**Setting up a VAR Model in Stata**

*Starting Point*

The starting point of a VAR analysis is the unrestricted model, i.e. the reduced-form VAR, in which the modeler has not imposed any restrictions on the model. The reduced-form *k*th-order VAR can be denoted as follows:

🡺

Where is a (*)*-vector of *K-*endogenous variables,is the inverse of the *()-*matrix of contemporaneous parameters, is a *()-*matrix with the *k-*lag polynomials, , is a *()-*vector of structural shocks, and which are reduced-form shocks with the following variance-covariance matrix: .

**Step 1:** *Data selection N-variable kth-order reduced-form*

Of course, the first step would be to decide on which variables to include in and on the frequency of the data. We have decided to first only work with variables within the current database that CEPAL uses for ARIMAX forecasts.

The first thing we have to take into account is that the database consists of variables with different frequencies, namely quarterly data (see Table 1) and monthly data (see Table 2). Hence, before estimating the model we should convert all monthly data to quarterly data.

**Table 1** All variables in CEPAL’s existent database with a quarterly frequency

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Defintion** | **Period** | **Source** |
| *rgdp* |  | 1993q1 - 2017q1 |  |
| *rpc* |  | 1993q1 - 2017q1 |  |
| *rgc* |  | 1993q1 - 2017q1 |  |
| *ri* |  | 1993q1 - 2017q1 |  |
| *rx* |  | 1993q1 - 2017q1 |  |
| *rm* |  | 1993q1 - 2017q1 |  |
| *primario* |  | - |  |
| *manuf* |  | - |  |
| *serv* |  | - |  |
| *rgdp\_sa* |  | 1993q1 - 2017q1 |  |

**Table 2** All variables in CEPAL’s existent database with a monthly frequency

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Defintion** | **Period** | **Source** |
| *igae* |  | Jan 1993 - Aug 2016 |  |
| *retail* |  | Jan 2008 - Aug 2016 |  |
| *ip* |  | Jan 1993 - Sep 2016 |  |
| *ip\_mineral* |  | Jan 1993 - Sep 2016 |  |
| *ip\_energy* |  | Jan 1993 - Sep 2016 |  |
| *ip\_construction* |  | Jan 1993 - Sep 2016 |  |
| *ip\_manufacturing* |  | Jan 1993 - Sep 2016 |  |
| *finv* |  | Jan 1993 - Aug 2016 |  |
| *consumo* |  | Jan 1993 - Aug 2016 |  |
| *confianza\_emp* |  | Jan 2004 - Oct 2016 |  |
| confianza\_con |  | Apr 2001 - Oct 2016 |  |
| p\_exp |  | Jan 1993 - Dec 2016 |  |
| p\_imp |  | Jan 1993 - Dec 2016 |  |
| tot |  | Jan 1993 - Dec 2016 |  |
| exp |  | Jan 1993 - Sep 2016 |  |
| exp\_petro |  | Jan 1993 - Sep 2016 |  |
| imp |  | Jan 1993 - Sep 2016 |  |
| imp\_consumer |  | Jan 1993 - Sep 2016 |  |
| imp\_intermediate | | Jan 1993 - Sep 2016 |  |
| imp\_capital |  | Jan 1993 - Sep 2016 |  |
| ing\_gob |  | Jan 1993 - Sep 2016 |  |
| gto\_gob |  | Jan 1993 - Sep 2016 |  |
| remesa |  | Jan 1995 - Sep 2016 |  |
| cred |  | Dec 1994 - Sep 2016 |  |
| tcr |  | Jan 1994 - Sep 2016 |  |
| m1 |  | Jan 1993 - Sep 2016 |  |
| m2 |  | Jan 1993 - Sep 2016 |  |
| bolsa |  | Jan 1993 - Oct 2016 |  |
| cta\_fin |  | Feb 2000 - Sep 2016 |  |

Secondly, we have to decide whether we have to transform the data, for example to first differences. I assume we are going to work with first-differences?

**Step 2:** Test for stationarity: Augmented Dickey-Fuller (ADF) test.

After we decided on the N-variables to include in the VAR we should run an ADF test on each variables in order to determine whether the variables are a stationary I(0) process or a non-stationary I(1) process. A notion must be made that the ADF-test has one significant disadvantage. That is, it models the time-series either as a full non-stationary unit root process or as a full stationary process. This strict assumption implies that fractionally integrated series, i.e. those series that are close to a non-stationary unit root process, are frequently misjudged as a unit root process (see e.g. Cochrane, 1991; Diebold and Rudebusch, 1991).

Stata command:

dfuller varname [if] [in] [, options]

Options: - noconstant suppress constant term in regression, - trend include trend term in regression, - drift include drift term in regression, - regress display regression table, - lags(#) include # lagged differences.

