Introduction to Time Series Analysis:

Laura Mayoral[1]

[1]Institute for Economic Analysis and Barcelona School of Economics

DG-ECFIN, Nov 2021

Handout 2

Estimation of univariate stationary time series

Introduction

Goal:

Given a univariate stationary process $\{y_t\}$, we want to use data to

1. estimate moments: mean, variances, correlations, long-run variance...
2. An application of the above: regression with serially correlated errors; We will compute standard errors robust to serial correlation
3. model serial correlation: fit an ARMA process

Roadmap

1. Introduction
2. Estimation of moments of serially correlated processes. Limit theorems for serially correlated processes
3. OLS regression with serially correlated errors
4. Estimation of ARMA models

II. Estimating moments: The LLN and the CLT for serially correlated processes

Consider the problem of estimating the mean of a stationary univariate process, $\{y_t\}$.

- Since $\{y_t\}$ is stationary, the mean μ is constant for all t. (But think of what happens outside of stationarity!)
- An obvious estimator for μ is $\bar{y} = \sum_{t=1}^{T} y_t$, the sample mean of $\{y_t\}$
- Is $\bar{y}$ consistent? $\Rightarrow$ We need a a law of large numbers (LLN) for serially correlated processes!
- Inference on $\bar{y}$: We need a a central limit theorem (CLT) for serially correlated processes!

LLN for i.i.d. sequences

- The LLN is a central result in statistics
- It's typically stated for i.i.d. sequences like this:

Theorem

(Weak law of large numbers for iid sequences) If $\{y_t\}$ is an $i.i.d$ sequence of random variables with finite mean μ then

$$\bar{y}_T = T^{-1} \sum_{t=1}^{T} y_t \xrightarrow{p} \mu$$

- But our data is no longer $i.i.d$. so we need to extend this theorem. Can we?

Limit theorems for stationary processes

- Yes, we can! This is the L.L.N. for stationary and ergodic processes

Theorem

If $\{y_t\}$ *is stationary with mean* μ *and autocovariance function* $\gamma(.)$, *then*
i) $E(\bar{y}_T) = \mu$
ii) If $\gamma(T) \to 0$ *as* $T \to \infty$, $Var(\bar{y}_T) = E(\bar{y}_T - \mu)^2 \to 0$.
iii) If $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, $TE(\bar{y}_T - \mu)^2 \to \sum_{h=-\infty}^{\infty} \gamma(h)$.
Proof. See Brockell and Davis, (1991), p. 219.

LLN for dependent processes: comments

- Condition (ii) introduces the ergodicity condition.
- A stationary process is ergodic for the mean provided $\gamma(T) \to 0$ as $T \to \infty$
- Ergodicity: (Loosely speaking) it limits the dependence between the conditions.
- $i)$ and $ii)$ together imply that $\bar{y}_T$ converges –in mean square– to μ.
- Bottom line: for the LLN to hold y_t needs to be stationary AND ergodic for the mean (i.e., $\gamma(T) \to 0$).
- Good news: The type of ARMA processes we're considering are stationary and ergodic, thus a LLN holds!

LLN for dependent processes: comments II

- Condition (iii) is going to be used while stating the central limit theorem for dependent processes.
- Recall: CLT states that $T^{1/2}(\bar{y}_T - \mu)$ converges in distribution to a Normal variable with zero mean and variance given by the limit of the variance of $T^{1/2}\bar{y}_T$
- Condition (iii) is about the limit of the variance of $T^{1/2}\bar{y}_T$:
  - Provided: $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then the variance of $T^{1/2}\bar{y}_T$ converges to the long run variance:

$$\sum_{h=-\infty}^{\infty} \gamma(h).$$

  - $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ is also an ergodicity condition: Ergodicity for second moments

Central limit theorem for dependent processes

Theorem

(Central limit theorem for dependent processes) Let $\{y_t\}$ *be a stationary sequence given by* $y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ *where* $\{\varepsilon_t\}$ *is an iid(0,*σ^2*) sequence of random variables and* $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$ *and* $\sum_{j=0}^{\infty} \psi_j \neq 0$ *then*

$$\sqrt{T}(\bar{y}_T - \mu) \xrightarrow{d} N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right).$$

Proof. See Brockell and Davis, (1991), Section 7.3.

Summary: LLN and CLT for dependent processes

If y_t is serially correlated we need to introduce conditions so that the LLN and the CLT keep holding.

