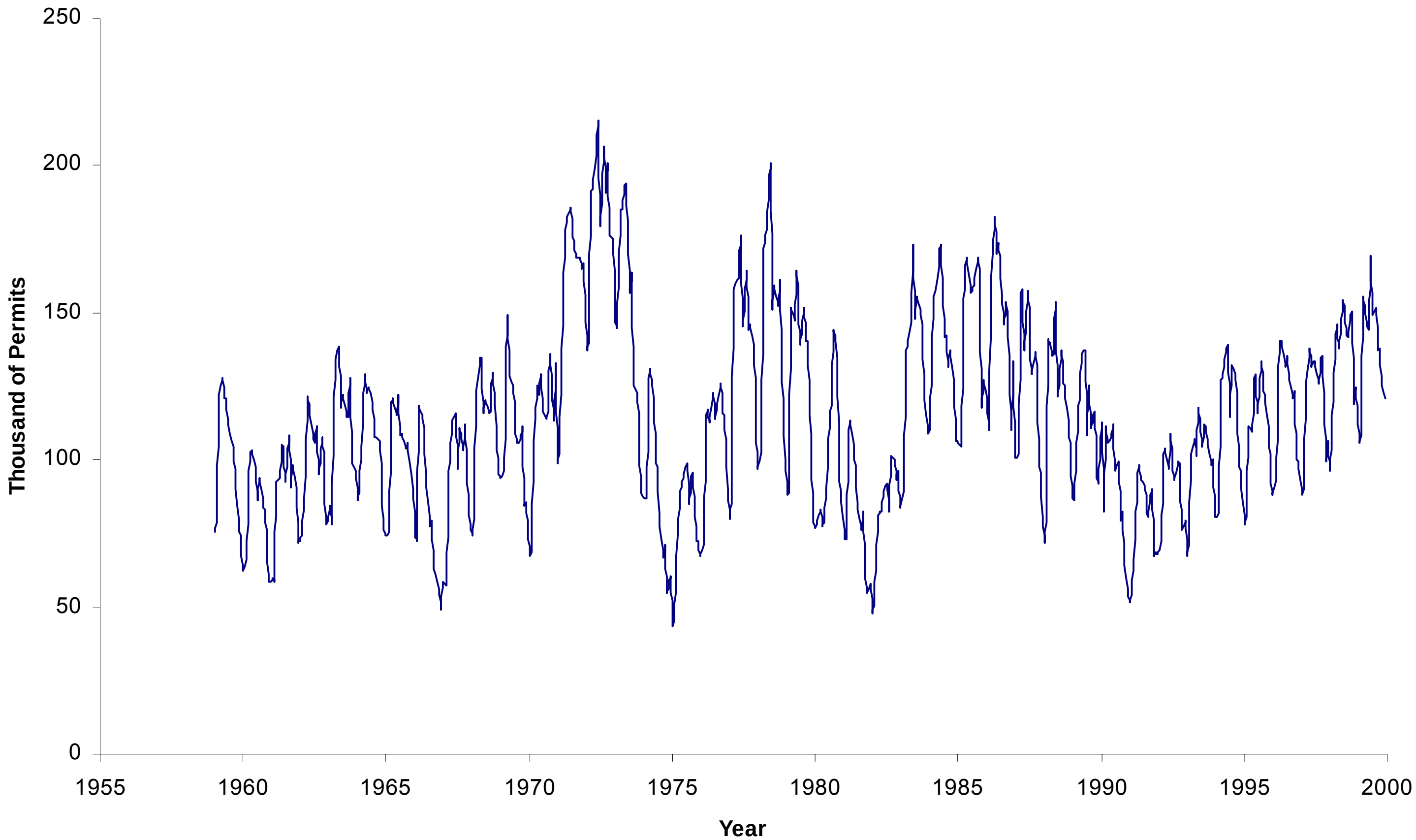
Topic 1: Linear Filters and Descriptive Statistics

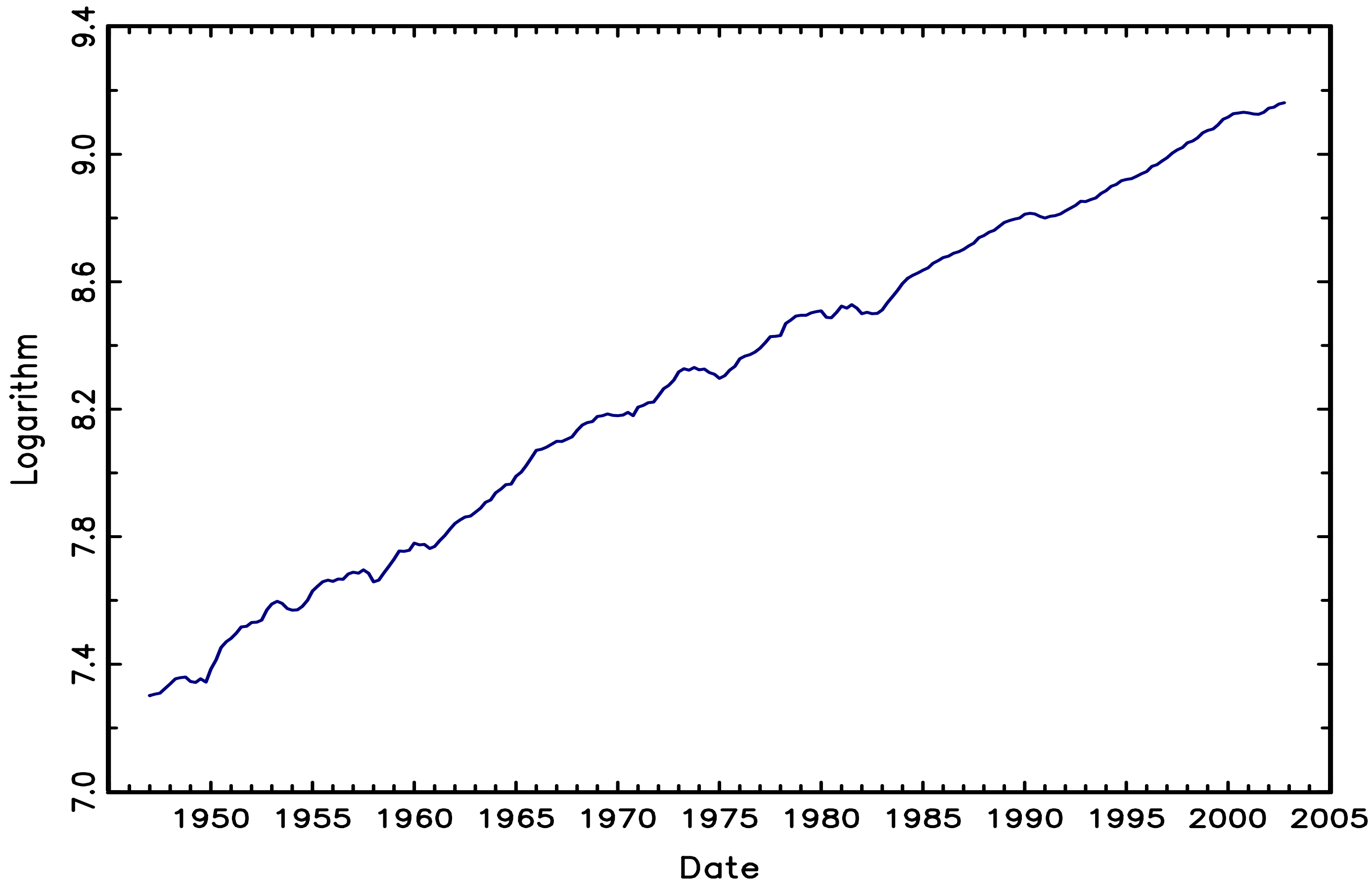
Examples:

