Vector Autoregressions

Gabriel Marin[1]

Inter-American Development Bank[1]

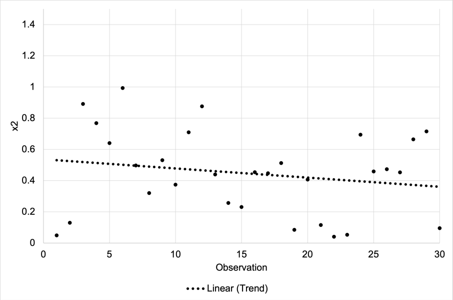
Macroeconomic Policy: Fiscal, Monetary and Finance

Outline

Table of Contents

Why does OLS is as good as econometricians make it sound?

Imagine a random sample of two variables x_1 and x_2, you know that x_2 is a variable linearly correlated with x_1 such that $\rho(x_1, x_2) \neq 0$ and you want to find the best way to explain x_2, what can you do? Well, a reasonable assumption would be to compute the trend of x_2 and expect that to the value it would converge too, wouldn't it?



Why does OLS is as good as econometricians make it sound?

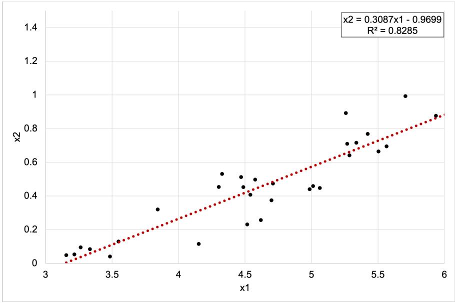
Yet, you know that x_1 is a variable linearly correlated with x_2, so, why don't we try and try to explain x_2 given x_1? Well, a way to introduce that into this problem is to create the linear projection of x_2 onto x_1 such as

$$x_{i,2} = c_0 + c_1 x_{i,1} + error_i \tag{1}$$

I'm guessing you guys know where we are heading...

OLS

Yes, c_0 is what we know as the intercept of the equation β_0 and c_1, the coefficient associated with x_1 is what we commonly know as β_1. Yet, we are missing out on how to estimate such relationship. For that, we use the methodology known as Ordinary Least Squares (OLS).



What is OLS?

OLS seeks to minimize the squared distance between of our predicted variable $\hat{x_2}$ and the variable itself x_2. In other words, OLS seeks to minimize the squared sum of the error term of the linear model shown before. Formally,

$$min \sum_{i}^{n} (x_2 - \hat{x_2})^2 = min \sum_{i}^{n} (\hat{u_i})^2. \tag{2}$$

How do we find the minimum of a variable? Furthermore, can the variables change to find this optimum value? What has to change to find this minimum value of u_i?

OLS Coefficients

Recall:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 x \tag{3}$$

and

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{4}$$

How would this look in matrix form? Say, we have another variable x_3 that also explains x_2?

OLS Assumptions - Introduction to Econometrics

1. Linearity in Parameters
2. Random Sampling
3. No Perfect Collinearity
4. Zero Conditional Mean
5. Constant Variance (Homoskedasticity)
6. Normality of the Error Term

Such assumptions are important to be able to perform statistical inference of your coefficient. Under Assumptions 1-4, the OLS coefficients are unbiased and consistent. If Assumption 5 also holds, OLS coefficients are asymptotically efficient.

Table of Contents

OLS Assumptions in Time-Series Data

1. Linearity in Parameters
2. ~~Random Sampling~~ Why?
2. No Perfect Collinearity (Allows for variables to be correlated, but not perfectly correlated)
3. Zero Conditional Mean $(E[U_t|X_t] = 0)$. Contemporaneously exogenous. Is this the only thing we need? What about future or past realizations of X?
4. Constant Variance (Homoskedasticity)
5. No Serial Correlation $(Cor(U_t, U_s) = 0)$
6. Normality of the Error Term

Stationarity - Wooldridge (2016)

