Stata Spain Conference

Introduction to Local Projection Impulse-Response functions (IRFs) in Stata

Gustavo Sánchez

StataCorp LLC

May 22, 2025
Seville, Spain

Outline

1. Local projection impulse-response functions: Overview
2. IRFs with `lpirf`
    - Example with Covid data
    - Univariate estimation compared to AR(p)
    - Multivariate estimation comparing to VAR(p)
3. IRFs Instrumenting endogenous impulse variable with `ivlpirf`
    - Example with Stock market data
    - Example for unemployment rate

Why impulse-response functions and local projections

- **How does a change in x_t affect y_t over time?**

$$\frac{\partial y_{t+h}}{\partial x_t} = E[y_{t+h}|x_t = 1, \mathbf{w}_t] - E[y_{t+h}|x_t = 0, \mathbf{w_t}] = IRF_{x \to y}(h)$$

Overview

Overview: Sims (1980) "Incredible" identification

- IRFs have been widely used with VAR models.
- IRFs from VAR estimation depend on model specification.
- Sequence of transformations are needed to obtain IRFs.

Local projections proposed by Jordà (2005): convenient alternative to obtain IRFs

- Are based on univariate methods not constrained by a whole system.
- Impulse can be the outcome (say y1), another endogenous variable (say y2), or an exogenous variable (say x).

Local projections estimates IRFs from regressions at different time horizons (h=1,2,...,H) :

$$y_{i,t+h} = \theta_{ijh} y_{j,t} + \mathbf{Z_t}\delta + u_{t+h}$$

Where:

y_i is the response variable

y_j is the impulse variable

θ_{ijh} is the impulse-response coefficient

$\mathbf{Z_t}$ are control variables (e.g. further lags of endogenous variables)

δ are nuisance parameters.

Dynamic multipliers for exogenous variables in the local projection equation at horizons h=0,1,2,...,H-1:

$$y_{i,t+h} = \phi_{ikh} x_{k,t} + \mathbf{Z_t}\delta + u_{t+h}$$

Where:

y_i is the response variable
$x_{k,t}$ is the exogenous variable
ϕ_{ikh} is the dynamic multiplier coefficient
$\mathbf{Z_t}$ are control variables (e.g. further lags of endogenous variables)
δ are nuisance parameters.

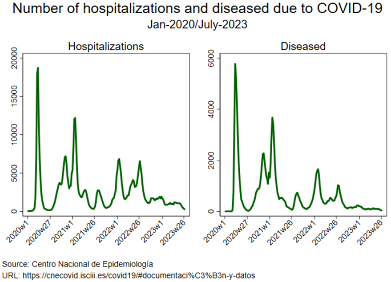
What does the local projections formulation imply? (taken from Jordà and Taylor (2024):

- "... linearity may appear restrictive."
- "... the error term is likely to be serially correlated."
- "... it can be extended to panel data settings."
- "... easier to implement useful nonlinear extensions."'

Example 1: Could we use IRFs to identify different stages for the Covid-19 pandemic in Spain?

- Covid-19 data (Centro Nacional de Epidemiología):

```
. describe
Contains data
 Observations:          183                     Covid-19 data for Spain
                                                  2020w1-2023w27
    Variables:            4
--------------------------------------------------------------------------------
Variable      Storage   Display     Value
    name         type    format     label      Variable label
--------------------------------------------------------------------------------
week            float   %tw
hospitalized    double  %10.0g                 Hospitalizations due to Covid-19
diseased        double  %10.0g                 Diseased due to Covid-19
trend           float   %9.0g
--------------------------------------------------------------------------------
Sorted by: week
     Note: Dataset has changed since last saved.
```



Let's first work with a univariate autorregressive model:

$$\Delta \mathit{diseased} = \alpha_1 + \beta_{L.\Delta \mathit{diseased}} * L.\Delta \mathit{diseased} + \mu_1$$

- We will obtain the impulse-response functions from:
  - an autoregressive model AR(2) fit with `arima`
  - a (univariate) VAR model with 2 lags with `var`
  - local projections with a second lag as a control variable `lpirf`

Fit univariate ARIMA model with `arima`

```
. arima D.diseased, ar(1/2) nolog
ARIMA regression
Sample: 2020w2 thru 2023w27                     Number of obs     =        182
                                                Wald chi2(2)      =    1227.39
Log likelihood = -1262.928                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
  D.diseased | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
diseased     |
       _cons |   .0264805   52.53416     0.00   1.000    -102.9386    102.9915
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |    .917799   .0275567    33.31   0.000     .8637889    .9718092
         L2. |  -.4564361   .0301977   -15.11   0.000    -.5156226   -.3972496
-------------+----------------------------------------------------------------
      /sigma |   249.0217   4.608127    54.04   0.000     239.9899    258.0534
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```

Use `irf` commands to get the IRF after `arima`

```
 . irf create myarimairf,set(myarimairf) replace
(file myarimairf.irf now active)
(file myarimairf.irf updated)

.
. irf table irf,impulse(D.diseased) response(D.diseased) noci
Results from myarimairf
```

Step	(1) irf
0	1
1	.917799
2	.385919
3	-.064721
4	-.235548
5	-.186645
6	-.06379
7	.026645
8	.053571