- Building Permits
- Growth and Cycles in U.S. GDP

Figure 1
Building Permits in the United States

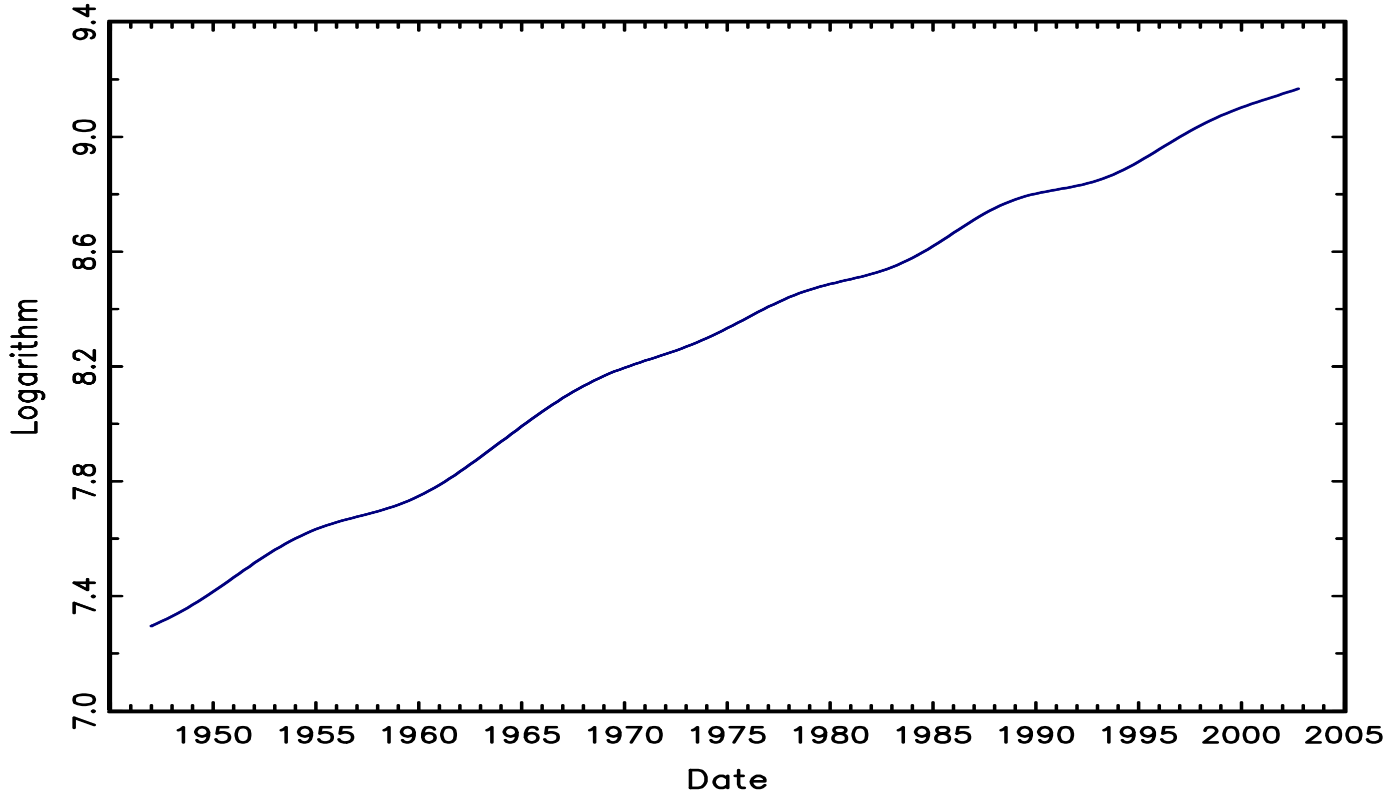


Logarithm of Real GDP in the United States

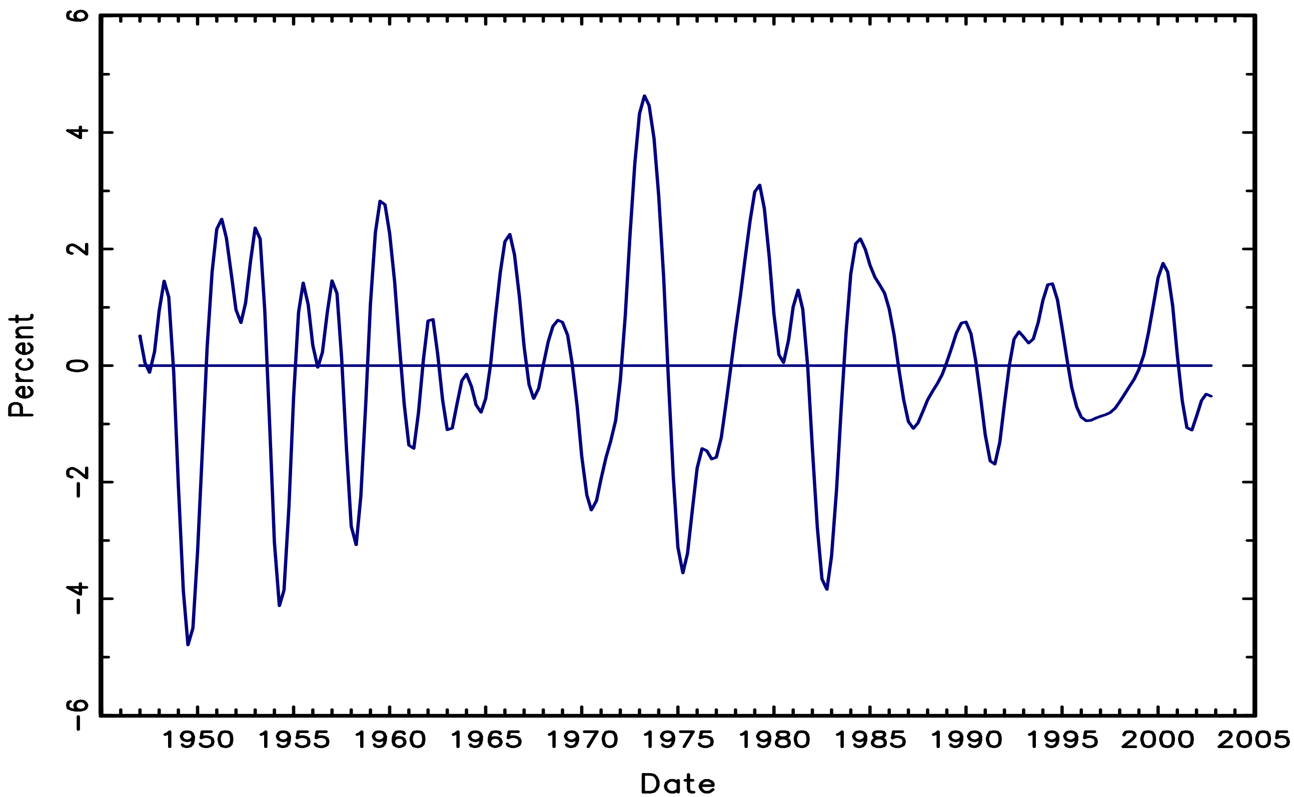


Trend Component

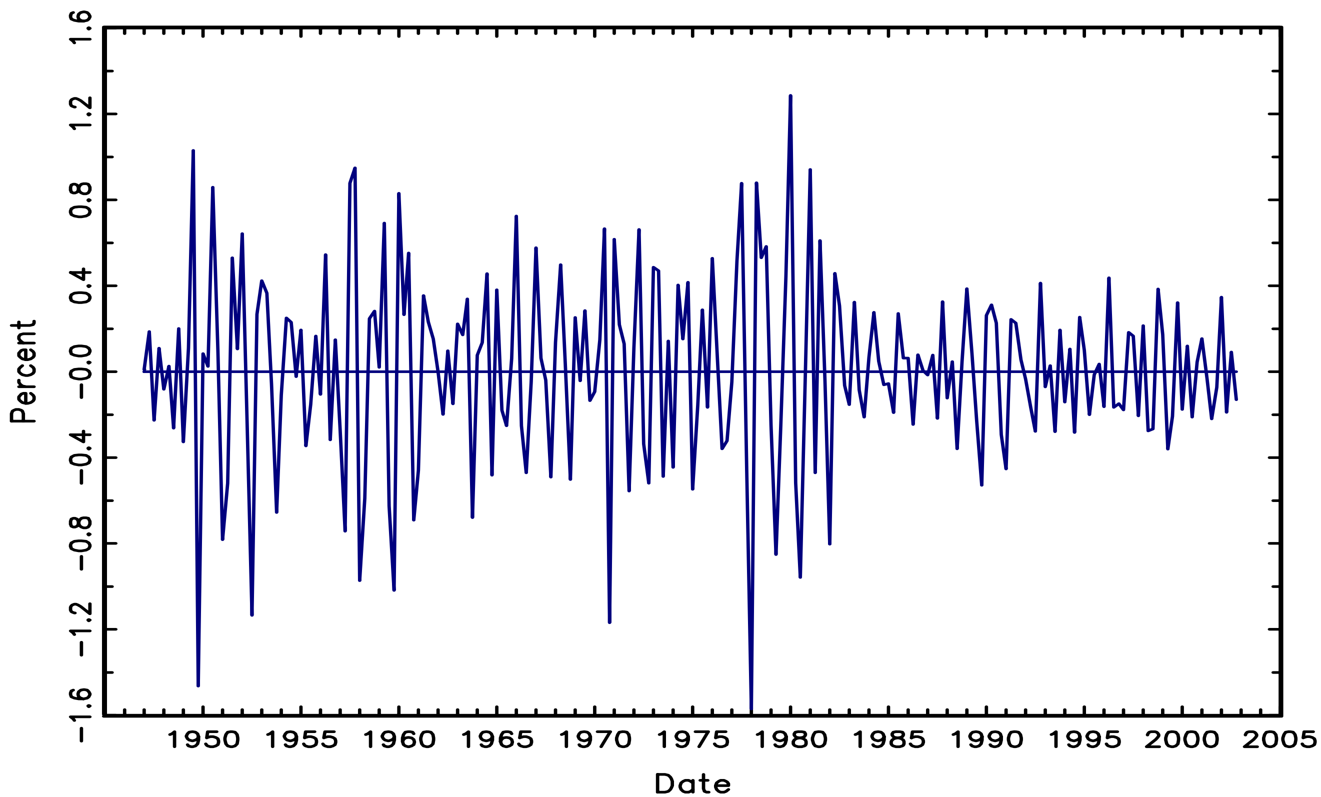
(Period > 32 Quarters)



Business Cycle Component
(6 Quarters ≤ Period ≤ 32 Quarters)



Irregular Component
(Period < 6 Quarters)



Time Domain Descriptive Statistics:

- Autocovariances: The autocovariances of a covariance stationary process are $\lambda_k = cov(Y_t, Y_{t+k}^{'})$
- Autocorrelations: The autocorrelations of a covariance stationary process are $\rho_k = cor(Y_t, Y_{t+k}^{'})$
- Autocovariance Generating Function: The autocovariance generating function is $\lambda(z) = \sum_{j=-\infty}^{\infty} \lambda_j z^j$

Autocovariance Generating Functions and Some ARMA algebra

The autocovariance generating function for a covariance stationary process is given by

$$\lambda(z) = \sum_{j=-\infty}^{\infty} \lambda_j z^j$$

so that the autocovariances are given by the coefficients on the argument z^j. Its purpose (or one of the purposes) is the same as the moment generating function – namely it is a convenient way to "store" the autocovariances of a covariance stationary stochastic process.

For the MA process, the ACGF is particularly easy to construct. Suppose

$$Y_t = \theta(L)\varepsilon_t$$

then

$$\lambda(z) = \sigma^2\theta(z)\theta(z^{-1})$$

We will verify this formula for the MA(1) process – you should verify it for higher order MA processes.

For an MA(1) process $\theta(z) = (1-\theta z)$, so that

$$\theta(z)\theta(z^{-1}) = (1-\theta z)(1-\theta z^{-1}) = -\theta z^{-1} + 1 + \theta^2 + \theta z$$

which implies autocovariances

$$\begin{aligned} \lambda_{-1} &= -\sigma^2\theta \\ \lambda_0 &= \sigma^2(1+\theta^2) \\ \lambda_1 &= -\sigma^2\theta \end{aligned}$$

with all other autocovariances equal to 0. Thus, the formula yields the correct answer for an MA(1) process.

Invertibility and non-uniqueness of MA processes

Consider the MA polynomial for the MA(q) model

$$\theta(z) = (1 - \theta_1 z - \theta_2 z^2 - \ldots - \theta_q z^q)$$

Suppose that this polynomial has zeros at $z = \gamma_1^{-1}, \gamma_2^{-1}, \ldots, \gamma_q^{-1}$. In this case, we can factor the polynomial as

$$\theta(z) = (1 - \gamma_1 z)(1 - \gamma_2 z) \ldots (1 - \gamma_q z)$$

and so the ACGF is given by

$$\lambda(z) = \sigma^2[(1 - \gamma_1 z)(1 - \gamma_2 z) \ldots (1 - \gamma_q z)(1 - \gamma_1 z^{-1})(1 - \gamma_2 z^{-1}) \ldots (1 - \gamma_q z^{-1})].$$

But, since $(1 - \gamma z)(1 - \gamma z^{-1})$ is proportional to $(1 - \gamma^{-1} z)(1 - \gamma^{-1} z^{-1})$ [that is $(1 - \gamma z)(1 - \gamma z^{-1}) = \gamma^2 \; (1 - \gamma^{-1} z)(1 - \gamma^{-1} z^{-1})$], we can "flip" or invert the roots of the MA polynomial, change σ^2 (to adjust for the factor of proportionality), and obtain the same ACGF – hence a model with the same autocovariances.

Thus, for example, in the MA(2) model, if

$$\theta(z) = (1 - \gamma_1 z)(1 - \gamma_2 z)$$

then models with

$$\theta_1(z) = (1 - \gamma_1^{-1} z)(1 - \gamma_2 z)$$

$$\theta_2(z) = (1 - \gamma_1 z)(1 - \gamma_2^{-1} z)$$

and

$$\theta_3(z) = (1 - \gamma_1^{-1} z)(1 - \gamma_2^{-1} z)$$

are observationally equivalent.

The ACGF for the ARMA can be derived as follows. Since

$$\phi(L)Y_t = \theta(L)\varepsilon_t$$

then

$$Y_t = c(L)\varepsilon_t$$

with $c(L) = \phi(L)^{-1}\theta(L)$ which is a well defined (mean square convergent) polynomial in positive powers of L (*i.e.*, backward looking) if the roots of $\phi(z)$ are greater than 1 in absolute value. Thus Y_t has the MA representation $Y_t = c(L)\varepsilon_t$ so that

$$\begin{aligned}\lambda(z) &= \sigma^2 c(z)c(z^{-1}) \\ &= \sigma^2\phi(z)^{-1}\theta(z)\phi(z^{-1})^{-1}\theta(z^{-1})\end{aligned}$$

Examples − Sums of ARMAs

$X_t = Y_t + W_t$ where Y and W are independent ARMA processes

1. Y ~ MA(1) and W ~ white noise

2. Y ~ AR(1) and W ~ white noise

3. Y ~ ARMA(p_Y, q_Y) and W ~ ARMA(p_W, q_W)

Frequency Domain Descriptive Statistics: Introduction to Spectral Analysis

- Primarily Reference: Hamilton pages 152-164
- Secondary Reference: Watson, M.W. "Time Series: Cycles," *International Encyclopedia of the Social and Behavioral Sciences*, (and the references cited therein).

Spectral Analysis is motivated by the Spectral Representation Theorem (sometimes called the Cramér Representation) of a covariance stationary stochastic process, given by

$$Y_t = \int_0^{\pi} \cos(\omega t) d\alpha(\omega) + \int_0^{\pi} \sin(\omega t) d\delta(\omega)$$

where $d\alpha(\omega)$ and $d\delta(\omega)$ are zero mean random variables that are mutually uncorrelated, uncorrelated across frequency, but having variances that depend on frequency. The representation decomposes Y into a set of strictly periodic components – each uncorrelated with the others, and each with it's own variance.. Thus for example, processes with important seasonal components will have large values of the components corresponding to the seasonal frequency, and similarly for series with strong "business cycle" components.

It is useful to digress and review some facts from trigonometry

- Let $Y_t = \cos(\omega t)$

  Then Y_t is strictly periodic with

  $period = \dfrac{2\pi}{\omega}$

- This follows from $\cos(a + 2\pi) = \cos(a)$. That is

  $Y_{t+\frac{j2\pi}{\omega}} = Y_t$ for $|j| = 1, 2, \ldots$

- A similar result holds for $Y_t = \sin(\omega t)$

- $\sin(0) = 0;\ \cos(0) = 1;\ \sin(\pi/2) = 1;\ \cos(\pi/2) = 0;\ \sin(\pi) = 0;\ \cos(\pi) = -1;$

  $\sin(\omega) = -\sin(-\omega);\ \cos(\omega) = \cos(-\omega)$

- $e^{i\omega t} = \cos(\omega t) + i\sin(\omega t)$ where $i = \sqrt{-1}$

To motivate the Cramér representation we begin by constructing a very simple periodic stochastic process

$$Y_t = \alpha \cos(\omega t) + \delta \sin(\omega t)$$

where α and δ are random variables with

$$E(\alpha) = E(\delta) = 0$$
$$E(\alpha\delta) = 0$$
$$E(\alpha^2) = E(\delta^2) = \sigma^2$$

While simple, this process has three attractive characteristics.

1. It is periodic: Y_t repeats itself with a period of $\frac{2\pi}{\omega}$.
2. The random components α and δ give Y a random amplitude (value of at its peak) and a random phase (value at $t = 0$). Thus, two realizations of Y will have different amplitudes and different timing of their peaks of troughs.
3. $E(Y_t) = 0$
4. $E(Y_t^2) = \sigma^2[\cos^2(\omega t) + \sin^2(\omega t)] = \sigma^2$
5. $E(Y_t Y_{t-k}) = \sigma^2[\cos(\omega t)\cos(\omega(t-k)) + \sin(\omega t)\sin(\omega(t-k))] = \sigma^2 \cos(\omega k)$

   where this final result uses the trigonometric identities

$$\cos(u+v) = \cos(u)\cos(v) - \sin(u)\sin(v)$$
$$\sin(u+v) = \sin(u)\cos(v) + \sin(v)\cos(u)$$

The final three results imply that the process is covariance stationary with $E(Y_t) = 0$ and $\lambda_k = \sigma^2 \cos(\omega k)$.

We can construct a more general process as

$$Y_t = \sum_{j=1}^{n} [\alpha_j \cos(\omega_j t) + \delta_j \sin(\omega_j t)]$$

with

$$E(\alpha_j) = E(\delta_j) = 0, \text{ for all } j$$
$$E(\alpha_j \delta_k) = 0, \text{ for all } j, k$$
$$E(\alpha_j \alpha_k) = E(\delta_j \delta_k) = 0, \text{ for } j \neq k$$
$$E(\alpha_j^2) = E(\delta_j^2) = \sigma_j^2$$

Then, as above

$$E(Y_t) = 0$$

$$Var(Y_t) = \sum_{j=1}^{n} \sigma_j^2$$

$$cov(Y_t Y_{t-k}) = \sum_{j=1}^{n} \sigma_j^2 \cos(\omega_j k)$$

Thus variance of Y can be decomposed in pieces, with each corresponding to a strictly periodic component ($\alpha_j \cos(\omega_j t) + \delta_j \sin(\omega_j t)$) with different period ($\frac{2\pi}{\omega_j}$) and with random amplitude, with variances that depend on frequency (σ_j^2).

The Spectral Representation is a generalization of this where the “increments” $d\alpha(\omega)$ and $d\delta(\omega)$ have orthogonality properties like those of α_j and δ_j. The variance of these components is frequency specific (like σ_j^2) which we now summarize with a density function $S(\omega)$. The function $S(\omega)$ is called the spectral density (or sometimes just the spectrum) of the process Y_t.

Sometimes the spectral representation is written using complex numbers as:

$$Y_t = \int_{-\pi}^{\pi} e^{i\omega t} d\gamma(\omega) \tag{1}$$

where $d\gamma(\omega) = \frac{1}{\sqrt{2}}[d\alpha(\omega) - id\delta(\omega)]$ for $\omega \geq 0$ and $d\gamma(\omega) = \overline{d\gamma(-\omega)}$ for $\omega < 0$. This representation simplifies notation for some calculations. The variance of $d\gamma(\omega)$ is $E(d\gamma(\omega)\overline{d\gamma(\omega)}) = S(\omega)d\omega$. (The use of a density function to summarize the variance of $d\gamma(\omega)$ is not completely general and rules out processes with deterministic (or perfectly predictable) components.)

You may wonder why the spectral representation doesn't use frequencies larger than π. The answer is that these are not needed to describe a process measured at discrete intervals. This discreteness means that periodic components associated with frequencies larger than π will look just like components associated with frequencies less than π. For example, a component with frequency 2π will have period of 1 and, because the series is measured only once every period, this component will appear to be constant – it will look just like (be *aliased*) with a component that has a frequency of $\omega = 0$.

