

TIME SERIES ECONOMETRICS

Vector Autoregressions & Local Projections

Ara Patvakanian*

First Version: April 2025

This Version: January 2026

Click [HERE](#) for the latest version.

* Contact: arap15@upenn.edu. The original version of this presentation was prepared for an employer; the views expressed herein are my own and do not necessarily represent those of the employer.

Contents

1 Time Series Data & Models

1.1 Growth Rates

1.2 Annualization

1.3 Autoregressive Models

1.4 Potential Difficulties in Time Series

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

Contents

1 Time Series Data & Models

- 1.1 Growth Rates
- 1.2 Annualization
- 1.3 Autoregressive Models
- 1.4 Potential Difficulties in Time Series

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

Contents

1 Time Series Data & Models

1.1 Growth Rates

1.2 Annualization

1.3 Autoregressive Models

1.4 Potential Difficulties in Time Series

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

Growth Rates

With time series data, we are usually interested in the rate of change of the data series through time—the growth rate.

The growth rate between any two observations at times $t - 1$ and t , denoted by g_t , is given by the formula

$$(1) \quad g_t := \frac{x_t - x_{t-1}}{x_{t-1}} = \frac{x_t}{x_{t-1}} - 1$$

which derives from the multiplicative growth relationship

$$(2) \quad x_t = (1 + g_t)x_{t-1}.$$

Growth Rates

Continued

Having a series of growth rates requires an “initial” value of the original series at a time period that overlaps with the growth rates. Once you have this value, you can recover the entirety of the original series.

Suppose that we have x_0 and corresponding growth rates $\{g_t\}_{t=1}^T$. Then we can recover x_τ for $\tau > 0$ by multiplying x_0 by the cumulative product of the growth rates:

$$(3) \quad x_\tau = x_0 \prod_{t=1}^{\tau} (1 + g_t).$$

Going back in time, we symmetrically have:

$$x_{-\tau} = x_0 \prod_{t=-1}^{-\tau} \frac{1}{(1 + g_{t+1})}.$$

Levels and Log-Levels

Time series data is typically given in levels. To transform it into log-levels, simply take the logarithm of the variable.¹

$$\{x_t\}_{t=0}^T \mapsto \{\ln(x_t)\}_{t=0}^T$$

Working with log-levels is very convenient because the first difference of a time series in log-levels, known as the log-difference, approximately equals the growth rate when the period-by-period growth rate is near 0.

$$\ln(x_t) - \ln(x_{t-1}) = \ln(1 + g_t) \approx g_t \quad (\text{by Taylor expanding around } g_t \approx 0)$$

Combining, we get the log-difference approximation of the growth rate, which is often denoted as $\Delta \ln(x_t)$.

$$(4) \quad g_t \approx \Delta \ln(x_t) := \ln(x_t) - \ln(x_{t-1})$$

Another nice result from log-differences is, for any two periods, 0 and $\tau > 0$,

$$(5) \quad \ln(x_\tau) - \ln(x_0) = \sum_{t=1}^{\tau} [\Delta \ln(x_t)] = \sum_{t=1}^{\tau} [\ln(x_t) - \ln(x_{t-1})]$$

1. Note that the domain of the logarithm function is \mathbb{R}^+ . Most measured economic quantities fall inside this subset.

Why Do We Use Log-Differences?

We often use log-differences instead of actual growth rates in practical applications because they

- Are symmetric while percent changes are not;
- Downsize (positive) outliers; and
 - In log-differences, cumulative changes of the same magnitude in different directions cancel out: $-10 \log \text{ p.p.} + 10 \log \text{ p.p.} = 0\%$ cumulative change.
 - With percents, cumulative changes must be multiplied and are thus not symmetric: A +10% increase followed by a 10% decrease is $1.1 \times 0.9 = 0.99$, a 1% cumulative decrease.
- Enable log-log regression specifications where the slope coefficient is interpreted as a percent.

We will exploit log-differences later when computing local projections.

Final Remarks on Log-Differences

Log-differences also makes growth rates for a product or quotient of two variables² quite easy to compute. Defining $x_t := \frac{y_t}{z_t}$ and substituting it into $g_t \approx \ln(x_t) - \ln(x_{t-1})$ yields

$$\begin{aligned} g_t &\approx \Delta \ln(x_t) = \Delta \ln\left(\frac{y_t}{z_t}\right) \\ &= \ln\left(\frac{y_t}{z_t}\right) - \ln\left(\frac{y_{t-1}}{z_{t-1}}\right) \\ &= \ln(y_t) - \ln(z_t) - \ln(y_{t-1}) + \ln(z_{t-1}) \\ &= [\ln(y_t) - \ln(y_{t-1})] - [\ln(z_t) - \ln(z_{t-1})] \\ &= \Delta \ln(y_t) - \Delta \ln(z_t) \end{aligned}$$

which is simply a difference of log-differences (growth rates).

Finally, remember that the Taylor approximation holds only for values of g_t close to 0. If the actual growth rate is large away from 0 (towards $\pm\infty$), we can no longer use log-differences to approximate the growth rate as in (4).

2. The latter shows up when one is calculating the growth rate of a per-capita variable, which is an aggregate quantity divided by the population.

Contents

1 Time Series Data & Models

1.1 Growth Rates

1.2 Annualization

1.3 Autoregressive Models

1.4 Potential Difficulties in Time Series

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

Why Annualize?

We can calculate growth rates between any two periods in the data (for example, t and $t + h$). The smallest period for which we can calculate growth rates for is that of the data itself (t and $t + 1$).

We need a common frame of reference to compare growth rates.

- For example, by how much do you think prices or GDP should grow in a month? How much should they grow by in a year?
- It's easier for the sake of comparison to benchmark on (and target) rates over a commonly-accepted time frame over which the variable in question grows by: 2% inflation, 3% GDP growth, et cetera.
- Annualizing recasts growth rates in terms of similar time frequency: "If the time series were to grow at the same rate over a year, what would the growth rate be?"

You can use the following method to "annualize" to any frequency in the data you wish (such as to 2-year growth rates). Here we will only cover annualizing monthly and quarterly data, although the general formula is provided, too.

Generalized Annualization Formulas

The growth rate between observations indexed at arbitrary t and $t - h$ and annualized over H -observation frequency is given by

$$(6) \quad g_t^* = \left(\frac{x_t}{x_{t-h}} \right)^{\frac{H}{h}} - 1.$$

The counterpart to this formula in logarithms is given by

$$(7) \quad g_t^* \approx \frac{H}{h} (\ln(x_t) - \ln(x_{t-h})).$$

Example Stata Code for Annualizing Monthly Data

Calculate PCE inflation rate at different horizons from the monthly PCE price index release, PCE. We directly apply the formula from above with $H = 12$ since there are 12 months in a year.

```
tsset date // Monthly
gen PCE_12m = ((PCE/L12.PCE)-1)*100           // Annual
gen PCE_1m = ((PCE/L1.PCE)^(12)-1)*100        // 1-month annualized
gen PCE_3m = ((PCE/L3.PCE)^(4)-1)*100         // 3-month annualized
gen PCE_dlog_3m = (log(PCE)-log(L3.PCE))*4*100
           // 3-month annualized using log-differences
```

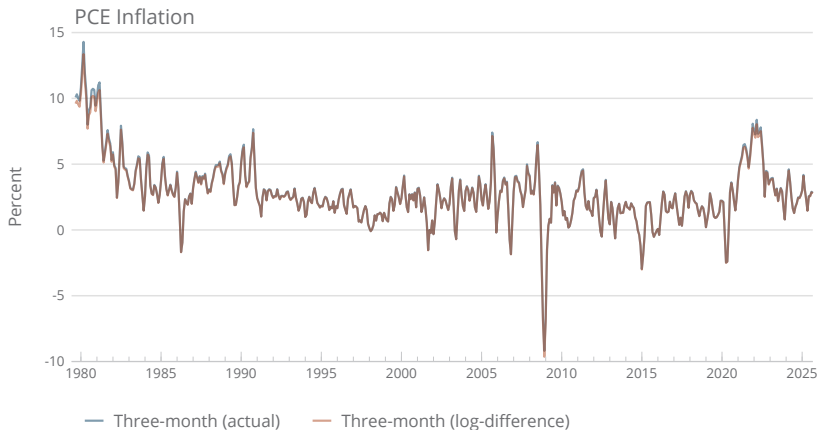
We multiply by 100 to express these growth rates as percents.

You can calculate the “error” between the actual growth rate and its log-difference approximation. Note that the error is largest when the actual growth rate is large.

```
gen difference_3m = PCE_3m - PCE_log_3m
br date PCE_3m PCE_log_3m difference_3m
```

Actual vs. Log-Difference Growth Rates

$$\pi_t = \frac{PCE_t}{PCE_{t-1}} - 1 \quad \pi_t \approx \log(PCE_t) - \log(PCE_{t-1})$$



Example Stata Code for Annualizing Quarterly Data

Calculate the annual and annualized 1-quarter, 2-quarter, 2-year, and 4-year (nominal/real) GDP growth rates from the quarterly (nominal/real) GDP release, GDP. We again apply the formula while noting that now $H = 4$ since there are 4 quarters in a year.

```
tsset date_q // Quarterly
foreach hh of numlist 1 2 4 8 16 { // Convenient for-loop implementation
    gen GDP_`hh'q = ((GDP/L`hh'.GDP)^(4/`hh')-1)*100
    gen GDP_dlog_`hh'q = (log(GDP)-log(L`hh'.GDP))*(4/`hh')*100
}
```

Remarks on Annualization

1. You can annualize “up”, expressing monthly growth rates as annualized growth rates, or annualize “down”, expressing 2-year growth rates as annualized growth rates. In both cases, the annualized growth rates are directly comparable because they are, by definition, expressed as though the same rate over the original period were calculated over a year.
2. Here, we are using the term “annualize” to refer to any transformation that expresses growth rates calculated over two disparate periods to the same period. The latter period does not necessarily need to be yearly. We usually annualize monthly or quarterly growth rates. For instance, we can also “annualize” (written here in quotes as there is no general term) second-by-second and minute-by-minute data to daily.
3. An annual growth rate should not be confused with an annualized growth rate. The former is calculated over a time period spanning a year ($H = 12$ for monthly data) whereas the latter is a growth calculated over a time period shorter or longer than a year expressed as though it grew over a year. However, an annualized growth rate is directly comparable with an annual growth rate whereas a non-annualized growth rate is not.
 - As such, it doesn't make sense to annualize an annual rate as the latter is already calculated over the base period.

Contents

1 Time Series Data & Models

1.1 Growth Rates

1.2 Annualization

1.3 Autoregressive Models

1.4 Potential Difficulties in Time Series

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

Basic Time Series Models

In this section we will briefly cover the following time series models:

- Autoregressions – $AR(p)$
- Random Walks (With and Without Drift) – RW

There are many more time series models. A common trend in time series econometrics is that simpler models are special cases of more general models.

- Moving Average Models – $MA(q)$
- Integrated Models – $I(d)$
- Autoregressive Moving Average Models – $ARMA(p,q)$
 - Combines the $AR(p)$ and $MA(q)$ models
- Autoregressive Integrated Moving Average Models – $ARIMA(p,d,q)$
 - Combines the $AR(p)$, $I(d)$, and $MA(q)$ models
- (Generalized) Autoregressive Conditional Heteroskedasticity – $ARCH(q)$ & $GARCH(p,q)$

Note that many of these models can be estimated using both frequentist (maximum likelihood, including OLS) and Bayesian (MCMC: Gibbs sampling, Metropolis-Hastings) methods.

Autoregressive Processes – AR(p)

Autoregressive (AR) processes describe the dynamics of a time series as an affine relationship between each value and its lags. The simplest of these is the AR(1) model:

$$y_t = \theta + \phi y_{t-1} + \varepsilon_t, \quad \varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

The AR(1) model can be generalized further by incorporating more lags of y_t . The more general AR(p) models each y_t as a linear combination of p of its lags.

$$y_t = \theta + \sum_{\tau=1}^p \phi_\tau y_{t-\tau} + \varepsilon_t, \quad \varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

These models can be directly estimated using OLS. Note that the unconditional moments of the AR(1) process when $0 < \phi < 1$ are given as follows.

$$\mathbb{E}[y_t] = \frac{\theta}{(1 - \phi)}$$
$$\mathbb{V}[y_t] = \frac{\sigma_\varepsilon^2}{(1 - \phi^2)}$$

Proof: $\mathbb{E}[y_t]$

Proof: $\mathbb{V}[y_t]$

Random Walks without Drift

Random walks are themselves special cases of AR models where $\phi = 1$.

$$y_t = y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

The model assumes that each subsequent value in the time series is some error added to the previous value. Note that all random walks have a unit root (since $\beta = 1$) and thus there is no mean-reversion in a random walk.

Random Walks with Drift

A random walk with drift is a random walk with an intercept term; or, equivalently, a random walk without drift is a special case of the random walk with drift model where $\theta = 0$.

$$y_t = \theta + y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

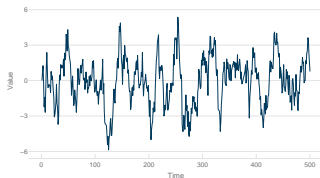
$$\theta \neq 0$$

You can test whether your data generating process is not a random walk (that is, it does not contain a “unit root”) using Dickey-Fuller or KPSS tests.

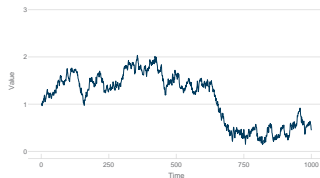
AR(1) & Random Walks

Eugene Fama famously described stock market prices as a random walk (see Fama (1965)). Here are some stochastic simulations of these processes.

