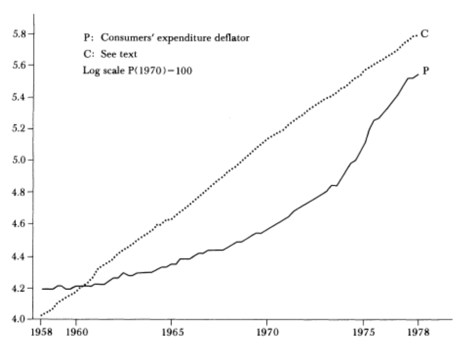
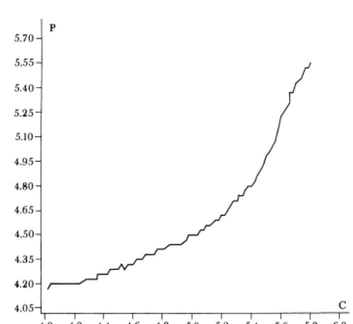
1 Data Issues

1.1 Stationarity

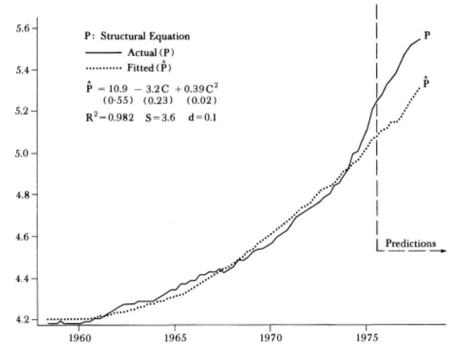
- Assumptions Needed for GMM
  - Since we're going to do a programming implementation of GMM today, we're going to start out talking about the assumptions needed for GMM
    - Last time, we introduced the concept of stationarity
  - The idea is that for, GMM, the data need to be 'weakly stationary' or 'covariance stationary'
    - Recall that this just means that the first and second moments of the data must be finite and independent of time
    - Why? ...
- Why Stationarity?
  - Consider the following example...
- Example: Hendry's Theory of Inflation
  - A certain variable, denoted "C", is the real cause of inflation in Great Britain.
    - Hendry is certain the variable is exogenous, causality is from C to P only, and C is outside of the government's control.
  - Quarterly Time Series (Seasonally Unadjusted)



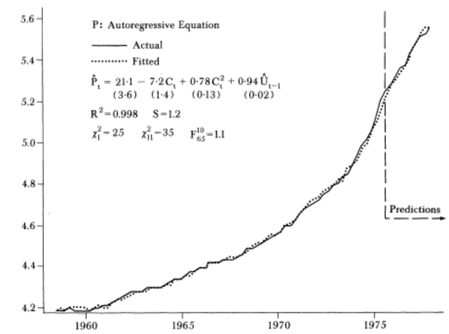
  - Cross-Plot of P against C (in logs)
    - There is a clear close, but nonlinear, relationship



- Regression analysis assuming a quadratic equation.
  - There is a "good fit", the coefficients are "significant", but autocorrelation remains and the equation predicts badly.



- Assuming a first-order autoregressive error process...
  - The fit is "spectacular", the parameters are "highly significant", and there is no obvious residual autocorrelation ("eyeball test"), and the predictive test does not reject the model.



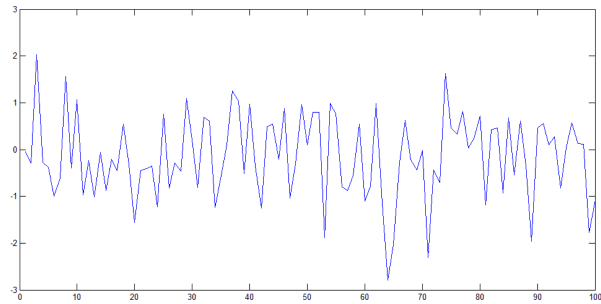
- Looks good, right?