There is a one-to-one relationship between the spectral density function and the autocovariances of the process. The autocovariances follow directly from the spectral representation:

$$\lambda_k = E(Y_t Y_{t-k}) = E(Y_t \overline{Y}_{t-k}) = E[\int_{-\pi}^{\pi} e^{i\omega t} d\gamma(\omega) \int_{-\pi}^{\pi} e^{-i\omega(t-k)} \overline{d\gamma(\omega)}] \tag{2}$$

$$= \int_{-\pi}^{\pi} e^{i\omega k} S(\omega) d\omega.$$

The spectrum can be determining by inverting this relation:

$$S(\omega) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \lambda_j e^{-i\omega j} \tag{3}$$

(To verify that (3) is the inverse of (2), use

$$\int_{-\pi}^{\pi} e^{i\omega k} d\omega = \begin{Bmatrix} 2\pi \text{ for } k = 0 \\ 0 \text{ for } k \neq 0 \end{Bmatrix} \tag{4}$$

so that $(2\pi)^{-1}\{\int_{-\pi}^{\pi} \sum_{j=-\infty}^{\infty} \lambda_j e^{-i\omega(j-k)} d\omega\} = \lambda_k$.)

A simpler formula for the spectrum is

$$S(\omega) = (2\pi)^{-1}[\lambda_0 + 2\sum_{j=1}^{\infty} \lambda_j \cos(j\omega)] \tag{5}$$

which follows from (3), since $\lambda_j = Cov(Y_t Y_{t-j}) = Cov(Y_{t+j} Y_t) = \lambda_{-j}$.

Since the "Autocovariance Generating Function" is defined as $\lambda(z) = \sum_{j=-\infty}^{\infty} \lambda_j z^j$, the spectrum can be written as $S(\omega) = (2\pi)^{-1} \lambda(e^{-i\omega})$.

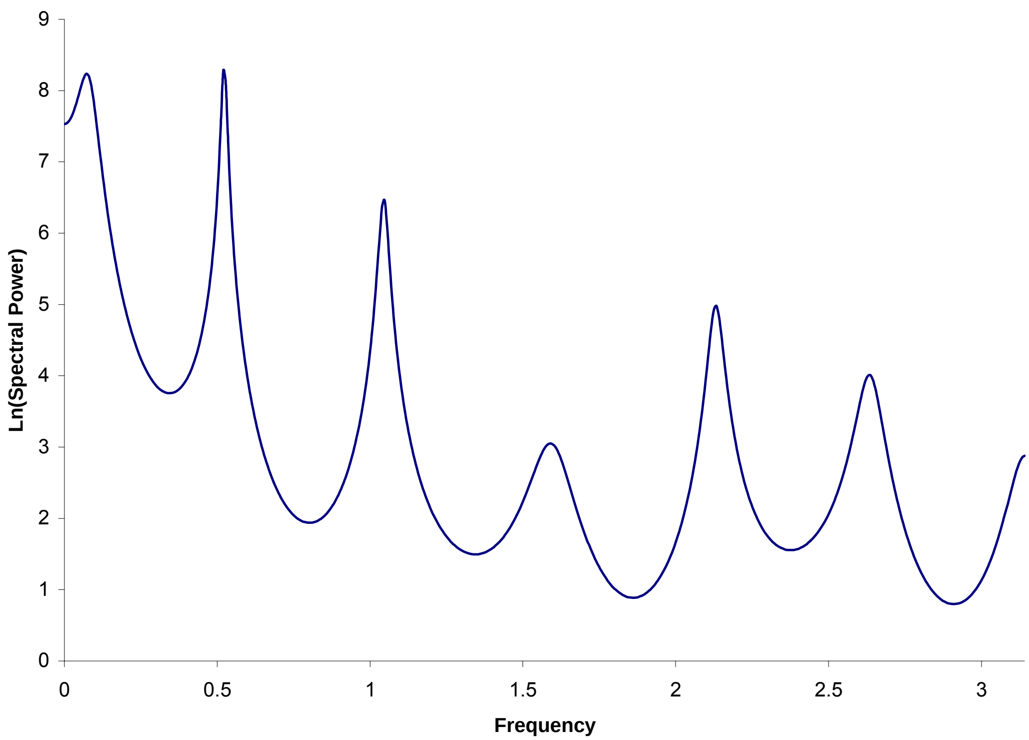
Summarizing the results presented above, the spectrum (or spectral density function) has 4 important properties.

6. $S(\omega)d\omega$ can be interpreted as the variance of the cyclical component of Y corresponding to the frequency ω. The period of this component is $\frac{2\pi}{\omega}$.
7. $S(\omega) \geq 0$. This follows, because $S(\omega)$ is a variance function.
8. $S(\omega) = S(-\omega)$. This follows from the definition of $\gamma(\omega)$ in the spectral representation (1), or from (5), since $\cos(a) = \cos(-a)$. Because of this symmetry, plots of the spectrum are presented $0 \leq \omega \leq \pi$.
9. $\lambda_0 = Var(Y) = \int_{-\pi}^{\pi} S(\omega)d\omega.$

Example: The Spectrum of Building Permits

Most of the mass in the spectrum is concentrated around the seven peaks evident in the plot. (These peaks are sufficiently large that spectrum is plotted on a log scale.) The first peak occurs at frequency $\omega = .07$ corresponding to a period of 90 months. This represents the business cycle variability in the series. The other peaks occur at frequencies $2\pi/12$, $4\pi/12$, $6\pi/12$, $8\pi/12$, and π. These are peaks for the seasonal frequencies: the first corresponds to a period of 12 months, and the others are the seasonal "harmonics" 6, 4, 3, and 2 months. (These harmonics are necessary to reproduce an arbitrary − not necessary sinesoidal − seasonal pattern.)

Figure 2
Spectrum of Building Permits



Linear Filters

Let

$$x_t = c(L)y_t$$

where

$$c(L) = c_{-r}L^{-r} + ... + c_{-1}L^{-1} + c_0L^0 + c_1L + ... + c_sL^s$$

is a time-invariant linear filter. It is useful to study how $c(L)$ changes the cyclical properties of y_t. To study this, suppose that y_t is strictly periodic

$$y_t = 2\cos(\omega t)$$

with period $p = \frac{2\pi}{\omega}$. Note that we can write

$$y_t = 2\cos(\omega t) = e^{i\omega t} + e^{-i\omega t}$$

Now

$$\begin{aligned} x_t &= \sum_{j=-r}^{s} c_j y_{t-j} \\ &= \sum_{j=-r}^{s} c_j [e^{i\omega(t-j)} + e^{-i\omega(t-j)}] \\ &= e^{i\omega t}\sum_{j=-r}^{s} c_j e^{-i\omega j} + e^{-i\omega t}\sum_{j=-r}^{s} c_j e^{i\omega j} \\ &= e^{i\omega t}c(e^{-i\omega}) + e^{-i\omega t}c(e^{i\omega}) \end{aligned}$$

Write the complex number $c(e^{i\omega})$ in polar form, as

$$c(e^{i\omega}) = a + ib$$

where

$$a = Re[c(e^{i\omega})] \text{ and } b = Im[c(e^{i\omega})]$$

or

$$c\left(e^{i\omega}\right) = (a^2 + b^2)^{\frac{1}{2}}[\cos(\theta) + i\sin(\theta)] = ge^{i\theta}$$

where

$$g = (a^2 + b^2)^{\frac{1}{2}} = [c(e^{i\omega})c(e^{-i\omega})]^{\frac{1}{2}} \text{ and}$$

$$\theta = \tan^{-1}[\frac{b}{a}] = \tan^{-1}[\frac{Im[c(e^{i\omega})]}{Re[c(e^{i\omega})]}]$$

Thus

$$
\begin{aligned}
x_t &= e^{i\omega t} g e^{-i\theta} + e^{-i\omega t} g e^{i\theta} \\
&= g[e^{i\omega[t-\frac{\theta}{\omega}]} + e^{-i\omega[t-\frac{\theta}{\omega}]}] \\
&= 2g\cos(\omega(t-\frac{\theta}{\omega}))
\end{aligned}
$$

So that the filter $c(L)$ "amplifies" y_t by the factor g and shifts y_t back in time by $\frac{\theta}{\omega}$ time units.

- Note that g and θ depend on ω, and so it makes since to write them as $g(\omega)$ and $\theta(\omega)$.
- $g(\omega)$ is called the filter *gain* (or sometimes the *amplitude gain*).
- $\theta(\omega)$ is called the filter *phase*.
- $g(\omega)^2 = [c(e^{i\omega})c(e^{-i\omega})]$ is called the *power transfer function* of the filter.

Examples

1. $c(L) = L^2$

$c(e^{i\omega}) = e^{2i\omega} = \cos(2\omega) + i\sin(2\omega)$, so that

$$\theta(\omega) = \tan^{-1}[\frac{\sin 2\omega}{\cos 2\omega}] = 2\omega$$

so that

$$\frac{\theta(\omega)}{\omega} = 2 \text{ time periods}$$

Also

$$g(\omega) = |c(e^{i\omega})| = 1$$

2. Kuznets Filter for annual data: Let

$$a(L) = (1/5)(L^{-2} + L^{-1} + L^0 + L^1 + L^2)$$

(which "smooths" the series) and

$$b(L) = (L^{-5} - L^5)$$

(which forms centered ten-year differences), then the Kuznets filter is

$$c(L) = b(L)a(L)$$

A useful computational note:

The gain of *c*(L) is the product of the gain of *a*(L) and *b*(L)
The phase of *c*(L) is the sum of the phase of *a*(L) and *b*(L).

Spectra of Commonly Used Stochastic Processes

Suppose that Y has spectrum $S_Y(\omega)$ and $X_t = c(L)Y_t$. What is the spectrum of X? As was shown above, the frequency components of X are the frequency components of Y scaled by the factor $g(\omega)e^{i\theta(\omega)}$, where $g(\omega)$ is the gain and $\theta(\omega)$ is the phase of $c(L)$. This means that spectra of X and Y are related by

$$S_X(\omega) = g(\omega)^2 S_Y(\omega) = | c(e^{i\omega}) |^2 S_Y(\omega)$$

which follows from $| e^{i\theta(\omega)} | = 1$ and the definition of $g(\omega)$.

Now, suppose that ε_t is a "white noise" process, defined by the properties $E(\varepsilon_t) = 0$, $\lambda_0 = \sigma^2$, and $\lambda_k = 0$ for $k \neq 0$. The spectrum of ε is then easily calculated from (5):

$$S_\varepsilon(\omega) = (2\pi)^{-1}\sigma^2$$

So, the spectrum of white noise is constant. (Which is why the process is called "white" noise.)

Now suppose that $Y_t = c(L)\varepsilon_t$. Then

$$S_Y(\omega) = |c(e^{i\omega})|^2 \, S_\varepsilon(\omega) = |c(e^{i\omega})|^2 \, (2\pi)^{-1}\sigma^2$$

This result can be used to determine the spectrum of any stationary ARMA process. If Y_t follows an ARMA process, then it can be represented as

$$\phi(L)Y_t = \theta(L)\varepsilon_t$$

The autoregressive operator, $\phi(L)$ can be inverted to yield $Y_t = c(L)\varepsilon_t$ with $c(L) = \phi(L)^{-1}\theta(L)$. This means that

$$S_Y(\omega) = |c(e^{i\omega})|^2 \, (2\pi)^{-1}\sigma^2 = \frac{2(\pi)^{-1}\sigma^2 \, |\theta(e^{i\omega})|^2}{|\phi(e^{i\omega})|^2} \tag{6}$$

As an example, consider the AR(1) model

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

equivalently written as

$$(1 - \phi L) Y_t = \varepsilon_t .$$

Applying (6) yields

$$S_Y(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \phi e^{-i\omega}|^2} = \frac{\sigma^2}{2\pi} \frac{1}{(1 + \phi^2 - 2\phi \cos(\omega))}$$

This spectrum is equal to $\sigma^2 [2\pi(1 + \phi^2 - 2\phi)]^{-1}$ at $\omega = 0$. When $0 < \phi < 1$, it falls steadily as ω increases from 0 to π. This means that, relative to white noise, the low frequency components of the AR(1) are more important than the high frequency components. Thus, realizations of the series appear smoother than white noise.

Band-Pass Filters

Reference: M.B. Baxter and King, R.G. (1999), "Measuring Business Cycles:Approximate Band-Pass Filters for Economic Time Series," *Review-of-Economics-and-Statistics* 81(4): 575-93.

Consider a description of y as

$$y_t = \tau_t + c_t$$

which decomposes y into a trend (τ) plus a cycle (c). There are three general approaches to estimating the components:

10. Band-Pass Filters: These differentiate τ and c by their cyclical characteristics and form estimates that attenuate certain frequencies to mask one of the components
11. Smoothing Filters such as the HP filter: These define the trend by certain smoothness characteristics and use this to construct estimates
12. Signal Extraction Filters: These define τ and c by specific stochastic processes and then apply signal extraction methods.

We will Signal Extraction filters later in the week. Here we focus on bandpass filters.

Let

$$c(L)=\sum_{j=-\infty}^{\infty} c_j L^j$$

with

$$c(e^{-i\omega})=\sum_{j=-\infty}^{\infty} c_j e^{-i\omega j}$$

and gain $|c(e^{i\omega})|$. Suppose that we set the phase of $c(L)$ to be equal to 0 (and thus $c(L)$ is symmetric: $c_j=c_{-j}$) and we want

$$gain(c(L))=|c(e^{i\omega})|=c(e^{i\omega})=\begin{Bmatrix} 1 \text{ for } -\underline{\omega}\le\omega\le\underline{\omega} \\ 0 \text{ elsewhere} \end{Bmatrix}$$

where the second equality follows because $c(e^{i\omega})$ is real, because $c(L)$ is symmetric. Since

$$c(e^{-i\omega})=\sum_{j=-\infty}^{\infty} c_j e^{-i\omega j}$$

then

$$c_j=(2\pi)^{-1}\int_{-\pi}^{\pi} e^{i\omega j}c(e^{-i\omega})d\omega$$

follows generally from $\int_{-\pi}^{\pi} e^{i\omega k}d\omega=\begin{Bmatrix} 2\pi \text{ for } k=0 \\ 0 \text{ for } k\neq 0\end{Bmatrix}$. Setting the gain equal to unity over the desired frequencies and carrying out the integration yields

$$c_j=(2\pi)^{-1}\frac{1}{ij}e^{i\omega j}\,|_{-\underline{\omega}}^{\underline{\omega}}=\begin{Bmatrix} \frac{1}{j\pi}\sin(\underline{\omega}j) \text{ for } j\neq 0 \\ \frac{\underline{\omega}}{\underline{\pi}} \text{ for } j=0 \end{Bmatrix}$$

Comments:

13. The values of c_j die out at the rate $\frac{1}{j}$
14. $1-c(L)$ passes everything except $-\underline{\omega}\le\omega\le\underline{\omega}$
15. Baxter and King show that $c_k(L)=\sum_{j=-k}^{k} c_j L^j$ is an optimal finite order approximation to $c(L)$ in the sense that the gain of $c_k(L)$ is as close as possible to the gain of $c(L)$ for a k-order filter.

