Core Econometrics - Time Series

Basics

Static and Dynamic Regressions #flashcard

- Static regression: do not account for temporal dependence. e.g. regress y_t on a constant.
- Dynamic regression: model the temporal dependence explicitly. e.g. AR(1)

Correlograms #flashcard

- **Autocorrelation / Correlogram / Autocorrelation Function (acf.) a_s**
  - Estimation:
    - Signed $\sqrt{R^2}$ in the regression (with the sign of β_s) / normalised version of β_s in

$$Y_t = \beta_0 + \beta_s Y_{t-s} + u_t$$

    - Explicit formula:

$$a_s = \frac{\sum_{t=s+1}^{T} \left(Y_t - \overline{Y}_{s+1}^{T}\right)\left(Y_{t-s} - \overline{Y}_{1}^{T-s}\right)}{\sqrt{\sum_{t=s+1}^{T} \left(Y_t - \overline{Y}_{s+1}^{T}\right)^2 \sum_{t=s+1}^{T} \left(Y_{t-s} - \overline{Y}_{1}^{T-s}\right)^2}}$$

  - Properties:
    - For $Y_t \sim^{iid} (\mu, \sigma^2)$, $\sqrt{T} a_s \sim N(0, 1)$.
    - For $Y_t = \epsilon_t + \theta \epsilon_{t-1}$ (MA1), $\sqrt{T} a_s \sim N(0, 1)$ for $s \geq 2$
    - For $Y_t = \alpha Y_{t-1} + \epsilon_t$ (AR1), $a_s = \alpha^s$
- **Partial Autocorrelation / Partial Correlogram / Partial Auctocorrelation Function (pacf.) p_s**
  - Measures the conditional correlation of Y_t and Y_{t-s} given the observation in between
  - Estimation:
  - A scaled version of β_s the least square regression:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_s Y_{t-s} + u_t$$

  - *Frisch-Waugh-Lovell* idea:
  - Estimate 2 aux regression of Y_t, Y_{t-s} on $Y_{t-1}, \ldots, Y_{t-s+1}$ and obtain residuals $\hat{v}_{0,t}$ and $\hat{v}_{s,t}$
  - Compute the coefficient of correlation between the residuals:

$$p_s = \frac{\sum_{t=s+1}^{T} \hat{v}_{0,t} \hat{v}_{s,t}}{\sqrt{\sum_{t=s+1}^{T} \hat{v}_{0,t}^2 \sum_{t=s+1}^{T} \hat{v}_{s,t}^2}}$$

  - Properties:
  - For $Y_t \sim^{iid} (\mu, \sigma^2)$, $\sqrt{T} p_s \sim N(0, 1)$.
  - For $Y_t = \epsilon_t + \theta \epsilon_{t-1}$ (MA1), $p_s = \theta^s$
  - For $Y_t = \alpha Y_{t-1} + \epsilon_t$ (AR1), $p_s = 0, \sqrt{T} p_s \sim N(0, 1)$ for $s \geq 2$
  - For an AR(k) process $\implies$ $p_s = 0$ for $s > k$

MA(1) Process #flashcard

- Moving Average MA(1):

$$Y_t = \epsilon_t + \theta \epsilon_{t-1}$$

  where ϵ_t is iid.
- The scaled autocorrelation coefficient $\sqrt{T} a_s$ will have an asymptotic normal distribution.
- Correlation coefficient $a_s = 0, \sqrt{T} a_s \sim N(0, 1)$ for $s \geq 2$
- Partial autocorrelation coefficient $p_s = \theta^s$

Markov Property #flashcard

- The condition density of current value Y_t only depends on the past through Y_{t-1}:

$$f(Y_t | Y_{t-1}, Y_{t-2}, \ldots) = f(Y_t | Y_{t-1})$$

Martingale Sequences and Martingale Difference Sequences #flashcard

- Martingale Sequences:

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] = X_t$$

- Martingale Difference Sequences:

$$\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = 0$$

AR(1)

AR(1) without Intercept

AR(1) without Intercept Setup and Estimation #flashcard

- AR(1) without intercept setup:

$$X_t = \alpha X_{t-1} + \epsilon_t, \epsilon_t \sim^{iid} N(0, \sigma^2)$$

  - This is a Markov process
- Estimation:
  - Standard SLR OLS estimator:

$$\hat{\alpha} = \frac{\sum_{t=1}^T X_t X_{t-1}}{\sum_{t=1}^T X_{t-1}^2}, s^2 = \frac{1}{T-1} \sum_{t=1}^T \hat{\epsilon}_t^2$$

  - Avar:

$$\widehat{Avar}(\hat{\beta}) = \frac{s^2}{\sum_{t=1}^T X_{t-1}^2}$$

  - t-like-stat:

$$Z = \frac{\widehat{\alpha} - 0}{\sqrt{\frac{s^2}{\sum_{t=1}^T X_{t-1}^2}}}$$

Prediction Decomposition and Likelihood for AR(1)

- *This applies to other TS models!*
- AR(1) without intercept setup:

$$X_t = \alpha X_{t-1} + \epsilon_t, \epsilon_t \sim^{iid} N(0, \sigma^2)$$

- Derive the Joint Log Likelihood using the Prediction Decomposition #flashcard
- Prediction decomposition $\rightsquigarrow$ Autoregressive Likelihood:

$$f(X^T, \ldots, X_1 | X_0) = \prod_{t=1}^T f(X_t | X_{t-1}, \ldots, X_0)$$

-
  - Markov Property $\rightsquigarrow$:

$$\prod_{t=1}^T f(X_t | X_{t-1}, \ldots, X_0) = \prod_{t=1}^T f(X_t | X_{t-1})$$

-
  - Functional Form Assumptions $\rightsquigarrow$:

$$\prod_{t=1}^T f(X_t | X_{t-1}) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_t - \alpha X_{t-1})^2}{2\sigma^2}\right)$$

- Take logs:

$$\mathcal{L} = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T (X_t - \alpha X_{t-1})^2$$

- Log-likelihood
  - Specific Derivation
  -

2.1.3 The prediction decomposition

The autoregressive estimators are maximum likelihood estimators. The key to deriving this result is the *prediction decomposition*. For any joint density, we have

$$\mathsf{f}(x_T,\ldots,x_1 \mid x_0) = \prod_{t=1}^{T} \mathsf{f}(x_t \mid x_{t-1},\ldots,x_0) \qquad (2.5)$$

-

This is proved by successive conditioning.

$$\mathsf{f}(x_T,\ldots,x_1 \mid x_0) = \frac{\mathsf{f}(x_T,\ldots,x_1,x_0)}{\mathsf{f}(x_0)} = \left\{\frac{\mathsf{f}(x_T,\ldots,x_1,x_0)}{\mathsf{f}(x_{T-1},\ldots,x_1,x_0)}\right\}\left\{\frac{\mathsf{f}(x_{T-1},\ldots,x_1,x_0)}{\mathsf{f}(x_0)}\right\}$$
$$= \mathsf{f}(x_T \mid x_{T-1},\ldots,x_0)\mathsf{f}(x_{T-1},\ldots,x_1 \mid x_0).$$

Recursion

Repeat the argument for the second factor in the last expression.

$$\frac{f(x_{T-1},\ldots,x_0)}{f_0} = \underbrace{\frac{f(x_{T-1},\ldots,x_0)}{f(x_{T-2},\ldots,x_0)}}_{f(x_{T-1}\mid x_{T-2},\ldots,x_0)} \underbrace{\frac{f(x_{T-2},\ldots,x_0)}{f(x_0)}}_{f(x_{T-2},\ldots,x_1\mid x_0)}$$

$$\ldots = \prod_{t=1}^{T} f(x_t \mid x_{t-1},\ldots,x_0)$$

Useful for more general TS models.

2.1.4 The autoregressive likelihood

We derive the likelihood for the Gaussian autoregressive model in § 2.1.1. For given values of the parameters α, σ^2 the joint density of outcomes $x_1,\ldots,x_T$ given x_0 is

$$\mathsf{f}_{\alpha,\sigma^2}(x_1,\ldots,x_T|x_0) = [\text{ prediction decomposition }] = \prod_{t=1}^{T} \mathsf{f}_{\alpha,\sigma^2}(x_t|x_{t-1},\ldots,x_0) \qquad (2.6)$$

Markov Process Property

$$= [\ (2.1)\ \&\ (i)\ ] = \prod_{t=1}^{T} \mathsf{f}_{\alpha,\sigma^2}(x_t|x_{t-1}) \qquad (2.7)$$

Distribution of ϵ_t

$$= [\ (ii)\ ] = \prod_{t=1}^{T} (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(x_t - \alpha x_{t-1})^2\right\} \qquad (2.8)$$

Put Π inside

$$= (2\pi\sigma^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(x_t - \alpha x_{t-1})^2\right\}. \qquad (2.9)$$

The log likelihood equals the log density evaluated in the observations and viewed as a function of the parameters:

$$\ell_{X_1,\ldots,X_T|X_0}(\alpha,\sigma^2) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(X_t - \alpha X_{t-1})^2. \qquad (2.10)$$

This is a classical regression likelihood. We recognise the sum of squared deviations (2.2) and maximize by least squares. Thus, the least square estimators are maximum likelihood, albeit the residual variance should be normalized by T instead of the degrees of freedom.

AR(1) without Intercept Interpretation #flashcard

- AR(1) without intercept setup:

$$X_t = \alpha X_{t-1} + \epsilon_t, \epsilon_t \sim^{iid} N(0,\sigma^2)$$

- Recursive Solution:

$$X_t = \sum_{j=0}^{t-1} \alpha^j \epsilon_{t-j} + \alpha^t X_0$$

- Conditional Distribution:

$$X_t|X_0 \sim N\left(\underbrace{\alpha^t X_0}_{\mathbb{E}[X_t|X_0]}, \underbrace{\frac{1-\alpha^{2t}}{1-\alpha^2}\sigma^2}_{Var[X_t|X_0]}\right)$$

- Long-run Mean, Variance, and Autocovariance for $\alpha \neq 1$:

$$\lim_{t\to\infty} \begin{cases} \mathbb{E}\left[X_t|X_0\right] & = 0 \\ Var[X_t|X_0] & = \sigma_X^2 = \frac{\sigma^2}{1-\alpha^2} \\ Cov\left[X_t^*, X_{t-s}^*\right] & = \alpha^s \sigma_X^2 \end{cases}$$

  - Same if we treat X_0 as random.
- Stationary/Invariant (LR) Distribution:

$$X_t|X_0 \sim^a N\left(0, \frac{\sigma^2}{1-\alpha^2}\right)$$

  - Same if we treat X_0 as random.
- Infinite Representation:

$$X_t = \sum_{i=0}^{\infty} \alpha^i \epsilon_{t-i}$$

Strict Stationarity #flashcard

A process X_t^* is **strictly stationary** if $\forall\ s$ the joint distribution $(X_{t+1}^*,\ldots,X_{t+s}^*)$ does not depend on t.

- i.e. stable joint distribution across time
  Preserve of Stationary under Transformations: if X_t^* is a stationary process and $g : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function, then the process $Z_t = g(X_t^*, X_{t-1}^*, \ldots)$ is also stationary.

Weak/Covariance Stationarity #flashcard

- A process X_t is **weak/covariance stationary** if:

$$\forall\ s, t : \begin{cases} \mathbb{E}\left[X_t\right] & = \mu \\ Cov\left[X_t, X_{t+s}\right] & = \psi_s \end{cases}$$

  - i.e. mean, autocovariance do not change with t
  - A covariate stationary process is strictly stationary if it is Normally distributed.

White Noise Process #flashcard

- A process $\{X_t\}$ is called **white noise** if:
  - It is covariate/weakly stationary
  - $\mathbb{E}\left[X_t\right] = 0$
  - $Cov[X_t, X_{t+s}] = 0\ \forall\ s \neq 0, t$

TSLLN - AR(1) without Intercept #flashcard

Theorem 2.4. *LLN for autoregressions. Let $X_t = \alpha X_{t-1} + \varepsilon_t$ for $t \in \mathbb{N}$. Suppose $|\alpha| < 1$ and that ε_t are i.i.d. with mean zero and finite variance σ^2 conditional on X_0. The initial observation X_0 is a fixed or a random variable. Then, for $T \to \infty$,*

New

$$T^{-1}\sum_{t=1}^{T} X_{t-j} \xrightarrow{\mathsf{P}} 0 \qquad \text{for } j = 0, 1, \tag{2.35}$$

$$T^{-1}\sum_{t=1}^{T} X_{t-j}^2 \xrightarrow{\mathsf{P}} \underbrace{\sigma_X^2 = \frac{\sigma^2}{1-\alpha^2}}_{\text{LR Variance}} \qquad \text{for } j = 0, 1, \tag{2.36}$$

$$T^{-1}\sum_{t=1}^{T} X_{t-1}\varepsilon_t \xrightarrow{\mathsf{P}} 0, \tag{2.37}$$

$\Rightarrow \mathbb{E}[X_{t-1}\varepsilon_t] \overset{\perp}{=} \mathbb{E}[X_{t-1}]\mathbb{E}[\varepsilon_t] = 0$

$$T^{-1}\sum_{t=1}^{T} \varepsilon_t^2 \xrightarrow{\mathsf{P}} \sigma^2. \tag{2.38}$$

Residual variance (standard LLN)

TSCLT - AR(1) without Intercept #flashcard

Theorem 2.5. *CLT for autoregressions. Let $X_t = \alpha X_{t-1} + \varepsilon_t$ for $t \in \mathbb{N}$. Suppose $|\alpha| < 1$ and that ε_t are i.i.d. with mean zero, variance σ^2 conditional on X_0 and $\mathsf{E}|\varepsilon_t|^{2+p} < \infty$ for some $p > 0$. The initial observation X_0 is a fixed or a random variable. Then, for $T \to \infty$ and $j = 0, 1$,*

$$T^{-1/2}\sum_{t=1}^{T} X_{t-j} = \frac{1}{1-\alpha}T^{-1/2}\sum_{t=1}^{T}\varepsilon_t + \mathrm{o_P}(1) \xrightarrow{\mathsf{D}} \mathsf{N}\Big\{0, \underbrace{\frac{\sigma^2}{(1-\alpha)^2}}_{\text{Not the LR Var!!!}}\Big\}, \tag{2.44}$$

$$T^{-1/2}\sum_{t=1}^{T}\begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix}\varepsilon_t \xrightarrow{\mathsf{D}} \mathsf{N}_2\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2\begin{pmatrix} \sigma_X^2 & 0 \\ 0 & 1 \end{pmatrix} \right\} \qquad \text{where } \sigma_X^2 = \sigma^2/(1-\alpha^2). \tag{2.45}$$

Independence

20

- Note this is different from the LR var of X_t!

