Lecture Note 2
Stationarity, sample means, and robust regressions

Lars A. Lochstoer
UCLA Anderson School of Management

Winter 2025

Overview of Lecture Note 2

Prices vs. returns, and methods for robust inference

- Efficient markets, i.i.d. returns, and the Random Walk hypothesis
- Covariance stationarity: returns vs. prices
- The standard error of the mean return revisited: the central limit theorem
- Time-varying volatility and Generalized Least Squares
- Robust standard errors

A useful benchmark model of returns

Write log returns as:

$$r_t = \mu + \sigma\varepsilon_t, \quad \text{for all } t$$

where the error term, ε_t, has the following properties

1. Independent across time: $f\left(\varepsilon_t, \varepsilon_{t+j}\right) = f\left(\varepsilon_t\right) f\left(\varepsilon_{t+j}\right)$ for any t, j
2. Has mean zero: $E_{t-1}\left[\varepsilon_t\right] = 0$ for all t
3. Has unit variance: $Var_{t-1}\left[\varepsilon_t\right] = 1$
4. Has finite skewness and kurtosis (so that typical Central Limit and Law of Large Numbers theorems hold)

Notice: Returns have constant conditional mean and variance, but are not necessarily Normally distributed

The Random Walk hypothesis

Given this model, consider the log value of a portfolio, p_t, that earns this return each period and that has no distributions (all wealth is reinvested)

$$\begin{aligned} p_t &= p_{t-1} + r_t \\ &= p_{t-1} + \mu + \sigma\varepsilon_t. \end{aligned}$$

This value process is said to follow a Random Walk with Drift

- A Random Walk with Drift is a process with unforecasteable increments, except for a constant drift term (μ)
  - In particular, $\Delta p_t \equiv p_t - p_{t-1} = \mu + \sigma\varepsilon_t$, so $E_{t-1}\left[\Delta p_t\right] = \mu$, and $E_{t-1}\left[p_t\right] = p_{t-1} + \mu$

This is the original Efficient Markets model of Gene Fama (1970)

- If markets are efficient, you cannot forecast returns (other than the constant risk premium component)
- We recognize now that the risk premium ($\mu_t - r_{f,t}$) could be time-varying. More on this later in the class.

The Random Walk hypothesis

α in the plot is our μ

Covariance Stationarity

In this model, **prices are nonstationary** while **returns are stationary**

- Technically, we will operate with a notion of stationarity that is called *covariance stationarity*
- Such stationarity is an important condition for most of the econometric techniques you will encounter

Definition

A process $\{x_t\}_{t=-\infty}^{\infty}$ is **covariance stationary** if $E[x_t] = \mu$ and $Cov(x_t, x_{t+j}) = \gamma_j$ for all t and j. That is, the **unconditional** mean and covariances exist and are not a function of time t.

A corollary of this is, using the Law of Large Numbers, that the sample mean and covariances are consistent estimates of the true mean and covariances.

Prices and Stationarity

Let's consider the Random Walk model of prices

- We get the unconditional expectation by conditioning on the initial observation, p_0, and taking the limit as $t \to \infty$:

$$\lim_{t \to \infty} E\left[p_t | p_0\right] = \lim_{t \to \infty} p_0 + \mu t = \begin{cases} -\infty & \text{if } \mu < 0 \\ p_0 & \text{if } \mu = 0 \\ \infty & \text{if } \mu > 0 \end{cases}$$

- Thus, if $\mu \neq 0$, the unconditional mean does not exist and it is clear that for any finite t the expectation is a function of t.
- For $\mu = 0$, it looks like we're fine. But, we need to check the covariances as well. Let's check for $j = 0$, i.e. the variance:

$$\lim_{t \to \infty} Var\left[p_t | p_0\right] = \lim_{t \to \infty} t\sigma^2 = \infty$$

- Thus, the unconditional variance of a Random Walk does not exist

$\Longrightarrow$ **The wealth process is nonstationary!**

Returns and Stationarity

Let's consider the return process:

$$\begin{aligned} E\left[r_t\right] &= E\left[\mu + \sigma\varepsilon_t\right] = \mu \text{ for all } t \\ Var\left(r_t\right) &= Var\left(\mu + \sigma\varepsilon_t\right) = \sigma^2 \text{ for all } t \end{aligned}$$

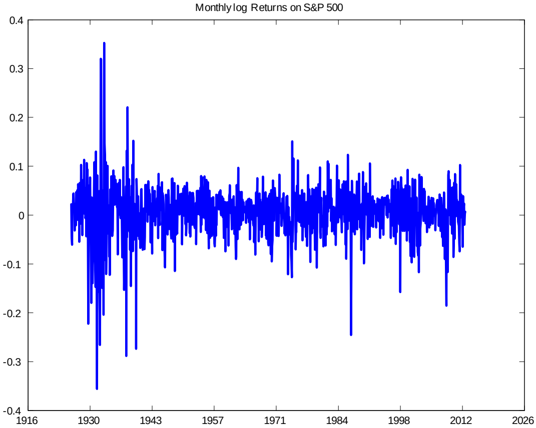
$\Longrightarrow$ **The return process is stationary!**

This is what holds in the data, as well. See next slide.

