# Empirical Methods in Finance MFE230E
# Week 5: Cointegration in Finance

**Martin Lettau**

**Spring 2019**

**Haas School of Business**



# This week

1. Objective: Understand joint dynamic behavior of stock prices, dividends and returns
2. Related to age-old question whether asset markets are **efficient or not**
3. Fama vs. Shiller – Nobel winner vs. Nobel winner

   Watch their Nobel lectures: `https://www.youtube.com/watch?v=WzxZGvrpFu4`!
4. How did you define market efficiency in *Fundamentals of Investments*?
5. Are stock markets predictable?

References:

- Ruppert ch. 15
- Cochrane's *Time Series for Macroeconomics and Finance* ch. 11
- Campbell, Lo, MacKinlay ch. 7
- More advanced: Hamilton chapters 18, 19, 20

# Outline

If $x_{1,t}$, $x_{2,t}$ are independent but both are $I(1)$, regression

$$x_{2,t} = \mu + \beta x_{1,t} + u_t$$

will yield spurious results.

Question: Can we properly analyze the relationships between $I(1)$ variables?

Answer: Sometimes, when they are **cointegrated**.

**Example 1 (Cointegrated I(1) variables).**

$$z_t = z_{t-1} + e_t \quad \leftarrow \quad \text{RW}$$

$$e_t, v_t, u_t \sim I(0)$$

$$x_{1,t} = z_t + v_t$$

$$x_{2,t} = z_t + u_t$$

$$\Rightarrow x_{1,t} - x_{2,t} = w_t + u_t \ \sim I(0)!!$$

- The $I(1)$ variable $z_t$ is present in $x_{1,t}$ and $x_{2,t}$, so they share the same source of non-stationarity
- Taking the difference $x_{1,t} - x_{2,t}$ removes the common $I(1)$ variable
- $z_t$ is called a **common stochastic trend**
- $x_{1,t} - x_{2,t}$ is **mean-reverting**
- Therefore $x_{1,t}$ and $x_{2,t}$ stay “close together”

- If $x_{1,t}$ and $x_{2,t}$ are $I(1)$ and **not** cointegrated, then the regression

$$x_{2,t} = \mu + \beta x_{1,t} + \epsilon_t$$

is spurious since $\epsilon_t \sim I(1)$! has component $z_t$

- If $x_{1,t}$ and $x_{2,t}$ are $I(1)$ and **cointegrated**, then the regression

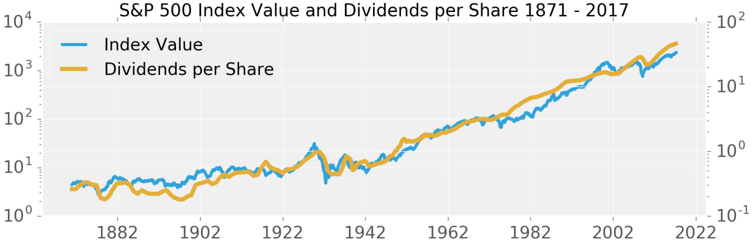
$$x_{2,t} = \mu + \beta x_{1,t} + \epsilon_t$$

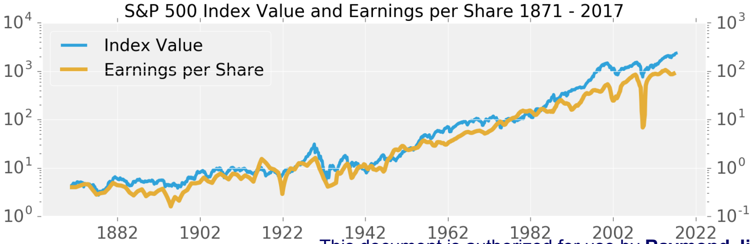
is **not** spurious since $\epsilon_t \sim I(0)$!

- Cointegration plays a central role in finance
- We will see shortly that **every present value relation** implies cointegration!
- The behavior of prices, returns and fundamentals (e.g. dividends and earnings) cannot be understood without cointegration
- Some popular trading strategies are based on cointegration:
  - Pairs trading two similar stocks in an industry. Look at gap between stocks, Buy low one and short high
  - Long Term Capital Management (LTCM)
- Great book about LTCM's rise and collapse:

  "When Genius Failed: The Rise and Fall of Long-Term Capital Management," by Roger Lowenstein

Stock prices and dividends/earnings are both *I*(1) but are linked together



**Definition 1.**

An $n \times 1$ vector time series $\mathbf{X}_t$ is **cointegrated** if each of the series is $I(1)$, but some linear combination(s) of the univariate series $\boldsymbol{\alpha}'\mathbf{X}_t$ is $I(0)$ for some nonzero $n \times 1$ vector $\boldsymbol{\alpha}$.

Note:

- The **cointegrating vector** $\boldsymbol{\alpha}'$ is not unique. If $\boldsymbol{\alpha}'\mathbf{X}_t$ is stationary, then $\delta\,\boldsymbol{\alpha}'\mathbf{X}_t$ is also stationary
- There may be $h$ linearly independent cointegration vectors $\boldsymbol{\alpha}_i$. We call $h$ the **cointegration rank**
- We will consider only $h = 1$ but $h$ can be estimated.

$\alpha$ usually $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

# Independent Random Walks

Recall from week 2: Independent random walks will diverge arbitrarily far away from each other.

Example:

$$x_{1,t} = x_{1,t-1} + e_t$$

$$x_{2,t} = x_{2,t-1} + u_t$$

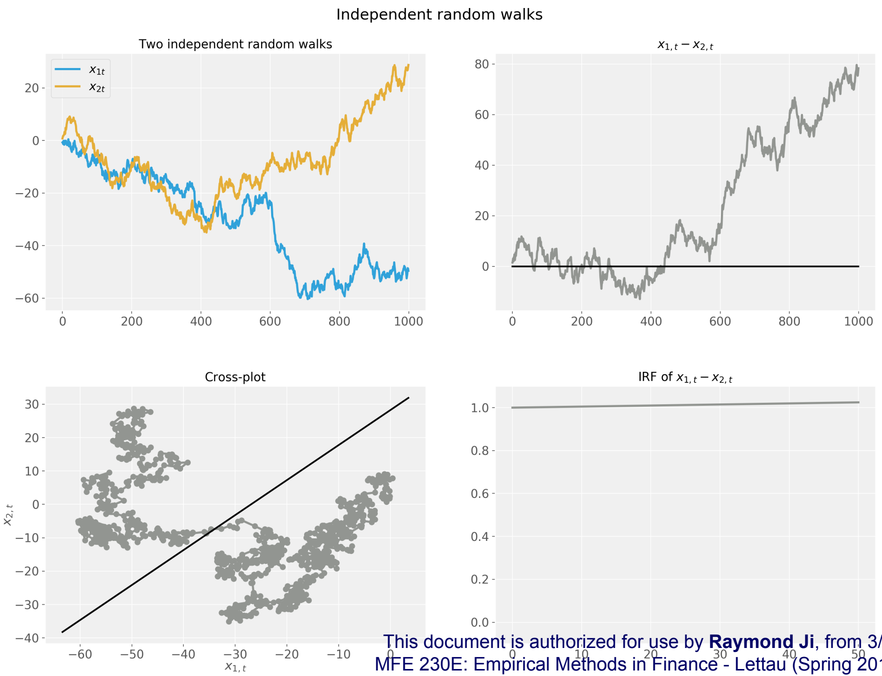
$$\text{Define } z_t = x_{1,t} - x_{2,t}$$

$$w_t = e_t - u_t$$

$$\Rightarrow z_t = z_{t-1} - w_t \sim I(1)$$

$$\text{Regression } x_{2,t} = \mu + \beta x_{1,t} + u_t$$

is spurious because the errors $u_t$ are $I(1)$.