Asymptotics for AR(1) without Intercept

- Prove Consistence and Asymp. Distribution of estimators:

$$\hat{\alpha} = \frac{\sum_{t=1}^{T} X_t X_{t-1}}{\sum_{t=1}^{T} X_{t-1}^2}, s^2 = \frac{1}{T-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2$$

- Derive the asymp distribution of Z-stat (no finite-sample argument, so use Z instead of t-stat):

$$Z_{\alpha=0} = \frac{\sqrt{T}\left(\hat{\alpha} - 0\right)}{\sqrt{\frac{s^2}{\sum_{t=1}^{T} X_{t-1}^2}}}$$

  - However, note that this seems to be the only place we have $\sqrt{T}$. In the later part of the notes, Bent seems to still use the standard t-stat, just denoted as Z. #flashcard
- Consistency of estimator

  Consistency result for the autoregresive estimator in ($\boxed{2.31}$). Rewrite the estimator using the model equation ($\boxed{2.1}$) as

$$\widehat{\alpha} = \frac{\sum_{t=1}^{T} X_{t-1}X_t}{\sum_{t=1}^{T} X_{t-1}^2} = \frac{\sum_{t=1}^{T} X_{t-1}(\alpha X_{t-1} + \varepsilon_t)}{\sum_{t=1}^{T} X_{t-1}^2} = \alpha + \frac{\sum_{t=1}^{T} X_{t-1}\varepsilon_t}{\sum_{t=1}^{T} X_{t-1}^2}. \tag{2.46}$$

  Thus, we first normalize by T, then appeal to the LLN for autoregressions for numerator and denominator, and finally combine the limits by appealing to the Slutsky result or Continuous Mapping Theorem to get

  TSLLN Slutsky Theorem

$$\widehat{\alpha} - \alpha = \frac{T^{-1}\sum_{t=1}^{T} X_{t-1}\varepsilon_t}{T^{-1}\sum_{t=1}^{T} X_{t-1}^2} = \frac{\overbrace{T^{-1}\sum_{t=1}^{T} X_{t-1}\varepsilon_t}^{\xrightarrow{\mathsf{P}} 0}}{\underbrace{T^{-1}\sum_{t=1}^{T} X_{t-1}^2}_{\xrightarrow{\mathsf{P}} \sigma_X^2\ \ \frac{\sigma^2}{(1-\alpha)^2}}} \xrightarrow{\mathsf{P}} \frac{0}{\sigma_X^2} = 0. \tag{2.47}$$

-

- Asymptotic distribution of estimator

Asymptotic distribution for the autoregresive estimator in (2.32). Use the identity established in (2.46) to get

$$\widehat{\alpha}-\alpha=\frac{\sum_{t=1}^{T}X_{t-1}\varepsilon_t}{\sum_{t=1}^{T}X_{t-1}^2} \tag{2.48}$$

Then normalize by $T^{1/2}$ on the left and distribute that normalization on numerator and denominator on the right to get

$$T^{1/2}(\widehat{\alpha}-\alpha)=\frac{T^{-1/2}\sum_{t=1}^{T}X_{t-1}\varepsilon_t}{T^{-1}\sum_{t=1}^{T}X_{t-1}^2} \tag{2.49}$$

First, apply the CLT for autoregressions for the numerator and the LLN for autoregressions for the denominator and then appeal to the Continuous Mapping Theorem to get

TSCLT CMT/Slutsky Var[ε] × Var[Xt] since they are independent

$$T^{1/2}(\widehat{\alpha}-\alpha)=\frac{\overbrace{T^{-1/2}\sum_{t=1}^{T}X_{t-1}\varepsilon_t}^{\xrightarrow{\mathsf{D}}\mathsf{N}(0,\sigma_X^2\sigma^2)}}{\underbrace{T^{-1}\sum_{t=1}^{T}X_{t-1}^2}_{\xrightarrow{\mathsf{P}}\sigma_X^2\ \ \frac{\sigma^2}{(1-\alpha)^2}}}\xrightarrow{\mathsf{D}}\mathsf{N}\left(0,\frac{\sigma^2\sigma_X^2}{\sigma_X^4}\right)=\mathsf{N}\left(0,\frac{\sigma^2}{\sigma_X^2}\right)=\mathsf{N}(0,1-\alpha^2). \tag{2.50}$$

TSLLN Squared cancel σ_X^2 $\frac{\sigma^2}{\frac{\sigma^2}{(1-\alpha)^2}}=(1-\alpha)^2$

We note that the limiting distribution does not depend on σ^2. This is because the regressand X_t and the regressor X_{t-1} have the same scale. Further, the limiting variance $1-\alpha^2$ is positive whenever $|\alpha|<1$, which is also the condition for stationarity.

- Consistency of residual variance estimator and LR X Var estimator:

Consistency result for the residual variance estimator in (2.31). Rewrite the sum of squared residuals noting that $\widehat{\varepsilon}_t=X_t-\widehat{\alpha}X_{t-1}=\varepsilon_t-(\widehat{\alpha}-\alpha)X_{t-1}$ so that

$$\sum_{t=1}^{T}\widehat{\varepsilon}_t^2=\sum_{t=1}^{T}\{\varepsilon_t-(\widehat{\alpha}-\alpha)X_{t-1}\}^2=\sum_{t=1}^{T}\left(\varepsilon_t-\frac{\sum_{t=1}^{T}X_{t-1}\varepsilon_t}{\sum_{t=1}^{T}X_{t-1}^2}X_{t-1}\right)^2$$
$$=\sum_{t=1}^{T}\varepsilon_t^2-\frac{(\sum_{t=1}^{T}X_{t-1}\varepsilon_t)^2}{\sum_{t=1}^{T}X_{t-1}^2}. \tag{2.51}$$

We normalize, then apply to the LLN for autoregressions and finally appeal to the Slutsky result or Continuous Mapping Theorem to get

0 (Slutsky)

$$s^2=\underbrace{\frac{1}{T-1}}_{\text{DoF Adj.}}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2=\underbrace{\frac{T}{T-1}}_{\to 1}\left\{\underbrace{T^{-1}\sum_{t=1}^{T}\varepsilon_t^2}_{\xrightarrow{\mathsf{P}}\sigma^2}-\overbrace{\frac{(T^{-1}\sum_{t=1}^{T}X_{t-1}\varepsilon_t)^2}{\underbrace{T^{-1}\sum_{t=1}^{T}X_{t-1}^2}_{\text{TSLLN }\xrightarrow{\mathsf{P}}\sigma_X^2}}}^{\text{TSLLN }\xrightarrow{\mathsf{P}}0}\right\}\xrightarrow{\mathsf{P}}\sigma^2. \tag{2.52}$$

Consistency result for the estimator of the variance of the stationary distribution of X_t in (2.31). We have directly from the LLN for autoregressions that

$$s_X^2=T^{-1}\sum_{t=1}^{T}X_{t-1}^2\xrightarrow{\mathsf{P}}\sigma_X^2. \tag{2.53}$$

- Asymptotic Distribution of t-stat:

  -

Asymptotic distribution for the t-statistic in (2.32). Combine the above results using the Continuous Mapping Theorem to get

$$Z_{\alpha=0}=\frac{\overbrace{T^{1/2}(\widehat{\alpha}-0)}^{\xrightarrow{\mathsf{D}}\mathsf{N}(0,\sigma^2/\sigma_X^2)}}{(\underbrace{s^2}_{\xrightarrow{\mathsf{P}}\sigma^2}/\underbrace{s_X^2}_{\xrightarrow{\mathsf{P}}\sigma_X^2})^{1/2}}\xrightarrow{\mathsf{D}}\mathsf{N}(0,1). \tag{2.54}$$

AR(1) with Intercept

AR(1) with Intercept Setup and Estimation

- Setup, Estimators, Test statistic, Likelihood #flashcard
- **AR(1) setup**:

$$Y_t=\alpha Y_{t-1}+\mu+\epsilon_t$$

where the innovations ϵ_t satisfies:

- **conditional independence (so serial correlation)**: $\epsilon_1,\ldots,\epsilon_T$ are mutually independent given Y_0
- **conditional normality**: $\epsilon_t\sim N(0,\sigma^2)$
- **Estimation and test**:
  - Same as cross-sectional regression
  - Stack the model:

$$Y=S\beta+\epsilon$$

- Estimators:

$$\hat{\beta}=(S^TS)^{-1}S^TY,s^2=\frac{\hat{\epsilon}^T\hat{\epsilon}}{T-2}$$

and

$$\hat{\alpha}=\frac{\sum_{t=1}^{T}Y_t(Y_{t-1}-\overline{Y}_-)}{\sum_{t=1}^{T}(Y_{t-1}-\overline{Y}_-)^2}$$

- Asymptotic variance:

$$\widehat{Avar}(\hat{\beta})=(S^TS)^{-1}s^2$$

- Test statistic for no autocorrelation ($H_0 : \alpha = 0$):

$$Z_{\alpha=0} = \frac{\hat{\alpha} - 0}{\sqrt{\frac{s^2}{\sum_{t=1}^{T}(Y_{t-1} - \overline{Y}_{-})^2}}}$$

- Joint Log-Likelihood:

$$l_{Y_1,\ldots,Y_T}(\alpha, \mu, \sigma^2) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(Y_t - \alpha Y_{t-1} - \mu)^2$$

AR(1) with Intercept Interpretation

- AR(1) setup:

$$Y_t = \alpha Y_{t-1} + \mu + \epsilon_t$$

and $\epsilon_t \sim^{iid} N(0, \sigma^2)$
- Stationarity, Unobserved Components Formulation, LR Mean, LR Variance #flashcard
- Stationarity requirement: $|a| < 1$
- Assuming stationarity:
- Long-run Mean:

$$\mu_Y = \frac{\mu}{1 - \alpha}$$

- Which can be derived by Unobserved Components Formulation:

Unobserved components formulation. We decompose the solution to the autoregressive equation (3.8) into a stochastic part and a deterministic part:

$$Y_t = \underbrace{X_t}_{\text{Stochastic}} + \underbrace{\mu_Y}_{\text{Deterministic}}, \qquad t = 1, \ldots, T, \tag{3.12}$$

$$\text{where} \qquad X_t = \alpha X_{t-1} + \varepsilon_t. \tag{3.13}$$

$Y_t = \alpha Y_{t-1} + \mu + \varepsilon_t$

$Y_t = X_t + \frac{\mu}{1-\alpha}$

$X_t = \alpha X_{t-1} + \varepsilon_t$

Here, μ_Y is a constant level parameter while X_t is an unobserved component satisfying an autoregressive equation without intercept. We recognize the equation for X_t from (2.1). In particular, we have seen that the X_t equation has a stationary solution when $|\alpha| < 1$. We say (3.12)-(3.13) is an *unobserved components* representation of the dynamic model.

We find an expression for the level μ_Y as follows. Note that the calculation will not require $|\alpha| < 1$. Subtract and add μ_Y from Y_t and Y_{t-1} in (3.8) to get

$$\underbrace{(Y_t - \mu_Y)}_{X_t} = \alpha\underbrace{(Y_{t-1} - \mu_Y)}_{X_{t-1}} + \underbrace{(\mu - \mu_Y + \alpha\mu_Y)}_{\text{set } = 0} + \varepsilon_t. \tag{3.14}$$

$Y_t = \alpha Y_{t-1} + \mu + \varepsilon_t$

$Y_t = \alpha Y_{t-1} - \alpha\mu_Y + \alpha\mu_Y + \mu + \varepsilon_t$

$\underbrace{Y_t - \mu_Y}_{X_t} = \alpha\underbrace{(Y_{t-1} - \mu_Y)}_{X_{t-1}} + \alpha\mu_Y + \mu + \varepsilon_t - \mu_Y$

$X_t = X_{t-1} + [\mu + (\alpha - 1)\mu_Y] + \varepsilon_t$

The equation

$$\mu - \mu_Y + \alpha\mu_Y = 0 \tag{3.15}$$

can be solved when $\alpha \neq 1$ giving the level parameter

Deterministic Part:

$$\mu_Y = \frac{\mu}{1 - \alpha}. \tag{3.16}$$

The unit root case where $\alpha = 1$ will be covered later. Note, that for now we cannot interpret the long-run mean as an expectation. Rather, it is interpreted through the unobserved component representation.