- Consistent estimator of μ (LLN): We need stationarity and ergodicity for the mean, (i.e., $\gamma(T) \to 0$).
- Distribution of the sample mean (CLT): We need stationarity and ergodicity for second moments $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$.

An application: OLS regression with autocorrelated errors

We will apply the theorems above to an usual problem:

- Consider a standard linear regression framework, estimated by OLS
- What happens when the residuals are serially correlated?
- Framework: linear regression, all other assumptions are maintained, in particular x_t (indep. variable) is exogeneous

OLS with autocorrelated errors

- Linear regression framework:

$$y_t = x_t'\delta_0 + \epsilon_t$$

- Assumptions:

  1. $E(X_t\epsilon_t) = 0$ (Exogeneity)
  2. ϵ_t might be autocorrelated

- OLS gives:

$$\hat{\delta} - \delta_0 = \sum_{t=1}^{T}(x_t x_t')^{-1}\sum_{t=1}^{T} x_t\epsilon_t$$

OLS with autocorrelated errors, II

1. In the "standard" framework: $g_t = x_t \epsilon_t$ is non-autocorrelated (it's typically assumed to be an IID or a MDS (=martingale difference sequence).
2. Now, g_t can be autocorrelated
3. Notice that the second term in the OLS formula, $\sum_{t=1}^{T} x_t \epsilon_t$, is the sample mean of g_t
4. Notice that: $E(g_t) = 0$ (by the law of iterated expectations.

OLS with autocorrelated errors

- Assume that $\{g_t\} = \{x_t \varepsilon_t\}$ is a stationary and ergodic (for second moment) process, given by

$$g_t = \Psi(L)\eta_t, \eta \sim i.i.d.(0, \Sigma)$$

- Then, by the LLN for stationary and ergodic processes:

$$\frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t \xrightarrow{p} 0$$

- and by the CLT for stationary and ergodic processes:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t \varepsilon_t \xrightarrow{d} N(0, LRV)$$

LRV: long run variance

Asymptotics with autocorrelated errors

Consistency:

$$\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t \xrightarrow{p} \boldsymbol{\Sigma}_{xx}^{-1} \times \mathbf{0} = \mathbf{0}$$

Asymptotic Distribution:

$$\sqrt{T} \left(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 \right) = \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{x}_t \varepsilon_t$$

$$\xrightarrow{d} \boldsymbol{\Sigma}_{xx}^{-1} \times N(\mathbf{0}, \mathsf{LRV}) \equiv N(\mathbf{0}, \boldsymbol{\Sigma}_{xx}^{-1} \mathsf{LRV} \boldsymbol{\Sigma}_{xx}^{-1})$$

Asymptotics with autocorrelated errors

Summarizing:

Provided all other assumptions are maintained:

- (Provided the x's are still exogenous!) the OLS estimator is still consistent
- However, the standard errors need to be adjusted: we need to compute HAC standard errors! (=heteroskedasticity- and autocorrelation-consistent std errors)

Computation of HAC standard errors

To compute correct standard errors, we need an estimator of the LRV.

Two types of estimators:

- Parametric: we need a model for g_t and from this model we estimate the covariances. For instance, $g_t \sim AR(1)$ and then we employ the formula of autocovariances for AR(1) processes
- Non-parametric: we estimate the covariances without imposing a parametric model. Newey-West estimator is the most popular one.

Computation of HAC standard errors: Newey-West estimator

- Recall that the LRV of $\bar{g}_t$ is given by

$$\sum_{h=-\infty}^{\infty} \gamma(h).$$

- Newey-West estimator: a weighted average of the first q covariances
- q: bandwith parameter (you need to choose a value)
- Many software programs use this rule:

$$q(T) = int(4(\frac{T}{100})^{1/4})$$

- "weights": Bartlett Kernel

Problems that may arise when errors are autocorrelated

- Are the x's still exogenous?
- Serial correlation in the residuals can make the regressors endogenous!
- Example:

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

- if ε_t is white noise $\rightarrow$, y_{t-1} is "predetermined" (exogenous, uncorrelated with ε_t)
- But if ε_t is autocorrelated, then y_{t-1} can become endogenous.

