POL 606 Time Series Analysis
Week 3 - Intervention and Transfer Function Models
February 13, 18, 20 2008

ARIMA models may be augmented by the inclusion of dummy (0-1) and continuous right-hand-side variables.

The former are known as “interventions,” – a term given by Box and Tiao (1975).

Continuous right-hand-side variables are called “transfer functions” – a term from chemical engineering.

In political science applications, intervention variables are typically used to measure the effects of public policy innovations (e.g. the introduction of seat-belt laws on traffic fatalities), the effects of various unanticipated events such as foreign policy crises or wars on levels of presidential approval; scandals, or the impact of economic shocks such as the OPEC oil embargo on economic growth.

Examples of continuous variables include levels of consumer confidence in models of public support for governing political parties and their leaders, and interest rates in models of unemployment rates.

These intervention and transfer-function components can be hypothesized to exert either *abrupt* or *gradual* effects.

And, in the case of interventions, the effects can be hypothesized to be either *permanent* or *temporary*.

**Intervention Variables**

Measured as dummy variables.

Although we can think about things impacting over more than one period.

Impact assessment is concerned only with *events*.

Impact assessment requires that the onset of an event be specified beforehand.

We begin with an ARIMA model before we specify these impacts.

We now refer to that portion of our explanation of the dependent variable as the *noise model*, and denote it: $N_t$.

So the full impact assessment model can be written as:

$$Y_t = f(I_t) + N_t$$

Where $N_t$ denotes the noise component – any combination of AR, MA and differencing components and $f(I_t)$ denotes a function of the variable $I_t$..

* Sometimes the impacts of interventions can be big enough that they distort the PACF and the ACF and make diagnosing the ARIMA model more difficult.

This is similar to the problems that outliers present in analyzing cross-sectional data.

One way to avoid this problem in identification is to estimate ACFs and PACFs from the pre- or post-intervention portion of the series only, or both separately.

The hypothesis for the effects of the intervention are generated theoretically – that is, we aren't just letting the data speak by themselves.

Then the parameters of the "full impact assessment model" – which includes both noise and intervention components) will be estimated.

The parameters are checked for significance and common sense.

Once a tentative model is specified and found significant with acceptable parameter estimates, move on to diagnosis.

Again, residuals should be white noise.

Then move on to interpretation – a big difference here from before is that with the building of the noise model we are not very concerned with the theoretical implications, here we are – impact assessment is done so that we can interpret the effects of events.

A general model:

$$Y_t = \frac{\omega_0}{(1-\delta L)} I_t + N_t$$

Where $I_t$ is the intervention scored 1 for the month it occurs and 0 otherwise.
$N_t$ is the noise models, in the simplest case $N_t = \varepsilon_t$ .
$\omega_0$ is the impact parameter – tells us the initial effects. ($\omega$ =omega)
$\delta$ is the decay parameter – we use this for "gradual permanent" effects and "gradual temporary" effects. ($\delta$ =delta)

With the most simple types of interventions, where $\delta$ =0, the impact assessment model is:

$$Y_t = \omega_0 I_t + N_t$$

If we subtract the noise model from the dependent variable we get:

$$Y_t^* = Y_t - N_t$$
$$Y_t^* = \omega_0 I_t$$

Think of $Y^*$ as the deviation from the noise model.

So, prior to the event, when $I_t$ =0:

$Y_t^* = \omega_0(0)$ or $Y_t^* = 0$

After the onset of the event:

$Y_t^* = \omega_0(1)$ or $Y_t^* = \omega_0$ .

We can conceive of $I_t$ as being measured in (at least) two possible ways giving us our first two types of interventions.

**Abrupt Permanent – a.k.a. step function.**

One large step leading to a new level. For example, the impact of the polio vaccine on polio deaths in the United States.

Here, $I_t$ is measured as 0 before the event occurs and 1 when it occurs and 1 ever after.

$I_t$ = …0 0 0 0 0 1 1 1 1 1 …where the first 1 appears at *t*.

We use this coding scheme for any permanent effect.

**Abrupt Temporary – a.k.a. "Pulse effect"**

One large increase that lasts a very short time. For example, the popularity of minor parties increases drastically when there are by-elections in the UK.

Any temporary effect is coded as: $I_t$ = …0 0 1 0 0 0 …

For either of these, we can have multiple $\omega$s.