Assuming stationarity and taking expectations. Suppose we assume that the autoregression is stationary. The unobserved components formulation (3.12)-(3.13) shows that Y_t^* has a stationary distribution when $|\alpha| < 1$. But for now, we only use the high level assumption that Y_t^* is stationary. To emphasize that this assumption is made we use the $*$ notation, so that the autoregressive equation is

$$Y_t^* = \alpha Y_{t-1}^* + \mu + \varepsilon_t. \tag{3.17}$$

Taking expecations, we get

$$\mathsf{E}(Y_t^*) = \alpha\mathsf{E}(Y_{t-1}^*) + \mu + \mathsf{E}(\varepsilon_t). \tag{3.18}$$

By stationarity we have $\mu_Y^* = \mathsf{E}(Y_t^*) = \mathsf{E}(Y_{t-1}^*)$. Noting that also $\mathsf{E}(\varepsilon_t) = 0$, we get that

$$\mu_Y^* = \alpha\mu_Y^* + \mu + 0, \tag{3.19}$$

-

matching the equation (3.15). When $\alpha \neq 1$, we have the same solution as before:

$$\mu_Y^* = \frac{\mu}{1 - \alpha} = \mu_Y. \tag{3.20}$$

- Long-run Variance:

$$\sigma_Y^{2*} = \frac{\sigma^2}{1 - \alpha^2}$$

- which can also be derived by Unobserved Components Formulation:

Assuming stationarity and taking variance. We can derive the variance of the stationary distribution in a similar way. Take variance on both sides of (3.17) to get

$$\mathsf{Var}(Y_t^*) = \mathsf{Var}(\alpha Y_{t-1}^* + \mu + \varepsilon_t). \tag{3.21}$$

The right hand side reduces to $\mathsf{Var}(\alpha Y_{t-1}^* + \varepsilon_t)$ as constants do not contribute to variance. Since αY_{t-1}^* and ε_t are independent, we get a further reduction to

$$\mathsf{Var}(Y_t^*) = \alpha^2 \mathsf{Var}(Y_{t-1}^*) + \mathsf{Var}(\varepsilon_t). \tag{3.22}$$

$\sigma_{Y^*}^2 = \alpha^2\sigma_{Y^*}^2 + \sigma^2$ $\quad (1-\alpha^2)\sigma_{Y^*}^2 = \sigma^2$

By stationarity, $\mathsf{Var}(Y_t^*) = \mathsf{Var}(Y_{t-1}^*) = \sigma_Y^{2*}$, say, while $\mathsf{Var}(\varepsilon_t) = \sigma^2$. Thus, we have the equation $\sigma_Y^{2*} = \alpha^2\sigma_Y^{2*} + \sigma^2$ with solution

$$\sigma_Y^{2*} = \frac{\sigma^2}{1-\alpha^2} = \sigma_X^2. \tag{3.23}$$

This shows that the stationary distributions of Y_t and X_t have the same variance. In fact, by the unobserved components representation we have $Y_t = X_t + \mu_Y$ so that the stationary distribution of Y_t is found by adding μ_Y to the stationary distribution of X_t.

AR(1) with Intercept Asymptotic Distribution #flashcard

Consider the first order autoregression

AR(1)
$$Y_t = \alpha Y_{t-1} + \mu + \varepsilon_t, \tag{3.26}$$

where the innovations ε_t are i.i.d.$(0, \sigma^2)$ and $|\alpha| < 1$. The details of the asymptotic theory are given in Appendix §ss:asymp:ar1c. In line with the results in §2.3, the least squares estimators are consistent, so that

Consistency:
$$\widehat{\alpha} \overset{\mathsf{P}}{\to} \alpha, \qquad \widehat{\mu} \overset{\mathsf{P}}{\to} \mu, \qquad s^2 \overset{\mathsf{P}}{\to} \sigma^2. \tag{3.27}$$

Further, the LR and t statistics on μ and α have standard asymptotic distributions

Distributions:
$$\mathsf{LR}_{\mu=0} \overset{\mathsf{D}}{\to} \chi_1^2, \qquad Z_{\mu=0} \overset{\mathsf{D}}{\to} \mathsf{N}(0,1), \tag{3.28}$$
$$\mathsf{LR}_{\alpha=0} \overset{\mathsf{D}}{\to} \chi_1^2, \qquad Z_{\alpha=0} \overset{\mathsf{D}}{\to} \mathsf{N}(0,1), \tag{3.29}$$
$$\mathsf{LR}_{\alpha=\mu=0} \overset{\mathsf{D}}{\to} \chi_2^2. \tag{3.30}$$

- Thus, the usual standard errors are valid in a well-specified autoregression with intercept when $|\alpha| < 1$.

(General) LR Tests in Time Series #flashcard

- LR Test:

$$-2\left\{l_{\text{Unstricted}} - l_{\text{Restricted}}\right\} \sim \chi^2_{\Delta \#\text{parameters}}$$

Misspecified Models

Estimation in a Misspecified Model: Autocorrelation in Errors and Its Correction

- Estimate:

$$Y_t = \mu_Y + u_t,\ u_t = \alpha u_{t-1} + \epsilon_t, \epsilon \sim^{iid} (0, \sigma^2)$$

- What happens to consistency of estimators $\tilde{\mu}_Y, \tilde{s}^2$? Impact on inference? How to deal with autocorrelations in errors? #flashcard
- In short: *no problem for consistency, but problem for inference. We need to add lags or use HAC SE.*
- Both estimators $\tilde{\mu}_Y, \tilde{s}^2$ are *still consistent* for the expectation and variance of the stationary distribution of Y_t:

$$\hat{\mu}_Y \to^p \mu_Y, \tilde{s}_Y^2 \to^p \sigma_Y^2 = \frac{\sigma^2}{1-\alpha^2}$$

- Further, $\tilde{\mu}_Y = \overline{Y}$ has an asymp. Normal Distribution:

$$\sqrt{T}(\tilde{\mu}_Y - \mu Y) \to^D N\left(0, \frac{\sigma^2}{(1-\alpha)^2}\right)$$

- However, *the t-(like)-stat for testing $\mu_Y = 0$ will not have the desired $\widehat{se}$ for a Standard Normal Dist!!*:

$$Z_{\mu_Y=0} = \frac{\sqrt{T}(\tilde{\mu}_Y - 0)}{\tilde{s}_Y} \to^D N\left(0, \frac{\frac{\sigma^2}{(1-\alpha)^2}}{\frac{\sigma^2}{1-\alpha^2}}\right) = N\left(0, \frac{1-\alpha}{1+\alpha}\right) \neq N(0,1)$$

- **Solutions**
  1. Add more lagged dependent variables
  - We can soak up dependencies in errors by adding more lagged dependent variables
  2. Use HAC SE for inference
  - We can estimate α and correct the t-(like)-stat:

$$\sqrt{\frac{1-\hat{\alpha}}{1+\hat{\alpha}}} Z_{\mu_Y=0} = \frac{\sqrt{T}(\tilde{\mu}_Y - 0)}{\tilde{s}_Y\sqrt{\frac{1-\hat{\alpha}}{1+\hat{\alpha}}}} \to^D N(0,1)$$

  - There exists non-parametric version.
  - Large HAC correction implies that we should properly form a dynamic model instead of trusting the HAC correction.

More AR Models

Linear Trends

AR(1) with a Linear Trend Setup and Basics

- Setup, Unobserved Components Formulation, Estimation #flashcard
- AR(1) with an intercept and a linear trend:

$$Y_t = \alpha Y_{t-1} + \mu_c + \mu_l t + \epsilon_t$$

- Unobserved Components Formulation:

$$\begin{cases} Y_t & = X_t + \underbrace{\frac{\mu_l}{1-\alpha}}_{\tau_l} t + \underbrace{\frac{1}{1-\alpha}\left(\mu_c - \frac{\alpha\mu_l}{1-\alpha}\right)}_{\tau_c} \\ X_t & = \alpha X_{t-1} + \epsilon_t \end{cases}$$

  - If X_t is stationary, then Y_t is "trend stationary"
  -

This model can be manipulated in the same way as before. For $\alpha \neq 1$ we have

$$Y_t = X_t + \underbrace{\tau_c + \tau_\ell t}_{\text{Deterministic}}, \qquad t = 1, \dots, T, \tag{4.2}$$

where
$$\underbrace{X_t = \alpha X_{t-1} + \varepsilon_t}_{\text{Stochastic Part (Stationary)}} \tag{4.3}$$

and $X_0 = Y_0 - \tau_c$. The subscripts in τ_c and τ_ℓ denote the constant and the *l*inear term. Subtract $\tau_c + \tau_\ell t$ from Y_t in equation (4.1) to get

$$\underbrace{(Y_t - \tau_c - \tau_\ell t)}_{X_t} = \alpha\{\underbrace{Y_{t-1} - \tau_c - \tau_\ell(t-1)}_{X_{t-1}}\} + (\underbrace{\mu_c - \tau_c + \alpha\tau_c - \alpha\tau_\ell}_{\text{set }=0}) + (\underbrace{\mu_\ell - \tau_\ell + \alpha\tau_\ell}_{\text{set }=0})t + \varepsilon_t, \tag{4.4}$$

☆ Trend slope is not the coefficient of t !

with solutions

$$\tau_\ell = \frac{\mu_\ell}{1-\alpha}, \qquad \tau_c = \frac{\mu_c - \alpha\tau_\ell}{1-\alpha} = \frac{1}{1-\alpha}\left(\mu_c - \frac{\alpha\mu_\ell}{1-\alpha}\right). \tag{4.5}$$

The expression for the *slope* τ_ℓ is similar to that for the *mean* (3.16) in the intercept model. The expression for the intercept τ_c is more fiddly, but also less important to remember.

A process that can be written as the sum of a stationary series and a linear trend as in (4.2) is often said to be *trend stationary*. Stationary around a linear trend.

- Estimation: OLS or MLE with standard TS likelihood

Higher Order Autoregressions (AR(k))

AR(2) with Intercept Setup, Estimation, Interpretation, and Asymptotics

- Setup, Estimation, LR Mean, Stationarity, Asymptotics #flashcard
- AR(2) setup:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \mu + \epsilon_t, \epsilon_t \sim^{iid} N(0, \sigma^2)$$

- This can be estimated by OLS or MLE with standard TS likelihood.
- Interpretation:
  - LR Mean:

$$\mu_Y = \frac{\mu}{1 - \alpha_1 - \alpha_2}$$

  - Unobserved Components Formulation

4.2.2 Interpretation: deterministic terms

Just as for the first order autoregression we get the unobserved components formulation

$$Y_t = X_t + \mu_Y \qquad t = 1, \dots, T \tag{4.8}$$
$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t. \tag{4.9}$$

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \mu + \varepsilon_t$$
$$\downarrow$$
$$\begin{cases} Y_t = X_t + \dfrac{\mu}{1-\alpha_1-\alpha_2} \\ \quad\; = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t \end{cases}$$

To see this introduce a parameter μ_Y in equation (4.7) to get

$$\underbrace{Y_t - \mu_Y}_{=X_t} = \alpha_1(\underbrace{Y_{t-1} - \mu_Y}_{=X_{t-1}}) + \alpha_2(\underbrace{Y_{t-2} - \mu_Y}_{=X_{t-2}}) + \underbrace{\mu - \mu_Y(1 - \alpha_1 - \cdots - \alpha_k)}_{\text{set }=0} + \varepsilon_t. \tag{4.10}$$

The equation $\mu - \mu_Y(1 - \alpha_1 - \alpha_2) = 0$ can be solved when $\sum_{\ell=1}^{2} \alpha_\ell \neq 1$. This gives the *long-run mean*

$$\mu_Y = \frac{\mu}{1 - \alpha_1 - \alpha_2}. \tag{4.11}$$

- This expression generalizes the expression (3.16) for first order autoregressions.

4.2.3 Interpretation: stochastic part

For analysis of the stochastic part of model (4.9), write the equation on *companion form*

$$\underbrace{\begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix}}_{X_t} = \underbrace{\begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix}}_{X_{t-1}} + \underbrace{\begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{\iota} \varepsilon_t. \tag{4.12}$$

Companion Matrix

Write this conformably in matrix form as

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \iota\varepsilon_t. \tag{4.13}$$

When does this equation have a stationary solution? In the first order case we needed the condition that $|\alpha| < 1$. Here, the condition is expressed in terms of the *eigenvalues* of the matrix $\mathbf{A}$. There are two eigenvalues solving the equation

$$\det(\mathbf{A} - \lambda I_2) = 0, \tag{4.14}$$

or equivalently

$$0 = (\alpha_1 - \lambda)(-\lambda) - (1)(\alpha_2) = \lambda^2 - \alpha_1\lambda - \alpha_2. \tag{4.15}$$

-

- The stability condition is that largest of the absolute values of the eigenvalues is less than unity. This is called the *spectral radius*. Thus, we get the condition

$$\rho(\mathbf{A}) = \max_{j=1,2} |\text{eigen}_j(\mathbf{A})| = \max_{j=1,2} |\lambda_j| < 1. \tag{4.16}$$

The solutions to (4.14) are referred to as the roots of the companion matrix. Negative roots and non-real roots are associated with seasonal patterns in the time series.

When the spectral radius of is less than unity, $\rho(\mathbf{A}) < 1$, and the innovations ε_t are i.i.d.$(0, \sigma^2)$ then it can be shown as for the first order autoregression that the stationary solution can be written as

$$\mathbf{X}_t^* \overset{\mathsf{D}}{=} \mathsf{N}(0, \Sigma_X) \quad \text{where} \quad \Sigma_X = \mathbf{A}\Sigma_X\mathbf{A}' + \sigma^2 \iota\iota', \tag{4.17}$$

where ι is a unit vector, see (4.12)/(4.13). The equation (4.17) can be derived along the lines of (3.22). It is linear in Σ_X and called a Lyapounov equation. It is usually solved numerically. The diagonal elements of Σ_X are constant and equal to the variance σ_X^2 of the stationary distribution of X_t^*.

