Empirical Macroeconomic Modelling: 1. Time Series and VARs

Karl Whelan

School of Economics, UCD

June 2014

Empirical Macroeconomic Modelling

- You have probably already seen lots of macroeconomic models. Many of them, like IS-LM or the Solow growth model teach useful principles but aren't very precise about how long it takes for things to happen or the size of the impacts of various events.
- We are going to discuss empirical economic modelling.
- Ideally, we would like an empirical economic model to deliver specific concrete answers to concrete questions.
- For example, if you are asked what be would be effect on the economy of a quarter point increase in the ECB policy rate, then an ideal answer would communicate how consumption, investment, output, and inflation would respond next quarter and the quarter after that, and so on.
- It would also communicate the amount of uncertainty surrounding this answer.
- It would also be nice to explain *why* you are giving this particular answer since whoever you're giving the answer to might be interested in understanding what's going on.

The Plan for this Course

- My plan is to first introduce some pure time series models that can be used for forecasting and analysis and then to give a fairly brief introduction to the methods underlying so-called **Dynamic Stochastic General Equilibrium** (DSGE) models.
- The course will be organised as follows
  - **Day One**
    - ★ Time series as a framework for modern macroeconomics.
    - ★ Vector Autoregressions (VARs): Identification and estimation.
    - ★ Some examples of VAR studies.
  - **Day Two**
    - ★ More examples of VAR studies.
    - ★ Analysing VAR models with RATS.
    - ★ Introduction to rational expectations modelling.
  - **Day Three**
    - ★ Introduction to DSGE: The real business cycle model
    - ★ Log-linearisation and solving for steady states.
    - ★ Simulating DSGE models.

Part I

Separating Trends from Cycles

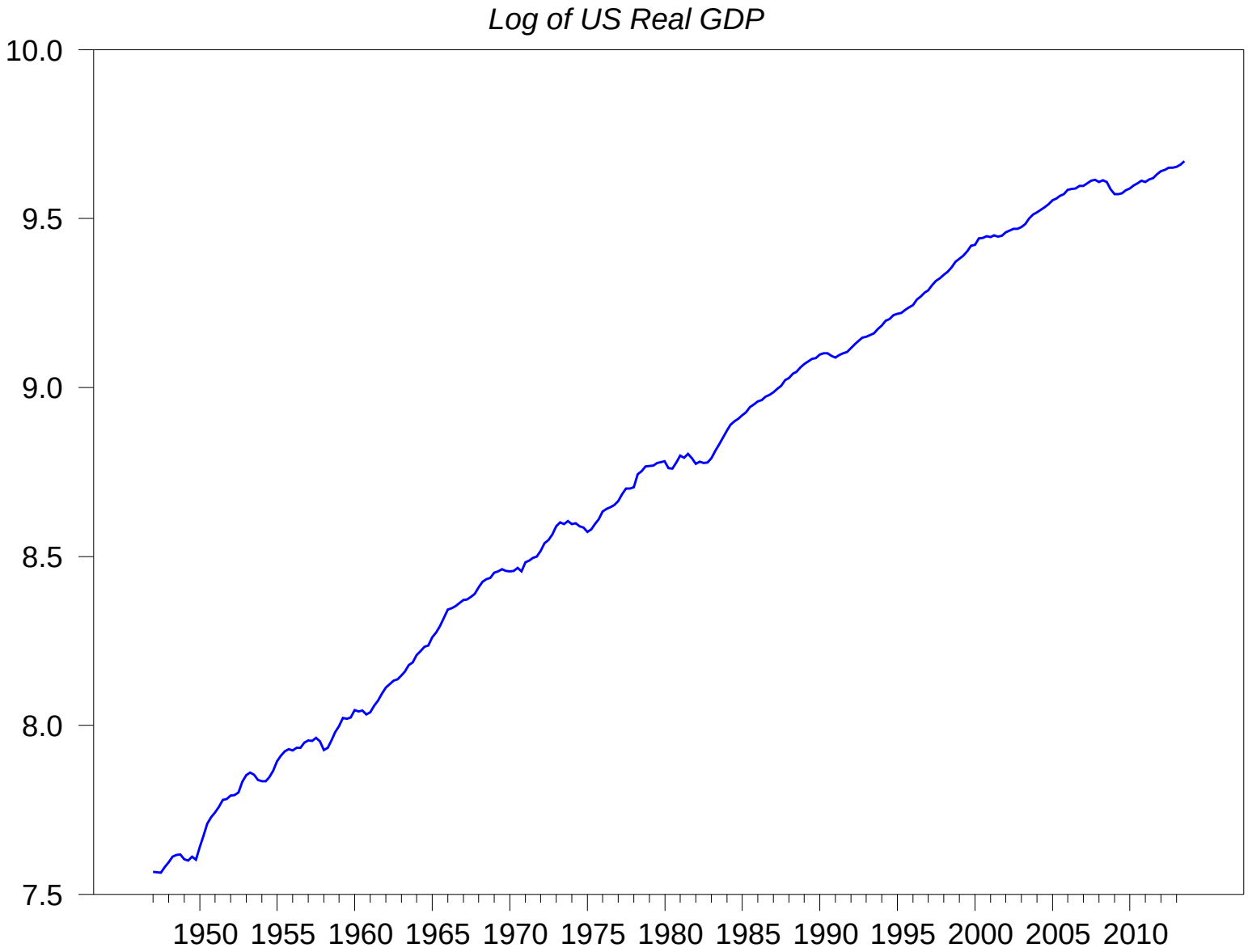
Trends and Cycles

- Macroeconomists tend to break series into a "non-stationary" long-run trend and a "stationary" cyclical component.
- "Business cycle analysis" relates to this modelling and explaining the cylical components of the major macroeconomic variables.
- Fine in theory, but how is this done in practice?
- Simplest method: Log-linear trend
  - Estimated from regression

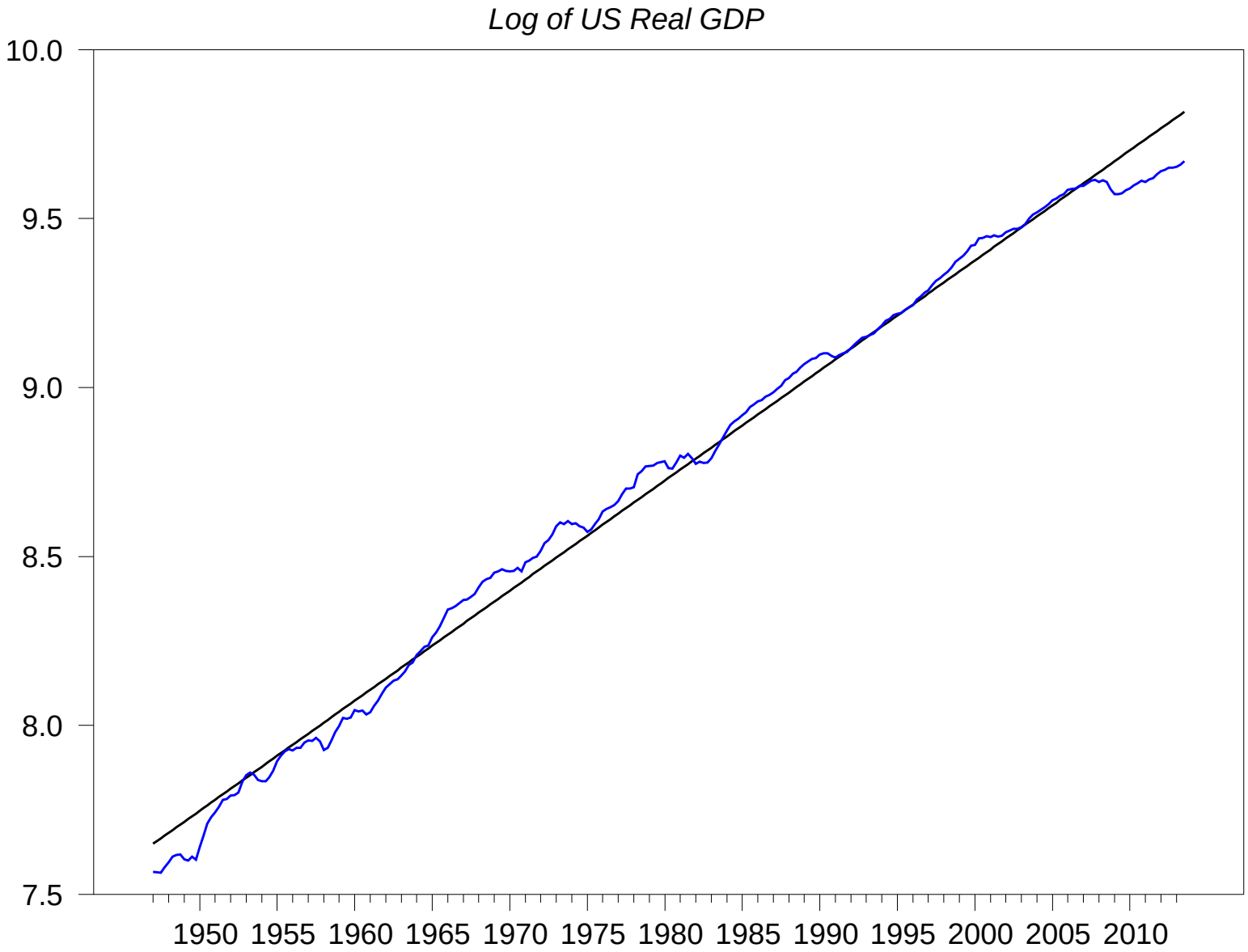
$$\log(Y_t) = y_t = \alpha + gt + \epsilon_t$$

  - Trend component $\alpha + gt$.
  - Zero-mean stationary cyclical component ϵ_t.
  - Log-difference Δy_t (equivalent to growth rate) has two components: Constant trend growth g and the change in cyclical component $\Delta\epsilon_t$.

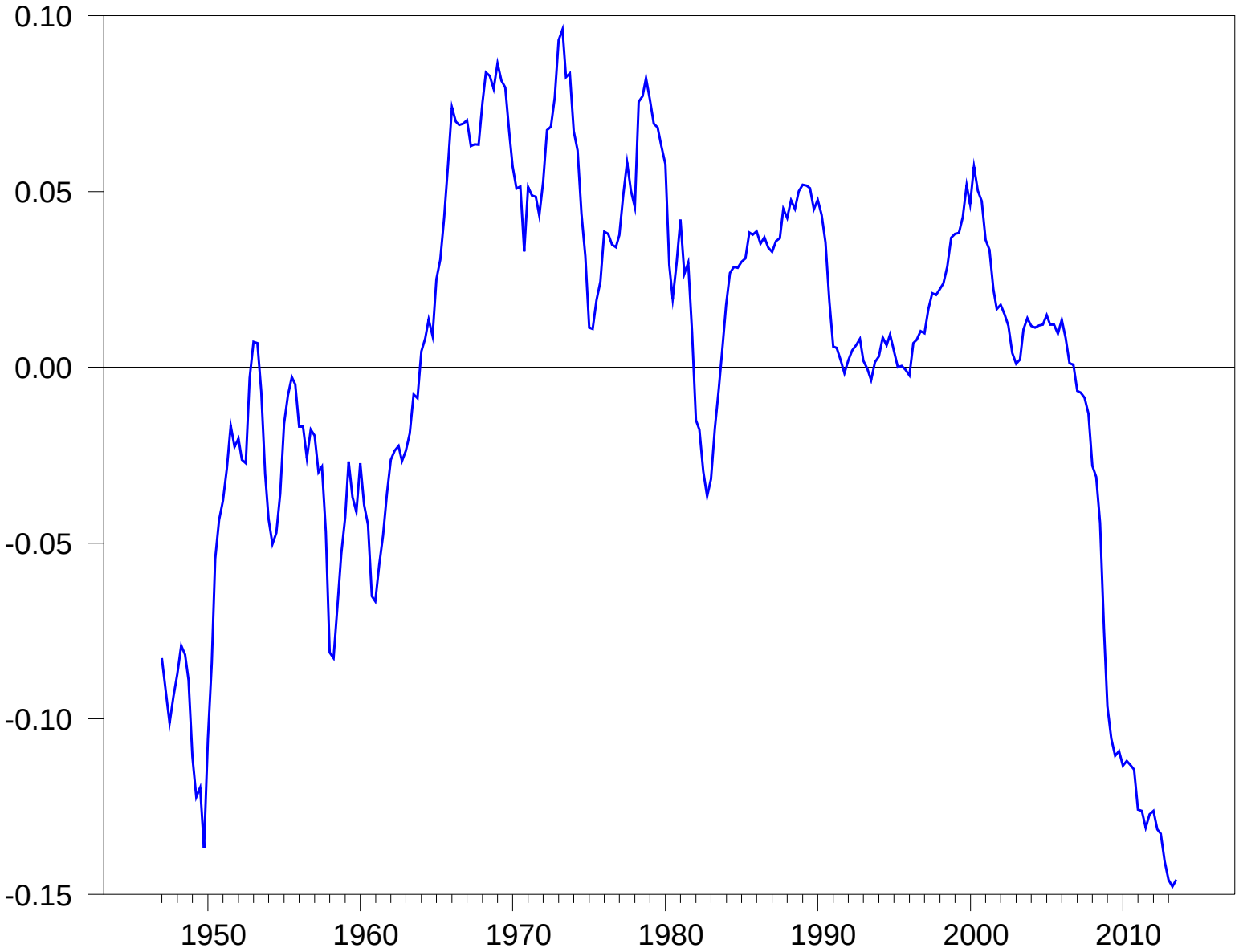
Trends and Cycles in US GDP: Cycles Are Pretty Small



Simplest Example: Log-Linear Trend



Cycles From a Log-Linear Trend Model



Potential Problems: A Stochastic Trend Model

- Drawing straight lines to detrend series can provide misleading results. For example, suppose the correct model is

$$y_t = g + y_{t-1} + \epsilon_t$$

- Growth has a constant component g and a random bit ϵ_t.
- Cycles are just the accumulation of all the random shocks that have affected Δy_t over time.
- There is no tendency to revert to the trend: Expected growth rate is always g no matter what has happened in the past.
- In this case Δy_t is stationary: First-differencing gets rid of the non-stationary stochastic trend component of the series.
- In this example, if we fitted a log-linear trend line through the series, there might appear to be a mean-reverting cyclical component but there is not.
- So detrending times series is not generally as simple as drawing a straight line.

