Title

stata.com

var intro — Introduction to vector autoregressive models

Description Remarks and examples References Also see

Description

Stata has a suite of commands for fitting, forecasting, interpreting, and performing inference on vector autoregressive (VAR) models and structural vector autoregressive (SVAR) models. The suite includes several commands for estimating and interpreting impulse–response functions (IRFs), dynamic-multiplier functions, and forecast-error variance decompositions (FEVDs). The table below describes the available commands.

Fitting a VAR or SVAR

`var`	[TS] **var**	Fit vector autoregressive models
`svar`	[TS] **var svar**	Fit structural vector autoregressive models
`varbasic`	[TS] **varbasic**	Fit a simple VAR and graph IRFs or FEVDs

Model diagnostics and inference

`varstable`	[TS] **varstable**	Check the stability condition of VAR or SVAR estimates
`varsoc`	[TS] **varsoc**	Obtain lag-order selection statistics for VARs and VECMs
`varwle`	[TS] **varwle**	Obtain Wald lag-exclusion statistics after `var` or `svar`
`vargranger`	[TS] **vargranger**	Perform pairwise Granger causality tests after `var` or `svar`
`varlmar`	[TS] **varlmar**	Perform LM test for residual autocorrelation after `var` or `svar`
`varnorm`	[TS] **varnorm**	Test for normally distributed disturbances after `var` or `svar`

Forecasting after fitting a VAR or SVAR

`fcast compute`	[TS] **fcast compute**	Compute dynamic forecasts after `var`, `svar`, or `vec`
`fcast graph`	[TS] **fcast graph**	Graph forecasts after `fcast compute`

Working with IRFs, dynamic-multiplier functions, and FEVDs

`irf`	[TS] **irf**	Create and analyze IRFs, dynamic-multiplier functions, and FEVDs

This entry provides an overview of vector autoregressions and structural vector autoregressions. More rigorous treatments can be found in Hamilton (1994), Lütkepohl (2005), and Amisano and Giannini (1997). Stock and Watson (2001) provide an excellent nonmathematical treatment of vector autoregressions and their role in macroeconomics. Becketti (2020) provides an excellent introduction to VAR analysis with an emphasis on how it is done in practice.

Remarks and examples

stata.com

Remarks are presented under the following headings:

Introduction to VARs
Introduction to SVARs
Short-run SVAR models
Long-run restrictions
IRFs and FEVDs

Introduction to VARs

A VAR is a model in which K variables are specified as linear functions of p of their own lags, p lags of the other $K-1$ variables, and possibly additional exogenous variables. Algebraically, a p-order VAR model, written VAR(p), with exogenous variables $\mathbf{x}_t$ is given by

$$\mathbf{y}_t = \mathbf{v} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{B}_0\mathbf{x}_t + \mathbf{B}_1\mathbf{x}_{t-1} + \cdots + \mathbf{B}_s\mathbf{x}_{t-s} + \mathbf{u}_t \qquad t \in \{-\infty, \infty\} \quad (1)$$

where

$\mathbf{y}_t = (y_{1t}, \ldots, y_{Kt})'$ is a $K \times 1$ random vector,
$\mathbf{A}_1$ through $\mathbf{A}_p$ are $K \times K$ matrices of parameters,
$\mathbf{x}_t$ is an $M \times 1$ vector of exogenous variables,
$\mathbf{B}_0$ through $\mathbf{B}_s$ are $K \times M$ matrices of coefficients,
$\mathbf{v}$ is a $K \times 1$ vector of parameters, and
$\mathbf{u}_t$ is assumed to be white noise; that is,
$E(\mathbf{u}_t) = \mathbf{0}$,
$E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{\Sigma}$, and
$E(\mathbf{u}_t\mathbf{u}_s') = \mathbf{0}$ for $t \neq s$

There are $K^2 \times p + K \times (M(s+1)+1)$ parameters in the equation for $\mathbf{y}_t$, and there are $\{K \times (K+1)\}/2$ parameters in the covariance matrix $\mathbf{\Sigma}$. One way to reduce the number of parameters is to specify an incomplete VAR, in which some of the $\mathbf{A}$ or $\mathbf{B}$ matrices are set to zero. Another way is to specify linear constraints on some of the coefficients in the VAR.