It is common express the model equation (4.9) in terms of a lag operator L representing the lagging $L^j X_t = X_{t-j}$ and a lag polynomial $A(L)$ given by

$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \varepsilon_t \Leftrightarrow$
$$\varepsilon_t = L^0 X_t - \alpha_1 L^1 X_t - \alpha_2 L^2 X_t = (\underbrace{L^0 - \alpha_1 L^1 - \alpha_2 L^2}_{=A(L)})X_t. \tag{4.18}$$
Lag Polynomial $\to A(z)$ Characteristic Polynomial

Replacing L in the lag polynomial $A(L)$ by a complex number z gives the characteristic polynomial $A(z)$. The solutions to $A(z) = 0$ are the characteristic roots. They are the inverses of the companion matrix roots. So the stability condition can also be expressed as the condition that the characteristic roots have absolute value larger than unity.

$z = \frac{1}{\lambda}$

- Stability $\Longleftrightarrow$ All eigenvalues of the companion matrix lie within the spectral radius $\Longleftrightarrow$ Roots of the characteristic equation lie outside the spectral radius
    - Eigenvalues of this lie within the spectral radius:

$$\lambda^2 - \alpha_1 \lambda - \alpha_2 = 0$$

    - Roots of this (characteristic equation) lie outside the spectral radius:

$$1 - \alpha_1 z - \alpha_2 z^2 = 0$$

- Asymptotics:
  -

We can apply the standard asymptotic inference in the kth order autoregression as long as the spectral radius is less than unity, $\rho(\mathbf{A}) < 1$. We may want to test a linear hypothesis of dimension q on the parameters in the autoregressive equation (4.7). If $q = 1$ we can conduct inference using a t-test statistic. For general q we can apply F or LR test statistics. It can be shown that

$$t \overset{\mathsf{D}}{\to} \mathsf{N}(0,1), \qquad F \overset{\mathsf{D}}{\to} \chi_q^2/q, \qquad LR \overset{\mathsf{D}}{\to} \chi_q^2. \tag{4.19}$$

In particular, we can test whether the second lag parameter α_2 is needed, whether the dynamic parameters α_1, α_2 are needed at all, and whether the intercept μ is zero.

Determine Lag Length

Likelihood Ratio Test for Lag Length Determination #flashcard

A test for mis-specification can be done by fitting the alternative model explicitly and perform a likelihood ratio test. For instance, we could test a first order autoregression

$$H_0: \quad Y_t = \alpha_1 Y_{t-1} + \mu + \varepsilon_t \tag{4.22}$$

against the more general second order autoregression

$$H_a: \quad Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \mu + \varepsilon_t. \tag{4.23}$$

The likelihood ratio test statistic is

$$\mathsf{LR}_{static|ADL} = -2(\widehat{\ell}_{AR(1)} - \widehat{\ell}_{AR(2)}). \quad \sim \chi_2^2 \tag{4.24}$$

The LR-test statistic is asymptotically χ_1^2, where the degrees of freedom is found as the difference in the number of parameters in the two models. An example was given above.

The χ^2 distribution applies quite generally and does not involve constraints on the unknown dynamic parameters α_1, α_2. This is a great advantage as the lag length determination is then not dependent on knowing that, for instance, the roots are stationary. Paulsen (1984), Tsay (1984) consider stationary and unit roots. (Nielsen, 2006b) includes explosive roots.

- LR test does not require stationarity!

Residual Based Test for Lag Length Determination #flashcard

$H_0: Y_t = \alpha_1 Y_{t-1} + \varepsilon_t, \ \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$

(1978). Suppose we fit a first-order autoregression like (4.22), but we worry that the errors may be autocorrelated of order 1 so that

$H_a: Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \varepsilon_t, \ \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2) \Leftrightarrow$
$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \qquad t = 1, \ldots, T, \tag{4.25}$$

where u_t are i.i.d. We can test for autocorrelation by running the *auxiliary regression*

Aux Reg:
$$\widehat{\varepsilon}_t = \underbrace{\rho \widehat{\varepsilon}_{t-1}}_{\text{Test}=0} + \overbrace{\delta_0}^{\text{Extra Term}} + \delta_1 Y_{t-1} + v_t, \qquad t = 2, \ldots, T. \tag{4.26}$$

and testing $\rho = \delta_1 = 0$ using an F-test statistic or the TR^2-statistic. It holds that F and TR^2 are asymptotically χ_1^2, where the degrees of freedom is found as the number of lags in (4.25) rather than the number of restrictions in the auxiliary regression (4.26). The regressor Y_{t-1} from the original regression (4.22) is included in the auxiliary regression (4.26) to mimic that the partial estimator for Y_{t-2}, in second order autoregression (4.23) involves a correction for the future regressor Y_{t-1}. Wooldridge (2015, §12.2) motivates the inclusion of Y_{t-1} in the auxiliary regression (4.26) as adressing an endogeneity problem.

Information Criteria for Lag Length Determination #flashcard

Information criteria are used to estimate the lag-length k from a range of possible values $j = 0, 1, \dots, K$. Here, we consider an autoregression, but the discussion generalizes.

Start by keeping the first K observations as initial values. For each j estimate an autoregression of order j giving the likelihood value $\widehat{\ell}_{AR(j)}$. The information criteria is constructed by penalizing the likelihood using a penalty function $f(T)$ that only depends on sample size. It is customary to multiply the penalized likelihood by $-2/T$ to get

Information Criterion

$$\Phi_j = -\frac{2}{T}\left\{\widehat{\ell}_{AR(j)} - j\frac{f(T)}{2}\right\} = \log(2\pi) + \log(\widehat{\sigma}^2_{AR(j)}) + j\frac{f(T)}{T}. \tag{4.27}$$

→ Penalise more lags

→ Reward more lags

The lag length is then estimated by minimising the information criteria over j, that is

$$\widehat{k} = \arg\min_{j=0,1,\dots K} \Phi_j. \tag{4.28}$$

The penalty is introduced to compensate for the improvement in likelihood by including more regressors. The three most common choices of penalties are

$$\begin{aligned} &\text{Asymptotic Information Criteria } AIC: && f(T) = 2; \\ &\text{Hannan-Quinn Information Criteria } HQ: && f(T) = 2\log\log(T); \\ &\text{Bayesian Information Criteria } BIC: && f(T) = \log(T). \end{aligned}$$

} Minimise MSFE when k grows with T

} Consistent

AIC, BIC and HQ were suggested by Akaike (1973), Schwarz (1978), Hannan and Quinn (1979), respectively. When the data generating process is autoregressive with a fixed number of lags, k_0 say, then $\widehat{k}$ is consistent as long as $f(T) \to \infty$. BIC and HQ are of this type. HQ is designed to give the smallest rate to ensure strong consistency. AIC can be shown to minimize forecast errors when the lag length of the data generating process grows with T (Ing and Wei, 2005).

Misspecification Tests

Test for Normality #flashcard

Test for normality.

Test against skewness and excess kurtosis. Often attributed to Jarque and Bera (1987), although the test was introduced 100 years earlier. A time series implementation by Kilian and Demiroglu (2000) allows stationary roots and unit roots.

For a random variable X, let $Y = \{X - \mathsf{E}(X)\}/\mathsf{sdv}(X)$. Then $\kappa_3 = \mathsf{E}(Y^3)$ is the *skewness* and $\kappa_4 = \mathsf{E}(Y^4) - 3$ is the *excess kurtosis*. If X is $\mathsf{N}(\mu, \sigma^2)$ then $\kappa_3 = \kappa_4 = 0$. Obtain residuals $\widehat{\varepsilon}_t$ from a regression including an intercept and compute sample versions of $\widehat{\kappa}_3$ and $\widehat{\kappa}_4$. If the innovations are independent $\mathsf{N}(\mu, \sigma^2)$ then

$$T\frac{\widehat{\kappa}_3^2}{6} \overset{\mathsf{D}}{\to} \chi^2_1, \qquad T\frac{\widehat{\kappa}_4^2}{24} \overset{\mathsf{D}}{\to} \chi^2_1, \qquad T\frac{\widehat{\kappa}_3^2}{6} + T\frac{\widehat{\kappa}_4^2}{24} \overset{\mathsf{D}}{\to} \chi^2_2. \tag{4.30}$$

$H_0: \varepsilon_t \overset{iid}{\sim} N(\mu, \sigma^2)$

Test for Heteroskedasticity #flashcard

The White (1980) test for heteroskedasticity. Designed for cross sections. Applies to stationary time series and to unit root series with normal errors.

Example: 1 Obtain residuals from regression $Y_t = \mu + \beta X_t + \varepsilon_t$. Run auxiliary regression

2 Aux Reg :

$$(\widehat{\varepsilon}_t)^2 = \gamma_1 + \gamma_2 X_t + \gamma_3 X_t^2 + u_t. \tag{4.31}$$

→ All regressors and their squares and interaction terms

3 Test $\gamma_2 = \gamma_3 = 0$ using $TR^2 \overset{\mathsf{D}}{\to} \chi^2_2$. F-form is often used.

→ Thus F-test is a finite-sample correction

Test for Functional Form (RESET) #flashcard

Regression specification test (RESET) for functional form. Suggested by Ramsey (1969) for cross sections.

Example: 1 Obtain predictors from regression $Y_t = \mu + \beta X_t + \varepsilon_t$. Run auxiliary regression

2

$$Y_t = \gamma_1 + \gamma_2 X_t + \gamma_3 (\widehat{Y}_t)^2 + u_t. \tag{4.32}$$

3 Test $\gamma_3 = 0$ using a t-statistic $TZ^2 \overset{\mathsf{D}}{\to} \chi^2_1$.

$H_0: Y_t = \mu + \beta X_t + \varepsilon_t$

Test for Autoregressive Conditioinal Heteroskedasticity #flashcard

A test has been suggested by Engle (1982). It tests for time varying variances.

Example: 1 Obtain residuals from regression $Y_t = \mu + \beta X_t + \varepsilon_t$. Run auxiliary regression

2

$$(\widehat{\varepsilon}_t)^2 = \gamma_1 + \gamma_2 (\widehat{\varepsilon}_{t-1})^2 + u_t. \tag{4.33}$$

3 Test $\gamma_2 = 0$ using $TR^2 \overset{\mathsf{D}}{\to} \chi^2_1$.

Multiple Time Series

Autoregressive Distributed Lags (ADL)

Autoregression Distributed Lags (ADL) Model Setup, Estimation, and Asymptotics

- Setup, Estimation, Joint Log Likelihood, Asymptotics #flashcard
- Setup:

  The data consist of the dependent variable Y_t and an explanatory vector Z_t. We start with the special case of one lag so that the indices are labelled $t = 0, 1, \dots T$. If, in addition, Z_t is univariate, we get the model equation

  $$Y_t = \omega Z_t + \alpha Y_{t-1} + \beta Z_{t-1} + \mu + \varepsilon_t \qquad \text{for } t = 1, \dots, T. \tag{5.1}$$

  In the general case with k lages, the indices are labelled $t = 1-k, \dots, 0, 1, \dots T$. If, in addition, Z_t is a vector, we get the model equation

  ADL(k,k)

  $$Y_t = \omega' Z_t + \sum_{\ell=1}^{k} \alpha_\ell Y_{t-\ell} + \sum_{\ell=1}^{k} \beta' Z_{t-\ell} + \mu + \varepsilon_t \qquad \text{for } t = 1, \dots, T. \tag{5.2}$$

  An autoregressive distribution model is a conditional model. Introduce the filtration $\mathcal{G}_{t-1}$ to describe what we condition on at time t, which is $Z_t, Y_{t-1}, Z_{t-1}, \dots, Y_1, Z_1$ as well as the initial values $Y_0, Z_0, \dots, Y_{1-k} Z_{1-k}$. The model assumptions is then that ε_t is independent of $\mathcal{G}_{t-1}$ and $\mathsf{N}(0, \sigma^2)$ distributed.

- Estimation:
  1. OLS
  2. Partial MLE

- Joint Log Likelihood Derivation:

$$
\begin{aligned}
\text{Full Likelihood} &= f(Y^T, Z_T, Z_{t-1}, Z_{t-1}, \dots, Y_1, Z_1 | Y_0, Z_0) \\
&= \prod_{t=1}^{T} f(Y_t, Z_t | Y_{t-1}, Z_{t-1}, \dots, Y_0, Z_0) \\
&= \prod_{t=1}^{T} f(Y_t | Y_{t-1}, Z_{t-1}, \dots, Y_0, Z_0) f(Z_t | Y_{t-1}, Z_{t-1}, \dots, Y_0, Z_0) \\
&= \prod_{t=1}^{T} f(Y_t | Y_{t-1}, Z_{t-1}, \dots, Y_0, Z_0) \underbrace{\cancel{\prod_{t=1}^{T} f(Z_t | Y_{t-1}, Z_{t-1}, \dots, Y_0, Z_0)}}_{\text{Ignore with weak exogeneity}} \\
\text{Partial Likelihood} &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_t - \omega Z_t - \alpha Y_{t-1} - \beta Z_{t-1} - \mu)^2}{2\sigma^2} \right)
\end{aligned}
$$

- Then take log on the partial likelihood:

$$
\mathcal{L} = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (Y_t - \omega Z_t - \alpha Y_{t-1} - \beta Z_{t-1} - \mu)^2
$$

-

We estimate the autoregressive distributed lag model by least squares in the same fashion as for the first order autoregression, see §3.1.2. Further, using the prediction decomposition and the normality assumption we will show that the least squares estimators maximize a *partial likelihood*, in the sense of Cox (1975).