```
(1) irfname = myarimairf, impulse = D.diseased, and response = D.diseased.
```

Impulse-response functions after `arima`

IRF table

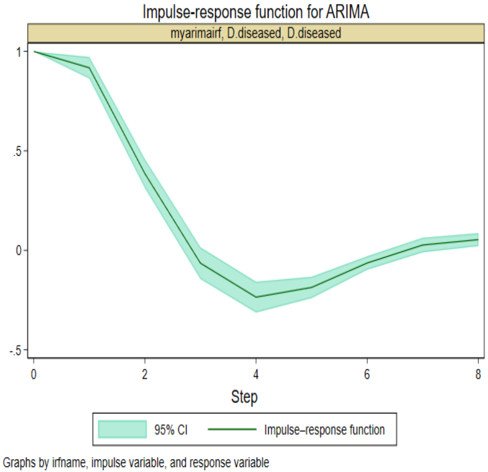
```
. irf table irf,          ///
>   impulse(D.diseased) ///
>   response(D.diseased) noci
```

Results from myarimairf

Step	(1) irf
0	1
1	.917799
2	.385919
3	-.064721
4	-.235548
5	-.186645
6	-.06379
7	.026645
8	.053571

(1) irfname = myarimairf, impuls

IRF graph



Fit univariate VAR model with `var`

```
. var D.diseased, lags(1/2)
Vector autoregression
Sample: 2020w4 thru 2023w27                    Number of obs     =        180
Log likelihood =  -1249.559                    AIC               =   13.91733
FPE            =   64825.36                    HQIC              =    13.9389
Det(Sigma_ml)  =   62699.94                    SBIC              =   13.97054
Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_diseased            3     252.513   0.5268   200.3559   0.0000
----------------------------------------------------------------

------------------------------------------------------------------------------
  D.diseased | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
D_diseased   |
    diseased |
         LD. |   .9228634   .0661313    13.96   0.000     .7932484    1.052478
        L2D. |  -.4613232   .0661315    -6.98   0.000    -.5909384   -.3317079
             |
       _cons |   .0234907   18.66369     0.00   0.999    -36.55667    36.60365
------------------------------------------------------------------------------
```

Use `irf` commands to get the IRF after `var`

```
. irf create myvarirf,set(myvarirf) replace
(file myvarirf.irf now active)
(file myvarirf.irf updated)

.
. irf table irf,impulse(D.diseased) response(D.diseased) noci
Results from myvarirf
```

Step	(1) irf
0	1
1	.922863
2	.390354
3	-.065495
4	-.240522
5	-.191755
6	-.066005
7	.027547
8	.055872

```
(1) irfname = myvarirf, impulse = D.diseased, and response = D.diseased.
```

Impulse-response functions after `var`

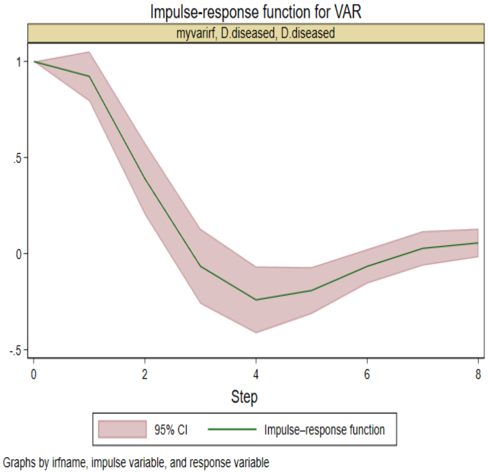
IRF table

```
. irf table irf,        ///
>   impulse(D.diseased) ///
>   response(D.diseased) noci
Results from myvarirf
```

Step	(1) irf
0	1
1	.922863
2	.390354
3	-.065495
4	-.240522
5	-.191755
6	-.066005
7	.027547
8	.055872

```
(1) irfname = myvarirf, impulse
```

IRF graph



Impulse-response with `lpirf`

- Use `lpirf` to get the impulse-response function with local projections

```
. lpirf D.diseased, lags(1/2)
Local-projection impulse responses
Sample: 2020w4 thru 2023w20                    Number of obs        = 173
                                               Number of impulses   =   1
                                               Number of responses  =   1
                                               Number of controls   =   1
```

	IRF coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
diseased						
FD.	.9230845	.0674477	13.69	0.000	.7908894	1.05528
F2D.	.379857	.0922328	4.12	0.000	.199084	.56063
F3D.	.0027866	.0944041	0.03	0.976	-.182242	.1878152
F4D.	-.1781756	.093747	-1.90	0.057	-.3619163	.005565
F5D.	-.2205025	.0946082	-2.33	0.020	-.4059311	-.0350739
F6D.	-.1665882	.0959982	-1.74	0.083	-.3547411	.0215648
F7D.	-.1695401	.0968207	-1.75	0.080	-.3593053	.020225
F8D.	-.1296116	.0975469	-1.33	0.184	-.3208	.0615768