For example, a two step process:

$$y_t = \omega_0 I_t + \omega_1 I_{t-1} + N_t$$

**Gradual Effects**

In addition to modeling abrupt permanent and temporary effects we can also model gradual effects.

So there are two dimensions.

1. Duration – some have permanent effects, some temporary.

2. Nature of impact – is it abrupt or gradual?

Four types of interventions to model – actually, there can be other types as well but these 4 cover most situations.

Each combination gives us one of our four types of interventions.

1. abrupt permanent – a.k.a. "step function"
2. gradual permanent
3. abrupt temporary – a.k.a. "pulse effect"
4. gradual temporary – a.k.a. "pulse decay"

Do these sound familiar to anyone? In "Issue Evolution", Carmines and Stimson use these to discuss the effects of new issues.

$$Y_t = \frac{\omega_0}{(1-\delta L)} I_t + N_t$$

The slope parameter is $\delta$ (delta) which estimates the adjustments subsequent to the change (pulse decay) or the overall rate of increase (gradual permanent).

$\delta$ s are only for gradual temporary and gradual permanent effects – the abrupt effects are thereby easier to deal with.

Typically not more than one adjustment parameter – $\delta$ .

Typically, estimates of $\delta$ are positive.

If negative, adjustment parameter is like a "declining saw."

$|\delta|<1$ for a stationary process. > 1 for a non-stationary process.

To model a permanent effect at time *t*:

$I_t$ = …0 0 0 1 1 1 …where the first one appears at *t*.

To model a temporary effect:

$I_t$ = …0 0 1 0 0 0 …

**Gradual Permanent**

A slow change that leads to a new permanent level. For example, an undergraduate adding an additional hour of studying to their grades.
So the coding here is: $I_t$ = …0 0 1 1 …where the first one appears at *t*.

$$Y_t = \frac{\omega_0}{(1-\delta L)} I_t + N_t$$

Where $-1 < \delta_1 < 1$.

These are called the *bounds of system stability*. If the value of $\delta_1$ lies outside these bounds, the impact assessment model is unstable – a condition identical to non-stationarity.

When $\delta_1$ =/> 1, the post-intervention time series is nonstationary. This is interpreted to mean that the event has triggered a trend in the series.

The above is equivalent to:

$$Y_t^*(1-\delta L) = \omega_0 I_t$$
$$Y_t^* - \delta Y_{t-1} = \omega_0 I_t$$
$$Y_t^* = \delta Y_{t-1} + \omega_0 I_t$$

Impact at *t* is equal to $\omega$.

Impact at *t+1*:

$$Y_{t+1}^* = \delta Y_t^* + \omega_0 I_{t+1}$$
$$= \omega_0 (1+\delta)$$

Impact over time is:
$$\omega_0 (1+\delta+\delta^2+\delta^3+\delta^4+\delta^5+\delta^6 ...)$$

Assume that $y_{t-1}^*$= 0, $\omega = 2$, and $\delta = 0.5$

So that:

$Y_t^* = 0.5(Y_{t-1}) + 2I_t = 0.5(0) + 2(1) = 2$
$Y_{t+1}^* = .5(Y_t) + 2I_{t+1} = .5(2) + 2(1) = 3$
$Y_{t+2}^* = .5(Y_{t+1}) + 2I_{t+2} = .5(3) + 2(1) = 3.5$
$Y_{t+3}^* = 3.75$

The change produced by a gradual permanent intervention is:

$$\omega_0(1 + \delta + \delta^2 + \delta^3 + \delta^4 + \delta^5 + \delta^6 ...)$$

So:

$$Y_{i+n}^* = \sum_{k=0}^{n} \delta_1^k \omega_0$$

Because $\delta < 1$, $\delta^{10}$ is infinitesimal.

The effects as time go on get smaller.

We can calculate the long-tem change (asymptotic change) produced by a gradual permanent intervention:

$$= \frac{\omega_0}{1-\delta}$$

The closer $\delta_1$ is to 1 the longer it takes for the impact to be fully realized.

Above example is:

$$= \frac{2}{1-.5} = 4$$

Back to the bounds of system stability:

What if $\delta = 1$?

Then, $Y_t = \frac{\omega_0}{(1-\delta L)} I_t + N_t$ becomes:

$Y_t = \frac{\omega_0}{(1-L)} I_t + N_t$ or:

$$Y_t^* - Y_{t-1}^* = \omega_0 I_t$$

Since $I_t$ is scored as 0 0 0 1 1 1 …. this creates a nonstationary series from a stationary one.