One-Sided Filters

One practical problem with these filters is that they are two-sided, with weights that can die out slowly. For example, in the exact band-pass filters the coefficients c_j die out at the rate $1/j$. This introduces important "endpoint" problems when applying these filters to finite realizations of data, say $\{Y_t\}_{t=1}^T$. In this case the minimum mean square error estimate of the infeasible two-sided estimates are easily constructed. For example, if

$$X_t = c(L)Y_t = \sum_{i=-\infty}^{\infty} c_i Y_{t-i},$$

the optimal estimate of X_t given $\{Y_t\}_{t=1}^T$ is

$$E(X_t \mid \{Y_j\}_{j=1}^T) = \sum_{i=-\infty}^{\infty} c_i E(Y_{t-i} \mid \{Y_j\}_{j=1}^T)$$

This can be accomplished by using backcasts and forecasts for the missing pre-sample and post-samples values of Y_t.

The variance of the error associated with using $\{Y_t\}_{t=1}^T$ is

$$var[X_t - E(X_t \mid \{Y_j\}_{j=1}^T] = var[\sum_{i=-\infty}^{\infty} c_i \{E(Y_{t-i} \mid \{Y_j\}_{j=1}^T) - Y_{t-i}\}].$$

The Functional Central Limit Theorem

The functional central limit theorem (FCLT) is an extension of the central limit theorem. It will help to solve some inference problems discussed below. We will not formally work through all of the details of this theorem, but the idea is quite simple. We begin with a particular continuous-time stochastic process, a Wiener Process (sometimes called Standard Brownian Motion).

Denote the process by $W(s)$, defined on $s \in [0,1]$, with the following properties

1. $W(0) = 0$

2. For any dates $0 \leq t_1 < t_2 < ... < t_k \leq 1$, $W(t_2) - W(t_1)$, $W(t_3) - W(t_4), ..., W(t_k) - W(t_{k-1})$ are independent normally distributed random variables with $W(\tau_2) - W(\tau_1) \sim N(0, \tau_2 - \tau_1)$.

3. Any realization of $W(s)$ is continuous $w.p.\ 1$.

From (2) $W(1) \sim N(0,1)$

Approximating a realization of $W(s)$

We can do this as follows. Take the unit interval and divide it into 10 equally spaced points $(1/10, 2/10, ..., 9/10, 10/10)$. Then draw 10 $Niid(0,1)$ random variables, e_1, e_2, ..., e_{10}. We can then approximate $W(s)$ at the grid points by

$$W(\frac{j}{10}) = \frac{1}{\sqrt{10}} \sum_{i=1}^{j} e_i,\ j = 1, ..., 10$$

note, with this construction

$$W(\tfrac{j}{10}) \sim N(0, j/10)$$

$$W(\tfrac{j+1}{10}) - W(\tfrac{j}{10}) \sim N(0, 1/10)$$

So that on these grid points we have created a set of random variables with the same distribution as $W(s)$.

We could create a better approximation by using a even finer grid. Thus, using a grid with T grid points we could construct

$$W(\frac{j}{T}) = \frac{1}{\sqrt{T}} \sum_{i=1}^{j} e_i,\ j = 1, \ldots, T$$

and letting $[\cdot]$ denote the "greatest integer $\leq$" function (so that $[1.2] = 1$, $[5.98] = 5$, etc.) then for any $s \in [0,1]$ we have

$$W(\frac{[sT]}{T}) \rightarrow W(s).$$

So that, in some sense, we have created an Wiener process.

The assumption that the underlying errors, e_i, are normal is not important, because the central limit theorem can applied to the standardized sums.

This kind of construction leads to the functional central limit theorem. Crudely it says: Suppose that e_t is a mds with finite fourth moment. Then the process $\overline{W}_T(s) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} e_t$ converges to the Wiener process on $s \in [0,1]$. (A careful statement of the theorem is given in Stock (Handbook of Econometrics, 1994).)

We will also use the Continuous Mapping Theorem:

Let $g(.)$ be a continuous function, then

$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

$$X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$$

$$X_n - Y_n \xrightarrow{p} 0 \text{ and } Y_n \xrightarrow{d} Y \text{ then } g(X_n) - g(Y_n) \xrightarrow{p} 0$$

Example:

Let

$$y_t = \sum_{i=1}^{t} \varepsilon_i$$

where ε_i is $iidN(0,1)$, and let $\overline{W}_T(s) = \frac{1}{\sqrt{T}} y_{[sT]}$ be a step function approximation of $W(s)$. Then

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} y_t = \frac{1}{T} \sum_{t=1}^{T} [\frac{1}{\sqrt{T}} y_t] = \int_0^1 \overline{W}_T(s) ds \Rightarrow \int_0^1 W(s) ds$$

Instability in Regression Models

Consider the model

$$y_t = x_t'\beta_t + \varepsilon_t$$

where β_t is not necessarily constant. There are two questions of interest: (1) how can you test the null hypothesis that $\beta_t = \beta_0$ for all t, and (2) assuming (1) is false, how can you estimate the time path of β_t. We will study each of these problems. For notational simplicity much of the analysis will be done in the simple model

$$y_t = \beta_t + \varepsilon_t$$

so that $x_t = 1$.

Models for coefficient instability

Tests and estimators in the TVP model depend on the model for instability. Two popular models are:

1. Discrete Break Model:

$$\beta_t = \beta \text{ for } t \le \tau$$
$$\beta_t = \beta + \gamma \text{ for } > \tau.$$

Under the null hypothesis $H_0 : \gamma = 0$ the regression coefficients are constant; while under the alternative $H_a : \gamma \neq 0$, there is a shift in the coefficients at time period τ.

Extensions to models with multiple discrete breaks are conceptually straightforward (although the calculations may not be).

2. Martingale TVP Variation:

$$\beta_t = \beta_{t-1} + v_t$$

where v_t is a white noise (or, more generally, an $I(0)$) process.

In this section I will discuss the discrete break model. The next section will discuss martingale variation.

Tests for discrete breaks

The null and alternative can be written as:

$$H_o : \gamma = 0$$

and

$$H_a : \gamma \neq 0$$

Tests for H_o vs. H_a depend on whether τ known or τ unknown. For simplicity we will assume that ε_t is $iid(0,1)$.

Chow Tests

If the break date τ was known then the MLE of γ is

$$\hat{\gamma} = \bar{Y}_2 - \bar{Y}_1$$

where

$$\bar{Y}_1 = \frac{1}{\tau}\sum_{t=1}^{\tau} y_t \text{ and } \bar{Y}_2 = \frac{1}{T-\tau}\sum_{t=\tau+1}^{T} y_t$$

and the Wald statistic is

$$\xi_W = \frac{\hat{\gamma}^2}{(\frac{1}{\tau} + \frac{1}{T-\tau})}$$

since

$$\bar{Y}_1 \sim N(\beta_1, \frac{1}{\tau})$$

and

$$\bar{Y}_2 \sim N(\beta_2, \frac{1}{T-\tau})$$

Under the null ξ_W is distributed as a χ_1^2 random variable. (This test is often called a "Chow test" after Gregory Chow.) With non-normal ε's, this result holds asymptotically.

Quandt Tests (Sup Wald or QLR)

If the break date is unknown, the analysis is more difficult. One idea, suggested by Quandt (*JASA* 1960), is to compute the statistic for a large number of possible values of τ and use the largest of these as the test statistics. Of course, the largest of sequence of these statistics will no longer have a χ_1^2 distribution and the problem is to find the distribution. One way to do this is as follows.

Define the statistic of interest as

$$\xi_Q = \max_{\tau_1 \le \tau \le \tau_2} \xi_W(\tau)$$

where the Chow statistic ξ_W is now indexed by the break date and the test is computed for all possible break dates between τ_1 and τ_2.

Under the null hypothesis, $\overline{Y}_1 = \beta + \frac{1}{\tau}\sum_{t=1}^{[sT]} \varepsilon_t$ and $\overline{Y}_2 = \beta + \gamma + \frac{1}{T-\tau}\sum_{t=\tau+1}^{T} \varepsilon_t$. Let s = τ/T. Under the null $\gamma = 0$ and we can then write $\xi_W(\tau)$ as

$$\begin{aligned}
\xi_W(\tau) \equiv G_T(s) &\overset{H_o}{=} \frac{[\frac{1}{[sT]}\sum_{t=1}^{[sT]} \varepsilon_t - \frac{1}{[(1-s)T]}\sum_{t=[sT]+1}^{T} \varepsilon_t]^2}{\frac{1}{[sT]} + \frac{1}{[(1-s)T]}} \\
&= \frac{[\frac{1}{s}\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \varepsilon_t - \frac{1}{(1-s)}\frac{1}{\sqrt{T}}\sum_{t=[sT]+1}^{T} \varepsilon_t]^2}{\frac{1}{s} + \frac{1}{(1-s)}} \\
&\Rightarrow \frac{[\frac{1}{s}W(s) - \frac{1}{(1-s)}(W(1) - W(s)]^2}{\frac{1}{s} + \frac{1}{(1-s)}} \\
&= \frac{[W(s) - sW(1)]^2}{s(1-s)}
\end{aligned}$$

where the last equality follows from multiplying the numerator and denominator by $s^2(1-s)^2$ and simplifying.

Suppose that τ_1 is chosen as δT and τ_2 is chosen as $(1-\delta)T$, where $0<\delta<0.5$. Then

$$\max_{\tau_1 \le \tau \le \tau_2} \xi_W(\tau) = \sup_{\delta \le s \le (1-\delta)} G_T(s).$$

Since the "sup" function is continuous

$$\xi_Q \Rightarrow \sup_{\delta \le s \le (1-\delta)} \frac{[W(s)-sW(1)]^2}{s(1-s)}$$

It has become standard practice to use a value of $\delta = .15$. Using the value of δ, the 1%, 5% and 10% critical values for the test are: 12.16, 8.68 and 7.12. (These can be compared to the corresponding critical values of the χ_1^2 of 6.63, 3.84 and 2.71).

The results have been derived here for the case of a single constant regressor. Exensions to the case of multiple (non-constant) regressors can be found in Andrews (1993) (Critical values for the test statistic are also given in Andrews (1991) with corrections in Andrews (2002), reprinted in Stock and Watson (2003).)

3. Optimal Tests when τ is unknown (covered as time allows)

It is useful to step back from this problem and think for a moment about "optimal" tests for this null and this alternative. This problem differs from standard problems because one of the important unknowns τ is unidentified under the null hypothesis. This means that the standard results and intuition about optimality of LR tests may not apply. Here we present an optimal test following the approach used in Andrews and Ploberger.

It helps to get a running start on the problem, and will do so by assuming that $\beta = 0$. (Later we will discuss the modifications required for uknown values of β and σ_ε^2.) Note with $\beta = 0$ and τ known,

$$\hat{\gamma} = \bar{Y}_2 \text{ and } \xi_W(\tau) = \frac{\hat{\gamma}^2}{[1/(T-\tau)]} = (1-\pi)\hat{\alpha}^2$$

with

$$\pi = \frac{\tau}{T}, \text{ and } \alpha = \sqrt{T}\gamma \text{ , and } \hat{\alpha} = \sqrt{T}\hat{\gamma}$$

Now, suppose that under the alternative the parameters α and π are random variables. Remember that $\hat{\gamma} \overset{a}{\sim} N(0, \frac{1}{T}\frac{1}{(1-\pi)})$ or $\hat{\alpha} = \sqrt{T}\hat{\gamma} \xrightarrow{d} N(0, \frac{1}{1-\pi})$, and is convenient to choose a "conjugate" distribution for α:

$$\alpha \,|\, \pi \sim N(0, \frac{c}{(1-\pi)}) = h_\pi(\alpha)$$

where c is a parameter that we will assume is known. The distribution for π is denoted G (more will said about G below).

We then have, under the null, the likelihood/density is

$$L_o = \kappa \exp[-\frac{1}{2}\sum_{t=1}^{T} y_t^2]$$

where κ is a constant. Under the alternative, the likelihood/density is

$$L_a = \kappa \int\int \exp[-\frac{1}{2}\sum_{t=1}^{\tau} y_t^2 - \frac{1}{2}\sum_{t=\tau+1}^{T} (y_t - \frac{\alpha}{\sqrt{T}})^2] h_\pi(\alpha) d\alpha dG$$

In this setting both the null and alternatives are simple hypotheses that do not depend on any unknown parameters. From the Neyman-Pearson lemma, the most powerful test rejects the null for large values of the $LR = L_a/L_o$ An alternative interpretation is that the LR test has best weighted power against alternatives $H_a : \gamma = \frac{\alpha}{\sqrt{T}},\ \tau = \pi T$ using the weight function $h_\pi(\alpha)dG$.

Now, some algebra to make the statistic more interpretable:

$$LR = L_a/L_o = \int\int A(\alpha,\pi)h_\pi(\alpha)d\alpha dG$$

where

$$A(\alpha,\pi) = \exp[-\frac{1}{2}\sum_{t=1}^{\tau} y_t^2 - \frac{1}{2}\sum_{t=\tau+1}^{T}(y_t - \frac{\alpha}{\sqrt{T}})^2 + \frac{1}{2}\sum_{t=1}^{T} y_t^2]$$
$$= \exp[-\frac{1}{2}(-\pi)\alpha^2 + (1-\pi)\alpha\hat{\alpha}]$$

where the second inequality follows from some algebra and $\hat{\alpha} = \sqrt{T}(T-\tau)^{-1}\sum_{t=\tau+1}^{T} y_t$

Thus,

$$A(\alpha,\pi)h_\pi(\alpha) = \kappa\sqrt{\frac{1-\pi}{c}}\exp[-\frac{1}{2}(1-\pi)\alpha^2 + (1-\pi)\alpha\hat{\alpha} - \frac{1}{2}\frac{1}{c}(1-\pi)\alpha^2]$$

$$= \kappa\sqrt{\tfrac{1}{1+c}}\sqrt{\tfrac{(1-\pi)(1+c)}{c}}\exp\{-\tfrac{1}{2}(1-\pi)\tfrac{1+c}{c}[\alpha - \tfrac{c}{1+c}\hat{\alpha}]^2 + \tfrac{1}{2}(1-\pi)\tfrac{c}{1+c}\hat{\alpha}^2\}$$

and

$$\int A(\alpha,\pi)h_\pi(\alpha)d\alpha =$$

$$\sqrt{\tfrac{1}{1+c}}\exp[\tfrac{1}{2}(1-\pi)\tfrac{c}{1+c}\hat{\alpha}^2]\int\sqrt{\tfrac{(1-\pi)(1+c)}{c}}\exp\{-\tfrac{1}{2}(1-\pi)\tfrac{1+c}{c}[\alpha - \tfrac{c}{1+c}\hat{\alpha}]^2\}d\alpha =$$
$$\sqrt{\frac{1}{1+c}}\exp[\frac{1}{2}(1-\pi)\frac{c}{1+c}\hat{\alpha}^2]$$

where the final equality follows because the integral term is equal to 1.