Stationary of returns in a picture

Note that it's the **unconditional** mean and variance that needs to be constant

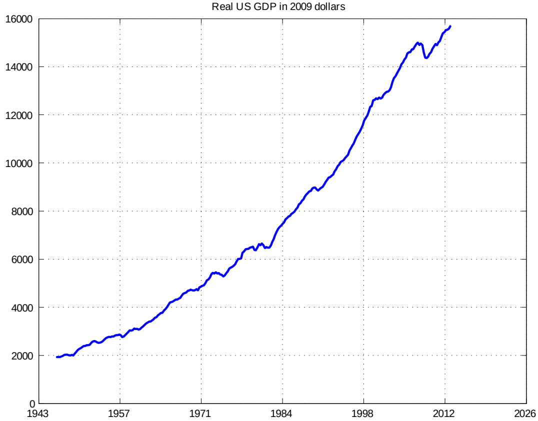
- The **conditional** mean and variance can move around



Nonstationarity in a picture

Aggregate output (GDP) and other macroeconomic series are nonstationary

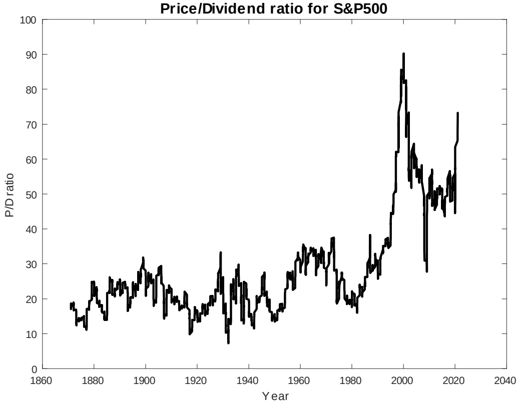
- At least, that's the consensus (more on this later)



What about valuation ratios?

Market price over cash dividends a stationary variable?

- S&P500 (Shiller data)
- Hard to say using eyeball econometrics...!



The sample mean revisited

Sample means are tremendously important in econometrics

- They make up the *moments* used for identification of parameters
- The mean and variance of a sample of returns $\{r_1, r_2, ..., r_T\}$ are:

$$
\begin{aligned}
m_T &\equiv E_T[r_t] = \frac{1}{T}\sum_{t=1}^{T} r_t \\
E[m_T] &= \frac{1}{T}\sum_{t=1}^{T} E[r_t] = \mu \\
E\left[(m_T - E[m_T])^2\right] &= E\left[\left(\frac{1}{T}\sum_{t=1}^{T}(r_t - \mu)\right)^2\right] \\
&= E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\sigma\varepsilon_t\right)^2\right] = \frac{\sigma^2}{T}
\end{aligned}
$$

Ergodicity

In order to do statistical inference on the sample mean, we need its distribution

- Enter the magic of the Central Limit Theorem!
- There are lots of them, with different assumptions. We will assume *ergocidity*, which is a condition that ensures that the variance of the sample mean is finite.
- In the scalar case we are operating in, it is sufficient to assume the infinite sum of the autocovariances is finite:

$$\sum_{j=0}^{\infty} \left|\gamma_j\right| < \infty$$

- This is trivially the case in our example, where returns are i.i.d. (so all $\gamma_j = 0$ for $j > 0$) with finite variance, σ^2

The Central Limit Theorem

Theorem

If the sample mean has finite variance and as $T \longrightarrow \infty$, the sample mean estimate $\bar{y}_T$ converges in distribution to a Normally distributed variable with mean equal to the true mean and variance equal to the infinite sum of autocovariances, S:

$$\sqrt{T}\left(\bar{y}_T - \mu\right) \sim N\left(0, S\right)$$

Thus, the sample mean in our example is distributed as follows:

$$m_T \sim N\left(\mu, \frac{\sigma^2}{T}\right)$$

and we can use the usual Normal 95% confidence band.

The sample mean with heteroskedasticity

Let's extend our model to include time-varying volatility:

$$r_t = \mu + \sigma_{t-1}\varepsilon_t$$

where $|\sigma_{t-1}| < \infty$ for all t and where $V[r_t] = \sigma^2$.

- Does this affect our test?

Notice that the central limit theorem only asks for the unconditional moments.

- Thus, the test is the same

In sum, despite the non-normalities found in the data, the Central Limit Theorem provides a robust testing framework as long as the sample is sufficiently large

OLS revisited

Let's next consider how time-varying volatility and non-normalities affects regressions

Recall OLS:

$$\underset{(T\times 1)}{Y} = \underset{(T\times K)}{X}\underset{(K\times 1)}{\beta} + \underset{(T\times 1)}{\varepsilon}.$$

The standard OLS assumption is that the error term is normally i.i.d. distributed: $\varepsilon_t \overset{i.i.d.}{\sim} N\left(0, \sigma^2\right)$ for all t

- Thus, the residuals' variance-covariance matrix is:

$$E\left[\varepsilon\varepsilon'\right] = \sigma^2 I_T$$

Heteroskedastic Error Terms

What if, as is typically the case for financial data, error terms are heteroskedastic?