- So Why Stationarity?
  - First: Spurious regressions
    - Two variables might be totally unrelated, but if they are - for example - both trending upwards over time, a regression of one on the other could have a high R^2
  - Second: The (non)stationarity of a time-series can influence its behavior and properties
    - E.g., the persistence of shocks can be infinite for a non-stationary series
  - Validity of standard assumptions for asymptotic analysis

* We cannot do hypothesis tests about regression parameters if the variables in the regression model are not stationary - standard assumptions are not valid, so, e.g., the usual 't-ratios' won't follow a t-distribution

- Example of Stationarity
  - An example of a stationary i.i.d. time-series is just

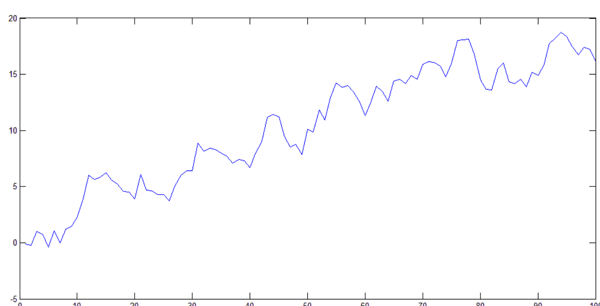
$$Y_t = a + \varepsilon_t \text{ where } \varepsilon_t \sim iid\ N\left(0, \sigma^2\right)$$



1.2 Non-Stationarity

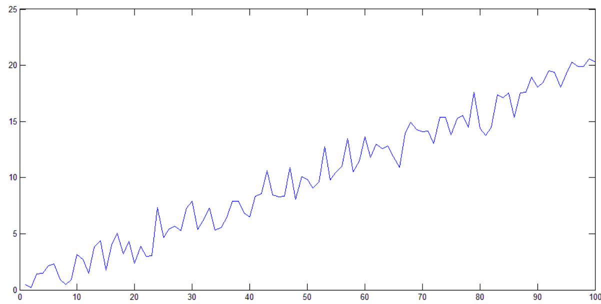
- Examples of Non-Stationarity
  - Two models are frequently used to characterize non-stationary:
  - The random walk model with drift

$$Y_t = \mu + Y_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim iid\ N\left(0, \sigma^2\right)$$



  - And the deterministic trend process

$$Y_t = \alpha + \beta t + \varepsilon_t \text{ where } \varepsilon_t \sim iid\ N\left(0, \sigma^2\right)$$



- Stochastic Non-Stationarity
  - We can generalize the first model to

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim iid\ N\left(0, \sigma^2\right)$$

  - Let's let ϕ be any value for now...
  - Consider an example of an $AR(1)$ with no drift

$$Y_t = \phi Y_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim iid\ N\left(0, \sigma^2\right)$$

  - We can write

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \varepsilon_t \\ Y_{t-1} &= \phi Y_{t-2} + \varepsilon_{t-1} \\ Y_{t-2} &= \phi Y_{t-3} + \varepsilon_{t-2} \end{aligned}$$

  - So that with successive substitutions, we get

$$Y_t = \phi^T Y_0 + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + ... + \phi^T \varepsilon_0 + \varepsilon_t$$

  - For the impact of shocks,
    * if $\phi < 1$, $\phi^T \longrightarrow 0$ as $T \longrightarrow \infty$ so the shocks eventually die out
    * if $\phi = 1$, $\phi^T = 1$ so we have shocks that are persistence and never die out. We then get a time series that is basically an infinite sum of past shocks
  - Given this, we therefore usually use $\phi = 1$ to characterize the non-stationarity because
    * If $\phi > 1$, we have an 'explosive' process
    * We usually ignore this case ($\phi > 1$) because $\phi > 1$ does not describe many data series in economics and finance and it has the (intuitively unappealing) property that shocks are propagated through time so that they have an increasingly large influence as time goes by.

- Detrending

- Going back to our two characterizations on non-stationarity

$$Y_t = \mu + Y_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \sim iid\ N\left(0, \sigma^2\right)$$

and

$$Y_t = \alpha + \beta t + \varepsilon_t \text{ where } \varepsilon_t \sim iid\ N\left(0, \sigma^2\right)$$

- We can "difference" the first case (called stochastic non-stationarity) so that we get a "difference stationary" series

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

- And we can "de-trend" the second case by running an OLS regression to estimate $Y_t = \widehat{\alpha} + \widehat{\beta} t + \varepsilon_t$ and then getting a "trend stationary" series

$$Y_t^{dt} = Y_t - \widehat{\alpha} - \widehat{\beta} t$$

- What do the graphs look like?
  - If you were to graph aggregate consumption and stock prices, they trend up over time...
  - However, time-series for consumption growth and stock returns do not...
- Which Method?
  - What if we use the wrong model?
  - Although trend stationary and difference stationary series both "trend" over time, we need to use the correct method
    - If we difference a trend-stationary series, we remove the non-stationarity, but we introduce a $MA(1)$ structure into the errors
    - If we detrend a series that has a stochastic trend, then we don't even remove the non-stationarity
  - (Keep in mind that most series in economics and finance are probably best described by the stochastic non-stationary model.)
  - We have to work with stationary time series, but we can deal with error terms in our models that are not i.i.d.
    - Here, we'll use our HAC errors that we introduced last time.
    - Note that tests are more easily derived for i.i.d. errors, but error terms in finance are never i.i.d.!