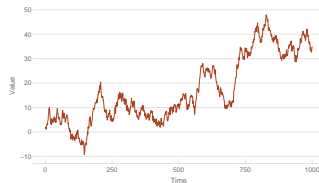
(a) AR(1) with $\beta = 0.9$



(b) Without Drift



(c) With Drift



Contents

1 Time Series Data & Models

1.1 Growth Rates

1.2 Annualization

1.3 Autoregressive Models

1.4 Potential Difficulties in Time Series

2 Motivation for VARs & LPs

3 Local Projections (LPs)

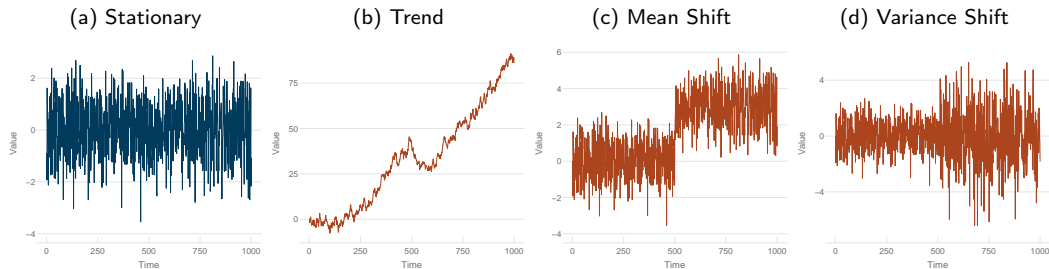
4 Vector Autoregressions (VARs)

References

Appendix

Stationary Processes

A stochastic process such as a time series whose moments (usually mean or variance) does not change is called stationary.



Figures (c) and (d) depict regime shifts or switches. You can use statistical tests to check for nonstationarity.

Autocorrelation (Serial Correlation)

Time series observations tend to be correlated across time with past observations: Intuitively, it is more likely that this month's inflation reading is related to last month's reading more than that from years ago.

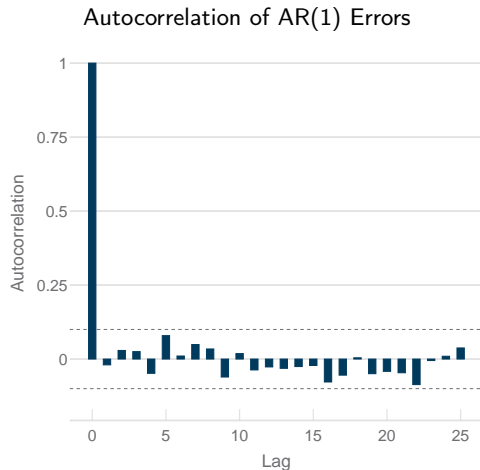
When the model's errors (residuals) are correlated over time, we might be omitted lagged effects or other time-varying variables from the model. Further, these errors may have time-varying variance; that is, $\varepsilon_t \sim \mathcal{N}(0, \sigma_{\varepsilon,t}^2)$. ARCH and GARCH models (not covered here) were developed to deal with this issue.

If the model is misspecified, the typical exclusion requirement necessary for OLS might not hold which biases standard error estimates: $\mathbb{E}[\varepsilon_t \varepsilon_{t+1}] \neq 0$.

Autocorrelation of Residuals

To check this, plot the autocorrelation function of the model's residuals.

- Note here that the autocorrelation of the errors is small because the estimated AR(1) model is exactly the same as the data generating process. This may not be the case when fitting real-world data!
- To address this, consider adding more lags of the LHS variable to the RHS of whatever model you are using.
- Adding too many lags results in an oversaturated model (i.e. overfits the variance in the data). Generally you can pick the number of lags to use using some information criteria or out-of-sample forecast errors (root mean squared forecast error).



Newey-West Estimators

A common fix to estimating robust standard errors with serially correlated residuals are a class of covariance matrix estimators known as heteroskedasticity- and autocorrelation-consistent (“HAC”) estimators. The most commonly used HAC estimators are the Newey-West(-Bartlett) estimators (Newey and West (1987)). For an introduction and references to the econometric literature, see chapter 10 of Hamilton (1994) as well as the original paper.

Newey-West estimators downweight the covariance between residuals further apart from one another according to a user-specified kernel. Setting the kernel to 0 lags reproduces the standard Huber-White (sandwich/heteroskedasticity-robust) covariance estimator.

There exists a literature on how to choose the optimal number of lags (known as the bandwidth) used in computing the kernel. In my experience, controlling for a year (12 months or 4 quarters depending on data frequency) is fine for applied work.

Newey-West Estimators

Continued

The Stata command for Newey-West is `newey` and the kernel is specified using option `lag(#)`. You can directly substitute this command for `regress [...], robust` when working with time series data. Note that it takes longer to run on Stata due to the extra matrix manipulations needed to calculate the standard errors.

```
newey depvar [indepvars] [if] [in] [weight], lag(#) [options]
```

A situation in which you would use HAC standard errors is when your regression variables are annual growth rates or moving averages: In such cases, the overlap between lagged values that are used in calculating the rates or MAs can induce autocorrelation in the residuals.

Trend

One can remove the trend from a time series (and thus, render it stationary) by differencing the series: The first difference removes a linear trend; the second difference removes a quadratic trend; and so on.

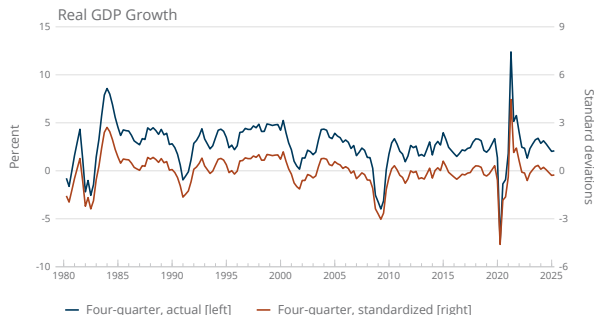
The traditional way in macroeconomics of separating a “trend” component from the “cyclical” component is by applying the HP filter (see Hodrick and Prescott (1981)). This method has fallen out of favor in recent years but many economists still use it.

Standardization

One final commonly-used transformation is to standardize the time series such that its units are in standard deviations around 0. Standardization does not affect inference but helps with the interpretation of coefficient estimates.

```
* These are equivalent
egen mean_GDP_4q = mean(GDP_4q)
egen sd_GDP_4q = sd(GDP_4q)
gen GDP_4q_std = ///
    (GDP_4q - ///
    mean_GDP_4q)/sd_GDP_4q

egen GDP_4q_std_alt = ///
    std(GDP_4q)
```



Contents

1 Time Series Data & Models

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

Identification in Macroeconomics

Identification in macro tends to be extremely difficult because most macroeconomic and financial time series are all related.

- For instance, consumption, investment, taxation, government spending, interest rates, inflation, exchange rates, global news, regional shocks all influence each other—how do you study (a subset of) these all together?
- How do you identify the effect that changes in one variable have on other variables through time?
- How do you identify *dynamic* causal effects?

An Overview of Approaches Used in Dynamic Causal Identification

1. Narrative Identification

- Uses plausibly-exogenous changes manually picked from the historical record as exogenous shocks
- Examples: Friedman and Schwartz (1963) and Romer and Romer (2004)

2. Structural Vector Autoregressions (SVARs)

- Cholesky identification (“internal instruments”): Short-run (Christiano, Eichenbaum, and Evans (1999, 2005)) and long-run (Blanchard and Quah (1989)) restrictions
- Proxy SVAR (“external instruments”): Stock (2008), Stock and Watson (2012), and Mertens and Ravn (2013)
- Sign restrictions: Uhlig (2005)

3. High-Frequency Identification

- Using high-frequency changes in federal funds rate or eurodollar futures around tight windows on announcement days to isolate the causal effect of policy changes
- Examples: Bauer and Swanson (2023)

4. Combinations of the Above

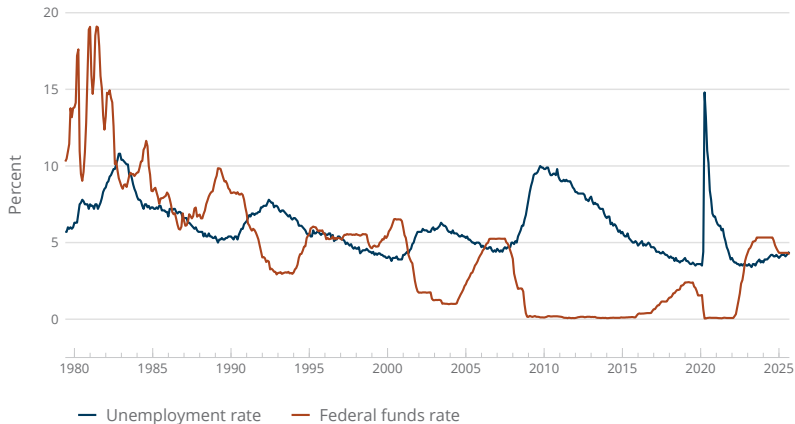
- Narrative with high-frequency data: Känzig (2021)
- SVARs with high-frequency data: Gertler and Karadi (2015), Miranda-Agrippino and Ricco (2021), and Nunes, Ozdagli, and Tang (2022)

From Macroeconomic Theory to Empirics

Have you ever heard something along the lines of “monetary policy affects the economy with long and variable lags”? This assertion derives from research that relies on either VARs or LPs.

Vector autoregressions (VARs) and local projections (LPs) are the most ubiquitous econometric methods used to estimate dynamic effects and perform causal identification when studying multiple endogenous time series. In these slides we will stick to empirical methods, but since (log-)linearized solutions of macroeconomic models can typically be represented in state-space form, they readily lend themselves to VAR or LP analysis.

Example: How Does Monetary Policy Affect Unemployment?



Historically, the unemployment rate rises when the federal funds rate is low! Does this mean \downarrow FFR \Rightarrow \uparrow UR?

Endogeneity in Simple Regressions

Estimating the naïve regression over 1980:01–2025:06

$$UR_t = \alpha + \beta i_t + \varepsilon_t$$

yields a statistically significant estimate of $\hat{\beta} = 0.05$ (0.02). This does not accord with our intuition. So what's going on?

- The UR and the FFR are endogenous!
- The regression is actually capturing the comovement of these variables in response to shocks in each other and in other variables not included in the regression.
- The regression is invalid because (1) it contains omitted variable bias and (2) it violates the exclusion requirement.

We will now look at how LPs and VARs can be used to overcome these issues.

Contents

1 Time Series Data & Models

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

Causal Identification in a Static Cross-Sectional Context

Consider the following simple linear regressions.

$$y_i = \beta x_i + \varepsilon_i$$
$$\log(y_i) = \gamma \log(x_i) + \nu_i$$

If x_i is exogenous to y_i , then β yields the estimated causal effect on y_i in y_i 's units of raising x_i by 1 unit. Also, γ yields the estimated causal effect of y_i *in percent* of raising x_i by 1 percent.

Moreover, if x_i is endogenous to y_i but you have an instrumental variable z_i , you can identify the causal effect of x_i on y_i by using an instrumental variables (IV) or two-stage least squares (2SLS) regression.

Impulse responses (whether from LPs or an identified SVAR) are an extension of this procedure when (1) the system includes multiple variables related across time and contemporaneously and (2) a shock to one variable takes time to affect the other variables.

Impulse Responses

An impulse response function (IRF) gives the causal effect that one variable (the impulse) has on another (the response) in the dynamic system. Since we are working with time series data, impulse responses are calculated and reported at horizons $h \geq 0$ where $h = 0$ is known as the “on impact” response.

Impulse response functions are the dynamic equivalent of static estimates of a causal effect; for that reason, they make sense only when a horizon is provided. The statement “a 1 percentage point increase in interest rates decreases inflation by 3 percentage points” begs the follow-up question “how long after the increase in interest rates do we see the fall in the inflation rate?” The timing matters!

Instruments & Shocks

When running LPs, we usually rely on identified shocks that come from some other economist's website or replication package. There exists a huge literature on identifying shocks, the surface of which we will barely scrape here.

A shock is a variable that is exogenous to a dynamic system. As such, it can be used to identify causal effects because it satisfies the exclusion requirement that caused our naïve regression to fail. In regressions we can either use shocks directly or as instruments for an endogenous variable.

Macroeconomic shocks can broadly be binned into shifters of aggregate demand and aggregate supply. The demand- and supply-side shocks that I have most often seen used are monetary policy shocks³ and oil supply shocks, respectively.

3. As you can imagine, it's very difficult to identify exogenous changes to monetary policy: After all, the Fed hires hundreds of economists to make sure that monetary policy is as endogenous as can be!

Local Projections

Local projections (LPs) were first introduced by Jordà (2005) as an alternative to VARs for estimating impulse responses. LPs are just a bunch of regressions estimated at various horizons that are especially easy to use once one has an identified shock.

LPs...

- estimate the same IRFs as VARs (in population: see discussion later); and
- tend to be easier to set up and estimate than VARs;
- are easier to use in an external instrument context;
- are easier to use in the presence of potential nonlinearities and/or interaction effects; and
- do not have a potential noninvertibility problem as SVARs.

LPs take their name from the fact that instead of compounding the estimated dynamics of the system as in VARs, they instead compute IRFs by calculating responses locally to each forecast horizon.

Local Projections

The general form of a local projection is

(Local Projections)

$$y_{t+h} = \alpha^{(h)} + \beta^{(h)} z_t + w_t' \gamma^{(h)} + \sum_{\tau=1}^p \eta_{t-\tau}^{(h)} y'_{t-\tau} + \sum_{\tau=1}^p x'_{t-\tau} \delta_{t-\tau}^{(h)} + \varepsilon_{t+h}^{(h)}, \quad h \in \{0, \dots, H\}$$

where

- y_t is the scalar variable whose response we are interested in;
- z_t is a scalar shock occurring at time t ;
- w_t is a vector of contemporaneous control variables;
- $x_{t-\tau}$, $\tau \in \{0, \dots, p\}$ are lagged vectors of controls that might include $z_{t-\tau}$ and $w_{t-\tau}$;
- $\varepsilon_{t+h}^{(h)}$ is the regression residual; and
- $\alpha^{(h)}, \beta^{(h)}, \gamma^{(h)}, \delta_{t-\tau}^{(h)}, \eta_{t-\tau}^{(h)}$ are regression coefficients.