Independent random walks
Two independent random walks
$x_{1t}$
$x_{2t}$
$x_{1,t} - x_{2,t}$
Cross-plot
$x_{1,t}$
$x_{2,t}$
IRF of $x_{1,t} - x_{2,t}$

**Example 2 (Cointegrated I(1) variables).**

$$x_{1,t} = x_{1,t-1} + e_t$$

$$x_{2,t} = x_{1,t} - z_t$$

$$\text{where } \; z_t = \psi z_{t-1} + u_t$$

AR(1)

- $x_{1,t}$ is a random walk
- $x_{2,t}$ is the sum of $x_{1,t}$ (a random walk) and a stationary AR(1) $z_t$:

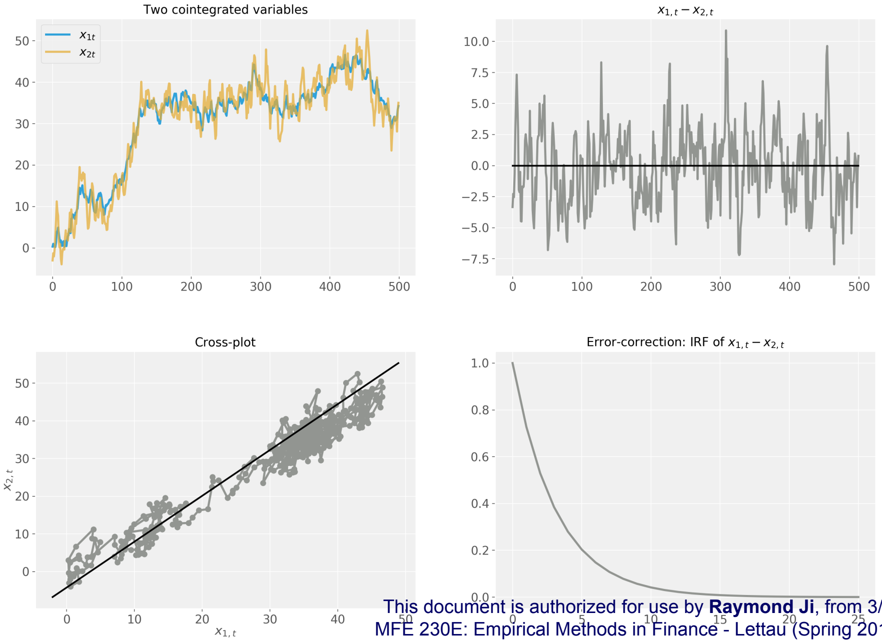
  AR($p_1$) + AR($p_2$) = ARMA($p_1 + p_2, \max(p_1, p_2)$):

$$(1-L)(1-\psi L)x_{2,t} = (1+\theta L)w_t \rightarrow \; x_{2,t} \sim \text{ARIMA}(1,1,1)$$

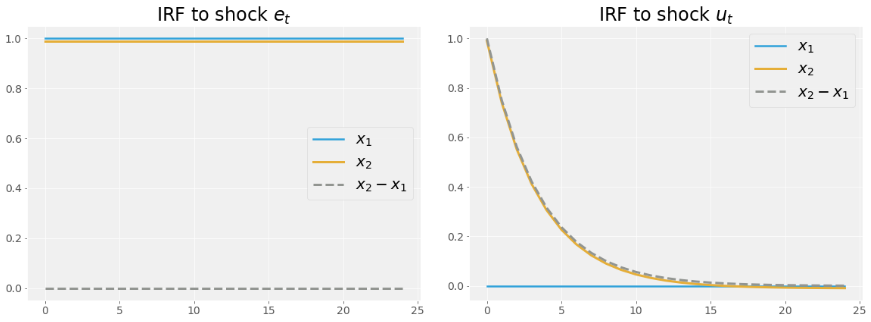
- But $x_{1,t} - x_{2,t} = z_t \sim$ AR(1) is stationary!

  $\longrightarrow$ $x_{1,t}$ and $x_{2,t}$ are **cointegrated**

# Cointegration



# Impulse response functions



## Cointegration: Error correction

- Two cointegrated nonstationary series cannot diverge forever, e.g. in the long run prices are tied to dividends and earnings
- $x_{1,t} - x_{2,t} = z_t$ is called the **cointegration error**
- If the series are currently far apart, eventually they will come back together: **error correction**
- The vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'\mathbf{X}_t$, $\mathbf{X}_t = (x_{1,t}, x_{2,t})'$ is $I(0)$ is called the **cointegration vector**
- Example: $\boldsymbol{\alpha} = (1, -1)'$

$x_{1,t} - x_{2,t} \sim I(0)$

- **Cointegration → mean reversion → forecastability**

Cointegrated systems can be written in different ways:

$$z_t = \psi z_{t-1} + u_t$$
$$x_{1,t} = x_{1,t-1} + e_t$$
$$x_{2,t} = x_{1,t} - z_t = x_{1,t-1} + e_t - \psi \underbrace{z_{t-1}}_{x_{1,t}-x_{2,t}} - u_t$$

$$X_t = \begin{pmatrix} 1 & 0 \\ 1-\psi & \psi \end{pmatrix} X_{t-1} + \begin{pmatrix} e_t \\ u_t \end{pmatrix}$$

$\Leftrightarrow$ VAR:

$$x_{1,t} = x_{1,t-1} + e_t$$
$$x_{2,t} = (1-\psi)x_{1,t-1} + \psi x_{2,t-1} + u_t$$

each row must sum to 1

$\rightarrow$ we can use MLE with restric.

$\Leftrightarrow$ VECM:

$$\Delta x_{1,t} = e_t \quad \leftarrow \text{white noise}$$
$$\Delta x_{2,t} = \underbrace{(1-\psi)(\overbrace{x_{1,t-1} - x_{2,t-1}}^{z_{t-1}})}_{\text{error correction term!}} + u_t$$

- **VAR** with parameter restrictions
- **VECM: Vector Error Correction Model**

$x_{1,t} = a x_{1,t-1} + b x_{2,t-1} + e_t$

$x_{2,t} = c x_{1,t-1} + d x_{2,t-1} + u_t$

$x_1 - x_2 = (a-c) x_{1,t-1} + (b-d) x_{2,t-1} + e_t - u_t$

- Any cointegrated system can be written in two forms:
  - VAR with coefficient restrictions: MLE
  - VECM: OLS

$a - c = -(b-d)$

$= -b + d$

$\boxed{a + b = c + d}$

1. **Restricted VAR** in $\mathbf{X}_t$ (e.g. $\phi_{21} + \phi_{22} = 1$):

$$x_{1,t} = \phi_{11} x_{1,t-1} + \phi_{12} x_{2,t-1} + \epsilon_{1,t}$$

$$x_{2,t} = \phi_{21} x_{1,t-1} + \phi_{22} x_{2,t-1} + \epsilon_{2,t}$$

2. **VECM**:

$$\Delta x_{1,t} = \gamma_1 z_{t-1} + \phi_{11} \Delta x_{1,t-1} + \phi_{12} \Delta x_{2,t-1} + \epsilon_{1,t}$$

$$\Delta x_{2,t} = \gamma_2 z_{t-1} + \phi_{21} \Delta x_{1,t-1} + \phi_{22} \Delta x_{2,t-1} + \epsilon_{2,t}$$

without first term, VAR in differences

$z_t$ error correction

Note: A VECM($p$) can be written as VAR($p+1$) (see Hamilton ch. 20)

Cointegration implies **forecastability**!

$$\text{VECM:} \qquad \Delta x_{1,t} = e_t$$
$$\Delta x_{2,t} = (1-\psi)(x_{1,t-1} - x_{2,t-1}) + u_t$$

VECM: **Vector Error Correction Model** since the $x_{1,t} - x_{2,t-1}$ "error-corrects" the deviation between $x_{1,t}$ and $x_{2,t-1}$

- If $x_{2,t} > x_{1,t}$, what can you say about future $x_{1,t+j}$ and $x_{2,t+j}$?
- Are $x_{1,t+j}$ or $\Delta x_{1,t+j}$ forecastable?
- Are $x_{2,t+j}$ or $\Delta x_{2,t+j}$ forecastable?

- Question: What happens if $\psi \longrightarrow 1$?



