Lecture 2: Time series IV regression

Geert Mesters (Universitat Pompeu Fabra, Barcelona GSE and VU Amsterdam)

Motivation

Motivation

- In this set of slides we revisit instrumental variable regression from a time series perspective
- This is mainly motivated by the recent development of new instruments in macroeconomics
- For instance we now have observable proxies for oil price shocks, monetary policy shocks, fiscal policy shocks, productivity shocks and more, which can all be used as instrumental variables to solve endogeneity problems
- This is a very exciting new area of research that essentially relies on old methodology

Motivation

- Different from cross-sectional IV literature, the macro time series setting has some peculiarities that deserve attention
  - very short samples
  - dependent variables
  - often weak instruments
- Some key papers include: Kuttner (2001), Romer & Romer (2004), Jorda (2005), Gurkaynak, Sack & Swanson (2005), Mertens & Ravn (2013,2015), Gerler & Karadi (2015), Ramey & Zubairy (2018), Stock & Watson (2018), Plagborg-Moller & Wolf (2018), Olea, Stock & Watson (2019)

Motivating example

Effects of monetary policy

- How large are the effects of monetary policy on the economy?

Effects of monetary policy

- How large are the effects of monetary policy on the economy?
- For who is this important?
  - Central banks who needs to achieve their mandates (e.g. price stability and maximum employment)
  - Producers/consumers/everybody who need to make investment decisions

Effects of monetary policy

Using lecture 1 we could consider

$$M_{t+h} = F_t \beta_h^{MP} + \epsilon_t$$

where M_t is some macro variable and F_t is the Fed funds rate

Effects of monetary policy

Using lecture 1 we could consider

$$M_{t+h} = F_t \beta_h^{MP} + \epsilon_t$$

where M_t is some macro variable and F_t is the Fed funds rate

But

- Monetary policy is endogenous
- $\mathbb{E}(F_t \epsilon_t) \neq 0$
- Fed responds to changes in economy
- $\hat{\beta}_h^{MP}$ is not consistent for the true parameter !!!

Modern solution: High Frequency Identification

- Idea: exploit changes in interest rates futures around the FOMC[1]

$\Delta f_t =$ Change in fed funds futures in 30 min around FOMC

- If no other shocks hit the economy in these 30 min: then the "surprises" are due to monetary policy
- Seems like a clean identification strategy

[1] See Kuttner (2001), Gertler & Karadi (2015), Nakamura & Steinson (2018)

Modern solution: High Frequency Identification

Implies

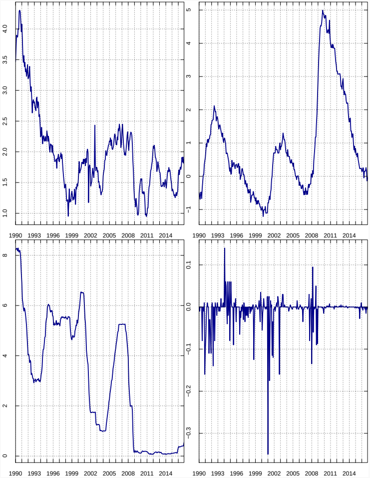
$$M_{t+h} = \Delta f_t \beta_h^{MPS} + \text{rest}$$

where M_{t+h} is the macro economic variable of interest (h periods ahead)

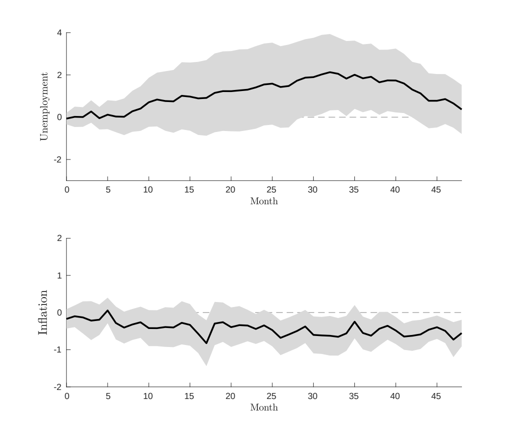
- β_h^{MPS} captures the "effect" of the monetary policy surprise
- We may consider this regression for $h = 0, 1, 2, \ldots, H$, to obtain the impulse response function $\beta_0^{MPS}, \ldots, \beta_H^{MPS}$

Data series

Inflation, Unemployment gap,
Fed funds rate, Fed funds futures surprises



Answer



Great !!! unemployment increases and inflation decreases

Modern solution: High Frequency Identification

$$M_{t+h} = \Delta f_t \beta_h^{MPS} + \text{rest}$$

Why do I not look happy?