The impulse response function to the shock z_t is given by estimating the LP for each h and plotting the coefficients $\beta^{(h)}$ with their associated confidence bands.

LP Estimation Assumptions

3 Kinds of LP Specifications

The impulse response coefficient has a different interpretation depending on whether the response variable is a difference and the units of the response variable. The response variable can either be given

1. in levels (percents) or log-levels;
2. as a a period-by-period difference ($t + h - 1$ to $t + h$); or
3. as a cumulative multiplier ($t - 1$ to $t + h$).

We will cover all 3 of these cases individually in the coming slides. Before we do, let's review the Frisch-Waugh-Lovell theorem, which will help us with understanding these different specifications.

Frisch-Waugh-Lovell Theorem (FWL)

Frisch-Waugh-Lovell Theorem: Controlling for a set of variables on the right-hand side of a regression is equivalent to the regression where the left- and right-hand-side variables are residualized with respect to the control variables.

The regression $Y = X\beta + Z\gamma + \varepsilon$ produces the same coefficients and standard errors for β as the regression $\tilde{Y} = \tilde{X}\beta + \nu$ where $\tilde{Y} := Y - Z\hat{\gamma}^Y$ and $\tilde{X} := X - Z\hat{\gamma}^X$. (\tilde{Y} and \tilde{X} are the residuals from the regressions $Y = Z\gamma^Y + \tilde{Y}$ and $X = Z\gamma^X + \tilde{X}$.) Controlling for Z is equivalent to partialing it out from both Y and X .

Specification #1: Response Variable in Levels or Percent

Suppose y is in levels or is a percent.⁴ For $\beta^{(h)}$ to have the interpretation of a change in y from $t - 1$ to $t + h$, we need to control for lags of y on the RHS. The estimates of $\beta^{(h)}$ will be in the same units as y .

$$y_{t+h} = \alpha^{(h)} + \beta^{(h)} z_t + \sum_{\tau=1}^p \lambda_{t-\tau}^{(h)} y_{t-\tau} + \sum_{\tau=1}^p x'_{t-\tau} \delta_{t-\tau}^{(h)} + \varepsilon_{t+h}^{(h)}, \quad h \in \{0, \dots, H\}$$

Note that this is equivalent to subtracting y_{t-1} from both sides of the regression equation and writing the LP as follows. This specification is typical if y is a rate such as the unemployment rate or a credit spread.

$$y_{t+h} - y_{t-1} = \alpha^{(h)} + \beta^{(h)} z_t + \hat{\lambda}_{t-1}^{(h)} y_{t-1} + \sum_{\tau=2}^p \lambda_{t-\tau}^{(h)} y_{t-\tau} + \sum_{\tau=1}^p x'_{t-\tau} \delta_{t-\tau}^{(h)} + \varepsilon_{t+h}^{(h)}, \quad h \in \{0, \dots, H\}$$

where $\hat{\lambda}^{(h)} = \lambda^{(h)} - 1$. Note that by FWL, the coefficient of interest, $\beta^{(h)}$, is equivalent between the two specifications!

4. Note: Here, “percent” does not refer to the units of the variable. The variable could be measured in percents, such as the unemployment rate, or in percentage points, such as the corporate bond spread.

Specification #1: Response Variable in Log-Levels

Suppose y is in log-levels. Controlling for lags of y on the RHS then yields $\beta^{(h)}$ in percentage points units. This specification is typically used for variables which are not already percents in their “level” form such as indices or measured quantities.

This specification is the exact same except that the units of the coefficient $\beta^{(h)}$ is in percentage points. Note, however, that it does not make sense to take logs of a percent variable since simple differences are already in percentage points units.

Specification #1: Interpretation

When using specification #1, if z is in levels (a percent) [in log-levels] and y is...

- ...in levels, then y changes by $\beta^{(h)}$ *units* for each unit (percentage point) [percent] increase in z ...
- ...in log-levels, then y changes by $\beta^{(h)}$ *percent* for each unit (percentage point) [percent] increase in z ...
- ...a percent (such as a yield or spread), then y changes by $\beta^{(h)}$ *percentage points* for each unit (percentage point) [percent] increase in z ...

... h periods after the impact of the shock.

Specification #2: Response Variable as a Difference

Define the period-by-period difference as

$$\begin{aligned}\Delta y_t &:= y_t - y_{t-1} \\ \implies \Delta y_{t+h} &= y_{t+h} - y_{t+h-1}.\end{aligned}$$

The LP where both the response variable and its controls are differences is given as follows.

$$\Delta y_{t+h} = \alpha^{(h)} + \beta^{(h)} z_t + \sum_{\tau=1}^p \lambda_{t-\tau}^{(h)} \Delta y_{t-\tau} + \sum_{\tau=1}^p x'_{t-\tau} \delta_{t-\tau}^{(h)} + \varepsilon_{t+h}^{(h)}, \quad h \in \{1, \dots, H\}$$

If y is in levels (a percent) [in log-levels], then $\beta^{(h)}$ gives the units (percentage points) [percent] change in y to an increase in z^5 h periods after the shock.

5. Similarly, this can either be for a unit (if z is a level), a percentage point (if z is a percent), or a percent (if z is a log-level) increase in z .

Specification #3: Response Variable as a Cumulative Sum

Let y be either a level, log-level or a percent. Define the cumulative sum of Δy from $t - 1$ to $t + h$ as

$$\tilde{\Delta}y_{t-1:t+h} := \sum_{\tau=-1}^h \Delta y_{t+\tau}.^6$$

We run this specification with $\tilde{\Delta}y_{t-1:t+h}$ on the LHS while controlling for lags of Δy_{t+h} on the RHS.

$$\tilde{\Delta}y_{t-1:t+h} = \alpha^{(h)} + \beta^{(h)}z_t + \sum_{\tau=1}^p \lambda_{t-\tau}^{(h)} \Delta y_{t-\tau} + \sum_{\tau=1}^p x'_{t-\tau} \delta_{t-\tau}^{(h)} + \varepsilon_{t+h}^{(h)}, \quad h \in \{0, \dots, H\}$$

In this specification, $\beta^{(h)}$ carries the interpretation of a cumulative multiplier. In other words, $\beta^{(h)}$ is the cumulative effect of the shock on y h periods after impact.

6. Note that if y is a log difference (as is usually the case when using this specification), the sum telescopically collapses to $\tilde{\Delta}y_{t-1:t+h} \equiv y_{t+h} - y_{t-1}$ which is the percent change in y from $t - 1$ and $t + h$.

Specification #3: Interpretation

Specification #3 is useful for answering the question, “what is the total impact of the shock z on y after h periods have passed” instead of “how does y respond to z after h periods”. While these two questions appear similar, there is a substantive and quantitative difference between them. Saying that an oil supply shock raises prices by 1 percentage point *over* 12 months is much different than saying that it raises prices by 1 percentage point 12 months after impact.

There are variations on this specification depending on y 's units and the kinds of lags we control for. Keeping track of whether variables are flows and stocks is essential. Specification #1 gives the IR of a stock; #2 gives the IR of a flow; and #3 gives the cumulative impact of flows on the underlying stock.

Remarks on LP Specifications

Each of the three specifications carries a different interpretation for $\beta^{(h)}$, so it's important to check whether the specification you are using corresponds with your desired interpretation.

For example, it is possible to recover the IRF of specification #3 by estimating #2 and, at each h , showing $\sum_{\tau=0}^h \beta^{(\tau)}$. However, you cannot simply sum over the standard errors. It is thus usually advisable to directly estimate specification #3.

Picking H and p

We have a choice over H and p —the number of horizons for which we compute the LPs and the number of lags of the control variables we include, respectively.

In the monthly or quarterly macro context, IRFs are usually reported to horizons of 2 to 5 years (24 to 60 months or 8 to 20 quarters) because LPs and VARs are meant to capture the system's short-term dynamics. Going too far renders the responses susceptible to noise as the control variables no longer effectively capture trends and comovements in the outcome variables at long horizons.

As such, one typically controls for a year or two of lags. Including less lags would fail to capture trends in variables and the short- to medium-term dynamics of the system. Including more lags runs the risk of soaking up the cyclical variation in the system in the control variables instead of the exogenous shock. In principle, it is possible to use the Akaike (AIC) or Bayesian (BIC) information criteria to choose the number of lags.

Example: Transmission of MP to Real and Financial Variables

We are interested in the response of 6 variables to a monetary policy shock MP_t (sourced from Nunes, Ozdagli, and Tang (2022)):

- Employment (log-level $\times 100$)
- Core PCE price index (log-level $\times 100$)
- House price index (log-level $\times 100$)
- Outstanding corporate bonds (log-level $\times 100$)
- S&P 500 stock index (log-level $\times 100$)
- Gilchrist and Zakrajšek (2012) corporate bond credit spread (percentage points)

The LP equations are different for the variables depending on whether it is a percent (GZ spread) or log-level (all other variables after taking logarithms and multiplying by 100). Due to the log-level transformation, all response units are in percentage points.

Example: MP Transmission

Continued

Here are more implementation details about the LPs.

- All time series are balanced over 1985:Q1–2019:Q4.
- $p = 4$ quarterly lags
- RHS variables thus span 1985:Q1–2018:Q4; LHS variables span 1986:Q1–2019:Q4
- Confidence intervals were calculated with the Newey-West correction with 5 autocorrelated lags.

Since MP_t is scaled to raise the 1-year bond yield by 1 percentage point (this is a implausibly huge shock), we rescale it to raise the 1-year bond yield by 25 basis points. An easy shortcut way of doing this is by simply multiplying it by the reciprocal of the factor you wish you rescale it by: $MP_t \mapsto 4 \times MP_t$.⁷

7. The non-shortcut way involves calculating the on-impact ($h = 0$) LP and rescaling the shock by the estimated coefficient times the reciprocal of the desired rescaling factor.

Example: MP Transmission

Continued

The LP for the GZ credit spread at horizon h takes the form

$$\begin{aligned} \text{GZS}_{t+h} = & \alpha^{(h)} + \beta^{(h)} \text{MP}_t + \sum_{\tau=1}^4 \left(\text{Empl}_{t-\tau} \delta_{t-\tau}^{\text{Empl},(h)} + \text{PCEX}_{t-\tau} \delta_{t-\tau}^{\text{PCEX},(h)} + \text{OCB}_{t-\tau} \delta_{t-\tau}^{\text{OCB},(h)} \right. \\ & \left. + \text{S\&P}_{t-\tau} \delta_{t-\tau}^{\text{S\&P},(h)} + \text{HPI}_{t-\tau} \delta_{t-\tau}^{\text{HPI},(h)} + \text{GZS}_{t-\tau} \delta_{t-\tau}^{\text{GZS},(h)} \right) + \varepsilon_{t+h}^{\text{GZS},(h)}. \end{aligned}$$

For the 5 other variables, I instead used the cumulative log-difference on the LHS.

$$\tilde{\Delta} \text{S\&P500}_{t-1:t+h} := \text{S\&P500}_{t+h} - \text{S\&P500}_{t-1}$$

I kept the RHS the same. Note that I should have included fewer lags of the LHS variable on the RHS for all variables except the GZ credit spread (as these are computed as differences on the LHS), but by the Frisch-Waugh-Lowell theorem this does not affect the coefficient of interest, $\beta^{(h)}$.

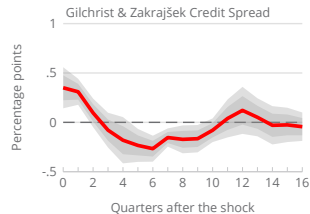
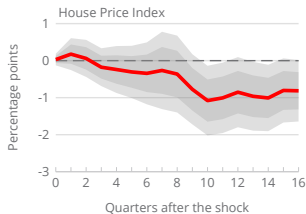
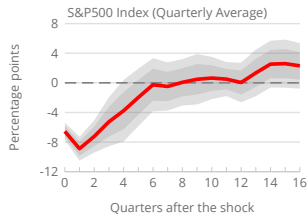
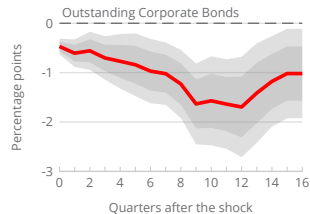
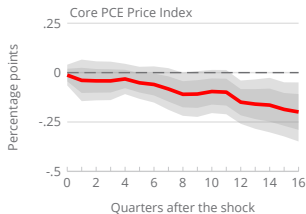
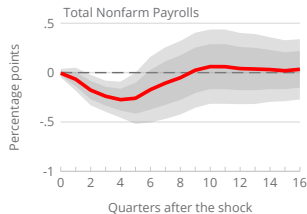
Exercise: Local Projections

Let's run these LPs using both monetary policy (MP) shocks from Nunes, Ozdagli, and Tang (2022) and oil supply news (OSN) shocks from Känzig (2021).

The MP shock is scaled such that it raises 1-year Treasury yield by 1 percentage point on impact. The OSN shock, on the other hand, is scaled such that it raises the real price of oil by 10% on impact.