This means that the LR statistic can be written as

$$LR = \frac{1}{\sqrt{1+c}} \int \exp[\frac{1}{2}\frac{c}{1+c}\xi_W(\pi)]dG(\pi)$$

Some cases of interest

16. $G(\pi)$ puts point mass at $\pi = \overline{\pi}$. The test rejects for large values of $\xi_W(\overline{\pi})$ (same a Chow test).
17. $G(\pi)$ is uniform.
    (a) c small $\exp[-\frac{1}{2}\frac{c}{1+c}\xi_W(\pi)] \approx 1 + \frac{1}{2}\frac{c}{1+c}\xi_W(\pi)$. The test rejects for large values of $\int \xi_W(\pi)d\pi$, which is an average of the Wald Statistics.
    (b) c large The test rejects for large values of $\int \exp[\frac{1}{2}\xi_W(\pi)]d\pi$ which is sometimes called the exponential Wald statistic (or the Andrews-Ploberger statistics). In practice, this behaves much like the sup-Wald or QLR test.

Complications arising from an unknown value of β is handled by considering tests that are invariant to the value of β. Here, optimal tests are constructed using demeaned values of y_t in place of y in the expressions above. These yield statistics of the same form, but with ξ_W computed as in the beginning of this section. Complications arising from an unknown value of σ_ε^2 are handled by dividing ξ_W by an estimate of σ_ε^2.

Estimating Break Dates

Continuing with the simplifying example. Let $\overline{Y}_\tau = \frac{1}{\tau}\sum_{t=1}^{\tau} Y_t$ and $\overline{Y}_\tau^* = \frac{1}{T-\tau}\sum_{t=\tau+1}^{T} Y_t$

As above, let $\pi = \frac{\tau}{T}$ and let $\pi_0 = \frac{\tau_0}{T}$ where the subscript "0" denotes the true value. Note that the Chow/Wald statistic for a break at τ can be written as

$$\xi(\pi) = \pi(1-\pi)(\overline{Y}_\tau - \overline{Y}_\tau^*)^2$$

The least squares/MLE estimator of τ is the minimizer of

$$SSR(\tau) = \sum_{t=1}^{\tau}(Y_t - \overline{Y}_\tau)^2 + \sum_{t=\tau+1}^{T}(Y_t - \overline{Y}_\tau^*)^2$$

Let $TSS = \sum_{t=1}^{T}(Y_t - \overline{Y}_T)^2$. Then, straightforward algebra yields:

$$TSS - SSR(\tau) = \xi(\pi)$$

so that minimizing $SSR(\tau)$ is the same as maximizing the test statistic $\xi(\pi)$. Thus, the value of τ (equivalently the value of π) that maximizes the Chow statistic is the least squares estimator of the break date.

Some results that are useful for inference:

- (Bai shows) $\hat{\pi} - \pi \sim O_p(T^{-1}\gamma^{-2})$, so that
    - $T\gamma^2(\hat{\pi} - \pi_0) \sim O_p(1)$
    - $\gamma^2(\hat{\tau} - \tau_0) \sim O_p(1)$

Thus, $\hat{\pi}$ is consistent for π_0, but $\hat{\tau}$ is not consistent for τ_0.

- In general, the distribution of $\hat{\pi}$ and $\hat{\tau}$ depends on the distribution of the errrors ε_t. This is true even and T grows large. Thus, robust inference is problematic.
- There are approximations that can be used with γ is appropriately small.

Suppose $\gamma \to 0$ but $T\gamma^2 \to \infty$ (so $\gamma = T^{-1/4}$ for example). In this case, the random variable $T\gamma^2(\hat{\pi} - \pi_0)$ has a limiting distribution, and this distribution can be used to construct confidence intervals for π_0 and τ_0. I'll now sketch the derivation of this limiting distribution.

First, let's restrict attention to values of π that are "close" to π_0. We can do this because of the consistency result above. Thus, we will write

$$\pi = \pi_0 + \upsilon T^{-1}\gamma^{-2}, \text{ for } -m \le \upsilon \le m$$

where m is a large but finite number. Now, let's study the behavior of $\xi(\pi)$ in the vicinity of $\xi(\pi_0)$. For $\pi > \pi_0$, a little algebra shows

$$\xi(\pi) - \xi(\pi_0) = A1 + A2 + A3$$

where

$$A1 = T\gamma^2\{\pi_0(1-\pi)(\pi_0/\pi) - \pi_0(1-\pi_0)\}$$

$$A2 = 2\{\pi_0(1-\pi)\gamma T(\overline{\varepsilon}_\tau - \overline{\varepsilon}^*_\tau) - \pi_0(1-\pi_0)\gamma T(\overline{\varepsilon}_{\tau_0} - \overline{\varepsilon}^*_{\tau_0})\}$$

and

$$A3 = \pi(1-\pi)T(\overline{\varepsilon}_\tau - \overline{\varepsilon}^*_\tau)^2 - \pi_0(1-\pi_0)T(\overline{\varepsilon}_{\tau_0} - \overline{\varepsilon}^*_{\tau_0})^2$$

First, let's study $A1$. With $\pi = \pi_0 + \upsilon T^{-1}\gamma^{-2}$, $A1 \to \upsilon$.

Second, let's study $A2$. First note that $\pi_0(1-\pi)\gamma T = \pi_0(1-\pi_0)\gamma T - \upsilon\gamma^{-1}$ and $\gamma^{-1}(\overline{\varepsilon}_\tau - \overline{\varepsilon}^*_\tau)p \to 0$. Thus

$$A2 = 2\pi_0(1-\pi_0)\gamma T[(\overline{\varepsilon}_\tau - \overline{\varepsilon}^*_\tau) - (\overline{\varepsilon}_{\tau_0} - \overline{\varepsilon}^*_{\tau_0})] + o_p(1)$$

Now.

$$T(\overline{\varepsilon}_\tau - \overline{\varepsilon}_{\tau_0}) = (\pi_0/\pi)\frac{1}{\pi_0}\sum_{t=1}^{\tau}\varepsilon_t - \frac{1}{\pi_0}\sum_{t=1}^{\tau_0}\varepsilon_t = \frac{1}{\pi_0}\sum_{t=\tau_0+1}^{\tau}\varepsilon_t + o_p(1)$$

and

$$\frac{1}{\pi_0}\sum_{t=\tau_0+1}^{\tau}\varepsilon_t = \frac{1}{\pi_0}\sum_{t=\tau_0+1}^{\tau_0+\upsilon\gamma^{-2}}\varepsilon_t$$

Similarly

$$T(\overline{\varepsilon}^*_\tau - \overline{\varepsilon}^*_{\tau_0}) = -\frac{1}{1-\pi_0}\sum_{t=\tau_0+1}^{\tau_0+\upsilon\gamma^{-2}}\varepsilon_t$$

So that

$$T(\overline{\varepsilon}_{\tau}-\overline{\varepsilon}_{\tau_0})-T(\overline{\varepsilon}_{\tau}^{*}-\overline{\varepsilon}_{\tau_0}^{*})=\frac{1}{\pi_0(1-\pi_0)}\sum_{t=\tau_0+1}^{\tau_0+\upsilon\gamma^{-2}}\varepsilon_t+o_p(1)$$

Recognize that with $\gamma\to 0$, $\gamma\sum_{t=\tau_0+1}^{\tau_0+\upsilon\gamma^{-2}}\varepsilon_t\Rightarrow W(\upsilon)$. Thus,

$$A2\Rightarrow 2W(\upsilon)$$

Finally, a straightforward calcution shows that $A3p\to 0$. Putting these results together, for $\pi=\pi_0+\upsilon T^{-1}\gamma^{-2}$ with $\upsilon>0$.

$$\xi(\pi)-\xi(\pi_0)\Rightarrow 2W(\upsilon)-\upsilon$$

Searching over values of $\pi>\pi_0$ we find $T\gamma^2(\hat{\pi}-\pi_0)=\upsilon$ is the value of υ that maximizes $2W(\upsilon)-\upsilon$. A similar result holds for value of $\pi<\pi_0$.

The value of υ that maximize the criterion have a very non-gaussian shape. Here are some values

$$\Pr(|\hat{\upsilon}|<c)$$

Probability	c	Standard Normal
50%	2.8	
67%	4.4	1.0
80%	6.7	
90%	10.0	1.64
95%	13.8	1.96
99%	23.5	2.56

1. A Summary Calculation

Suppose that $T = 300$, $\hat{\gamma} = 0.33$, $\hat{\sigma} = 0.6$, and $\hat{\pi} = 0.51$ (so that $\hat{\tau} = 153$). Then a 67% confidence interval for π_0 satistifies

$$|\frac{1}{\hat{\sigma}} T \hat{\gamma}^2 (\hat{\pi} - \pi_0)| < 4.4$$

or

$$\hat{\pi} - 4.4 \frac{\hat{\sigma}}{T\hat{\gamma}^2} < \pi_0 < \hat{\pi} - 4.4 \frac{\hat{\sigma}}{T\hat{\gamma}^2}$$

or

$$0.43 < \pi_0 < 0.59$$

and

$$0.43 \times 300 < \tau_0 < 0.59 \times 300$$

A 95% confidence interval satisfies

$$|\frac{1}{\hat{\sigma}} T \hat{\gamma}^2 (\hat{\pi} - \pi_0)| < 13.8$$

or

$$0.26 < \pi_0 < 0.76$$

and

$$0.26 \times 300 < \tau_0 < 0.76 \times 300$$

Empirical Example … Stock-Watson NBER MacroAnnual 2002

	Conditional Mean			Conditional Variance: break only			Conditional Variance: trend and break		
Series	***p*-value**	**break date**	**67% confidence interval**	***p*-value**	**break date**	**67% confidence interval**	***p*-value: trend**	***p*-value: break**	**break date**
GDP	0.98	.	. - .	0.00	1983:2	1982:4 - 1985:3	0.63	0.00	1983:2
consumption	0.55	.	. - .	0.00	1992:1	1991:3 - 1994:1	0.00	0.11	.
consumption – durables	0.04	1987:3	1987:1 - 1988:1	0.00	1987:3	1987:2 - 1990:2	0.68	0.03	1987:3
consumption – nondurables	0.00	1991:4	1991:2 - 1992:2	0.08	.	. - .	0.96	0.80	.
consumption – services	0.00	1969:4	1969:2 - 1970:2	0.18	.	. - .	0.03	0.00	1971:3
investment (total)	0.05	.	. - .	0.13	.	. - .	0.06	0.25	.
fixed investment – total	0.69	.	. - .	0.01	1983:3	1983:1 - 1986:4	0.65	0.07	.
nonresidential	0.47	.	. - .	0.70	.	. - .	0.69	0.60	.
residential	0.10	.	. - .	0.00	1983:2	1983:1 - 1985:2	0.08	0.00	1983:2
Δinventory investment/GDP	0.91	.	. - .	0.04	1988:1	1987:3 - 1992:2	0.00	0.10	.
exports	0.09	.	. - .	0.00	1975:4	1975:2 - 1978:2	0.95	0.75	.
imports	0.00	1972:4	1972:2 - 1973:2	0.00	1986:2	1986:1 - 1988:1	0.96	0.05	1986:2
government spending	0.06	.	. - .	0.45	.	. - .	0.33	0.65	.
Production									
goods (total)	0.92	.	. - .	0.00	1983:4	1983:2 - 1986:4	0.54	0.03	1983:3
nondurable goods	0.09	.	. - .	0.00	1983:4	1983:3 - 1987:1	0.00	0.29	.
durable goods	0.77	.	. - .	0.02	1985:2	1984:3 - 1989:1	0.33	0.02	1985:2
services	0.00	1968:3	1968:1 - 1969:1	0.98	.	. - .	0.69	0.92	.
structures	0.02	1991:3	1991:1 - 1992:1	0.02	1984:2	1983:4 - 1988:1	0.42	0.03	1984:2
nonagricultural employment	0.03	1981:2	1980:4 - 1981:4	0.00	1983:2	1982:4 - 1985:3	0.00	0.02	1973:3
price inflation (GDP deflator)	0.00	1973:2	1972:4 - 1973:4	0.11	.	. - .	0.00	0.00	1971:2
90-day T-bill rate	0.00	1981:1	1980:3 - 1981:3	0.01	1984:4	1984:2 - 1988:1	0.00	0.00	1984:4
10-year T-bond rate	0.02	1981:1	1980:3 - 1981:3	0.00	1979:3	1972:2 - 1980:1	0.02	0.00	1979:3

Notes: The test results are based on the QLR test for changes in the coefficients of an AR(4). The first column shows the *p*-value for the QLR test break test statistic. The second column shows the least squares estimate of the break date (when the QLR statistic is significant at the 5% level), and the final column shows the 67% confidence interval for the break date. The results in the “Conditional Mean Coefficients” columns correspond to the parameters α and ϕ in Equation (1). The results in the “Conditional Variance” columns refer to the variance of ε_t in Equation (1), either with or without a time trend in the QLR regression.

Martingale TVP Variation

Tests for martingale time variation

To derive the tests, write the model for β_t as

$$\beta_t = \beta_{t-1} + \gamma e_t$$

where e_t is i.i.d. $N(0,\sigma_\varepsilon^2)$ and is independent of ε_j for all t and j. (Note, as a normalization, the variance of ε and the variance of e are assumed to be equal.) Again, to get things rolling, suppose that $\beta_0 = 0$. Let $Y = (y_1 \;\; y_2 \;\; \ldots \;\; y_T)'$. Then $Y \sim N(0,\sigma_\varepsilon^2\Omega(\gamma))$, where $\Omega = I + \gamma^2 A$, where $A = [a_{ij}]$, with $a_{ij} = \min(i, j)$. The optimal test of $H_o : \gamma = 0$ (no time variation) versus $H_a : \gamma = \gamma_a$ (some time variation) can be constructed using the likelihhod ratio statistic. The LR statistic is given by

$$LR = |\Omega(\gamma_a)| \exp\{-\frac{1}{2\sigma_\varepsilon^2}[Y'\Omega(\gamma_a)^{-1}Y - Y'Y]\}$$

so that the test rejects the null for large values of

$$\frac{Y'\Omega(\gamma_a)^{-1}Y}{Y'Y}$$

A well known version of this test uses the local approximation $\Omega(\gamma_a)^{-1} = [I + \gamma_a^2 A]^{-1} = 1 - \gamma_a^2 A$. In this case, the test rejects for large values of

$$\psi = \frac{Y'AY}{Y'Y}$$

which is a version of the locally best test of Nyblom. Note $A = PP'$, where P is a lower triangular matrix of $1's$, so the test statistic can be written as

$$\psi = \frac{Q'Q}{Y'Y}$$

where $Q = P'Y$, so that $q_t = \sum_{i=t}^{T} y_i$. The statistic can then be written as

$$\psi = \frac{\sum_{t=1}^{T}(\sum_{i=t}^{T} y_i)^2}{\sum_{t=1}^{T} y_t^2}.$$

To derive the distribution of the statistic under the null, write

$$T^{-1}\psi = \frac{\frac{1}{T}\sum_{t=1}^{T}(\frac{1}{\sqrt{T}}\sum_{i=t}^{T} y_i)^2}{\frac{1}{T}\sum_{t=1}^{T} y_t^2} \overset{H_o}{\Rightarrow} \int_0^1 (W(1)-W(s))^2\, ds$$

Complications associated with an unknown value of β_0 are handled just as in the discrete break case. Attention is focused on tests that are invariant to the value of β_0. M. King shows that these tests have the same form as the LR tests constructed above, but using demeaned values of y.