Variations in Trend Growth: The Hodrick-Prescott Filter

- A more realistic model should be one in which we accept that growth rate of the trend probably varies a bit over time leaving a cycle that moves up and down over time.
- Hodrick and Prescott (1981) suggested choosing the time-varying trend Y_t^* so as to minimize

$$\sum_{t=1}^{N}\left[\left(Y_t - Y_t^*\right)^2 + \lambda\left(\Delta Y_t^* - \Delta Y_{t-1}^*\right)\right]$$

- This method tries to minimize the sum of squared deviations between output and its trend $\left(Y_t - Y_t^*\right)^2$ but also contains a term that emphasises minimizing the change in the trend growth rate $\left(\lambda\left(\Delta Y_t^* - \Delta Y_{t-1}^*\right)\right)$.
- How do we choose λ and thus weight the goodness-of-fit of the trend versus smoothness of the trend?
- $\lambda = 1600$ is the standard value used in business cycle detrending. We will discuss this in a minute.
- Many DSGE modellers apply a HP filter to their data and then analyse only the cyclical components.

Another Angle on the Hodrick-Prescott Filter

- Another way to view the HP filter is to see it as a way of estimating an unobserved or latent variable.
- For example, an underlying model that fits well with the HP filter is

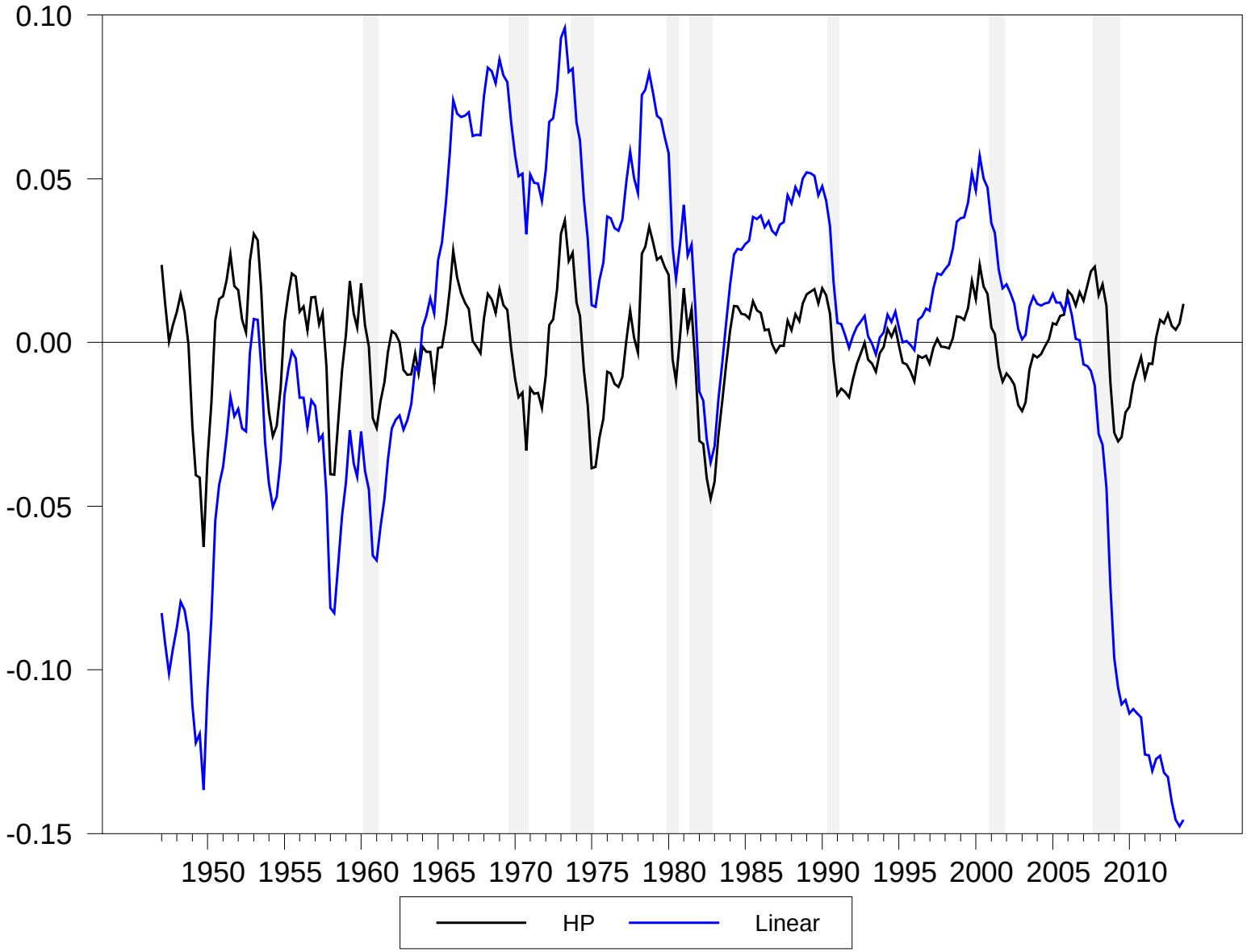
$$\begin{aligned} Y_t &= Y_t^* + C_t \\ \Delta Y_t^* &= \Delta Y_{t-1}^* + \epsilon_t^g \end{aligned}$$

- In this model, there is an observed output variable, Y_t and an unobserved trend variable, Y_t^*.
- There is also shock each period to the both trend growth component and the cyclical component.
- The problem of extracting an unobserved trend from this series is similar to the classical statistical problem of extracting a true "signal" from a measured series that includes "noise".
- One formal method for dealing with models like this is the so-called Kalman filter.

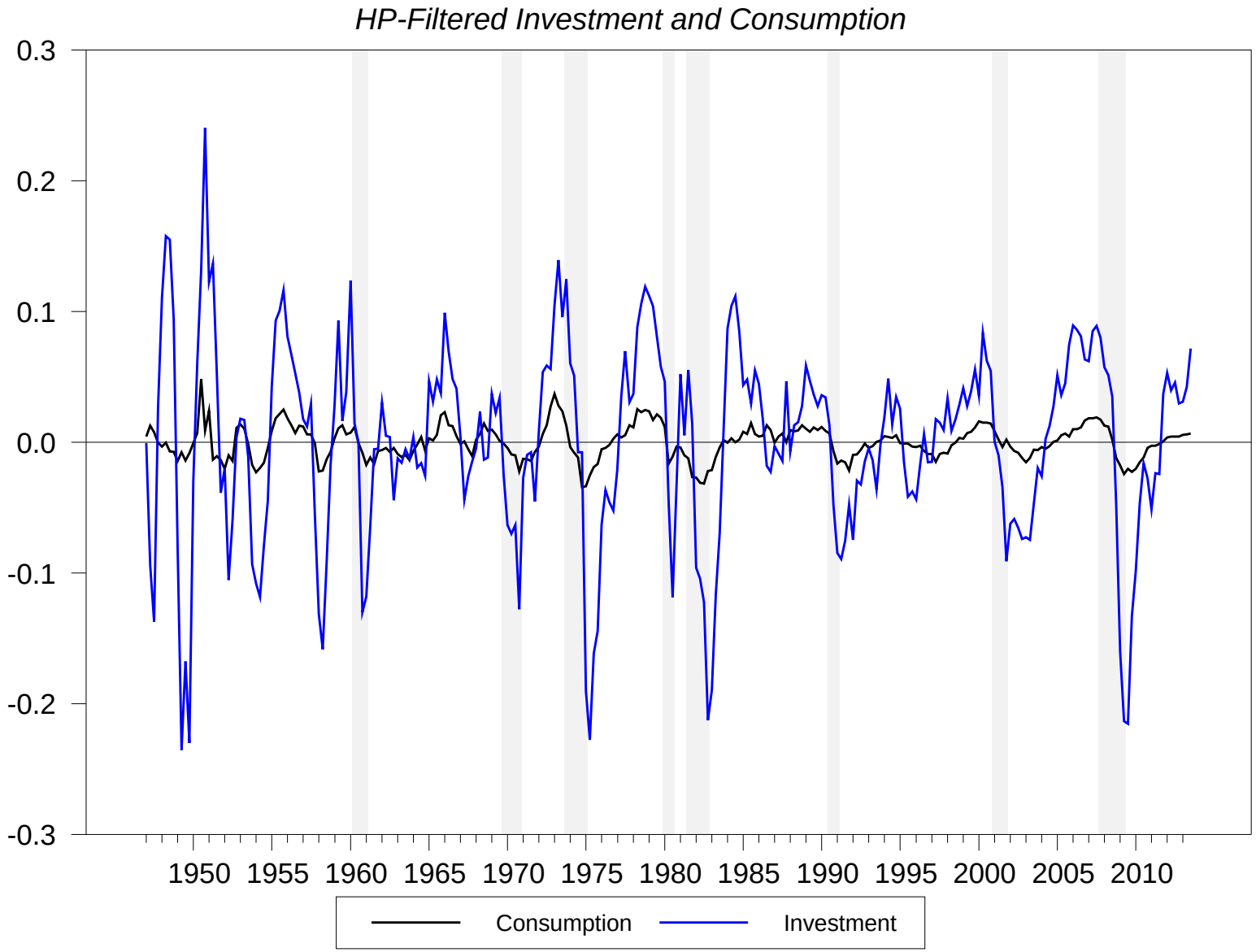
The Kalman Filter and $\lambda = 1600$

- We won't have time to cover the Kalman filter in this course but this method estimates the unobserved variable as follows.
  1. An estimate for the starting value of the unobserved variable is made and assumptions made about the uncertainty surrounding this guess.
  2. Then, after each data point is observed, the estimate of the unobserved variable is updated.
  3. How much the estimate is updated depends on the estimated variance of the shocks to the trend and cycle. If the trend growth rate is assumed to be highly variable, then a lot of weight is given to observations suggesting a potential higher or lower trend (and the opposite if the trend growth rate is assumed to vary little over time).
- Over long samples and under certain conditions, the HP filter trend is the same as the optimal Kalman filter estimates with λ being the ratio of the variance of C_t to the variance of ϵ_t^g.
- Hodrick and Prescott assumed C_t had a standard deviation of 5 percentage points while ϵ_t^g had a standard deviation of one-eighth of a percentage point. Hence $\lambda = \frac{5^2}{\left(\frac{1}{8}\right)^2} = (25)(64) = 1600$.

HP-Filtered Cycles Correspond Well to NBER Recessions



Investment Cycles Are Bigger than Consumption Cycles



Part II

Modelling Cycles

The AR(1) Model and Impulse Responses

- Cyclical components are positively autocorrelated (i.e. positively correlated with their own lagged values). and also exhibit random-looking fluctuations.
- One simple model that captures these features is the AR(1) model (Auto-Regressive of order 1):

$$y_t = \rho y_{t-1} + \epsilon_t$$

- Suppose an AR(1) series starts out at zero. Then there is a unit shock, $\epsilon_t = 1$ and then all shocks are zero afterwards.
- Period t, we have $y_t = 1$, period $t+1$, we have $y_{t+1} = \rho$, period $t+n$, we have $y_{t+n} = \rho^n$ and so on.
- The shock fades away gradually. How fast depends on the size of ρ. The time path of y after this hypothetical shock is known as the **Impulse Response Function**.
- Can think of this as the path followed from t onwards when shocks are $(\epsilon_t + 1, \epsilon_{t+1}, \epsilon_{t+2}, ...)$ instead of $(\epsilon_t, \epsilon_{t+1}, \epsilon_{t+2}, ...)$, i.e. the incremental effect in all future periods of a unit shock today.
- IRF graphs are commonly used to illustrate dynamic properties of macro data.

Volatility: Shocks and Propagation Mechanisms

- Consider the AR(1) model

$$y_t = \rho y_{t-1} + \epsilon_t$$

- Suppose the variance of ϵ_t is σ_ϵ^2.
- The long run variance of y_t is the same as the long-run variance of y_{t-1} and (remembering that ϵ_t is independent of y_{t-1}) this is given by