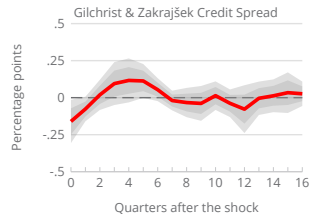
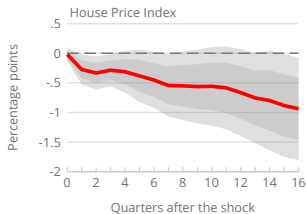
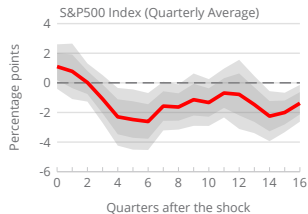
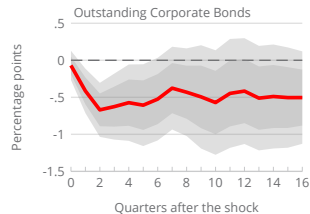
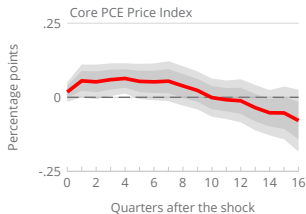
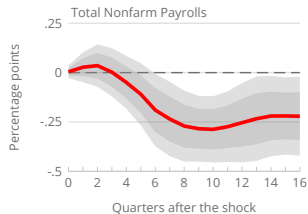
```
run_LP.sdo
```


Transmission of Monetary Policy Shocks to Real and Financial Variables



■ Estimate ■ 90% confidence interval ■ 68% confidence interval

Transmission of Oil Supply News Shocks to Real and Financial Variables



■ Estimate ■ 90% confidence interval ■ 68% confidence interval

Extensions to LPs

The most commonly-used extensions to LPs are:

- Panel LPs, used for estimating dynamic impulse responses while controlling for cross-sectional variation;
- Local projections with instrumental variables (LP-IV), first introduced by Stock and Watson (2018) and used for instrumenting one of the endogenous variables with a shock (instead of running with the shock directly on the RHS of the LP). These are equivalent to proxy SVARs;
- Difference-in-differences local projections (LP-DiD), introduced by Dube et al. (2025) and used for estimating differential responses among categories dynamically; and
- Bayesian LPs (BLPs): introduced by Miranda-Agrippino and Ricco (2021) and further expanded upon in Ferreira, Miranda-Agrippino, and Ricco (2025), estimation with informative priors.

Panel LPs and LP-IVs can easily be implemented in Stata by using the `reghdfe`, `ivreg2`, and `ivreghdfe` (or substitute) commands.

Packages and References for Local Projections

Òscar Jordà's [page on local projections](#) has plenty of papers and references including for many extensions of LPs. Some packages he highlights are:

- Stata
 - locproj
 - lpdid
- R
 - lpirfs

It is fairly straightforward to manually program local projections by either running multiple looping through horizons or by appending and running a single regression.

Contents

1 Time Series Data & Models

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

Vector Autoregressions

A vector autoregression (VAR) is a generalization of an AR process to multiple variables. VARs are used for identifying shocks,⁸ deriving impulse responses, historically decomposing innovations in dynamic systems, forecast error variance decompositions, forecasting in the short- to medium-runs, and much more.

In this section, we will cover the basics of VARs with as little notation as possible. For a more thorough treatment, please see the supplementary material section of the appendix. For an excellent treatment of these topics (and a slightly-different approach to my coverage here), see [Ambrogio Cesa-Bianchi's slides](#).

8. VARs are usually used to identify the shocks that we drive LPs with.

Impulse Responses in an AR(1)

Suppose x follows the following mean-0 AR(1) process. All of the forthcoming results can be extended to accommodate for $p > 1$ and a constant, but let's proceed with the simple case for the sake of exposition.

$$x_t = \phi x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1)$$

The impulse response to a 1 percentage point increase in x at time $t = 0$ is given by setting $\varepsilon_0 = 1$ and iterating the AR(1) process forward (i.e., applying it over and over again). We let $x_{t-1} = 0$ and the sequence $\varepsilon_t = 0$ for $t > 0$. Then:

$$x_0 = \phi x_{-1} + \varepsilon_0 = \phi \cdot 0 + 1 = 1$$

$$x_1 = \phi x_0 + \varepsilon_1 = \phi \cdot 1 + 0 = \phi$$

$$x_2 = \phi x_1 + \varepsilon_2 = \phi \cdot \phi + 0 = \phi^2$$

$$\implies x_H = \phi^H$$

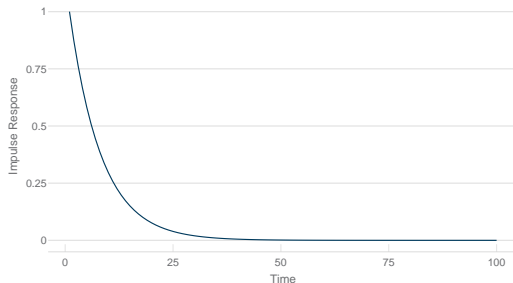
Impulse Responses in an AR(1)

Continued

If the AR process is stable—meaning $\phi < 1$ —then a shock in the variable dissipates to 0 since

$$\lim_{h \rightarrow \infty} x_h = \lim_{h \rightarrow \infty} \phi^h = 0.$$

The closer to 0 that ϕ is, the faster the rate of dissipation. Similarly, the closer to 1 that ϕ is, the longer takes for the effects of the shock to dissipate. Here is the empirical IRF of a stochastic simulation when $\phi = 0.9$.



If $\phi > 1$, then the AR is explosive and the limit does not exist, meaning the effects of a shock are greater the further away from the impact period we go. In most contexts, this does not make sense intuitively.

From AR(1) to VAR(1)

Now suppose we have two variables, inflation (π) and interest rates (i), which follow the following VAR(1) process written in scalar form:

$$\begin{aligned}\pi_t &= \phi_{1,1}\pi_{t-1} + \phi_{1,2}i_{t-1} + \nu_t^\pi \\ i_t &= \phi_{2,1}\pi_{t-1} + \phi_{2,2}i_{t-1} + \nu_t^i\end{aligned}$$

We can stack these equations up and express the VAR compactly in companion form.

$$\underbrace{\begin{bmatrix} \pi_t \\ i_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix}}_{\Phi} \underbrace{\begin{bmatrix} \pi_{t-1} \\ i_{t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} \nu_t^\pi \\ \nu_t^i \end{bmatrix}}_{\nu_t}$$

As its name suggests, a VAR is an AR in vectors. Just like in an AR, if we included p lags of X_t on the RHS, we would have a VAR(p) model. Note that since every VAR(p) can be written in VAR(1) form, we will stick to the latter here.

Reduced-Form VARs

First, we make sure that π and i are stationary (otherwise, we would need to include more lags and/or a constant term). Once we have the reduced-form VAR set up, we can directly estimate it using OLS or Bayesian linear regression. Thus, we estimate

(Reduced-Form VAR)
$$X_t = \Phi X_{t-1} + \nu_t$$

where

- X_τ , $\tau \in \{t-1, t\}$ represents the system variables at time τ ;
- Φ is the reduced-form coefficient matrix;
- ν_t represents the reduced-form residuals/innovations; and
- $\Omega := \mathbb{E}[\nu_t \nu_t']$ is the reduced-form covariance matrix.

Forecasting with VARs

Suppose we are at time T and wish to produce a forecast for H periods from now. We can generate forecasts for all of the variables in the VAR⁹ by iterating it forward.¹⁰

$$\begin{aligned}X_T &= \Phi X_{T-1} + \nu_t \\ \mathbb{E}[X_{T+1}] &= \hat{\Phi} X_T \\ \mathbb{E}[X_{T+2}] &= \hat{\Phi} \mathbb{E}[X_{T+1}] = \hat{\Phi}^2 X_T \\ \mathbb{E}[X_{T+3}] &= \hat{\Phi} \mathbb{E}[X_{T+2}] = \hat{\Phi}^3 X_T \\ &\vdots \\ \mathbb{E}[X_{T+H}] &= \hat{\Phi}^H X_T\end{aligned}$$

The general result is obtained by induction. Producing confidence or credibility bands in this context is slightly more complex so we will not cover these here.

9. Note: You can also forecasting using LPs (not covered in these slides).

10. Many policy institutions nowadays rely on Bayesian VARs (BVARs) to produce their forecasts. One reason for this is because credible intervals produced by BVARs permit probabilistic claims about the future paths.

The Identification Problem in VARs

Though we can use the reduced-form VAR to produce forecasts, we cannot use it to perform causal identification or produce IRFs.

$$\begin{aligned}\pi_t &= \phi_{1,1}\pi_{t-1} + \phi_{1,2}i_{t-1} + \nu_t^\pi \\ i_t &= \phi_{2,1}\pi_{t-1} + \phi_{2,2}i_{t-1} + \nu_t^i\end{aligned}$$

Here, ν_t^π and ν_t^i are the amount in π_t and i_t that cannot be explained by lagged values of π and i . We have no way of knowing how much of these is due to a shock in π or a shock in i . This is known as the identification problem. Another way of seeing this issue is by further decomposing the vector ν_t into what are known as the structural errors or innovations, represented by ε .

$$\begin{aligned}\nu_t^\pi &= a'_{1,1}\varepsilon_t^\pi + a'_{1,2}\varepsilon_t^i \\ \nu_t^i &= a'_{2,1}\varepsilon_t^\pi + a'_{2,2}\varepsilon_t^i\end{aligned}$$

For now, we ignore the \cdot' notation.

The Identification Problem in VARs

Continued

We can do the same thing to each of the ϕ 's by decomposing them into coefficients a' and b . Collecting everything into matrices, we have:

$$\underbrace{\begin{bmatrix} \pi_t \\ i_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} a'_{1,1} & a'_{1,2} \\ a'_{2,1} & a'_{2,2} \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix}}_B \underbrace{\begin{bmatrix} \pi_{t-1} \\ i_{t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} a'_{1,1} & a'_{1,2} \\ a'_{2,1} & a'_{2,2} \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} \varepsilon_t^\pi \\ \varepsilon_t^i \end{bmatrix}}_{\varepsilon_t}.$$

$\underbrace{\hspace{10em}}_{\Phi} \qquad \underbrace{\hspace{10em}}_{\nu_t}$

In all, this is just a decomposed version of what we had before:

$$X_t = \Phi X_{t-1} + \nu_t$$

where

$$\Phi \equiv A^{-1}B,$$

$$\nu_t \equiv A^{-1}\varepsilon_t.$$

Structural Vector Autoregressions

Recall that multiplying a matrix by its inverse (if the latter exists) yields the identity matrix: $AA^{-1} = I$. Also recall that any matrix multiplied by the identity matrix yields the original matrix: $IB = BI = B$. By left-multiplying both sides of the reduced-form VAR by the matrix A , we recover the structural form of the VAR (SVAR).

$$\begin{aligned}X_t &= \Phi X_{t-1} + \nu_t \\ \iff X_t &= A^{-1}BX_{t-1} + A^{-1}\varepsilon_t \\ \implies AX_t &= AA^{-1}BX_{t-1} + AA^{-1}\varepsilon_t \\ &= IBX_{t-1} + I\varepsilon_t \\ \text{(Structural VAR)} \quad &AX_t = BX_{t-1} + \varepsilon_t\end{aligned}$$

Written out, this looks like

$$\underbrace{\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}}_A \underbrace{\begin{bmatrix} \pi_t \\ i_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix}}_B \underbrace{\begin{bmatrix} \pi_{t-1} \\ i_{t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t^\pi \\ \varepsilon_t^i \end{bmatrix}}_{\varepsilon_t}.$$

Here, B is the structural coefficient matrix while A is the inverse impact matrix (meaning A^{-1} is the impact matrix).

SVAR Identification

In order to estimate IRFs, the reduced-form VAR is not enough: IRFs using the response to a reduced-form innovation tell us little about shocks, which are definitionally orthogonal with respect to movements in the other variables. We thus need to identify the SVAR.

Notice that from the reduced form of the VAR, we retrieve estimates for the coefficient matrix Φ and the reduced-form covariance matrix Ω . We are, however, interested in the structural matrices A^{-1} and B . Since $\Phi = A^{-1}B$, once we have A^{-1} we immediately get B , too.

Identifying the SVAR amounts to placing restrictions or obtaining estimates for the matrix A^{-1} . Here, we will only focus on the intuition behind Cholesky identification—a technique that was famously featured in Christiano, Eichenbaum, and Evans (2005) but fallen out of use since in favor of more agnostic approaches. Regardless, it is useful as a simple way of understanding how SVARs work.

Cholesky Identification

With some algebra (see the appendix), we can show that $\Omega = A^{-1} (A^{-1})'$. The Cholesky decomposition of Ω yields $\Omega = PP'$ where P is a positive semi-definite, lower-triangular matrix. Setting $A^{-1} = P$ identifies the SVAR.

$$\begin{bmatrix} \pi_t \\ i_t \end{bmatrix} = \begin{bmatrix} a'_{1,1} & 0 \\ a'_{2,1} & a'_{2,2} \end{bmatrix} \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ i_{t-1} \end{bmatrix} + \begin{bmatrix} a'_{1,1} & 0 \\ a'_{2,1} & a'_{2,2} \end{bmatrix} \begin{bmatrix} \varepsilon_t^\pi \\ \varepsilon_t^i \end{bmatrix}.$$

A^{-1} is known as the impact matrix because it reflects the period- t impact that a structural shock in π or i has on π or i . This procedure essentially sets $a'_{1,2} = 0$ in the simple 2-variable VAR(1) model. Economically what this amounts to is a contemporaneous timing restriction.

Cholesky Identification

Continued

Without loss of generality, we can also set $a'_{1,1} = a'_{2,2} = 1$. Writing out the product of the last term in the identified SVAR, we see how the timing restriction works in this case.

$$\nu_t = A^{-1}\varepsilon_t = \begin{bmatrix} 1 & 0 \\ a'_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_t^\pi \\ \varepsilon_t^i \end{bmatrix} \implies \begin{cases} \nu_t^\pi = \varepsilon_t^\pi \\ \nu_t^i = a'_{2,1}\varepsilon_t^\pi + \varepsilon_t^i \end{cases}$$

Contemporaneously (that is, in the same period), a structural shock in ε_t^π is the only thing that affects the value of π_t . However, the value of i_t is contemporaneously affected by both shocks to π_t as well as to the interest rate, i_t .

The restriction states that interest rates respond immediately to changes in the inflation rate, but that a change in the interest rate in period t only starts affecting inflation in period $t + 1$.

