Time Series Models

1 Introduction

Let's start with our basic model:[1]

$$y_t = x_t\beta + \epsilon_t, \quad t = 1, 2, \ldots, T \tag{1}$$

On the whole, our sample sizes tend to be smaller in time series models than some of the other models that we have looked at. It is important to remember that we still need to worry about all the things that we normally worry about. However, we now need to worry about whether our disturbances are serially correlated.[2] We say that our errors are serially correlated if ϵ_t is not independent of ϵ_t' i.e. $E[\epsilon_i, \epsilon_j] \neq 0, i \neq j$. In practice, we consider only certain ways in which our disturbances are correlated. Specifically, we consider (i) spatial autocorrelation - correlation across cross-sectional units, and (ii) serial correlation across time. We will only consider time series issues here.[3]

1.1 Old-Fashioned View

The old fashioned view is that serial correlation is a nuisance that causes the Gauss-Markov assumptions to be violated and leads to the non-optimality of OLS. Basically, if you employ OLS and your disturbances show evidence of serial correlation, then

- OLS standard errors are wrong, possibly very wrong
- OLS estimates are no longer efficient
- OLS estimates are still unbiased and consistent though.

The bottom line is that the OLS estimator is no longer BLUE. As a result, the usual t, F, and χ^2 statistics may be wrong. It turns out that serial correlation is not 'knife-edged'. In other words, small amounts of serial correlation cause only small problems, whereas a large amount of serial correlation can cause big problems.

1.2 Modern View

The more modern view is to model the correlation or dynamics as part of the model rather than treat it as an estimation nuisance.[4] Let's take a quick look at how this would work. Consider the following simple model.

$$y_t = x_t\beta + \epsilon_t \tag{2}$$

$$\epsilon_t = \rho\epsilon_{t-1} + \nu_t \qquad \text{where } -1 < \rho < 1 \tag{3}$$

[1] Notes are based on Beck and Zorn.

[2] We will use the terms 'serial correlation' and 'autocorrelation' interchangeably. This is the most common practice. However, some authors prefer to distinguish the two terms, with autocorrelation referring to correlation between the disturbances from the same series and serial correlation referring to correlation between the disturbances from two different series.

[3] For a useful introduction to spatial autocorrelation, see Beck, Gleditsch and Beardsley (2006).

[4] This is often referred to as the London School of Economics approach.

where ν_t is iid (independent and identically distributed) i.e. $E[\nu_t, \nu_{t-1}] = 0$.[5] As we'll learn in a moment, the disturbances or errors in this model are said to follow a first order autoregressive (AR1) process. Thus, the current error is part of the previous error plus some shock. We don't usually think of the second equation as being part of the model. We can obviously rewrite these two equations as

$$y_t = x_t\beta + \rho\epsilon_{t-1} + \nu_t \tag{4}$$

We know that

$$y_{t-1} = x_{t-1}\beta + \epsilon_{t-1} \tag{5}$$

and

$$\epsilon_{t-1} = y_{t-1} - x_{t-1}\beta \tag{6}$$

We can now substitute Eq. (6) into Eq. (4) to get the following model:

$$\begin{aligned} y_t &= x_t\beta + \rho(y_{t-1} - x_{t-1}\beta) + \nu_t \\ &= x_t\beta + \rho y_{t-1} - \rho x_{t-1}\beta + \nu_t \end{aligned} \tag{7}$$

The model in Eq. (6) looks like something that we'll see a little later known as the 'autoregressive distributed lag' (ADL) model:

$$y_t = x_t\beta + \rho y_{t-1} + \gamma x_{t-1}\beta + \nu_t \tag{8}$$

This model is 'autoregressive' because it includes a lagged y and it is a 'distributed lag' model because it includes a lagged x. The point to note here is that the model starts to look much more like a standard linear problem where you can use OLS. The modern approach, then, is to model the dynamics directly rather than treat autocorrelation as a nuisance.

2 Testing for Serial Correlation

2.1 Error Processes

When we test for serial correlation, we typically assume something about the error process. The usual assumption is either that the errors follow a first-order autoregressive process (AR1) or a first order moving average process (MA1).

2.1.1 AR1 Error Process

$$\text{AR1:} \quad \epsilon_t = \rho\epsilon_{t-1} + \nu_t \tag{9}$$

[5] iid errors are often described as 'white noise'.