```
Impulse:   D.diseased
Response:  D.diseased
Control:   L2D.diseased
```

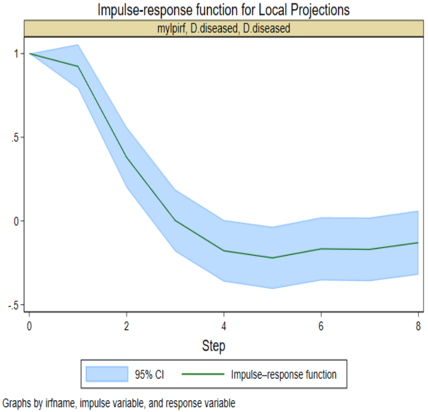
Impulse-response functions after `lpirf`

IRF table

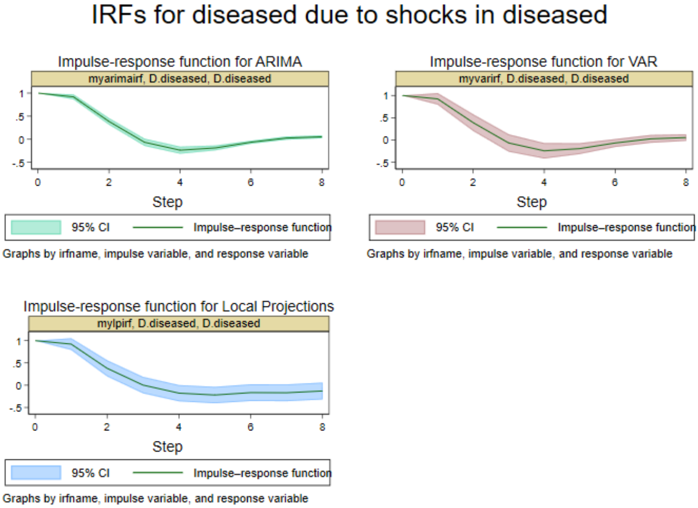
```
. quietly lpirf D.diseased, ///
>      lags(1/2)
. etable,cstat(_r_b)
```

	D.diseased
FD.diseased	0.923
F2D.diseased	0.380
F3D.diseased	0.003
F4D.diseased	−0.178
F5D.diseased	−0.221
F6D.diseased	−0.167
F7D.diseased	−0.170
F8D.diseased	−0.130
Number of observations	173

IRF graph



Impulse-response functions after `arima` `var` and `lpirf`



We will analyze the impact of shocks in the number of hospitalizations on the number of diseased patients.

- Let's consider the following linear model specification:

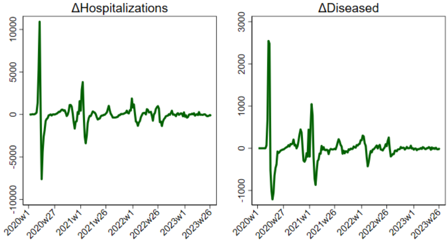
$$\Delta \textit{diseased} = \alpha_1 + \beta_{\Delta \textit{hospitalized}} * \Delta \textit{hospitalized} + \epsilon_1$$

Where:

diseased	: Monthly number of deseased due to Covid-19.
hospitalized	: Monthly number of hospitalized due to Covid-19.
Δ	: First difference

Number of hospitalizations and diseased due to COVID-19

Jan-2020/July-2023



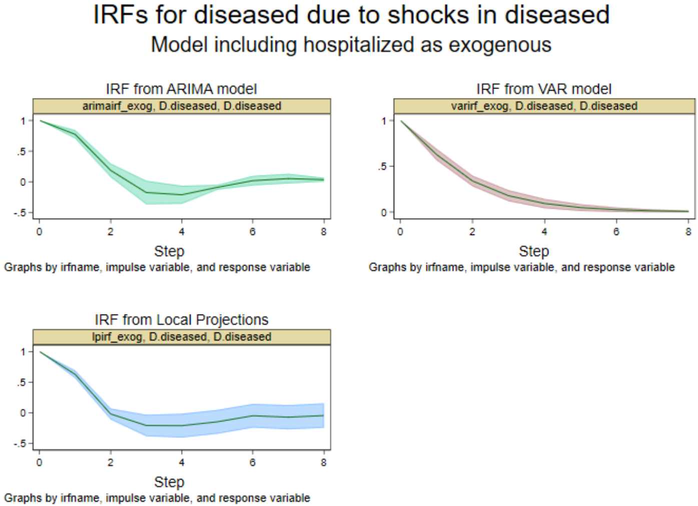
Source: Centro Nacional de Epidemiología

URL: https://cnecovid.isciii.es/covid19/#documentaci%C3%B3n-y-datos

Including hospitalized as an exogenous variable

```
 . /* ARIMA */
. quietly arima D.diseased D.hospitalized, ar(1/2)
. irf create arimairf_exog,set(arimairf_exog,replace)
. irf graph irf, impulse(D.diseased) response(D.diseased)      ///
>     title("IRF from ARIMA model") byopts(legend(off))        ///
>     ci ciopts(color(stgreen%30)) name(irf_arima_ex,replace)
.
. /* VAR */
. quietly var D.diseased, exog(D.hospitalized) lags(1/2)
. irf create varirf_exog, set(varirf_exog,replace)
. irf graph irf, impulse(D.diseased) response(D.diseased)      ///
>     title("IRF from VAR model") byopts(legend(off))          ///
>     ci ciopts(color(maroon%30)) name(irf_var_ex,replace)
.
. /* LOCAL PROJECTIONS */
. quietly lpirf D.diseased, exog(D.hospitalized) lags(1/2)
. irf create lpirf_exog, set(lpirf_exog,replace)
. irf graph irf, impulse(D.diseased) response(D.diseased)      ///
>     title("IRF from Local Projections") byopts(legend(off)) ///
>     ci ciopts(color(stblue%30)) name(irf_lpirf_ex,replace)
.
. graph combine irf_arima_ex irf_var_ex irf_lpirf_ex,          ///
>    title("IRFs for diseased due to shocks in diseased")      ///
>    subtitle("Model including hospitalized as exogenous")
```

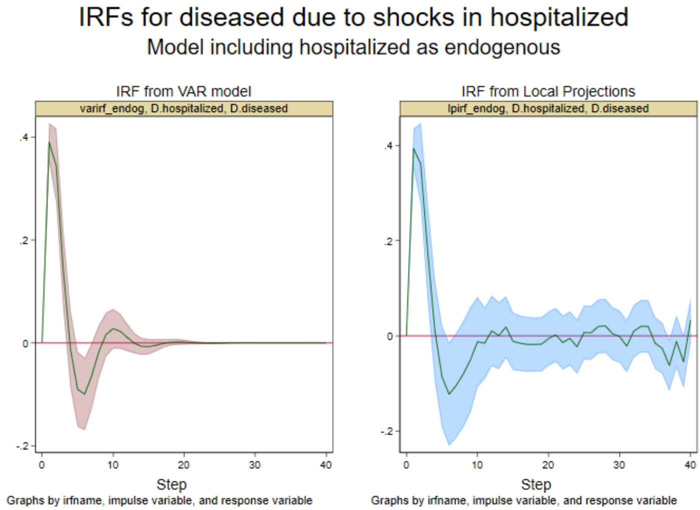
Impulse-response functions after `arima` `var` and `lpirf`



Including hospitalized as an endogenous variable