Cholesky Identification

Continued

In other words, Cholesky identification assumes that variables *ordered later* are unaffected by shocks in variables *ordered earlier*. In our example, inflation is ordered first in the VAR, so it is only contemporaneously affected by shocks to itself (and shocks to it affect all other variables contemporaneously); interest rates are ordered last in the VAR, so they contemporaneously respond to shocks to all earlier variables but do not contemporaneously affect any other variable.

Cholesky identification produces radically different results depending on the ordering of the variables in the VAR because when using Cholesky identification, this ordering encodes the contemporaneous restrictions placed on the variables. The structure is thus not agnostic with respect to the data, but rather imposed by the econometrician, who likely has outside evidence or a theoretical model that suggests which restrictions to use.

Impulse Responses in an Identified SVAR

Once we have an identified SVAR, we can calculate impulse responses almost in the same way as in the AR(1). I write the IRFs in matrix notation because the algebra becomes complex.

Suppose there is a structural shock at $t = 0$ that raises π_t by 1 percentage point when all variables are at steady state (X_0 is the 0 matrix here). There is no structural shock in i at time $t = 0$. Thus, $\varepsilon_t^\pi = 1$ and $\varepsilon_t^i = 0$.

$$\mathbb{E}[X_0^\pi] = \Phi \cdot 0 + A^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ a'_{2,1} \end{bmatrix}$$

$$\mathbb{E}[X_1^\pi] = \Phi \mathbb{E}[X_0] + A^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \Phi \begin{bmatrix} 1 \\ a'_{2,1} \end{bmatrix}$$

$$\mathbb{E}[X_2^\pi] = \Phi \mathbb{E}[X_1] + A^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \Phi^2 \begin{bmatrix} 1 \\ a'_{2,1} \end{bmatrix}$$

$$\implies \mathbb{E}[X_H^\pi] = \Phi^H \begin{bmatrix} 1 \\ a'_{2,1} \end{bmatrix}$$

Remarks on IRFs Derived from a SVAR

Similarly, the impulse response to a 1 percentage point shock in i_t is given as follows. Notice that now there is no contemporaneous response in π because it is ordered first.

$$\mathbb{E} [X_H^i] = \Phi^H \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Also notice that both sets of impulse responses are 2×1 vectors (in general, they will be $N \times 1$ where N is the number of variables in the VAR). The first element is the response of the first variable, π , while the second element is the response of the second variable, i . In both cases, we iterate the VAR forward by applying the reduced-form autoregressive coefficient matrix Φ to the initial response.

Remarks on IRFs Derived from a SVAR

Continued

Also notice that forecasts and impulse responses are identical except that they use different initial values.

- For real-time forecasts, we iterate the VAR forward starting at the latest observations of the series that we have. For impulse responses, we iterate forward starting from the contemporaneous impact of the structural shocks.
- For impulse responses, we need to know how the reduced-form innovations relate to the structural innovations of the system (via A^{-1}) in order to start at the correct initial values. When forecasting, we only need to know the dynamics of the system as encoded in Φ .

Last, recall that in the AR(1) model we required $\phi < 1$ for the AR(1) to be considered stable. The analogue in VARs is the requirement that all of the eigenvalues of Φ lie within the unit circle in modulus.

Other Results from Identified SVARs

Historical Decompositions: Derives the contributions of the structural shocks to the endogenous variables in driving them away from their steady-state.

Forecast Error Variance Decompositions: Derives how much of the variance of the forecast errors at some horizon h of the endogenous variables is attributable to each structural shock.

Equivalence of IRFs Derived from VARs and LPs

“Specifically, any LP impulse response function can be obtained through an appropriately ordered recursive VAR, and any (possibly nonrecursive) VAR impulse response function can be obtained through a LP with appropriate control variables.” (Plagborg-Møller and Wolf (2021))

The main result of recent work is an equivalence result between IRFs derived from LPs and VARs in population (up to a scaling factor). Moreover, any identification scheme implementable in a SVAR can be done in an LP.¹¹ Note, however, that there are differences between the two approaches in small samples/with limited lags, and that IRFs from LPs (since they are estimated locally) can be erratic whereas IRFs from VARs impose a structure on their shape. This means that VAR IRFs tend to be smoother than their LP counterparts.

A current working paper including the same researchers shows that IRFs from VARs and LPs are on opposite ends of the bias-variance tradeoff (Montiel Olea et al. (2025)): LPs have lower bias, higher variance whereas VARs have higher bias, lower variance. There exist mixed recommendations on which estimator to use in which context (number of lags, horizons, et cetera).

11. Think about how you would do this!

Intuition for Equivalence of IRFs Derived from VARs and LPs

First notice that by FWL, the LP regression reduces to the univariate regression of Y on the shock where both variables are orthogonalized with respect to lags of the other variables in the system.

Running an LP with a shock is equivalent to running an Cholesky-identified SVAR with the shock ordered first. The matrix algebra works out such that the lags of the other variables are again partialled out with respect to the shock and we are left with the same estimator as the above times a constant that does not depend on the horizon h .

IRFs from SVARs and LPs as Forecasts

“Since impulse responses are just forecasts, LP and VAR impulse response estimands coincide in population.” (Plagborg-Møller and Wolf (2021))

One way of thinking about IRFs is as a forecast in which the system is shocked at time t . In a VAR, the successive application of the matrix Φ —the estimated autoregressive dynamics of the system—yields a forecast as to what the shock will cause in the system in subsequent periods. The LP does this same thing without estimating the system’s autoregressive dynamics directly.

Common Misconceptions about VARs vs. LPs

According to Plagborg-Møller and Wolf (2021), the following are common misconceptions:

1. “VAR impulse response estimators are generally more efficient than local projection estimators”;
2. “local projections are generally more robust to misspecification than VARs”;
3. “SVAR analysis is required when implementing nonrecursive, non-IV identification schemes”; and
4. “simple SVAR methods cannot be used when the structural shock of interest is noninvertible”.

Some Useful Resources

- See the Ramey (2016) chapter in the Handbook of Macroeconomics for the definitive treatment of macroeconomic shock identification.
- See Nakamura and Steinsson (2018) for an overview of identification in macro.
- Much of this presentation's coverage of VARs is indebted to Ambrosio Cesa-Bianchi's slides, [A Primer on Vector Autoregression](#). He has also published a codebase called the VAR Toolbox that you might find helpful.
- For a thorough treatment of time series methods prior to the 2000s, see Hamilton (1994).

Contents

1 Time Series Data & Models

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

References I

- Bauer, Michael D., and Eric T. Swanson. 2023. "An Alternative Explanation for the "Fed Information Effect"." *American Economic Review* 113 (3): 664–700. <https://doi.org/10.1257/aer.20201220>.
<https://www.aeaweb.org/articles?id=10.1257/aer.20201220>.
- Blanchard, Olivier Jean, and Danny Quah. 1989. "The Dynamic Effects of Aggregate Demand and Supply Disturbances." *American Economic Review* 79 (4): 655–673. <https://doi.org/None>.
<https://ideas.repec.org/a/aea/aecrev/v79y1989i4p655-73.html>.
- Cesa-Bianchi, Ambrogio. 2024. *A Primer on Vector Autoregressions*. Technical report.
- Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans. 1998. *Monetary Policy Shocks: What Have We Learned and to What End?* Working Paper, Working Paper Series 6400. National Bureau of Economic Research, February. <https://doi.org/10.3386/w6400>. <http://www.nber.org/papers/w6400>.
- . 1999. "Chapter 2 Monetary policy shocks: What have we learned and to what end?," 1:65–148. *Handbook of Macroeconomics*. Elsevier. [https://doi.org/https://doi.org/10.1016/S1574-0048\(99\)01005-8](https://doi.org/https://doi.org/10.1016/S1574-0048(99)01005-8).
<https://www.sciencedirect.com/science/article/pii/S1574004899010058>.
- . 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113 (1): 1–45. ISSN: 00223808, 1537534X. <http://www.jstor.org/stable/10.1086/426038>.

References II

- Dube, Arindrajit, Daniele Girardi, Òscar Jordà, and Alan M. Taylor. 2025. "A Local Projections Approach to Difference-in-Differences." *Journal of Applied Econometrics*, <https://doi.org/https://doi.org/10.1002/jae.70000>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.70000>.
- Fama, Eugene F. 1965. "Random Walks in Stock Market Prices." *Financial Analysts Journal* 21 (5): 55–59. ISSN: 0015198X, accessed August 6, 2025. <http://www.jstor.org/stable/4469865>.
- Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson. 2007. "ABCs (and Ds) of Understanding VARs." *American Economic Review* 97 (3): 1021–1026. <https://doi.org/10.1257/aer.97.3.1021>. <https://www.aeaweb.org/articles?id=10.1257/aer.97.3.1021>.
- Ferreira, Leonardo N., Silvia Miranda-Agrippino, and Giovanni Ricco. 2025. "Bayesian Local Projections." *The Review of Economics and Statistics* (July): 1–15. ISSN: 0034-6535. https://doi.org/10.1162/rest_a_01334. eprint: https://direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01334/2112661/rest_a_01334.pdf. https://doi.org/10.1162/rest%5C_a%5C_01334.
- Friedman, Milton, and Anna Jacobson Schwartz. 1963. *A Monetary History of the United States, 1867-1960* [in en]. National Bureau of Economic Research Publications. Princeton, NJ: Princeton University Press.
- Gertler, Mark, and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7, no. 1 (January): 44–76. <https://doi.org/10.1257/mac.20130329>. <https://www.aeaweb.org/articles?id=10.1257/mac.20130329>.

References III

- Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102, no. 4 (June): 1692–1720. <https://doi.org/10.1257/aer.102.4.1692>.
<https://www.aeaweb.org/articles?id=10.1257/aer.102.4.1692>.
- Hamilton, James Douglas. 1994. *Time Series Analysis*. Princeton University Press.
- Hodrick, Robert J., and Edward Prescott. 1981. *Post-War U.S. Business Cycles: An Empirical Investigation*. Discussion Papers 451. Northwestern University, Center for Mathematical Studies in Economics and Management Science, May. <https://doi.org/None>. <https://ideas.repec.org/p/nwu/cmsems/451.html>.
- Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, <https://www.aeaweb.org/articles?id=10.1257/0002828053828518>.
- Jordà, Òscar, and Alan M. Taylor. 2025. "Local Projections." *Journal of Economic Literature* 63 (1): 59–110. <https://doi.org/10.1257/jel.20241521>. <https://www.aeaweb.org/articles?id=10.1257/jel.20241521>.
- Känzig, Diego R. 2021. "The Macroeconomic Effects of Oil Supply News: Evidence from OPEC Announcements." *American Economic Review* 111 (4): 1092–1125.
- Mertens, Karel, and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103, no. 4 (June): 1212–47. <https://doi.org/10.1257/aer.103.4.1212>. <https://www.aeaweb.org/articles?id=10.1257/aer.103.4.1212>.

References IV

- Miranda-Agrippino, Silvia, and Giovanni Ricco. 2021. "The Transmission of Monetary Policy Shocks." *American Economic Journal: Macroeconomics* 13, no. 3 (July): 74–107. <https://doi.org/10.1257/mac.20180124>.
<https://www.aeaweb.org/articles?id=10.1257/mac.20180124>.
- Montiel Olea, José Luis, Mikkel Plagborg-Møller, Eric Qian, and Christian K. Wolf. 2025. "Local Projections or VARs? A Primer for Macroeconomists," arXiv: 2503.17144 [econ.EM]. <https://arxiv.org/abs/2503.17144>.
- Nakamura, Emi, and Jón Steinsson. 2018. "Identification in Macroeconomics." *Journal of Economic Perspectives* 32 (3): 59–86. <https://doi.org/10.1257/jep.32.3.59>. <https://www.aeaweb.org/articles?id=10.1257/jep.32.3.59>.
- Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–708. ISSN: 00129682, 14680262, accessed September 11, 2025. <http://www.jstor.org/stable/1913610>.
- Nunes, Ricardo, Ali Ozdagli, and Jenny Tang. 2022. "Interest Rate Surprises: A Tale of Two Shocks." *Federal Reserve Bank of Boston Research Department Working Papers* 22 (2).
- Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. "Local Projections and VARs Estimate the Same Impulse Responses." *Econometrica* 89 (2): 955–980. <https://doi.org/https://doi.org/10.3982/ECTA17813>.
<https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17813>.

References V

- Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propagation," edited by John B. Taylor and Harald Uhlig, 2:71–162. *Handbook of Macroeconomics*. Elsevier.
<https://doi.org/https://doi.org/10.1016/bs.hesmac.2016.03.003>.
<https://www.sciencedirect.com/science/article/pii/S1574004816000045>.
- Romer, Christina D., and David H. Romer. 1989. "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz." *NBER Macroeconomics Annual* 4:121–170. <https://doi.org/10.1086/654103>.
<https://doi.org/10.1086/654103>.
- . 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 94, no. 4 (September): 1055–1084. <https://doi.org/10.1257/0002828042002651>.
<https://www.aeaweb.org/articles?id=10.1257/0002828042002651>.
- Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48 (1): 1–48. ISSN: 00129682, 14680262, accessed April 28, 2025. <http://www.jstor.org/stable/1912017>.
- Stock, James H. 2008. "What's new in econometrics: time series." *NBER Summer Institute*, http://www.nber.org/minicourse_2008.html.
- Stock, James H., and Mark W. Watson. 2012. "Disentangling the Channels of the 2007-2009 Recession." *Brookings Papers on Economic Activity*, <https://www.brookings.edu/articles/disentangling-the-channels-of-the-2007-2009-recession/>.

References VI

- Stock, James H., and Mark W. Watson. 2018. "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments." *Economic Journal* 128, no. 610 (May): 917–948. <https://doi.org/10.1111/eoj.12593>. <https://ideas.repec.org/a/wly/econjl/v128y2018i610p917-948.html>.
- Uhlig, Harald. 2005. "What are the effects of monetary policy on output? Results from an agnostic identification procedure." *Journal of Monetary Economics* 52 (2): 381–419. ISSN: 0304-3932. <https://doi.org/https://doi.org/10.1016/j.jmoneco.2004.05.007>. <https://www.sciencedirect.com/science/article/pii/S0304393205000073>.