- A stochastic process is stationary if for every collection of time indices $1 \leq t_1 < t_2 < ... < t_m$, the joint distribution of $(x_t, x_{t2}, ..., x_{tm})$ is the same as the joint distribution of $(x_{t+h}, x_{t2+h}, ..., x_{tm+h})$. In other words, this implies that the distribution of, say, x_t has the same distribution of x_1 for all $t = 2, 3, ...$
- What is the implication of that in terms of expected values? Would $E[x_{t+1}] = E[x_t]$ hold if the time series were not stationary?
- Sometimes a weaker form of stationarity suffices, if $(x_t, x_{t2}, ..., x_{tm})$ has a finite second moment, i.e $E[x_t^2] < \infty$ for all t, then the series is **covariance stationary**

Covariance Stationarity - Wooldridge (2016)

- Covariance stationarity implies that the mean and variance of the time series process is constant across time, and the covariance between x_t and x_{t+h} depend only on the distance h.
- Why does this matter? Well, if the relationship between y_t and x_t changes over time, then how can we establish a reliable relationship between them?

Weak Dependence - Wooldridge (2016)

- A time series process is said to be **weakly dependent** if x_t and x_{t+h} are "almost independent" as h increases.
- In other words, this implies the relationship between future values of x_{t+h} and x_t "dies" over time. This ensures stability in the model!
- This assumption replaces the Random Sampling from OLS!

A Cool Example - The AR(1) Model

Suppose the variable y_t has the following process

$$y_t = \rho_1 y_{t-1} + e_t, t = 1, 2, ..., \tag{5}$$

where e_t is an i.i.d sequence with zero mean and variance σ_e^2. If we assume that e_t is independent of y_0, the starting point of the sequence, and that $E[y_0]$, we have an **autoregressive process of order one AR(1)**.

- In this case, weak dependence necessarily requires $|\rho_1| < 1$, called the *stability condition*. If $|\rho_1| = 1$, the process is called a *random walk*.
- In practice, time-series are usually *trend-stationary*, such that when including a trend term in the regression, you can obtain β coefficients that are asymptotically valid.

Break

Why did you study economics? What do you want to work in after graduation? Are you planning on pursuing a graduate degree?

Table of Contents

Vector Autoregressions Explained

- They are the expansion of the AR model we just previously saw in the multivariate case.
- They are extremely powerful tools, once the assumptions we saw are fulfilled.
- Pioneered in Sims (1980).

Suppose we have two variables x and z that behave in such manner

$$x_t = \phi_{1,x} x_{t-1} + \phi_{2,x} z_{t-1} + e_{x,t} \tag{6}$$

$$z_t = \phi_{1,z} x_{t-1} + \phi_{2,z} z_{t-1} + e_{z,t} \tag{7}$$

How would this look in matrix form?

VAR(1) Framework in Matrix Form

In matrix form,

$$AY_t = \Phi_1 Y_{t-1} + CE_t, \tag{8}$$

where

$$Y_t = \begin{bmatrix} x_t \\ z_t \end{bmatrix}, \Phi_1 = \begin{bmatrix} \phi_{1,x} & \phi_{2,x} \\ \phi_{1,z} & \phi_{2,z} \end{bmatrix}, E_t = \begin{bmatrix} e_{x,t} \\ e_{z,t} \end{bmatrix}, \tag{9}$$

and

$$A = C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{10}$$

VAR(p) Framework

If we were to expand to p lags, the VAR framework in matrix form would look as follows

$$AY_t = \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + ... + \Phi_p Y_{t-p} + CE_t, \tag{11}$$

- This is an **unrestricted** VAR model, as the RHS of the equations are the same for both variables.
- In a VAR framework, selecting p can be quite tricky. A safe bet would be to choose a type of *lag selection criteria*. Which use the models' residuals to identify whether more lags add more "information" to the model. This might be a useful link to understand.

VAR Estimation

- VARs are estimated through a Maximum Likelihood Method, assuming the data is conditionally normally distributed. Formally,

$$Y_t | Y_{t-1}, Y_{t-2}, ..., Y_{t-p} \sim N(\Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + ... + \Phi_p Y_{t-p}, \Omega). \quad (12)$$

- In an unrestricted VAR setting, the $\hat{\Phi_k}$ coefficients are the same as OLS equation by equation!
- If we were to estimate a restricted VAR, things would have to be quite different.