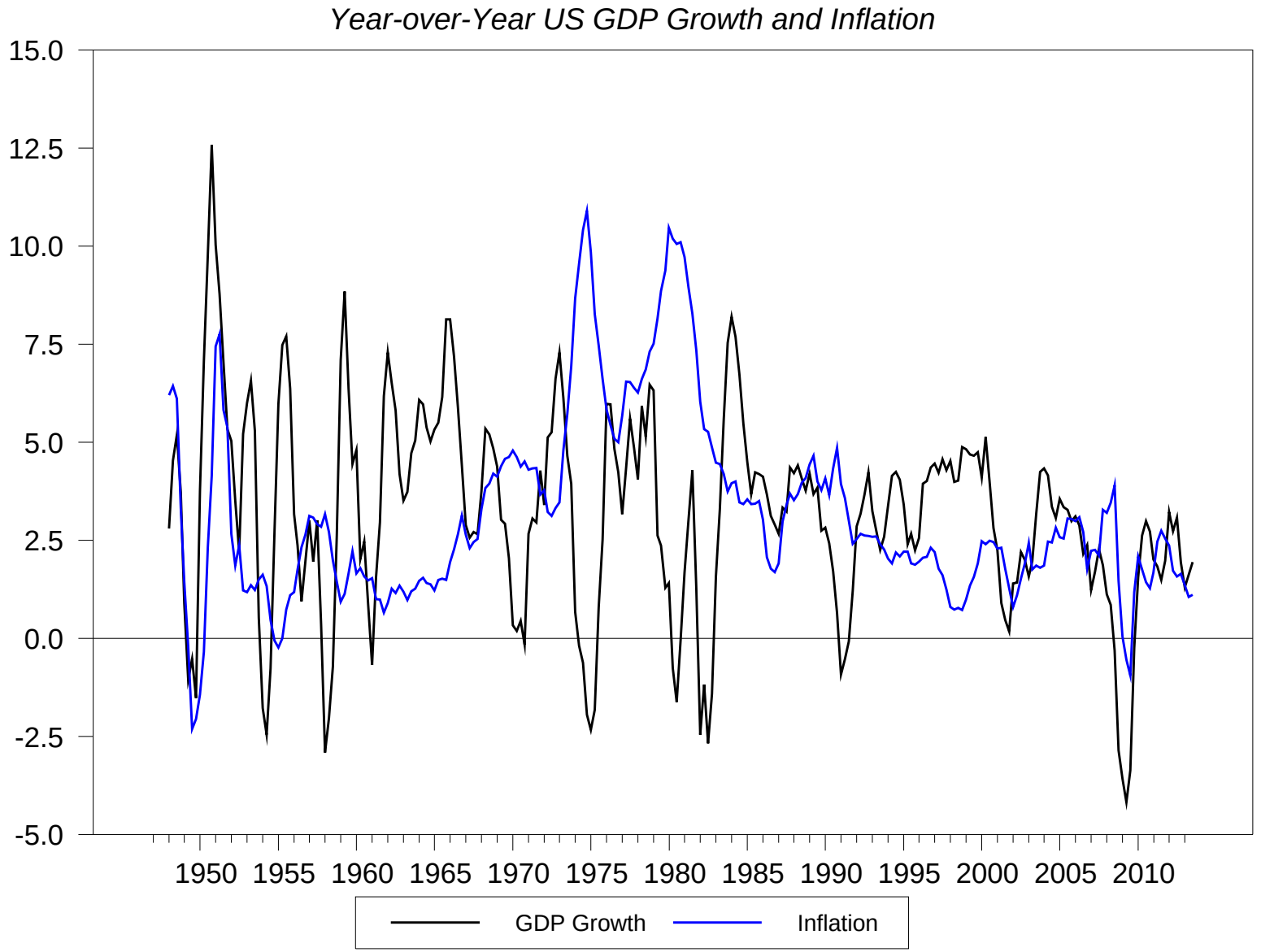
$$\sigma_y^2 = \rho^2 \sigma_y^2 + \sigma_\epsilon^2$$

- Simplifies to $\sigma_y^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}$
- The variance of output depends positively on both shock variance σ_ϵ^2 and also on the persistence parameter ρ.
- So the volatility of the series is partly due to size of shocks but also due to the strength of the **propagation mechanism**.

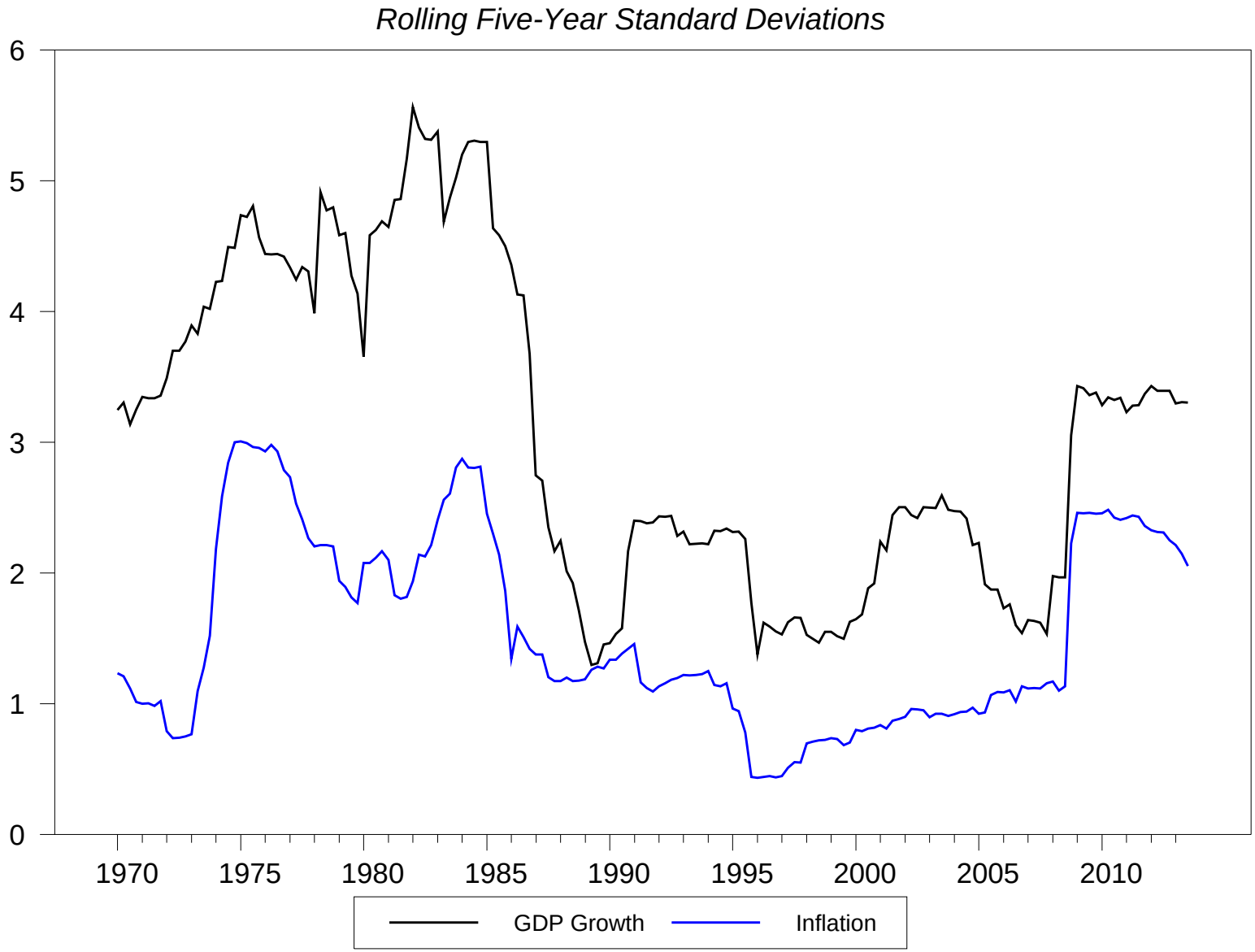
Example: The Great Moderation

- An interesting pattern: Output and inflation became substantially less volatile after the mid-1980s. This was widely dubbed "The Great Moderation"
- This pattern occurred in all the world's major economies.
- What was the explanation?
- Smaller shocks? (Smaller values of ϵ_t)
  1. Less random policy shocks?
  2. Smaller shocks from goods markets or financial markets?
  3. Smaller supply shocks?
- Weaker propagation mechanisms? (Smaller values of ρ)
  1. Did policy become more stabilizing?
  2. Did the economy become more stable, e.g. better inventory management, increased share of services?
  3. Some had thought that financial modernization had stabilized the economy. Less clear now!
- Does the 2008-2009 global recession and subsequent slow recovery spell the end for the Great Moderation?

Less Extreme Movements in Output Growth and Inflation



The Great Moderation: Substantial Reductions in Volatility



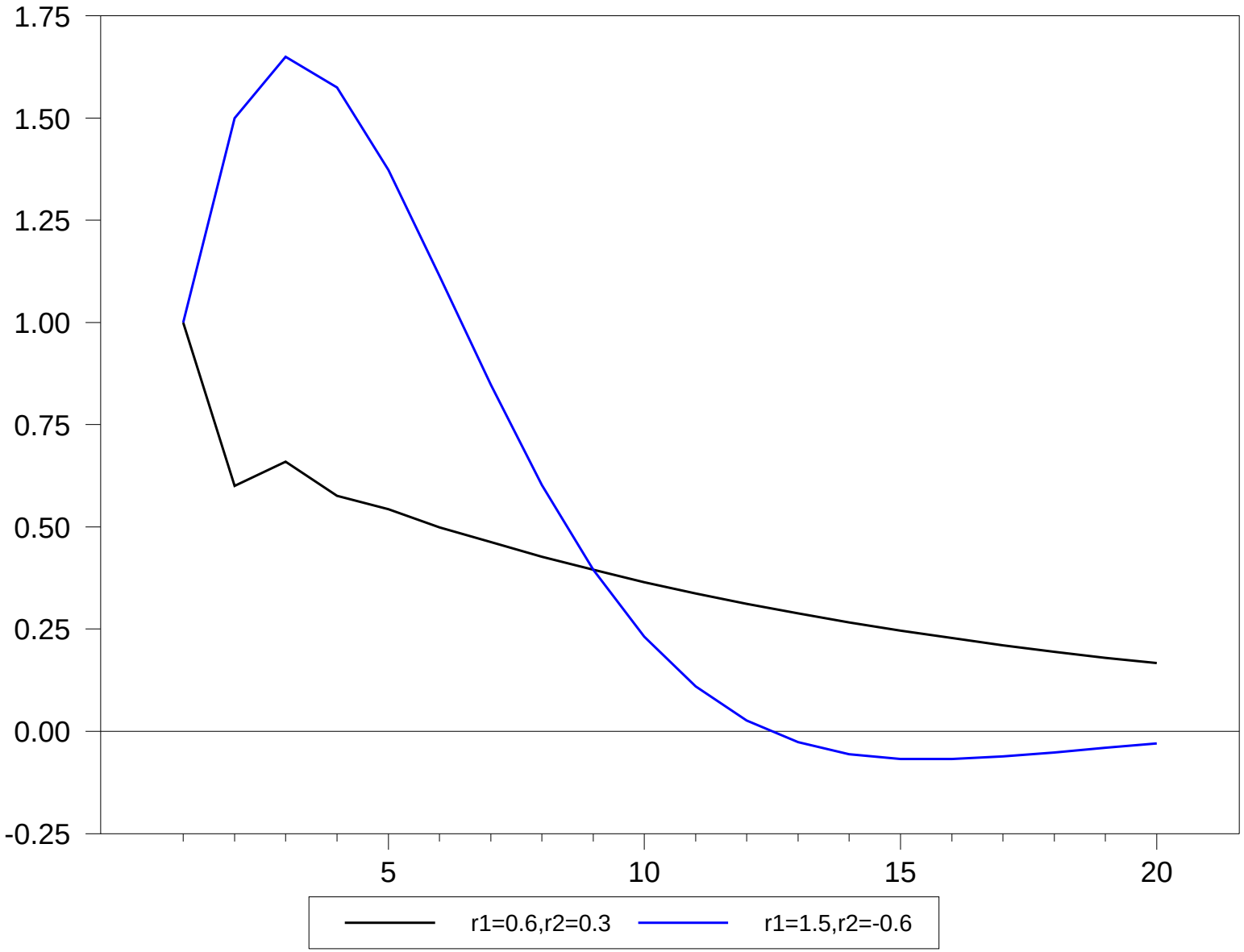
More Complex Dynamics: The AR(2) Model

- Not all impulse response functions just erode gradually of time as in the AR(1) model.
- Macroeconomic dynamics can often be far more complicated.
- Consider the AR(2) model:

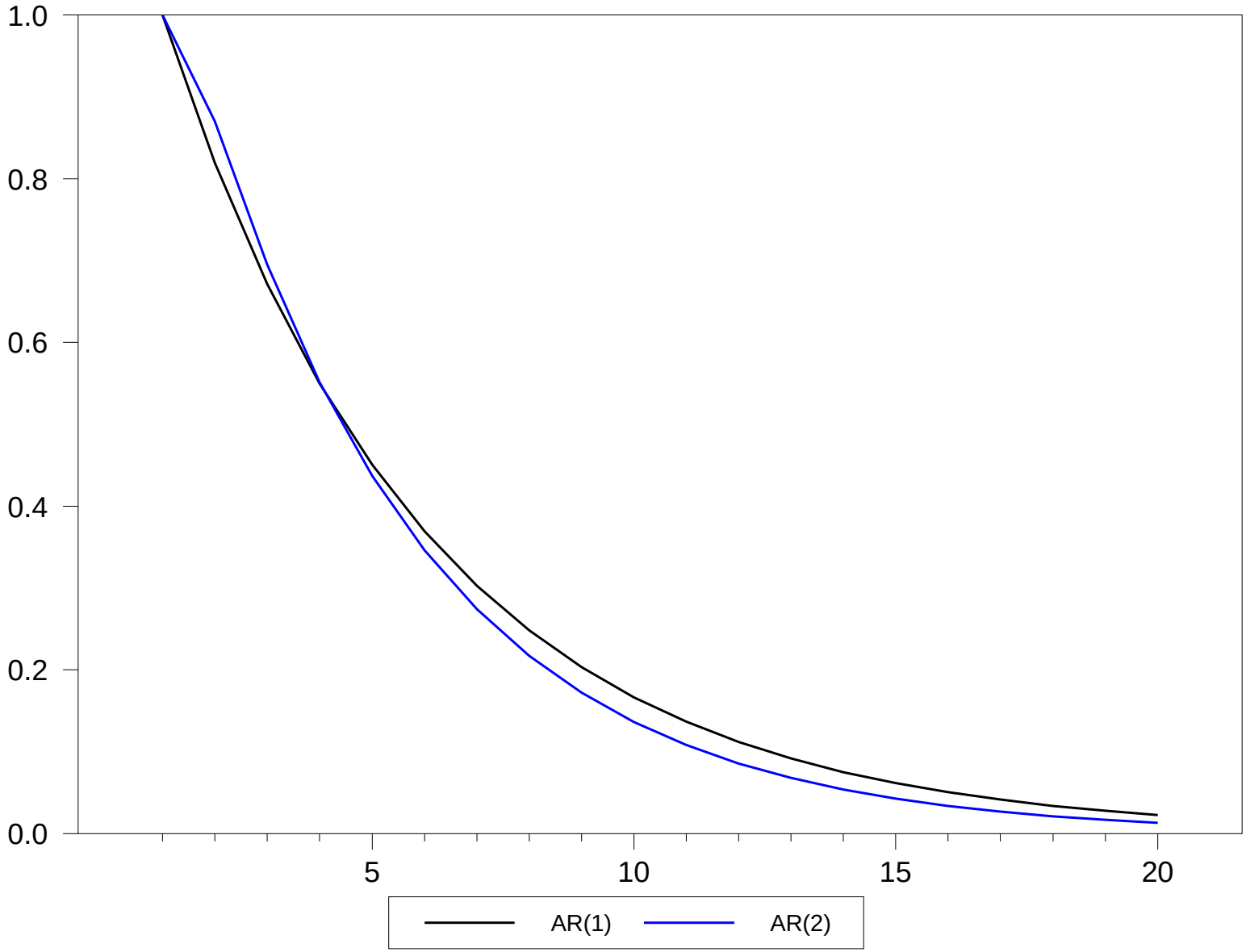
$$y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \epsilon_t$$

- This type of model can generate various types of impulse response functions such as oscillating or hump-shaped responses.
- AR(3) and higher models can generate even more complex responses.
- Lesson: The dynamic properties of your model will depend upon how many lags you allow.
- Practitioners constructing empirical models often run battery of lag selection tests to decide upon the appropriate lag length.

