The Stata Journal (2020)
20, Number 2, pp. 426–434 DOI: 10.1177/1536867X20931004

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

Muhammad Asali
International School of Economics at Tbilisi State University, Tbilisi, Georgia
IZA, Bonn, Germany
School of International and Public Affairs, Columbia University, New York, NY
muhammad.asali@gmail.com

Abstract. Vector autoregression (VAR) estimation is a vital tool in economic studies. VARs, however, can be dimensionally cumbersome and overparameterized. The `vgets` command allows for a general-to-specific 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.

Keywords: st0602, vgets, general-to-specific vector autoregressions, Granger causality, steady-state effects, cumulative impulse–responses

1 Introduction

Since Sims (1980) popularized it, the vector autoregression (VAR) model has proven to be one of the most successful models for analyzing economic (and noneconomic) time-series data, offering descriptive analysis of the dynamics, forecasting, structural inference, and policy analysis. However, because a many-variables, higher-order VAR tends to be overparameterized—yielding weak inference—a general-to-specific (GETS) approach has been suggested to improve the inference after VAR (Campos, Ericsson, and Hendry 2005). Moreover, Asali, Abu-Qarn, and Beenstock (2017) provided additional estimators of long-run (LR) and steady-state effects that can be calculated after the VAR or the GETS VAR. Asali and Gurashvili (Forthcoming) use this framework to study the relationship between discrimination in the labor market and the macroeconomy.

To facilitate this type of analysis, from estimating the original balanced VAR to its parsimonious GETS version, and inference after each model, I developed the command `vgets`, which standardizes the steps in this type of analysis and simplifies the process by 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 LR and the cumulative impulse–response (CIR) 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 analyses with `vgets`, compared with carrying them out manually, is immeasurable.

Jaeger and Paserman (2008) (henceforth, JP) used 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 LR kill ratio or the CIR), overturned previous findings. In this article, I use the JP and AAB studies as examples of applying the `vgets` command.

2 Statistical background: VAR, GETS VAR, Granger causality, and LR effects

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

$$
\begin{aligned}
x_t &= \pi_1 + \sum_{i=1}^{L_{x1}} \alpha_{1i} x_{t-i} + \sum_{i=1}^{L_{y1}} \beta_{1i} y_{t-i} + \mathbf{z}\boldsymbol{\gamma}_1 + e_{xt} \\
y_t &= \pi_2 + \sum_{i=1}^{L_{x2}} \alpha_{2i} x_{t-i} + \sum_{i=1}^{L_{y2}} \beta_{2i} y_{t-i} + \mathbf{z}\boldsymbol{\gamma}_2 + e_{yt}
\end{aligned}
$$

where $\mathbf{z}$ 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 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 (that is, $L_{x1}, L_{y1}, L_{x2}, L_{y2}$) can be different. We then define the short-run (SR) effect of y on x as the sum of the coefficients of all the lagged y variables in the x equation; namely,

$$
\mathrm{SR}_{xy} = \sum_{i=1}^{L_{y1}} \beta_{1i} = \beta_1
$$

Likewise, we can define α_1, α_2, and β_2. These effects are grouped in the matrix $\mathbf{A}$:

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

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

$$
\mathrm{LR}_{xy} = \frac{\beta_1}{1-\alpha_1}
$$

Likewise, the LR y/x ratio from the second equation is given by

$$\mathrm{LR}_{yx} = \frac{\alpha_2}{1-\beta_2}$$

Finally, the CIR is defined as the relevant element in the matrix:

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

For the 2×2 case above, for example, the CIR of y in the x equation is given by

$$\mathrm{CIR}_{xy} = \frac{\beta_1}{(1-\alpha_1)(1-\beta_2) - \alpha_2\beta_1}$$

Statistical inference is then carried out for the effects in question (SR, LR, and CIR): 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 CIR effects are always provided for any n-variable VAR, statistical inference is limited to up-to-four-variables VAR systems because calculating the standard errors for the different CIR elements in many-variables (≥ 5) VARs is computationally prohibitive.

3 Automatic inference after VAR: vgets

The above steps can, in principle, be applied manually, but that is extremely time consuming and tedious, and the process is prone to many subtle errors of execution that are difficult to discover or fix retroactively. The command `vgets` is aimed at addressing this need. In addition, it offers 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, can give the Granger causality results a genuine causality interpretation (for example, refer to the AAB article).