$\gamma$ gives sensitivity to error $x_{1,t} - x_{2,t} = z_t$

**Theorem 1 (Granger's Representation Theorem).**

*If the system $\mathbf{X}_t$ is cointegrated,*

1. *then the VECM representation*

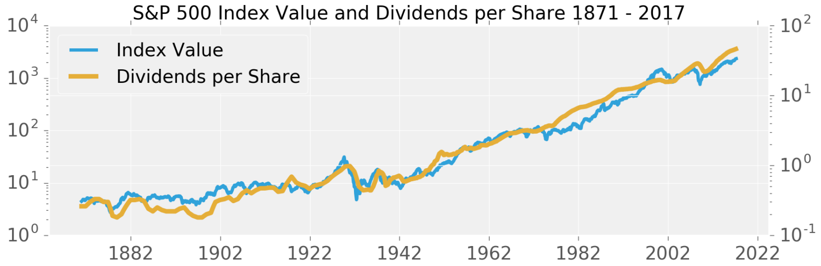
polynomial of gaps

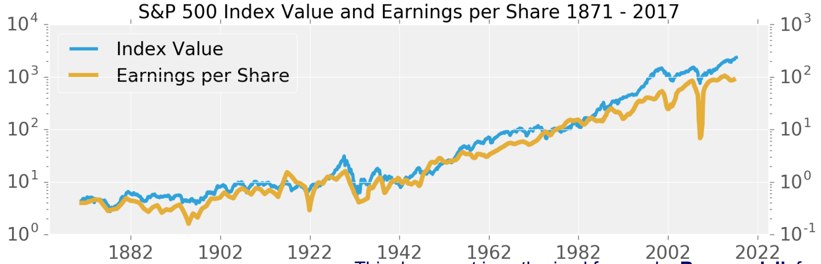
$$\begin{pmatrix} x_{1t} - x_{1t-1} \\ \vdots \\ x_{nt} - x_{nt-1} \end{pmatrix} = \underbrace{\Delta \mathbf{X}_t}_{\in \mathbb{R}^n} = \boldsymbol{\gamma}(\underbrace{\boldsymbol{\alpha}' \mathbf{X}_{t-1}}_{\text{creates stationary gap, } \in \mathbb{R}}) + \overbrace{\boldsymbol{\Phi}(L)}^{\text{polynomial of gaps}} \Delta \mathbf{X}_{t-1} + \boldsymbol{\epsilon}_t$$

   *exist, and*

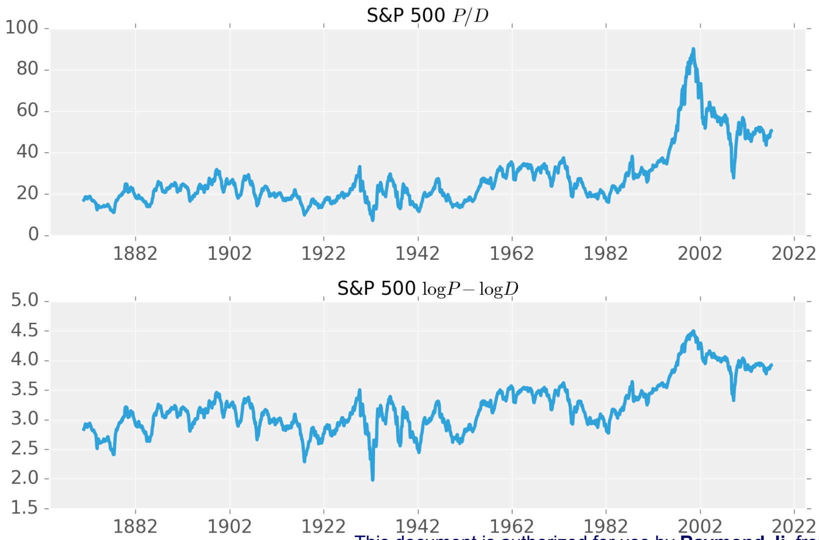
2. *at least one one of the coefficients $\gamma_i$ is non-zero.* → depends on at least one gap
3. *Hence, at least one of the series $X_{it}$ in $\mathbf{X}_t$ is forecastable.*

# S&P 500 DIVIDEND, EARNINGS AND PRICES



# S&P 500 PRICE/DIVIDEND RATIO



# COINTEGRATION IN PRACTICE

1. Before running any econometric test, use economic logic/theory to form a **prior** about how the data should behave (e.g. stationarity? cointegration? cointegration vector?).
2. Run a **unit root test** on each individual series.
3. Evidence that all series are **stationary** → Estimate **VAR**.
4. Evidence of **unit root(s)**:
   - Cointegrating vector is known: Unit root test on the cointegrating error $z_t = \boldsymbol{\alpha}' \mathbf{X}_t$
   - If the cointegrating vector is not known, estimate $\boldsymbol{\alpha}$ and test for cointegration (e.g. Johansen procedure, Ruppert p. 417)
5. Evidence for cointegration?
   - No: First difference the variables → Estimate **VAR in first differences**
   - Yes: Estimate **VECM**

**Algorithm:** Multivariate non-stationarity, cointegration and estimation

**Data:** $\mathbf{X}_t = (x_{1,t}, x_{2,t})'$

**Prior:** Cointegrated with $\boldsymbol{\alpha} = (1, -1)'$

**Test:** Dickey-Fuller or Johansen tests: $x_{1,t}, x_{2,t}$

**if** $x_{1,t}, x_{2,t} \sim I(0)$ **then**

  **Stationarity** $\rightarrow$ **VAR**$(\mathbf{X}_t)$

**else**

  **Non-stationarity**;

  **Test:** Dickey-Fuller test: $z_t = x_{1,t} - x_{2,t}$

  **if** $z_t \sim I(1)$ **then**

    **No cointegration** $\rightarrow$ **VAR**$(\Delta\mathbf{X}_t)$

  **else**

    **Cointegration** $\rightarrow$ **VECM**

# CASE 1: DATA IS STATIONARY

$\mathbf{X}_t$: **Stationary → Standard methods apply**

**Unrestricted VAR** in $\mathbf{X}_t$:

$$x_{1,t} = \phi_{11} x_{1,t-1} + \phi_{12} x_{2,t-1} + \epsilon_{1,t}$$

$$x_{2,t} = \phi_{21} x_{1,t-1} + \phi_{22} x_{2,t-1} + \epsilon_{2,t}$$

Estimation: OLS equation-by-equation

**$\mathbf{X}_t$: Non-stationary, no cointegration → Take first differences**

**Unrestricted VAR** in differences $\Delta\mathbf{X}_t$:

$$\Delta x_{1,t} = \phi_{11}\Delta x_{1,t-1} + \phi_{12}\Delta x_{2,t-1} + \epsilon_{1,t}$$
$$\Delta x_{2,t} = \phi_{21}\Delta x_{1,t-1} + \phi_{22}\Delta x_{2,t-1} + \epsilon_{2,t}$$

Estimation: OLS equation-by-equation

Note: VAR in $\mathbf{X}_t$ is also OK but less efficient (see Hamilton)

**$\mathbf{X}_t$: Non-stationary but cointegrated → VECM**

**VAR with restrictions**:

$$x_{1,t} = \phi_{11} x_{1,t-1} + \phi_{12} x_{2,t-1} + \epsilon_{1,t}$$
$$x_{2,t} = \phi_{21} x_{1,t-1} + \phi_{22} x_{2,t-1} + \epsilon_{2,t}$$

or **VECM**:

$$\Delta x_{1,t} = \gamma_1 z_{t-1} + \phi_{11} \Delta x_{1,t-1} + \phi_{12} \Delta x_{2,t-1} + \epsilon_{1,t}$$
$$\Delta x_{2,t} = \gamma_2 z_{t-1} + \phi_{21} \Delta x_{1,t-1} + \phi_{22} \Delta x_{2,t-1} + \epsilon_{2,t}$$

$$\Leftrightarrow \ \Delta \mathbf{X}_t = \boldsymbol{\gamma}(\boldsymbol{\alpha}' \mathbf{X}_t) + \boldsymbol{\Phi}(L) \Delta \mathbf{X}_t + \boldsymbol{\epsilon}_t$$

Estimation: OLS equation-by-equation

# Example: Putting everything together

$$z_t = \phi z_{t-1} + u_t$$
$$x_{1,t} = x_{1,t-1} + e_t$$
$$x_{2,t} = x_{1,t-1} - z_t$$

- **Case 1**: $x_{1,t}, x_{2,t} \sim I(0)$ ⟶ Unrestricted VAR(1)
- **Case 2**: $|\phi| < 1$ ⟶ $x_{1,t}, x_{2,t} \sim I(0)$ ⟶ Cointegration

  ⟶ Restricted VAR(1) or VECM(0):

$$\begin{aligned} x_{1,t} &= x_{1,t-1} + e_t \\ x_{2,t} &= (1-\phi)x_{1,t-1} + \phi x_{2,t-1} + u_t \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \Delta x_{1,t} &= e_t \\ \Delta x_{2,t} &= (1-\phi)\left(x_{1,t-1} - x_{2,t-1}\right) + u_t \end{aligned}$$

  Note: VECM is easier to estimate. (Why?)