Optimal tests require a choice of γ_a, which measures the amount of time variation under the alternative. A common way to choose γ_a is to use a value so that, when $\gamma = \gamma_a$, the test has a pre-specified power, often 50%. Generally, this test (called a "point optimal" or "point optimal invariant" test) has near optimal power for a wide range of values of γ. A good rule of thumb (from Stock-Watson (JASA 1998) is to set $\gamma_a = 7/T$.)

The assumption of martingale time variation with Gaussian innovations turns out to be quite general as shown in Elliott and Mueller (2005).

Estimating the amount of time variation

When coefficients vary as martingale, the paramter γ indexes the amount of time variation and is often a parameter of interest. When γ is large, it MLE has the usual desirable properties, and should be used. However, in many cases of economic interest the amount of time variation is not large, in the sense that TVP tests like those discussed above yield p-values of say 5%, 10% or even larger. When γ is small, the MLE performs poorly and often produces an estimate of $\gamma = 0$, even when γ differs from zero.

For these cases an alternative estimator can be constructed by "inverting" one of the test statistics discussed above (Stock and Watson (1998)). Here's the idea.

Suppose we write $\gamma = g/T$ and work out the asymptotic behavior of the TVP test statistics as T gets large, but with g fixed. This asymptotic calculation will turn out to provide a good approximation to the distribution of test statistics for moderate sized T but with small γ. Write

$$\begin{aligned} y_t &= \beta_0 + \varepsilon_t + (\beta_t - \beta_0) \\ &= \beta_0 + \varepsilon_t + \frac{g}{T}\sum_{i=1}^{t} e_t \end{aligned}$$

For notational simplicity, suppose that $\beta_0 = 0$. In this case all of the statistics above are functionals of the partial sums $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} y_t$, and the distributions of the statistics under the null hyptohesis were developed using $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} y_t \Rightarrow W(s)$. Under the alternative with $\gamma = g/T$ we can also work out the limiting distribution of $\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} y_t$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} y_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \varepsilon_t + \frac{g}{T^{3/2}}\sum_{t=1}^{[sT]}\sum_{i=1}^{t} e_t$$

Now

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[sT]} \varepsilon_t \Rightarrow W(s)$$

and

$$\frac{g}{T^{3/2}}\sum_{t=1}^{[sT]}\sum_{i=1}^{t} e_t = g\frac{1}{T}\sum_{t=1}^{[sT]}\left[\frac{1}{\sqrt{T}}\sum_{i=1}^{t} e_t\right] \Rightarrow g\int_0^s B(r)dr$$

where $B(r)$ is a Weiner process that is independent of $W(s)$. Notice that the distribution of this last term will depend on g. This means that each of the TVP test statistics discussed above will have a probability distribution that depends on g.

For a give statistic, say X, let $A(g,x)$ denote its cdf, so that $\Pr(X \le x) = A(g,x)$, Let $\hat{g}$ solve $A(\hat{g}, X) = .5$, then $\Pr(\hat{g} \le .g) = .5$ and $\hat{g}$ is a "median unbiased estimator" of g. Also $(g \,|\, .95 \le A(g,X) \le .05)$ is a 90% confidence interval for g.

Empirical Example: US GDP Growth Rates

$400\times\Delta y_t = \beta_t + v_t$

$\beta_t = (g/T)e_t$

$v_t \sim \text{AR}(1)$

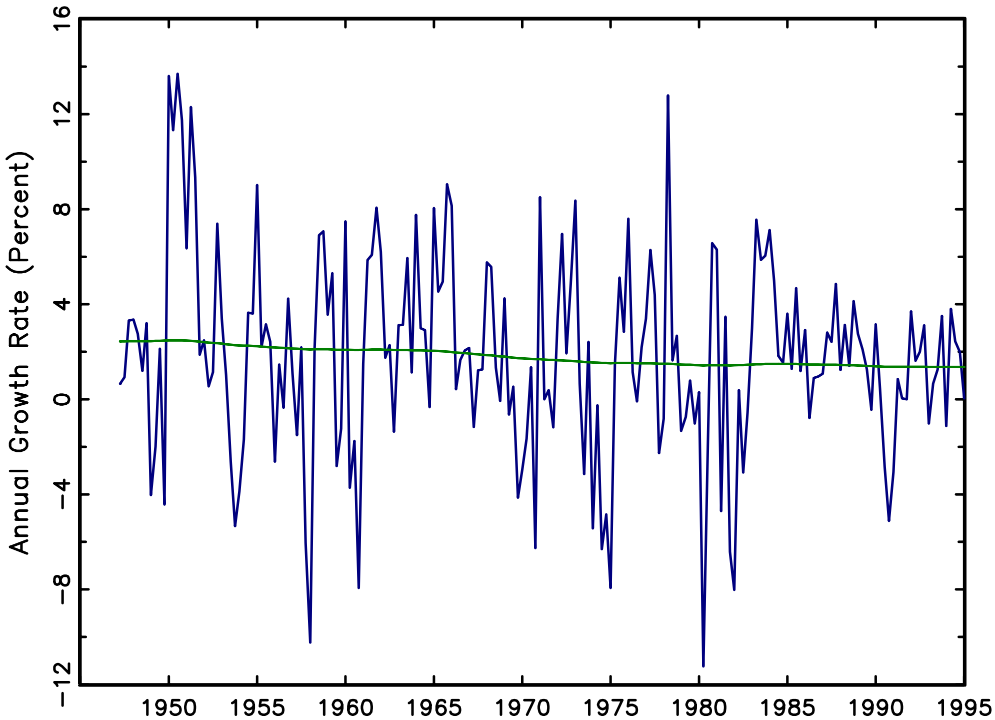
$\sigma_e = \omega_v$ = "long-run" standard deviation of v

Data 1947 -1995

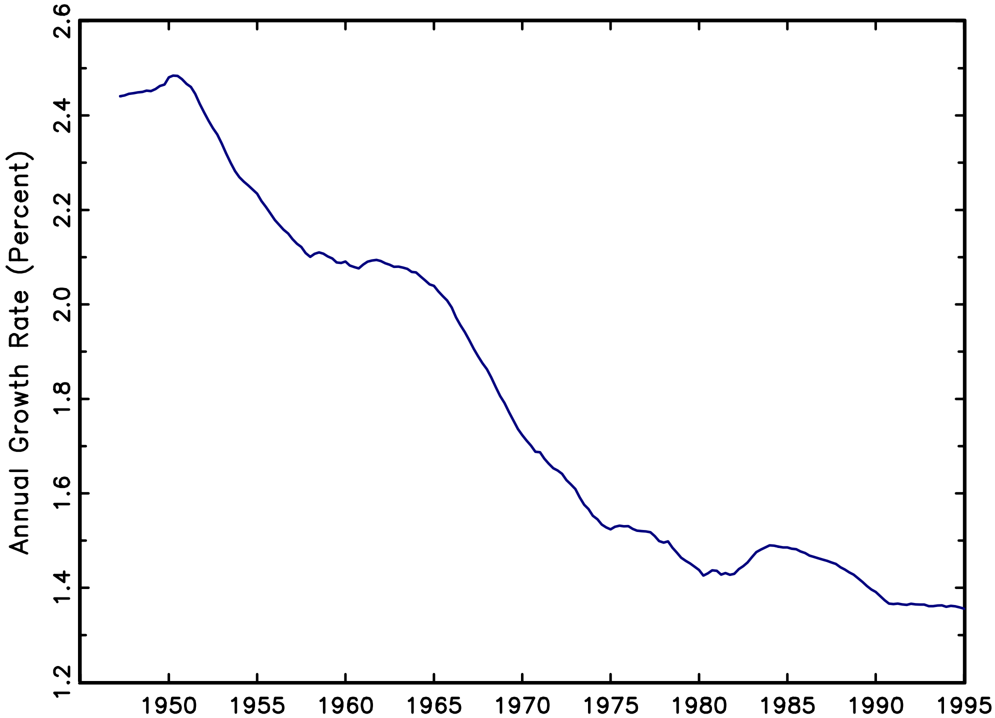
$\hat{g}_{\text{MLE}} = 0$ ($\hat{\sigma}_{\Delta\beta} = 0$)

$\hat{g}_{\text{Mub}} = 4$ ($\hat{\sigma}_{\Delta\beta} = 0.13$ => value of $\hat{\sigma}_{\beta_{1995}-\beta_{1947}} = 5\%$)

GDP Growth Rates and estimated time varying "Mean"



Estimated time varying "Mean"



Estimating the time paths of the time varying coefficients

This will be discussed in our discussion of Kalman Filtering later in the week.

Unit Roots in Autoregressions

References: Hamilton, Chapter 15,17 and Hayashi, Chapter 9

Secondary Reference: Stock, J.H. (1994) "Unit Roots, Structural Breaks and Trends" Handbook of Econometrics, Vol. 4

In this section we carry out analysis for a specific non-stationary model – the autoregressive model with a largest root of unity. This model is important because it provides a better approximation to many economic time series than the stationary model. It must be studied separately because statistical inference differs from the stationary model.

The Gaussian AR(1) Model

We begin with the Gaussian AR(1) model

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

with $y_0 = 0$ and where $\varepsilon_t \sim Niid(0,1)$. Let

$$\hat{\phi} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}$$

denote the least squares estimator or ϕ.

In the stationary model ($|\phi|<1$), standard results are:

1. $\hat{\phi} \to \phi$ at rate $T^{\frac{1}{2}}$

2. The limiting distribution of $\hat{\phi}$ is normal : $\hat{\phi} \overset{a}{\sim} N(\phi, \frac{\sigma^2}{\sum y_{t-1}^2})$,

so that classical linear regression procedures can be used for inference

However, the limiting variance converges to 0 as $\phi \to 1$ suggesting the normal approximation is poor for ϕ close to 1. Indeed when $\phi = 1$ these result change in an important way.

Specifically:

1. $\hat{\phi} \to \phi$ at rate T
2. The limiting distribution of $\hat{\phi}$ is Non-normal

(2) implies that classical linear regression procedures cannot be used for inference. We must develop new approximations.

We will work out some results for a model in which $\phi = 1$. This makes it easy to contrast limiting results to the stationary model. The knife-edge feature of this results ($\phi = 1$ vs. $\phi < 1$) should not be taken seriously. The normal approximation breaks down when ϕ is close to 1 when sample sizes are moderate ($T = 100$ and $\phi = 0.9$, for example.)

To begin, as usual we have the identity

$$\hat{\phi}-\phi=\frac{\sum y_{t-1}\varepsilon_t}{\sum y_{t-1}^2}$$

so that

$$T(\hat{\phi}-\phi)=\frac{\frac{1}{T}\sum y_{t-1}\varepsilon_t}{\frac{1}{T^2}\sum y_{t-1}^2}$$

(Note the non-standard standardization factors involving T.) We will analyze the numerator and denominator separately. We begin with the numerator.

Since $\phi=1$,

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

$$\sum y_t^2=\sum y_{t-1}^2+2\sum y_{t-1}\varepsilon_t+\sum \varepsilon_t^2$$

so that

$$\begin{aligned}\frac{1}{T}\sum y_{t-1}\varepsilon_t &= \frac{1}{2}[\frac{1}{T}\sum y_t^2-\frac{1}{T}\sum y_{t-1}^2-\frac{1}{T}\sum \varepsilon_t^2] \\ &= \frac{1}{2}[\frac{1}{T}y_T^2-\frac{1}{T}\sum \varepsilon_t^2]\end{aligned}$$

Also

$$y_T=\sum_{t=1}^{T}\varepsilon_t$$

so that

$$\frac{1}{T}y_T^2=(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t)^2\overset{d}{\rightarrow}[N(0,1)]^2\sim\chi_1^2$$

and

$$\frac{1}{T}\sum \varepsilon_t^2\overset{p}{\rightarrow}1$$

Putting these together

$$\frac{1}{T}\sum y_{t-1}\varepsilon_t\overset{p}{\rightarrow}\frac{1}{2}[\chi_1^2-1]$$

And thus the numerator is non-normal. From this expression we can learn about the median bias of $\hat{\phi}$. Note that $sign(\hat{\phi}-\phi)=sign(\frac{1}{T}\sum y_{t-1}\varepsilon_t)$, so that

$\Pr(\hat{\phi}<\phi)\rightarrow\Pr(\chi_1^2<1)=0.68$.

The denominator can be analyzed using the FCLT. Write

$$\frac{1}{T^2}\sum_{t=1}^{T} y_t^2 = \frac{1}{T}\sum_{t=1}^{T}[\frac{1}{\sqrt{T}} y_t]^2$$

and let $\bar{W}_T(s) = \frac{1}{\sqrt{T}} y_{[sT]}$ be a step function approximation of $W(s)$. Then

$$\frac{1}{T}\sum_{t=1}^{T}[\frac{1}{\sqrt{T}} y_t]^2 = \int_0^1 \bar{W}(s)^2 ds \Rightarrow \int_0^1 W(s)^2 ds$$

Putting this together with the expression for the numerator, we have shown that

$$T(\hat{\phi}-1) \Rightarrow \frac{\frac{1}{2}[\chi_1^2 - 1]}{\int_0^1 W(s)^2 ds}$$

There is another limiting representations of the random variable that make the dependence of the numerator and denominator clearer:

$$\frac{1}{T}\sum y_{t-1}\varepsilon_t \Rightarrow \int_0^1 W(s)dW(s) \sim \frac{1}{2}[\chi_1^2 - 1]$$

But, the important thing to see, however, is the limiting random variable is non-normal and so standard inference procedures cannot be used.

The non-standard distribution carries over to the t-statistic for testing $H_o : \phi = 1$. A calculation like the one carried above shows

$$t - stat = \frac{\int_0^1 W(s)dW(s)}{[\int_0^1 W(s)^2 ds]^{\frac{1}{2}}}$$

This random variable has a non-normal distribution. Critical values can be found in Hamilton Table B.6

Tests for unit autoregressive roots are sometimes called "Dickey-Fuller" tests, because of the important work that Dickey and Fuller did on this problem.