An example: ε_t is AR(1)

$$\varepsilon_t = \beta \varepsilon_{t-1} + \eta_t$$

- y_{t-1} is endogeneous:

$$corr(y_{t-1}, \varepsilon_t) = corr(\alpha y_{t-2} + \varepsilon_{t-1}, \varepsilon_t) \neq 0$$

- Thus, the OLS estimator is NOT consistent anymore.
- In fact, the model is misspecified. Notice that

$$\varepsilon_t = \frac{\eta_t}{1 - \beta L}$$

- Which implies:

Summarizing I: running regressions serial correlation in the residuals

- When you run regressions with time series variables, always suspect serial correlation in the residuals
- What to do? test for serial correlation in the residuals
- If serial correlation is found:
- Provided regressors are exogenous: OLS is consistent
- But, standard errors need to be adjusted: HAC standard errors

Summarizing II: running regressions with serial correlation in the residuals

- Be careful with lagged dependent variables in the presence of serial correlation in the residuals.
- OLS can become inconsistent
- Advice: Try to specify the model in such a way that the residuals look uncorrelated (you can test for this)
- Serial correlation in the residuals doesn't always mean than lagged dependent variables will become endogenous (see Wooldridge (Introduction to Econometrics), Chapter 12 for an example).

Summarizing III: HAC with other estimation approaches

- The analysis above only considers OLS estimation
- However, similar conclusions are reached when other estimation approaches are used: for instance, same if GMM is employed instead.
- For a good coverage of regression with autocorrelated errors in a general setup, see Hayashi, Chapter 6.

Stata Example

You can find an example and Stata codes here: Lecture2.do

(http://mayoral.iae-csic.org/timeseries2021/Lecture2.do)

Estimation of ARMA(p,q) processes

Estimation of ARMA(p,q) processes

In estimating ARMA processes, 3 STEPS are necessary:

1. Identification: select values for p and q.
2. Estimation: In general, Maximum likelihood. Exception: if process is AR(p), OLS can be employed.
3. Diagnostic checking: check the accuracy of the proposed model by analyzing the residuals of the estimated model.

We will start by considering STEP 2: estimation

STEP 2: Estimation of ARMA(p,q) processes

We will begin by considering Step 2: Estimation

- For now, assume (p, q) are known.
- Estimation: we will consider first estimation of AR processes (it's simpler) and then we will move to estimation of general ARMA(p,q) processes.

Estimation of stationary AR(p) processes

Two cases:

- X_t follows an AR(p) $(= ARMA\,(p, 0))$, for some finite p.
- X_t follows an AR(∞) process (notice that this includes any invertible stationary process!).

Estimation of stationary AR(p) for finite (and known) p

- Let $\{X_t\}$ be a stationary AR(p) process with $p < \infty$, *i.e.*,

$$X_t = c + \phi_0 + \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \varepsilon_t,$$

  where $\{\varepsilon_t\}$ is a zero-mean innovation (either w.n., m.d.s. or $i.i.d.$) with variance σ^2.

- Estimation: OLS.

- The resulting estimator has the usual properties: $\sqrt{T}-$consistent and asymptotically normally distributed.

Estimation of AR(p) for finite (and known) p

- Stationarity is a key assumption. (Remember that AR processes might not be stationary).
- To see its importance, consider for instance the asymptotic distribution of $\hat{\phi}$ in the AR(1) case,

Estimation of stationary AR(p) for finite (and known) p

- Model: $y_t = \phi y_{t-1} + \epsilon_t$
- Notice: $corr(y_{t-1}, \epsilon_t)$ and ϵ_t is white noise: standard OLS assumptions
- It can be shown that the distribution of the OLS estimator of ϕ is given by

$$T^{1/2}\left(\hat{\phi}_{ols} - \phi\right) \xrightarrow{d} N\left(0, \frac{\sigma^2}{var(X_t)}\right) = N\left(0, \left(1 - \phi^2\right)\right).$$

- This distribution is only valid under stationarity, as we know (CLT not valid otherwise!)
- Notice that if $\phi = 1$, variance is zero!.

Estimation of AR(∞) processes

- We know that any invertible ARMA process can be written as an AR(∞) process.
- Is it possible to estimate an AR(∞) process (that involves an infinite number of parameters!) with a finite sample?
- Answer: yes! Berk (1974) showed that under some assumptions one can obtain consistent estimates for the AR coefficients when the underlying process is AR(∞).

Estimation of AR(∞) processes, II

- Consider a stationary and invertible process X_t given by

$$X_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

where $\sum_{i=0}^{\infty} |\psi_i| < \infty$ and $\{\varepsilon_t\}$ is an $\left(0, \sigma^2\right)$ $i.i.d.$ sequence.