- **Case 3**: $\phi = 1$ ⟶ $x_{1,t}, x_{2,t} \sim I(1)$ ⟶ No cointegration

  ⟶ Unrestricted VAR(0) in first differences:

$$\Delta x_{1,t} = e_t$$
$$\Delta x_{2,t} = u_t$$

  Note: VAR in levels is also OK but less efficient.

$$z_t = \phi z_{t-1} + u_t$$
$$x_{1,t} = x_{1,t-1} + e_t$$
$$x_{2,t} = x_{1,t-1} - z_t$$

$$\Leftrightarrow \text{ VECM: } \quad \Delta x_{1,t} = e_t$$
$$\Delta x_{2,t} = (1-\phi)(x_{1,t} - x_{2,t-1}) + u_t$$

Estimation: VECM(1)

crm difference!

$$\Delta x_{1,t} = \gamma_1 z_{t-1} + \phi_{11}\Delta x_{1,t-1} + \phi_{12}\Delta x_{2,t-1} + \epsilon_{1,t}$$
$$\Delta x_{2,t} = \gamma_2 z_{t-1} + \phi_{21}\Delta x_{1,t-1} + \phi_{22}\Delta x_{2,t-1} + \epsilon_{2,t}$$

Question: What are the expected point estimates for $\gamma_i$ and $\phi_{jk}$?

$$z_t = \phi z_{t-1} + u_t$$
$$x_{1,t} = x_{1,t-1} + e_t$$
$$x_{2,t} = x_{1,t-1} - z_t$$

$$\Leftrightarrow \text{ VECM:} \quad \Delta x_{1,t} = e_t$$
$$\Delta x_{2,t} = (1-\phi)(x_{1,t} - x_{2,t-1}) + u_t$$

Simulation with $\phi = 0.75 \Rightarrow \gamma_1 = 0, \gamma_2 = 0.25$

| | $\gamma$ | $\phi_{j1}$ | $\phi_{j2}$ |
|---|---|---|---|
| $\Delta x_{1,t}$ | 0.00 | -0.02 | -0.00 |
| $\Delta x_{2,t}$ | 0.24 | -0.01 | -0.01 |

Note: $\Delta x_{2,t}$ is forecastable by $z_{t-1} = x_{1,t-1} - x_{2,t-2}$!

## EXAMPLE, CONTINUED

- In this simple example, the VECM and VAR in differences are of order 0
- If we add terms to the example, the resulting VECM and VAR will have more lages
- Extended example:

$$
\begin{aligned}
z_t &= \phi z_{t-1} + u_t \\
\Delta x_{1,t} &= \psi \Delta x_{1,t-1} + e_t \\
x_{2,t} &= x_{1,t-1} - z_t
\end{aligned}
$$

## Outline

- Starting with the definition of a (net) return:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1$$

$$\begin{aligned} P_t &= \frac{P_{t+1} + D_{t+1}}{1 + R_{t+1}} \\ &= \frac{D_{t+1}}{1 + R_{t+1}} + \frac{1}{1 + R_{t+1}} \frac{P_{t+2} + D_{t+2}}{1 + R_{t+2}} \\ &= \sum_{i=1}^{K} \frac{D_{t+i}}{\prod_{j=1}^{i}(1 + R_{t+j})} + \frac{P_{t+K}}{\prod_{j=1}^{i}(1 + R_{t+j})} \end{aligned}$$

- Since $P_t$ is known at time $t$:

$$P_t = \mathrm{E}_t\left[\sum_{i=1}^{K} \frac{D_{t+i}}{\prod_{j=1}^{i}(1+R_{t+j})}\right] + \mathrm{E}_t\left[\frac{P_{t+K}}{\prod_{j=1}^{K}(1+R_{t+j})}\right]$$

- Assumption: $\lim_{k\to\infty} \mathrm{E}_t\left[\frac{P_{t+K}}{\prod_{j=1}^{K}(1+R_{t+j})}\right] = 0$ (i.e. prices do not explode)

- **Present-value (PV) relationship** between prices, dividends and returns:

$$P_t = \mathrm{E}_t\left[\sum_{i=1}^{\infty} \frac{D_{t+i}}{\prod_{j=1}^{i}(1+R_{t+j})}\right]$$

# Discussion

PV:

$$P_t = \mathrm{E}_t\left[\sum_{i=1}^{\infty} \frac{D_{t+i}}{\prod_{j=1}^{i}(1+R_{t+j})}\right].$$

Discussion:

- We started from the definition of returns
- The only assumption we made is that prices are assumed to not explode
- This equation holds whether you believe markets are efficient or not
- Even Shiller would agree!
- So, what do Fama and Shiller (and the rest of the finance profession) argue about?

# Outline





## CONSTANT DISCOUNT RATES (OR WHAT WE TEACH MBAS)

- Now we make an (important) assumption: $\mathrm{E}_t[R_{t+1}] = R \ \forall t$:

$$P_t = \mathrm{E}_t \sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+R)^i}$$

- We will come back to the precise meaning of this assumption later
- **Question: What causes price to be high?**
- Important role of financial markets:

**Prices reveal information of market participants**

- Note: Assets with higher expected returns have lower prices (given dividend expectations)
- Let's look at the properties of prices and dividends

The observed price should be the expected value of discounted dividends:

$$P_t = \mathrm{E}_t \sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+R)^i}$$

Shiller (1982): compute the (pseudo) ex post perfect-forecast prices using realized dividends

$$P_t^* = \sum_{i=1}^{K} \frac{D_{t+i}}{(1+R)^i}$$

$$P_t = \mathrm{E}_t P_t^*$$

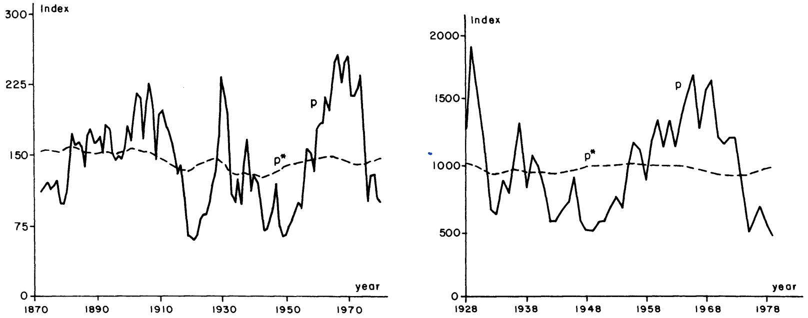
$$P_t^* = P_t + \epsilon_t$$

$$\mathrm{Var}(P_t^*) \geq \mathrm{Var}(P_t) \quad \Rightarrow \quad \textbf{variance bound}$$

→ The price is too volatile for dividends

General property: A forecast cannot be more volatile than the variable it is forecasting.

# Shiller's excess volatility



# WORKING ASSUMPTION: PRICES AND DIVIDENDS ARE NON-STATIONARY,

- Data: Prices and dividends are both $I(1)$, differences are stationary:

$$P_t, D_t \sim I(1)$$

$$\Delta P_t, \Delta D_t, R_t \sim I(0)$$

- For log variables:

$$p_t, d_t \sim I(1)$$

$$\Delta p_t, \Delta d_t, r_t \sim I(0)$$



## Working assumption: Prices and dividends are non-stationary,

- Nonstationarity makes statistical analysis difficult
  - spurious regressions
  - unconditional (population) variance of a non-stationary process is infinite
- Strictly speaking, Shiller's volatility bounds are statistically invalid!
- But we will see that his basic point is correct: Prices are too volatile relative to dividends
- Recall: Shiller's volatility bounds are derived from the assumption $E_t R_{t+1} = R$

- Problem set: Show that

$$P_t = \mathrm{E}_t \sum_{i=1}^{\infty} \frac{D_{t+i}}{(1+R)^i}$$

$$\Rightarrow P_t - \frac{D_t}{R} = \frac{1}{R}\,\mathrm{E}_t \sum_{i=0}^{\infty} \frac{\Delta D_{t+1+i}}{(1+R)^i}$$

- The rhs depends on $\Delta D_t$, which is $I(0)$.
- Even though $P_t, D_t \sim I(1)$, a linear combination of $P_t$ and $D_t$ depends on $\Delta D_{t+i}$ and is stationary!

$$P_t - \frac{D_t}{R} \sim I(0) \quad \Rightarrow \quad \text{cointegration}$$



Why is cointegration important for finance?

