaped response).
 - Argue results imply that Fed announcements provide information that affects beliefs about economic fundamentals beyond interest rates.
- Which do we prefer?
 - Some issues of time aggregation here. N-S is more convincing for real interest rates and inflation expectations.
 - But sometimes we want to look at outcomes that are not as high frequency (e.g. output, realized inflation, credit spreads) and have to bite the time aggregation bullet.

Natural Experiments

- Ideal Evidence: Experiment where randomly change money supply in some places.
- Problem: Central banks are run by economists.
 - Changes in money supply are not random!
- Solution: Natural experiments. Examples:
 - Hyperinflations: inflation tracks money supply.
 - U.S. Great Depression (Friedman and Schwartz, 1963).
 - Gold Standard and Great Depression (Eichengreen and Sachs, 1985).
 - Breakdown of Bretton Woods (Mussa, 1986).
 - Volcker disinflation in early 1980s.

Our Problem: Economics PhDs (and Their Friends)



- We could really use an idiot in charge of the money supply.

Notre Idiot: Louis Henri, Duke of Bourbon



Notre Idiot: Louis Henri, Duke of Bourbon



- Wikipedia: Surviving descriptions of the duke's personality are highly uncomplimentary. They fall under the general categories of greed, bad manners, and stupidity.
 - "He had a very limited mind, knows nothing, and only likes pleasure and hunting."
 - "Lacks the necessary experience."
 - "Immensely inferior in talent."

Louis Henri: Prime Minister 1723-1726

- Why is Louis Henri in charge?

Louis Henri: Prime Minister 1723-1726

- Why is Louis Henri in charge?
- Because Louis XV is 13!



Louis Henri: Prime Minister 1723-1726

- What is Louis Henri obsessed with?

Louis Henri: Prime Minister 1723-1726

- What is Louis Henri obsessed with?
- Making heirs!

Louis Henri: Prime Minister 1723-1726

- What is Louis Henri obsessed with?
- Making heirs!
 - Louis is engaged to his 5 year old cousin the Infanta of Spain, and that means no heirs for too long.



Louis Henri: Prime Minister 1723-1726

- What is Louis Henri obsessed with?
- Making heirs!
 - Louis is engaged to his 5 year old cousin the Infanta of Spain, and that means no heirs for too long.

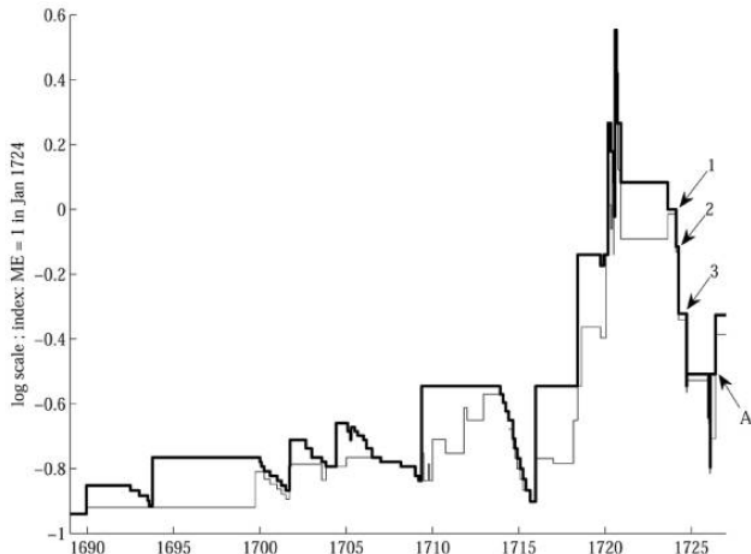


- He is also engaged in a favorite pastime: Fighting Protestants.

