Examining Vector Autoregressive Modeling and its Applications to Macroeconomic Data to Hypothesize the Impacts of De-Dollarization

A Senior Comprehensive Project
by
Nahayan Hussain Minhas
Allegheny College
Meadville, PA

March 26, 2024

Submitted to the Department of Mathematics in partial fulfillment of the requirements for the degree of Bachelor of Science.

Project Advisor: Dr. Anthony J. Lo Bello
Second Mathematics Reader: Professor Rachel Weir

I hereby recognize and pledge to fulfill my responsibilities, as defined in the Honor Code, and to maintain the integrity of both myself and the College community as a whole.

Pledge:

Nahayan Hussain Minhas

Acknowledgments

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Contents

1 Introduction

Vector Autoregression (VAR), a stochastic process model, is a statistical modeling technique that captures the relationship between multiple quantities as they change over time. VAR models generalize the univariate Autoregression Model by allowing for a multivariate time series. Due to its simplicity and success at modeling macroeconomic indicators, VAR has become a standard tool for central bankers to construct economic forecasts. In this project, we will explore Vector Autoregressive Modeling and its various forms, followed by its application to the slowly, but surely, occurring monetary phenomenon of De-dollarization.

First, we will explore the specifications of Vector Autoregressive Modeling and define the phenomenon of De-dollarization as it occurs in the world today.

In the second chapter, we will dive down deeper into VAR by examining its sub-forms such as structural VAR and reduced VAR, followed by an insight into the techniques of Bayesian VAR (BVAR) and Deep-learning VAR (DVAR). Applying this to past macroeconomic data, we will forecast recent observations and compare to actual recent observations using Limited Least Squares (LLS) to select the best VAR technique for modeling economic outcomes.

Finally, we will apply it to the problem of De-dollarization by predicting future values of economic indicators and formulating a hypothesis as to the large-scale monetary impact of De-dollarization.

2 Vector Autoregression & De-dollarization

Definition 1 (**Stochastic Process**)**.** A stochastic process is defined as a collection of random variables defined on a common **probability space** $(\Omega, \mathcal{F}, P)$, where Ω is a **sample space**, $\mathcal{F}$ is a σ**-algebra**, and **P** is a probability measure; and the random variables, indexed by some **T**, all take values in the same mathematical space **S**, which must be measurable with respect to some σ**-algebra** Σ.

Definition 2 (**Sample Space**)**.** The sample space Ω is the set of all possible outcomes of a probabilistic experiment.

Definition 3 (**Sigma-algebra**)**.** Let $\mathcal{F}$ be a set whose elements are subsets of Ω. Then, $\mathcal{F}$ is a sigma-algebra if and only if it satisfies the following axioms:

1 $\Omega \in \mathcal{F}$

2 If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (where A^c is the complement of A, also denoted by $\Omega \backslash A$)

3 If $\{A_n\}$ is a countable collection of elements of $\mathcal{F}$ and $A = \cup_{n-1}^{\infty} A_n$, then $A \in \mathcal{F}$.

Definition 4 (**Probability Measure**)**.** Let P be a function that associates a real number to each element of the sigma-algebra $\mathcal{F}$. Then, P is a probability measure if and only if it satisfies the following axioms:

1 If $A \in \mathcal{F}$, then $\mathrm{P}(A) \geq 0$

2 $\mathrm{P}(\Omega) = 1$

3 If $\{A_n\}$ is a countable collection of disjoint elements of $\mathcal{F}$ (i.e., $A_j \cap A_k = \emptyset$ if $j \neq k$), then $\mathrm{P}\left(\cup_{n-1}^{\infty} A_n\right) = \sum_{n-1}^{\infty} \mathrm{P}\left(A_n\right)$.

Definition 5 (**Conditional Likelihood Function for an Unrestricted Vector Autoregression**)**.** The conditional likelihood for a pth order Gaussian Vector Autoregression is given by

$$f_{Y_T, Y_{T-1}, \ldots, Y_1 | Y_0, Y_{-1}, \ldots, Y_{-p+1}}(Y_T, Y_{T-1}, \ldots, Y_1 | Y_0, Y_{-1}, \ldots, Y_{-p+1}; \theta)$$

Definition 6 (**Realization**)**.** A specific sequence or set of observations from a time series process.