A VAR can be viewed as the reduced form of a system of dynamic simultaneous equations. Consider the system

$$\mathbf{W}_0\mathbf{y}_t = \mathbf{a} + \mathbf{W}_1\mathbf{y}_{t-1} + \cdots + \mathbf{W}_p\mathbf{y}_{t-p} + \widetilde{\mathbf{W}}_1\mathbf{x}_t + \widetilde{\mathbf{W}}_2\mathbf{x}_{t-2} + \cdots + \widetilde{\mathbf{W}}_s\mathbf{x}_{t-s} + \mathbf{e}_t \quad (2)$$

where $\mathbf{a}$ is a $K \times 1$ vector of parameters, each $\mathbf{W}_i$, $i = 0, \ldots, p$, is a $K \times K$ matrix of parameters, and $\mathbf{e}_t$ is a $K \times 1$ disturbance vector. In the traditional dynamic simultaneous equations approach, sufficient restrictions are placed on the $\mathbf{W}_i$ to obtain identification. Assuming that $\mathbf{W}_0$ is nonsingular, (2) can be rewritten as

$$\begin{aligned}\mathbf{y}_t =& \mathbf{W}_0^{-1}\mathbf{a} + \mathbf{W}_0^{-1}\mathbf{W}_1\mathbf{y}_{t-1} + \cdots + \mathbf{W}_0^{-1}\mathbf{W}_p\mathbf{y}_{t-p} \\ &+ \mathbf{W}_0^{-1}\widetilde{\mathbf{W}}_1\mathbf{x}_t + \mathbf{W}_0^{-1}\widetilde{\mathbf{W}}_2\mathbf{x}_{t-2} + \cdots + \mathbf{W}_0^{-1}\widetilde{\mathbf{W}}_s\mathbf{x}_{t-s} + \mathbf{W}_0^{-1}\mathbf{e}_t\end{aligned} \quad (3)$$

which is a VAR with

$$\begin{aligned}\mathbf{v} &= \mathbf{W}_0^{-1}\mathbf{a} \\ \mathbf{A}_i &= \mathbf{W}_0^{-1}\mathbf{W}_i \\ \mathbf{B}_i &= \mathbf{W}_0^{-1}\widetilde{\mathbf{W}}_i \\ \mathbf{u}_t &= \mathbf{W}_0^{-1}\mathbf{e}_t\end{aligned}$$

The cross-equation error variance–covariance matrix $\mathbf{\Sigma}$ contains all the information about contemporaneous correlations in a VAR and may be the VAR's greatest strength and its greatest weakness. Because no questionable a priori assumptions are imposed, fitting a VAR allows the dataset to speak for itself. However, without imposing some restrictions on the structure of $\mathbf{\Sigma}$, we cannot make a causal interpretation of the results.

If we make additional technical assumptions, we can derive another representation of the VAR in (1). If the VAR is stable (see [TS] **varstable**), we can rewrite $\mathbf{y}_t$ as

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{D}_i \mathbf{x}_{t-i} + \sum_{i=0}^{\infty} \mathbf{\Phi}_i \mathbf{u}_{t-i} \tag{4}$$

where $\boldsymbol{\mu}$ is the $K \times 1$ time-invariant mean of the process and $\mathbf{D}_i$ and $\mathbf{\Phi}_i$ are $K \times M$ and $K \times K$ matrices of parameters, respectively. Equation (4) states that the process by which the variables in $\mathbf{y}_t$ fluctuate about their time-invariant means, $\boldsymbol{\mu}$, is completely determined by the parameters in $\mathbf{D}_i$ and $\mathbf{\Phi}_i$ and the (infinite) past history of the exogenous variables $\mathbf{x}_t$ and the independent and identically distributed (i.i.d.) shocks or innovations, $\mathbf{u}_{t-1}, \mathbf{u}_{t-2}, \ldots$. Equation (4) is known as the vector moving-average representation of the VAR. The $\mathbf{D}_i$ are the dynamic-multiplier functions, or transfer functions. The moving-average coefficients $\mathbf{\Phi}_i$ are also known as the simple IRFs at horizon i. The precise relationships between the VAR parameters and the $\mathbf{D}_i$ and $\mathbf{\Phi}_i$ are derived in *Methods and formulas* of [TS] **irf create**.

The joint distribution of $\mathbf{y}_t$ is determined by the distributions of $\mathbf{x}_t$ and $\mathbf{u}_t$ and the parameters $\mathbf{v}$, $\mathbf{B}_i$, and $\mathbf{A}_i$. Estimating the parameters in a VAR requires that the variables in $\mathbf{y}_t$ and $\mathbf{x}_t$ be covariance stationary, meaning that their first two moments exist and are time invariant. If the $\mathbf{y}_t$ are not covariance stationary, but their first differences are, a vector error-correction model (VECM) can be used. See [TS] **vec intro** and [TS] **vec** for more information about those models.

