ISET

WORKING PAPER SERIES

vgets: A program to estimate general-to-specifc VARs, Granger causality, steady-state effects, and cumulative impulse responses

Muhammad Asali

vgets: A program to estimate general-to-specific VARs, Granger causality, steady-state effects, and cumulative impulse responses

Muhammad Asali

International School of Economics at TSU

IZA, Germany

SIPA, Columbia University

Abstract

Vector Autoregression (VAR) estimation is a vital tool in economic studies. VARs, however, can be dimensionally cumbersome and overparameterized. `vgets` command allows for a general-to-specific (GETS) estimation of VARs, overcoming the potential overparameterization, and provides tests for Granger causality, estimates of the long-run effects and the cumulative impulse response of each variable in the system; it also offers diagnostics that facilitate a genuine-causality interpretation of the Granger causality tests.

1 Introduction

Since Sims (1980) has popularized it, the vector autoregression (VAR) model has proven to be one of the most successful models for analyzing economic (and non-economic) time series data, offering descriptive analysis of the dynamics, forecasting, structural inference, and policy analysis. As a many-variables, higher-order VAR tends to be overparameterized, yielding weak inference, however, a general-to-specific approach (GETS) has

been suggested to improve the inference after VAR (Hendry et al. 2005). Moreover, Asali et al. (2017) provided additional estimators of long-run and steady-state effects that can be calculated after the VAR or the GETS-VAR.

To facilitate this type of analysis, from estimating the original balanced VAR, to its parsimonious GETS version, and inference after each model, we developed the command `vgets`, which standardizes the steps in this type of analysis as well as simplifies the process in making it easier and less prone to calculation mistakes. The `vgets` command estimates the full-specification VAR, arrives at its best reduced version (GETS), tests for Granger causality of each variable in the system, and provides estimates of the long-run and the cumulative impulse response effects of each variable, for both specifications (the full and the GETS). It also provides additional diagnostics and tests that can facilitate genuine-causality interpretation of the Granger causality tests. The time saved in doing the analysis with `vgets`, as compared to carrying them out manually, is immeasurable.

Jaeger and Paserman (2008) (henceforth JP) utilized the VAR framework to analyze the cycle of violence in the Israeli-Palestinian conflict. Asali, Abu-Qarn, and Beenstock (2017) (henceforth AAB), using the same JP data, but devising the use of GETS specifications, and offering new steady-state estimators (like the long-run kill-ratio, or the cumulative impulse response), overturned previous findings. In this paper, we use the JP and the AAB studies as an example of applying the `vgets` command.

2 Statistical Background: VAR, GETS VAR, Granger Causality, and Long-Run Effects

To illustrate the method and test statistics, we use the simplest two-variables VAR system:

$$x_t = \pi_1 + \sum_{i=1}^{L_{x1}} \alpha_{1i} x_{t-i} + \sum_{i=1}^{L_{y1}} \beta_{1i} y_{t-i} + z\gamma_1 + e_{xt}$$

$$y_t = \pi_2 + \sum_{i=1}^{L_{x2}} \alpha_{2i} x_{t-i} + \sum_{i=1}^{L_{y2}} \beta_{2i} y_{t-i} + z\gamma_2 + e_{yt}$$

where z is a vector of exogenous control variables. The full-specification refers to the balanced VAR, with the same lag-length for all the main variables. The General-to-Specific (GETS) VAR is the system we get from eliminating some of the insignificant lagged variables in each of the system equations; this is the so-called near-VAR system. In this case, the GETS VAR does not have to be balanced, and the lag length with respect to each variable (i.e., $L_{x1}, L_{y1}, L_{x2}, L_{y2}$) can be different. We then define the short-run effect of y on x as the sum of the coefficients of all the lagged y variables in the x equation, namely:

$$SR_{xy} = \sum_{i=1}^{L_{y1}} \beta_{1i} = \beta_1$$

likewise are defined α_1, α_2, and β_2. These effects are grouped in the matrix A:

$$A = \begin{pmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{pmatrix}$$

The long-run effects of one variable on the other is calculated by solving that dependent variable's equation for the long-run ratios between the variables—all the short-run lags being summed up over all the lags. The long-run effect of y on x from the first equation (which is the long-run x/y ratio) is given by:

$$LR_{xy} = \frac{\beta_1}{1 - \alpha_1}$$

likewise, the long-run y/x ratio from the second equation is given by:

$$LR_{yx} = \frac{\alpha_2}{1 - \beta_2}$$

Finally, the Cumulative Impulse Response is defined as the relevant element in the matrix:

$$CIR = (I - A)^{-1}$$

For the 2x2 case above, for example, the cumulative impulse response of y in the x equation is given by:

$$CIR_{xy} = \frac{\beta_1}{(1 - \alpha_1)(1 - \beta_2) - \alpha_2 \beta_1}$$

Statistical inference is then carried out for the effects in question (short-run, long-run, and cumulative impulse response): the standard errors of these estimates are calculated using the delta-method.

The extension to higher-order VARs is straightforward. While estimates of the LR and the CIR effects are always provided for any n-variable VAR, statistical inference is, however, limited to up-to-4-variables VAR systems because calculating the standard errors for the different CIR elements in many-variables (≥ 5) VARs is computationally prohibitive.[1]

3 Automatic Inference after VAR: vgets

The above mentioned steps can, in principle, be applied manually. That, however, is extremely time consuming, tedious, and the process is prone to many subtle errors of execution, that are difficult to discover or fix retroactively. The command offered here, `vgets`, is aimed at addressing this need. Besides, it offers all kinds

---

[1] For example, in a 4x4 system, the CIR effect of the third variable on the first variable is given by:

$(\gamma_1 + \beta_1\gamma_2 - \beta_2\gamma_1 - \gamma_1\delta_4 + \delta_1\gamma_4 - \beta_1\gamma_2\delta_4 + \beta_1\delta_2\gamma_4 + \beta_2\gamma_1\delta_4 - \beta_2\delta_1\gamma_4 - \gamma_1\beta_4\delta_2 + \gamma_2\delta_1\beta_4)/[\alpha_1\beta_2 - \beta_2 - \gamma_3 - \delta_4 - \alpha_1...$

$-\alpha_2\beta_1 + \alpha_1\gamma_3 - \alpha_3\gamma_1 + \alpha_1\delta_4 + \beta_2\gamma_3 - \alpha_4\delta_1 - \beta_3\gamma_2 + \beta_2\delta_4 - \beta_4\delta_2 + \gamma_3\delta_4 - \gamma_4\delta_3 - \alpha_1\beta_2\gamma_3 + \alpha_1\beta_3\gamma_2 + \alpha_2\beta_1\gamma_3 - \alpha_2\gamma_1\beta_3...$

$-\beta_1\alpha_3\gamma_2 + \alpha_3\beta_2\gamma_1 - \alpha_1\beta_2\delta_4 + \alpha_1\beta_4\delta_2 + \alpha_2\beta_1\delta_4 - \alpha_2\delta_1\beta_4 - \beta_1\alpha_4\delta_2 + \beta_2\alpha_4\delta_1 - \alpha_1\gamma_3\delta_4 + \alpha_1\gamma_4\delta_3 + \alpha_3\gamma_1\delta_4 - \alpha_3\delta_1\gamma_4...$

$-\gamma_1\alpha_4\delta_3 + \alpha_4\delta_1\gamma_3 - \beta_2\gamma_3\delta_4 + \beta_2\gamma_4\delta_3 + \beta_3\gamma_2\delta_4 - \beta_3\delta_2\gamma_4 - \gamma_2\beta_4\delta_3 + \beta_4\gamma_3\delta_2 + \alpha_1\beta_2\gamma_3\delta_4 - \alpha_1\beta_2\gamma_4\delta_3 - \alpha_1\beta_3\gamma_2\delta_4...$

$+\alpha_1\beta_3\delta_2\gamma_4 + \alpha_1\gamma_2\beta_4\delta_3 - \alpha_1\beta_4\gamma_3\delta_2 - \alpha_2\beta_1\gamma_3\delta_4 + \alpha_2\beta_1\gamma_4\delta_3 + \alpha_2\gamma_1\beta_3\delta_4 - \alpha_2\gamma_1\beta_4\delta_3 - \alpha_2\beta_3\delta_1\gamma_4 + \alpha_2\delta_1\beta_4\gamma_3 + \beta_1\alpha_3\gamma_2\delta_4...$