```
 . /* VAR */
. quietly var D.diseased D.hospitalized, lags(1/2)
. irf create varirf_endog, set(varirf_endog,replace) step(40)
. irf graph irf, impulse(D.hospitalized) response(D.diseased)   ///
>     title("IRF from VAR model") byopts(legend(off))           ///
>     ci ciopts(color(maroon%30)) name(irf_var_end,replace) yline(0)
.
.
. /* LOCAL PROJECTIONS */
. quietly lpirf D.diseased D.hospitalized, lags(1/2) step(40)
. irf create lpirf_endog, set(lpirf_endog,replace)
. irf graph irf, impulse(D.hospitalized) response(D.diseased)    ///
>   title("IRF from Local Projections") byopts(legend(off))      ///
>   ci ciopts(color(stblue%30)) name(irf_lpirf_end,replace) yline(0)
.
.
. graph combine irf_var_end irf_lpirf_end,                       ///
>   title("IRFs for diseased due to shocks in hospitalized")     ///
>   subtitle("Model including hospitalized as endogenous")
```

Impulse-response functions after `var` and `lpirf`

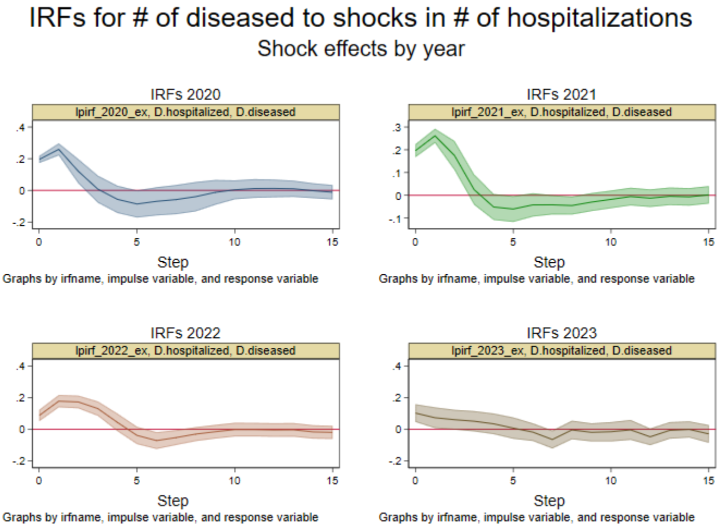


Including hospitalized as an exogenous variable

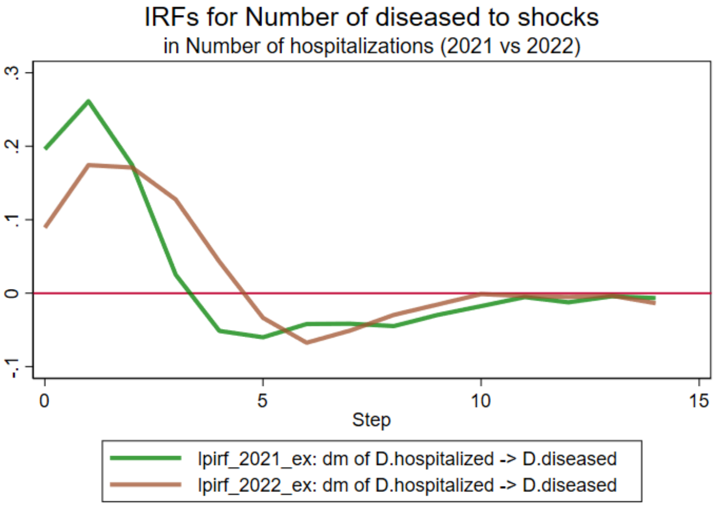
- Let's analyze the effects of the shocks at different stages of the pandemic.