This is a very unlikely scenario – it means going from a state of equilibrium to a state of growth.

However it may appear that this is what is going on in our data if the period after our intervention is too short to see the return to equilibrium. One solution is to wait for more data to arrive.

So, be wary of $\delta$ if it is negative or very close to 1.

**Gradual Temporary Effect**

a.k.a. "pulse decay". A very quick change followed by a gradual shift back to its expected value prior to the intervention. For example, foreign policy effects, rally around the flag, desert storm, 9/11.

Coded as: $I_t$ = …0 0 1 0 0 0 …

Looks like a jump up (or down) followed by a gradual decline (or rise) to its original position.

Impact at *t*= $\omega_0$

At *t+1*: $\delta\omega$

At *t+2*: $\delta^2\omega$

At *t+i*: $\delta^i\omega$

For example, if $\delta_1$=0.5

The impact is:
At *t:* $\omega_0$
At *t+1:* $\delta\omega = 0.5\,\omega$

So, the $\delta$ parameter tells us how much of the initial impact remains in the first period after the intervention and how much of the remaining impact remains in the next period and so on.

We see an event have an impact but that impact is only temporary.

Having gradual permanent effects might be seen more in public policy where some policy change has an initial impact and continues to have an impact while the policy is adopted. Common in health care – the improvements based on a drug treatment. for example.

**Differencing Interventions**

What we do to our dependent variable, we also do to our independent variables.

We want to have the same level of analysis for our independent variables as we have for our dependent variable.

So, if we difference our dependent variable and our unit of analysis is the differences from one time point to the next, we must do the same for our interventions.

$I_t$ = …0 0 0 0 0 1 1 1 1 1 … becomes: $\Delta I_t$ = …0 0 0 0 -1 0 0 0…

and:

$I_t$ = …0 0 0 0 1 0 0 0 … becomes: $\Delta I_t$ = …0 0 0 -1 1 0 0 …

What if we fail to do this?

Example:

$\Delta y_t = y_t - y_{t-1}$

$\Delta y_t = \omega_0 I_t + N_t$ where $N_t = \varepsilon_t$

Suppose an abrupt permanent effect is left in level form:

$I_t$ = …0 0 0 0 0 1 1 1 1 1

And suppose that $\omega_0 = 2$ and $y_{t-1} = 0$.

Then: $\Delta y_t = 2I_t$

or: $y_t - y_{t-1} = 2I_t$
$y_t = y_{t-1} + 2I_t$
$y_t = 0 + 2(1)$
So $y_t = 2$

And at time $t$+1:

$$y_{t+1} = y_t + 2I_{t+1}$$
$$y_{t+1} = 2 + 2(1)$$
$$y_{t+1} = 4$$

And at time *t*+2:

$$y_{t+2} = y_{t+1} + 2I_{t+2}$$
$$y_{t+2} = 4 + 2(1)$$
$$y_{t+2} = 6$$

And so on.

We want to model the intervention as a permanent step increase in $y_t$ but instead it seems to be trending.

So, remember to difference!

Here's the right way to do it:

Again, suppose that $\omega_0 = 2$ and $y_{t-1} = 0$.

Now:

$$y_t - y_{t-1} = \omega_0 I_t - \omega_0 I_{t-1}$$
$$y_t = y_{t-1} + \omega_0 I_t - \omega_0 I_{t-1}$$
$$y_t = 0 + 2(1) - 2(0)$$
$$y_t = 2$$

and at $y_{t+1}$:

$$y_{t+1} = y_t + \omega_0 I_{t+1} - \omega_0 I_t$$
$$y_{t+1} = 2 + 2(1) - 2(1)$$
$$y_{t+1} = 2$$

By differencing the intervention variable we are subtracting out that additional effect and modeling what we want, an abrupt permanent increase of 2 at time *t*.

Same is true to model temporary effects – you can work your way through a couple of these.

**Testing Rival Impact Hypotheses**

Now that we have different types of impacts, how do we determine which applies?

Ideally we would have a theoretical reason for expecting a particular type of impact.

When theory is lacking, however, we can test rival hypotheses.

Looking at our general model, we can make tests of parameters.

$$y_t^* = \frac{\omega_0}{1-\delta_1 L} I_t$$

We can set the intervention in two ways – as 1 only at the time of the event or as a 1 at the time of the event and ever after.