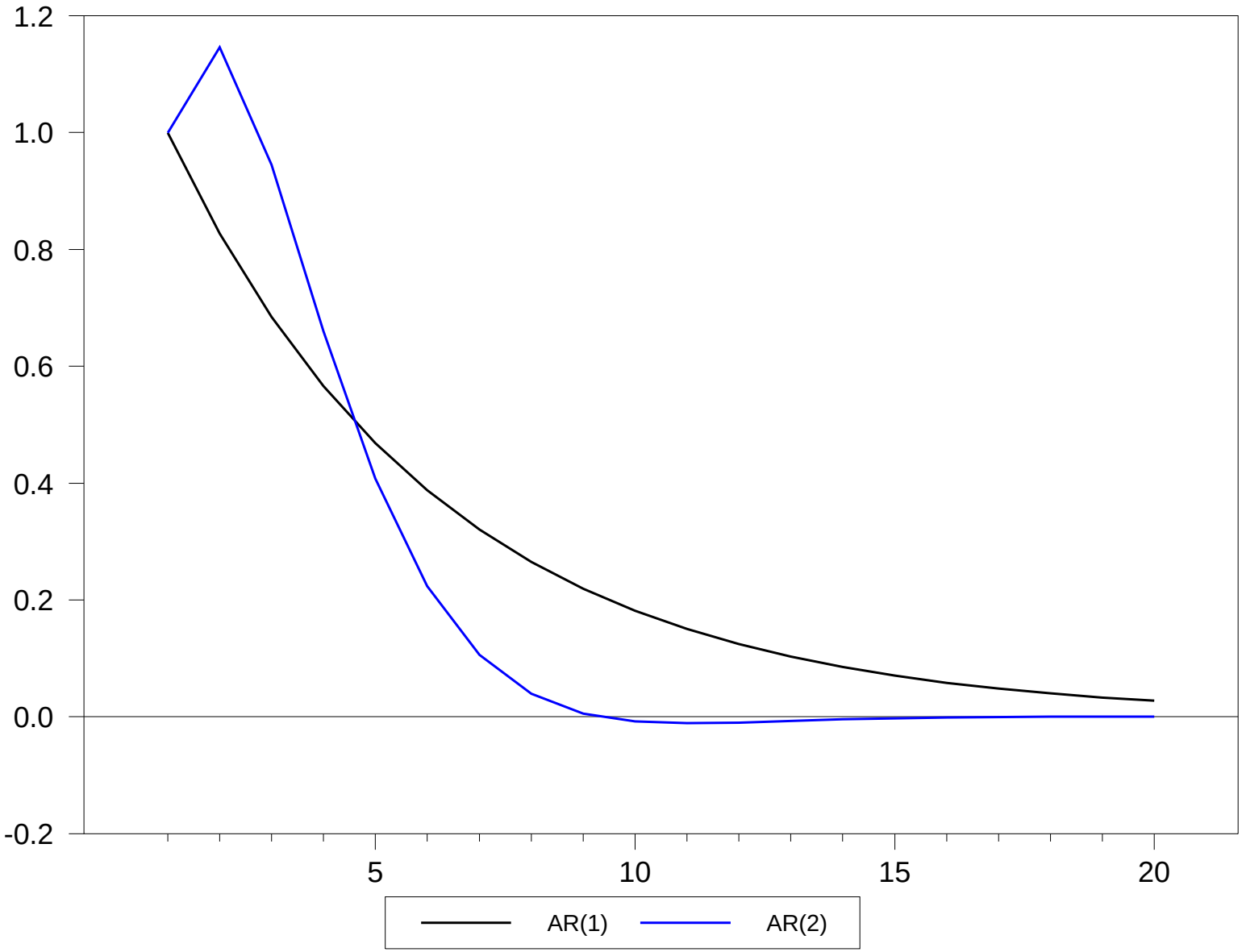
Two Examples of AR(2) Impulse Responses



Consumption Dynamics Seem to be AR(1)



Output AR(2) Model Shows A Small Humped-Shape IRF



Vector Autoregressions

- AR models are a very useful tool for understanding the dynamics of individual variables.
- But they ignore the *interrelationships* between variables.
- Vector Autoregressions (VARs) model the dynamics of n different variables, allowing each variable to depend on lagged values of all of the variables.
- Can examine impulse responses of all n variables to all n shocks.
- Simplest example is two variables and one lag:

$$\begin{aligned} y_{1t} &= a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + e_{1t} \\ y_{2t} &= a_{21}y_{1,t-1} + a_{22}y_{2,t-1} + e_{2t} \end{aligned}$$

- Invented by Chris Sims (1980). Now used as a central tool in applied macroeconomics.

What Are These Shocks?

Macroeconomists now spend a lot of time examining the shocks in VAR models and their effects. But what are the shocks? Lots of possibilities:

1. Policy changes not due to the systematic component of policy captured by the VAR equation.
2. Changes in preferences, such as attitudes to consumption versus saving or work versus leisure.
3. Technology shocks: Random increases or decreases in the efficiency with which firms produce goods and services.
4. Shocks to various frictions: Increases or decreases in the efficiency with which various markets operate, such as the labour market, goods markets, or financial markets.

Time Series as a Framework for Empirical Macro

- The time series perspective—cycles being determined by various random shocks which are propagated throughout the economy over time—is central to how modern macroeconomists now view economic fluctuations.
- VARs are a very common framework for modelling macroeconomic dynamics and the effects of shocks.
- But while VARs can describe how things work, they cannot explain *why* things work that way.
- To have real confidence in a description of how the economy works, we ideally want to know how people in the economy behave and why they they behave that way.
- That's where economic theory comes in.
- DSGE models aim to have the dynamic structure of VARs (shocks and propagation mechanisms, IRFs) but are derived from economic theory in which all agents are rational and optimizing.

Part III

Introducing VAR Methods

Background on VARs

- These model were introduced to the economics profession by Christopher Sims (1980) in a path-breaking article titled "Macroeconomics and Reality."
- Sims was indeed telling the macro profession to "get real."
- He criticized the widespread use of highly specified macro-models that made very strong identifying restrictions (in the sense that each equation in the model usually excluded most of the model's other variables from the right-hand-side) as well as very strong assumptions about the dynamic nature of these relationships.
- VARs were an alternative that allowed one to model macroeconomic data accurately, without having to impose lots of incredible restrictions. In the phrase used in an earlier paper by Sargent and Sims (who shared the Nobel prize award) it was "macro modelling without pretending to have too much a priori theory."
- We will see that VARs are not theory free. But they do make the role of theoretical identifying assumptions far clearer than was the case for the types of models Sims was criticizing.

Matrix Formulation of VARs

- The simplest possible VAR features two variables and one lag:

$$\begin{aligned} y_{1t} &= a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + e_{1t} \\ y_{2t} &= a_{21}y_{1,t-1} + a_{22}y_{2,t-1} + e_{2t} \end{aligned}$$

- The most compact way to express a VAR system like this is to use matrices. Defining the matrices

$$\begin{aligned} Y_t &= \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} \\ A &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \\ e_t &= \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \end{aligned}$$

- This system can be written as

$$Y_t = AY_{t-1} + e_t$$

Vector Moving Average (VMA) Representation

- VARs express variables as function of what happened yesterday and today's shocks.
- But what happened yesterday depended on yesterday's shocks and on what happened the day before.
- This VMA representation is obtained as follows

$$\begin{aligned} Y_t &= e_t + AY_{t-1} \\ &= e_t + A\left(e_{t-1} + AY_{t-2}\right) \\ &= e_t + Ae_{t-1} + A^2\left(e_{t-2} + AY_{t-3}\right) \\ &= e_t + Ae_{t-1} + A^2 e_{t-2} + A^3 e_{t-3} + ... + A^t e_0 \end{aligned}$$

- This makes clear how today's values for the series are the cumulation of the effects of all the shocks from the past.
- It is also useful for deriving predictions about the properties of VARs.

Impulse Response Functions

- Suppose there is an initial shock defined as

$$e_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

  and then all error terms are zero afterwards, i.e. $e_t = 0$ for $t > 0$.

- Recall VMA representation

$$Y_t = e_t + Ae_{t-1} + A^2 e_{t-2} + A^3 e_{t-3} + ... + A^t e_0$$

- This tells us that the response after n periods is $A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

- So IRFs for VARs are directly analagous to the IRFs for AR(1) models that we looked at before.

Using a VAR to Forecast

- VARs are often used for forecasting.
- Supppose we observe our vector of variables Y_t. What's our forecast for Y_{t+1}?
- The model for next period is

$$Y_{t+1} = AY_t + e_{t+1}$$

- Because $E_t e_{t+1} = 0$, an unbiased forecast at time t is AY_t. In other words, $E_t Y_{t+1} = AY_t$.
- The same reasoning tells us that $A^2 Y_t$ is an unbiased forecast of Y_{t+2} and $A^3 Y_t$ is an unbiased forecast of Y_{t+3} and so on.
- So once a VAR is estimated and organised to be in this form, it is very easy to construct forecasts.

Generality of the First-Order Matrix Formulation: I

- The model we've been looking at may seem like a small subset of all possible VARs because it doesn't have a constant term and only has lagged values from one period ago.
- However, one can add a third variable here which takes the constant value 1 each period. The equation for the constant term will just state that it equals its own lagged values. So this formulation actually incorporates models with constant terms.
- We would also expect most equations in a VAR to have more than one lag. Surely this makes things much more complicated?
- Not really. It turns out, the first-order matrix formulation can represent VARs with longer lags.
- Consider the two-lag system

$$\begin{aligned} y_{1t} &= a_{11}y_{1,t-1} + a_{12}y_{1,t-2} + a_{13}y_{2,t-1} + a_{14}y_{2,t-2} + e_{1t} \\ y_{2t} &= a_{21}y_{1,t-1} + a_{22}y_{1,t-2} + a_{23}y_{2,t-1} + a_{24}y_{2,t-2} + e_{2t} \end{aligned}$$

Generality of the First-Order Matrix Formulation: II

- Now define the vector

$$Z_t = \begin{pmatrix} y_{1t} \\ y_{1,t-1} \\ y_{2t} \\ y_{2,t-1} \end{pmatrix}$$

- This system can be represented in matrix form as

$$Z_t = AZ_{t-1} + e_t$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 1 & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad e_t = \begin{pmatrix} e_{1t} \\ 0 \\ e_{2t} \\ 0 \end{pmatrix}$$

- This is sometimes called the "companion form" matrix formulation.

Interpreting Shocks and Impulse Responses

- The system we've been looking at is usually called a *reduced-form* VAR model.
- It is a purely *econometric* model, without any theoretical element.
- How should we interpret it? One interpretation is that e_{1t} is a shock that affects only y_{1t} on impact and e_{2t} is a shock that affects only y_{2t} on impact.
- For instance, one can use the IRFs generated from an inflation-output VAR to calculate the dynamic effects of "a shock to inflation" and "a shock to output".
- But other interpretations are available.
- For instance, one might imagine that the true shocks generating inflation and output are an "aggregate supply" shock and an "aggregate demand" shock and that both of these shocks have a direct effect on both inflation and output.
- How would we identify these "structural" shocks and their impulse responses?

The Multiplicity of Shocks and IRFs

- Suppose reduced-form and structural shocks are related by

$$\begin{aligned} e_{1t} &= c_{11}\epsilon_{1t} + c_{12}\epsilon_{2t} \\ e_{2t} &= c_{21}\epsilon_{1t} + c_{22}\epsilon_{2t} \end{aligned}$$

- Can write this in matrix form as

$$e_t = C\epsilon_t$$

- These two VMA representations describe the data equally well:

$$\begin{aligned} Y_t &= e_t + Ae_{t-1} + A^2e_{t-2} + A^3e_{t-3} + ... + A^te_0 \\ &= C\epsilon_t + AC\epsilon_{t-1} + A^2C\epsilon_{t-2} + A^3C\epsilon_{t-3} + ... + A^tC\epsilon_0 \end{aligned}$$

- Can interpret the model as one with shocks e_t and IRFs given by A^n.
- Or as a model with structural shocks ϵ_t and IRFs are given by A^nC.
- And we could do this for any C: We just don't know the structural shocks.

Contemporaneous Interactions: I

- Another way to see how reduced-form shocks can be different from structural shocks is if there are contemporaneous interactions between variables, which is likely.
- Consider the following model:

$$\begin{aligned} y_{1t} &= a_{12}y_{2t} + b_{11}y_{1,t-1} + b_{12}y_{2,t-1} + \epsilon_{1t} \\ y_{2t} &= a_{21}y_{1t} + b_{21}y_{1,t-1} + b_{22}y_{2,t-1} + \epsilon_{2t} \end{aligned}$$

- Can be written in matrix form as

$$AY_t = BY_{t-1} + \epsilon_t$$

where

$$A = \begin{pmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{pmatrix}$$

Contemporaneous Interactions: II

- Now if we estimate the "reduced-form" VAR model

$$Y_t = DY_{t-1} + e_t$$

- Then the reduced-form shocks and coefficients are

$$\begin{aligned} D &= A^{-1}B \\ e_t &= A^{-1}\epsilon_t \end{aligned}$$

- Again, the following two decompositions both describe the data equally well

$$\begin{aligned} Y_t &= e_t + De_{t-1} + D^2e_{t-2} + D^3e_{t-3} + ... \\ &= A^{-1}\epsilon_t + DA^{-1}\epsilon_{t-1} + D^2A^{-1}\epsilon_{t-2} + ... + D^tA^{-1}\epsilon_0 \end{aligned}$$

- For the structural model, the impulse responses to the structural shocks from n periods are given by D^nA^{-1}.
- Again, this is true for any arbitrary A matrix.