**Step 3:** Number of lags to include in the VAR

After we have decided on the *N-*variables to include and dealt we the stationarity issues, we should determine the number of *k-*lags to include. For this purpose we should do a lag-test.

Stata command: varsoc 🡺 The varsoc command gives you a battery of lag-order selection tests.

**Step 4:** Stability of the VAR

After we have obtained the number of lags we have to examine whether the VAR is stable. This condition is satisfied if and only if all inverse roots of the determinant of the characteristic polynomial have a modulus .

If the VAR is stable (see command varstable) we can rewrite the VAR in moving average from as explained in the next step.

Stata command:

varstable [, options]

Options: estimates(estname): use previously stored results estname default is to use active results, - amat(matrix\_name); save the companion matrix as matrix\_name, - graph; graph eigenvalues of the companion matrix, - dlabel; label eigenvalues with the distance from the unit circle, - modlabel; label eigenvalues with the modulus, marker\_options; change look of markers (color, size, etc.), -rlopts(cline\_options); affect rendition of reference unit circle, - nogrid; suppress polar grid circles, - pgrid([...]); specify radii and appearance of polar grid circles.

**Step 5:** If we want to turn it into an SVAR, the final step would be the identification of the structural shocks.

The reduced-form VAR in **(G2)** is represented in the conventional autoregressive (AR) representation. In the AR representation, a set of *N* time-series are modeled as a function of their past values. However, equation **(G2)** cannot be used for structural economic inference due to two reasons. First, equation **(G2)** is written as a function of past values of the variables, but in order to analyze structural shocks, we must rewrite as a function of current and past shocks of the variables. This can be achieved by writing the VAR in its *moving average* (MA) representation:

🡺

Note that in the step from equation **(G3)** to equation **(G4)**, I took the inverse of . This step is only possible if is invertible (i.e. non-singular). Furthermore, equation **(G4)** is stable if and only if all inverse roots of the determinant of the characteristic polynomial have a modulus . In other words, the *k*th-VAR is stable, if and only if, all inverse roots of lie on the complex unit circle. This condition is required to make sure that the structural shocks peter-out in the long run (we have tested for this in step 4). Moreover, note that a stable VAR-process is always stationary but that an unstable process can also be stationary.

Second, equation **(G2)**, and equation **(G4)** as well, only consists of reduced shocks. These shocks are devoid of any economic content and can therefore not be used for economic inference. We can cope with this issue by rewriting the model such that it is expressed as a function of current and past *structural* shocks. For this transformation we make use of the condition that and write:

🡺 🡺 🡺

Where and **.** The *N-*variable *k*th-order VAR is now modeled in its MA representation and is now written as a function of current and past structural shocks. Nonetheless, the model is not ready yet to construct and interpret the IRFs because in order to compute , which contain the impulse response functions to the structural shocks, the matrix of the contemporaneous parameters, , must be known.

In order to identify , we must impose identifying restrictions. The first identifying restriction is a general restriction that must be satisfied in a general *N-*variable *k*th-order VAR and that is the *orthogonality* assumption of the structural shocks. Orthogonality implies that the structural shocks are uncorrelated. In other words, only one type of structural shock occurs at the time and there cannot be a commingling of shocks. Technically, this implies that the covariances, presented on the off-diagonal elements of , are all zero[[1]](#footnote-2). To understand how this imposes one non-linear restriction on , recall that the reduced-form shocks and structural shocks are linked by , and that the variance-covariance matrix of the reduced-form shocks and the structural shocks are related through: . Hence, you can see that the orthogonality restriction imposed on Induces one non-linear restriction on . An important remark must be made that the orthogonality assumption only implies that the structural shocks are uncorrelated, and it does not imply that the structural shocks only enter their own structural equation. Structural shocks are still allowed to enter other structural equations[[2]](#footnote-3).

Besides the orthogonality restriction we still need to impose restrictions on the matrix in order to identify the model. As with all structural modeling, the identification should be done based on economic theory and a helpful tool in determining these restriction is a Granger-causality test. The latter test helps with determining the contemporaneous links between the variables. If one variable is said to granger-cause the other this implies that this variable leads the other in time. The results from these tests can be used as rationale to order the contemporaneous links

Stata commands:

With the command svar you can impose many short-run and long-run restrictions in stata. I have attached the help file in the document.

The **vargranger** post estimation command performs a battery of Granger causality tests. As before, equations are distinguished by their dependent variable. For each equation, **vargranger** tests for the Granger causality of each variable in the VAR individually, then tests for the Granger causality of all added variables jointly.

1. Note that I apply another general transformation of the variance-covariance matrix and that is normalization. This imposes not a restriction but only functions as a practical way to generate one standard deviation structural shocks. [↑](#footnote-ref-2)
2. The interested reader is referred to Bernanke (1986) who clearly explains the implications of the orthogonality assumption. [↑](#footnote-ref-3)