Modern solution: High Frequency Identification

$$M_{t+h} = \Delta f_t \beta_h^{MPS} + \text{rest}$$

Why do I not look happy?

- We didn't answer the right question
- We typically do not care about monetary policy *surprises*
- We care about monetary policy β_h^{MP}
- Idea: let's use Δf_t as an instrumental variable to identify β_h^{MP}

IV regression

Endogenous regression

In general we consider

$$Y_t = X_t' \beta_0 + \epsilon_t$$

where again

- Y_t outcome variable
- $X_t = (X_{1\ ,t}, \ldots, X_{p\ ,t})'$ potentially endogenous regressors
- ϵ_t disturbance

Endogenous regression

In general we consider

$$Y_t = X_t'\beta_0 + \epsilon_t$$

where again

- Y_t outcome variable
- $X_t = (X_{1\ ,t}, \dots, X_{p\ ,t})'$ potentially endogenous regressors
- ϵ_t disturbance

Key difference: $\mathbb{E}(X_t\epsilon_t) \neq 0$

Sources of endogeneity

Why does $\mathbb{E}(X_t \epsilon_t) \neq 0$ happen?

- Simultaneity
- Omitted variables
- Measurement error

Instrumental variables

To solve the problem we need

- $L \times 1$ vector of instrumental variables Z_t,
- $L \geq p$ more instruments then endogenous regressors

Instrumental variables

To solve the problem we need

- $L \times 1$ vector of instrumental variables Z_t,
- $L \geq p$ more instruments then endogenous regressors

Key Assumptions

1. $\mathbb{E}(Z_t \epsilon_t) = 0$ for all t (Exogenous)
2. $\lim_{T \to \infty} \sum_{t=1}^{T} \mathbb{E}(Z_t X_t') / T$ full column rank (Relevant)

Instrumental variables

To solve the problem we need

- $L \times 1$ vector of instrumental variables Z_t,
- $L \geq p$ more instruments then endogenous regressors

Key Assumptions

1. $\mathbb{E}(Z_t \epsilon_t) = 0$ for all t (Exogenous)
2. $\lim_{T \to \infty} \sum_{t=1}^{T} \mathbb{E}(Z_t X_t') / T$ full column rank (Relevant)

In our monetary policy example

- $\mathbb{E}(\Delta f_t \epsilon_t) = 0$, why?
- $\mathbb{E}(F_t \Delta f_t) \neq 0$, why?

Deriving the GMM estimator

Simple method of moment approach would suggest

$$\mathbb{E}(Z_t \epsilon_t) = 0 \qquad \to \qquad \underbrace{\frac{1}{T}\sum_{t=1}^{T} Z_t(Y_t - X_t'\beta)}_{\bar{g}_T} = 0$$

If $L = p$ we can solve to get

$$\hat{\beta}^{IV} = \left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} Z_t Y_t$$

the simple IV estimator

Deriving the GMM estimator

If $L > p$ we have that $\bar{g}_T = 0$ has more equations (L) then unknown variables (p) $\rightarrow$ no solution, instead we solve

$$\hat{\beta}(\hat{W}) = \arg\min_{\beta \in \mathbb{R}^p} J(\beta; \hat{W}) \qquad J(\beta; \hat{W}) = \bar{g}_T' \hat{W} \bar{g}_T$$

where $\hat{W}$ is positive definite weight matrix.