Definition 7 (**Covariance**)**.** to be entered

Definition 8 (**Autocovariance**)**.** Given a particular realization such as $\left\{y_i^{(1)}\right\}_{t=-\infty}^{\infty}$ on a time series process, consider constructing a vector $\mathbf{x}_t^{(1)}$ associated with date t. This vector consists of the $[j+1]$ most recent observations on y as of date t for that realization:

$$\mathrm{x}_t^{(1)} \equiv \begin{bmatrix} y_t^{(1)} \\ y_{t-1}^{(1)} \\ \vdots \\ y_{i-j}^{(1)} \end{bmatrix}.$$

We think of each realization $\{y_t\}_{t=-x}^{x}$ as generating one particular value of the vector $\mathbf{x}_t$ and want to calculate the probability distribution of this vector $\mathbf{x}_f^{(f)}$ across realizations i. This distribution is called the joint distribution of $(Y_t, Y_{t-1}, \ldots, Y_{t-j}\ )$. From this distribution we can calculate the j th autocovariance of Y_t (denoted γ_{jt} :

$$\begin{aligned}\gamma_{jt} = &\int_{-\infty}^{\infty}\int_{-x}^{\infty}\cdots\int_{-\infty}^{x}\left(y_t-\mu_t\right)\left(y_{t-j}-\mu_{t-j}\right)\\ &\times f_{Y_t,Y_{t-i},\ldots,Y_{t-j}}\left(y_t, y_{t-1},\ldots,y_{t-j}\right)dy_t dy_{t-1}\cdots dy_{t-j}\\ =&E\left(Y_t-\mu_t\right)\left(Y_{t-j}-\mu_{t-j}\right).\end{aligned}$$

Definition 9 (**Stationarity**)**.** If neither the mean μ_t nor the autocovariances γ_{jt} depend on the date t, then the process for Y_f is said to be covariance-stationary or weakly stationary:

$$\begin{aligned} &E\left(Y_t\right)=\mu && \text{for all } t\\ &E\left(Y_t-\mu\right)\left(Y_{t-1}-\mu\right)=\gamma_j && \text{for all } t \text{ and any } j\end{aligned}$$

Definition 10 (**The Normal Distribution**)**.** The variable Y_t has a Gaussian, or *Normal*, distribution with mean and variance σ^2 if

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \tag{1}$$

We write

$$Y_t \sim N(, \sigma^2)$$

to indicate that the density of Y_t is given by (1).

Definition 11 (**Gaussian Process**). A process $\{Y_\ell\}$ is said to be Gaussian if the joint density

$$f_{\gamma_t,,Y_{t+j_1},\ldots,Y_{t+t_n}}\left(y_t, y_{t+j_t}, \ldots, y_{t+j_n}\right)$$

is Gaussian for any $j_1, j_2, \ldots, j_n$. Since the mean and variance are all that are needed to parameterize a multivariate Gaussian distribution completely, a covariance-stationary Gaussian process is strictly stationary.

Definition 12 (**Multivariate Gaussian Distribution**). Let

$$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)'$$

be a collection of n random variables. The vector $\mathbf{Y}$ has a multivariate Normal, or multivariate Gaussian, distribution if its density takes the form

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-n/2}|\mathbf{\Omega}|^{-1/2} \exp\left[(-1/2)(\mathbf{y}-\boldsymbol{\mu})'\mathbf{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right].$$

The mean of $\mathbf{Y}$ is given by the vector μ :

$$E(\mathbf{Y}) = \boldsymbol{\mu};$$

and its variance-covariance matrix is $\mathbf{\Omega}$:

$$E(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})' = \mathbf{\Omega}.$$

Note that $(\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})'$ is symmetric and positive semidefinite for any $\mathbf{Y}$, meaning that any variance-covariance matrix must be symmetric and positive semidefinite; the form of the likelihood assumes that $\mathbf{\Omega}$ is positive definite.

Definition 13 (**Positive Definite Matrices**)**.** An (n x n) real symmetric matrix **A** is said to be *positive semidefinite* if for any real (n x 1) vector **x**,

$$\mathbf{x'Ax} \geq 0$$

A real symmetric matrix **A** is *positive definite* if for any real nonzero (n x 1) vector **x**,

$$\mathbf{x'Ax} > 0;$$

hence, any positive definite matrix could also be said to be positive semidefinite.