We find the density of $Y_T, Z_T, \dots, Y_1, Z_1$ given initial values and parameters. For simplicity, we focus on the one lag model (5.1). We first use the prediction decomposition (2.5) on the pairs y_t, z_t, second apply that $\mathsf{f}(y, z) = \mathsf{f}(y \mid z)\mathsf{f}(z)$ and third reorder to get

Prediction Decomposition

$$
\mathsf{f}(y_T, z_T, \dots, y_1, z_1 \mid y_0, z_0) = \prod_{t=1}^{T} \mathsf{f}(y_t, z_t \mid y_{t-1}, z_{t-1}, \dots, y_0, z_0) \tag{5.3}
$$

$$
= \prod_{t=1}^{T} \mathsf{f}(y_t \mid z_t, y_{t-1}, z_{t-1}, \dots, y_0, z_0) \mathsf{f}(z_t \mid y_{t-1}, z_{t-1}, \dots, y_0, z_0) \tag{5.4}
$$

$$
= \left\{ \prod_{t=1}^{T} \mathsf{f}(y_t \mid z_t, y_{t-1}, z_{t-1}, \dots, y_0, z_0 1) \right\} \left\{ \cancel{\prod_{t=1}^{T} \mathsf{f}(z_t \mid y_{t-1}, z_{t-1}, \dots, y_0, z_0)} \right\}. \tag{5.5}
$$

Ignore this part with weak exogeneity assumption
(parameters in these 2 ∏'s are ⊥)

- Asymptotics

  -

We can apply the standard asymptotic inference in autoregressive distributed lag models as long as the innovations are i.i.d. with 2+ moments and independent of the regressors, while the vector of all regressors is stationary with 2+ moments. The argument would be similar to what we have seen before.

Analogue to $H\beta = \theta$

Let θ is the vector of all regression parameters. Suppose, θ has dimension m. We can test an affine hypothesis of the form $R'\theta = \theta_R$ for a known $q \times m$ matrix R and a known q vector θ_R. If $q = 1$ we can conduct inference using a t-test statistic. For general q we can apply F or LR test statistics. We have

$$
t \overset{\mathsf{D}}{\to} \mathsf{N}(0,1), \qquad F \overset{\mathsf{D}}{\to} \chi^2_q / q, \qquad LR \overset{\mathsf{D}}{\to} \chi^2_q. \tag{5.8}
$$

Autoregression Distributed Lags (ADL) Model Interpretation: ML Mean and Error Correction Form

- Model:

$$
Y_t = \omega Z_t + \alpha Y_{t-1} + \beta Z_{t-1} + \mu + \epsilon_t \tag{ADL(1,1)}
$$

- Find the ECF, and LR Mean #flashcard
- Error Correction Form:

$$
\Delta Y_t = \omega \Delta Z_t \underbrace{-(1-\alpha)}_{\pi} \underbrace{\left\{ Y_{t-1} - \underbrace{\frac{\omega + \beta}{1 - \alpha}}_{\kappa} Z_{t-1} - \underbrace{\frac{\mu}{1-\alpha}}_{\tau} \right\}}_{ecm_{t-1}} + \epsilon_t
$$

  - where:
    - ω is the SR impact coefficient
    - ecm_{t-1} measures the dis-equilibrium
    - π is the adjustment to disequilibrium (stationary systems typically have $\pi < 0$)
  - Derivation

The long-run means relation (5.11) can also be found through reparametrisation of the autoregressive distributed lags model equation (5.1) in equilibrium correction form. Recall the model equation

$$Y_t = \omega Z_t + \alpha Y_{t-1} + \beta Z_{t-1} + \mu + \varepsilon_t. \tag{5.12}$$

$Y_t - Y_{t-1} = wZ_t - wZ_{t-1} + \alpha Y_{t-1} + \beta Z_{t-1} + \mu - Y_{t-1} + wZ_{t-1} + \varepsilon_t$

① Get ΔY_t, ΔZ_t

Subtract Y_{t-1} on both sides, add and subtract ωZ_{t-1} on the right hand side, introduce the notation for the growth rates $\Delta Y_t = Y_t - Y_{t-1}$ and $\Delta Z_t = Z_t - Z_{t-1}$ to get

$$\Delta Y_t = \omega \Delta Z_t \underbrace{- (1-\alpha) Y_{t-1} + (\omega + \beta) Z_{t-1} + \mu}_{\text{matches (5.10)}} + \varepsilon_t. \tag{5.13}$$

$\Delta Y_t = w\Delta Z_t - (1-\alpha) Y_{t-1} + (\beta + w) Z_{t-1} + \mu + \varepsilon_t$

② Use a factor to normalize the coefficient on Y_{t-1}

The underbraced terms relate to levels of the processes Y_{t-1} and Z_{t-1} and the intercept μ. Assuming $\alpha \neq 1$, we can take common factor $1-\alpha$ to get

$$\Delta Y_t = \omega \Delta Z_t - (1-\alpha)\Big(Y_{t-1} - \underbrace{\frac{\omega+\beta}{1-\alpha}}_{=\kappa} Z_{t-1} - \underbrace{\frac{\mu}{1-\alpha}}_{=\tau}\Big) + \varepsilon_t.$$

$\Delta Y_t = w\Delta Z_t \underbrace{- \frac{1}{1-\alpha}}_{\pi} \Big(Y_{t-1} - \overbrace{\frac{\beta + w}{1-\alpha}}^{\kappa} Z_{t-1} - \underbrace{\frac{\mu}{1-\alpha}}_{\tau} \Big) + \varepsilon_t$

$\Delta Y_t = w\Delta Z_t - \pi \overbrace{(Y_{t-1} - \kappa Z_{t-1} - \tau)}^{ecm} + \varepsilon_t$

③ SR Impact Coefficient

Renaming the parameters we get the *equilibrium correction form*

$$\Delta Y_t = \underbrace{\omega}_{\text{SR Dynamics}} \Delta Z_t + \pi \underbrace{(Y_{t-1} - \kappa Z_{t-1} - \tau)}_{=ecm_{t-1}} + \varepsilon_t, \tag{5.14}$$

$\to (\alpha - 1)$: it measures the speed of adjustment towards LR equilibrium

with *equilibrium-correction mechanism*

All terms here LR Mean $= 0$ => Relations between LR Means

$$ecm_{t-1} = Y_{t-1} - \kappa Z_{t-1} - \tau. \tag{5.15}$$

$\pi < 0$ Converge
$\pi = 0$ No Correction
$\pi > 0$ Diverge

The long-run mean of the equilibrium-correction mechanism is zero due to (5.11). At the same time the growth rates ΔY_t and ΔZ_t have long-run means of zero.

The equilibrium correction form shows how ΔY_t adjusts when Y_{t-1} and Z_{t-1} deviate from the equilibrium of their long-run means. The dis-equilibrium is positive, $ecm_{t-1} > 0$, if Y_{t-1} is to large relative to Z_{t-1}. In stationary systems the adjustment coefficients $\pi = \alpha - 1$ is usually negative. Accordingly, ΔY_t responds negatively to a positive dis-equilibrium $ecm_{t-1} > 0$.

We say that the parameters π and ω are short-run adjustment coefficients. Since we do only have a partial model we cannot say anything about the joint adjustment of the system of Y_t and Z_t to dis-equilibrium.

- Long-run Mean
  -

We can interpret an autoregressive distributed lag model in terms of a relation between the long-run means of the regressand Y_t and the regressor Z_t. Working with the simpler model (5.1) and denoting the long-run means by μ_Y and μ_Z, respectively, we get

$$(Y_t - \mu_Y) = \omega(Z_t - \mu_Z) + \alpha(Y_{t-1} - \mu_Y) + \beta(Z_{t-1} - \mu_Z) + \underbrace{\mu - (1-\alpha)\mu_Y + (\omega+\beta)\mu_Z}_{\text{Set to 0}} + \varepsilon_t. \tag{5.9}$$

Transform all variables to their deviation from the LR mean and set the remaining terms to 0

40

  -

This reduces to a homogenous equation in the variables $Y_t - \mu_Y$ and $Z_t - \mu_Z$ and without deterministic terms if the intercept and long-run means μ_Y and μ_Z satisfy the relation

$$(1-\alpha)\mu_Y = (\omega+\beta)\mu_Z + \mu, \tag{5.10}$$

Jointly Identified with 2DOF

or equivalently

$$\mu_Y = \underbrace{\frac{\omega+\beta}{1-\alpha}}_{=\kappa}\mu_Z + \underbrace{\frac{\mu}{1-\alpha}}_{=\tau}. \tag{5.11}$$

This is an equilibrium relation for the long-run means implied by the autoregressive distributed lag model. When the variables Y_t and Z_t are stationary then (5.11) is a relation between the expected values. It is not possible to identify the individual long-run means μ_Y and μ_Z from the intercept μ in the partial model.

Vector Autoregressions (VARs)

Vector Autoregressions (VARs) Model Setup, Estimation, Interpretation, and Asymptotoics

- Setup, Interpretation, Estimation and Inference #flashcard
- Setup:

We consider the first order vector autoregressive model, also called the $VAR(1)$ model. The available data is a time series of p-dimensional vectors $X_0, X_1, \ldots, X_T$. Note, we are now using the notation X_t for a different purpose than earlier. The model equation is

$$X_t = AX_{t-1} + \mu + \varepsilon_t \qquad t = 1, \ldots T. \tag{5.22}$$

Repeat the equation with an indication of the dimensions of variables and parameters:

VAR(1)

$$\underbrace{X_t}_{p\times 1} = \underbrace{A}_{p\times p}\underbrace{X_{t-1}}_{p\times 1} + \underbrace{\mu}_{p\times 1} + \underbrace{\varepsilon_t}_{p\times 1}. \tag{5.23}$$

We assume that ε_t is $\mathsf{N}_p(0, \Omega)$ distributed and independent of the past vectors $(X_{t-s})_{s\geq 1}$.

The innovation variance is now a square matrix. In the two-dimensional case then

Cov-Var(ε_t)

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}. \tag{5.24}$$

It is common to report standard deviations and correlations:

$$\sigma_1 = \sqrt{\Omega_{11}}, \qquad \sigma_2 = \sqrt{\Omega_{22}}, \qquad \rho_{12} = \frac{\Omega_{12}}{\sqrt{\Omega_{11}\Omega_{22}}}. \tag{5.25}$$

- Interpretation

The interpretation of vector autoregressions is similar to that of scalar autoregressions. For a first order vector autoregression as here, the stability condition is that the spectral radius of the coefficient matrix A is less than unity, $\rho(A) < 1$, see (4.16).

We can solve the autoregressive equation recursively and get the stationary solution, for $t \to \infty$,

$$(X_t|X_0) \overset{\mathsf{D}}{\to} X^* = \sum_{j=0}^{\infty} A^j(\varepsilon_{-j} + \mu) \overset{\mathsf{D}}{=} \mathsf{N}(\mu_X, \Sigma_X), \tag{5.26}$$

where the mean and variance of the stationary distribution are given by

$$\mu_X = (I_p - A)^{-1}\mu, \qquad \Sigma_X = \sum_{j=0}^{\infty} A^j \Omega (A^j)'. \tag{5.27}$$

Similar as before

The variance solves the Lyapounov equation, $\Sigma_X = A\Sigma_X A' + \Omega$, see (4.17).

$\to$ Stationarity + VAR Equation

- Estimation and Inference
  - Unrestricted $\to$ Unpack and estimate by OLS; inference with standard tools

-

An unrestricted vector autoregression is estimated by regression. As an example consider the bivariate case. The model equation (5.22) is then

Unpooled VAR

$$Y_{1,t} = \alpha_{11}Y_{1,t-1} + \alpha_{12}Y_{2,t-1} + \mu_1 + \varepsilon_{1,t}, \tag{5.28}$$
$$Y_{2,t} = \alpha_{21}Y_{1,t-1} + \alpha_{22}Y_{2,t-1} + \mu_2 + \varepsilon_{2,t}. \tag{5.29}$$

The equations (5.28) and (5.29) are estimated separately by least squares. The standard deviations σ_1 and σ_2 in (5.25) are estimated by the two residual standard deviations. The correlation ρ_{12} is estimated by the sample correlation between the residuals from the two regressions. The estimators can be shown to be maximum likelihood, see §A.5 or Lütkepohl (2005, §3.4).

We can apply standard inference in the stationary vector autoregressions. That is, t-statistics are asymptotically standard normal and LR-statistics are asymptotically χ^2.

- Testing

-

5.2.4 Linear restrictions → Regression + LR Test

Linear restrictions on a vector autoregression are restriction that affect all equations in the same way, such as dropping an explanatory variable from all equations. In that situation the restricted model estimated by regression.