3.1 Syntax

`vgets` *varlist* [*if*] [`, maxlag(`*#*`) t(`*#*`) exog(`*varlist*`) diagnostics quick nois format(%`*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 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 LR effects (or ratios) of the different variables, and the CIRs 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 (that is, robust standard errors and p-values) is 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. The default is `maxlag(1)`.

`t(#)` specifies the maximum level of the t statistic at which lagged variables will be dropped in the GETS procedure. Guided by the Haitovsky (1969) rule, the default is `t(1)`: lagged variables with absolute t statistic below 1 will be dropped iteratively to reach the system GETS final specification.

`exog(`*varlist*`)` specifies the list of exogenous or control variables that are not subject to elimination in the GETS procedure (that is, maintained regardless of their t statistic). These variables are included in only 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, it reports robust Breusch–Godfrey tests of serial correlation (up to the sixth order), Jarque–Bera asymptotic tests for normality of the residuals in each equation of the system, skewness and kurtosis of the respective residuals set, and Akaike information criterion and Bayesian information criterion for model selection.

`quick` shows only the effects (SR, LR, CIR) 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 five 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 command stores the results in `r()`. The SR effects are stored in the `r(SR)` matrix and in the `r(getsSR)` matrix for the GETS specification. The LR effects (ratios) are stored in the `r(LR)` matrix for the full specification and in the `r(getsLR)` matrix for the GETS specification. Likewise, the CIR matrix is stored under the name `r(CIR)` for the full specification and `r(getsCIR)` for the GETS specification. If Granger causality tests are carried out, then the respective test statistics and p-values will be stored under the scalars `r(chi`XY`)` and `r(pchi`XY`)`, respectively; 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 `r(chi32)` will represent the test statistic of Granger causality of `var2` on `var3`. For the GETS specification, these results are stored under the names `r(Gchi`XY`)` and `r(Gpchi`XY`)`.

Similarly, the effects, their standard errors, and their p-values for the LR effects and the CIRs will be stored under the scalars `r(LR`XY`)`, `r(LRse`XY`)`, `r(LRp`XY`)`, `r(CIR`XY`)`, `r(CIRse`XY`)`, and `r(CIRp`XY`)` for the full specification. For the GETS specification, the respective scalars are `r(GLR`XY`)`, `r(GLRse`XY`)`, `r(GLRp`XY`)`, `r(GCIR`XY`)`, `r(GCIRse`XY`)`, and `r(GCIRp`XY`)`. The matrix `r(TABLE)` summarizes the results for all equations in the full specification—Granger causality, LR effect, and CIR of each variable for each equation. The GETS specification results are stored in the matrix `r(getsTABLE)`.

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 because of the overparameterization in the full, balanced specification.[1] 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

$$
\begin{aligned}
\text{pal}_t &= \alpha_{p,1}\text{pal}_{t-1} + \cdots + \alpha_{p,t-14}\text{pal}_{t-14} + \beta_{p,1}\text{isr}_{t-1} + \cdots + \beta_{p,t-14}\text{isr}_{t-14} \\
&\quad + \mathbf{z}_t\boldsymbol{\gamma}_p + e_{pt} \\
\text{isr}_t &= \alpha_{i,1}\text{pal}_{t-1} + \cdots + \alpha_{i,t-14}\text{pal}_{t-14} + \beta_{i,1}\text{isr}_{t-1} + \cdots + \beta_{i,t-14}\text{isr}_{t-14} \\
&\quad + \mathbf{z}_t\boldsymbol{\gamma}_i + e_{it}
\end{aligned}
$$

where $\mathbf{z}$ is a vector of control, exogenous variables that include dummy variables indicating the weekday, dummy variables indicating the period in the Israeli–Palestinian context (for example, Oslo accords, Barak government, Sharon government, pre- and post-September 11), 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 equation 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.

1. The data used in these studies, and in this example, can be downloaded from http://qed.econ.queensu.ca/jae/datasets/asali001/.

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:

```
. use intifada_extended_data
. 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 GETS specification. The vgets command reports both:

FULL Specification

	pal_tot isr_tot	isr_tot pal_tot
GC	24.952	17.974
(pv)	0.035	0.208
LR	1.320	0.094
se	0.435	0.042
(pv)	0.002	0.026
CIR	1.615	0.179
se	0.655	0.096
(pv)	0.014	0.062

GETS Specification

	pal_tot isr_tot	isr_tot pal_tot
GC	21.519	13.332
(pv)	0.011	0.064
LR	1.233	0.084
se	0.408	0.041
(pv)	0.003	0.042
CIR	1.523	0.158
se	0.582	0.089
(pv)	0.009	0.075

GC refers to Granger causality, so in the full specification, the own (Israeli) fatalities Granger-cause Palestinian fatalities; that is, Israel reacts to violence, significant at the 5% level (pv = 0.035), while Palestinians do not react to violence (pv = 0.208). 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 (seemingly unrelated estimation), as opposed to the equation-by-equation ordinary least squares.

LR refers to the LR effect. In the current context, the LR effect 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 CIR, which is the response of one side to the violence of the other, accounting for 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, respectively. (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 the system residuals, 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