Why Care?

- There is no problem with forecasting with reduced-form VARs: Once you know the reduced-form shocks and how they have affected today's value of the variables, you can use the reduced-form coefficients to forecast.
- The problem comes when you start asking "what if" questions? For example, "what happens if there is a shock to the first variable in the VAR?"
- In practice, the error series in reduced-form VARs are usually correlated with each other. So are you asking "What happens when there is a shock to the first variable only?" or are you asking "What usually happens when there is a shock to the first variable given that this is usually associated with a corresponding shock to the second variable?"
- Most likely, the really interesting questions about the structure of the economy relate to the impact of different types of shocks that are uncorrelated with each other.
- A structural identification that explains how the reduced-form shocks are actually combinations of uncorrelated structural shocks is far more likely to give clear and interesting answers.

Structural VARs: A General Formulation

- In its general formulation, the structural VAR is

$$AY_t = BY_{t-1} + C\epsilon_t$$

- The model is fully described by the following parameters:
  1. n^2 parameters in A
  2. n^2 parameters in B
  3. n^2 parameters in C
  4. $\frac{n(n+1)}{2}$ parameters in Σ, which describes the pattern of variances in covariances underlying the shock terms.
- Adding all these together, we see that the most general form of the structural VAR is a model with $3n^2 + \frac{n(n+1)}{2}$ parameters.

Identification of Structural VARs: The General Problem

- Estimating the reduced-form VAR

$$Y_t = DY_{t-1} + e_t$$

gives us information on $n^2 + \frac{n(n+1)}{2}$ parameters: The coefficients in D and the estimated covariance matrix of the reduced-form errors.

- To obtain information about structural shocks, we thus need to impose $2n^2$ *a priori* theoretical restrictions on our structural VAR.
- This will leave us with $n^2 + \frac{n(n+1)}{2}$ known reduced-form parameters and $n^2 + \frac{n(n+1)}{2}$ structural parameters that we want to know.
- This can be expressed as $n^2 + \frac{n(n+1)}{2}$ equations in $n^2 + \frac{n(n+1)}{2}$ unknowns, so we can get a unique solution.
- Example: Asserting that the reduced-form VAR is the structural model is the same as imposing the $2n^2$ a priori restrictions that $A = C = I$.

Recursive SVARs

- SVARs generally identify their shocks as coming from distinct independent sources and thus assume that they are uncorrelated.
- The error series in reduced-form VARs are usually correlated with each other. One way to view these correlations is that the reduced-form errors are combinations of a set of statistically independent structural errors.
- The most popular SVAR method is the recursive identification method.
- This method (used in the original Sims paper) uses simple regression techniques to construct a set of uncorrelated structural shocks directly from the reduced-form shocks.
- This method sets $A = I$ and constructs a C matrix so that the structural shocks will be uncorrelated.

The Cholesky Decomposition

- Start with a reduced-form VAR with three variables and errors e_{1t}, e_{2t}, e_{3t}.
- Take one of the variables and assert that this is the first structural shock, $\epsilon_{1t} = e_{1t}$.
- Then run the following two OLS regressions involving the reduced-form shocks

$$\begin{aligned} e_{2t} &= c_{21} e_{1t} + \epsilon_{2t} \\ e_{3t} &= c_{31} e_{1t} + c_{32} e_{2t} + \epsilon_{3t} \end{aligned}$$

- This gives us a matrix equation $Ge_t = \epsilon_t$.
- Inverting G gives us C so that $e_t = C\epsilon_t$. Identification done.
- Remember that error terms in OLS equations are uncorrelated with the right-hand-side variables in the regressions.
- Note now that, by construction, the ϵ_t shocks constructed in this way are uncorrelated with each other.

Interpreting the Cholesky Decomposition

- The method posits a sort of "causal chain" of shocks.
- The first shock affects all of the variables at time t. The second only affects two of them at time t, and the last shock only affects the last variable at time t.
- The reasoning usually relies on arguments such as "certain variables are sticky and don't respond immediately to some shocks." We will discuss examples next week.
- A serious drawback: The causal ordering is not unique. Any one of the VARs variables can be listed first, and any one can be listed last.
- This means there are $n! = (1)(2)(3)....(n)$ possible recursive orderings.
- Which one you like will depend on your own prior thinking about causation.

Another Way to Do Recursive VARs

- The idea of certain shocks having effects on only some variables at time t can be re-stated as some *variables* only having effects on some variables at time t.
- In our 3 equation example this method sets $C = I$ and directly estimates the A and B matrices using OLS:

$$\begin{aligned} y_{1t} &= b_{11}y_{1,t-1} + b_{12}y_{2,t-1} + b_{13}y_{3,t-1} + \epsilon_{1t} \\ y_{2t} &= b_{21}y_{1,t-1} + b_{22}y_{2,t-1} + b_{23}y_{3,t-1} - a_{21}y_{1t} + \epsilon_{2t} \\ y_{3t} &= b_{31}y_{1,t-1} + b_{32}y_{2,t-1} + b_{33}y_{3,t-1} - a_{31}y_{1t} - a_{32}y_{2t} + \epsilon_{3t} \end{aligned}$$

- See how the first shock affects all the variables while the last shock only affects the last variable.
- This method delivers shocks and impulse responses that are identical to the Cholesky decomposition.
- Shows that different combinations of A, B and C can deliver the same structural model.

Part IV

Estimating VARs

How to Estimate a VAR's Parameters?

- VARs consist of a set of linear equations, so OLS is an obvious technique for estimating the coefficients.
- However, it turns out that OLS estimates of a VAR's parameters are biased.
- Here we discuss a set of econometric issues relating to VARs. Specifically, we discuss

1. The nature of the bias in estimating VARs with OLS.
2. Methods for adjusting OLS estimates for their bias.
3. An asymptotic justification for OLS, stemming from the fact that they are Maximum Likelihood Estimators for VAR models when errors are normal.
4. A method for calculating standard errors for impulse response functions.
5. Problems due to have large amounts of parameters to estimate.
6. Bayesian estimation of VAR models as a solution to this problem.

OLS Estimates of VAR Models Are Biased

- Consider the AR(1) model

$$y_t = \rho y_{t-1} + \epsilon_t$$

- The OLS estimator for a sample of size T is

$$\begin{aligned}
\hat{\rho} &= \frac{\sum_{t=2}^{T} y_{t-1} y_t}{\sum_{t=2}^{T} y_{t-1}^2} \\
&= \rho + \frac{\sum_{t=2}^{T} y_{t-1} \epsilon_t}{\sum_{t=2}^{T} y_{t-1}^2} \\
&= \rho + \sum_{t=2}^{T} \left(\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} \right) \epsilon_t
\end{aligned}$$

- ϵ_t is independent of y_{t-1}, so $(y_{t-1}\epsilon_t) = 0$. However, ϵ_t is **not independent** of the sum $\sum_{t=2}^{T} y_{t-1}^2$. If ρ is positive, then a positive shock to ϵ_t raises current and future values of y_t, all of which are in the sum $\sum_{t=2}^{T} y_{t-1}^2$. This means there is a negative correlation between ϵ_t and $\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$, so $\hat{\rho} < \rho$.
- This argument generalises to VAR models: OLS estimates are biased.

The Bias of VAR Estimates

- Recall that for the AR(1) model, the OLS estimate can be written as

$$\hat{\rho} = \rho + \sum_{t=2}^{T} \left(\frac{y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} \right) \epsilon_t$$

- Bias stems from correlation of ϵ_t with $\sum_{t=2}^{T} y_{t-1}^2$.
- The size of the bias depends on two factors:
  1. The size of ρ: The bigger this is, the stronger the correlation of the shock with future values and thus the bigger the bias.
  2. The sample size T: The larger this is, the smaller the fraction of the observations sample that will be highly correlated with the shock and thus the smaller the bias.
- More generally, the bias in OLS estimates of VAR coefficients will be larger the higher are the "own lag" coefficients and the smaller the sample size.

A Bootstrap Bias Adjustment

- If OLS bias is likely to be a problem, one solution is to use "bootstrap methods".
- These use the estimated error terms to simulate the underlying sampling distribution of the OLS estimators when the data generating process is given by a VAR with the estimated parameters. These calculations can be used to apply an adjustment to the OLS bias.
- In practice, this can be done roughly as follows:
  1. Estimate the VAR $Z_t = AZ_{t-1} + \epsilon_t$ via OLS and save the errors $\hat{\epsilon}_t$.
  2. Randomly sample from these errors to create, for example, 10,000 new error series ϵ_t^* and simulated data series generated by the recursion $Z_t^* = \hat{A}^{OLS} Z_{t-1}^* + \epsilon_t^*$. (Need some starting assumption about Z_0.)
  3. Estimate a VAR model on the simulated data and save the 10,000 different sets of OLS estimate coefficients $\hat{A}^*$.
  4. Compute median values of each entry in $\hat{A}^*$ as $\bar{A}$ and compare this to $\hat{A}$ to get an estimate of the bias of the OLS estimates.
  5. Formulate new estimates $\hat{A}^{BOOT} = \hat{A}^{OLS} - \left(\bar{A} - \hat{A}^{OLS}\right)$
- See the paper by Killian on the webpage.

Maximum Likelihood Estimation

- Let θ be a potential set of parameters for a model and let f be a function such that, when the model parameters equal θ, the joint probability density function that generates the data is given by $f(y_1, y_2, ..., y_n|\theta)$.
- In other words, given a value of θ, $f(y_1, y_2, ..., y_n|\theta)$ describes the probability density of the sample $(y_1, y_2, ..., y_n)$ occurring.
- For a particular observed sample $(y_1, y_2, ..., y_n)$, we call $f(y_1, y_2, ..., y_n|\theta)$ the **likelihood** of this sample occurring if the true value of the parameters equalled θ.
- The maximum likelihood estimator (MLE) is the estimator θ^{MLE} that maximises the value of the likelihood function for the observed data.
- MLE estimates may be biased but it can be shown that they are consistent and asymptotically efficient, i.e. they have the lowest possible asymptotic variance of all consistent estimators.
- In general, MLEs cannot be obtained using analytical methods, so numerical methods are used to estimate the set of coefficeints that maximise the likelihood function.

MLE with Normal Errors

- Suppose a set of observations $(y_1, y_2, ..., y_n)$ were generated by a normal distribution with an unknown mean and standard deviation, μ and σ. Then the MLEs are the values of μ and σ that maximise the joint likelihood obtained by multiplying together the likelihood of each of the observations

$$f(y_1, y_2, ..., y_n|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_i - \mu)^2}{2\sigma^2}\right]$$

- Econometricians often work with the log of the likelihood function (this is a monotonic function so maximising the log of the likelihood produces the same estimator). In this example, the log-likelihood is

$$\log f(y_1, y_2, ..., y_n|\mu, \sigma) = -\frac{n}{2}\log 2\pi - n\log\sigma + \sum_{i=1}^{n}\left[\frac{-(y_i - \mu)^2}{2\sigma^2}\right]$$

- More generally, the log-likelihood function of n observations, $y_1, y_2, ..., y_n$, of a vectors of size k drawn from a $N(\mu, \Sigma)$ multivariate normal distribution is

$$\log f(y_1, y_2, ..., y_n|\mu, \Sigma) = -\frac{(kn)}{2}\log 2\pi - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{k=1}^{n}(y_i - \mu)'\Sigma^{-1}(y_i - \mu)$$

MLE of Time Series with Normal Errors

- Suppose $\epsilon_t \sim N\left(0, \sigma^2\right)$. Consider the AR(1) model

$$y_t = \rho y_{t-1} + \epsilon_t$$

- Figuring out the joint unconditional distribution of a series $y_1, y_2, ..., y_n$ is tricky but we can say $y_2 \sim N\left(\rho y_1, \sigma^2\right)$ and $y_3 \sim N\left(\rho y_2, \sigma^2\right)$ and so on.
- Let $\theta = (\rho, \sigma)$. Conditional on the first observation, we can write the joint distribution as

$$\begin{aligned} f\left(y_2, ..., y_n | \theta, y_1\right) &= f\left(y_n | \theta, y_{n-1}\right) f\left(y_{n-1} | \theta, y_{n-2}\right) ... f\left(y_2 | \theta, y_1\right) \\ &= \prod_{i=2}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\left(y_i - \rho y_{i-1}\right)^2}{2\sigma^2}\right] \end{aligned}$$