- Since X_t is invertible, it can be written as

$$X_t = \sum_{i=1}^{\infty} \phi_i X_{t-i} + \varepsilon_t.$$

Estimation of AR(∞) processes, III

- Berk showed that if an AR(k) process is fitted to X_t, where k **tends to infinity** at T, the sample size, tends to ∞, OLS delivers consistent and asymptotically normally distributed estimators of the k coefficients introduced in the regression.

- Key aspect: k has to be chosen carefully in order to get estimators with the properties above.

- Berk (1994) showed that k has to verify two conditions for the OLS estimators to have good properties.

- These conditions are quite theoretical, they don't give us much information about how to choose k in real applications!.

  - an upper bound condition: $k^3/T \to 0$ (that says that k should not increase too quickly),
  - and a lower bound one, $\mathrm{T}^{1/2}\sum_{j=k+1}^{\infty}\psi_i \to c \neq 0$ (that says that k must not increase too slowly).

- How to choose k in practice?

AR(∞) estimation: How to choose k in practice?

- Standard information criteria (AIC, BIC) tend to choose values of k that are too small (do not satisfy the lower bound condition), (Ng and Perron, 1995). In these cases, estimates are consistent but not asymptotically normal.

- The **General-to-Specific** model selection technique verifies Berk's conditions. Then, the resulting OLS estimators are consistent and asymptotically normal when k is choosing using that criterion (Kuersteiner, 2005).

Estimation of ARMA(p,q) process: Maximum likelihood estimation

- Let $\{X_t\}$ be an ARMA(p,q) process

$$\phi_p(L) X_t = \alpha + \theta_q(L) \varepsilon_t$$

and let $\delta = \left(\phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \alpha, \sigma^2\right)'$ be the vector containing all the unknown parameters.

- Suppose that we have observed a sample of size T: (x_1, x_2, ...).
- Since the errors are not observed, this model cannot be estimated by OLS.
- Alternative: maximum likelihood.

Maximum likelihood estimation

- The ML approach will amount to calculate the joint density of $(X_T, ..., X_1)$

$$f_{X_T,\ X_{T-1},...,X_1}\left(x_T,\ x_{T-1},...,x_1;\delta\right) \tag{1}$$

which might be loosely interpreted as the probability of having observed this particular sample.

- The Maximum likelihood estimator (MLE) of δ is the value that maximizes (1), i.e., the probability of having observed this sample.

3 steps to obtain the ML estimator

- Specify a distribution for the innovations of the process, ε_t. Typically, ε_t is a Gaussian white noise, $\varepsilon_t \sim iid \ \mathrm{N}\left(0, \sigma^2\right)$.
- Compute the log likelihood function (1).
- Obtain the values of δ that maximize the log likelihood.
- In general there is no a closed-form solution. Thus, numerical optimization procedures should be applied in order to obtain the values of the estimates.

Asymptotic properties of ML Estimators

ML estimators have very good asymptotic properties.

- Under certain conditions it can be shown that they are
  - consistent
  - asymptotically normal
  - efficient, since the variance-covariance matrix equals the inverse of the Fisher information matrix.

Asymptotic properties of ML Estimators, II

- Let $\{X_t\}$ be an ARMA(p,q) process, δ_0 be the vector containing the true parameter values and $\hat{\delta}$ is the MLE of δ. Assuming that neither δ_0 nor $\hat{\delta}$ falls on the boundary of the parameter space then

$$\sqrt{T}\left(\hat{\delta}-\delta_0\right) \xrightarrow{d} N\left(0, \mathcal{I}^{-1}\right)$$

where $\mathcal{I}$ is the Fisher information matrix

$$\mathcal{I} = -\mathrm{E}\left(\left.\frac{\partial^2 \mathcal{L}\left(\delta\right)}{\partial\delta\partial\delta'}\right|_{\delta=\delta_0}\right).$$

STEP 1: Identification of an ARMA process- Selecting p and q

- How to select p and q?
- ACF and PACF are of help but their usefulness is limited in the general ARMA(p, q) case.
- Other tools: model selection criteria (information criteria.
  - Information criteria: are rules to select a statistical model among a set of candidate models, based on observed data.
  - A model selection criterion that identifies the "correct" statistical model asymptotically with probability 1 as T tens to infinity is said to be consistent.