**Every present value relation implies cointegration**

Example: $(P_t, D_t)$ are cointegrated with cointegrating vector $(1, -1/R)'$.

**Cointegration implies forecastability (cf. Granger)**

Question: What does the model

$$P_t - \frac{D_t}{R} = \frac{1}{R}\,\mathrm{E}_t \sum_{i=0}^{\infty} \frac{\Delta D_{t+1+i}}{(1+R)^i}$$

imply for $\gamma_P, \gamma_D$ and $\alpha$ in the VECM

$$\Delta P_t = \gamma_P(P_{t-1} - \alpha D_{t-1}) + \phi_{11}\Delta P_{t-1} + \phi_{12}\Delta D_{t-1} + e_{P,t}$$
$$\Delta D_t = \gamma_D(P_{t-1} - \alpha D_{t-1}) + \phi_{21}\Delta P_{t-1} + \phi_{22}\Delta D_{t-1} + e_{D,t}$$

# Outline

- The PV above assumed that expected returns are constant: $\mathrm{E}_t R_{t+1} = R$
- Let's consider the PV without assuming $\mathrm{E}_t[R_{t+1}] = R$ (back to identities)

$$P_t = \mathrm{E}_t \sum_{i=1}^{\infty} \frac{D_{t+i}}{\prod_{j=1}^{i}(1 + R_{t+j})}$$

- Another trick to obtain stationarity: Convert to **ratios**

$$\frac{P_t}{D_t} = \mathrm{E}_t \sum_{i=1}^{\infty} \prod_{j=1}^{i} \frac{D_{t+j}/D_{t-1+j}}{(1 + R_{t+j})}$$

- How does this help with non-stationarity?

- Logs: $p_t = \log P_t$, $d_t = \log D_t$, $r_t = \log(1 + R_t)$

$$p_t - d_t = \log \left[ \mathrm{E}_t \sum_{i=1}^{\infty} \prod_{j=1}^{i} \frac{D_{t+i}/D_{t-1+i}}{(1 + R_{t+j})} \right]$$

- RHS "looks" stationary but is complicated: log of a sum of a product
- We will see that the RHS is indeed $I(0)$ under mild conditions
- **Log prices and dividends are cointegrated with $\alpha = (1, -1)'$**
- RHS: Log of expectation of a sum of a product $\Rightarrow$ nonlinear
- Next: Convert to an (approximate) linear model in logs

## An (approximate) log-linear framework

Start from return definition

$r_t = \log(1+R_t)$

$$1+R_{t+1}=\frac{P_{t+1}+D_{t+1}}{P_t}=\frac{P_{t+1}}{P_t}\left(1+\frac{D_{t+1}}{P_{t+1}}\right)$$

$$r_{t+1}=p_{t+1}-p_t+\log(1+\exp(d_{t+1}-p_{t+1}))$$

$$f(x_{t+1})\approx f(\bar{x})+f'(\bar{x})(x_{t+1}-\bar{x})$$

$$\rho\equiv 1/(1+\exp(\overline{d-p}))\approx 0.96$$

$$r_{t+1}\approx k+\rho p_{t+1}+(1-\rho)d_{t+1}-p_t$$

$$\Rightarrow p_t\approx k+\rho p_{t+1}+(1-\rho)d_{t+1}-r_{t+1}$$

Problem set: The approximate expression for log returns can be solved for log prices (ignoring constants):

$$p_t \approx \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j \left[ (1-\rho) d_{t+1+j} - r_{t+1+j} \right]$$

log

Compare to standard PV:

$$P_t = \mathrm{E}_t \sum_{i=1}^{\infty} \frac{D_{t+i}}{\prod_{j=1}^{i} (1+R_{t+j})}$$

## A LOG-LINEAR PRESENT VALUE RELATION

- Advantage: linear in logs $\Rightarrow$ we can use linear time series methods
- Note: if $d_t \sim I(1)$, $r_t \sim I(0) \Rightarrow p_t \sim I(1)$
- Standard effect of expected dividends on today's price: Higher expected dividends in the future $\longrightarrow$ Today's prices goes up
- But: This PV does NOT assume constant expected returns!
- Price can change for two reasons:
  - Dividend expectation change
  - Expected returns are changing
- If expected returns go up, prices decline (holding $E[d]$ constant)
- Same mechanism as for bonds: Yields and prices are inversely related
- **Question: How does all of this relate to Shiller's excess volatility?**

## Why might expected returns change over time?

- Expected returns depend on the **amount of risk and risk aversion**
- Example: Market return in the CAPM

$$ER^M = R^f + \bar{A}\,\sigma_M^2$$

- Recall: The standard CAPM is a static model
- In dynamic versions of the CAPM:

$$E_t R_{t+1}^M = R_t^f + \bar{A}_t\,\sigma_{M,t}^2$$

- If $R_t^f$, $\bar{A}_t$, $\sigma_{M,t}^2$ vary over time, the market discount rate changes as well
- **Question: Can returns be predictable in an efficient market?**

# OUTLINE

1. The econometrics of cointegration
2. Returns, prices, and dividends
3. Constant expected returns
4. The price-dividend ratio
5. **What is an efficient market?**
6. Price-dividend example
7. Long-horizon regressions

# What is an efficient market?

- **Weak form of market efficiency**: Prices follow a martingale

$$P_{t+1} = P_t + \epsilon_{t+1}$$

or (almost) equivalently returns are unforecastable:

$$R_t = R + \epsilon_t$$

$$\mathrm{E}_{t-1} R_t = R \qquad E[R_t | F_{t-1}] = R \quad \text{forecast}$$

- Implicit **assumption**: risk and risk aversion are constant (e.g. CAPM holds)
- Important insight: This is an assumption, **not a definition**



The modern definition of market efficiency has three parts:

1. Model assumptions about what determines expected returns (i.e. how risky the asset is): SDF $M_{t+1}$ in $E_t[M_{t+1}R_{t+1}] = 1$ (e.g. CAPM, Consumption CAPM, ...)
2. Model assumptions about how expectations are formed: Rational, behavioral, ...
3. PV holds:

$$P_t = \mathrm{E}\left[\sum_{i=1}^{\infty} \frac{D_{t+i}}{\prod_{j=1}^{i}(1 + R_{t+j})} \,\Big|\, F_t\right]$$

Example: Unconditional CAPM

$$ER^M_{t+1} - R^f = \bar{A}\,\sigma^2_M = \text{const.} \quad \Rightarrow \quad R^M_t - R = \epsilon_t,$$

where $\bar{A}$ is the average risk aversion of investors

$\longrightarrow$ Market risk premium: **Price of risk** $\times$ **amount of risk**

Extension: Dynamic CAPM

$$E_t R^M_{t+1} - R^f_t = \bar{A}_t\,\sigma^2_{M,t}$$



# IMPORTANT QUESTION: MARKET EFFICIENY?

If the data show that prices are not a random walk and stock returns are forecastable, can we conclude that the stock market market is inefficient?

NO!

We can say that the model, along with its assumptions about how expected returns are determined, is rejected. But we cannot draw the general conclusion that markets are inefficient.

Any test of market efficiency is a **joint** test of the PV **and** the assumed model of risk!

# IMPORTANT QUESTION: MARKET EFFICIENY?

Example:

- Model assumption: $M_{t+1} \equiv M = 1/R$
- Implication: $1 = E_t[M_{t+1}R_{t+1}] = M\, E_t R_{t+1} \;\Rightarrow\; R_{t+1} = R$
- Data: If $H_0 : E_t R_{t+1} = R$ is rejected, then
- **Proper conclusion**: The model with assumption $M_{t+1} = M$ is rejected
- **Incorrect conclusion**: Markets are inefficient!