```
. 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) diagnostics
Diagnostics: FULL Specification
```

	pal_tot	isr_tot
AR(1)	1.505	0.118
pv	0.220	0.731
AR(2)	1.376	1.104
pv	0.253	0.332
AR(3)	1.436	0.850
pv	0.230	0.467
AR(4)	1.828	0.877
pv	0.121	0.477
AR(5)	1.508	0.718
pv	0.184	0.610
AR(6)	1.277	0.630
pv	0.265	0.706
JB	1.20e+05	1.59e+05
pv	0.000	0.000
Skewness	4.451	6.107
Kurtosis	44.842	50.819
AIC	8264.603	7070.572
BIC	8489.674	7295.643

```
System AIC: 15329.555. SBIC: 15779.697
```

```
Diagnostics: GETS Specification
             |   pal_tot    isr_tot
-------------+----------------------
       AR(1) |     0.141      0.010
          pv |     0.707      0.919
       AR(2) |     0.292      0.083
          pv |     0.747      0.920
       AR(3) |     0.209      0.157
          pv |     0.890      0.925
       AR(4) |     0.380      0.294
          pv |     0.823      0.882
       AR(5) |     0.329      0.303
          pv |     0.896      0.912
       AR(6) |     0.279      0.325
          pv |     0.947      0.924
          JB |  1.25e+05   1.58e+05
          pv |     0.000      0.000
    Skewness |     4.493      6.114
    Kurtosis |    45.743     50.631
         AIC |  8245.753   7039.634
         BIC |  8406.518   7168.246
System AIC: 15279.695. SBIC: 15569.071
FULL Specification
```

(*output omitted*)

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 $= 0.035$) can be interpreted as genuine causality. The same can be said about the Palestinian reaction function.

4 Conclusion

The VAR model has proven to be a useful and successful model for describing the dynamics of time-series data, offering accurate forecasting and 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 (Campos, Ericsson, and Hendry 2005). AAB have shown that using a GETS approach to the VAR analysis is important, leading not only to more accurate inference but also to actually overturning findings that are based on the fully specified VAR.

The command `vgets` that I present here streamlines the process of working with VARs and GETS VARs and carrying out statistical inference from these models, like studying the Granger causality of the different variables and their steady-state LR effects and CIRs. It also provides statistical tests whose results might render the found Granger causality genuine.

5 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 20-2
. net install st0602      (to install program files, if available)
. net get st0602          (to install ancillary files, if available)
```

6 References

Asali, M., A. S. Abu-Qarn, and M. Beenstock. 2017. The cycle of violence in the Second Intifada: Causality in nonlinear vector autoregressive models. *Journal of Applied Econometrics* 32: 1197–1205. https://doi.org/10.1002/jae.2563.

Asali, M., and R. Gurashvili. Forthcoming. Labor market discrimination and the macroeconomy. *Economics of Transition and Institutional Change*.

Campos, J., N. R. Ericsson, and D. F. Hendry, eds. 2005. *General-to-Specific Modelling*. Cheltenham, UK: Edward Elgar Publishing.

Haitovsky, Y. 1969. A note on the maximization of $\overline{R}^2$. *American Statistician* 23: 20–21. https://doi.org/10.1080/00031305.1969.10481815.

Jaeger, D. A., and M. D. Paserman. 2008. The cycle of violence? An empirical analysis of fatalities in the Palestinian–Israeli conflict. *American Economic Review* 98: 1591–1604. https://doi.org/10.1257/aer.98.4.1591.

Sims, C. A. 1980. Macroeconomics and reality. *Econometrica* 48: 1–48. https://doi.org/10.2307/1912017.

About the author

Muhammad Asali (PhD in economics, Columbia University) is an associate professor of economics at the International School of Economics at Tbilisi State University, Georgia. He is also the head of graduate studies at the school. Asali is an adjunct associate professor of international affairs at Columbia University and a research fellow at IZA, the Institute of Labor Economics, in Germany. His research and teaching interests include applied microeconomics, statistics, econometrics, time-series econometrics, and labor economics.