Information criteria

- Let k be the number of variables included in a candidate model.
- The main idea is that you only want to include in the model variables with strong explanatory power
- Information criteria are expressions that typically combine two elements:
  1. the variance of the error term: this term tends to decrease when more variables are included
  2. a penalty term: a term that is increasing in the number of variables

Information criteria; II

- Goal: find the model that minimizes the chosen information criterion.
- Why? the penalty term makes the introduction of variables that have small explanatory power costly
- Formulas for the IC are obtained by minimizing a distance between the "true" model and the different "candidates" (the Kullback-Leibler discrepancy, see Gourieux and Monfort)

Information criteria; II

- General expression of IC

$$I_T(k) = \log(\hat{\sigma}_k^2) + \underbrace{k\frac{C(T)}{T}}_{\text{penalty term}}, \quad (2)$$

Some popular IC are:

1. Akaike: $\mathrm{C}(T) = 2$

2. Schwartz or Bayesian (BIC): $C(T) = \log T$

3. Hannan and Quinn (HQ): C(T)=$2\log(\log T)$

- The number of variables $\hat{k}$ to be included in the model is chosen as

$$\hat{k} = \arg \min_{k \leq m} I_T(k)$$

where m is some pre-specified maximum number of variables.

Properties of the different IC: Consistency

- A model selection criterion that identifies the "correct" statistical model asymptotically with probability 1 as $T \to \infty$ is said to be consistent.
- AIC:
  - AIC is not consistent (i.e, it might not choose asymptotically the correct model with probability one).
  - The AIC tends to overestimate the number of parameters to be included in the model. However, this does not mean that this method is useless. It minimizes the MSE for one-step ahead forecasting.

Properties of the different IC: Consistency

- The BIC and the HQ criteria are consistent, that is

$$\lim_{T\to\infty} P(\hat{k} = k_0) = 1. \tag{3}$$

  In fact, the consistency result (3) holds for any criterion of the type (2) with $\lim_{T\to\infty} C(T)/T = 0$ **and** $\lim_{T\to\infty} C(T) = \infty$.

- Notice, however, that BIC and HQ are more likely to underestimate k_0 than the AIC.

- Underestimating k_0 has worse consequences than overestimating it!

Other model selection approaches: General-to-specific criterion.

General-to-specific criterion, (Ng and Perron, 1995).

- This method amounts to
  - set the maximum number of parameters to be estimated. For instance, an AR(k) model with $k = 7$.
  - Estimate the 'general' model, (an AR(7) in this case).
  - test for the statistical significance of the coefficients associated with the higher order lags –with either a t or an F test– (y_{t-7} in this case).
  - Exclude the nonsignificant parameters, reestimate the smaller model and test again for significance. Repeat until all the remaining estimates are significantly different from zero.

Summarizing: steps you should follow to fit an ARMA(p,d) process

- You should always start by plotting the data, ACFs and PACFs.
- What are the main patterns you see in the time plot, in the ACFs and the PACFs?
- Most likely it will look nonstationary: the first step will be to find a stationary transformation. We'll see how to do this.
- Plot again the transformed data, as well as its ACF and PACF.
- Use information criteria to choose p and q
- Estimate the resulting model by maximum likelihood.

Summarizing: steps you should follow to fit an ARMA(p,d) process, II

- You can use different software to do this: MATLAB (but you need the econometrics toolbox), STATA, EVIEWS, GRETL...

STEP 3: Diagnostic checking

- If the fitted model is appropriate, the residuals should behave in a similar way as the true innovations of the process.
- Thus, the last step will be to look at the residuals and check whether they are like the innovations in the assumed model.
- Innovations are typically assumed to be

  i.) zero-mean

  ii) homokedastic (ie. they should have constantnvariance)

  iii) uncorrelated.

- Signs of misspecification: cyclic or trended behavior, non-constant variance, etc.

- Among the conditions above: i) is not very restrictive (i.e., models that can be very "wrong" can satisfy this property), ii) is more stringent, iii) lag of correlation for any lag is central for ensuring that the model is suitable.

Autocorrelation tests

- Goal: we want to test whether the residuals are autocorrelated.
- Let $\{e_t\}$ be the sequence of residuals, given by

$$e_t = X_t - \hat{X}_t,$$

where $\hat{X}_t$ are the fitted values.