Solving gives Generalized Method of Moments estimator:

$$\hat{\beta}(\hat{W}) = (S_{zx}' \hat{W} S_{zx})^{-1} S_{zx}' \hat{W} s_{zy}$$

where

$$S_{zx} = \frac{1}{T} \sum_{t=1}^{T} Z_t X_t' \qquad s_{zy} = \frac{1}{T} \sum_{t=1}^{T} Z_t Y_t$$

The GMM estimator

$$\hat{\beta}(\hat{W}) = (S'_{zx}\hat{W}S_{zx})^{-1}S'_{zx}\hat{W}s_{zy}$$

- Different choices of $\hat{W}$ can be chosen, e.g.
  - $\hat{W} = I_L$
  - $\hat{W} = S_{zz}^{-1}$, with $S_{zz} = \frac{1}{T}\sum_{t=1}^{T} Z_t Z_t'$ gives 2SLS
  - $\hat{W} = \hat{V}^{-1}$, with $\hat{V} \xrightarrow{p} V$ where $V = \lim_{T\to\infty} \mathbb{V}ar(T^{-1/2}\sum_{t=1}^{T} Z_t\epsilon_t)$ gives Efficient GMM
- Importantly, given exogeneity and relevance of the instruments any $\hat{W}$ will lead to a consistent estimator, but the differences in efficiency can be large

Side note on the History of IV

Side note on history

- Before we go into the details of the estimator we briefly look at the history of IV estimation
- Surprisingly, the first detailed treatment of instrumental variable regression was found in 1928 in an appendix of a book on oils

A classic in econometrics:

the problem of identifying and estimating one or more parameters of a system of simultaneous equations

First appeared in Phillip Wright's book review of Henry Moore's *Economic Cycles: Their Law and Cause* (e.g. Moore 1914, Wright, 1915)

Moore claimed to have documented an upward sloping demand curve. Wright argued that this could just as well be a supply curve traced out by a shifting demand curve

History of IV

Identification problem

Consider

$$q_t^d = \alpha_0 + \alpha_1 p_t + u_t \qquad \text{demand equation}$$

$$q_t^s = \beta_0 + \beta_1 p_t + v_t \qquad \text{supply equation}$$

$$q_t^d = q_t^s \qquad \text{equilibrium}$$

The shocks u_t and v_t are mean zero and shift the demand and supply imposed price-quantity planes

If we impose that $q_t^d = q_t^s = q_t$

$$q_t = \alpha_0 + \alpha_1 p_t + u_t \qquad \text{demand equation}$$

$$q_t = \beta_0 + \beta_1 p_t + v_t \qquad \text{supply equation}$$

Now to estimate we need

- for α_1 consistent $\mathrm{E}(p_t u_t)$ must be equal to zero
- for β_1 consistent $\mathrm{E}(p_t v_t)$ must be equal to zero

But when solving the equations for p_t and q_t we find

$$
\begin{aligned}
p_t &= \tfrac{\beta_0-\alpha_0}{\alpha_1-\beta_1} + \tfrac{v_t-u_t}{\alpha_1-\beta_1} \\
q_t &= \tfrac{\alpha_1\beta_0-\alpha_0\beta_1}{\alpha_1-\beta_1} + \tfrac{\alpha_1 v_t-\beta_1 u_t}{\alpha_1-\beta_1}
\end{aligned}
$$

p_t is a function of u_t and v_t and thus $\mathrm{E}(p_t u_t) \neq 0$ and $\mathrm{E}(p_t v_t) \neq 0$

History of IV

Summary identification problem

- Prices are a function for both demand and supply shifters
- Implies that slopes of demand and supply (α_1 and β_1) cannot be estimated consistently
- Phillip Wright was fully aware of this problem in 1915
- In 1928 Phillip Wright wrote a book that proposed the first solution to this problem