$-\beta_1\alpha_3\delta_2\gamma_4 - \beta_1\alpha_4\gamma_2\delta_3 + \beta_1\alpha_4\gamma_3\delta_2 - \alpha_3\beta_2\gamma_1\delta_4 + \alpha_3\beta_2\delta_1\gamma_4 + \alpha_3\gamma_1\beta_4\delta_2 - \alpha_3\gamma_2\delta_1\beta_4 + \beta_2\gamma_1\alpha_4\delta_3 - \beta_2\alpha_4\delta_1\gamma_3 - \gamma_1\alpha_4\beta_3\delta_2 +$

$\alpha_4\beta_3\gamma_2\delta_1 + 1]$, whose standard error calculation by the delta method in Stata is resources and time consuming. In the 5x5 case we reckon Stata might take days to calculate the standard error of one element of the CIR matrix, if at all possible.

of diagnostic checks that are essential for the interpretation of results and the validity of statistical inference after the VAR estimation. The results of these checks, for example the absence of serial correlation in the system errors, moreover, can give the Granger causality results a genuine causality interpretation (e.g., see AAB).

3.1 Syntax

`vgets` *varlist* [if] [, maxlag(*int*) t(#) exog(*varlist*) <u>d</u>iagnostics <u>q</u>uick <u>no</u>is <u>f</u>ormat(*%fmt*)]

3.2 Description

`vgets` estimates a balanced VAR of the listed variables, with lags of each variable from 1 to "maxlag." It also estimates the general-to-specific (GETS) VAR system of the given variables, using a stepwise elimination procedure, where the omitted lagged variables are those whose coefficient's t-statistics are, in absolute value, below the specified level in the `t()` option. The command then reports, for both the full specification and the GETS specification, Granger causality tests of the different variables, the long-run effects (or ratios) of the different variables, and the cumulative impulse responses of the different variables. The exogenous variables listed in the `exog()` option are not subject to elimination in the GETS procedure.

While the point estimates of the effects are reported for any set of main variables, statistical inference (i.e., standard errors and p-values) are calculated and reported only for systems of two, three, or four variables.

3.3 Options

`maxlag(#)` specifies the maximum lag length of the main variables in the balanced VAR. If not specified, the maximum lag is assumed to be 1.

`t()` specifies the maximum level of the t-statistic at which lagged variables will be dropped in the general-to-specific procedure; guided by Haitovsky (1969) rule, the default level is 1: lagged variables with absolute

t-statistic below 1 will be dropped iteratively, to reach the system general-to-specific final specification.

`exog(varlist)` the list of exogenous/control variables that are not subject to elimination in the general-to-specific procedure (i.e., maintained regardless of their t-statistic). These variables are included in their current (time t) values, and are not subject to any lag structure.

`diagnostics` reports diagnostic measures of the residuals in each equation. In particular, robust Breusch-Godfrey tests of serial correlation (up to the 6th order), Jarque-Bera asymptotic test for normality of the residuals in each equation of the system, skewness and kurtosis of the respective residuals set, and AIC and BIC information criteria for model selection.

`quick` shows only the effects (short-run, long-run, cumulative impulse response) matrices, without statistical inference (standard errors, p-values, tests for Granger causality). This is the default and only option if the main set of variables includes 5 or more variables.

`nois` shows all the underlying regressions, tests, and intermediate steps.

`format(%fmt)` specifies the display format for coefficients, effects, standard errors, and p-values.

3.4   Stored Results

The long-run effects (ratios) are saved in the LR matrix for the full specification, and in the getsLR matrix for the GETS specification. Likewise, the cumulative impulse response matrix is saved under the name CIR for the full specification, and getsCIR for the GETS specification. If Granger causality tests are carried out then the respective test statistics, and p-values, will be saved under the scalars "chiXY" and "pchiXY," respectively; where XY refers to the order of the variables in the list *varlist*; that is, the effect of variable number Y on variable number X. So, if we type `vgets var1 var2 var3`, then "chi32" will represent the test statistic of Granger causality of var2 on var3. For the GETS specification, these results are saved under the names GchiXY and GpchiXY.