Fama (1991):

> *"**Market efficiency per se is not testable**. It must be **tested jointly with some model of equilibrium**, an asset-pricing model. ... As a result, when we find anomalous evidence on the behavior of returns, the way it should be **split between market inefficiency or a bad model of market equilibrium is ambiguous**."*

(My highlights)

Next:

1. Derive a log-linear model of prices and returns that is **stationary**.
2. Take the model to the data.

We derived

$$p_t \approx \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j \left[(1-\rho) d_{t+1+j} - r_{t+1+j}\right]$$

Problem set: The log price-dividend ratio can written as:

$$p_t - d_t \approx E_t \left[ \sum_{j=0}^{\infty} \rho^j \left( \Delta d_{t+1+j} - r_{t+1+j} \right) \right]$$

- $\Delta d_t, r_t \sim I(0) \Rightarrow (p_t, d_t)$ are cointegrated with coint. vector $(-1, 1)$
- Current prices are high relative to current dividends if future expectations of dividend growth are high or expectations of future expected returns are low



# LOG-LINEAR PRICE-DIVIDEND RATIO

Examples:

- The Fed releases good news about the **future** economy

  $\longrightarrow$ future expected dividends are higher

  $\longrightarrow$ **current** $p_t - d_t$ will rise

- Concerns about **future** economic uncertainty are heightened

  $\longrightarrow$ future expected risk premia rise

  $\longrightarrow$ **current** $p_t - d_t$ will fall

- Market expectations about future dividends and returns are not directly observable.
- But current $p_t - d_t$ is observable!
- How can we exploit the information that is contained in $p_t - d_t$?
- What regressions are implied by the $p_t - d_t$ expression?
- Hint: Think about the equation in reverse:

$$E_t\left[\sum_{j=0}^{\infty}\rho^j \Delta d_{t+1+j}\right] - E_t\left[\sum_{j=0}^{\infty}\rho^j r_{t+1+j}\right] = p_t - d_t$$

- If $r_t$ and/or $\Delta d_t$ are i.i.d., what regression would you run?

# Outline

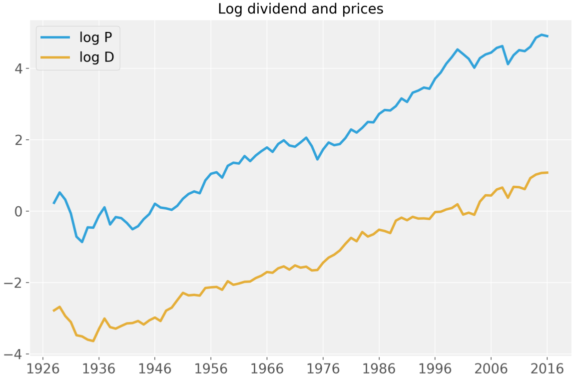




# EXAMPLE: LOG PRICES AND DIVIDENDS

Data: CRSP-VW index, 1927-2013 (available from WRDS)




**Question: Why annual data rather than monthly?**

Robert Shiller maintains a data set that starts in 1871:
`http://www.econ.yale.edu/~shiller/data/`

Great to have a longer data set but Shiller's data construction of dividends and earnings uses interpolations:

> *"Monthly dividend and earnings data are computed from the S&P four-quarter totals for the quarter since 1926, with linear interpolation to monthly figures."*

Interpolations introduces weird short-run dynamics to the data.

OK for using for unit root tests but not for short-term forecasting!

(Important: Always read data description!)



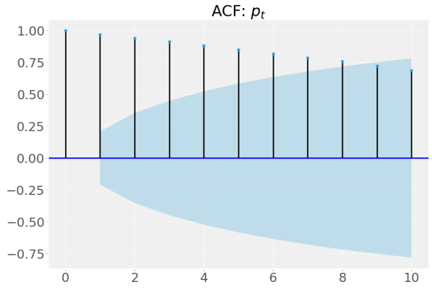
**Step 1: Before running any econometric test, use economic theory to form a prior about how the data should behave (e.g. PV relations should imply cointegration)**

We derived: Implications for stock prices and dividends:

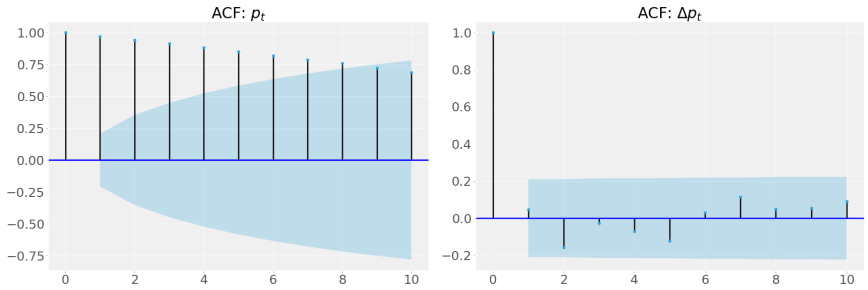
$$p_t - d_t = E_t\left[\sum_{j=0}^{\infty}\rho^j\left(\Delta d_{t+1+j} - r_{t+1+j}\right)\right]$$

Prior: $p_t$ and $d_t$ should be cointegrated with cointegration vector $(1, -1)'$

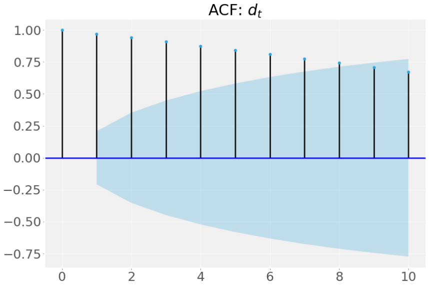
# Example: Log prices and dividends

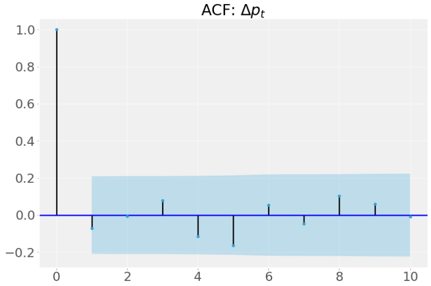


# Example: Log prices and dividends



# Example: Log prices and dividends



## Step 2: Run a unit root test on each individual series

Augmented Dickey-Fuller test:

$$\Delta y_t = \alpha + \delta y_{t-1} + \beta t + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

$$H_0: \delta = 0 \quad \text{(i.e. a unit root is present)}$$

$$ADF = \frac{\hat{\delta}}{SE(\hat{\delta})}$$

Question: Include time trend in ADF regression?

# Example: Log prices and dividends

| | $p$ | $\Delta p$ | $d$ | $\Delta d$ |
|---|---|---|---|---|
| ADF | 0.33 | -7.38 | 0.81 | -9.91 |
| p-value | 0.98 | 0.00 | 0.99 | 0.00 |
| lags | 0.00 | 1.00 | 1.00 | 0.00 |
| cv 5% | -2.89 | -2.90 | -2.89 | -2.90 |

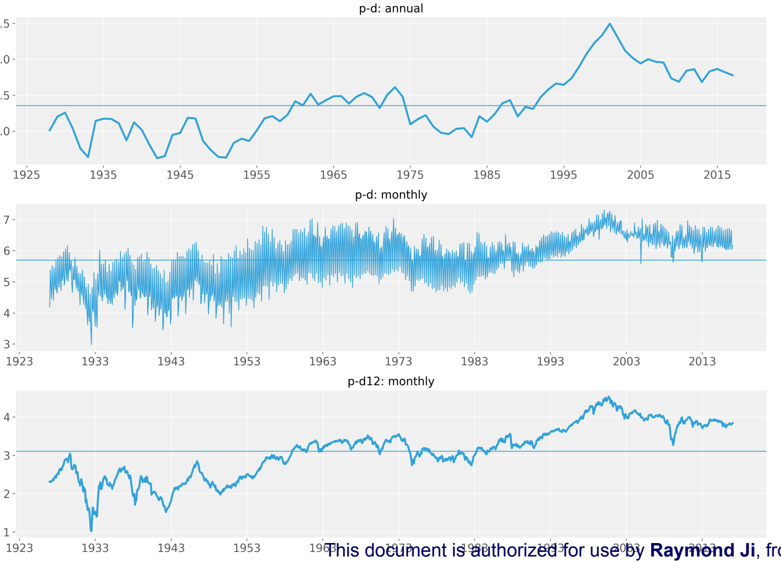
Based on the results in Step 2, unit roots in $p_t$ and $d_t$ cannot be rejected but $\Delta p_t$ and $\Delta d_t$ are stationary