1.3 Other Questions

- There are lots of important questions that we need to ask of our data before we even being to look at implementing asset pricing models...
- Stationarity is one, but timing is also a huge issue

  - Consider the consumption-based model.
  - One problem is that consumption is measured as the **average** over a quarter, while returns are measured **point-to-point**
  - So how do you line these up? It's not completely clear. If you get this wrong, the model totally fails.
  - E.g., Suppose the correlation between R_{t+1} and $\triangle c_{t+1}$ is strong and your data for consumption and returns are roughly iid. Then the correlation between R_{t+1} and $\triangle c_{t+2}$ or $\triangle c_t$ is zero.

- Other issues include a consideration of real vs. nominal measurements - how do we deflate?
- What measure of consumption should we use?
- And whose consumption do we use?
  - Do we use aggregate consumption, or per capital consumption?
  - Should we use everyone's consumption, or only stockholder consumption levels?
- Units?

2 GMM with Matlab

2.1 A Simple Example for GMM

- Let's suppose that we are considering the consumption-based asset pricing model, so that we have, generally

$$p_t = E_t \left(m_{t+1} x_{t+1} \right)$$

  and

$$m_{t+1} = \beta \frac{u' \left(c_{t+1} \right)}{u' \left(c_t \right)}$$

- Now, we are going to consider a very simple way to implement this model:
  - Suppose that we have a series of excess returns
  - And we specify power utility and assume no time discounting
    - Then we are only estimating γ

2.2 Solving for the Parameters

- Population Moment (The $f(\cdot)$'s)

$$E \left(m_{t+1} R^e_{t+1} \right) = 0$$

  with

$$m_{t+1} = \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma}$$

  Sample Analogue's (The $g(\cdot)$'s)

$$g_T = \frac{1}{T} \sum_{t=1}^{T} \left(m_{t+1} R^e_{t+1} \right) = 0$$

- GMM will use this equation to pick γ to set the average pricing errors to zero
  - Note that this does NOT mean that per-period errors will be small!
- We're going to use consumption data and data on different sets of excess returns, so in the previous equations
  - Each R^e_{t+1} is $T \times 1$
  - m_{t+1} is $T \times 1$
  - 0 is 1×1
- If we have more than one set of returns for different assets, each one with have a separate moment equations. I.e., if we have two sets of excess returns,

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\left(m_{t+1}R^{e1}_{t+1}\right) &= 0 \\ \frac{1}{T}\sum_{t=1}^{T}\left(m_{t+1}R^{e2}_{t+1}\right) &= 0 \end{aligned}$$

- The computer code we are going to build will solve for γ.

2.3 Calculating Standard Errors

- For **each** moment equation, we have a set of residuals. Again, if we have two sets of excess returns, we have

$$\begin{aligned} u^1_{t+1} &= R^{e1}_{t+1}\left(\frac{c_{t+1}}{c_t}\right)^{-\widehat{\gamma}} \\ u^2_{t+1} &= R^{e2}_{t+1}\left(\frac{c_{t+1}}{c_t}\right)^{-\widehat{\gamma}} \end{aligned}$$

- Now, to just calculate the variance-covariance matrix, $\widehat{\Gamma}_0$, we write

$$z = \left[\begin{array}{cc} u^1 & u^2 \end{array}\right]$$

and

$$\widehat{\Gamma}_0 = \frac{1}{T}\sum_{t=1}^{T} z'z$$

- If we just have one set of residuals, this is just the formula for the variance of a sample mean $(1/T)\,E\,(u'u)$
- Is the variance-covariance matrix, $\widehat{\Gamma}_0$, good enough?
  - If the residuals from the moment conditions are i.i.d...
  - I.e., the u's would need to be uncorrelated over time
- But we can do more...
  - It's easy enough to correct for autocorrelation and heteroskedasticity...