Example: Fulton. In the vector autoregression (5.30), (5.31) it appears that the lagged quantities, q_{t-1}, are insignificant in both equations. Re-estimate to get

$$\underset{(\text{s.e.})}{\widehat{p}_t} = +\underset{(0.06)}{0.67}p_{t-1} - \underset{(0.04)}{0.17} + \underset{(0.06)}{0.25}S_t + \underset{(0.05)}{0.14}M_t + \underset{(0.14)}{0.12}H_t, \tag{5.33}$$
$$\underset{(\text{s.e.})}{\widehat{q}_t} = +\underset{(0.17)}{0.37}p_{t-1} + \underset{(0.11)}{8.88} - \underset{(0.15)}{0.56}S_t - \underset{(0.14)}{0.24}M_t - \underset{(0.37)}{2.16}H_t, \tag{5.34}$$
$$\widehat{\sigma}_p = 0.24, \quad \widehat{\sigma}_q = 0.63, \quad \widehat{\rho}_{pq} = -0.45, \qquad \widehat{\ell} = -84.74, \qquad T = 110.$$

The likelihood ratio test statistic $\mathsf{LR} = -2(-84.74 + 83.11) = 3.2$ is small when judging against a χ^2_2-distribution with $\mathsf{p} = 0.20$.

5.2.5 Cross-equation restrictions → MLE + (LR Test ?)

Cross equation restrictions are restrictions that link the parameters from different equations. For such restrictions, least squares estimation is no longer maximum likelihood. Nonetheless, inference remains standard.

As an example, consider the coefficients for the weather variables in (5.33), (5.34):

$$\widehat{\Gamma}\begin{pmatrix} S_t \\ M_t \end{pmatrix} = \begin{pmatrix} \widehat{\gamma}_{p,s} & \widehat{\gamma}_{p,m} \\ \widehat{\gamma}_{q,s} & \widehat{\gamma}_{q,m} \end{pmatrix}\begin{pmatrix} S_t \\ M_t \end{pmatrix} = \begin{pmatrix} \underset{(0.06)}{0.25} & \underset{(0.05)}{0.14} \\ \underset{(0.15)}{-0.57} & \underset{(0.14)}{-0.25} \end{pmatrix}\begin{pmatrix} S_t \\ M_t \end{pmatrix}. \tag{5.35}$$

It appears that the second row is approximately minus twice the first row. This indicates an approximate linear dependence of the rows of this coefficient matrix. Imposing exact linear dependence on the rows of Γ is a non-linear cross-equation restriction. This is also known as a reduced rank restriction. It implies that a linear relation of prices and quantities does not depend on the weather variables. That relation can be thought of as demand function with the weather variables as instruments. Persuing this idea leads to a maximum likelihood approach to instrument variable estimation. This is known as *limited information maximum likelihood* (LIML) or *reduced rank regression* (Hendry and Nielsen, 2007, §14) carries this idea through for the Fulton fish data example. → $\text{Rank}(A) = 2 \xrightarrow{\text{Restrict}} \text{Rank}(A) = 1$

Conditioning Properties of Bivariate Normal Distributions

- Let:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}\right)$$

and $\rho = \frac{\Omega_{12}}{\sqrt{\Omega_{11}\Omega_{22}}} \iff \Omega_{12} = \Omega_{21} = \rho\sqrt{\Omega_{11}\Omega_{22}}$

- Then, we have the following **Conditional Distribution**: #flashcard
  -

$$(Y_1|Y_2 = y_2) \sim N\left(\mu_{1|2}, \Omega_{1|2}\right)$$

where

$$\begin{cases} \mu_{1|2} & = \mu_1 + \frac{\Omega_{12}}{\Omega_{22}}(Y_2 - \mu_2) \\ \Omega_{1|2} & = \Omega_{11}(1 - \rho^2) \end{cases}$$

- More succinctly:

$$(Y_1|Y_2) \sim N\left(\mu_1 + \frac{\Omega_{12}}{\Omega_{22}}(Y_2 - \mu_2), \Omega_{11}(1 - \rho^2)\right)$$

- Note that conditioning reduces uncertainty(variance).

VAR → ADL by Conditioning #flashcard

☆ Conditioning Properties of BNV

$\rho = \frac{\Omega_{12}}{\sqrt{\Omega_{11}\Omega_{22}}}$

$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N_2\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Omega_{11}, & \Omega_{12} \\ \Omega_{21}, & \Omega_{22} \end{bmatrix}\right)$, $\Omega_{12} = \rho\sqrt{\Omega_{11}\Omega_{22}}$, $\Omega_{21} = \rho\sqrt{\Omega_{11}\Omega_{22}}$

Conditional Distribution

$(Y_1 | Y_2 = y_2) \sim N(\mu_{1|2}, \Omega_{1|2})$

where $\mu_{1|2} = \mu_1 + \frac{\Omega_{12}}{\Omega_{22}}(y - \mu_2)$

$\Omega_{1|2} = \Omega_{11}(1 - \rho^2)$, $\rho = \frac{\Omega_{12}}{\sqrt{\Omega_{11}\Omega_{22}}}$

* Conditioning reduces uncertainty.

The linear transformation comes about by exploiting the conditioning property of the bivariate normal distribution. Suppose

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \overset{\mathsf{D}}{=} \mathsf{N}_2\left\{\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\right\}, \tag{5.37}$$

This is not ρ!

Defining the population regression coefficient $\omega = \Omega_{12}\Omega_{22}^{-1}$ and the conditional variance $\sigma^2 = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$, we get by a linear transformation

Or $\Omega_{11}(1 - \rho^2)$

$$\begin{pmatrix} Y_1 - \omega Y_2 \\ Y_2 \end{pmatrix} \overset{\mathsf{D}}{=} \mathsf{N}_2\left\{\begin{pmatrix} \mu_1 - \omega\mu_2 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \Omega_{22} \end{pmatrix}\right\}. \tag{5.38}$$

Let the joint density of $Z = Y_1 - \omega Y_2$ and Y_2 be denoted $\mathsf{f}(z, y_2)$. In general we have $\mathsf{f}(z, y_2) = \mathsf{f}(z|y_2)\mathsf{f}(y_2)$. Thus, (5.38) shows that the conditional distribution of Z given Y_2 is $\mathsf{N}(\mu_1 - \omega\mu_2, \sigma^2)$. Since $Y_1 = Z - \omega Y_2$ and Y_2 is fixed when conditioning, we deduce that

- Conditional Distribution $(Y_1 \mid Y_2) \overset{\mathsf{D}}{=} \mathsf{N}\{\mu_1 + \omega(Y_2 - \mu_2), \sigma^2\}. \tag{5.39}$
  - Apply:

$$(Y_1|Y_2) \sim N\left(\mu_1 + \frac{\Omega_{12}}{\Omega_{22}}(Y_2 - \mu_2), \Omega_{11}(1 - \rho^2)\right)$$

- Example:
  -

Example: Fulton prices. The estimated vector autoregression (5.30), (5.31) was

VAR

$$\underset{(\text{s.e.})}{\widehat{p}_t} = \overbrace{\underset{(0.07)}{0.65}p_{t-1} - \underset{(0.03)}{0.03}q_{t-1} + \underset{(0.28)}{0.10} + \underset{(0.06)}{0.25}S_t + \underset{(0.05)}{0.14}M_t + \underset{(0.14)}{0.09}H_t}^{\mu_1}, \tag{5.40}$$

$$\underset{(\text{s.e.})}{\widehat{q}_t} = \underbrace{\underset{(0.17)}{0.45}p_{t-1} + \underset{(0.09)}{0.15}q_{t-1} + \underset{(0.72)}{7.64} - \underset{(0.15)}{0.57}S_t - \underset{(0.14)}{0.25}M_t - \underset{(0.37)}{2.02}H_t}_{\mu_2}, \tag{5.41}$$

$$\widehat{\sigma}_p = 0.24, \quad \widehat{\sigma}_q = 0.62, \quad \widehat{\rho}_{pq} = -0.44, \quad \widehat{\ell}_{p,q} = -83.11, \quad T = 110.$$

-0.065472 $\quad \omega = \frac{-0.065472}{0.24^2} = 1.15$

$$\begin{pmatrix} 0.24^2 & -0.44\times 0.24\times 0.62 \\ -0.44\times 0.24\times 0.62 & 0.62^2 \end{pmatrix}$$

Recall also the estimated autoregressive distributed lag model (5.16) given by

ADL

$\mu_2 + \omega(p_t - \mu_1) = \mu_2 - 1.15(p_t - \mu_1)$

$$\underset{(\text{s.e.})}{\widehat{q}_t} = \underset{(0.23)}{-1.15}p_t + \underset{(0.22)}{1.20}p_{t-1} + \underset{(0.08)}{0.11}q_{t-1} + \underset{(0.65)}{7.75} - \underset{(0.15)}{0.28}S_t - \underset{(0.13)}{0.09}M_t - \underset{(0.34)}{1.91}H_t, \tag{5.42}$$

ω $\quad 0.45 - (-1.15)\times 0.65$

$$\widehat{\sigma}_{q|p} = 0.56, \quad \widehat{\ell}_{q|p} = -88.66, \quad T = 110.$$

We can derive the latter equation (5.42) from the vector autoregressive equations (5.40), (5.41) as follows. Matching these equations with (5.37) we let μ_1, μ_2 represent the right hand side of (5.40), (5.41). Moreover $\Omega_{11} = \widehat{\sigma}_p^2$, $\Omega_{22} = \widehat{\sigma}_q^2$ and $\Omega_{12} = \widehat{\rho}_{pq}\widehat{\sigma}_p\widehat{\sigma}_q$. Thus, from (5.39), we can deduce (5.42) so that $\omega = \Omega_{21}/\Omega_{11} = \widehat{\rho}_{pq}\widehat{\sigma}_q/\widehat{\sigma}_p$ is found as the coefficient to p_t in (5.42) and $s^2 = \Omega_{22} - \Omega_{12}^2/\Omega_{11}$ is found as $\widehat{\sigma}_{q|p}$.

Now, estimate the marginal model for p_t alone to get

Marginal Model

$$\underset{(\text{s.e.})}{\widehat{p}_t} = \underset{(0.07)}{0.65}p_{t-1} - \underset{(0.03)}{0.03}q_{t-1} + \underset{(0.28)}{0.10} + \underset{(0.06)}{0.25}S_t + \underset{(0.05)}{0.14}M_t + \underset{(0.14)}{0.09}H_t, \tag{5.43}$$

$$\widehat{\sigma}_p = 0.24, \quad \widehat{\ell}_p = 5.54, \quad T = 110.$$

The estimates match those of the first vector autoregressive equation (5.40).

We see that in line with (5.36) the maximum likelihood values satisfy the relation

$$\widehat{\ell}_{p,q} = -83.11 = 5.54 - 88.66 = \widehat{\ell}_p + \widehat{\ell}_{q|p}.$$

This comes about because the equations (5.42), (5.43) are estimated without any cross equation restrictions. The cross product of residuals from the two equations is therefore zero by construction.

Super Exogeneity, Causality, and Identification in VAR #flashcard

A variable X is **super exogenous** for the parameters of interest if:

1. X is **weakly exogenous** (it helps estimate the parameters of the conditional model), **and**
2. The **parameters of the conditional model** (e.g., $Y|X$) **remain invariant** when the distribution of X changes (e.g., due to policy).

> In short: **You can use X in a model to predict Y *even if a policy shock changes X*** — because the relationship between Y and X doesn't change.

Unit Roots

Unit Root/Random Walk with a Level Setup, Estimation, Interpretation, and Equilibrium Correction Form #flashcard

- Setup

We repeat the autoregressive model. The data is the univariate time series $Y_0, Y_1, \ldots, Y_T$. The first order autoregression is

Unit Root: $Y_t = Y_{t-1} + \mu_c + \varepsilon_t$

$$Y_t = \alpha Y_{t-1} + \mu_c + \varepsilon_t, \qquad \alpha = 1 \tag{6.3}$$

assuming the innovations ε_t are $\mathsf{N}(0, \sigma^2)$ and independent of the past, $Y_{t-1}, Y_{t-2}, \ldots, Y_1, Y_0$.

We know that model has a stationary solution when $|\alpha| < 1$. Here, we will be interested in the case where $\alpha = 1$. This is referred to as the unit root case, since the characteristic polynomial $\lambda - \alpha = 0$ has a root at unity when $\alpha = 1$.

When finding the long-run mean in stationary series, we have seen that precisely the case $\alpha = 1$ brings complications. There is an interesting interaction between deterministic terms and random walks. We will therefore argue that the hypothesis of interest is not simply $\alpha = 1$ but rather the joint hypothesis

$$\text{Random Walk:} \quad \alpha = 1, \qquad \mu_c = 0. \qquad Y_t = Y_{t-1} + \varepsilon_t \quad \alpha = 1,\ \mu_c = 0 \tag{6.4}$$

In any case, we can estimate the unrestricted model (6.3) by least squares. We will consider a t-test statistic for $\alpha = 1$ and an F-test statistic for $\alpha = 1$, $\mu_c = 0$.

- Interpretation:
  - Unobserved Components Formation

When interpreting the stationary autoregression we derived an unobserved components representation. Here, we will start from the unobserved components representation

$$Y_t = X_t + \tau_c, \tag{6.5}$$

$$X_t = \alpha X_{t-1} + \varepsilon_t. \tag{6.6}$$

For the latent autoregressive equation (6.6) the natural unit root hypothesis is $\alpha = 1$.

We derive an autoregressive representation for Y_t with a view to understanding the joint hypothesis in (6.4). First, solve (6.5) for X_t to get $X_t = Y_t - \tau_c$. Second, insert this in (6.6) to get

$$(Y_t - \tau_c) = \alpha(Y_{t-1} - \tau_c) + \varepsilon_t. \tag{6.7}$$

Third, rearrange as

$$Y_t = \alpha Y_{t-1} + \underbrace{(1-\alpha)\tau_c}_{=\mu_c} + \varepsilon_t. \tag{6.8}$$

We see that when $\alpha = 1$ in (6.6) then the intercept in (6.8) satisfies $\mu_c = (1-\alpha)\tau_c = 0$. Thus, the joint hypothesis $\alpha = 1$, $\mu_c = 0$ is natural in the autoregressive equation.