If the $\mathbf{u}_t$ form a zero mean, i.i.d. vector process, and $\mathbf{y}_t$ and $\mathbf{x}_t$ are covariance stationary and are not correlated with the $\mathbf{u}_t$, consistent and efficient estimates of the $\mathbf{B}_i$, the $\mathbf{A}_i$, and $\mathbf{v}$ are obtained via seemingly unrelated regression, yielding estimators that are asymptotically normally distributed. When the equations for the variables $\mathbf{y}_t$ have the same set of regressors, equation-by-equation OLS estimates are the conditional maximum likelihood estimates.

Much of the interest in VAR models is focused on the forecasts, IRFs, dynamic-multiplier functions, and the FEVDs, all of which are functions of the estimated parameters. Estimating these functions is straightforward, but their asymptotic standard errors are usually obtained by assuming that $\mathbf{u}_t$ forms a zero mean, i.i.d. Gaussian (normal) vector process. Also, some of the specification tests for VARs have been derived using the likelihood-ratio principle and the stronger Gaussian assumption.

In the absence of contemporaneous exogenous variables, the disturbance variance–covariance matrix contains all the information about contemporaneous correlations among the variables. VARs are sometimes classified into three types by how they account for this contemporaneous correlation. (See Stock and Watson [2001] for one derivation of this taxonomy.) A reduced-form VAR, aside from estimating the variance–covariance matrix of the disturbance, does not try to account for contemporaneous correlations. In a recursive VAR, the K variables are assumed to form a recursive dynamic structural equation model in which the first variable is a function of lagged variables, the second is a function of contemporaneous values of the first variable and lagged values, and so on. In a structural VAR, the theory you are working with places restrictions on the contemporaneous correlations that are not necessarily recursive.

Stata has two commands for fitting reduced-form VARs: `var` and `varbasic`. `var` allows for constraints to be imposed on the coefficients. `varbasic` allows you to fit a simple VAR quickly without constraints and graph the IRFs.

Because fitting a VAR of the correct order can be important, `varsoc` offers several methods for choosing the lag order p of the VAR to fit. After fitting a VAR, and before proceeding with inference, interpretation, or forecasting, checking that the VAR fits the data is important. `varlmar` can be used to check for autocorrelation in the disturbances. `varwle` performs Wald tests to determine whether certain lags can be excluded. `varnorm` tests the null hypothesis that the disturbances are normally distributed. `varstable` checks the eigenvalue condition for stability, which is needed to interpret the IRFs and IRFs.

Introduction to SVARs

As discussed in [TS] **irf create**, a problem with VAR analysis is that, because $\mathbf{\Sigma}$ is not restricted to be a diagonal matrix, an increase in an innovation to one variable provides information about the innovations to other variables. This implies that no causal interpretation of the simple IRFs is possible: there is no way to determine whether the shock to the first variable caused the shock in the second variable or vice versa.

However, suppose that we had a matrix $\mathbf{P}$ such that $\mathbf{\Sigma} = \mathbf{P}\mathbf{P}'$. We can then show that the variables in $\mathbf{P}^{-1}\mathbf{u}_t$ have zero mean and that $E\{\mathbf{P}^{-1}\mathbf{u}_t(\mathbf{P}^{-1}\mathbf{u}_t)'\} = \mathbf{I}_K$. We could rewrite (4) as

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\mu} + \sum_{s=0}^{\infty} \mathbf{\Phi}_s \mathbf{P}\mathbf{P}^{-1}\mathbf{u}_{t-s} \\ &= \boldsymbol{\mu} + \sum_{s=0}^{\infty} \mathbf{\Theta}_s \mathbf{P}^{-1}\mathbf{u}_{t-s} \\ &= \boldsymbol{\mu} + \sum_{s=0}^{\infty} \mathbf{\Theta}_s \mathbf{w}_{t-s} \end{aligned} \tag{5}$$

where $\mathbf{\Theta}_s = \mathbf{\Phi}_s\mathbf{P}$ and $\mathbf{w}_t = \mathbf{P}^{-1}\mathbf{u}_t$. If we had such a $\mathbf{P}$, the $\mathbf{w}_k$ would be mutually orthogonal, and the $\mathbf{\Theta}_s$ would allow the causal interpretation that we seek.

SVAR models provide a framework for estimation of and inference about a broad class of $\mathbf{P}$ matrices. As described in [TS] **irf create**, the estimated $\mathbf{P}$ matrices can then be used to estimate structural IRFs and structural FEVDs. There are two types of SVAR models. Short-run SVAR models identify a $\mathbf{P}$ matrix by placing restrictions on the contemporaneous correlations between the variables. Long-run SVAR models, on the other hand, do so by placing restrictions on the long-term accumulated effects of the innovations.