2.4 Robust Standard Errors: Empirical Implementation of S

- "Newey-West" errors are HAC
- $\widehat{S}$ is now defined as

$$\widehat{S} = \widehat{\Gamma}_0 + \sum_{t=1}^{T} \left(1 - \frac{v}{q+1}\right) \left(\widehat{\Gamma}_v + \widehat{\Gamma}_v'\right)$$

where

$$\widehat{\Gamma}_v = \frac{1}{T-v} \sum_{t=v+1}^{T} g\left(\overline{Y}_t, \widehat{b}\right)' g\left(\overline{Y}_t, \widehat{b}\right)$$

- Note that q is the maximum number of lags to use
  - Rule of thumb: $q =$ cube root of the square root of the number of observations
  - We're using the "Bartlett" kernel here (efficient, but biased)
  - Note when $q = 0$, we just back to the variance-covariance matrix for the moment conditions
- For

$$\widehat{S} = \widehat{\Gamma}_0 + \sum_{t=1}^{T} \left(1 - \frac{v}{q+1}\right) \left(\widehat{\Gamma}_v + \widehat{\Gamma}_v'\right)$$

- $\widehat{\Gamma}_0$ is just the variance-covariance matrix (like OLS)
- The second term in that expression is the adjustment for autocorrelation and heteroskedasticity
  - This correction is easy enough to add to our program...

2.5 Standard Errors: The d Matrix

- Now, we need the d matrix for our standard errors. Recall,

$$\widehat{b}_{GMM} \overset{a}{\sim} N\left[b, \frac{1}{T}\left(d\widehat{S}^{-1}d'\right)^{-1}\right]$$

where d is just the derivative of the moment condition(s) w.r.t. b,

$$d = \frac{\partial g\left(\overline{Y}_T, b\right)}{\partial b}$$

- For 2 moment conditions and 1 parameter to estimate,

$$d = \frac{\partial g\left(\cdot\right)}{\partial b}|_{\widehat{b}}$$

is 2×1.

- For γ, with each i^{th} moment equation, we have that

$$\begin{aligned} \frac{\partial g_i\left(\cdot\right)}{\partial \gamma}|_{\widehat{\gamma}} &= \frac{\partial\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{t+1}^{ei}\right)}{\partial \gamma}|_{\widehat{\gamma}} \\ &= -\frac{1}{T}\sum_{t=1}^{T} \ln\left(\frac{c_{t+1}}{c_t}\right)\left(\frac{c_{t+1}}{c_t}\right)^{-\widehat{\gamma}} R_{t+1}^{ei} \end{aligned}$$

- Recalling, that $\left[\frac{d(a^x)}{dx} = \ln a \cdot a^x\right]$
- And just for good measure, suppose that we were actually estimating two parameters, β and γ.
- Then our d matrix would be 2×2 and we would also have

$$\begin{aligned}
\frac{\partial g_i(\cdot)}{\partial \beta}|_{\widehat{\beta},\widehat{\gamma}} &= \frac{\partial \left(\frac{1}{T}\sum_{t=1}^{T} \beta \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{t+1}^{ei}\right)}{\partial \beta}|_{\widehat{\beta},\widehat{\gamma}} \\
&= \frac{\partial \left(\frac{1}{T}\sum_{t=1}^{T} \beta \left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R_{t+1}^{i} - 1\right)}{\partial \beta}|_{\widehat{\beta},\widehat{\gamma}} \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(\frac{c_{t+1}}{c_t}\right)^{-\widehat{\gamma}} R_{t+1}^{(e)i}
\end{aligned}$$

and similarly

$$\frac{\partial g_i(\cdot)}{\partial \gamma}|_{\widehat{\beta},\widehat{\gamma}} = -\frac{1}{T}\sum_{t=1}^{T} \widehat{\beta} \ln\left(\frac{c_{t+1}}{c_t}\right)\left(\frac{c_{t+1}}{c_t}\right)^{-\widehat{\gamma}} R_{t+1}^{(e)i}$$

- What would we need to estimate β too?
  - (Is having just a set of excess returns enough?)

2.6 The VCV Matrix

- So we have that

$$d = \begin{bmatrix} \frac{\partial g_1(\cdot)}{\partial \gamma} \\ \frac{\partial g_2(\cdot)}{\partial \gamma} \\ \vdots \end{bmatrix} \text{ or } = \begin{bmatrix} \frac{\partial g_1(\cdot)}{\partial \gamma} & \frac{\partial g_1(\cdot)}{\partial \beta} \\ \frac{\partial g_2(\cdot)}{\partial \gamma} & \frac{\partial g_2(\cdot)}{\partial \beta} \\ \vdots & \vdots \end{bmatrix}$$

- And now we can get our VCV matrix (2×2) for b

$$V = inv\left(d'\widehat{S}^{-1}d\right)$$

  - So, the $(1,1)$ element of V is the variance of γ, and the $(2,2)$ element of V is the variance of β
- If we are only estimating γ, V is just 1×1 (or a scalar)