Odds and Ends

We have worked out results for the simple AR(1) model with no deterministic components. However, to model economic time series, deterministic components must be added to the model.

Two standard formulations are

$$y_t = \alpha + u_t$$
$$u_t = \phi u_{t-1} + \varepsilon_t$$

so that y_t varies around a constant level α. A formulation like this seems appropriate (in the U.S. anyway) for interest rates or the rate of price inflation.

The second formulation is

$$y_t = \alpha + \beta t + u_t$$
$$u_t = \phi u_{t-1} + \varepsilon_t$$

so that y_t varies around a linear trend. A formulation like this seems appropriate (in the U.S. anyway) for the logarithm of output, investment and consumption. Including these components in the model changes the distribution of the estimators for ϕ and for test statistics. See Hamilton, pages 487-501 for details.

The results for the AR(1) model carry over directly to the higher order AR(p) model. We won't go into detail, but the basic idea is as follows. Suppose

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-1} + \varepsilon_t$$

notice that $\sum_{i=1}^{p} \phi_i y_{t-1}$ can be rewritten as

$$\sum_{i=1}^{p} \phi_i y_{t-1} = \phi y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-i}$$

where

$$\phi = \sum_{i=1}^{p} \phi_i \text{ and } \delta_i = -\sum_{j=i+1}^{p} \phi_j$$

(Example: in the AR(2) model, $\phi_1 y_{t-1} + \phi_2 y_{t-2}$ is rewritten as $(\phi_1 + \phi_2) y_{t-1} - \phi_2 (y_{t-1} - y_{t-2})$.)

Thus the AR(p) model can be written as

$$y_t = \phi y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-i} + \varepsilon_t$$

where ϕ denotes the sum of the AR coefficients. Notice that when the AR(p) model contains a unit root, then $\phi = 1$. Thus, one can test for a unit root by regressing y_t onto y_{t-1} and $p-1$ lags of Δy_t and testing the coefficient on y_{t-1} to see whether it is unity. (This is sometimes called an "Augmented Dickey-Fuller Regression.") As in the AR(1) model, the estimated value of ϕ will have a non-standard (non-normal) distribution, as will its t-statistic. The distribution of the t-statistic testing $H_o : \phi = 1$ turns out to be the same as in the unit root AR(1) model, so that the tables in Hamilton can be used for critical values in this formulation.

In the AR(p) regression it turns out that the estimated regression coefficients $\hat{\delta}_i$ and their t-statistics have the usual asymptotic normal properties. You can read about this in Hamilton or in Watson (1994). If time permits, we will discuss the implications of this in class.

As you might imagine the apparent discontinuity in the asymptotic distributions ($\phi = 1$ and $\phi < 1$) makes it difficult to construct confidence intervals for ϕ (where a range of possible true values of ϕ must be entertained). Stock (1991) develops a method for constructing confidence intervals based on local-to unity asymptotic approximations.

Local-to-Unity Asymptotic Approximations

Suppose that

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

where ϕ is close to, but perhaps not exactly equal to 1. For the asymptotic approximations, consider the approximation

$$\phi = 1 - \frac{c}{T}$$

This has the attractive feature that $\lim_{T\to\infty} \phi^T = e^{-c}$, which is continuous in the parameter c. Now, using the same construction as we used above, with $W(\tfrac{t}{T}) = \tfrac{1}{\sqrt{T}} \sum_{j=1}^{t} \varepsilon_j$, and denoting $J_{c,T}(\tfrac{t}{T}) = \tfrac{1}{\sqrt{T}} y_t$ we can rewrite AR(1) model as

$$J_{c,T}(\frac{t}{T}) = (1 - \frac{c}{T}) J_{c,T}(\frac{t-1}{T}) + W(\frac{t}{T}) - W(\frac{t-1}{T})$$

or

$$J_{c,T}(\frac{t}{T}) - J_{c,T}(\frac{t-1}{T}) = (\frac{c}{T}) J_{c,T}(\frac{t-1}{T}) + W(\frac{t}{T}) - W(\frac{t-1}{T})$$

which, as T grows large converges to the diffusion process

$$dJ_c(s) = cJ_c(s)ds + dW(s)$$

Local-to-unity asymptotics use this stochastic process as the key building block for constructing distributions. Since the diffusions depend on the parameter c (which governs the amount of mean reversion), the distributions will depend on c.

We will go through a few examples using this approximation in class.

Empirical Example 1: Volatility of long rates and short rates

Largest AR root

	1965:1-1987:9			1985:1 – 1998:9		
Interest Rate	ols	mub	90% CI	ols	mub	90% CI
Federal Funds	0.97	0.96	0.91-1.01	0.98	1.00	0.94-1.02
3-Month TBill	0.96	0.98	0.93-1.02	0.98	0.99	0.94-1.02

Implied Standard Deviation of Interest Rates

	1965:1-1987:9			1985:1 – 1998:9		
	Actual	OLS	MUB	Actual	OLS	MUB
1 Month	**2.44**	2.39	2.39	**1.50**	1.36	1.41
12 Month	1.40	1.80	1.76	1.54	1.31	1.46
60 Month	0.89	0.57	0.52	1.40	0.76	1.22
120 Month	**0.69**	0.31	0.29	**1.29**	0.46	0.98

Example 2: Largest autogressive roots for real exchange rates

$e_t = \rho e_{t-1} + \varepsilon_t$

Let h solve $\rho^h = ½$ (h is the "half-life" of a shock)

$H_0 : \rho = 1$ cannot be rejected for many real echange rates, so $h = \infty$ cannot be rejected.

But, what is a lower bound on the 95% confidence interval for h ?

Rossi (2005) JBES

Country	Lower bound of 95% CI
Australia	61.3
Canada	60.6
Germany	4.5
Japan	21.9
Switzerland	4.0
UK	8.1

Filtering, State Space Models and the Kalman Filter

Primary Reference: Hamilton, Chapter 13

The Linear State Space Model is a general framework for analyzing linear time series models. Indeed, it so general that every linear model that I can think of is a special case of this model. It, and the Kalman Filter, are useful for two reasons:

1. Sometimes models contain "latent" variables, and the Kalman filter is a algorithm for constructing minimum mse estimates of these variables

2. The likelihood function for any unknown parameters is easily calculated using the Kalman filter.

The Basic Model

Measurement Equation

$$y_t = Ax_t + H'\xi_t + w_t$$

with

$$E(w_t w_t) = R$$

Transition (or State) Equation

$$\xi_t = F\xi_{t-1} + v_t$$

with

$$E(v_t v_t') = Q$$

Variables

y_t is a vector of observed variables

x_t is a vector or deterministic components (constants, trends, seasonal components, etc.)

ξ_t is an unobserved vector of “state” variables

w_t and v_t are unobserved, mutually uncorrelated and serially uncorrelated noise variables.

A, H, R, F and Q are “system” matrices that may depend on unknown parameters.

Simple example: $y_t \sim AR(p)$

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \varepsilon_t$$

can be represented with the following state space model

$$\xi_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}$$

$$F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & 1 & 0 \end{bmatrix}$$

$$v_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and $w_t = 0$, $A = 0$ and $H' = [1 \;\; 0 \ldots 0]$

Signal Extraction and the Kalman Filter

Let

$\xi_{t/k} = E(\xi_t \mid \{y_i\}_{i=1}^k)$

$P_{t/k} = Var(\xi_t \mid \{y_i\}_{i=1}^k)$

The Kalman filter is a recursive algorithm for constructing $\xi_{t/t}$ and $P_{t/t}$. That is, the Kalman Filter is a function that constructs $\xi_{t/t}$ and $P_{t/t}$ from $(\xi_{t-1/t-1}, P_{t-1/t-1}, y_t, x_t)$. To derive the filter, we assume

$$\begin{bmatrix} w_t \\ v_t \end{bmatrix} \sim Niid(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R & 0 \\ 0 & Q \end{bmatrix}).$$

Recall the following fact from the multivariate normal distribution:

Suppose

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \sim Niid(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}),$$

then

$$E(z_1 \mid z_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(z_2 - \mu_2)$$

and

$$Var(z_1 \mid z_2) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

The Kalman Filter is an application of these formula.

We suppose $\xi_{t-1/t-1}$ and $P_{t-1/t-1}$ are known and that

$$\begin{bmatrix} \xi_t \\ y_t \end{bmatrix} \mid \{y_i\}_{i=1}^{t-1} \text{ is Normally Distributed}$$

We then use the multivariate normal formula with $z_1 = \xi_t$ and $z_2 = y_t$. The details are

18. $\xi_{t/t-1} = F\xi_{t-1/t-1}$ $\quad (\mu_1)$

19. $y_{t/t-1} = Ax_t + H'\xi_{t/t-1}$ $\quad (\mu_2)$

20. $P_{t/t-1} = FP_{t-1/t-1}F' + Q$ $\quad (\Sigma_{11})$

21. $h_t = H'P_{t/t-1}H + R$ $\quad (\Sigma_{22})$

22. $K_t = P_{t/t-1}Hh_t^{-1}$ $\quad (\Sigma_{12}\Sigma_{22}^{-1})$

23. $\nu_t = y_t - y_{t/t-1}$ $\quad (z_2 - \mu_2)$

24. $\xi_{t/t} = \xi_{t/t-1} + K_t\nu_t$ $\quad (\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(z_2 - \mu_2))$

25. $P_{t/t} = P_{t/t-1} - K_tH'P_{t/t-1}$ $\quad (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

Beginning the recursion – two cases:

ξ_t is covariance stationary, then

$\xi_{0/0} = E(\xi_0) = 0$

$P_{0/0} = Var(\xi_0)$ so that

$$P_{0/0} = FP_{0/0}F' + Q$$

or

$$vec(P_{0/0}) = [I - F \otimes F]^{-1} vec(Q)$$

where this uses the result that $vec(ABC) = (C' \otimes A)vec(B)$ for comformable matrices A, B, and C. (see Hamilton)

ξ_t is not covariance stationary (see the discussion in Hamilton)

A related recursion, called a "Smoother" can be used to construct $\xi_{t/T}$ and $P_{t/T}$ (again, see Hamilton)

Likelihood function

$y_t \mid \{y_i\}_{i=1}^{t-1}$ is normal with mean $y_{t/t-1}$ and variance h_t.

Log Likelihood is therefore

$$L = \kappa - \frac{1}{2}\sum_{t=1}^{T}\{\ln(h_t) + \frac{\nu_t^2}{h_t}\}$$

where $\nu_t = y_t - y_{t/t-1}$. Notice that both h_t and ν_t are computed by the Kalman filter.

Kalman Filter Examples

Here I present a series of examples. To avoid confusion with variables defined in the state-space model ($y, x,$ etc.), I will use (often) ~ over many of the variables.

Example 1. Standard Signal Extraction

Write the scalar $\tilde{y}$ as

$$\tilde{y}_t = \tilde{x}_t + \varepsilon_t$$

with

$$\tilde{x}_t = \phi\tilde{x}_{t-1} + e_t$$

In this model $\tilde{x}_t$ is the "signal" which is not directly observed. Instead the "measurement" $\tilde{y}_t$ is observed. The component ε is the measurement noise. To write this in state space form, use the definitions: $y = \tilde{y}$, $A = 0, x = 0, H = 1, \xi = \tilde{x}, F = \phi, w = \varepsilon$ and $v = e$.

Example 2: Random Walk Unobserved Components Model

This is the same as Example 1 but with $\phi = 1$.

Empirical Example: US Price Inflation

$\pi_t = \tau_t + \eta_t$
$\tau_t = \tau_{t-1} + \varepsilon_t$

ε_t and η_t are uncorrelated white noises

or IMA(1,1)

$\Delta\pi_t = \varepsilon_t + \Delta\eta_t = (1 - \theta B)a_t$

θ is an increasing function of $\sigma_\eta^2 / \sigma_\varepsilon^2$

UC: $\pi_t = \tau_t + \eta_t,\ \tau_t = \tau_{t-1} + \varepsilon_t$

IMA: $\Delta\pi_t = \varepsilon_t + \Delta\eta_t = (1 - \theta B)a_t$

	GDPD		PCE-core		PCE-all	
	1960:I – 1983:IV	1984:I – 2004:IV	1960:I – 1983:IV	1984:I – 2004:IV	1960:I – 1983:IV	1984:I – 2004:IV
UC parameters						
σ_ε	0.91	0.26	**0.79**	**0.20**	0.96	0.30
σ_η	0.66	0.61	**0.53**	**0.50**	0.64	0.80
IMA parameters: $\Delta\pi_t = (1 - \theta B)a_t$						
θ	0.28	0.66	**0.25**	**0.68**	0.25	0.69
σ_a	1.26	0.75	1.05	0.60	1.27	0.97

Example 3: Missing data:

Suppose that the scalar $\tilde{y}$ is generated by the ARMA(1,1) model

$$\tilde{y}_t = \phi \tilde{y}_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

and suppose that observations are available on $\tilde{y}_t$ for $t = 1,...,45,56,...,100$, but not available for $t = 46,...,55$. How can the likelihood be formed? How can the missing values be estimated? The answer is the use the prediction error decomposition of the likelihood and the Kalman filter/smoother. To write the model in state-space form define

$$\xi_t = \begin{bmatrix} \tilde{y}_t \\ \varepsilon_t \end{bmatrix}, F = \begin{bmatrix} \phi & -\theta \\ 0 & 0 \end{bmatrix}, v_t = \begin{bmatrix} \varepsilon_t \\ \varepsilon_t \end{bmatrix}$$

With this definition of v then

$$Q = \sigma_\varepsilon^2 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

The other parameters are $A = 0, x_t = 0$ and $w_t = 0$ (so that $R = 0$). The matrix H_t is time varying with $H_t' = [10]$ for $t = 1,...,45,56,...,100$ and $H_t' = [00]$ otherwise. Finally $y_t = \tilde{y}_t$ for $t = 1,...,45,56,...,100$ and $y_t = 0$ otherwise.

Empirical Example:

Constructing a Monthly version of GDP for the U.S. (Chow and Lin (1971), and Bernanke, Gertler, Watson (1998) BPEA).

Example 4: Time Varying Parameter Regression model

Suppose that the scalar $\tilde{y}$ is generated by the regression model

$$\tilde{y}_t = \tilde{x}_t^{'}\beta_t + \varepsilon_t$$

with

$$\beta_t = \beta_{t-1} + e_t$$

This model can be estimated using the state-space model with $\xi_t = \beta_t$, $y_t = \tilde{y}_t, F = I$, $v_t = e_t, w_t = \varepsilon_t$, $H_t = x_t$, $A = 0$ and $x_t = 0$.