Short-run SVAR models

A short-run SVAR model without exogenous variables can be written as

$$\mathbf{A}(\mathbf{I}_K - \mathbf{A}_1 L - \mathbf{A}_2 L^2 - \cdots - \mathbf{A}_p L^p)\mathbf{y}_t = \mathbf{A}\boldsymbol{\epsilon}_t = \mathbf{B}\mathbf{e}_t \tag{6}$$

where L is the lag operator; $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are $K \times K$ matrices of parameters; $\boldsymbol{\epsilon}_t$ is a $K \times 1$ vector of innovations with $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{\Sigma})$ and $E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_s'] = \mathbf{0}_K$ for all $s \neq t$; and $\mathbf{e}_t$ is a $K \times 1$ vector of orthogonalized disturbances; that is, $\mathbf{e}_t \sim N(\mathbf{0}, \mathbf{I}_K)$ and $E[\mathbf{e}_t\mathbf{e}_s'] = \mathbf{0}_K$ for all $s \neq t$. These transformations of the innovations allow us to analyze the dynamics of the system in terms of a change to an element of $\mathbf{e}_t$. In a short-run SVAR model, we obtain identification by placing restrictions on $\mathbf{A}$ and $\mathbf{B}$, which are assumed to be nonsingular.

Equation (6) implies that $\mathbf{P}_{\rm sr} = \mathbf{A}^{-1}\mathbf{B}$, where $\mathbf{P}_{\rm sr}$ is the $\mathbf{P}$ matrix identified by a particular short-run SVAR model. The latter equality in (6) implies that

$$\mathbf{A}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\mathbf{A}' = \mathbf{B}\mathbf{e}_t\mathbf{e}_t'\mathbf{B}'$$

Taking the expectation of both sides yields

$$\boldsymbol{\Sigma} = \mathbf{P}_{\rm sr}\mathbf{P}_{\rm sr}'$$

Assuming that the underlying VAR is stable (see [TS] **varstable** for a discussion of stability), we can invert the autoregressive representation of the model in (6) to an infinite-order, moving-average representation of the form

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{s=0}^{\infty} \boldsymbol{\Theta}_s^{\rm sr}\mathbf{e}_{t-s} \tag{7}$$

whereby $\mathbf{y}_t$ is expressed in terms of the mutually orthogonal, unit-variance structural innovations $\mathbf{e}_t$. The $\boldsymbol{\Theta}_s^{\rm sr}$ contain the structural IRFs at horizon s.

In a short-run SVAR model, the $\mathbf{A}$ and $\mathbf{B}$ matrices model all the information about contemporaneous correlations. The $\mathbf{B}$ matrix also scales the innovations $\mathbf{u}_t$ to have unit variance. This allows the structural IRFs constructed from (7) to be interpreted as the effect on variable i of a one-time unit increase in the structural innovation to variable j after s periods.

$\mathbf{P}_{\rm sr}$ identifies the structural IRFs by defining a transformation of $\boldsymbol{\Sigma}$, and $\mathbf{P}_{\rm sr}$ is identified by the restrictions placed on the parameters in $\mathbf{A}$ and $\mathbf{B}$. Because there are only $K(K+1)/2$ free parameters in $\boldsymbol{\Sigma}$, only $K(K+1)/2$ parameters may be estimated in an identified $\mathbf{P}_{\rm sr}$. Because there are $2K^2$ total parameters in $\mathbf{A}$ and $\mathbf{B}$, the order condition for identification requires that at least $2K^2 - K(K+1)/2$ restrictions be placed on those parameters. Just as in the simultaneous-equations framework, this order condition is necessary but not sufficient. Amisano and Giannini (1997) derive a method to check that an SVAR model is locally identified near some specified values for $\mathbf{A}$ and $\mathbf{B}$.

Before moving on to models with long-run constraints, consider these limitations. We cannot place constraints on the elements of $\mathbf{A}$ in terms of the elements of $\mathbf{B}$, or vice versa. This limitation is imposed by the form of the check for identification derived by Amisano and Giannini (1997). As noted in *Methods and formulas* of [TS] **var svar**, this test requires separate constraint matrices for the parameters in $\mathbf{A}$ and $\mathbf{B}$. Also, we cannot mix short-run and long-run constraints.