Definition 14 (**Endogenous Variable**)**.** An endogenous variable is a variable that is correlated with the regression error term ϵ_t

Definition 15 (**Trace of a Matrix**)**.** The trace of an ($n \times n$) matrix is defined as the sum of the elements along the principal diagonal:

$$\text{trace}(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{nn}.$$

If **A** is an $(m \times n)$ matrix and **B** is an $(n \times m)$ matrix, then **AB** is an ($m\times$ m) matrix whose trace is

$$\text{trace}(\mathbf{AB}) = \sum_{j=1}^{n} a_{1j}b_{j1} + \sum_{j=1}^{n} a_{2j}b_{j2} + \cdots + \sum_{j=1}^{n} a_{mj}b_{jm} = \sum_{k=1}^{m}\sum_{j=1}^{n} a_{kj}b_{jk}.$$

The product BA is an $(n \times n)$ matrix whose trace is

$$\text{trace}(\mathbf{BA}) = \sum_{k=1}^{m} b_{1k}a_{k1} + \sum_{k=1}^{m} b_{2k}a_{k2} + \cdots + \sum_{k=1}^{m} b_{nk}a_{kn} = \sum_{j=1}^{n}\sum_{k=1}^{m} b_{jk}a_{kj}.$$

Thus,

$$\text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA}).$$

If $\mathbf{A}$ and $\mathbf{B}$ are both $(n \times n)$ matrices, then

$$\text{trace}(\mathbf{A} + \mathbf{B}) = \text{trace}(\mathbf{A}) + \text{trace}(\mathbf{B}).$$

If $\mathbf{A}$ is an $(n \times n)$ matrix and λ is a scalar, then

$$\text{trace}(\lambda\mathbf{A}) = \sum_{i=1}^{n} \lambda a_{ii} = \lambda \cdot \sum_{i=1}^{n} a_{ii} = \lambda \cdot \text{trace}(\mathbf{A})$$

Definition 16 (Kronecker Product). For $\mathbf{A}$ an $(m \times n)$ matrix and B a $(p \times q)$ matrix, the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ is defined as the following $(mp) \times (nq)$ matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$$

The following properties of the Kronecker product are readily verified. For any matrices $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$,

$$(A \otimes B)' = A' \otimes B'$$

$(A \otimes B) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$. Also, for $\mathbf{A}$ and $\mathbf{B}$ both $(m \times n)$ matrices and $\mathbf{C}$ any matrix,

$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C})$$

$$\mathbf{C} \otimes (\mathbf{A} + \mathbf{B}) = (\mathbf{C} \otimes \mathbf{A}) + (\mathbf{C} \otimes \mathbf{B}).$$

Let $\mathbf{A}$ be $(m \times n)$, B be $(p \times q)$, $\mathbf{C}$ be $(n \times k)$, and $\mathbf{D}$ be $(q \times r)$. Then

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD});$$

that is,

$$\begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{nn}\mathbf{B} \end{bmatrix} \begin{bmatrix} c_{11}\mathbf{D} & c_{12}\mathbf{D} & \cdots & c_{1k}\mathbf{D} \\ c_{21}\mathbf{D} & c_{22}\mathbf{D} & \cdots & c_{2k}\mathbf{D} \\ \vdots & \vdots & \cdots & \vdots \\ c_{n1}\mathbf{D} & c_{n2}\mathbf{D} & \cdots & c_{nk}\mathbf{D} \end{bmatrix}$$

$$= \begin{bmatrix} \sum a_{1j}c_{j1}\mathbf{BD} & \sum a_{1j}c_{j2}\mathbf{BD} & \cdots & \sum a_{1j}c_{jk}\mathbf{BD} \\ \sum a_{2j}c_{j1}\mathbf{BD} & \sum a_{2j}c_{j2}\mathbf{BD} & \cdots & \sum a_{2j}c_{jk}\mathbf{BD} \\ \vdots & \vdots & \cdots & \vdots \\ \sum a_{mj}c_{j1}\mathbf{BD} & \sum a_{mj}c_{j2}\mathbf{BD} & \cdots & \sum a_{mj}c_{jk}\mathbf{BD} \end{bmatrix}$$

For $\mathbf{A}(n \times n)$ and $\mathbf{B}(p \times p)$ both nonsingular matrices we can set $\mathbf{C} = \mathbf{A}^{-1}$ and $\mathbf{D} = \mathbf{B}^{-1}$ in [A.4.38] to deduce that

$$(\mathbf{A} \otimes \mathbf{B})\left(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}\right) = \left(\mathbf{AA}^{-1}\right) \otimes \left(\mathbf{BB}^{-1}\right) = \mathbf{I}_n \otimes \mathbf{I}_p = \mathbf{I}_{np}.$$