If we are blind to what to expect, we can begin by estimating two parameters, $\omega_0, \delta_1$, and setting the intervention as …0 0 0 1 0 0 0….

If we get a very small value for $\delta_1$ it tells us that our impact is a temporary one – and if it equals zero we can conclude it is an abrupt temporary one.

$y_t^* = \frac{\omega_0}{1-L} I_t$ reduces to: $y_t^* = \omega_0 I_t$

If we had a value of $\delta_1$ close to or equal to 1, $y_t^* = \frac{\omega_0}{1-\delta_1 L} I_t$ reduces to

$$y_t^* = \frac{\omega_0}{1-L} I_t$$
$$y_t^*(1-L) = \omega_0 I_t$$

So the effect is remembered permanently.

So our method here with no prior information is to first hypothesize a temporary effect and if $\delta_1$ is too large, near 1, a temporary impact can be ruled out.

Next, we can hypothesize about a permanent but gradual pattern of impact.

Finding values of $\delta_1$ with standard errors showing it not to be equal to 0 or 1 would lead us to conclude a gradual effect.

**Interpreting Interventions with a Logged Dependent Variable**

When our data are untransformed or even just differenced, the interpretation of intervention parameters is straightforward.

We can interpret an impact parameter as the change in basic level of the series from one series to the next.

When our series has been transformed using logs, however, interpretation is more difficult. Again, we would transform our data using logarithms if our data are variance non-stationary.

What our impact parameter $\omega_0$ tells us is the change in the log of the series.

Presentation of time series results can be difficult enough without explaining to people that when $I_t$ occurs the natural logarithm of the series changes by $\omega_0$.

Simple convention allows us to perform the analysis in the log metric but state our findings in terms of the raw metric.

First, determine the ARIMA model and the intervention parameters.

Using the simplest intervention, we have:

$$Ln(y_t) = \omega_0 I_t + ARIMA$$

We can exponentiate this to:

$$e^{Ln(y_t)} = e^{\omega_0 I_t + ARIMA}$$
$$Yt = e^{\omega_0 I_t} e^{(ARIMA)}$$

The term $e^{(ARIMA)}$ denotes a multiplicative shock form of the ARIMA model.

Lets skip a bunch of steps here so we don't get bogged down in the math (but see Macleary and Hay pages 173-174 for the rest).

Think of the pre- and post-equilibrium levels of the series. The ratio of post- to pre-intervention equilibrium is:

$$\frac{post\,\mathrm{int}\,ervention..equilibrium}{pre\,\mathrm{int}\,ervention..equilibrium} = \frac{e^{(\varpi_0)} e^{(ARIMA)}}{e^{(ARIMA)}} = e^{(\varpi_0)}$$

While the parameter $\omega_0$ is not easy to interpret in the log metric, the term $e^{\omega_0}$ can be interpreted as the ratio of the post-intervention series level to the pre-intervention series level.

This ratio can be expressed as the percent change in the expected value of the process that is associated with the intervention:

Percent change = ($e^{\omega_0}$ -1)*100

Remember that $e \approx 2.7183$

Lets say that $\widehat{\omega}_0 = -.20$.

$$\widehat{\omega}_0 = -.20$$
$$e^{-.2} = 0.819$$
$$\%\Delta = (0.819 - 1)100 = -0.181(100) = -18.1\%$$

Thus, if the impact parameter for the logged series is -.2, the series decreases by 18.1% due to the intervention.

With 2 impact parameters we can sum them:

Percent change = $(e^{\omega_0+\omega_1} - 1)100$.

With two impacts and an adjustment parameter:

$$(e^{\frac{\omega_0+\omega_1}{\delta}} - 1)100.$$

**Transfer Function Analysis**

Box-Jenkins model with continuous variables on the right-hand-side of the model.

For example, explaining popularity in terms of employment, income, etc.

If $X_t$ is a continuous variable:

$$y_t = f(x_t) = \frac{(\omega_0 + \omega_1 + ... + \omega_t)}{(1 - \delta_1 L - \delta_2 L - ... - \delta_T L)} x_{t-i}$$

An example of a (0,1,1) ARIMA model with one intervention and one transfer function component is:

$$(1-L)Y_t = \frac{\omega_1}{(1-\delta_1 L)}(1-L)I_t + \frac{\omega_2}{(1-\delta_2)}(1-L)X_t + \varepsilon_t + \varepsilon_{t-1}$$

An important step is to ensure that what is done to the left side is also done to the right.