Contents

1 Time Series Data & Models

2 Motivation for VARs & LPs

3 Local Projections (LPs)

4 Vector Autoregressions (VARs)

References

Appendix

Appendix

I Supplemental Material by Section

I.I Time Series Data & Models

I.II Motivation for VARs & LPs

I.III Local Projections (LPs)

I.IV Vector Autoregressions (VARs)

II Derivations & Proofs

III Data Sources

Appendix

I Supplemental Material by Section

I.I Time Series Data & Models

I.II Motivation for VARs & LPs

I.III Local Projections (LPs)

I.IV Vector Autoregressions (VARs)

II Derivations & Proofs

III Data Sources

Appendix

I Supplemental Material by Section

I.I Time Series Data & Models

I.II Motivation for VARs & LPs

I.III Local Projections (LPs)

I.IV Vector Autoregressions (VARs)

II Derivations & Proofs

III Data Sources

Terminology & Notation

Types of Data

We call $\{x_i\}_{i=1}^N$ cross-sectional data if each observation is indexed by a cross-sectional variable.

We call $\{x_t\}_{t=0}^T$ a time series if each observation is indexed by time.

Data series that are indexed by time and a cross-sectional variable are known as panel data and are denoted using a subscript for time and for the cross-sectional variable: $\{x_{i,t}\}_{i=1,t=0}^{N,T}$.

Due to characteristics particular to time series data, we cannot naïvely use the exact same econometrics when estimating effects in the time series.

Terminology & Notation

Leads & Lags

Consider the following time series (where some of the notation from the previous slide is omitted to reflect a general case):

$$\{x_t\} := \{\dots, x_{t-5}, \dots, x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2}, \dots, x_{t+5}, \dots\}$$

The lags of observation x_t are comprised by the set $\{x_{t-h}\}_{h>0} \subset \{x_t\}$. Similarly, the leads of observation x_t comprised by the set $\{x_{t+h}\}_{h>0} \subset \{x_t\}$.

As such, we call x_{t-1} and x_{t-5} the first and fifth lags of x_t , respectively, while x_{t+1} and x_{t+5} are the first and fifth leads of x_t , respectively.

The h^{th} difference of the time series is defined as recursively mapping $\{x_t\} \mapsto \{x_t - x_{t-1}\}$ for h times. For example, the first difference is $\{x_t - x_{t-1}\}$; the second difference is $\{x_t - 2x_{t-1} + x_{t-2}\}$, et cetera.

Cumulating Log-Differences

Note that for any two periods $t = 0$ and $t = \tau > 0$

$$\begin{aligned}\sum_{t=1}^{\tau} (\ln(x_t) - \ln(x_{t-1})) &= \ln(x_{\tau}) + \cdots - \cancel{\ln(x_t)} + (\cancel{\ln(x_t)} - \cancel{\ln(x_{t-1})}) + \cancel{\ln(x_{t-1})} + \\ &\quad \cdots - \ln(x_0) \\ &= \ln(x_{\tau}) - \ln(x_0).\end{aligned}$$

We start from (3):

$$\begin{aligned}x_0 \prod_{t=1}^{\tau} (1 + g_t) &= x_{\tau} \\ \ln \left(\prod_{t=0}^{\tau} (1 + g_t) \right) &= \ln(x_{\tau}) - \ln(x_0) \\ \Rightarrow \sum_{t=0}^{\tau} g_t \approx \prod_{t=0}^{\tau} (1 + g_t) &= \exp(\ln(x_{\tau}) - \ln(x_0)) = \exp \left(\sum_{t=1}^{\tau} [\ln(x_t) - \ln(x_{t-1})] \right)\end{aligned}$$

Back

Appendix

I Supplemental Material by Section

I.I Time Series Data & Models

I.II Motivation for VARs & LPs

I.III Local Projections (LPs)

I.IV Vector Autoregressions (VARs)

II Derivations & Proofs

III Data Sources

The 3-Equation New Keynesian Model

Consider the following variation on the canonical log-linearized 3-equation New Keynesian model:

(Euler/DIS)
$$\tilde{y}_t = \mathbb{E}_t [\tilde{y}_{t+1}] - \frac{1}{\sigma} (i_t - \mathbb{E}_t [\pi_{t+1}] - r_t^*)$$

(NK Phillips Curve)
$$\pi_t = \kappa \tilde{y}_t + \beta \mathbb{E}_t [\pi_{t+1}]$$

(Taylor Rule)
$$i_t = \rho_t + \phi^\pi \pi_t + \phi^y \tilde{y}_t$$

where

\tilde{y}_t : output gap

i_t : nominal interest rate

π_t : the inflation rate

r_t^* : the natural rate of interest

ρ_t : monetary policy shocks

β : discount factor

σ : inverse intertemporal elasticity of substitution

ϕ^π, ϕ^y : parameters governing forcefulness of monetary policy response to π and \tilde{y}

κ contemporaneously relates the output gap and inflation

Note that this model has many flaws (for instance, supply-side constraints and financial constraints are not modeled; an output gap tomorrow is the same as an output gap today, there are many equilibria if the Taylor rule is misspecified), but it's a useful starting point.

Back

The New Keynesian Model

Continued

(Euler/DIS)
$$\tilde{y}_t = \mathbb{E}_t [\tilde{y}_{t+1}] - \frac{1}{\sigma} (i_t - \mathbb{E}_t [\pi_{t+1}] - r_t^*)$$

(NK Phillips Curve)
$$\pi_t = \kappa \tilde{y}_t + \beta \mathbb{E}_t [\pi_{t+1}]$$

(Taylor Rule)
$$i_t = \rho_t + \phi^\pi \pi_t + \phi^{\tilde{y}} \tilde{y}_t$$

Our main variables of interest are \tilde{y}_t , i_t , and π_t . How do these comove?

- According to this model, \tilde{y}_t and π_t both positively feed into each other.
- The central bank follows the Taylor Rule, setting nominal interest rates in order to either stimulate ($i_t \downarrow \implies \tilde{y}_t, \pi_t \uparrow$) or restrain ($i_t \uparrow \implies \tilde{y}_t, \pi_t \downarrow$) the economy.
- This should align with your intuition: During recessions, the Fed stimulates demand in the economy by cutting interest rates. Too-low rates for too long can cause inflation, which the Fed combats by raising rates. Higher rates then slow down real activity in the economy and induce a recession.

The New Keynesian Model

Continued

(Euler/DIS)
$$\tilde{y}_t = \mathbb{E}_t [\tilde{y}_{t+1}] - \frac{1}{\sigma} (i_t - \mathbb{E}_t [\pi_{t+1}] - r_t^*)$$

(NK Phillips Curve)
$$\pi_t = \kappa \tilde{y}_t + \beta \mathbb{E}_t [\pi_{t+1}]$$

(Taylor Rule)
$$i_t = \rho_t + \phi^\pi \pi_t + \phi^{\tilde{y}} \tilde{y}_t$$

The model parametrizes the intuition that all of these macroeconomic variables are correlated.

- So if we were to see both output and inflation rising in the data, how would we know which came first...
- ...or did both rise because interest rates were low?
- This simultaneity is the macro rehash of the chicken and egg problem.

From Macroeconomic Theory to Empirics

Redux

The 3-equation New Keynesian model is a simple example of a dynamic stochastic general equilibrium model (DSGE). DSGE models are typically *calibrated* by matching model predictions with empirical estimates (e.g., stable long-run trends, micro-data evidence, IRF matching) or *estimated* (e.g., via maximum likelihood of the model). Models are necessary to understand causal mechanisms and run counterfactual analyses, but they must first be optimized and evaluated against actual data.

Appendix

I Supplemental Material by Section

I.I Time Series Data & Models

I.II Motivation for VARs & LPs

I.III Local Projections (LPs)

I.IV Vector Autoregressions (VARs)

II Derivations & Proofs

III Data Sources

LP Estimation Assumptions

It is most common to directly estimate LPs as OLS or IV regressions. As such, the standard OLS assumptions must be met:

- Linearity: The dynamics of the system can locally be represented or approximated by a linear regression
- No Omitted Variables: $\rho_{x_t, \varepsilon_{t+h}^{(h)}} = 0 \quad \forall t$
- No Multicollinearity: $\rho_{x_i, x_j} = 0 \quad \forall i \neq j$
- Normality of Residuals (Optional): $\varepsilon_{t+h}^{(h)} \sim \mathcal{N}(0, \sigma_{\varepsilon^{(h)}}^2) \quad \forall t$
 - Homoskedasticity: $\sigma_{\varepsilon^{(h)}}^2 = \sigma_{\varepsilon_{t+h}^{(h)}}^2 = \sigma_{\varepsilon_{s+h}^{(h)}}^2 \quad \forall t \neq s$
 - No Autocorrelation: $|\rho_{\varepsilon_{t+h}^{(h)}, \varepsilon_{s+h}^{(h)}}| \neq 1 \quad \forall t \neq s$
 - One typically always uses robust standard errors when computing LPs which corrects for heteroskedasticity.
 - The Newey-West (heteroskedasticity- and autocorrelation-consistent) estimator is sometimes used to calculate standard errors when autocorrelation between the residuals is a concern. Jordà (2005) used the same number of lags in the Newey-West kernel as the forecast horizon, but you can just use the same number of lags as included of the control variables in the RHS of the LP. Usually, if you estimate the LP in log-levels and control for lags of the LHS variable, you do not need to worry about this.

Appendix

I Supplemental Material by Section

I.I Time Series Data & Models

I.II Motivation for VARs & LPs

I.III Local Projections (LPs)

I.IV Vector Autoregressions (VARs)

I.IV.I Theory

I.IV.II Estimation, Identification, & Causal Inference

II Derivations & Proofs

III Data Sources

Appendix

I Supplemental Material by Section

I.I Time Series Data & Models

I.II Motivation for VARs & LPs

I.III Local Projections (LPs)

I.IV Vector Autoregressions (VARs)

I.IV.I Theory

I.IV.II Estimation, Identification, & Causal Inference

II Derivations & Proofs

III Data Sources

Vector Autoregressions

A VAR is a system of jointly-estimated equations which models the evolution and dynamics of the variables it contains as a linear combination of their residuals. VARs are an extension of autoregressive processes. Let

- N be the number of variables included in the VAR;
- T the total number of time periods for which data is available; and
- p the number of lags of the variables included in the VAR.

A VAR(1) process takes the following form:

$$AX_t = BX_{t-1} + \varepsilon_t$$

As with AR(p) processes, one can have VAR(p) processes:

$$AX_t = B_1X_{t-1} + B_2X_{t-2} + \cdots + B_{p-1}X_{t-(p-1)} + B_pX_{t-p} + \varepsilon_t$$

For simplicity's sake, we will focus on VAR(1) for a reason that will become apparent in a few slides.

Structural Vector Autoregressions

Assume that the data is balanced. The structural form of a VAR is given as

$$\text{(Structural VAR}(p)) \quad AX_t = \sum_{\tau=1}^p B_{\tau} X_{t-\tau} + \varepsilon_t$$

$$\text{(Structural VAR}(1)) \quad AX_t = BX_{t-1} + \varepsilon_t$$

where

- $X_{t-\tau}$ ($N \times (T - p)$ for $\tau \in \{0, \dots, p - 1\}$ ¹²) is a matrix of endogenous variables;
- A ($N \times N$), the inverse impact matrix, encodes the contemporaneous relationship between the variables in X (thus, if it exists, A^{-1} is the impact matrix);
- B ($N \times N$), known as the structural coefficient matrix, describes the autoregressive dynamics of the system; and
- ε_t ($N \times (T - p)$) are the structural residuals/innovations (shocks) of the VAR; and
- $\Sigma := \mathbb{E}[\varepsilon_t \varepsilon_t']$ ($N \times N$) is the structural covariance matrix.

12. Note that in the this appendix section, I write VARs in the way one would mechanically set up the matrices. To convert from this version to the standard representation, simply substitute the time series vectors in the matrices (bold) for their scalar counterparts (normal typeface). For example, the dimensions of $X_{t-\tau}$ would thus be $N \times 1$.

VARs as Linear Systems of Equations

and the Etymology of “*Vector* Autoregression”

The matrix representation of X as stacked vectors is equivalent to writing down the VAR as a system of linear equations. In the example where $N = 2$, $p = 1$, and $x_t^1, x_t^2, \varepsilon_t^1, \varepsilon_t^2 \in \mathbb{R}^{T-1}$:

$$\underbrace{\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}}_A \underbrace{\begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix}}_B \underbrace{\begin{bmatrix} x_{t-1}^1 \\ x_{t-1}^2 \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix}}_{\varepsilon_t}$$
$$\iff \begin{cases} a_{1,1}x_t^1 + a_{1,2}x_t^2 = b_{1,1}x_{t-1}^1 + b_{1,2}x_{t-1}^2 + \varepsilon_t^1 \\ a_{2,1}x_t^1 + a_{2,2}x_t^2 = b_{2,1}x_{t-1}^1 + b_{2,2}x_{t-1}^2 + \varepsilon_t^2 \end{cases}$$

where $x_t = [x_T \ x_{T-1} \ \cdots \ x_3 \ x_2]$ and $x_{t-1} = [x_{T-1} \ x_{T-2} \ \cdots \ x_2 \ x_1] \ \forall \cdot \in \{1, 2\}$ since $p = 2$. Writing out VARs this way also makes it apparent where the “vector” part comes in. Here, x_t does not indicate a generic observation at time t , but the $1 \times (T - p)$ vector of observations of variable \cdot beginning at time $p + 1$ and ending at time T .

More generally, $x_{t-\tau} := [x_{T-\tau} \ \cdots \ x_{p-\tau}] \ \forall \tau \in \{0, \dots, p-1\}, \ \forall \cdot \in \{1, 2\}$.