THE TARIFF ON ANIMAL AND VEGETABLE OILS

BY

PHILIP G. WRIGHT

WITH THE AID OF THE COUNCIL AND STAFF OF THE INSTITUTE OF ECONOMICS

History of IV

Content

CONTENTS

History of IV
Content

CHAPTER III

CHAPTER IV

CHAPTER V

CHAPTER VI

CHAPTER VII

History of IV

Content

CONTENTS xvii

PAGE

History of IV

Content

APPENDIX B

APPENDIX C

APPENDIX D

APPENDIX E

APPENDIX F

History of IV

Appendix B

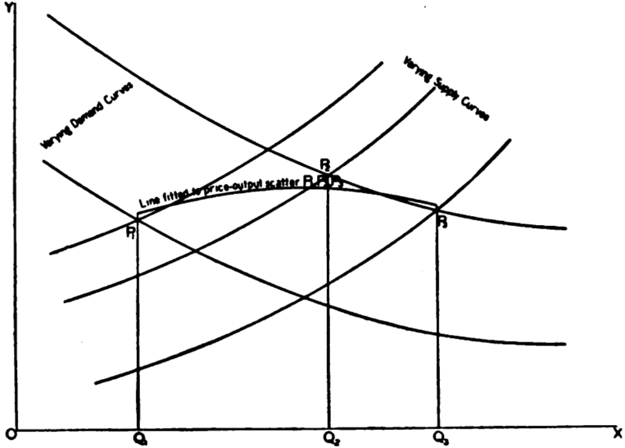
Out of nowhere appears Appendix B. It contains:

- A concise explanation on why data for prices and quantities alone are in general inadequate for estimating either supply or demand
- Two separate and correct derivations of instrumental variable estimators for supply and demand elasticities (limited and full information derivations)
- And an empirical application for butter and flaxseed

History of IV

The identification problem in Appendix B

FIGURE 4. PRICE-OUTPUT DATA FAIL TO REVEAL EITHER SUPPLY OR DEMAND CURVE.



History of IV
Solution to the identification problem in Appendix B

Page 311-312:

In the absence of intimate knowledge of demand and supply conditions, statistical methods for imputing fixity to one of the curves while the other changes its position must be based on the introduction of additional factors. Such additional factors may be factors which (A) affect demand conditions without affecting cost conditions or which (B) affect cost conditions without affecting demand conditions.

History of IV

Appendix B did exactly this

Exhibit 2

The Single-Equation Derivation of the Instrumental Variable Estimator in Appendix B

The derivation in Appendix B of the instrumental variable estimator of the coefficients of a single equation has two steps. The author tackled the supply curve first. Adopt his original notation, let O be the percentage deviation of output from its mean (then, as now, typically computed by taking the logarithm of the original quantity data, relative to its sample mean) and let P be the percentage deviation of price from its mean. Starting with the familiar supply and demand diagram, he first derived the supply curve with an additive disturbance,

$$O = eP + S_1,$$

where e is the elasticity of supply, S_1 represents the shift in the supply curve "brought about by a change in supply conditions," relative to when prices and output are at their long-run mean value, and the intercept is zero because the variables are deviated from their means. The author rearranges this expression as $eP = O - S_1$, then writes (p. 314):

> Now multiply each term in this equation by A (the corresponding deviation in the price of a substitute) and we shall have:
>
> $$eA \times P = A \times O - A \times S_1.$$
>
> Suppose this multiplication to be performed for every pair of price-output deviations and the results added, then:
>
> $$e \sum A \times P = \sum A \times O - \sum A \times S_1 \text{ or } e = \frac{\Sigma A \times O - \Sigma A \times S_1}{\Sigma A \times P}.$$
>
> But A was a factor which did not affect supply conditions; hence it is uncorrelated with S_1; hence $\Sigma A \times S_1 = 0$; and hence $e = (\Sigma A \times O)/(\Sigma A \times P)$.

(The shading has been added for the stylometric work carried out later.) The final expression for e is the formula for the instrumental variable estimator with a single instrument and a single included endogenous variable.