```
 . lpirf D.diseased if tin(2020w1,2020w52), ///
>       exog(D.hospitalized) lags(1/2) step(16)
.
.
. lpirf D.diseased if tin(2021w1,2021w52), ///
>       exog(D.hospitalized) lags(1/2) step(16)
.
.
. lpirf D.diseased if tin(2022w1,2022w44), ///
>       exog(D.hospitalized) lags(1/2) step(16)
.
.
. lpirf D.diseased if tin(2022w45,2023w27), ///
>       exog(D.hospitalized) lags(1/2) step(16)
```

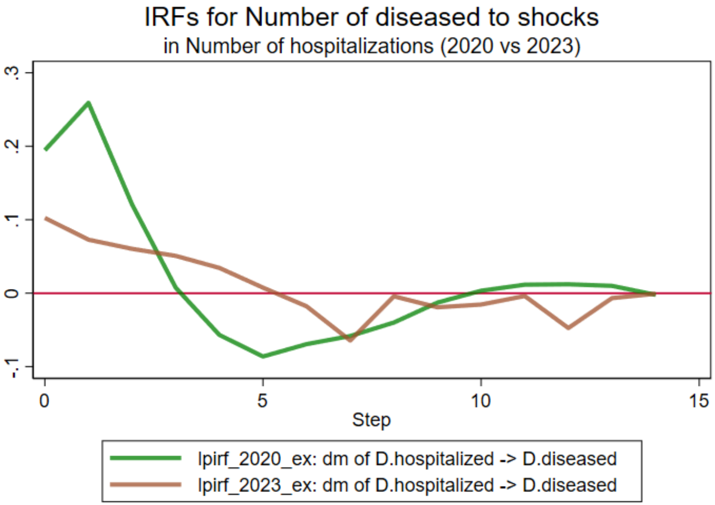
Stronger effect observed in early stages of the pandemic.



Compare now the IRFs during the second and third years of the pandemic



Compare now the IRFs during the first and fourth years of the pandemic



Local projections instrumenting endogenous impulse variable

By instrumenting the endogenous impulse variable, local projections can be interpreted as causal IRFs

- We look for exogenous variables that can be used as instruments for the endogenous impulse variable
- The estimation is performed using GMM with the moment conditions:

$$E[(y_{i,t+h} - \beta_h x_t + \gamma_h^{'} \mathbf{w_t})\mathbf{z_t}] = \mathbf{0}$$

Syntax for `ivlpirf`

`ivlpirf` depvars [if] [in] [, options]

Important options:

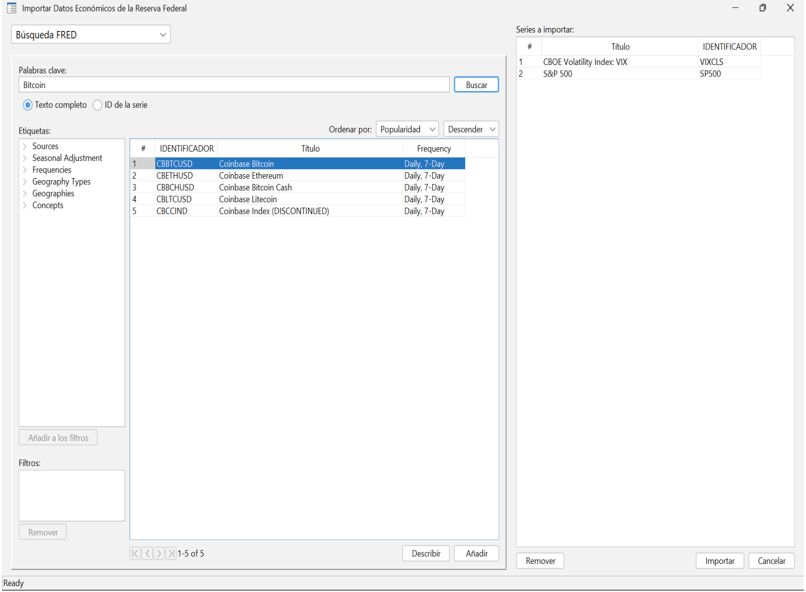
- endog(endovar = instvar) ; only one endogenous variable
- lags(numlist) - las of depvars and endovar
- exog(varlist) - control variables
- step() number of steps for IRFs
- cumulative - request cumulative coefficients

Example 2: Is Bitcoin affected by the stock market?

- Fit linear model for Bitcoin price on S&P500.
- Instrument impulse variable (S&P500) with the Chicago Board Options Exchange Volatility Index (VIX).
- We use `import fred` to load the data:

```
. import fred SP500  CBBTCUSD VIXCLS,     ///
>     daterange(2017-01-03 2025-04-30)   ///
>     aggregate(daily) clear
. rename SP500      sp500
. rename VIXCLS     vix
. rename CBBTCUSD   bitcoin
. tsset daten,daily
. generate bdate=bofd("bcalapr30",daten)
. format bdate %tbbcalapr30
. tsset bdate
. keep if bdate~=.
. tsappend,add(3)
. tsset bdate
```

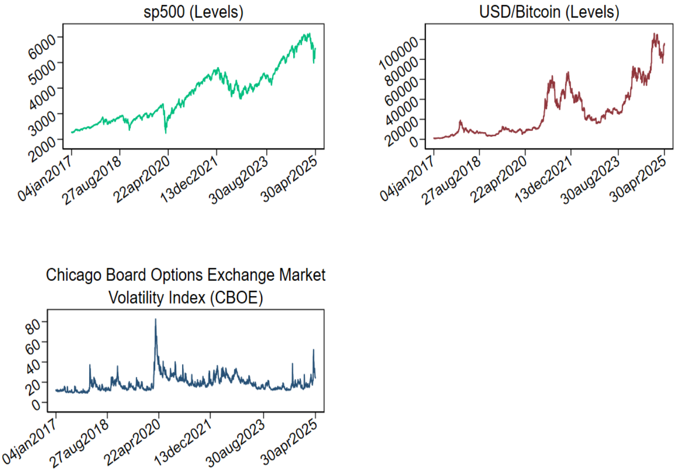
`import fred`: Dialog box



Data for Bitcoin price, S&P500, and VIX

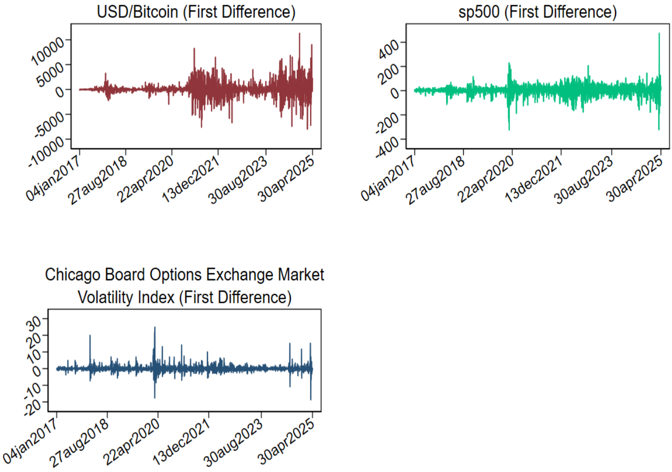
```
. describe
Contains data
 Observations:         2,097
    Variables:             5
--------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
--------------------------------------------------------------------
bdate           float   %tb..
daten           int     %td                   numeric (daily) date
sp500           float   %9.0g                 S&P 500
bitcoin         float   %9.0g                 Coinbase Bitcoin
vix             float   %9.0g                 CBOE Volatility Index
--------------------------------------------------------------------
Sorted by: bdate
     Note: Dataset has changed since last saved.
```

Levels of bitcoin price, S&P500, and VIX



Source: Federal Reserve Economic Data (FRED)
Downloaded using -import fred- in Stata

Bitcoin price, S&P500, and VIX (First difference)



Source: Federal Reserve Economic Data (FRED)

Downloaded using -import fred- in Stata

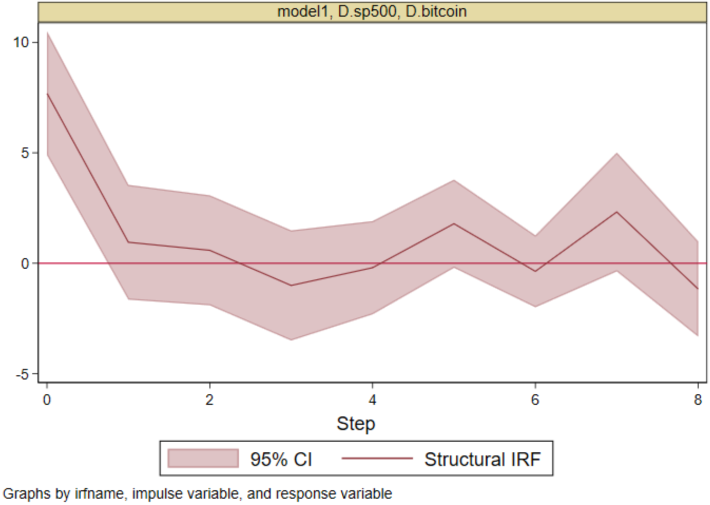
Specify the instrument in "endog()" option

```
. ivlpirf D.bitcoin, endog(D.sp500 = D.vix) step(4) nolog
Final GMM criterion Q(b) = 4.69e-35
note: model is exactly identified.
Instrumental-variables local-projection impulse responses
Sample: 06jan2017 thru 24apr2025, but with gaps
                                                    Number of obs = 2,071
 ( 1)  [D.sp500]D.sp500 = 1
------------------------------------------------------------------------------
             |      IRF      Robust
             | coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     bitcoin |
         D1. |    7.37798   1.431555     5.15   0.000     4.572184    10.18378
         FD. |   .9250945   1.312917     0.70   0.481    -1.648176    3.498365
        F2D. |   .5219293   1.255784     0.42   0.678    -1.939363    2.983221
        F3D. |  -1.031773   1.255343    -0.82   0.411    -3.492201    1.428654
        F4D. |  -.1344953   1.068617    -0.13   0.900    -2.228945    1.959955
             |
       sp500 |
         D1. |          1  (constrained)
         FD. |  -.1409887   .0647379    -2.18   0.029    -.2678727   -.0141047
        F2D. |   .1173217   .0663722     1.77   0.077    -.0127654    .2474088
        F3D. |  -.0212424   .0763746    -0.28   0.781    -.1709338    .1284491
        F4D. |   -.038634   .0616034    -0.63   0.531    -.1593746    .0821065
------------------------------------------------------------------------------
Note: Structural impulse-response functions are reported.
Impulse:     D.sp500
Responses:   D.bitcoin D.sp500
Instrument:  D.vix
Controls:    LD.bitcoin L2D.bitcoin LD.sp500 L2D.sp500
```

Sprecify 8 steps for the IRF

```
 . quietly ivlpirf D.bitcoin, endog(D.sp500 = D.vix) step(8)
. irf set ivlpirf_stocks.irf, replace
. irf create model1
. irf graph sirf, impulse(D.sp500) response(D.bitcoin)       ///
>         plot(color(maroon)) ciopts(color(maroon%30))       ///
>         yline(0) xlabel(0(2)8) name(irf_bitcoin,replace)
```

The dynamic shock shows a significant impact at step 0



Example 3: Unemployment rate for Spain

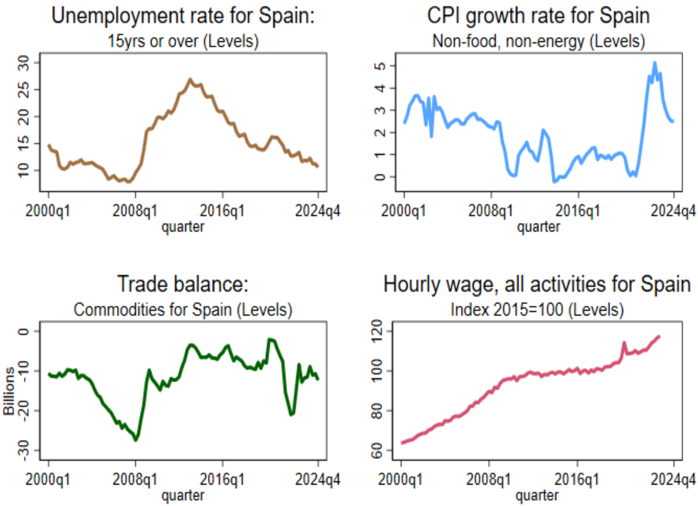
- We now fit a linear regression for the Spanish unemployment rate on inflation (considered to be endogenous).
- We use `import fred` to load the data:

```
. import fred LRUNTTTTESQ156N CPGRLE01ESQ659N  ///
>             LCEATT01ESQ661S XTNTVA01ESQ664S, ///
>             daterange(2000-01-01 .)
.
. rename  LRUNTTTTESQ156N urate
. rename  CPGRLE01ESQ659N cpi_growth
. rename LCEATT01ESQ661S wage
. rename XTNTVA01ESQ664S trade_bal
. replace trade_bal=trade_bal/1000000000
.
. generate year=year(daten)
. generate q=quarter(daten)
. generate quarter=yq(year,q)
. tsset quarter,quarterly
```

Data for unemployment rate, inflation, wage, and trade balance

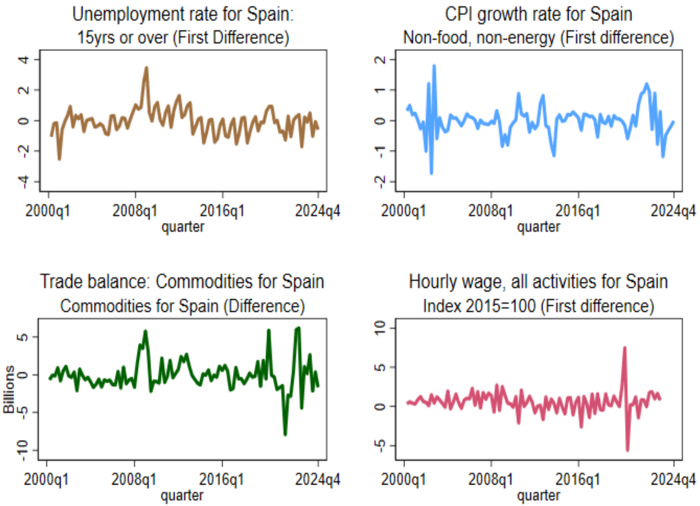
```
. describe

Contains data
 Observations:           101
    Variables:             7
--------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
--------------------------------------------------------------------------------
datestr         str10   %-10s                 observation date
daten           int     %td                   numeric (daily) date
urate           float   %9.0g                 Interannual unemployment rate for
                                                Spain
cpi_growth      float   %9.0g                 Interannual CPI growth rate,
                                                non-food non-energy for Spain
wage            float   %9.0g                 Labor Compensation: Earnings: All
                                                Activities: Hourly for Spain
trade_bal       float   %9.0g                 International Merchandise Trade
                                                Statistics: Trade Balance:
                                                Commodities for Spain
quarter         float   %tq
--------------------------------------------------------------------------------
Sorted by: quarter
     Note: Dataset has changed since last saved.
```

Levels of the variables



Source: World Bank and Organization for Economic Co-operation and Development via FRED®
Downloaded using -import fred- in Stata

First difference of the variables



Source: World Bank and Organization for Economic Co-operation and Development via FRED®

Downloaded using -import fred- in Stata

Multivariate regression for Unemployment rate and inflation

```
. ivlpirf D.urate,endog(D.cpi_growth = D.trade_bal D.wage) step(4) nolog
Final GMM criterion Q(b) = .1104853
Instrumental-variables local-projection impulse responses
Sample: 2000q4 thru 2023q3                         Number of obs = 92
 ( 1)  [D.cpi_growth]D.cpi_growth = 1
```

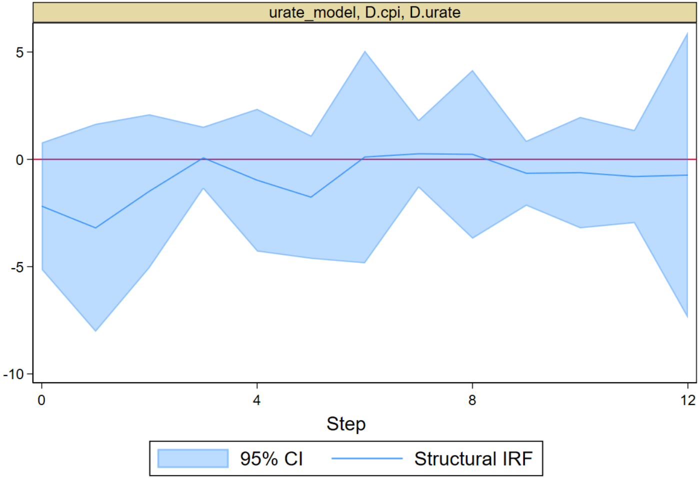
	IRF coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
urate						
D1.	-.4181043	1.386301	-0.30	0.763	-3.135205	2.298996
FD.	-.9768771	1.461507	-0.67	0.504	-3.841378	1.887624
F2D.	.1825374	1.173817	0.16	0.876	-2.118101	2.483176
F3D.	.5261707	.6972943	0.75	0.450	-.8405009	1.892842
F4D.	.4256272	1.110444	0.38	0.702	-1.750803	2.602057
cpi_growth						
D1.	1	(constrained)				
FD.	.0281609	.7530734	0.04	0.970	-1.447836	1.504158
F2D.	.7015666	.6476111	1.08	0.279	-.5677278	1.970861
F3D.	-.1469591	.769057	-0.19	0.848	-1.654283	1.360365
F4D.	-.1189874	.7288474	-0.16	0.870	-1.547502	1.309527