Thus,

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \left(\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}\right).$$

Definition 17 (**Vector Autoregression** - to be reworked)**.** A VAR describes the evolution of k endogenous variables (an explanatory variable correlated with an error term) over time. Each time period is numbered, $t = 1, ..., T$. The variables are stored in a vector y_t of length k, which might also be called a (k x 1) matrix. This vector y_t is modeled as a function of its previous value. The components of the vector are $y_{i,t}$, i.e., the observation at time t of the i-th variable. For example, if the third variable of the model measures the real interest rate over time, then $y_{3,2008}$ would indicate the real interest rate in the year 2008.

VAR models are characterized by their order p, which refers to the number of previous periods that the model uses. A lag is the value of a variable in a previous time period. So, in general, a pth-order VAR refers to a VAR model that includes lags for the previous p time periods. A pth-order VAR is denoted by VAR(p), or called a VAR with p lags. A pth-order VAR is written as

$$y_t = B_0 + B_1 y_{t-1} + B_2 y_{t_2} + ... + B_p y_{t-p} + \epsilon_t,$$

where variables y_{t-i} indicate the variable's value i time periods before t and are called the ith-lag of y_t. The variable B_0 is a k-vector of constants serving as the intercept of the model. B_i is a time-invariant (an indirectly time-dependent function) (k x k)-matrix and ϵ_t is a k-vector of error terms. The error terms must satisfy the following conditions:

- $E(\epsilon_t) = 0$, i.e., every error term has a mean of 0.
- $E(\epsilon_t \epsilon_t^{'}) = \Omega$, i.e., the contemporaneous covariance matrix is a k x k positive-semidefinite matrix denoted Ω
- $E(\epsilon_t \epsilon_{t-k}^{'}) = 0$ for any non-zero k. There is no correlation across time and, in particular, there is no autocorrelation in individual error terms.

Definition 18 (**The Impulse-Response Function**). A VAR was written in vector $MA(\infty)$ form as

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1 \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Psi}_2 \boldsymbol{\varepsilon}_{t-2} + \cdots$$

Thus, the matrix $\boldsymbol{\Psi}_s$ has the interpretation

$$\frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\varepsilon}_i^{\prime}} = \boldsymbol{\Psi}_s$$

that is, the row i, column j element of $\boldsymbol{\Psi}_s$ identifies the consequences of a oneunit increase in the j th variable's innovation, i.e., an external shock to the system, at date $t\,(\varepsilon_{ji})$ for the value of the i th variable at time $t+s\,(y_{its})$, holding all other innovations at all dates constant.

A plot of the row i, column j element of Ψ_s,

$$\frac{\partial y_{i,t+s}}{\partial \varepsilon_{jt}},$$

as a function of s is called the impulse-response function. It describes the response of $y_{i,t+s}$ to a one-time impulse in y_j with all other variables dated t or earlier held constant.

Definition 19 (**Autocorrelation**)**.** The j th autocorrelation of a covariance-stationary process (denoted ρ_j) is defined as its th autocovariance divided by the variance:

$$\rho_j = \gamma_j / \gamma_0.$$

The terminology arises from the fact that ρ_j is the correlation between Y_t and Y_{r-j} :

$$\operatorname{Corr}(Y_t, Y_{r-j}) = \frac{\operatorname{Cov}(Y_t, Y_{r-j})}{\sqrt{\operatorname{Var}(Y_t)}\sqrt{\operatorname{Var}(Y_{t-j})}} = \frac{\gamma_j}{\sqrt{\gamma_0 \sqrt{\gamma_0}}} = \rho_j.$$

Since ρ_j is a correlation, $|\rho_j| \leq 1$ for all j, by the Cauchy-Schwarz inequality. Notice also that the 0th autocorrelation ρ_0 is equal to unity for any covariance-stationary process by definition.

3 Picking the best VAR model

4 Macroeconomic Applications - De-dollarization

5 Conclusion

References

[1] Taboga, Marco. "Probability space." *Lectures on probability theory and mathematical statistics.* Kindle Direct Publishing, 2021. Online appendix. `https://www.statlect.com/glossary/probability-space`.

[2] Hamilton, James D. *Time Series Analysis.* 1st ed. Princeton University Press, 1994.

[3] Freund, John E. *Mathematical Statistics.* 2nd ed. Prentice Hall, 1971.