If $Y_t$ is differenced, the right-hand-side variables should be as well.

Once a model is estimated, diagnostics are used to check its adequacy.

Again, re-specification, re-estimation, and re-diagnosis may be required.

Popular applications in political science include modeling the dynamics of presidential approval and other popularity measures using political events and economic measures.

Large literature in both U.S. and British political science with smaller amounts in other developed countries – depending on the availability of data.

We will identify the nature of the model, estimate parameters and diagnose the adequacy of the model.

Steps in Transfer Function Analysis:

1. Univariate ARIMA analyses of dependent variable and continuous independent variables.
2. Identify the transfer function components.
    - prewhitening
    - cross correlation function (CCF) analysis.
3. Estimate Model
    - all independent variables at the same time.
4. Diagnostics on residuals.
5. Revise noise model if needed & revise specification.
6. Re-estimate and re-diagnose until we have white noise residuals.

Step 1 we know already – find the appropriate ARIMA model for *each* series.

Step 2 can be complicated.

Specification of the effects of continuous variables in transfer function models is often accomplished by "cross-correlating" the dependent and independent variables.

We do this if we are unsure about "lags or leads" that is, how is the independent variable affecting the dependent variable. Is there a lag? If so, of how many periods?

Cross correlations are correlations between $Y_t$ and $X_{t-i}$ where $i$=0,1,…$p$.

The aim is to detect at what lag X might affect Y – that is, we do not assume that X can affect Y only contemporaneously – there may be a delay of one or several periods. This adds a great number of possible model combinations.

Typically, the variables are "pre-whitened" (i.e. filtered) to purge them of possible spurious correlations before cross correlations are computed.

**Prewhitening**

Part of the second step of transfer function analysis.

Prewhitening involves taking the noise models of our dependent variable *and our independent variable* and seeing the correlations between them at various lags and leads.

The residuals of the two noise models are compared in a cross correlation function.

The CCF shows the correlations of 2 sets of residuals at various lags and leads.

We can see if some type of process is driving variation in both X and Y.

For example: if we find that the noise model of the independent variable is:

$$X_t = \frac{\theta(L)}{\phi(L)} a_t$$

And the noise model of the dependent variable is:

$$Y_t = \frac{\theta^*(L)}{\phi^*(L)} b_t$$

We get two sets of residuals.

Residuals for $X_t$: $$\hat{a}_t = \frac{\phi(L)}{\theta(L)} X_t$$

Then, we use the noise model of $X_t$ to filter the dependent variable and give another set of residuals.

Residuals for $Y_t$: $$\hat{b}_t = \frac{\phi^*(L)}{\theta^*(L)} Y_t$$

We take the two set of residuals and do a cross-correlation function.

CCF shows the correlations of 2 sets of residuals @ various lags and leads.

We want half of the CCF to show X lagging behind Y and half to show Y lagging behind X.

This helps us specify our transfer function model – variables can affect each other now or with lags.

We may see a spike at one lag or more.

We may also see a declining set of spikes indicating that there is an adjustment process for which we would want to include delta parameters.

Although useful when theory is weak or absent, cross correlations are less important when the analyst has theoretical reasons and hypotheses about when effects should occur.

Since cross correlations are bivariate measures of strength of relationship, they always should be viewed as guides in the model specification process.

That is, cross correlations don't take account of the full model so we should be a little wary about seeing direct effects in them.

This is especially true if we expect a relationship and don't see one in the CCF.

We should estimate at least a contemporaneous effect to see if it exists.

NOTE: Difference the non-stationary variables before beginning this process.

Again, we will want to have white noise residuals.

One further diagnostic tool we can use is to look at a CCF between the filtered independent variable and the residuals of the transfer function model.

Once we decide what lags to use for our independent variables we can proceed to include them in our model.

An important note here is that the only aspects of the ARIMA model we use for the dependent variable is $d$. We don't care about the $p$ and $q$ for the dependent variable for this. Thus, we are not using the complete ARIMA model developed for our dependent variable here.

Think of these as antecedent variables.

We are only interested in how the variation in the independent variable affects the variance in the noise model of the dependent variable.

To summarize this process (from Macleary and Hay):

1. Preliminary Univariate Analysis. Univariate models are built for the two time series. If modeling indicates that either series is nonstationary, the series must be differenced appropriately.