Similarly, the standard errors and p-values for the long-run effects and the cumulative impulse responses

will be saved under the scalars LRseXY, LRpXY, CIRseXY, CIRpXY, for the full-specification. For the GETS specification the respective scalars are: GLRseXY, GLRpXY, GCIRseXY, GCIRpXY

3.5 Example

JP provided an interesting use of balanced VARs to study the cycle of violence in the Middle East. Their conclusion, regarding the absence of a cycle, however, has been overturned in AAB, who suggested the use of GETS VAR, due to the overparameterization in the full, balanced specification.[2] In the Israeli-Palestinian conflict, the number of Palestinian and Israeli fatalities in day t are, respectively, pal_t and isr_t. The full specification of the system is given by:

$$pal_t = \alpha_{p,1}pal_{t-1} + ... + \alpha_{p,t-14}pal_{t-14} + \beta_{p,1}isr_{t-1} + ... + \beta_{p,t-14}isr_{t-14} + z_t\gamma_p + e_{pt}$$

$$isr_t = \alpha_{i,1}pal_{t-1} + ... + \alpha_{i,t-14}pal_{t-14} + \beta_{i,1}isr_{t-1} + ... + \beta_{i,t-14}isr_{t-14} + z_t\gamma_i + e_{it}$$

where z is a vector of control, exogenous variables that includes dummy variables indicating the weekday, dummy variables indicating the period in the Israeli-Palestinian context (e.g., Oslo accords, Barak government, Sharon government, pre and post September eleventh, etc...), and the length of the barrier built between Israel and the West Bank.

The first equation of the VAR is called the "Israeli reaction function," and the second one is called the "Palestinian reaction function." Israel "reacts to violence" if the lags of Israeli fatalities in the Israeli reaction function Granger-cause Palestinian fatalities. Likewise the Palestinian reaction is defined from the second equation.

The main findings of JP and AAB can be easily replicated with the `vgets` command. For example, reproducing columns 2 and 4 of Table 1 from JP, as well as columns (1) and (3) in both panels of Table 1 from AAB, is carried out as follows:

---

[2] The data used in these studies, and in this example, can be downloaded from here: <http://qed.econ.queensu.ca/jae/datasets/asali001/>.

7

```
. vgets pal_tot isr_tot if date>=14882 & date<=16451, maxlag(14) t(1) exog(Period2 Period3
Period4 Period5 Period6 Period7 completed sunday monday tuesday wednesday thursday friday)
format(%9.3f)
```

To see the underlying regressions and statistical tests add the option `nois`. While JP refer only to the full specification, AAB refer also to the general-to-specific specification (GETS). The `vgets` command reports both:

```
_____FULL Specification_____

Equation:  pal_tot
--------+---------------------------------
Variable| isr_tot
--------+---------------------------------
GC      |      24.952
    pv  | (     0.035)
--------+---------------------------------
LR      |       1.320
    se  | (     0.435)
    pv  | (     0.002)
--------+---------------------------------
CIR     |       1.615
    se  | (     0.655)
    pv  | (     0.014)
--------+---------------------------------

Equation:  isr_tot
--------+---------------------------------
Variable| pal_tot
--------+---------------------------------
GC      |      17.974
    pv  | (     0.208)
--------+---------------------------------
LR      |       0.094
    se  | (     0.042)
    pv  | (     0.026)
--------+---------------------------------
CIR     |       0.179
    se  | (     0.096)
    pv  | (     0.062)
--------+---------------------------------
_____
_____GETS SPECIFICATION_____

Equation:pal_tot
--------+---------------------------------
Variable| isr_tot
--------+---------------------------------
GC      |      21.519
    pv  | (     0.011)
--------+---------------------------------
LR      |       1.233
    se  | (     0.408)
    pv  | (     0.003)
--------+---------------------------------
CIR     |       1.523
    se  | (     0.582)
    pv  | (     0.009)
--------+---------------------------------

Equation:isr_tot
--------+---------------------------------
Variable| pal_tot
--------+---------------------------------
GC      |      13.332
    pv  | (     0.064)
--------+---------------------------------
LR      |       0.084
    se  | (     0.041)
    pv  | (     0.042)
--------+---------------------------------
CIR     |       0.158
    se  | (     0.089)
    pv  | (     0.075)
--------+---------------------------------
```

"GC" refers to Granger causality, so in the full specification the own (Israeli) fatalities Granger cause Palestinian fatalities, i.e., Israel reacts to violence, significant at the 5% level (pv=.035); while Palestinians do not react to violence (pv=.208). However, once the GETS (AAB) specification is considered, we see that

both sides react with respective p-values of: 0.011 and 0.064. The slight difference between the reported p-values and those appearing in the original studies is that the `vgets` command uses the more efficient system-estimation approach (Zellner's seemingly unrelated regression), as opposed to the equation-by-equation OLS.