- The log-likelihood is

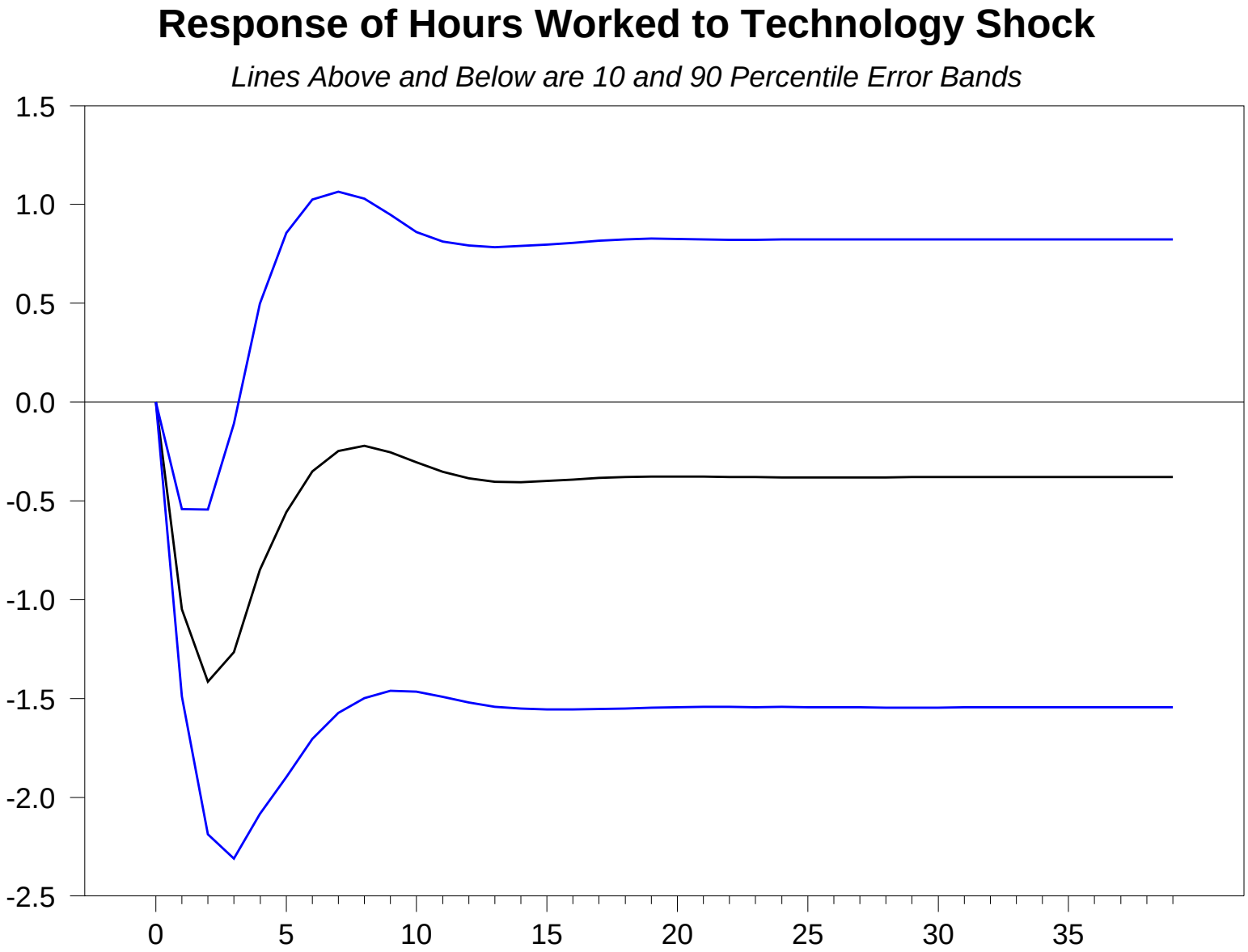
$$\log f\left(y_1, y_2, ..., y_n | \theta, y_1\right) = -\left(n-1\right)\left(\frac{\log 2\pi}{2} + \log\sigma\right) + \sum_{i=2}^{n}\left[\frac{-\left(y_i - \rho y_{i-1}\right)^2}{2\sigma^2}\right]$$

- Can easily show OLS provides the MLE for ρ. Generalises to VAR estimation.

Standard Error Bands for Impulse Response Functions

- When you estimate a VAR model such as $Z_t = AZ_{t-1} + \epsilon_t$ via OLS to obtain estimates $\hat{A}$, you can use the estimated coefficients in $\hat{A}$ to construct estimates of impulse responses $I, \hat{A}, \hat{A}^2, \hat{A}^3, ...$
- As with the underlying coefficient estimates, the estimated IRFs are just our best guesses. We would like to know how confident we can be about them. For example, how sure can we be that the response of a variable to a shock is always positive?
- For this reason, researchers often calculate confidence intervals for each point in an impulse response graph.
- Graphs in VAR papers thus usually show three lines for each impulse response function: The estimated set of impulse responses at each horizon and lines above and below showing the upper and lower ends of some confidence interval. See example on next page.
- For some reason, some researchers sometimes show plus or minus one standard deviation. However, if you want to be confident that responses have a particular sign, then 10th and 90th percentiles (or 5th and 95th) are a better idea.

Example of a Chart With IRF Error Bands



Bootstrapping Standard Errors for IRFs

- Analytical results can be derived to obtain asymptotic (i.e. large sample) distributions for impulse response functions. Unfortunately, these distributions are not very accurate representations of the actual distributions obtained in finite samples of the size used in most empirical work.
- Bootstrap methods are now commonly to derive the standard error bands for IRFs.
- In practice, this is done as follows:
  1. Estimate the VAR via OLS and save the errors $\hat{\epsilon}_t$.
  2. Randomly sample from these errors to create, for example, 10,000 simulated data series $Z_t^* = \hat{A}Z_{t-1}^* + \epsilon_t^*$.
  3. Estimate a VAR model on the simulated data and save the 10,000 different IRFs associated with these estimates.
  4. Calculate quantiles of the simulated IRFs, e.g. of the 10,000 estimates of the effect in period 2 on variable i of shock j.
  5. Use the n-th and $(100 - n)$-th quantiles of the simulated IRFs as confidence intervals.

A Problem: Lots of Parameters

- One problem with classical estimation of VAR systems is that there are lots of parameters to estimate.
- Estimating a Cholesky decomposition VAR with n variables with k lags involves direct estimation of $n^2k + \frac{n(n-1)}{2}$ parameters.
- For 3 variable VAR with one lag, this is already 12 parameters.
- Consider a 6 variable VAR with 6 lags. You've got $(36)(6) + (6)(5)/2 = 231$ coefficients.
- Because many of the coefficients are probably really zero or close to it, this can lead to a severe "over-fitting" problem that can result in poor-quality estimates and bad forecasts.
- This problem can lead researchers to limit the number of variables or number of lags used, perhaps resulting in mis-specification (leaving out important variables or missing important dynamics.)
- This can also lead to poor inference and bad forecasting performance.

Bayes's Law

- Bayes's Law is a well-known result from probability theory. It states that

$$\Pr(A \mid B) \propto \Pr(B \mid A) Pr(A)$$

- For example, suppose you have prior knowledge that A is a very unlikely event (e.g. an alien invasion). Then even if you observe something, call it B, that is likely to occur if A is true (e.g. a radio broadcast of an alien invasion), you should probably still place a pretty low weight on A being true.
- In the context of econometric estimation, we can think of this as relating to variables Z and parameters θ. When we write

$$\Pr(\theta = \theta^* \mid Z = D) \propto \Pr(Z = D \mid \theta = \theta^*) Pr(\theta = \theta^*)$$

we are calculating the probability that the a vector of parameters θ takes on a particular value, θ^* given the oberved data, D, as a function of two other probabilities: (i) the probability that $Z = D$ if it was the case that $\theta = \theta^*$ and (ii) the probability that $\theta = \theta^*$.

Bayesian Probability Density Functions

- Since coefficients and data in VARs are continuous, we need to write the Bayes relationship in form of probability density functions:

$$f_\theta\left(\theta^* \mid D\right) \propto f_Z\left(D \mid \theta^*\right) f_\theta\left(\theta^*\right)$$

- The function $f_Z\left(D \mid \theta^*\right)$ is the likelihood function—for each possible value of θ^*, it tells you the probability of a given dataset occurring if the true coefficients $\theta = \theta^*$.
- The likelihood functions can be calculated once you have made assumptions about the distributional form of the error process.
- Bayesian analysis specifies a "'prior distribution", $f_\theta(\theta^*)$, which summarises the researcher's pre-existing knowledge about the parameters θ.
- This is combined with the likelihood function to produce a "posterior distribution" $f_\theta\left(\theta^* \mid D\right)$ that specifies the probability of all possible coefficient values given both the observed data and the priors.
- Posterior distributions cannot generally be calculated analytically. Recent progress in computing power and numerical algorithms (via "Markov Chain Monte Carlo" algorithms — see the paper on the website for a discussion) have made Bayesian methods easier to implement.

Bayesian Estimation

- An obvious way to derive a "best estimator" from the posterior distribution is to calculate the mean of the distribution:

$$\hat{\theta} = \int_{-\infty}^{\infty} x f_{\theta}\left(x \mid D\right) dx$$

- You can show that this estimator is a weighted average of the "maximum likelihood estimator" and the mean of the prior distribution, where the weights depend on the covariances of the likelihood and prior functions: The more confidence the researcher specifies in the prior, the more weight will be placed on the prior mean in the estimator.
- With normally distributed errors, the maximum likelihood estimates are simply the OLS estimates, so Bayesian estimators of VAR coefficients are weighted averages of OLS coefficients and the mean of the prior distribution.

Bayesian VARs

- Typically, researchers specify priors so that coefficients are expected to get smaller for longer-lagged variables and that cross-equation coefficients (e.g. effect of lagged X_2 on X_1) are smaller than own-lag effects.
- A common approach is to specify the mean of the prior probability distribution for the first own-lag coefficient to be a large positive figure while specifying the prior mean for all other coefficients to be zero (e.g. the so-called "Minnesota prior").
- The researcher must also decide how confident they are in this prior e.g. how quickly the prior probabilities move towards zero as you move away from the prior mean. This is referred to as the "tightness" of the prior. The "tighter" the prior, the higher will be the weight on the prior in calculating the posterior Bayesian estimator.
- This sounds sort of complicated but in practice these days it is not. Various computer packages make it easy to specify priors of a particular form, with the tightness usually summarised by a couple of parameters.

Large Bayesian VARs

- Because of the problem with having to estimate so many parameters, most VAR papers have tended to use a small number of macroeconomic variables.
- However, economists in engaged in forecasting tend to look at a huge range of variables that are available at a monthly or weekly frequency.
- For instance, someone forecasting the US economy may look at employment, unemployment claims, personal income and consumption, industrial production, durable goods orders, figures on inventories, trade data, incoming fiscal data, sentiment surveys and indicators from financial markets.
- All could be useful for forecasting but a 20 variable monthly VAR with 12 lags could not be estimated by traditional methods. How can such a data set be cut down to a few variables without losing valuable information?
- Banbura, Giannone and Reichlin (2008) show that standard Bayesian VAR methods work very well for forecasting with VAR systems that incorporate large numbers of variables, provided that the tightness of the priors is increased as more variables are added.

Banbura-Giannone-Reichlin Evidence on Forecasting

Table 1: Relative MSFE, BVAR

		SMALL	CEE	MEDIUM	LARGE
h=1	EMPL	1.14	0.67	0.54	0.46
	CPI	0.89	0.52	0.50	0.50
	FFR	1.86	0.89	0.78	0.75
h=3	EMPL	0.95	0.65	0.51	0.38
	CPI	0.66	0.41	0.41	0.40
	FFR	1.77	1.07	0.95	0.94
h=6	EMPL	1.11	0.78	0.66	0.50
	CPI	0.64	0.41	0.40	0.40
	FFR	2.08	1.30	1.30	1.29
h=12	EMPL	1.02	1.21	0.86	0.78
	CPI	0.83	0.57	0.47	0.44
	FFR	2.59	1.71	1.48	1.93
λ		∞	0.262	0.108	0.035

Notes: Table reports MSFE relative to that from the benchmark model (random walk with drift) for employment (EMPL), CPI and federal funds rate (FFR) for different forecast horizons h and different models. SMALL, CEE, MEDIUM and LARGE refer to the VARs with 3, 7, 20 and 131 variables, respectively. λ is the shrinkage hyperparameter and is set so that the average in-sample fit for the three variable of interest is the same as in the SMALL model estimated by OLS. The evaluation period is 1971-2003.

Part V

Examples of VAR Studies

Examples of VAR Studies

We will look at three different examples of studies that use recursive VARs:

1. Lutz Killian (2009): Not All Oil Price Shocks are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. *American Economic Review*.
2. James Stock and Mark Watson (2001): Vector Autoregressions, *Journal of Economic Perspectives*. This paper examines the effect of monetary policy shocks.
3. Glenn Rudebusch (1998). "Do Measures of Monetary Policy in a Var Make Sense?" *International Economic Review* and a reply by Chris Sims.

Paper 1: Killian on Oil Shocks

- Oil shocks—large run-ups and subsequent declines in the price of oil—regularly receive a lot of attention.
- Many recent recessions were preceded by an increase in the price of oil. Why exactly this has occurred is not obvious: Oil usage is actually a relatively small input compared to GDP.
- Previous empirical work has generally asked the question "what are the effects of an oil price shock?"
- Killian asks "what is an oil price shock and are there different kinds of oil price shocks?"
- He uses VAR analysis to distinguish between shocks to oil prices due to global demand, shocks due to oil supply, and shocks due to speculation in the oil price market.
- Let's see how he does it.

Killian's Model

- Three variable monthly VAR in the growth rate of oil production, real global economic activity, and the real price of oil: $z_t = (\Delta prod_t, rea_t, rpo_t)'$.
- VAR structure is

$$A_0 z_t = \alpha + \sum_{i=1}^{24} A_i z_{t-i} + \epsilon_t$$

where ϵ_t are the structural shocks, and A_0 is lower-rectangular

$$A_0 = \begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix}$$

- Identifying assumptions:
  1. Oil production does not respond within the month to world demand and oil prices
  2. World demand is affected within the month by oil production, but not by oil prices.
  3. Oil prices respond immediately to oil production and world demand.

Interpreting the Structural Shocks

- If A_0 is lower-triangular, then so is A_0^{-1}.
- Reduced-form model is

$$z_t = A_0^{-1}\alpha + \sum_{i=1}^{24} A_0^{-1} A_i z_{t-i} + A_0^{-1}\epsilon_t$$

- Reduced-form shocks e_t related to structural shocks as $e_t = A_0^{-1}\epsilon_t$:

$$\begin{pmatrix} e_t^{\Delta prod} \\ e_t^{rea} \\ e_t^{rpo} \end{pmatrix} = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \epsilon_t^{\Delta prod} \\ \epsilon_t^{rea} \\ \epsilon_t^{rpo} \end{pmatrix}$$

- The oil production reduced-form shock is a structural shock; the reduced-form economic activity shock combines the structural oil shock and the structural activity shock; the reduced-form oil price shock combines all three structural shocks.

Checking the Identification Restrictions

Relative to the general model

$$AY_t = BY_{t-1} + C\epsilon_t$$

where are our $2n^2 = 18$ identifying restrictions?