```
Note: Structural impulse-response functions are reported.
Impulse:     D.cpi_growth
Responses:   D.urate D.cpi_growth
Instruments: D.trade_bal D.wage
Controls:    LD.urate L2D.urate LD.cpi_growth L2D.cpi_growth
```

Create the .irf dataset to get the plot for the IRFs

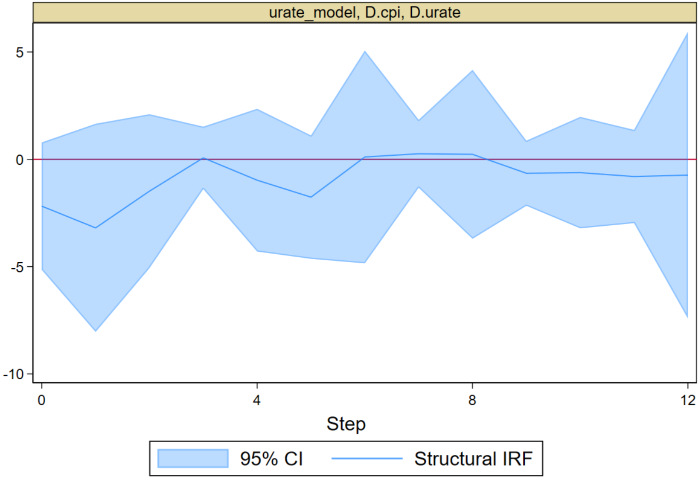
```
 . quietly ivlpirf D.urate,                          ///
>    endog(D.cpi_growth = D.trade_bal D.wage) ///
>    step(12)
.
. irf set urate_model.irf, replace
. irf create urate_model
. irf graph sirf, impulse(D.cpi_growth)       ///
>                 response(D.urate)            ///
>                 plot(color(stblue%75))       ///
>                 ciopts(color(stblue%30))     ///
>                 yline(0) xlabel(0(4)12)      ///
>                 name(urate_model,replace)
```

Impulse-response functions for Unemployment rate to shocks on inflation



Graphs by irfname, impulse variable, and response variable

Impulse-response functions for Unemployment rate to shocks on inflation

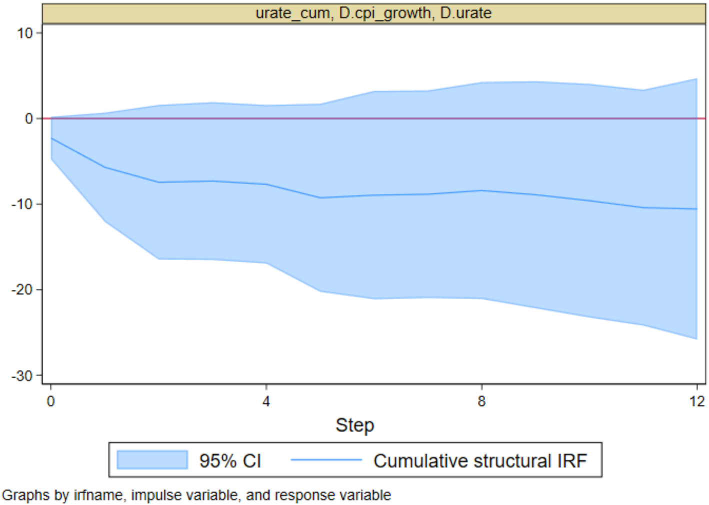


Graphs by irfname, impulse variable, and response variable

Create the .irf dataset to get the plot for the CSIRFs

```
. quietly ivlpirf D.urate,                                ///
>   endog(D.cpi_growth = D.trade_bal D.wage) ///
>   step(12) cumulative
.
. irf set urate_cum.irf, replace
. irf create urate_cum
.
. irf graph csirf, impulse(D.cpi_growth)       ///
>                 response(D.urate)            ///
>                 plot(color(stblue%75))        ///
>                 ciopts(color(stblue%30))      ///
>                 yline(0) xlabel(0(4)12)        ///
>                 name(urate_csirf,replace)
```

The cumulative SIRF provide the change in level because the outcome veriable is a rate



Final Remarks

- Local projections:
  - don't fit a full simultaneous equation (VAR) model.
  - simplify IRFs for multiple endogenous variables and horizons.
  - get CI with better small sample coverage than VAR models (Kilian and Kim 2011).
  - estimates IRFs and DM jointly, allowing more flexibility for inference.
  - can be extended to panel data and nonlinear models.

References

- Jordà, Ò (2005) Estimation and inference of impulse responses by local projections. American Economic Review 95:161–182.
- Jordà, Ò and Taylor, Alan M. (2024) Local Projections. Working paper 32822, National Bureau of Economic Research.
- Kilian, L., Kim, Yun Jung 2011, *How reliable are local projection estimators of impulse responses* The Review of Economics and Statistics, Vol. 93, No. 4, pp. 1460-1466