"LR" refers to the long-run effect. In the current context, LR from the Israeli reaction function refers to the number of Palestinian fatalities inflicted by Israel per each Israeli fatality. In the full-specification this is 1.32, and in the GETS specification this is 1.23; both are highly statistically significant.

Finally, "CIR" refers to the cumulative impulse response, which is the response of one side to the violence of the other, taking into account the other's response. These are also statistically significant and valued at 1.615 and 1.523 for the Israeli reaction function in the full and GETS specifications. (For the Palestinian reaction functions, the respective estimates are: 0.179 and 0.158.)

Adding the option `diagnostics` at the end of the command, reports all the diagnostic tests that AAB suggested for testing the performance and interpretation of the VAR in these contexts, like serial correlation tests of the system residuals, normality tests of these, and information criteria for each model. This pertains to the upper panel of Table 1 from AAB. For example, for the full specification of the Israeli reaction function the diagnostics output will be:

```
_____FULL Specification_____

Diagnostics... Equation: pal_tot
----------------------------------------
AR(1) (pv)|    1.505 (    0.220)
AR(2) (pv)|    1.376 (    0.253)
AR(3) (pv)|    1.436 (    0.230)
AR(4) (pv)|    1.828 (    0.121)
AR(5) (pv)|    1.508 (    0.184)
AR(6) (pv)|    1.277 (    0.265)
----------+------------------------------
JB (pv)   | 119711.4 (    0.000)
Skewness  |    4.45
Kurtosis  |   44.84
----------+-------------------------------
AIC       |    8264.6
BIC       |    8489.7
----------------------------------------
```

The absence of serial correlation in the residuals of the first equation, the long specification of the Israeli reaction function, renders the lagged variables weakly exogenous, and therefore suggests that the

Granger causality (from own Israeli fatalities to Palestinian fatalities) that we found earlier ($\chi^2 = 24.95$, p-value= .035) can be interpreted as genuine causality. The same can be said about the Palestinian reaction function.

4 Conclusion

Vector autoregression (VAR) model has proven to be a very useful and successful model for describing the dynamics of time series data, offering accurate forecasting, structural inference, and providing solid grounds for policy analysis. Nonetheless, the literature has argued for more parsimonious, and thus more accurate, specifications that do not suffer from overparameterization by applying the theory of reduction (Hendry et al. 2005). Asali et al. (2017) have shown that using a general-to-specific (GETS) approach to the VAR analysis matters in important ways, not only leading to more accurate inference, but can actually overturn findings that are based on the fully specified VAR.

The command `vgets` that we are suggesting here streamlines the process of working with VARs, GETS VARs, and carrying out statistical inference from these models, like studying the Granger causality of the different variables, and their steady-state long-run effects and cumulative impulse responses. It also provides statistical tests whose results might render the found Granger causality genuine.

References

[1] Asali, Muhammad; Abu-Qarn, Aamer; Beenstock, Michael. 2017. "The Cycle of Violence in the Second Intifada: Causality in Nonlinear Vector Autoregressive Models." *Journal of Applied Econometrics* 32(6): 1197–1205.

[2] Haitovsky, Yoel. 1969. "A note on the maximization of $\bar{R}^2$." *American Statistician* 23(1): 20–21.

[3] Hendry, David F.; Ericsson, Neil R.; Campos, Julia. 2005. *General-to-Specific Modelling.* Edward Elgar Publishing: Cheltenham, UK.

[4] Jaeger, David; Paserman, Daniel. 2008. "The cycle of violence? An empirical analysis of fatalities in the Palestinian-Israeli conflict." *American Economic Review* 98(4): 1591–1604.

[5] Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48(1): 1-48.