Further, under the joint hypothesis $\alpha = 1$, $\mu_c = 0$ we get that the Y_t process from (6.3) has the form

$$H_{RW}: Y_t = Y_{t-1} + \varepsilon_t \iff Y_t = Y_0 + \sum_{s=1}^{t} \varepsilon_s. \tag{6.9}$$

This corresponds to letting $X_0 + \tau_c = Y_0$ in (6.5) when $\alpha = 1$. The latent X_0 and the level τ_c are not separately identifiable.

Finally, when $\alpha = 1$ but we do not necessarily have $\mu_c = 0$ then the Y_t process from (6.3) has the form

$$H_{UR}: Y_t = Y_{t-1} + \mu_c + \varepsilon_t \iff Y_t = \underbrace{\sum_{s=1}^{t} \varepsilon_s}_{\textit{random walk}} + \underbrace{Y_0 + \mu_c t}_{\textit{linear trend}} \tag{6.10}$$

This is a rather extreme outcome, considering that the process is stationary around a level when $|\alpha| < 1$. This is another reason for preferring the joint hypothesis (6.4) that $\alpha = 1$ and $\mu_c = 0$. In other words, the parameter combination $\alpha = 1$ and $\mu_c \neq 0$ appears in the 'alternative hypothesis', but not in the 'null hypothesis'.

  - Equilibrium Correction Form

It is convenient to reparametrise the model equation (6.3) in equilibrium correction form. Subtract Y_{t-1} on both sides of (6.3) to get

$$\underbrace{Y_t - Y_{t-1}}_{=\Delta Y_t} = \underbrace{(\alpha - 1)}_{=\pi}Y_{t-1} + \mu_c + \varepsilon_t.$$

This brings the model into equilibrium correction form

$$\Delta Y_t = \pi Y_{t-1} + \mu_c + \varepsilon_t. \tag{6.11}$$

We can then write the unit root hypothesis (6.4) as

$$H_0: \text{Random Walk} \iff \pi = \mu_c = 0. \tag{6.12}$$

Under this hypothesis, the model equation reduces to $\Delta Y_t = \varepsilon_t$, so that the differenced process is stationary and even i.i.d. here.

- Estimation and Inference
  - We can still estimate using least squares/MLE, but specialised inference tools are needed
  -

In the model with intercept

$$\Delta Y_t = \pi Y_{t-1} + \mu_c + \varepsilon_t, \tag{6.22}$$

the unit root hypothesis is

H_{UR} in FD :
$$\pi = \mu_c = 0. \tag{6.23}$$

Limiting distributions of least squares statistics can be expressed in terms of Brownian motions as above. Here, we will just note that the likelihood ratio test statistic for the joint hypothesis $\pi = \mu_c = 0$ converges in distribution as

(preferred over t-test)

$$\mathsf{LR}_{\pi=\mu_c=0} \overset{\mathsf{D}}{\to} \mathsf{DF}_c^2. \tag{6.24}$$

Here DF honours Dickey and Fuller (1979, 1981). The super-script of 2 is as in the χ^2 distribution. The sub-script "c" is to indicate that it is the limiting distribution for the model with constant level. The distribution is tabulated in Table 6.2, which is extracted from (Johansen, 1995, § 15).

  -

H_{RW} in FD

Alternatively we could use the t-test statistic for testing $\pi = 0$. Its asymptotic distribution depends on whether $\mu_c = 0$ or not. Usually, we use the distribution assuming $\pi = \mu_c = 0$ in light of the discussion of (6.10). We get

$$\mathsf{t}_{\pi=0} \overset{\mathsf{D}}{\to} \mathsf{DF}_c \qquad \text{assuming } \pi = \mu_c = 0. \tag{6.25}$$

Unit Root/Random Walk with a Linear Time Trend #flashcard

6.4.1 The model with a linear trend UR/RW + Time Trend

For GDP data we need a model with an intercept and a linear trend. The model is

$$\Delta Y_t = \pi Y_{t-1} + \mu_c + \mu_\ell t + \varepsilon_t, \tag{6.26}$$

with unit root hypothesis

H_{UR} w/ Trend :
$$\pi = \mu_\ell = 0. \tag{6.27}$$

Interpretation: Under the unit root hypothesis the trend model satisfies

H_{UR} w/T (alt):
$$Y_t = \underbrace{\sum_{s=1}^{t} \varepsilon_s}_{\text{random walk}} + \underbrace{Y_0 + \mu_c t}_{\text{linear trend}}. \tag{6.28}$$

Inference: We can test the hypothesis $\pi = \mu_\ell = 0$ using a likelihood ratio test statistic or a t-statistic. We get a new set of limit distributions.

Test $H_{UR/RW}$ with Trend

$$\mathsf{LR}_{\pi=\mu_\ell=0} \overset{\mathsf{D}}{\to} \mathsf{DF}_\ell^2, \tag{6.29}$$

DF adjusted for linear trend

$$\mathsf{t}_{\pi=0} \overset{\mathsf{D}}{\to} \mathsf{DF}_\ell \qquad \text{assuming } \pi = \mu_\ell = 0. \tag{6.30}$$

- Note that test for RW/UR is one-sided!
  - Example:

Consider the UK GDP series shown in Figure 1.1. There are some rather large fluctuations from 2020Q1 when the pandemic started. Thus, we focus on the period 2009Q4 to 2019Q4. We will return to modelling post 2019Q4 in §8.

Estimate an unrestricted $AR(1)$ in equilibrium correction form to get

$$\underset{\text{se}}{\Delta \widehat{x}_t} = -\underset{(0.086)}{0.129} x_{t-1} + \underset{(1.02)}{1.54} + \underset{(0.00045)}{0.00065} t \tag{6.41}$$

$$\widehat{\sigma} = 0.0031, \quad \widehat{\ell} = 179.757, \quad R^2 = 0.060, \quad T = 41.$$

$$\chi^2_{\mathsf{norm}}[2] = 1.6[0.44] \qquad F_{\mathsf{ar}(1-3)}[3, 35] = 2.6[0.07]$$

$$F_{\mathsf{het}}[4, 36] = 1.5[0.23] \qquad F_{\mathsf{arch}(1-3)}[3, 35] = 2.4[0.08]$$

$-1.51 > -3.41$

The t-test statistic for a unit root is $-0.129/0.086 = -1.51$. Compare with the DF_ℓ-distribution with 5% critical value of -3.41. Unit root is not rejected. This assumes the coefficient to the linear trend is insignificant - but for this test it is insisted not to test it!

The restricted model with a unit root and linear trend term is

$$\underset{\text{se}}{\Delta \widehat{x}_t} = +\underset{(0.00049)}{0.00489} \tag{6.42}$$

$$\widehat{\sigma} = 0.0032, \quad \widehat{\ell} = 178.481, \quad T = 41.$$

The likelihood ratio test statistic is $\mathsf{LR} = -2(179.757 - 178.481) = 2.55$. Compare with the DF_ℓ^2-distribution with 95% quantile of 12.39. The hypothesis cannot be rejected.

Unit Root in Higher Order Autoregressions Setup, Interpretation, and Inference, ADF Test, I(2) Process #flashcard

Formulation

Consider a second order autoregression with an intercept

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \mu_c + \varepsilon_t. \tag{6.31}$$

$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \mu_c + \varepsilon_t$
$-Y_{t-1} \quad -Y_{t-1} \quad +\alpha_2 Y_{t-1} \quad -\alpha_2 Y_{t-1}$
$\underbrace{Y_t - Y_{t-1}}_{\Delta Y_t} = \underbrace{(\alpha_1 + \alpha_2 - 1)}_{\pi} Y_{t-1} + \alpha_2 \underbrace{(Y_{t-2} - Y_{t-1})}_{\Delta Y_{t-1}} + \mu_c + \varepsilon_t$

In the context of a unit root, this can be reformulated as

$$\Delta Y_t = \pi Y_{t-1} + \gamma \Delta Y_{t-1} + \mu_c + \varepsilon_t, \tag{6.32}$$

where $\pi = \alpha_1 + \alpha_2 - 1$ and $\gamma = -\alpha_2$.

Interpretation: If $\pi = \mu_c = 0$ then

UR AR(2)

$$\Delta Y_t = \gamma \Delta Y_{t-1} + \varepsilon_t, \qquad \text{(UR AR(2) FD)} \tag{6.33}$$

which is a first order autoregression. Thus, if $|\gamma| < 1$ then the differenced series, ΔY_t, can be given a stationary representation. To get at a representation for the levels of the series, rewrite (6.33) by subtracting $\gamma \Delta Y_t$ on both sides. This gives

Interpretation

$$(1-\gamma)\Delta Y_t = -\gamma \Delta^2 Y_t + \varepsilon_t, \tag{6.34}$$

where $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$ is a second difference. Replace t by s and cumulate over $s = 1, \dots t$ noting that $\sum_{s=1}^t \Delta Y_s = Y_t - Y_0$ and $\sum_{s=1}^t \Delta^2 Y_s = \Delta Y_t - \Delta Y_0$ to get

$$(1-\gamma)(Y_t - Y_0) = -\gamma(\Delta Y_t - \Delta Y_0) + \sum_{s=1}^t \varepsilon_s. \tag{6.35}$$

Thus, Y_t is expressed in terms of ΔY_t, which has a stationary representation, a random walk and initial values. Reorganizing, we get a representation of the form

UR in AR(2) ⇔

$$Y_t = C \sum_{s=1}^t \varepsilon_s + \text{stationary process} + \text{level}, \tag{6.36}$$

where $C = 1/(1-\gamma)$.

If $\gamma = 1$ then the differenced series, ΔY_t is itself a random walk:

Double UR in AR(2) / I(2) :

$$\Delta Y_t = \Delta Y_0 + \sum_{r=1}^t \varepsilon_r. \tag{6.37}$$

Replace t by s and cumulate over s to get

$$Y_t = Y_0 + t\Delta Y_0 + \sum_{s=1}^t \sum_{r=1}^s \varepsilon_r. \tag{6.38}$$

This is a cumulated random walk combined with a linear trend. The series is said to be integrated of order 2 – or I(2) in short – since double differencing is needed to achieve stationarity.

Inference

Inference: Stationary and random walk components can be argued to be asymptotically uncorrelated. ② Thus, we get the 'usual' unit root distributions when testing $\pi = \mu_c = 0$:

(DF distributions)

Testing UR → ADF

$$\mathsf{LR}_{\pi=\mu_c=0} \xrightarrow{\mathsf{D}} \mathsf{DF}_c^2, \qquad \mathsf{t}_\pi \xrightarrow{\mathsf{D}} \mathsf{DF}_c \qquad \text{when } \pi = \mu_c = 0. \tag{6.39}$$

The t-type test is often called the *augmented Dickey-Fuller test*. The asymptotics in (6.39) actually applies as long as $\gamma \neq 1$ (Nielsen, 2001). When $\gamma = 1$ is a realistic possibility in empirical work one may proceed a long the lines of Pantula (1989). ② Test statistics for hypotheses that do not alter the number of unit roots or the degree of a deterministic polynomial are asymptotically normal. For instance, the likelihood ratio test statistic for the lag length hypothesis $\gamma = 0$ satisfies

(use standard inference tools)

Testing Non-UR → Standard Tools

$$\mathsf{LR}_{\gamma=0} \xrightarrow{\mathsf{D}} \chi_1^2, \tag{6.40}$$

for all values of π (Nielsen, 2006b).

- Test for hypothesis that does not alter the number of unit roots ⇝ standard inference tool (normal or χ^2)
- Test for unit roots:

$$LR_{\pi=\mu_c=0} \rightarrow^D DF_c^2,\ t_\pi \rightarrow^D DF_c \text{ under } H_0 : \pi = \mu_c = 0$$

Conintegration

Cointegration: Granger-Johansen Representation → ADL

- 3 Cases of Modelling

  -

We derive a moving average representation for the vector autoregression. We start from the vector equilibrium correction form

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t. \tag{7.8}$$

Here, X_t could be the vector of c_t and y_t. We distinguish between the cases where Π is invertible, where $\Pi = 0$ and where Π has reduced rank.

For likelihood analysis, ε_t is assumed i.i.d. $\mathsf{N}_2(0, \Omega)$.

① **7.2.1 The fully stationary case** Full Rank

When Π is invertible, we can exploit the theory for stationary vector autoregressions developed in §5.2.2. Suppose the autoregressive matrix $A = I_2 + \Pi$ has spectral radius less than unity, $\rho(A) < 1$. Then, we have the stationary solution

$$(X_t \mid X_0) \xrightarrow{\mathsf{D}} X_t^* = \sum_{j=0}^\infty A^j \varepsilon_{t-j} \overset{\mathsf{D}}{=} \mathsf{N}_2(0, \Sigma_X), \tag{7.9}$$

where Σ_X solves $\Sigma_X = A\Sigma_X A' + \Omega$.

This representation would be relevant for the Fulton fish data.

② **7.2.2 The pure random walk case** Zero Rank

When $\Pi = 0$, we have $\Delta X_t = \varepsilon_t$. We generalize the unit root theory from §6.2.2 and get the random walk solution

$$X_t = X_0 + \sum_{s=1}^t \varepsilon_s. \tag{7.10}$$

This representation would be relevant for Yule's data.