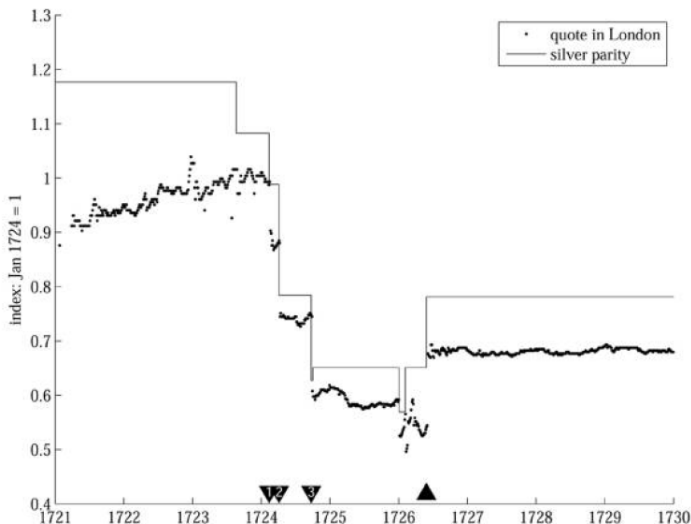
France in 1724: A Surprise 45% Deflation

- Money: coins with no face value.
 - Government sets nominal value by decree, can change it overnight and without warning.
- Velde (2009) examines an episode where three times in 1724, French cut value of currency overnight by a cumulative 45%.
 - Ex: September 22, 1724 at 8am, all 5 livre coins are now 4 livre coins.
 - “The high price level reduced the real value of soldiers’ wages and harmed government creditors.”
 - Why? Because the King and his ministers wanted to and there were no economists to advise them otherwise.
 - Revalue some in 1726.
- Expectations:
 - Had done before, but always fast inflations and gradual deflations.
 - Velde argues these three deflations were “unforetold.” Kept secret to reduce capital losses by state.

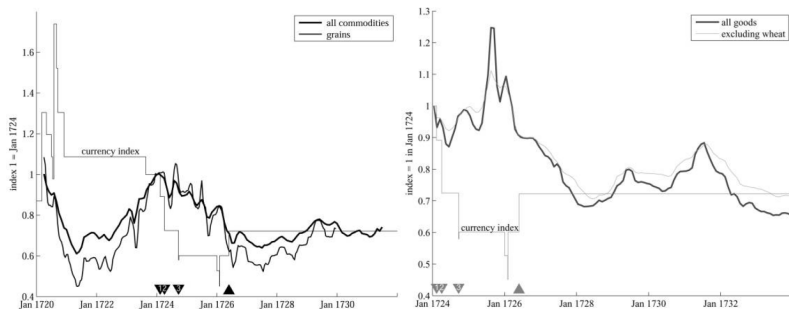
Value of a Coin



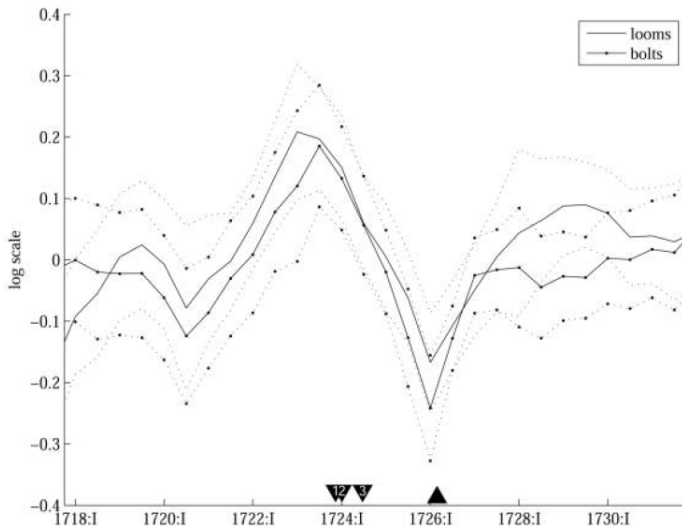
Foreign Exchange Prices Adjust Instantaneously



Commodities and Goods Prices Fall Slowly



Industrial Sector Contracts 30%



Summary: Strong Evidence of Non-Neutrality

- Introduced tool of VAR.
 - Recursive and structural.
 - Discussed assumptions, flaws, and benefits.
- Looked at VAR, narrative, high-frequency, and natural experiment evidence for monetary non-neutrality.
 - Strong evidence that money is non-neutral: it has effects on real economy.
 - Strong evidence of hump-shaped inflation responses.
- Next class: Introduce money and add it to RBC framework.
 - Can it explain the facts presented here?
 - Read Gali Ch. 2.