$\longrightarrow$ proceed to Step 4

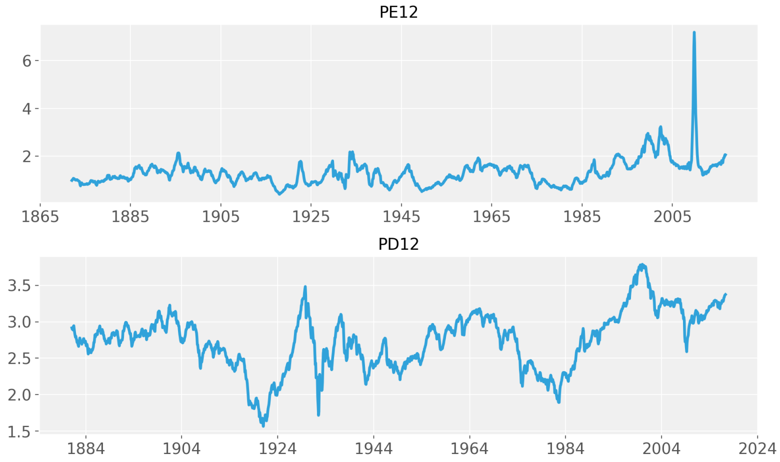
**Step 4: If there is strong evidence of unit roots in the variables, consider whether the series might be cointegrated; does theory imply a particular cointegration vector (e.g. $\alpha = (1, -1)'$ for dividends and prices)?**

Since theory implies a specific cointegration vector, let's look at $p_t - d_t$

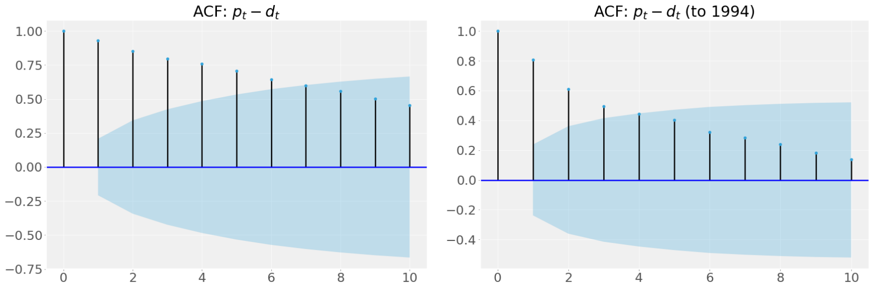
# Example: Log prices and dividends



# Example: Dividends/earnings.



# EXAMPLE: LOG PRICES AND DIVIDENDS



CRSP data:

| | $p-d$ | $p-d12$ | $p-d$ 94 | $p-d12$ 94 |
|---|---|---|---|---|
| AC(12) | 0.93 | 0.93 | 0.81 | 0.88 |
| ADF | -1.61 | -1.71 | -2.16 | -1.88 |
| p-value | 0.48 | 0.43 | 0.22 | 0.34 |
| lag(AIC) | 0.00 | 5.00 | 0.00 | 5.00 |

Shiller's data:

| | $p-d12$ | $p-e12$ | $p-e120$ |
|---|---|---|---|
| AC(12) | 0.89 | 0.76 | 0.89 |
| ADF | -2.89 | -5.14 | -2.76 |
| p-value | 0.05 | 0.00 | 0.06 |
| lag(AIC) | 5.00 | 5.00 | 5.00 |

### Step 6: Estimate a VECM

$$\Delta d_t = c_d + \gamma_d(p_{t-1} - d_{t-1}) + b_{11}\Delta d_{t-1} + b_{12}\Delta p_{t-1} + e_{d,t}$$

$$\Delta p_t = c_p + \gamma_p(p_{t-1} - d_{t-1}) + b_{21}\Delta d_{t-1} + b_{22}\Delta p_{t-1} + e_{p,t}$$

```
-------------------VECM-------------------
params             p            d
const          0.199       -0.0009
             (1.2498)     (-0.0069)
p_lag1        -0.0576      -0.1873
            (-0.4097)     (-1.9652)
d_lag1         0.246        0.097
             (1.5414)      (0.8632)
coinEq         0.0465      -0.0151
             (0.9906)     (-0.3911)
Rsq            0.0347       0.0427
------------------------------------------
```

# EXAMPLE: LOG PRICES AND DIVIDENDS

For data up to 1994:

```
-------------------VECM-------------------
          params           p             d
           const      0.5488        0.0315
                    (2.4091)      (0.1944)
          p_lag1       0.014       -0.1179
                    (0.0896)     (-1.2332)
          d_lag1      0.2194        0.1224
                    (1.0403)      (0.8686)
          coinEq      0.1627       -0.0021
                    (2.2553)     (-0.0426)
             Rsq      0.0775        0.0181
------------------------------------------
```



## EXAMPLE: LOG PRICES AND DIVIDENDS

If the sample ended in 1994 we would conclude:

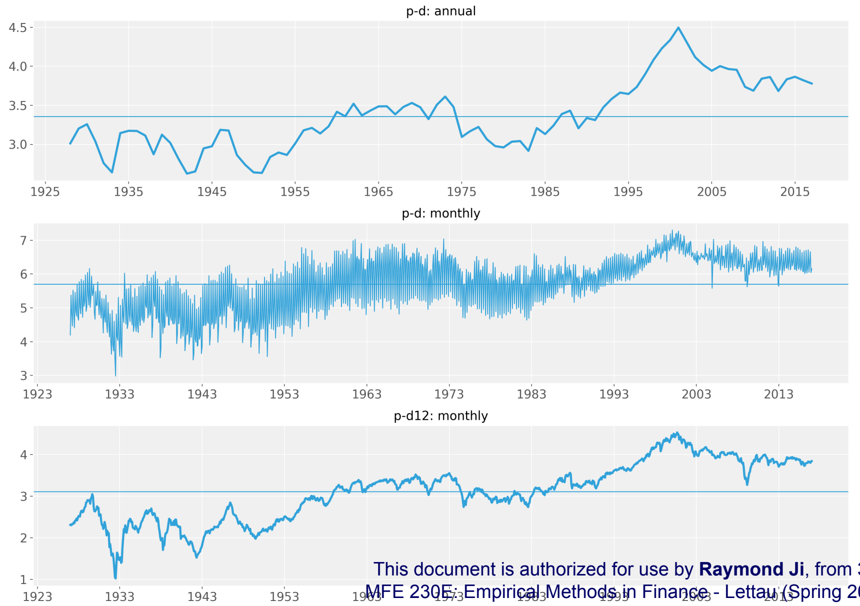
- Dividends are not forecastable by $p_t - d_t$
- Prices are forecastable by $p_t - d_t$
- Prices, not dividends, "error-correct"

Questions:

- Hence if prices are currently high relative to dividends, then what is likely to happen?
- What does this imply for market efficiency?

Open question: Why do the results change if we include data up to 2014?

# Example: Log prices and dividends



# Outline





We derived

$$p_t - d_t = E_t\left[\sum_{j=0}^{\infty} \rho^j \left(\Delta d_{t+1+j} - r_{t+1+j}\right)\right]$$

The VECM tests forecastability one-period ahead but $p_t - d_t$ depends on expectations in $t + j$.