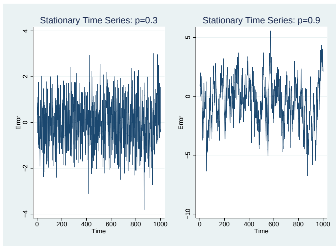
where ν_t is iid. As I noted earlier, the current error is part of the previous error plus some shock. It is a first order process because ϵ_t only depends on its immediate lagged value. It is relatively easy to see that this is a model of exponential decay. For example,

$$\begin{aligned}
\epsilon_t &= \rho\epsilon_{t-1} + \nu_t \\
&= \rho(\rho\epsilon_{t-2} + \nu_{t-1}) + \nu_t \\
&= \rho^2\epsilon_{t-2} + \rho\nu_{t-1} + \nu_t \\
&= \rho^2(\rho\epsilon_{t-3} + \nu_{t-2}) + \rho\nu_{t-1} + \nu_t \\
&= \rho^3\epsilon_{t-3} + \rho^2\nu_{t-2} + \rho\nu_{t-1} + \nu_t \ldots
\end{aligned} \tag{10}$$

It is easy to see that the current error is just the error of all the previous shocks weighted by some coefficient (ρ) that is exponentially declining. How fast the effect of these previous errors dies out depends on the value of ρ.

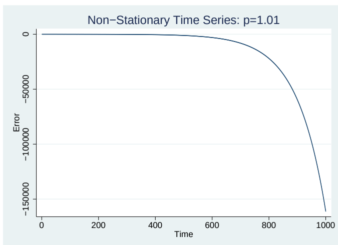
- When $\rho < 1$, then we have what is called **stationary** time series. In this setup, the time series looks jagged, it has structure, and it never wanders too far from the mean. The effect of the errors decay and disappear (well, not quite) over time. Things that happened recently are relatively more important than things that happened a long time ago.

Figure 1: Stationary Time Series with $\rho = 0.3$ and $\rho = 0.9$



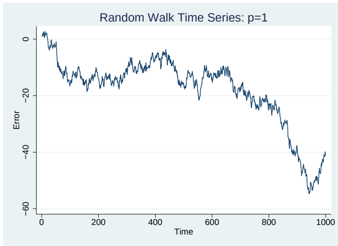
- When $\rho > 1$, then we have what is called **non-stationary** time series. In this setup, you get a smoother time series that eventually explodes. Things that occurred a long time ago have a large impact compared to things that occurred more recently.

Figure 2: Non-Stationary Time Series with $\rho = 1.01$



- When $\rho = 1$, then we have a **random walk** i.e. $y_t = y_{t-1} + \epsilon_t$. In this setup, the time series moves up and down, but slowly. Things that happened in any period have an equal impact.

Figure 3: Random Walk Time Series with $\rho = 1$



We'll discuss stationarity and non-stationarity in more detail next week. Suffice to say that a *small* change in ρ from 0.9 say to 1 or to 1.01 can lead to *huge* changes. As you'll already have noticed, things get weird in time series when $\rho \simeq 1$.

2.1.2 MA1 Error Process

$$\text{MA1:} \quad \epsilon_t = \mu\nu_{t-1} + \nu_t \tag{11}$$

where the ν_ts are iid. Thus, the current error is really part of the error from the previous period plus the error from this period. An MA process is just a linear combination of white noise or iid error terms. In contrast to the AR1 error process, the effect of the MA1 error process is short-lived and finite - it is affected by the current and prior values of ν only.

2.1.3 More Complicated Error Processes

So far, we have only looked at first order processes. It is quite easy to think of second (and higher) order processes. Below are second order autoregressive and moving average processes.

$$\text{AR2:} \quad \epsilon_t = \rho_1\epsilon_{t-1} + \rho_2\epsilon_{t-2} + \nu_t \tag{12}$$

$$\text{MA2:} \quad \epsilon_t = \mu_1\nu_{t-1} + \mu_2\nu_{t-2} + \nu_t \tag{13}$$

It is also possible to think of processes that combine autoregressive (AR) and moving average (MA) processes. These are called ARMA processes. Depending on the order of these processes, we can talk of ARMA(p, q) processes. For example, an ARMA(1,1) process would be

$$\epsilon_t = \rho\epsilon_{t-1} + \mu\nu_{t-1} + \nu_t \tag{14}$$

2.2 Testing for Serial Correlation

There are a variety of tests for serial correlation. We are going to focus on two: (i) Durbin-Watson d-test and (ii) Breusch-Godfrey or Lagrange Multiplier (LM) test.