Companion Representation of a VAR

From VAR(p) to VAR(1)

Any (S)VAR(p) can cleverly written in (S)VAR(1) form. For example,

$$AX_t = B_1X_{t-1} + B_2X_{t-2} + \cdots + B_pX_{t-p} + \varepsilon_t$$

becomes

$$\underbrace{\begin{bmatrix} A & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & I \end{bmatrix}}_{A'} \underbrace{\begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{bmatrix}}_{X'_t} = \underbrace{\begin{bmatrix} B_1 & B_2 & \cdots & B_p \\ I & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & I & 0 \end{bmatrix}}_{X'_{t-1}} \underbrace{\begin{bmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \end{bmatrix}}_{X'_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\varepsilon'_t}$$

where I is an identity matrix. As this notation is simpler, cleaner, and more convenient, it is typical in the literature to report the companion (VAR(1)) form of a VAR(p). The remainder of these notes will focus on VAR(1) processes with the understanding that the two representations are equivalent.

Reduced-Form Vector Autoregressions

If the matrix A is invertible, we can obtain the reduced form of the VAR.

$$X_t = A^{-1}BX_{t-1} + A^{-1}\varepsilon_t$$

(Reduced-Form VAR(p))

$$X_t = \sum_{\tau=1}^p \Phi_{\tau}X_{t-\tau} + \nu_t$$

(Reduced-Form VAR(1))

$$X_t = \Phi X_{t-1} + \nu_t$$

where

- Φ ($N \times N$) is the reduced-form coefficient matrix and
- $A^{-1}\varepsilon_t \equiv \nu_t$ ($N \times (T - p)$) are the reduced-form residuals/innovations of the VAR; and
- $\Omega := \mathbb{E}[\nu_t\nu_t']$ ($N \times N$) is the reduced-form covariance matrix.

Note that unlike in the SVAR, the reduced-form coefficient (Φ) and covariance (Ω) matrices can be directly estimated. To retrieve the latter, we need the reduced-form residuals which are given by $\nu_t = X_t - \Phi X_{t-1}$.

The acronym “VAR” usually refers to the reduced-form VAR while “SVAR” is used when referring to the structural VAR.

The Non-invertibility Problem

As stated previously, one can generally write down the (log-)linearized equilibrium to a macroeconomic model as a system of difference equations. However, we might not be able to estimate the structural model as a VAR because there is no linear rotation of the reduced-form innovations that allows us to recover the structural shocks. Non-invertibility is equivalent to the observables of the model not containing enough information to reveal the state variables (and their associated structural shocks). For more on non-invertibility, see Fernández-Villaverde et al. (2007).

The SVARs considered here will not encounter this issue as we will not be mapping the SVAR to a DSGE model.

Augmented VARs

(S)VARs can be augmented by including a constant term, deterministic terms, and exogenous variables. Consider the following reduced-form VAR.

$$X_t = \alpha + \Phi X_{t-1} + \Gamma D_t + \Psi Z_t + \nu_t$$

where

- α is a deterministic constant trend added to all variables in each period;
- D_t is a matrix of other deterministic terms with associated coefficient matrix Γ ; and
- Z_t is a matrix of variables exogenous to the system X_t with associated coefficient matrix Ψ .

Note that the constant term can also be modeled by including a column of 1's in front of X_t

The Wold Representation of a VAR

We retrieve the Wold representation of a VAR by recursively substituting for past values in the reduced-form VAR equation.

$$\begin{aligned}X_t &= \Phi X_{t-1} + \nu_t \\&= \Phi(\Phi X_{t-2} + \nu_{t-1}) + \nu_t = \Phi^2 X_{t-2} + \Phi \nu_{t-1} + \nu_t \\&\vdots\end{aligned}$$

$$\begin{aligned}(8) \quad (\text{Wold Representation}) \quad X_t &= \Phi^t X_0 + \sum_{\tau=0}^{t-1} \Phi^\tau \nu_{t-\tau} \\&= \underbrace{\Phi^t X_0}_{\text{Initial Condition}} + \underbrace{\sum_{\tau=1}^{t-1} \Phi^\tau \nu_{t-\tau}}_{\text{Past Residuals}} + \underbrace{\nu_t}_{\text{Last Residual}}\end{aligned}$$

The Wold representation of the VAR expresses the system of difference equations as the sum of an initial condition and past residuals.

VAR Stability

We require that the variables in a VAR be covariance stationary. However, if this is not the case, then iteratively applying Φ to X_t causes the system to explode. Taking $t \rightarrow \infty$ in the Wold representation, we get:

$$X_t = \Phi^\infty X_{t-\infty} + \sum_{\tau=0}^{\infty} \Phi^\tau \nu_{t-\tau}$$

VAR Stability: A VAR is called stable if and only if all of the eigenvalues of Φ are less than 1 in modulus, which causes the unconditional mean of X_t to equal 0 or the constant $(I - \Phi)^{-1}\alpha$.

$$\|\Phi - \lambda I\| = 0; \lambda < 1 \implies \Phi^\infty = 0; \sum_{\tau=0}^{\infty} \Phi^\tau = (I - \Phi)^{-1} \quad (\text{geometric series})$$

$$\mathbb{E}[X_t] = \Phi^\infty X_{t-\infty} + \sum_{\tau=0}^{\infty} \Phi^\tau \mathbb{E}[\nu_{t-\tau}] = 0 \quad (\text{since } \mathbb{E}[\nu_{t-\tau}] = 0)$$

If X_t contains a constant intercept term (α), then the unconditional mean of X_t equals:

$$\mathbb{E}[X_t] = \Phi^\infty X_{t-\infty} + \sum_{\tau=0}^{\infty} \Phi^\tau \mathbb{E}[\alpha] + \sum_{\tau=0}^{\infty} \Phi^\tau \mathbb{E}[\nu_{t-\tau}] = (I - \Phi)^{-1}\alpha \quad (\text{since } \mathbb{E}[\nu_{t-\tau}] = 0).$$

Appendix

I Supplemental Material by Section

I.I Time Series Data & Models

I.II Motivation for VARs & LPs

I.III Local Projections (LPs)

I.IV Vector Autoregressions (VARs)

I.IV.I Theory

I.IV.II Estimation, Identification, & Causal Inference

II Derivations & Proofs

III Data Sources

Estimating VARs

Recall that the matrices A and B and shocks ε_t in the SVAR are all unknown and need to be estimated. We proceed by estimating the reduced-form VAR and employing various identification strategies to back out the structural coefficient and covariance matrices.

$$AX_t = BX_{t-1} + \varepsilon_t$$

We are able to estimate the VAR using OLS or Bayesian methods if the following conditions are met:

- $\mathbb{E}[\varepsilon_t] = 0$
- $\mathbb{E}[\varepsilon_t \varepsilon_t'] = \Sigma$ where Σ is positive semi-definite and diagonal
 - The structural shocks are not correlated with each other: $\mathbb{E}[\varepsilon_t \varepsilon_\tau'] = 0 \forall \tau \neq t$
 - More specifically, we require that $\Sigma = I$, meaning the structural shocks do not exhibit autocorrelation. If the structural shocks are autocorrelated, this calls into question the identification of the SVAR.
- No serial correlation: $\mathbb{E}[\varepsilon_t \varepsilon_{t-\tau}'] = 0$ for $\tau > 0$

If the above assumptions are met, we are able to receive unbiased estimates for $\Phi := A^{-1}B$ and $\Omega := \mathbb{E}[\nu_t \nu_t']$ using OLS.

The Identification Problem in VARs

The estimated reduced-form VAR is useful for describing the dynamics of the data and for forecasting among other things, but it cannot be used for causal inference: It contains information about both structural contemporaneous relationships between the covariates and their relationships to lagged values.

In other words, we retrieve estimates for Φ and Ω but we are interested in B and Σ . Conducting causal inference using VARs requires us to move from the reduced-form VAR (which can be directly estimated) to the SVAR (which remains unobserved). For some identification strategies, this amounts to placing economically-meaningful (model-motivated) restrictions on the matrix A .

Causal Identification in VARs

We will cover the following identification strategies, all of which are still referenced or extended in various directions in contemporary research.

1. Cholesky identification (“internal instruments”)
 - Short-run restrictions
 - Long-run restrictions
2. Sign restrictions
3. Proxy SVAR (“external instruments”)

While the basic implementations of these strategies have fallen out of favor, they are nevertheless important for understanding (1) how economic methodology and knowledge has improved; (2) the extensions of these identification strategies; and (3) what qualifies as a credible identification technique in macroeconomics.

Causal Inference in VARs

Taking stock, we have:

$$\begin{aligned} B = A\Phi &\iff \Phi = A^{-1}B \\ \Sigma = A\Omega A' &\iff \Omega = \mathbb{E}[\nu_t \nu_t'] \\ &= \mathbb{E}\left[(A^{-1})\varepsilon_t \varepsilon_t' (A^{-1})'\right] \\ &= (A^{-1})\mathbb{E}[\varepsilon_t \varepsilon_t'] (A^{-1})' \\ &= (A^{-1})\Sigma(A^{-1})' \\ &= A^{-1}(A^{-1})' \end{aligned}$$

Note that we get N^2 estimates for Φ and $\frac{N(N+1)}{2}$ for Ω due to its symmetry as the covariance matrix. In all, we obtain $\frac{3N^2+N}{2}$ coefficient estimates. However, B and A are both N^2 so the SVAR contains $2N^2$ coefficients. We are thus lacking $\frac{N(N-1)}{2}$ estimates.

For the next section, we will need to review some linear algebra.

Linear Algebra Review

Triangle Factorization (LDU Decomposition): Any symmetric positive definite square matrix Ω has a unique representation of the form

$$\Omega = T\Sigma T'$$

where T is a lower triangular matrix with 1's along the main diagonal (meaning T' is an upper triangular matrix with 1's along the main diagonal) and Σ is a positive definite diagonal matrix.

Cholesky Decomposition: Any symmetric positive definite square matrix Ω has a unique representation of the form

$$\Omega = T\Sigma^{1/2}(T\Sigma^{1/2})' = PP'$$

where $P := T\Sigma^{1/2}$ is positive, lower-triangular and $\Sigma^{1/2}$ is defined as the positive definite square matrix whose diagonal is the square root of the diagonal of Σ from the triangle diagonalization. In other words, $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$.

Linear Algebra Review

Continued

LQ Decomposition: Any real square matrix Ω can be decomposed as

$$\Omega = LQ$$

where Q is an orthonormal matrix ($Q' = Q^{-1}$) and L is a lower-triangular matrix. If Ω is invertible and the diagonal of L is positive, then the decomposition is unique.

Short-Run Restrictions via Cholesky Identification

In the short run, suppose we have an economic theory or intuition that suggests that some variables do not contemporaneously affect other variables. For example, it might take a while for a change in interest rates to affect real variables.

Consider the simple 3-variable VAR that contains interest rates, inflation, and output growth. We assume that interest rates contemporaneously respond to inflation and output growth but that it takes a while for their impact to be seen. How can we implement this assumption by manipulating the values in the matrix A^{-1} ?

$$\underbrace{\begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} a'_{1,1} & a'_{1,2} & a'_{1,3} \\ a'_{2,1} & a'_{2,2} & a'_{2,3} \\ a'_{3,1} & a'_{3,2} & a'_{3,3} \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ b_{2,1} & b_{2,2} & b_{2,3} \\ b_{3,1} & b_{3,2} & b_{3,3} \end{bmatrix}}_B \underbrace{\begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ i_{t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} a'_{1,1} & a'_{1,2} & a'_{1,3} \\ a'_{2,1} & a'_{2,2} & a'_{2,3} \\ a'_{3,1} & a'_{3,2} & a'_{3,3} \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^\pi \\ \varepsilon_t^i \end{bmatrix}}_{\varepsilon_t}$$

Short-Run Restrictions via Cholesky Identification

Continued

We implement the short-run restrictions by making A^{-1} a lower-triangular matrix. In other words, we set $a'_{1,2} = a'_{1,3} = a'_{2,3} = 0$. Also note that the diagonal entries of A^{-1} must be 1's since a structural shock of some magnitude in any given variable can only raise the value by that amount: thus, $a'_{1,1} = a'_{2,2} = a'_{3,3} = 1$. The other coefficients in A^{-1} are still unknown.

$$\underbrace{\begin{bmatrix} y_t \\ \pi_t \\ i_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ a'_{2,1} & 1 & 0 \\ a'_{3,1} & a'_{3,2} & 1 \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ b_{2,1} & b_{2,2} & b_{2,3} \\ b_{3,1} & b_{3,2} & b_{3,3} \end{bmatrix}}_B \underbrace{\begin{bmatrix} y_{t-1} \\ \pi_{t-1} \\ i_{t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ a'_{2,1} & 1 & 0 \\ a'_{3,1} & a'_{3,2} & 1 \end{bmatrix}}_{A^{-1}} \underbrace{\begin{bmatrix} \varepsilon_t^y \\ \varepsilon_t^\pi \\ \varepsilon_t^i \end{bmatrix}}_{\varepsilon_t}$$

We have $\Omega = A^{-1} (A^{-1})'$ where A^{-1} is lower-triangular. We can retrieve the rest of the variables $\Omega = PP'$. Since P is unique, the lower-triangular assumption on A^{-1} implies $A^{-1} = P$.

Cholesky Identification

In general, we implement the short-run restrictions in a VAR using a two-step process.

1. First, we order the variables (when stacking the X_t matrix) such that the variables that respond with lags to other variables are first while variables that respond immediately to other variables are last. In general, when using Cholesky identification, for all variables in the VAR system, variables ordered later are assumed to not affect earlier variables contemporaneously.
2. Next, we use the Cholesky decomposition of Ω to identify A^{-1} . Being a covariance matrix, Ω is symmetric and positive semi-definite so it admits a unique Cholesky decomposition such that $\Omega = PP'$.

$$X_t = A^{-1}BX_{t-1} + A^{-1}\varepsilon_t = \Phi X_{t-1} + \nu_t$$

From the reduced-form VAR, $\Phi = A^{-1}B$ and $\Omega = A^{-1}(A^{-1})'$. Applying the Cholesky decomposition to Ω , we have $\Omega = PP'$ where P is positive semi-definite, and lower-triangular. Setting $A^{-1} = P$, we have identified the SVAR and can back out B and ε_t as follows:

$$B = A\Phi = P^{-1}\Phi$$

$$\varepsilon_t = A\nu_t$$

Economic Interpretation of Cholesky Decomposition

The key to understanding the identification is connecting the lower-triangular matrix that we get from the Cholesky decomposition to the impact matrix A^{-1} .