Implementation: Reverse the equation

$$E_t\left[\sum_{j=0}^{\infty} \rho^j \left(\Delta d_{t+1+j} - r_{t+1+j}\right)\right] = p_t - d_t$$

Recall: Derived from an (approximate) identity $\longrightarrow$ Not directly testable



The $p_t - d_t$ equations suggest the following regressions:

$$r_{t+1} + \cdots + r_{t+k} = \alpha_k + \beta_k (p_t - d_t) + u_{t+k,k}$$
$$\Delta d_{t+1} + \cdots + \Delta d_{t+k} = \alpha_k + \beta_k (p_t - d_t) + u_{t+k,k}$$

These specifications are called "long-horizon regressions"

- How to choose $k$?
- $p_t - d_t$ is very persistent $\rightarrow$ spurious regressions
- Overlapping vs. non-overlapping data

Appendix: Hansen-Hodrick correction for standard errors in long-horizon regressions.

$$r_{t+1} + \cdots + r_{t+k} = \alpha_k + \beta_k (p_t - d_t) + u_{t+k,k}$$

| | 1 | 2 | 3 | 4 | 5 | 10 |
|---|---|---|---|---|---|---|
| $\beta_h$ | -0.10 | -0.17 | -0.23 | -0.29 | -0.35 | -0.64 |
| OLS se | 0.05 | 0.07 | 0.08 | 0.09 | 0.09 | 0.10 |
| White se | 0.05 | 0.06 | 0.07 | 0.08 | 0.07 | 0.07 |
| NW se | 0.04 | 0.07 | 0.09 | 0.10 | 0.09 | 0.10 |
| HH se | 0.05 | 0.08 | 0.10 | 0.10 | 0.08 | 0.15 |
| $R^2$ | 0.04 | 0.07 | 0.09 | 0.12 | 0.16 | 0.36 |

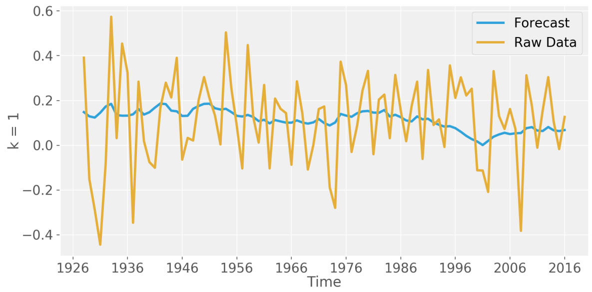
$$r_{t+1} + \cdots + r_{t+k} = \alpha_k + \beta_k (p_t - d_t) + u_{t+k,k}$$

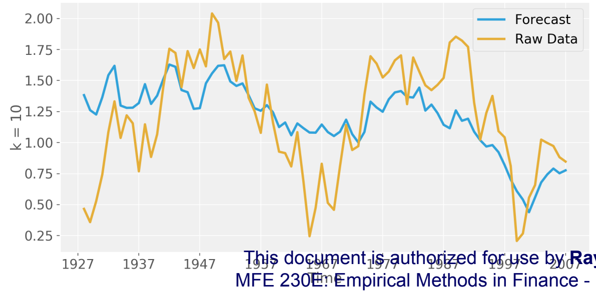
| | 1 | 2 | 3 | 4 | 5 | 10 |
|---|---|---|---|---|---|---|
| $\beta_h$ | -0.24 | -0.42 | -0.58 | -0.68 | -0.78 | -1.12 |
| OLS se | 0.09 | 0.13 | 0.15 | 0.16 | 0.16 | 0.17 |
| White se | 0.08 | 0.09 | 0.11 | 0.12 | 0.11 | 0.15 |
| NW se | 0.06 | 0.10 | 0.11 | 0.14 | 0.14 | 0.20 |
| HH se | 0.07 | 0.10 | 0.10 | 0.16 | 0.15 | 0.12 |
| $R^2$ | 0.10 | 0.15 | 0.20 | 0.22 | 0.28 | 0.43 |

$$\Delta d_{t+1} + \cdots + \Delta d_{t+k} = \alpha_k + \beta_k (p_t - d_t) + u_{t+k,k}$$

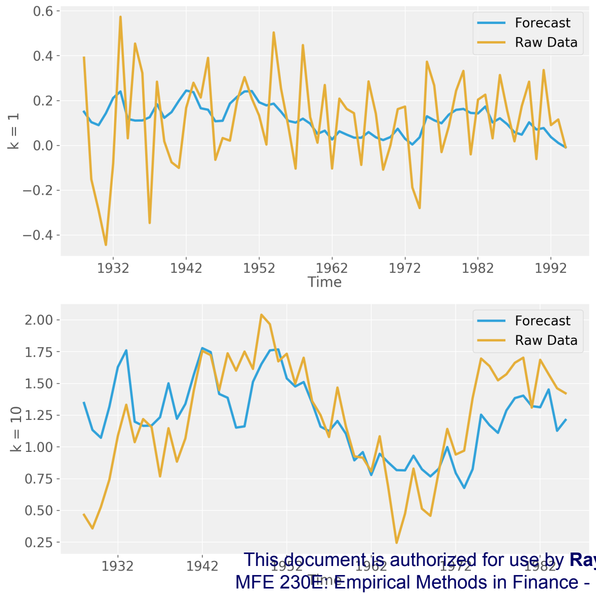
| | 1 | 2 | 3 | 4 | 5 | 10 |
|---|---|---|---|---|---|---|
| $\beta_h$ | 0.00 | 0.03 | 0.04 | 0.04 | 0.05 | 0.07 |
| OLS se | 0.04 | 0.05 | 0.06 | 0.07 | 0.07 | 0.08 |
| White se | 0.04 | 0.05 | 0.06 | 0.07 | 0.06 | 0.06 |
| NW se | 0.04 | 0.06 | 0.07 | 0.09 | 0.08 | 0.09 |
| HH se | 0.04 | 0.07 | 0.09 | 0.09 | 0.07 | 0.10 |
| $R^2$ | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 |

# Fitted returns

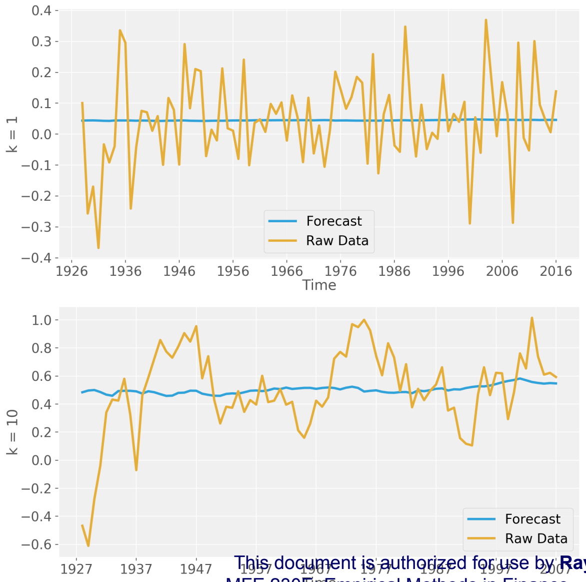


# Fitted returns: Data up to 1994



# FITTED DIVIDEND GROWTH



# Summary

- Any present-value relationship implies cointegration
- Granger representation theorem: Cointegration implies predictability
- Foundation of price-based valuation ratios
- Approximate framework for price-dividend ratio
- $p - d$ should forecast dividend growth and/or returns
- Historical data: Returns are forecastable but dividend growth is not
- **NOT** a violation of market efficiency!

Example: $k = 3$

$$r_1 + r_2 + r_3 = \alpha_k + \beta_3 (p_0 - d_0) + u_{3,3}$$

$$r_2 + r_3 + r_4 = \alpha_k + \beta_3 (p_1 - d_1) + u_{4,3}$$

$$r_3 + r_4 + r_5 = \alpha_k + \beta_3 (p_2 - d_2) + u_{5,3}$$

$$r_4 + r_5 + r_6 = \alpha_k + \beta_3 (p_3 - d_3) + u_{3,3}$$

- The left-hand-side variables have an $k-1$ overlap
- So will the residuals (at least under the null hypothesis of no forecastability)
- We need to correct the standard errors for heteroskedasticity in errors
- Standard choices: White and Newey-West corrections
- Here: autocorrelation of errors has a certain structure: $k-1$ periods
- The correction for $k-1$ overlap can be computed using GMM



## Appendix: Hansen-Hodrick correction in long-horizon regressions

Recall: OLS as GMM estimator

$$y_i = x_i'\beta + u_i$$

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) \stackrel{d}{\Rightarrow} N(\mathbf{0}, (\mathbf{D}\mathbf{S}^{-1}\mathbf{D}')^{-1})$$

$$\mathbf{D}' = E\left[\mathbf{x}_t\mathbf{x}_t'\right]$$

$$\mathbf{S} = \sum_{j=-\infty}^{\infty} E\left[(\mathbf{x}_t u_t)(\mathbf{x}_{t-j}u_{t-j})'\right]$$

Hansen and Hodrick (1980): Estimate the LH regression using OLS (which is equivalent to GMM) and use the estimated residuals to compute **S** as

$$\mathbf{S} = \sum_{j=-(k-1)}^{k-1} E\left[(\mathbf{x}_t u_t)(\mathbf{x}_{t-j}u_{t-j})'\right]$$

where $\mathbf{x}_t = (1, p_t - d_t)'$