Long-run restrictions

A general short-run SVAR has the form

$$\mathbf{A}(\mathbf{I}_K - \mathbf{A}_1L - \mathbf{A}_2L^2 - \cdots - \mathbf{A}_pL^p)\mathbf{y}_t = \mathbf{B}\mathbf{e}_t$$

To simplify the notation, let $\overline{\mathbf{A}} = (\mathbf{I}_K - \mathbf{A}_1L - \mathbf{A}_2L^2 - \cdots - \mathbf{A}_pL^p)$. The model is assumed to be stable (see [TS] **varstable**), so $\overline{\mathbf{A}}^{-1}$, the matrix of estimated long-run effects of the reduced-form VAR shocks, is well defined. Constraining $\mathbf{A}$ to be an identity matrix allows us to rewrite this equation as

$$\mathbf{y}_t = \overline{\mathbf{A}}^{-1}\mathbf{B}\mathbf{e}_t$$

which implies that $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}'$. Thus $\mathbf{C} = \overline{\mathbf{A}}^{-1}\mathbf{B}$ is the matrix of long-run responses to the orthogonalized shocks, and

$$\mathbf{y}_t = \mathbf{C}\mathbf{e}_t$$

In long-run models, the constraints are placed on the elements of $\mathbf{C}$, and the free parameters are estimated. These constraints are often exclusion restrictions. For instance, constraining $\mathbf{C}[1,2]$ to be zero can be interpreted as setting the long-run response of variable 1 to the structural shocks driving variable 2 to be zero.

Stata's `svar` command estimates the parameters of structural VARs. See [TS] **var svar** for more information and examples.

IRFs and FEVDs

IRFs describe how the K endogenous variables react over time to a one-time shock to one of the K disturbances. Because the disturbances may be contemporaneously correlated, these functions do not explain how variable i reacts to a one-time increase in the innovation to variable j after s periods, holding everything else constant. To explain this, we must start with orthogonalized innovations so that the assumption to hold everything else constant is reasonable. Recursive VARs use a Cholesky decomposition to orthogonalize the disturbances and thereby obtain structurally interpretable IRFs. Structural VARs use theory to impose sufficient restrictions, which need not be recursive, to decompose the contemporaneous correlations into orthogonal components.

FEVDs are another tool for interpreting how the orthogonalized innovations affect the K variables over time. The FEVD from j to i gives the fraction of the s-step forecast-error variance of variable i that can be attributed to the jth orthogonalized innovation.

Dynamic–multiplier functions describe how the endogenous variables react over time to a unit change in an exogenous variable. This is a different experiment from that in IRFs and FEVDs because dynamic-multiplier functions consider a change in an exogenous variable instead of a shock to an endogenous variable.

`irf create` estimates IRFs, Cholesky orthogonalized IRFs, dynamic-multiplier functions, and structural IRFs and their standard errors. It also estimates Cholesky and structural FEVDs. The `irf graph`, `irf cgraph`, `irf ograph`, `irf table`, and `irf ctable` commands graph and tabulate these estimates. Stata also has several other commands to manage IRF and FEVD results. See [TS] **irf** for a description of these commands.

`fcast compute` computes dynamic forecasts and their standard errors from VARs. `fcast graph` graphs the forecasts that are generated using `fcast compute`.

VARs allow researchers to investigate whether one variable is useful in predicting another variable. A variable x is said to Granger-cause a variable y if, given the past values of y, past values of x are useful for predicting y. The Stata command `vargranger` performs Wald tests to investigate Granger causality between the variables in a VAR.

References

Amisano, G., and C. Giannini. 1997. *Topics in Structural VAR Econometrics*. 2nd ed, revised and enlarged. Heidelberg: Springer.

Asali, M. 2020. vgets: A command to estimate general-to-specific VARs, Granger causality, steady-state effects, and cumulative impulse–responses. *Stata Journal* 20: 426–434.

Becketti, S. 2020. *Introduction to Time Series Using Stata*. Rev. ed. College Station, TX: Stata Press.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*. New York: Springer.

Stock, J. H., and M. W. Watson. 2001. Vector autoregressions. *Journal of Economic Perspectives* 15: 101–115. https://doi.org/10.1257/jep.15.4.101.

Watson, M. W. 1994. Vector autoregressions and cointegration. In Vol. 4 of *Handbook of Econometrics*, ed. R. F. Engle and D. L. McFadden. Amsterdam: Elsevier. https://doi.org/10.1016/S1573-4412(05)80016-9.

Also see

[TS] **irf** — Create and analyze IRFs, dynamic-multiplier functions, and FEVDs

[TS] **var** — Vector autoregressive models

[TS] **var svar** — Structural vector autoregressive models

[TS] **vec** — Vector error-correction models

[TS] **vec intro** — Introduction to vector error-correction models

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.