- Let's stick with Normal distribution for now: $\varepsilon_t \sim N\left(0, \sigma_t^2\right)$
- Also, let error terms be uncorrelated across time: $E\left[\varepsilon_t \varepsilon_{t+j}\right] = 0$ for all $j \neq 0$.
- Now the residual variance-covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^2 \end{bmatrix}$$

Intuitively, when estimating the regression coefficients, you want to weight observations with lower residual variance (less noisy observations) more than observations with higher residual variance

Generalized Least Squares (GLS)

Matrix inversion can be tricky, but not with diagonal matrices

- Consider the matrix $\Sigma^{-1/2}$:

$$\Sigma^{-1/2} = \begin{bmatrix} \sigma_1^{-1} & 0 & \cdots & 0 \\ 0 & \sigma_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^{-1} \end{bmatrix}$$

Redefine the independent and dependent variables:

$$\tilde{Y} = \Sigma^{-1/2} Y \quad \text{and} \quad \tilde{X} = \Sigma^{-1/2} X$$

Consider the GLS regression:

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$$

- What is the covariance matrix of the GLS residuals?
  - Simply, I_T. Thus, OLS is optimal in this alternative regression!
- The regression coefficients thus can be written:

$$\hat{\beta}^{GLS} = \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1}Y \sim N\left(\beta_{null}, \left(X'\Sigma^{-1}X\right)^{-1}\right)$$

Feasible GLS

Issue: we need to know the variance-covariance matrix of the residuals before running the regresion

- **Feasible GLS** is a two-pass approach
  1. First pass: Run OLS, estimate σ_j^2 using $\hat{\sigma}_j^2 = \hat{\varepsilon}_{OLS,j}^2$ for $j = 1, ..., T$
  2. Second pass: Run GLS using $\hat{\sigma}_j^2$ instead of (the unknown) σ_j^2
- Issue: The $\hat{\sigma}_j^2$ are quite noisy estimates, can lead to very noisy $\hat{\beta}^{GLS}$ estimates
  - Defeats the purpose, which was efficiency gain

Many researchers prefer to run OLS and instead adjust the standard errors for the heteroskedasticity

- So-called 'robust standard errors'
- An asymptotic adjustment that relies on the Central Limit Theorem is also robust to unconditionally non-normal residuals

Asymptotic OLS

Consider the OLS regression:

$$y_t = x_t' \beta + \varepsilon_t,$$

where x_t and β are $K \times 1$ vectors

- ε_t is a mean-zero error term with variance $\sigma_t^2 < \infty$. It need not be Normally distributed.
- Assume as before that $E\left[\varepsilon_t \varepsilon_{t+j}\right] = 0$ for all $j \neq 0$

We still need the OLS identifying assumption:

$$E\left[x_t' \varepsilon_t\right] = 0$$

The OLS Moment Condition

Define the OLS moment condition for the estimated $\hat{\beta}$:

$$f_t\left(\hat{\beta}\right) = x_t\left(y_t - x_t'\hat{\beta}\right)$$

Let the sample mean of the moment condition be:

$$g_T\left(\hat{\beta}\right) = \frac{1}{T}\sum_{t=1}^{T} f_t\left(\hat{\beta}\right) = 0$$

From the Central Limit Theorem:

$$\sqrt{T} g_T\left(\hat{\beta}\right) \sim N\left(0, S_T\right)$$

where

$$S_T = \frac{1}{T}\sum_{t=1}^{T} f_t\left(\hat{\beta}\right) f_t\left(\hat{\beta}\right)' = \frac{1}{T}\sum_{t=1}^{T} x_t x_t' \hat{\varepsilon}_t^2$$

- In standard OLS, the squared error term is uncorrelated with $x_t x_t'$ as the variance is constant.

White (robust) standard errors

In the end, we want the distribution of $\hat{\beta}$

- Note that, asymptotically

$$g_T(\beta) = E[x_t y_t] - E[x_t x_t'] \beta$$

Thus

$$\hat{\beta} - \beta \sim N\left(0, \frac{1}{T} E_T[x_t x_t']^{-1} S_T E_T[x_t x_t']^{-1}\right)$$

With constant variance OLS, $S_T = E_T[x_t x_t'] E_T[\hat{\varepsilon}_t^2]$.

In sum, OLS regressions in large samples

- Are unbiased
- Standard errors need to be adjusted for heteroskedasticity
- Do not require normally distributed errors

We will deal with cases where $E[\varepsilon_t \varepsilon_{t+1}] \neq 0$ later

Take-aways

1. Make sure you are estimating your model using stationary data
   - Historical samples "representative," Central Limit Theorem applies
   - Possible to work with non-stationary data using cointegration analysis, but in practice not much used

2. OLS regressions are unbiased and yield correct inference if sample is large
   - Do not need Normally distributed errors or constant variance of residuals (homoscedasticity)
   - Adjust standard errors (robust; White (1980)), we will do Newey-West for autocorrelation later
   - Explicitly modeling non-normalities and heteroskedasticity in a small sample often entails estimation error that outweighs potential benefit