1. We set $C = I$ instead assuming contemporaneous interactions between variables: 9 restrictions.
2. Lower-diagonal assumption on A_0: 3 zero restrictions.
3. Unit coefficient normalization on diagonal of A_0: 3 restrictions.
4. Orthogonal structural shocks: 3 off-diagonal elements of Σ are zero.

Decomposing the Variables

- In addition to the standard impulse response analysis, Killian shows how the real price of oil can be decomposed into components related to these three shocks. How did he do this?
- Recall the VMA representation:

$$Y_t = \epsilon_t + A\epsilon_{t-1} + A^2\epsilon_{t-2} + A^3\epsilon_{t-3} + ... + A^t\epsilon_0$$

- One can do this calculation three times, each time with only one type of shock "turned on" and the other set to zero. Adding these up, one will get the realized values of Y_t.
- Alternatively, one can do a dynamic simulation of the model

$$Y_t = AY_{t-1} + \epsilon_t$$

in each case letting the ϵ_t represent one of the realized historical shocks with the others set to zero.

Some of Killian's Findings

1. Despite getting a lot of attention, shocks to oil supply have limited effects on oil prices and have been of negligible importance in driving oil prices over time.
2. Both global demand and speculative oil price shocks can have significant effects on oil prices, but speculative oil price shocks have limited effects on global economic activity.
3. Speculative oil-market shocks have accounted for most of the month-to-month movements in oil prices.
4. But the steady increase in oil prices from 2000 onwards was almost solely due to strong global demand.
5. Main Lesson: How the economy reacts to an "oil price shock" will depend on the origins of that shock.
6. Helps to explain why the world economy survived a long period of increasing oil prices in the 2000s without going into recession. (And when it did go into recession, it had little to do with oil prices.)

Paper 2: A Monetary Policy VAR

- Stock and Watson's 2001 *JEP* paper is a very useful introduction to VAR methods.
- The paper contains an important application: What are the effects of monetary policy shocks?
- You can think of these VARs as useful in two ways:
  1. From a scientific perspective: Monetary policy co-moves with lots of other macro variables. Only by identifying the structural or exogenous shocks to policy can we discover its true effects.
  2. From a policy perspective, helps to answer the question "if I choose to raise interest rates by an extra quarter point today, what is likely to happen over the next year to inflation and output relative to the case where I keep rates unchanged?" Essentially, this is a question about impulse responses.

Stock and Watson's VAR

- Monthly data on inflation (π_t), the unemployment rate (u_t) and the federal funds rate (i_t).
- Posits a lower-triangular causal chain of the form

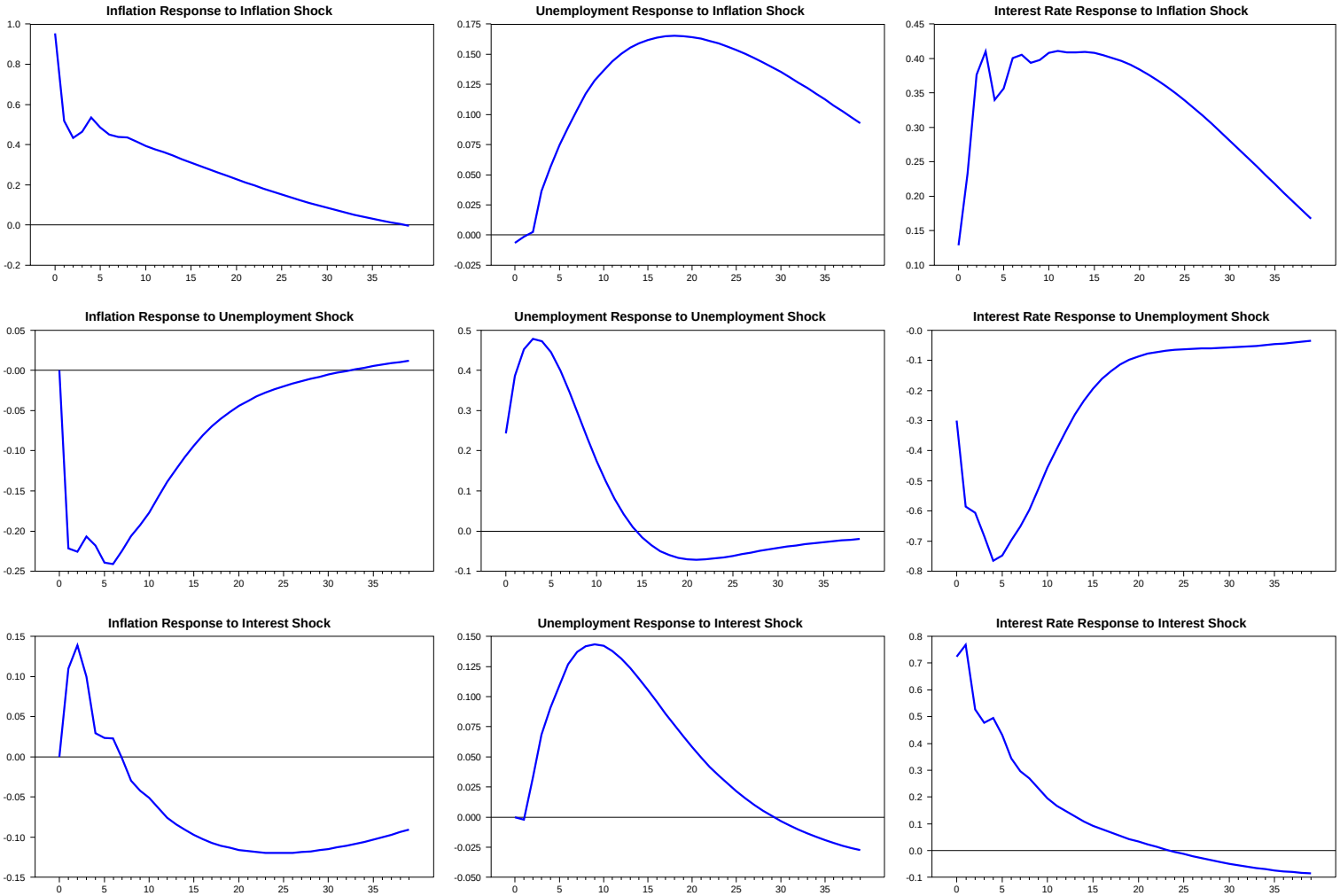
$$AZ_t = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \pi_t \\ u_t \\ i_t \end{pmatrix} = BZ_{t-1} + \epsilon_t$$

- Identifying assumptions
  1. Inflation depends only on lagged values of the other variables (perhaps motivated by the idea of sticky prices.)
  2. Unemployment depends on contemporaneous inflation but not the funds rate.
  3. The funds rate depends on both contemporaneous inflation and unemployment. (Fed using its knowledge about the current state of the economy when it is setting interest rates).
- I have used quarterly data to re-do Stock and Watson's VAR analysis and have reported the Impulse Response Functions on the next page.

IRFs From Recursive VAR, First Identification

Figure 1: IRFs From Recursive VAR, First Identification

Order is Inflation, Unemployment, Interest Rate



A Puzzle?

- These results generally make sense.
- In particular, a positive interest rate shock raises the unemployment rate, and over time it reduces the inflation rate.
- The short-run response of the inflation rate, however, is a bit puzzling: The interest rate increase seems to raise the inflation rate for a few periods.
- This "price puzzle" result has been obtained in a number of VAR studies. It provides a good illustration of the potential limitations of VAR analysis.
- Some think the explanation is that the Fed is acting on information not captured in the VAR (for example, information about commodity prices) and that this information may provide signals of future inflationary pressures.
- Thus, interest rate increases can tend to occur just before an increase in inflation. The VAR may be capturing this pattern and confusing causation and correlation.
- Indeed, adding commodity prices to the VAR has often been found to eliminate the "price puzzle." But is this improving the specification is it data mining to find the result we want?

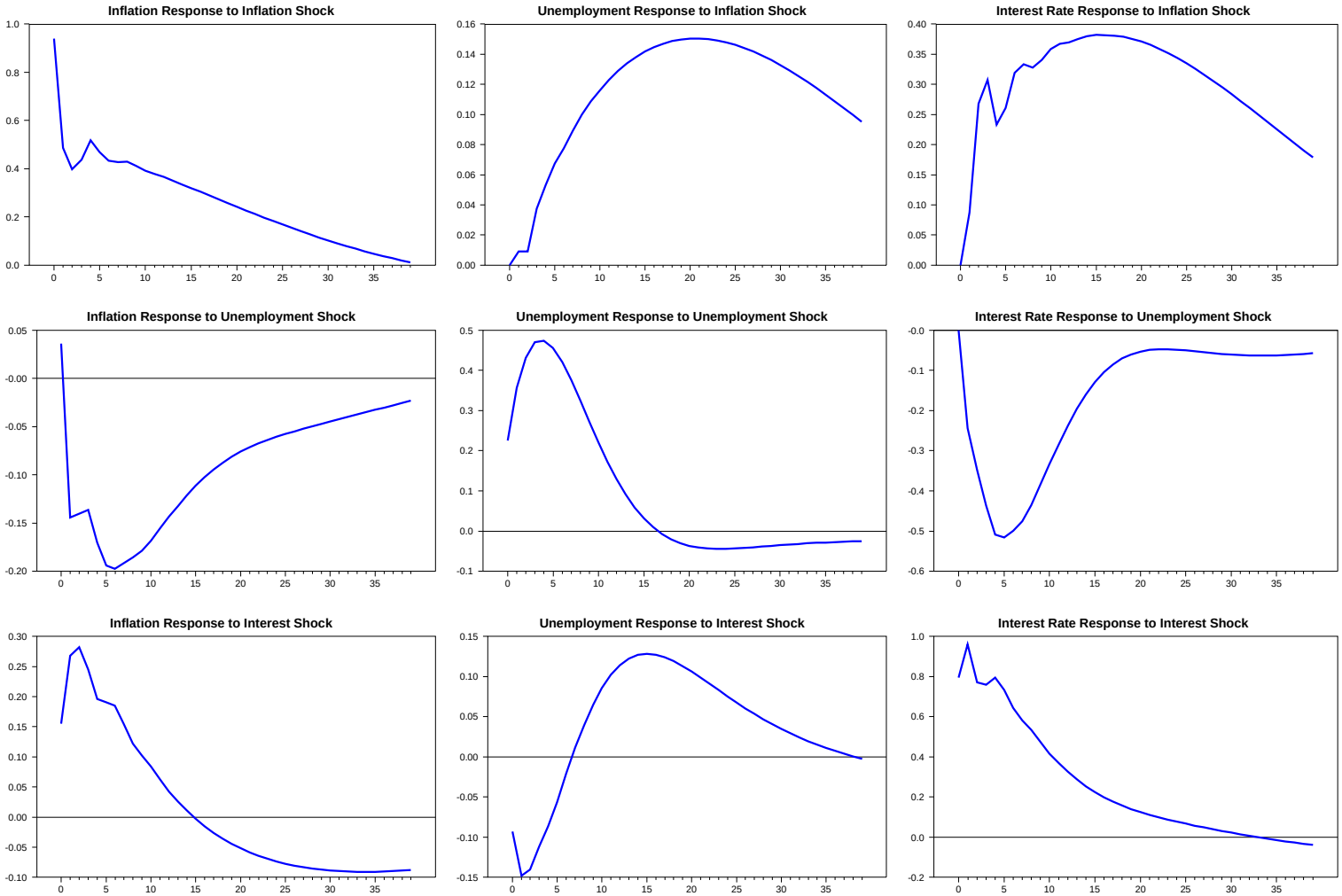
A Second Identification

- The results from the Stock-Watson identification generally make sense but you could imagine other possible identifications.
- The next slide shows results from an alternative ordering with interest rates first, unemployment second, and inflation last.
- You could argue for this case on the grounds that the Fed can only respond to the economy with a lag because it takes a while to receive data about the current state of the economy (so they are reacting to lagged information) but that inflation should be able to respond immediately to economic events.
- This may sound reasonable enough but the results from this identification don't make as much sense: The interest rate shock raises inflation now for almost four years and unemployment drops for a while after the increase in interest rates.
- Which identification a researcher settles on may depend on how "sensible" they believe results are. This can cause problems.
- RATS code for both identifications are available on the website.

IRFs From Recursive VAR, Second Identification

Figure 2: IRFs From Recursive VAR, First Identification

Order is Interest Rate, Unemployment, Inflation



Paper 3: The Rudebusch-Sims Debate

- A lot of interesting material on monetary policy VARs can be found in a 1998 exchange between Glenn Rudebusch and Chris Sims.
- Rudebusch's paper contains a strong critique of VARs used to assess monetary policy. Among his points:
  1. The VARs ignore changes over time in the formulation of monetary policy.
  2. They use final published data instead of the preliminary estimates the Fed has available when it makes decisions.
  3. They greatly underestimate the information available to the Fed when it takes decisions.
  4. They incorporate long lags but he argues it is not credible that the Fed responds to information from over a year prior to taking a decision.
  5. The monetary policy shocks don't look anything like the surprise element of monetary policy decisions obtained from looking at financial contracts like fed funds futures.
  6. Models with very different monetary policy shocks report similar IRFs, suggesting that perhaps the models have been data-mined to give these answers.

Final Example: The Rudebusch-Sims Debate

- Sims responded in detail to Rudebusch's critiques. Some of his points were as follows.
  1. VAR models may differ in their shocks but agree on their effects. For example, one model may include more variables in a supply equation than another so its supply shocks are, by construction, smaller in size but both models could still capture roughly the same effect of a shock to supply.
  2. Financial market "surprises" are not necessarily the best measures of the exogenous element of monetary policy. A Fed governor could give a speech the week before an FOMC meeting indicating that the Fed is going to raise rates (even if this isn't predicted by inflation or GDP or other standard VAR variables.) When this rate increase happens, we'd like to know what effect it has even though, on the day, it is not a surprise for financial markets.
  3. Issues like "time invariance, linearity, and variable selection are universal in macroeconomic modeling" and are not special to VARs.
- The exchange is well worth reading in full as it sheds light on a lot of the issues that economists need to think about when doing VAR analysis.