- Then, the autocorrelation function of the residuals is given by

$$\hat{\rho}_e(h) = \sum_{t=1}^{T-h} (e_t - \bar{e})(e_{t+h} - \bar{e}) \,/\, \sum_{t=1}^{T} (e_t - \bar{e})^2, \text{ h=1,2,...}$$

Box-Pierce statistic Q-statistic

- Under the null hypothesis that the model is correctly specified, then

$$Q_e(H) = T \sum_{i=1}^{H} \hat{\rho}_e^2(i) \xrightarrow{d} \chi^2_{(H-p-q)}.$$

- Thus, we reject that the model is correctly specified at level α if

$$Q_e(H) = T \sum_{i=1}^{H} \hat{\rho}_e^2(i) > \chi^2_{1-\alpha(H-p-q)}.$$

Ljung-Box test

- Ljung and Box (1978) suggest *replacing* Q_e by

$$Q_e{}^*(H) = n\,(n+2)\sum_{i=1}^{H}\hat{\rho}_e^2\,(i)\,/\,(n-i),$$

- This statistic offers a better approximation to the χ^2 distribution.

$$Q_e{}^*(H) = T\sum_{i=1}^{H}\hat{\rho}_e^2\,(i) \xrightarrow{d} \chi^2_{(H-p-q)}.$$

Tests for zero-mean

- The estimated residuals are not subject to the restriction that the average of the residuals is equal to zero.
- This condition is only imposed when estimating AR models including a constant.
- We can test this condition using the statistic

$$\frac{\bar{e}}{\bar{\sigma}_e/\sqrt{T}}$$

- And will reject the null of zero-mean is the value of this statistic is large when compared to the critical values of a normal distribution.

An example using STATA

The file Lecture2_ARMA.do will lead you through the 3 steps you need to follow to fit an ARMA process.

You can find that file here:

http://mayoral.iae-csic.org/timeseries2021/Lecture2_ARMA_Estimation.do

Wrapping up

1. Under stationary and ergodicity: LLNs and CLTs still work
2. If you run regressions with time series, beware of autocorrelated residuals! HAC standard errors
3. 3 steps in fitting ARMA models to stationary data:
   - Identification: information criteria (AIC, BIC, general to specific...)
   - Estimation: ML
   - Don't forget to look at the residuals to check whether your assumptions were right!

Wrapping up, II

Now: what can we do with these models?

- Many things!

For instance:

- Compute Impulse response functions
- Compute forecasts

The Impulse Response Function

Definition

Let $\{y_t\}$ be a time series process. The IRF is the path y follows if it is kicked by a unitary shock at time t i.e, $\varepsilon_t = 1$, assuming all other posterior shocks are equal to zero, i.e., $\varepsilon_{t+j} = 0$, for all $j > 1$.

- For univariate processes: How does y_t evolve over time when it is hit at time t by a shock, ε_t. Useful to understand the dynamics of the process and see how persistent a process is.

- For multivariate processes: When we are modeling several variables, the IRF describes the evolution of one variable, say GDP, when is perturbed by another variable at time t, say interest rates. Useful to understand the relationships among variables, the impact of economic policies, etc..

Computation

- The IRF can be computed as the derivative of y_{t+h} w.r.t to the shock that has been perturbed, ε_t

$$IRF(h) = \partial y_{t+h}/\partial \varepsilon_t, \text{ h} \geq 0.$$

If y_t is stationary

$$y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots$$

and then,

$$IRF(h) = \partial y_{t+h}/\partial \varepsilon_t = \psi_h.$$

Therefore, for stationary processes, the coefficients of the Wold representation are the values of the IRF.

Computation, II

For nonstationary processes, this is not true, since the Wold representation doesn't exist.

An alternative way of computing this function (for linear processes) is as follows:

$$IRF(h) = E(y_{t+h} | \varepsilon_t = 1) - E(y_{t+h} | \varepsilon_t = 0)$$

i.e., the IRF at lag h is the difference of two conditional expectations.

This is usually the simplest way of computing the IRF.

Estimation and inference of the IRF

Estimation.

The IRF is a function of the parameters of the model.

To estimate this function, you need first to estimate the model, then you use these values to compute the IRF, simply by replacing the unknown parameter by the estimate you obtained.

Inference

- for stationary processes, since the estimates of the parameters are asymptotically normal, estimates of the IRF are also asymptotically normal (by the delta method).
- However, the normal approximation works poorly for highly persistent processes.
- This is an important problem since economic data is usually very persistent.
- Solutions: Monte carlo confidence bands and bootstrap confidence bands.
- We'll see these methods in more detail when deriving IRFs for multivariate processes.

A few references

Hamilton, 1994, Chapter 5.
Hayashi, Chapter 6
Brockwell and Davis, (1991), Chapters 8, 9