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ a'_{2,1} & 1 & 0 \\ a'_{3,1} & a'_{3,2} & 1 \end{bmatrix}$$

Since the ordering and Cholesky decomposition forces A^{-1} to be lower-triangular, you can see that the first structural shock ε_t^y contemporaneously affects y_t , π_t , and i_t ; ε_t^π affects both π_t and i_t contemporaneously but not y_t ; and the last structural shock ε_t^i contemporaneously affects only i_t .

Note that if we did not order interest rates last (below output growth and inflation), then the contemporaneous-exclusion restriction implicit in the Cholesky decomposition would not hold and the Cholesky decomposition of Ω would have a different interpretation. Also note that the Cholesky decomposition also implicitly assumes that inflation contemporaneously responds to output growth but output growth does not contemporaneously respond to inflation.

Long-Run Restrictions

In short, we assume that the cumulative impact of some of the variables on others is 0 in the long run. For example, if in the long run we assume that supply-side shocks—not demand-side shocks—cumulatively determine output, we can use this as a long-run restriction.

From the Wold representation, the cumulative impact of the vector of structural shocks ε_t on X_t if the VAR is stable is given by:

$$X_t = A^{-1}\varepsilon_t + BA^{-1}\varepsilon_t + B^2A^{-1}\varepsilon_t + \cdots + B^\infty A^{-1}\varepsilon_t = \sum_{\tau=0}^{\infty} B^\tau A^{-1}\varepsilon_t = (I - B)^{-1}A^{-1}\varepsilon_t$$

To implement the long-run restriction, define

$$\begin{aligned}\Lambda &:= (I - B)^{-1}A^{-1} \\ \Omega^* &:= \Lambda\Lambda' \\ &= (I - B)^{-1}A^{-1}((I - B)^{-1}A^{-1})' \\ &= (I - B)^{-1}\left(A^{-1}(A^{-1})'\right)((I - B)^{-1})' \\ &= (I - B)^{-1}\Omega((I - B)^{-1})'. \end{aligned}$$

Long-Run Restrictions

Continued

Since Ω^* is positive semi-definite, it admits a (unique) Cholesky decomposition such that $\Omega^* = P^* (P^*)'$. To achieve identification, we set $A^{-1} = (I - B) P^*$. Note that A^{-1} is not guaranteed to be triangular in this case. As before, once we have A^{-1} , we have $\varepsilon_t = A\nu_t$ and $\Sigma = A^{-1} (A^{-1})'$, and so the SVAR is fully identified.

In this case, one should be careful in ordering the variables in such a way that P^* reflects the desired long-run restrictions: Variables not affected in the long-run by certain shocks should be ordered first while variables that affect other variables only in the short-run should be ordered last.

Final Remarks on Cholesky Identification

In essence, Cholesky decompositions instrument shocks ordered later with shocks ordered earlier. For this reason, identification strategies that rely on the Cholesky decomposition are known as “internal instruments” approaches.

Notice that applying the Cholesky decomposition for short- and long-run restrictions requires assumptions or information beyond the data: It is not an agnostic way of achieving identification. Using a different Cholesky ordering with the same data can produce radically different results. For this reason, it is seldom used in academia or policy analysis nowadays.

Proxy Structural Vector Autoregressions

External Instruments

Work-In-Progress

Sign Restrictions

We impose sign restrictions on the elements of candidate matrices for A^{-1} to achieve set identification. The sign-restrictions are typically informed by theoretical models or economic intuition. For instance, assume that interest rates and inflation are negatively related, interest rates and output growth are negatively related, and inflation and output growth are positively related.

$$A^{-1} = \begin{bmatrix} 1 & a'_{1,2} & a'_{1,3} \\ a'_{2,1} & 1 & a'_{2,3} \\ a'_{3,1} & a'_{3,2} & 1 \end{bmatrix}$$

This sign restriction assumption mechanically looks like:

- $a'_{1,2} < 0, a'_{1,3} < 0$
- $a'_{2,1} < 0 < a'_{2,3}$
- $a'_{3,1} < 0 < a'_{3,2}$

Recall that the diagonal terms are always equal to 1: $a'_{1,1} = a'_{2,2} = a'_{3,3} = 1$.

Sign Restrictions

Continued

To implement the sign restrictions, first recall that $\Omega = A^{-1} (A^{-1})' = PP'$ by Cholesky. For any orthonormal matrix, we have that $PP' = PQQ'P' = (PQ)(PQ)'$

1. Draw a random orthonormal matrix $Q^{(i)}$.
2. For that draw $Q^{(i)}$, calculate the matrix $PQ^{(i)}$.
3. If it conforms with the sign restrictions assumption, $PQ^{(i)}$ is a valid candidate matrix for A^{-1} . Note that $PQ^{(i)}$ no longer has to be lower-triangular.
4. Store $PQ^{(i)}$ and repeat the procedure.

The sign restriction method lends itself nicely to Bayesian estimation because it can be easily implemented in a Gibbs sampler step. Applications and extensions of the sign restrictions used in identification are ubiquitous.

Impulse Responses in a VAR

The impulse response of variable i to a shock in variable j (where $i, j \in \{1, \dots, N\}$ at horizon h is given by iterating the VAR forward:

$$\text{IRF}_{i,j}^{(h)} := \frac{\partial X_i^{(h)}}{\partial \varepsilon_j} = s_i' \Phi^h A^{-1} s_j \quad \text{where} \quad s. := \begin{cases} 1 & \text{if row} = . \\ 0 & \text{otherwise} \end{cases}$$

Note: $s.$ is the $N \times 1$ column selection vector, which equals 1 if the row number is $.$ and 0 otherwise. The ordering of the variables in X determines the order of the structural shocks ε_j .

Confidence intervals for VARs are typically calculated using a bootstrapping procedure that is provided in the estimation package that one is using to estimate the VAR.

Work-In-Progress

Forecast Error Variance Decompositions (FEVDs)

Work-In-Progress

Appendix

I Supplemental Material by Section

II Derivations & Proofs

III Data Sources

Proof: $g_t \approx \ln(x_t) - \ln(x_{t-1})$

We start from the growth rates formula from earlier and take logarithms of both sides.

$$x_\tau = x_0 \prod_{t=1}^{\tau} (1 + g_t)$$

$$\ln(x_\tau) = \ln(x_0) + \sum_{t=1}^{\tau} \ln(1 + g_t)$$

$$\ln(x_\tau) - \ln(x_0) = \sum_{t=1}^{\tau} \ln(1 + g_t)$$

$$\ln(x_t) - \ln(x_{t-1}) = \ln(1 + g_t)$$

The Taylor expansion of $\ln(1 + x)$ around $x^* \approx 0$ ($1 + x^* \approx 1$) yields $\ln(1 + x) \approx \ln(x^*) + \frac{x}{1+x^*} - \frac{x^{*2}}{2(1+x^*)^2} = \ln(1) + x + 0 = x$. Thus:

$$\ln(x_t) - \ln(x_{t-1}) \approx g_t \quad \square$$

Proof: Annualizing Time Series Data

Suppose you have a period-by-period growth rate g_t and a “secant” growth rate g_t^* between arbitrary points x_t and x_{t-h} and x_{t-H} , respectively. We wish to cast g_t in terms of g_t^* .

$$x_t = (1 + g_t)^h \cdot x_{t-h}$$

$$x_t = (1 + g_t^*) \cdot x_{t-H}$$

The period-by-period relationship between x_{t-h} and x_{t-H} is given by:

$$x_{t-h} = (1 + g_t)^{H-h} \cdot x_{t-H}.$$

Combining the first 2 equations yields:

$$\implies (1 + g_t)^h \cdot x_{t-h} = (1 + g_t^*) \cdot x_{t-H}.$$

Combining with the 3rd equation above yields:

$$\implies (1 + g_t)^h (1 + g_t)^{H-h} \cdot \frac{x_{t-h}}{x_{t-h}} = (1 + g_t^*)$$

$$\implies (1 + g_t)^H = (1 + g_t^*)$$

$$\implies (1 + g_t) = (1 + g_t^*)^{\frac{1}{H}} \quad \square$$

Derivation of Generalized Annualization Formulas

The consecutive period-by-period growth rate formula is:

$$x_t = (1 + g_t) \cdot x_{t-1}$$

Between two arbitrary periods x_{t-1} and x_t , this formula is:

$$x_t = (1 + g_t)^h \cdot x_{t-h}$$

Combining with the result from the previous slide yields

$$\begin{aligned} x_t &= (1 + g_t^*)^{\frac{h}{H}} \cdot x_{t-h} \\ \implies g_t^* &= \left(\frac{x_t}{x_{t-h}} \right)^{\frac{H}{h}} - 1 \end{aligned}$$

Rearranging, taking logarithms, and using the Taylor expansion yields the log-difference version of annualization.

$$\begin{aligned} \ln(1 + g_t^*) &= \frac{H}{h} (\ln(x_t) - \ln(x_{t-h})) \\ \implies g_t^* &\approx \frac{H}{h} (\ln(x_t) - \ln(x_{t-h})) \end{aligned}$$

Derivation of Unconditional Mean & Variance of AR(1) Processes

Let $y_t = \theta + \phi y_{t-1} + \varepsilon_t$, $\varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$, $0 < \phi < 1$. Then

$$\begin{aligned}\mathbb{E}[y_t] &= \mathbb{E}[\theta + \phi y_{t-1} + \varepsilon_t] \\ &= \mathbb{E}[\theta] + \mathbb{E}[\phi y_{t-1}] + \mathbb{E}[\varepsilon_t] \\ &= \theta + \phi \mathbb{E}[y_{t-1}] + 0 \\ \implies \mathbb{E}[y_t] &= \theta + \phi \mathbb{E}[y_{t-1}]\end{aligned}$$

Since $\mathbb{E}[y_t] = \mathbb{E}[y_{t-1}]$, rearranging terms yields:

$$\begin{aligned}\mathbb{E}[y_t] - \phi \mathbb{E}[y_{t-1}] &= \theta \\ (1 - \phi) \mathbb{E}[y_t] &= \theta \\ \mathbb{E}[y_t] &= \frac{\theta}{(1 - \phi)} \quad \square\end{aligned}$$

Derivation of Unconditional Variance of AR(1) Processes

Let $y_t = \theta + \phi y_{t-1} + \varepsilon_t$, $\varepsilon_t \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$, $0 < \phi < 1$. Then

$$\begin{aligned}\mathbb{V}[y_t] &= \mathbb{V}[\theta + \phi y_{t-1} + \varepsilon_t] \\ &= \mathbb{V}[\theta] + \mathbb{V}[\phi y_{t-1}] + \mathbb{V}[\varepsilon_t] \\ &= 0 + \phi^2 \mathbb{V}[y_{t-1}] + \sigma_\varepsilon^2 \\ \implies \mathbb{V}[y_t] &= \phi^2 \mathbb{V}[y_{t-1}] + \sigma_\varepsilon^2\end{aligned}$$

Since $\mathbb{V}[y_t] = \mathbb{V}[y_{t-1}]$, rearranging terms yields:

$$\begin{aligned}\mathbb{V}[y_t] - \phi^2 \mathbb{V}[y_{t-1}] &= \sigma_\varepsilon^2 \\ (1 - \phi^2) \mathbb{V}[y_t] &= \sigma_\varepsilon^2 \\ \mathbb{V}[y_t] &= \frac{\sigma_\varepsilon^2}{(1 - \phi^2)} \quad \square\end{aligned}$$

Appendix

I Supplemental Material by Section

II Derivations & Proofs

III Data Sources

Illustrative Figure Sources

The data used in this presentation is open-source and comes from the following sources.

- PCEPI: PCE Price Index
 - Source: U.S. Bureau of Economic Analysis
 - Retrieved from: FRED — PCEPI
- UR: Unemployment Rate
 - Source: U.S. Bureau of Labor Statistics
 - Retrieved from: FRED — UNRATE
- FFR: Federal Funds Effective Rate
 - Source: Federal Reserve Board of Governors
 - Retrieved from: FRED — FEDFUNDS
- GDP: Real Gross Domestic Product
 - Source: U.S. Bureau of Economic Analysis
 - Retrieved from: FRED — GDPC1

Response Variable Sources

- SP500: S&P500 Composite Stock Market Index
 - Source: Standard & Poor's
 - Retrieved from: FRED — SP500 (from 2016 onwards)
 - Retrieved from: Robert Shiller's website (prior to 2016): <https://shillerdata.com/>
- employment: Total Nonfarm Payrolls
 - Source: U.S. Bureau of Labor Statistics
 - Retrieved from: FRED — PAYEMS
- PCEPIX: Core PCE Price Index
 - Source: U.S. Bureau of Economic Analysis
 - Retrieved from: FRED — PCEPILFE
- GZCS: Gilchrist & Zakrajšek Credit Spread
 - Source: Federal Reserve Board of Governors
 - Retrieved from: Federal Reserve Board of Governors Website
- corporate_bonds: Nonfinancial Corporate Bonds
 - Source: Federal Reserve Board of Governors
 - Retrieved from: FRED — CBLBSNNCB
- house_prices: U.S. House Price Index
 - Source: U.S. Federal Housing Finance Agency
 - Retrieved from: FRED — USSTHPI

Impulse Variable (Shock) Sources

- NOT_MP: Monetary policy shock from Nunes, Ozdagli, and Tang (2022)
- K_OSN: Oil Supply News Shocks from Känzig (2021)