- The book by Phillip Wright (*The Tariff on Animal and Vegetable Oils*) can be found here: https://catalog.hathitrust.org/Record/001118057
- Stock and Trebbi (2003) provide more discussion: https://scholar.harvard.edu/files/stock/files/wr_5_w.pdf

Standard inference for GMM

Some theoretical results

The next slides discuss the properties of GMM estimator

$$\hat{\beta}(\hat{W}) = (S'_{zx}\hat{W}S_{zx})^{-1}S'_{zx}\hat{W}s_{zy}$$

- Consistency
- Asymptotic normality
- Variance estimation

The tools are the same as in Lecture 1 !!!

GMM Estimator - Consistency

Theorem

Assume that

1. $\{X_t, Z_t, \epsilon_t\}$ *is an* α*-mixing sequence of size* $-r/(r-1)$ *for* $r > 1$*;*
2. $\mathbb{E}|Z_{i\,t}\epsilon_t|^{r+\delta} < \Delta < \infty$ *and* $\mathbb{E}|X_{i\,t}Z_{j\,t}|^{r+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *for all* t *and* i, j*;*
3. $\mathbb{E}Z_t\epsilon_t = 0$ *for each* t*,* $\lim_{T\to\infty} \mathbb{E}Z'X/T \to \Sigma_{zx}$ *full column rank and* $\hat{W} \to W$ *with* W *positive definite*

Then

$$\hat{\beta}(\hat{W}) \xrightarrow{p} \beta_0 \ .$$

GMM Estimator - Asymptotic Normality

Theorem

1. $\{X_t, Z_t, \epsilon_t\}$ *is an* α*-mixing sequence of size* $-r/(r-2)$ *for* $r > 2$;
2. $\mathbb{E}|Z_{i\,t}\epsilon_t|^{r+\delta} < \Delta < \infty$ *and* $\mathbb{E}|X_{i\,t}Z_{j\,t}|^{(r/2)+\delta} < \Delta < \infty$ *for some* $\delta > 0$ *for all* t *and* i, j;
3. $\mathbb{E}Z_t\epsilon_t = 0$ *for each* t, $\lim_{T\to\infty} \mathbb{V}ar T^{-1/2} Z'\epsilon \to V$ *positive definite,* $\lim_{T\to\infty} \mathbb{E}Z'X/T \to \Sigma_{zx}$ *full column rank and* $\hat{W} \to W$ *with* W *positive definite*

Then

$$\sqrt{T} D^{-1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} \mathcal{N}(0, I_{p+1}) \ ,$$

where

$$D = (\Sigma_{zx}' W \Sigma_{zx})^{-1} \Sigma_{zx}' W V W \Sigma_{zx} (\Sigma_{zx}' W \Sigma_{zx})^{-1}$$

GMM Estimator - Efficiency

- The asymptotic variance depends on W
- Opportunity: choose W to get the smallest variance
- Take $W = V^{-1}$ which implies smallest variance $D = (\Sigma'_{zx} V^{-1} \Sigma_{zx})^{-1}$
- Issue: V can be estimated consistency using HAC (see lecture 1), but this requires residuals; and residuals require an estimator and ...

GMM Estimator - Efficiency

- The asymptotic variance depends on W
- Opportunity: choose W to get the smallest variance
- Take $W = V^{-1}$ which implies smallest variance $D = (\Sigma'_{zx} V^{-1} \Sigma_{zx})^{-1}$
- Issue: V can be estimated consistency using HAC (see lecture 1), but this requires residuals; and residuals require an estimator and ...

Solution: two-step GMM

1. Estimate β using 2SLS (e.g. use $\hat{W} = S_{zz}^{-1}$); compute HAC estimator $\hat{V}$
2. Estimate β using GMM with $\hat{W} = \hat{V}^{-1}$

GMM Estimator - Variance estimation

In general the Asymptotic variance of $\hat{\beta}(\hat{W})$ is given by

$$D = (\Sigma'_{zx} W \Sigma_{zx})^{-1} \Sigma'_{zx} W V W \Sigma_{zx} (\Sigma'_{zx} W \Sigma_{zx})^{-1}$$