Part VI

VARs With Long-Run Restrictions

An Alternative Approach: Long-Run Restrictions

- The identifying assumptions in the recursive VAR approach require knowledge of how certain variables react in an instantaneous way to certain shocks.
- Sometimes, because certain variables are "sluggish" or because information about some variables is only available with a lag, we can be pretty confident about these restrictions. But often they are pure guesswork.
- And economic theory gives very little guidance.
- In fact, economic theory usually tells us a lot more about what will happen in the longer-run, rather than exactly what will happen today.
- For instance, theory tells us that whatever positive aggregate demand shocks do in the short-run, in the long-run they have no effect on output and a positive effect on the price level.
- This suggests an alternative approach: Use these theoretically-inspired long-run restrictions to identify shocks and impulse responses.
- I will explain one of these methods over the next four slides. The general idea is more important than the technical details.

Information in the Reduced-Form Covariance Matrix

- Consider the VAR model

$$Z_t = BZ_{t-1} + C\epsilon_t$$

where the covariance matrix of the structural shocks is

$$E(\epsilon_t \epsilon_t') = \begin{pmatrix} E(\epsilon_1^2) & E(\epsilon_1 \epsilon_2) \\ E(\epsilon_1 \epsilon_2) & E(\epsilon_2^2) \end{pmatrix} = I$$

so the structural shocks are uncorrelated and have unit variance (this is just a harmless normalization).

- Note that the covariance matrix of the observed reduced-form errors is

$$\Sigma = E(e_t e_t') = E\{(C\epsilon_t)(C\epsilon_t)'\} = CE(\epsilon_t \epsilon_t')C' = CC'$$

- Thus, the observed covariance structure of the reduced-form shocks tells us something about how they are related to the uncorrelated, unit-variance, structural shocks.

Calculating Long-Run Effects in an SVAR

- Suppose $Z_t = (\Delta y_t, \Delta x_t)'$
- Then the long-run effect of the shock on y_t is the *sum* of its effects of Δy_t, Δy_{t+1}, Δy_{t+1} and so on.
- The long-run effect is the sum of the impulse responses.
- The impulse responses for the model $Z_t = BZ_{t-1} + C\epsilon_t$ are
  1. C in impact period.
  2. BC after one period.
  3. B^2C after two periods, ... B^nC after n periods.
- Long-run level effects are $D = \left(I + B + B^2 + B^3 + ...\right) C$.
- If eigenvalues of B are inside unit circle then $I + B + B^2 + B^3 + ... = (I - B)^{-1}$.
- This is the matrix equivalent of the multiplier formula $1 + c + c^2 + c^3 + ... = \frac{1}{1-c}$.
- So the long-run responses are $D = (I - B)^{-1} C$.

The Blanchard-Quah Method: I

- Now note that $DD' = (I - B)^{-1} CC' \left((I - B)^{-1}\right)'$
- But two slides ago, we established that $CC' = \Sigma$, the covariance matrix of the reduced-form shocks, which can be estimated.
- So $DD' = (I - B)^{-1} \Sigma \left((I - B)^{-1}\right)'$
- Now make a restriction about the long-run effects described in D: Assume that D is lower triangular: Only the first shock has a long-run effect on the first variable, and only the first and second shocks have long-run effects on the second variable and so on.
- In the two variable case, this is just

$$D = \begin{pmatrix} d_{11} & 0 \\ d_{21} & d_{22} \end{pmatrix}$$

The Blanchard-Quah Method: II

- $DD' = (I - B)^{-1} \Sigma \left((I - B)^{-1} \right)'$ is a symmetric matrix (the i, j entry is identical to the j, i entry)
- All symmetric matrices have a unique lower-diagonal matrix D so that DD' equals the symmetric matrix. This is known as the *Cholesky factor* of the symmetric matrix.
- D can be calculated directly in RATS and other software as the Cholesky factor of the known matrix $(I - B)^{-1} \Sigma \left((I - B)^{-1} \right)'$.
- Now remember that $D = (I - B)^{-1} C$.
- So the crucial matrix C defining the structural shocks can then be calculated as

$$C = (I - B)D$$

- Now, we can calculate the impulse response functions to the structural shocks.

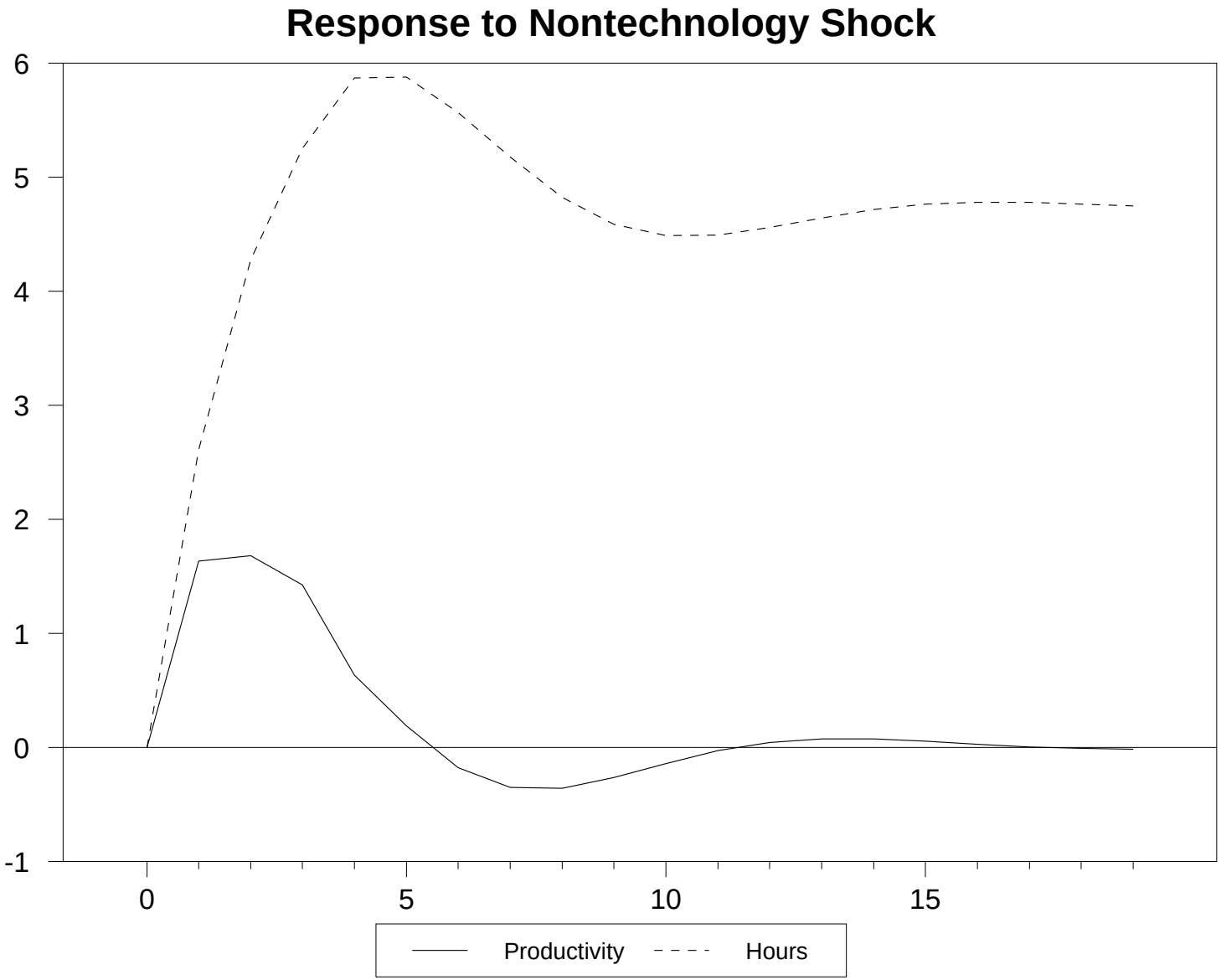
Blanchard-Quah: Identifying Supply and Demand Shocks

- Blanchard-Quah (1989) used a two-variable VAR in the log-difference in GDP Δy_t and the unemployment rate U_t.
- Unemployment was entered in *levels* form. Because the VAR is estimated to be stationary (eigenvalues inside unit circle) both structural shocks have zero long-run effect on the unemployment rate.
- The lower-diagonal assumption thus meant that of the two structural shocks only one of them could have a long-run effect on the level of output. BQ labelled this the "supply" shock while the shock that has no effect on long-run output was labelled the "demand" shock.
- The relative importance of supply versus demand shocks in determining output is a long-running theme in macroeconomics. Keynesians emphasize the importance of demand shocks while more classically-oriented economists, such as advocates of the Real Business Cycle approach, see supply shocks as being more important.
- BQ's results implied that demand shocks were responsible for the vast majority of short-run fluctuations.

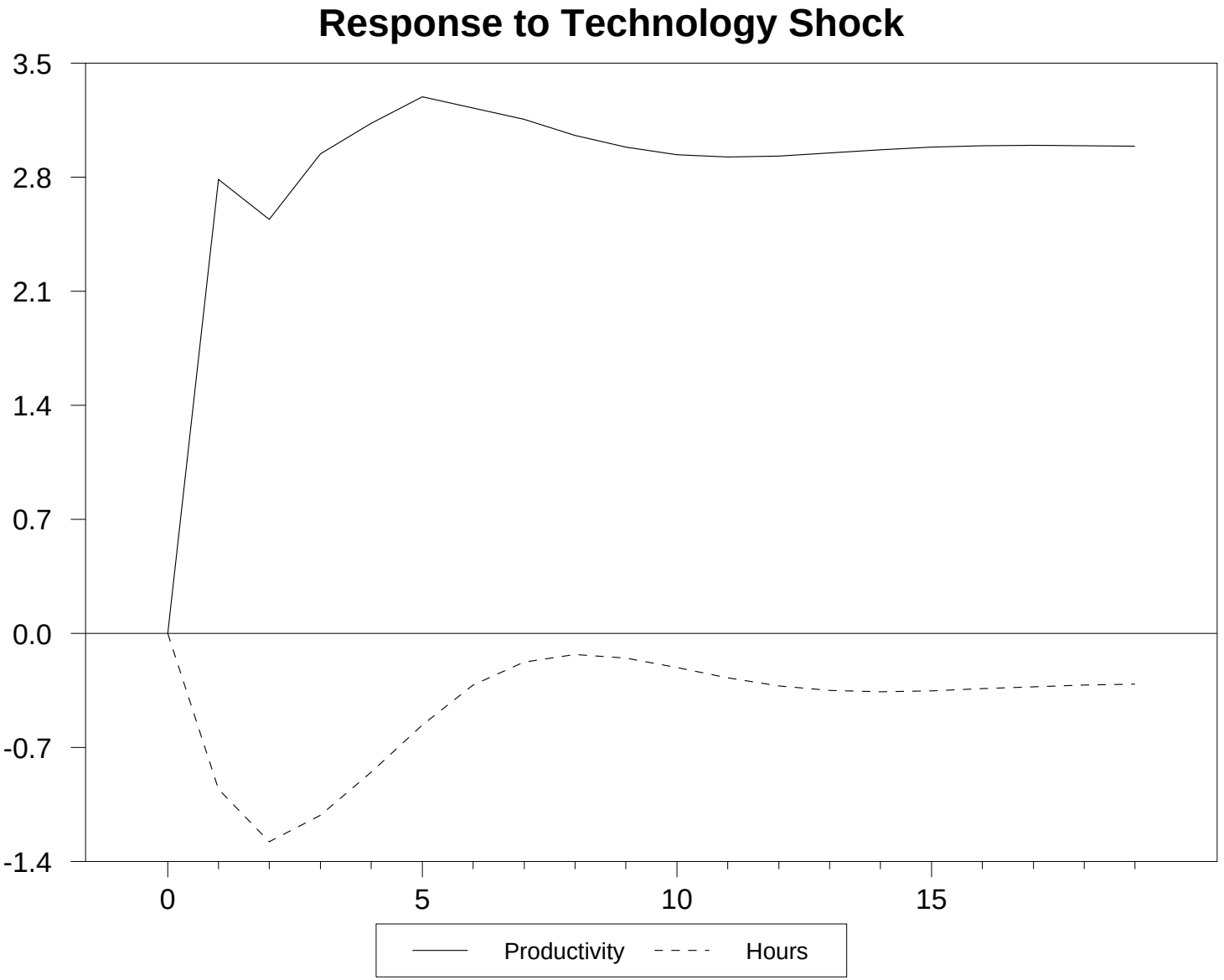
Galí (1999): Technology Shocks and Hours Worked

- BQ's formulation is a little bit restrictive: The assumption that neither supply or demand shocks can change the unemployment rate in the long-run may not be correct.
- Galí's paper applied a similar analysis to BQ, but for a formulation that moved a bit closer to the debate about RBC models and their predictions for the labour market,
- RBC models assume technology shocks drive the business cycle and explain why hours worked are higher in booms than in recessions: Better to work when you are productive than unproductive.
- Galí VAR: Log-difference of output per hour worked (labour productivity), Δz_t and the log-difference of hours worked, Δn_t.
- The lower-diagonal assumption about long-run responses now means that the supply shock (now called the "technology" shock) can affect productivity in the long-run, while the non-techology shock cannot.
- The model lets the data dictate the long-run effects of technology and non-technology shocks on hours worked.

Replicating Galí (1999)



Replicating Galí (1999)



Interpreting Galí's Results

- Nontechnology shocks cause both output and productivity to rise in the short-run.
- Evidence that short-run cyclical movements in productivity are not just due to technology shocks.
- Explanation? Costs of adjusting labour input. Rather than hire new labour in a boom, temporarily work existing labour a bit harder. In recession, employed labour is more likely to be under-utilized.
- Also shows nontechnology shocks raising labour input. Interpretation unclear. Some have specified this VAR with some stationary transformation of labour input, so shocks have no long-run effects.
- Technology shocks cause productivity to go up but hours to go down.
- Interpretation: Short-run output is demand-driven not supply-driven. More efficiency means that demanded output can be supplied with less labour.
- Bad news for technology-driven stories of the business cycle such as RBCs. These results have generated some controversy but I believe they are correct. See my own small contribution to this debate which is posted on the website.