Restricted VAR Model

- Recall equation (8), what would happen if A or C were not diagonal?

$$AY_t = \Phi_1 Y_{t-1} + CE_t, \tag{13}$$

- We would have more parameters than equations! For example, let A be of the following form

$$\begin{bmatrix} 1 & a_{12} \\ a_{21} & 1 \end{bmatrix}. \tag{14}$$

- Helpful link to understand VARs from Professor Susmel in University of Houston.

Restricted VAR Model - Continued

- Let us expand the system to understand where the identification issue arises, remember, we are estimating the variance of the error terms and the covariance!

$$Y_t = \begin{bmatrix} x_t \\ z_t \end{bmatrix}, \Phi_1 = \begin{bmatrix} \phi_{1,x} & \phi_{2,x} \\ \phi_{1,z} & \phi_{2,z} \end{bmatrix}, E_t = \begin{bmatrix} \epsilon_{x,t} \\ \epsilon_{z,t} \end{bmatrix}, \tag{15}$$

and

$$A = \begin{bmatrix} 1 & a_{12} \\ a_{21} & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{16}$$

- To see this problem better, let us compute the reduced form VAR.

$$Y_t = A^{-1}\Phi_1 Y_{t-1} + A^{-1}CE_t, \tag{17}$$

Restricted VAR Model - Continued

Therefore, we have

$$Y_t = \tilde{\Phi}_1 Y_{t-1} + \tilde{C} E_t, \tag{18}$$

In particular, we know from our matrix algebra courses that

$$A^{-1} = \frac{1}{1 - a_{12}a_{21}} \begin{bmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{bmatrix}, \tag{19}$$

Therefore,

$$\tilde{\Phi}_1 = \begin{bmatrix} \tilde{\phi_{1,x}} & \tilde{\phi_{2,x}} \\ \tilde{\phi_{1,z}} & \tilde{\phi_{2,z}} \end{bmatrix} = \frac{1}{1 - a_{12}a_{21}} \begin{bmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{bmatrix} \begin{bmatrix} \phi_{1,x} & \phi_{2,x} \\ \phi_{1,z} & \phi_{2,z} \end{bmatrix}, \tag{20}$$

$$\tilde{C} = \begin{bmatrix} \tilde{c_{1,1}} & \tilde{c_{1,2}} \\ \tilde{c_{2,1}} & \tilde{c_{2,2}} \end{bmatrix} = \frac{1}{1 - a_{12}a_{21}} \begin{bmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{21}$$

Restricted VAR Model - Continued

Expanding the previous matrix problem for $\tilde{\Phi}_1$ yields

$$\tilde{\phi}_{1,x} = \left(\frac{1}{1 - a_{12}a_{21}}\right)\phi_{1,x} - \left(\frac{a_{12}}{1 - a_{12}a_{21}}\right)\phi_{1,z}, \tag{22}$$

$$\tilde{\phi}_{2,x} = \left(\frac{1}{1 - a_{12}a_{21}}\right)\phi_{2,x} - \left(\frac{a_{12}}{1 - a_{12}a_{21}}\right)\phi_{2,z}, \tag{23}$$

$$\tilde{\phi}_{1,z} = -\left(\frac{a_{21}}{1 - a_{12}a_{21}}\right)\phi_{1,x} + \left(\frac{1}{1 - a_{12}a_{21}}\right)\phi_{1,z}, \tag{24}$$

$$\tilde{\phi}_{2,z} = -\left(\frac{a_{21}}{1 - a_{12}a_{21}}\right)\phi_{2,x} + \left(\frac{1}{1 - a_{12}a_{21}}\right)\phi_{2,z}, \tag{25}$$

Restricted VAR Model - Continued

Analogously, for $\tilde{C}$

$$c_{\tilde{1,1}} = \left(\frac{1}{1 - a_{12}a_{21}}\right) \tag{26}$$

$$c_{\tilde{1,2}} = -\left(\frac{a_{12}}{1 - a_{12}a_{21}}\right) \tag{27}$$

$$c_{\tilde{2,1}} = -\left(\frac{a_{21}}{1 - a_{12}a_{21}}\right) \tag{28}$$

$$c_{\tilde{2,2}} = \left(\frac{a_{21}}{1 - a_{12}a_{21}}\right) \tag{29}$$

Finally, we have to also estimate

$$\hat{\sigma_{\epsilon_x}}^2 \tag{30}$$

$$\hat{\sigma_{\epsilon_z}}^2 \tag{31}$$

$$Cov(\epsilon_x, \epsilon_z) \tag{32}$$

Restricted VAR Model - Continued

This implies that we have 12 parameters and 11 equations! How do we solve this?