A consistent estimator is given by

$$\hat{D} = (S'_{zx} \hat{W} S_{zx})^{-1} S'_{zx} \hat{W} \hat{V} \hat{W} S_{zx} (S'_{zx} \hat{W} S_{zx})^{-1}$$

where the Newey-West HAC estimator is given by

$$\hat{V} = \sum_{t=1}^{T} \hat{\epsilon}_t^2 Z_t Z_t' + \sum_{l=1}^{L} \sum_{t=1}^{T} w_l \, \hat{\epsilon}_t \hat{\epsilon}_{t-l} \left(Z_t Z_{t-l}' + Z_{t-l} Z_t' \right)$$

The problem of weak instruments

Instruments are often weak

- Very often the instruments that we use in macroeconomics do not explain a lot of variance in the endogeneous regressors
- This means that the correlation between X_t and Z_t is small
- The are lots of famous examples in the literature that have this feature, e.g.
  - Phillips curves (Mavroeidis, Plagborg-Moller & Stock 2014)
  - Elasticity of substitution (Yogo 2004)
  - Monetary policy rules (Mavroeidis 2010)
- The consequences of weak instruments are severe as we show in a Monte Carlo study

Monte Carlo Study

We generate data from the standard IV model (all scalars)

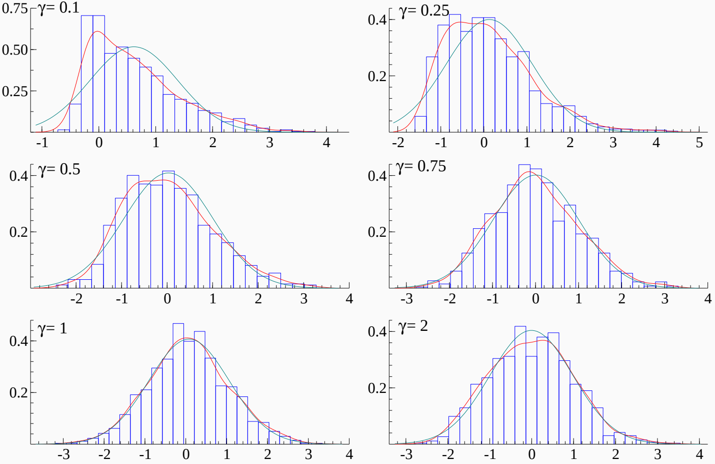
$$\begin{aligned} Y_t &= X_t\beta + \epsilon_t \\ X_t &= Z_t\gamma + \epsilon_t + \eta_t \end{aligned}$$

We take $\beta = 1$, $Z_t, \epsilon_t, \eta_t \sim NID(0,1)$ and vary

$$\gamma = 0.1, 0.25, 0.5, 0.75, 1, 2$$

- We show the empirical distribution of the t-statistics: $t = \sqrt{T}(\hat{\beta} - 1)/\sqrt{\hat{D}}$ for $S = 1000$ simulations
- Our theory immediately implies that $t \overset{a}{\sim} N(0,1)$ under $H_0 : \beta = 1$

Monte Carlo Study



For small γ the size of the test is completely wrong

Some comments

- The figure for the t-statistic was first shown in Nelson-Startz (1990 a,b)
- It implies that the standard normal approximation can be really bad when the correlation between X_t and Z_t is small
- Importantly, this is not a small sample issue!!!

Implications

There are roughly two ways to move forward

- Detect whether the correlation is large enough
  - if YES $\rightarrow$ using normal approximating distribution
  - if NO $\rightarrow$ find better instruments, or
- Use test statistics that do not depend on the correlation between Z_t and X_t; examples include the Anderson & Rubin (1949) statistics, the LM statistic of Kleibergen (2002) and the LR statistic of Moreira (2001)

Detecting weak instruments

It is convenient to have a way to decide if instruments are strong (IV works) or weak

Three options

1. IID homoskedastic errors: F-statistic from first stage regression of Z_t on X_t
2. IID heteroskedastic errors: robust F^R-statistic, sometimes called Kleibergen-Paap statistic
3. Serially correlated and heteroskedastic errors: effective F^E-statistic of Montiel Olea & Pflueger (2013)

Detecting weak instruments

In our time series settings we should always report Montiel Olea & Pflueger (2013) effective F statistic