  -

③ **7.2.3 The cointegrated case: stylized version** Reduced Rank

We start from the cointegrated vector autoregression in (7.7) with $\alpha_2 = 0$. This is

Just a Special Case

$$\Delta \begin{pmatrix} c_t \\ y_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix} \underbrace{(1, -\kappa)}_{\begin{bmatrix} \alpha_1 & -\kappa\alpha_1 \\ 0 & 0 \end{bmatrix}} \begin{pmatrix} c_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{c,t} \\ \varepsilon_{y,t} \end{pmatrix}. \tag{7.11}$$

Note that this expression for the Π matrix does not capture all possible Π matrices with rank 1, so the following analysis is not fully general.

We consider two linear combinations of the system in (7.11). First, we find the common trend. The second line of equation (7.11) for y_t does not involve lagged variables:

$$\Delta y_t = \varepsilon_{y,t}. \tag{7.12}$$

Solving this equation gives the random walk

$$y_t = \sum_{s=1}^t \varepsilon_{y,s} + y_0. \tag{7.13}$$

  - *Derive the Granger-Johansen Representation and the implied ADL Model for inference* #flashcard
  -

$[1, -\kappa]\, \Delta \begin{bmatrix} c_t \\ y_t \end{bmatrix} = [1, -\kappa] \begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix} [1, -\kappa] \begin{bmatrix} c_{t-1} \\ y_{t-1} \end{bmatrix} + [1, -\kappa] \begin{bmatrix} \varepsilon_{c,t} \\ \varepsilon_{y,t} \end{bmatrix}$

$\Delta (c_t - \kappa y_t) : \alpha_1 (c_{t-1} - \kappa y_t) + (\varepsilon_{c,t} - \kappa \varepsilon_{y,t})$

Second, we find a stationary combination. Taking the cue from the factor $(1, -\kappa)$ on the right hand side of (7.11), pre-multiply that equation by $(1, -\kappa)$ to get

$$\Delta(c_t - \kappa y_t) = \alpha_1 (c_{t-1} - \kappa y_{t-1}) + (\varepsilon_{c,t} - \kappa \varepsilon_{y,t}). \tag{7.14}$$

This is an autoregression for $c_t - \kappa y_t$. When $|1 + \alpha_1| < 1$, the stationary solution is

$$(c_t - \kappa y_t)^* = \sum_{s=0}^{\infty} (1 + \alpha_1)^s (\varepsilon_{c,t-s} - \kappa \varepsilon_{y,t-s}). \tag{7.15}$$

$c_t = \kappa y_t + \sum_{s=0}^{\infty} (1+\alpha_1)^s (\varepsilon_{c,t-s} - \kappa \varepsilon_{y,t-s})$
$= \kappa \left(\sum_{s=1}^{t} \varepsilon_{y,s} + y_0 \right)$
$+ \sum_{s=0}^{\infty} (1+\alpha_1)^s (\varepsilon_{c,t-s} - \kappa \varepsilon_{y,t-s})$

Third, in (7.15), move the κy_t^* term to the right and replace y_t^* by the random walk solution in (7.13) to get a moving average representation for c_t. Writing this together with the expression for y_t in (7.13), we get the Granger-Johansen representation

RW Component

$$c_t = \boxed{\kappa \sum_{s=1}^{t} \varepsilon_{y,s}} + \kappa y_0 + \sum_{s=0}^{\infty} (1 + \alpha_1)^s (\varepsilon_{c,t-s} - \kappa \varepsilon_{y,t-s}), \tag{7.16}$$

$$y_t = \boxed{\sum_{s=1}^{t} \varepsilon_{y,s}} + y_0. \tag{7.17}$$

These equations match (7.1), (7.2). Each variable has a random walk component, so both variables are I(1). This random walk is common - it is called a *common stochastic trend* - but the linear combination $c_t - \kappa y_t$ has no random walk component - it is I(0) as analyzed in (7.14), (7.15). This linear combination is called a *cointegrating relation.*

- Derive the ADL:

7.2.4 Back to the autoregressive distributed lag model

We can move from the cointegrated vector autoregressive model (7.11) to an autoregressive distributed lag model as in (7.4) by conditioning as in §5.2.6. Suppose the innovations in the cointegrated vector autoregressive model (7.11) are jointly normal so that

$$\begin{pmatrix} \varepsilon_{c,t} \\ \varepsilon_{y,t} \end{pmatrix} = \mathsf{N}_2 \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{cc} & \Omega_{cy} \\ \Omega_{yc} & \Omega_{yy} \end{pmatrix} \right\}. \tag{7.18}$$

60

$$\Delta \begin{pmatrix} c_t \\ y_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix} \underbrace{(1, -\kappa)}_{\begin{bmatrix} \alpha_1 & -\kappa\alpha_1 \\ 0 & 0 \end{bmatrix}} \begin{pmatrix} c_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{c,t} \\ \varepsilon_{y,t} \end{pmatrix}.$$

-

$\begin{bmatrix} \Delta c_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \alpha_1 & -\kappa\alpha_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{c,t} \\ \varepsilon_{y,t} \end{bmatrix} \Rightarrow \begin{bmatrix} \Delta c_t \\ \Delta y_t \end{bmatrix} \sim N\left(\begin{bmatrix} \alpha_1 c_{t-1} - \kappa\alpha_1 y_{t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} \Omega_{cc} & \Omega_{cy} \\ \Omega_{yc} & \Omega_{yy} \end{bmatrix} \right)$

$\Rightarrow \Delta c_t | \Delta y_t \sim N\left(\alpha_1 c_{t-1} - \kappa\alpha_1 y_{t-1} + \underbrace{\frac{\Omega_{cy}}{\Omega_{yy}}}_{\equiv \omega} [\Delta y_t - 0], \Omega_{cc}(1-\rho^2) \right)$

$\Rightarrow$ ADL Eqn: $\Delta c_t = \omega \Delta y_t + \alpha_1 c_{t-1} - \kappa\alpha_1 y_{t-1} + \varepsilon_{c \cdot y, t} = \omega \Delta y_t + \alpha_1 \underbrace{(\overbrace{c_{t-1}}^{I(1)} - \kappa \overbrace{y_{t-1}}^{I(1)})}_{I(0)} + \varepsilon_{c \cdot y, t}$

-

We want to argue that the conditional distribution of c_t given y_t and the past and the marginal distribution of y_t given the past satisfy

$$\Delta c_t = \omega \Delta y_t + \alpha_1 (c_{t-1} - \kappa y_{t-1}) + \varepsilon_{c \cdot y, t}, \tag{7.19}$$
$$\Delta y_t = \varepsilon_{y,t}. \tag{7.20}$$

Cov / Var

where $\omega = \Omega_{cy} \Omega_{yy}^{-1}$ and where $\varepsilon_{c \cdot y, t} = \varepsilon_{c,t} - \omega \varepsilon_{y,t}$ has variance $\sigma^2 = \Omega_{cc} - \Omega_{cy} \Omega_{yy}^{-1} \Omega_{yc}$ and is independent of $\varepsilon_{y,t}$. We recognize the autoregressive distributed lag model (7.4).

The autoregressive distributed lag model is said to be balanced. The idea is that I(1) variables are 'balanced' by only entering through the cointegrating relation. If we repeat the autoregressive distributed lag equation (7.19) and indicate the order of integration of the variables, we see that it is a balanced equation of I(0) variables:

I(1) I(1) (Balanced)

ADL :

$$\underbrace{\Delta c_t}_{\mathsf{I}(0)} = \omega \underbrace{\Delta y_t}_{\mathsf{I}(0)} + \alpha_1 \underbrace{(c_{t-1} - \kappa y_{t-1})}_{\mathsf{I}(0)} + \underbrace{\varepsilon_{c \cdot y, t}}_{\mathsf{I}(0)}. \tag{7.21}$$

- Estimation and Inference in the implied ADL
- Practically, *just test the coefficient on c_{t-1}*:

$$\begin{cases} H_0 : & \alpha = 0 \iff \text{No Cointegration (pure RW)} \\ H_1 : & \alpha < 0 \iff \text{Cointegration} \end{cases}$$

- use special critical values

-

7.3.1 Estimation and inference in the ADL model

The ADL model is univariate and appeals to standard multiple regression. With the ADL model it is often possible to add sufficiently many lags so that autocorrelation can be avoided in the residuals. This simplifies inference to some extent as pointed out by Davidson et al. (1978). However, inference will also depend on the dynamics of the regressors, which is left unspecified.

Consider an autoregressive distributed lag model as in (7.4) and (7.19). For practical purposes, we add an intercept. Following (5.14), we then get the model equation

Model

$$\Delta c_t = \omega \Delta y_t + \alpha (c_{t-1} - \kappa y_{t-1} - \tau) + \varepsilon_t. \tag{7.22}$$

The corresponding regression equation is

Regression

$$\Delta c_t = \omega \Delta y_t + \alpha c_{t-1} + \overset{\alpha\kappa}{\beta} y_{t-1} + \overset{-\mu/\alpha}{\mu} + \varepsilon_t, \tag{7.23}$$

which can be estimated by least squares and from which we can back out the parameters $\kappa = -\beta/\alpha$ and $\tau = -\mu/\alpha$.

→ DF distribution needed

∃ Cointegration

The hypothesis that $\alpha = 0$ is a unit root hypothesis. Going over the representations in §7.2, we can see what happens in the background. The present α corresponds to α_1 in cointegrated vector autoregression (7.11). When $|1 + \alpha| < 1$, we have cointegration with $c_t - \kappa y_t$ as cointegrating relation. But, when $\alpha = 0$, the equation (7.23) reduces to the pure random walk case in (7.10).

Table 7.1 reports simulated quantiles of the asymptotic distribution of the t-test for $\alpha = 0$ as reported in Banerjee et al. (1998). Hendry (1995, p. 548) refers to this test as the PcGive unit root test. Table 7.1 reports quantiles for different dimensions of the regressor y_t. In all cases it is assumed that the regressor y_t is a pure random walk, so

-

that Δy_t is i.i.d.$(0, \sigma_y^2)$. This assumption is increasingly tenuous as the dimension of y_t increases. Quantiles are also reported for the t-test for $\alpha = 0$ in the linear trend model

$$\Delta c_t = \omega \Delta y_t + \alpha \big(c_{t-1} - \kappa y_{t-1} - \tau_\ell t \big) + \mu_c + \varepsilon_t. \tag{7.24}$$

Here, y_t can be a random walk with a linear trend, so that Δy_t is i.i.d.(μ_y, σ_y^2). The reported quantiles also apply if additional lags are added to the ADL equation, while maintaining that y_t is a pure random walk.

Once it has been established that there is cointegration, inference on the remaining parameters can be done using standard inference.

The joint hypothesis that the level parameters are zero in (7.23), that is $\alpha = \beta = \mu = 0$, can be tested using a likelihood ratio test statistic. Asymptotic theory is developed in Harbo et al. (1998).

- Alternative: Static Engle-Granger Regression:

-

Engle and Granger (1987) suggested to use a two-step residual based test for the hypothesis of no cointegration. In the first step, run the regression

(1)

$$c_t = \kappa y_t + \mu + u_t \tag{7.27}$$

by least squares. If the residuals $\hat{u}_t = c_t - \hat{\kappa} y_t - \hat{\mu}$ appear stationary, then under mild regularity assumptions, $T(\hat{\kappa} - \kappa)$ converges in distribution. The limiting distribution depends on the specification of u_t (Phillips and Durlauf, 1986). In any case, we have a fast consistency rate of T and we say that $\hat{\kappa}$ is super-consistent. If there is no cointegration, the regression (7.27) is spurious and κ is not identified.

In the second step, we test for cointegration. Run an autoregression

(2) Test H_0: $\rho = 1$ No Cointegration

$$\hat{u}_t = \rho \hat{u}_{t-1} + v_t, \tag{7.28}$$

and compute a t-test statistic for a unit root, $\rho = 1$. In this second regression, (7.28), we would like the errors v_t to be i.i.d., so the regression may be augmented by lags and possibly a linear term. The asymptotic distribution is of Dickey-Fuller type and depends on the assumptions to the variables (Phillips and Ouliaris, 1990).

Table 7.2 reports critical values assuming that c_t and y_t are pure random walks.

Misc

Time Series MCQ

let $\epsilon_t \sim^{iid} N(0, \sigma^2)$

Process	iid?	Strict stationary?	Weak stationary?	Temp. uncorr.?	Cond. hetero.?	Random walk?	MDS?	Martingale?	Marginally normal?
$X_t = \varepsilon_t + \varepsilon_{t-1}$	No	Yes	Yes	No	No	No	No	No	Yes
$Y_t = \varepsilon_t - \varepsilon_{t-2}$	No	Yes	Yes	No	No	No	No	No	Yes
$Z_t = \varepsilon_t \varepsilon_{t-1}^2$	No	Yes	Yes	Yes	Yes	No	Yes	No	No
$W_t = \varepsilon_t \varepsilon_{t-1}$	No	Yes	Yes	Yes	Yes	No	Yes	No	No
$V_t = \varepsilon_t^2$	Yes	Yes	Yes	Yes	No	No	No	No	No
$U_t = \varepsilon_t + \varepsilon_{t-1}^2$	No	Yes	Yes	Yes	No	No	No	No	No
$S_t = \sum_{s=1}^{t} \varepsilon_s$	No	No	No	No	No	Yes	No	Yes	Yes

- iid $\iff$ only t subscript
- strictly/weakly stationary $\iff$ everything unless RW
- temp correlated $\iff$ additive
- conditional homoskedastic $\iff$ additive
- normality $\iff$ additive / RW
- RW $\iff$ $\sum_{s=1}^{t} \epsilon_s$