- We need to impose a restriction to the system for it to be properly identified. In econometrics, these restrictions usually come from economic theory.
- Homework, suppose you found through economic theory that x_t does not affect z_t contemporaneously, such that $a_{21} = 0$. How would the equations change? Is the system identified? List the equations and parameters found. **No credit if no work is shown**. Send this to my email gmarinmunoz@iadb.org one week from now.

Table of Contents

Impulse-Response Functions

- The Vector Autoregression is a really powerful tool in econometrics, but we are not particularly interested in the coefficients found. We are interested in the forecasting capabilities of this system.
- The way we do this is called an *impulse-response function*
- Before heading into what that is, we first need to setup some interesting facts of VAR models.

Impulse-Response Functions - Stability Condition

- Recall the reduced form VAR

$$Y_t = \tilde{\Phi}_1 Y_{t-1} + \tilde{C}E_t, \tag{33}$$

Rearranging

$$Y_t - \tilde{\Phi}_1 Y_{t-1} = \tilde{C}E_t, \tag{34}$$

$$\tilde{\Phi}_1 L(1) Y_t = \tilde{C}E_t, \tag{35}$$

$$Y_t = (\tilde{\Phi}_1 L(1))^{-1} \tilde{C}E_t. \tag{36}$$

- The stability condition requires that $|I - \tilde{\Phi}_1| \neq 0$. If the values of $\tilde{\Phi}_1 = 1$, what would happen?

Impulse-Response Functions - Cholesky Decomposition

- Therefore, we can express a VAR in terms of its error term (MA representation), this is called a **Wold decomposition**. Remember this is only valid if the VAR is stable!
- With impulse-response functions, we want to see how the system behaves if there is a shock to one of the variables of the system.
- Yet, the variance-covariance matrix of this system does not have zeroes off the diagonal, this means that the error terms are correlated.
- This is quite the problem, we need to use a tool to be able to isolate the shock of only one variable to the system, this is done usually through a method called **Cholesky Decomposition**.

Impulse-Response Functions - Cholesky Decomposition

- Cholesky's theorem states that any positive semidefinite matrix can be decomposed into a lower triangular matrix such that

$$\Omega = AA'. \tag{37}$$

- From Equation (33), we can then redefine the error term as

$$\tilde{E}_t = \tilde{C}E_t = A^{-1}E_t. \tag{38}$$

  By construction, this error term is orthogonal, because its variance-covariance matrix is diagonal

$$V(\tilde{E}_t) = A^{-1}V(E_t)A^{-1\prime} = A^{-1}\Omega A^{-1\prime} = I. \tag{39}$$

- With this decomposition, we can now estimate the impulse-response functions of the system!

Impulse-Response Function - Estimation

- From the Reduced Form VAR

$$Y_t = \tilde{\Phi_1} Y_{t-1} + \tilde{C} E_t. \tag{40}$$

$$Y_t = (\tilde{\Phi_1} L(1))^{-1} \tilde{C} E_t. \tag{41}$$

$$Y_t = (\tilde{\Phi_1} L(1))^{-1} \tilde{E}_t, \tag{42}$$

- where $\tilde{E}_t$ satisfies equation (37)

Now we can shock the system through the modified orthogonal error terms and trace the path a variable takes!

- Wooldridge, J. M. (2016). Introductory econometrics: A modern approach. Cengage learning.

- Sims, C. A. (1980). Macroeconomics and reality. Econometrica: journal of the Econometric Society, 1-48.

- Susmel, R. (n.d.). Lecture 17 multivariate time series Var amp; Svar - Bauer College of Business. Retrieved 2022, from https://www.bauer.uh.edu/rsusmel/phd/ec2-6.pdf

Thank you!

E-mail: gmarinmunoz@iadb.org