For X_t scalar

$$F^E = \frac{\hat{\gamma}' S_{zz} \hat{\gamma}}{\text{Tr}\left(\hat{V}_{\gamma}^{1/2} S_{zz} \hat{V}_{\gamma}^{1/2'}\right)}$$

where $\hat{V}_{\gamma}$ is the variance estimate for $\hat{\gamma}$ (requires HAC)

- Rule of thumb: $F^E > 10$ proceed to use IV estimator with normal approximating limit
- Or compare F^E to Montiel Olea & Pflueger (2013) critical values (in Stata weakivtest.ado) and if the test is passed you can use the IV estimator

Weak instrument robust methods

- There are several statistics that do not depend on the strength of the instruments these can be used if $F^E < 10$
- Prominent examples include the Anderson-Rubin, LM and CLR statistics; note that these are all tests and not estimators
- They all exploit that under $H_0 : \beta = \beta_0$ we should have $\mathbb{E}(Z_t(y_t - X_t\beta_0)) = 0$; and this is testable !!!
- The review paper by Andrews, Stock & Sun (2019) provides a very nice introduction into this material

Effects of Monetary Policy

Effects of Monetary Policy

We revisit our empirical application

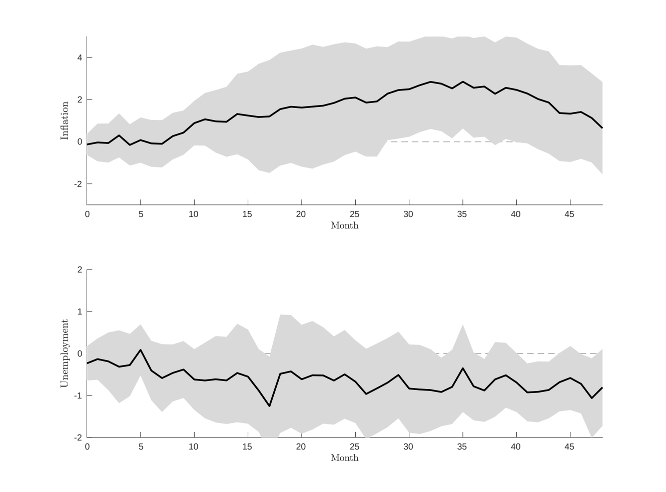
$$M_{t+h} = F_t \beta_h^{MP} + \text{controls} + \epsilon_t$$

where we now use Δf_t as an instrument and the controls include lags of observable macro variables (to soak up serial correlation)

First stage results

- For the sampling period $1990 - 2007$ we find that $F^E = 16.7$, so no problem
- However for $2007 - 2017$ we find that $F^E = 4.2$, so a big problem

Correct answer



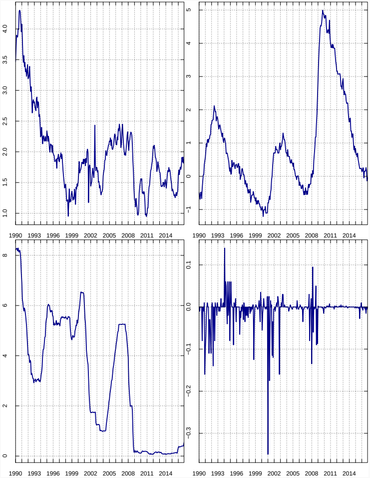
IRFs for 1990-2007; computed using Δf_t as an instrument

Some comments

- Using the monetary policy surprises as instruments is the correct thing to do !!!
- The identification strategy is much more credible when compared to the older literature that simply tries to control for omitted variables
- The current low interest rate environment poses challenges to this approach

We know what happened

Inflation, Unemployment gap,
Fed funds rate, Fed funds futures surprises



References & Material

Reference & Material

- References:
  H. White, "Asymptotic Theory for Econometricians", Chapters 1 to 6
  Andrews, Stock & Sun (2019), "Weak Instruments in IV Regression: Theory and Practice", forthcoming in Annual Review of Economics