2. Transfer function identification: The ARIMA model of the independent variable is applied to both series and the CCF between these prewhitened series is used to identify a transfer function component for the model.

3. Noise component identification. Parameters for the transfer function component are estimated. Residuals from this estimation are used to identify an ARIM model for the noise component.

4. Estimation. Parameters for the tentative model are estimated. If the parameters are not statistically significant, new components should be tried.

5. Noise component diagnosis. If residuals of the tentative model are not white noise, a new noise component must be identified.

6. Transfer function diagnosis. If residuals of the tentative model are correlated with the prewhitened independent variable, a new transfer function component must be identified.

7. Interpret the model.

**Granger Causality**

Notion of cause and effect is not available in cross-sectional analysis.

Because we are incorporating the notion of time in our analyses we can measure cause and effect and make firmer conclusions than in cross-sectional analyses.

Two series, X and Y.

One series X *Granger-causes* Y if by incorporating the past value if X we improve our prediction of Y compared to the prediction of Y we get using only the past history of Y.

If X does *not* Granger-cause Y it is *strongly exogenous* to Y.

This is different than a test for weak exogeneity.

For $x_t$ to be weakly exogenous we would require that it not be affected by the contemporaneous value of $y_t$.

But Granger causality refers only to the effects of past value of $x_t$ on the current value of $y_t$ .

Four possibilities (See Freeman 1983).

I. Y causes X: $\sigma^2(X_t \mid \overline{Y}_t, \overline{X}_t) < \sigma^2(X_t \mid \overline{X}_t)$

Here, the minimum predicted variance is smaller for X when past values of Y are included.

II. Y causes X instantaneously: $\sigma^2(X_t \mid \overline{\overline{Y}}_t, \overline{X}_t) < \sigma^2(X_t \mid \overline{Y}_t, \overline{X}_t)$

III. Feedback:

$$\sigma^2(X_t \mid \overline{X}_t, \overline{Y}_t) < \sigma^2(X_t \mid \overline{X}_t)$$
*and*
$$\sigma^2(Y_t \mid \overline{Y}_t, \overline{X}_t) < \sigma^2(Y_t \mid \overline{Y}_t)$$

IV. Independence: X and Y are not causally related:

$$\sigma^2(X_t \mid \overline{X}_t, \overline{\overline{Y}}_t) = \sigma^2(X_t \mid \overline{X}_t, \overline{Y}_t) = \sigma^2(X_t \mid \overline{X}_t)$$
*and*
$$\sigma^2(Y_t \mid \overline{Y}_t, \overline{\overline{X}}_t) = \sigma^2(Y_t \mid \overline{Y}_t, \overline{X}_t) = \sigma^2(Y_t \mid \overline{Y}_t)$$

Where $X_t$ and $Y_t$ are jointly covariance stationary.

$\overline{X}_t$ and $\overline{Y}_t$ denote all past values of X and Y, respectively.

$\overline{\overline{X}}_t$ and $\overline{\overline{Y}}_t$ denote all past *and* present values of X and Y, respectively.

$\sigma^2(X_t \mid Z)$ is the minimum predictive error variance of $X_t$ given Z, where Z is composed of the sets $[\overline{X}_t, \overline{Y}_t, \overline{\overline{X}}_t, \overline{\overline{Y}}_t]$.

So we predict Y in terms of its own history → develop an ARIMA model for it.

A second series, X, Granger Causes Y if prediction of Y improves with use of X over just ARIMA of Y.

Also, it is possible that Y also Granger-causes X.

If X and Y Granger-cause each other, we have Granger-causality running in both directions.

If we find that X Granger-causes Y but that Y does not Granger-cause X we say that:
X is strongly exogenous to Y and
Y is strongly endogenous to X.

Compare transfer function model of Y with the ARIMA model of Y.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + ...\alpha_p Y_{t-p} + \beta_1 X_{t-1} + ...\beta_p X_{t-p} + u_t$$

One test is a block F test for the joint significance of the Xs to see if X Granger-causes Y.

We can use Box-Jenkins methodology to do Granger Causality tests.
Called: Haugh-Pierce test.

**Haugh Pierce Test**
Develop univariate ARIMA models for two variables we are interested in.
Filter each variable in terms of its own ARIMA model.

$\hat{\varepsilon}_t^x = \frac{\phi L}{\vartheta L} X_t$ is the ARIMA model for X giving us a residual series $\hat{\varepsilon}_t^x$.