Example: TV NAIRU (Gordon, Staiger/Stock/Watson)

$$\Delta\pi_{t+1} = \alpha(L)\Delta\pi_t + \beta(u_t - u_t^N) + \gamma(L)\Delta u_t + \gamma Z_t + v_{t+1}$$

$$= \mu_t + \alpha(L)\Delta\pi_t + \beta u_t + \gamma(L)\Delta u_t + \gamma Z_t + v_{t+1}$$

where $\mu_t = -\beta u_t^N$

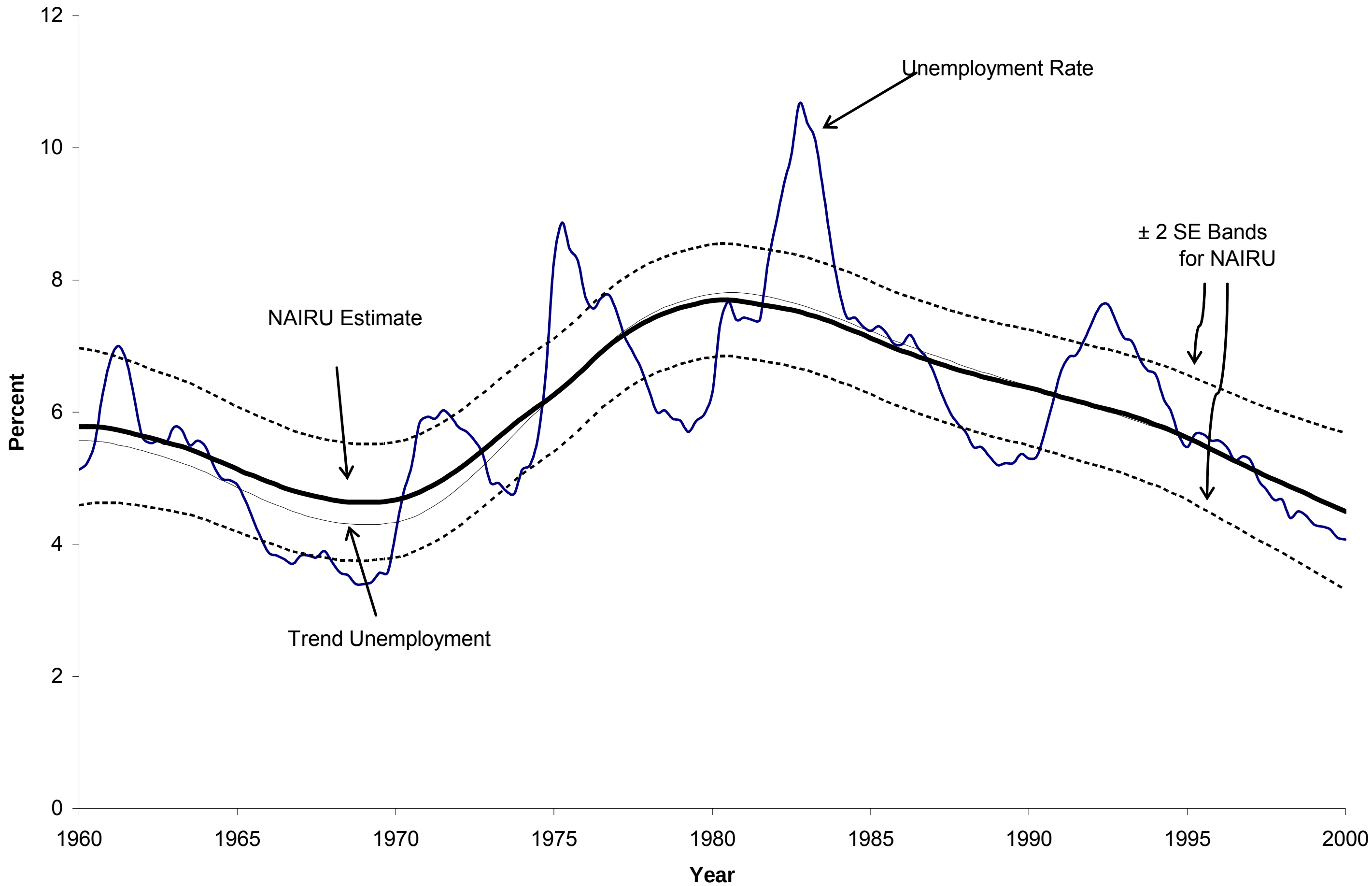
Modeled as

$\mu_t = \mu_{t-1} + e_t$

Smoothed estimate or NAIRU is

$$\hat{u}_{t/T}^n = -\frac{\mu_{t/t}}{\hat{\beta}}$$

Figure 3
NAIRU From Price Phillips Curve



Example 5: Dynamic Factor model:

Suppose the $n \times 1$ vector $\tilde{y}$ is generated by the model

$$\tilde{y}_t = \alpha \tilde{F}_t + \varepsilon_t$$

$$\tilde{F}_t = \phi \tilde{F}_{t-1} + e_t$$

where the unobserved variable $\tilde{F}_t$ is $m \times 1$ where $m < n$, and $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$ is a diagonal matrix. This model thus describes the covariability between the elements of $\tilde{y}$ in terms of a small number of unobserved factors. This model can analyzed by writing the model in state-space form with $y = \tilde{y}$, $H = \alpha$, $\xi = \tilde{F}, F = \phi, w = \varepsilon, v = e, A = 0$ and $x = 0$.

Empircal Example:

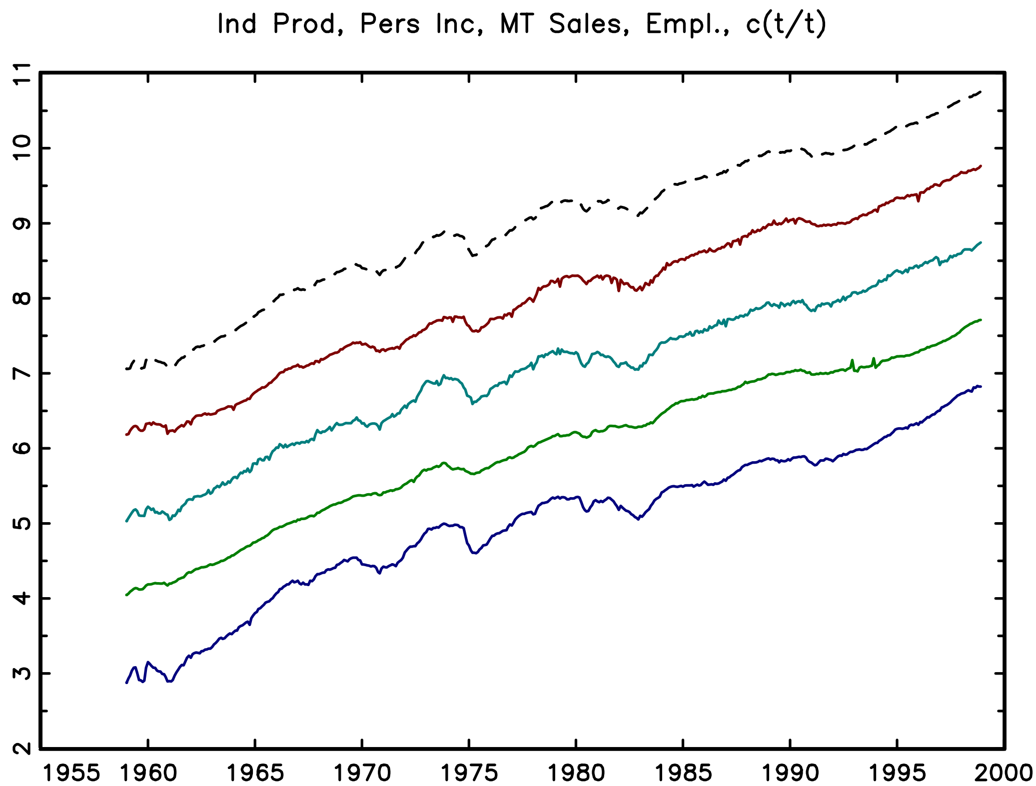
Rationalizing the "Index of Coincident Indicators"

$$\begin{bmatrix} \Delta PerIncome \\ \Delta MTSales \\ \Delta IP \\ \Delta Employment \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} C_t + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}$$

C_t ~ AR(2)

u_{it} ~ AR(2) (mutually independent)

All parameters estimated by ML.


Ind Prod, Pers Inc, MT Sales, Empl., c(t/t)
2
3
4
5
6
7
8
9
10
11
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000

Modifications of the Basic Model

Stochastic Volatility

Suppose

$y_t = \sigma_t \varepsilon_t$

$\ln(\sigma_t^2) = \ln(\sigma_{t-1}^2) + v_t$

where v_t and ε_t are iid Normal with means of zeros and variance of σ_v^2 and 1 (a normalization).

Squaring y and taking logs yields

$x_t = \beta_t + e_t$

$\beta_t = \beta_{t-1} + v_t$

where $x_t = \ln(y_t^2)$, $\beta_t = \ln(\sigma_t^2)$ and $e_t = \ln(\varepsilon_t)$.

This model has the same form as example 2 (the UC model), but now e_t is not Gaussian, rather it is the logarithm of a squared standard normal.

Shephard (1994, Biometrika) proposes a convenient method to approximated the optimal non-Gaussian filter in this model using a mixture of normals approximation for e_t. We will discuss this in class.

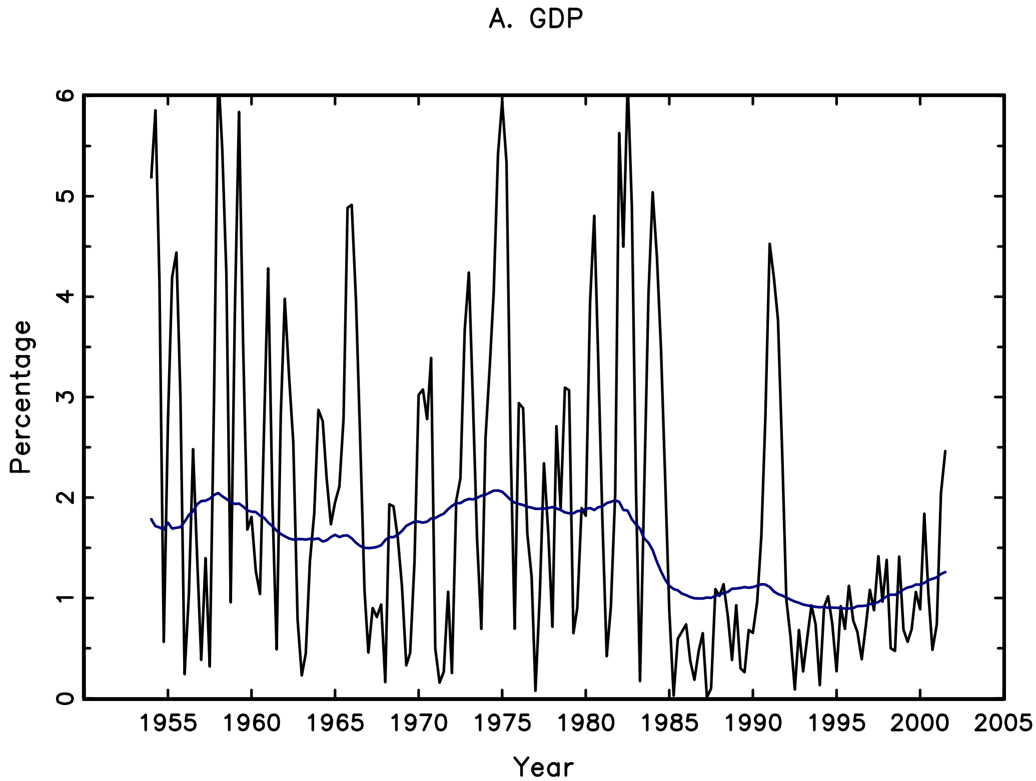
Emprical Example 1: (Stock-Watson, NBER MA 2002)

GDP Volatilty:

$\phi(\mathrm{L})\Delta y_i = \mu + \sigma_t\varepsilon_t$

$\ln(\sigma_t^2) = \ln(\sigma_{t-1}^2) + v_t$

where v_t has a mixture of normals distribution (e.g. a distribution with very fat tails), allowing large jumps in σ_t.


A. GDP
Percentage
Year
0
1
2
3
4
5
6
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005

Example 2: U.S. Inflation (Stock-Watson (2005))

$$\pi_t = \tau_t + \eta_t, \quad \text{where } \eta_t = \sigma_{\eta,t}\zeta_{\eta,t}$$

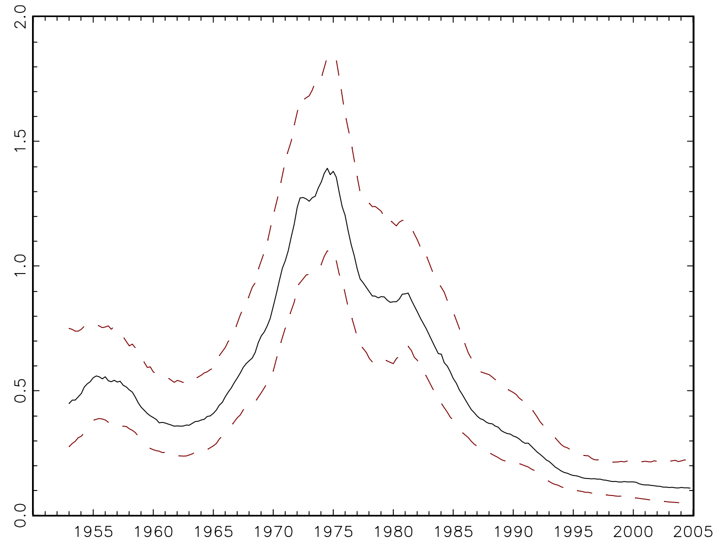
$$\tau_t = \tau_{t-1} + \varepsilon_t, \quad \text{where } \varepsilon_t = \sigma_{\varepsilon,t}\zeta_{\varepsilon,t}$$

$$\ln \sigma^2_{\eta,t} = \ln \sigma^2_{\eta,t-1} + \psi_{\eta,t}$$

$$\ln \sigma^2_{\varepsilon,t} = \ln \sigma^2_{\varepsilon,t-1} + \psi_{\varepsilon,t}$$

where $\zeta_t = (\zeta_{\eta,t},\ \zeta_{\varepsilon,t})$ is i.i.d. $N(0, I_2)$, $\psi_t = (\psi_{\eta,t},\ \psi_{\varepsilon,t})$ is i.i.d. $N(0, \gamma I_2)$, and ζ_t and ψ_t are independently distributed, and γ is a scalar parameter.

(a) Standard deviation of permanent innovation, $\sigma_{\varepsilon,t}$



(b) Standard deviation of transitory innovation, $\sigma_{\eta,t}$