$\hat{\varepsilon}_t^y = \frac{\phi L}{\vartheta L} Y_t$ is the ARIMA model for Y giving us a residual series $\hat{\varepsilon}_t^y$.

Now we can do a CCF for $\hat{\varepsilon}_t^x$ and $\hat{\varepsilon}_t^y$ to some plausible number of lags.

Once we have the CCF we can do the H.P. test.

We can do it generally – is there any Granger causality going in either direction or we can do it specifically, looking for Granger causality in one direction.

For the general test:

Null hypothesis is that $X_t$ and $Y_t$ are not causally related.

Test statistic S is distributed as $\chi^2$:

$S = N * \sum_{k=-m}^{m} r^2(k)$ -m goes from –m to -1. m goes from 1 to m.

That is, we do not include the correlation for m = 0.

Degrees of Freedom = 2m + 1.

This is another "Portmanteau" test.

We take the correlations from the CCF, square them and add up the squares to calculate S.

If we want to check for specific Granger-causality we look at half of the CCF.

$$S_{X \succ Y} = N * \sum_{k=1}^{m} r^2(k) \quad \text{df = m + 1}$$

Granger Causality tests have some problems.

1. They are based on bivariate correlations.
Significant relationships may appear in a well specified multivariate model that do not appear when looked at bivariately.

2. We ignore simultaneous correlations. But we need to think about the level of aggregation. There may indeed be Granger-causality that we can't test because of the level of aggregation.
For example, if X Granger-causes Y but only very quickly and both are measured monthly, we may not see the causality unless we were able to dis-aggregate and re-aggregate for weekly data.

3. Sometimes one of the variables can be static over time. Granger-causality over some periods is washed out over the periods where there is no effect.

For example, employment in the 1970s was very stable. Trying to find causal variables for it would be difficult.

For RATS examples see Conservative Party Transfer Function 1979-1992 and RATS User's Guide 197-198.

**Time Varying Parameter**

There are many points in the semester where this could be relevant.

Are our estimates of the relationships constant over time.

This can be an interesting revision of any major theory or empirical work.

Perhaps the way in which X affects Y has changed since some event occurs.

Or perhaps there are cyclical effects – X affects Y differently at different points in time.

Many ways to study time varying parameters.

Within the Box-Jenkins framework or regression framework we can estimate parameters using moving windows.

For example, use the first 30 data points to estimate coefficients, then use data points 2-31 and re-estimate, then 3-32 and so on.

Then we can see the dynamic properties of the coefficients.

Also, we'll see some of this in the readings on Rolling Cross-Sectional design.

So, its definitely an interesting question to ask – are our impact parameters stable through time?

One test of this is the Chow test – actually, there are many versions of this test.

Applicable in either regression analysis of time series or Box-Jenkins.

If we have some theoretical reason for suspecting parameter instability, divide the series in half and compare halves.

Steps:

1. Analyze all the data for the whole time period and calculate the residual sum of squares. Call this $S_1$.

2. Split data set into 2 parts – split may be in middle or depend on theoretical point in series – and call the halves $N_1$ and $N_2$.

Now run our model for each of $N_1$ and $N_2$.

Get the RSS for each and call these $S_2$ and $S_3$.

Add to $S_2$ and $S_3$ to get $S_4$.

Subtract $S_4$ from $S_1$: $S_1 – S_4 = S_5$.

If there is parameter instability at time point selected $S_5$ should be +ve.

$S_5$ tells us the addition RSS we have explained by dividing the sample.

Residuals for 2 halves add up to less because we can fit an explanation to the two halves much closer.

$S_5$ should never be negative.

We can do an F-test on $S_5$.

$$F = \frac{S_5 / k}{S_4 / (N_1 + N_2 - 2k)} \qquad \begin{aligned} df_1 &= k \\ df_2 &= N_1 + N_2 - 2*k \end{aligned}$$

Where *k* is the number of parameters in the model.

The null hypothesis is that there is no structural break.

If F > critical value, then reject the null.

If it makes little difference then $S_5$ will be small.

Problems with Chow Tests:

If the series are very short and k is very big the series can't be broken in half.

Also, it assumes we have the same variables on both sides of the break point.

If there is an important intervention that explains a lot of variation, it can only be in one of the halves. Thus, RSS of one half may be quite better than the other and lead to false rejection of the null.