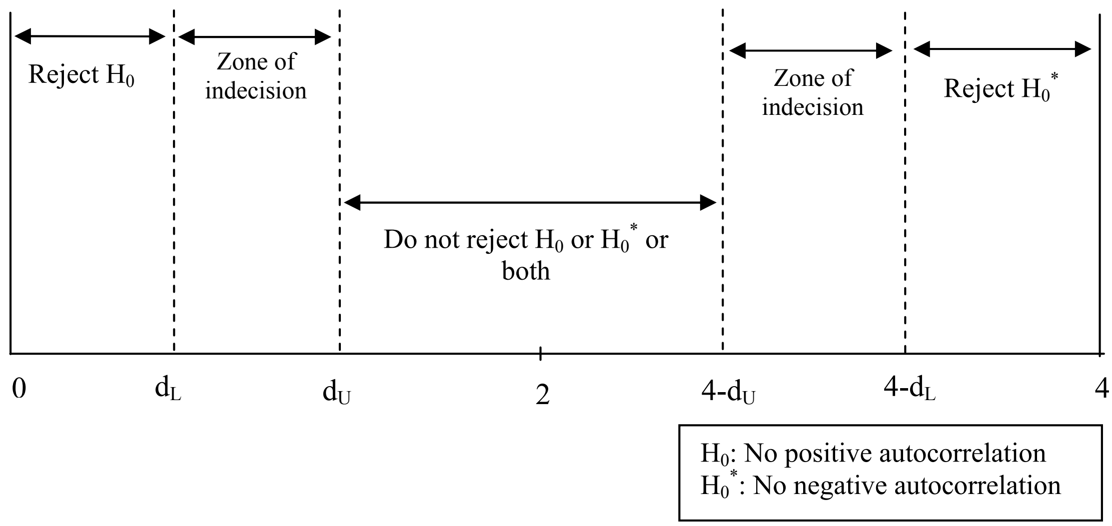
2.2.1 Durbin-Watson d-Test

The Durbin-Watson statistic is

$$d = \frac{\sum_{t=2}^{t=n}(\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{\epsilon}_t^2} \tag{15}$$

where $\hat{\epsilon}$ are the estimated residuals. It turns out that $d \approx 2(1-\hat{\rho})$. Thus, if $\hat{\rho} = 0$ i.e. no serial correlation, then d=2. If $\hat{\rho} = 1$ i.e. perfect positive correlation, then $d \approx 0$. If $\hat{\rho} = -1$ i.e. perfect negative correlation, then $d \approx 4$. Thus, d ranges from 0 to 4 with no serial correlation (the null hypothesis) being $d = 2$. Durbin and Watson were able to derive a lower bound d_L and an upper bound d_U such that if the computed d lies outside these critical values, then a decision can be made regarding the presence of positive or negative serial correlation. The precise values for d_L and d_U, which depend on the number of independent variables, can be found at the back of most textbooks such as Gujarati (2003). One drawback of the Durbin-Watson (DW) test is that there are zones of indecision - zones in which it is not possible to conclude that (first order) serial correlation does or does not exist. The decision rules to be used with the DW statistic are shown in Figure 4.

Figure 4: Durbin-Watson d Test: Decision Rules



To use the DW test, we have to assume that (i) the regression model contains an intercept, (ii) the error process is AR1 (the test says nothing about higher order autocorrelation), (iii) the error term is normally distributed, (iv) there is no lagged dependent variable,[6] and (v) there are no missing observations in the data. There are a number of different ways to get the DW statistic in STATA. One way is to type

[6]Durbin did propose an h test for models that include a lagged dependent variable. However, this test is not as powerful as the Breusch-Godfrey test we are about to see. You can find out more about the Durbin h test in Gujarati (2003, 503) or Johnston and Dinardo (1997, 182-183).

- regress Y X1 X2 etc.
- estat dwatson

Alternatively, you could type the following and this would report the DW statistic.

- prais Y X1 X2 etc.

2.2.2 Breusch-Godfrey or Lagrange Multiplier Test

The second test for serial correlation is the Breusch-Godfrey (BG) or Lagrange Multiplier (LM) Test. The BG test is useful in that it allows for (i) lagged dependent variables, (ii) higher order autoregressive processes as well as single or higher order moving average processes. The basic idea is to regress the residuals from the OLS regression on all of the independent variables and on the lagged values of the residuals. The steps for conducting this test are the following:

1. Estimate your model by OLS and obtain the residuals, $\hat{\epsilon}_t$.
2. Regress the residuals, $\hat{\epsilon}_t$, on all of the independent variables included in the original model and as many lags of the residuals as you think appropriate. For example, if you think that there is an AR(p) process going on, then you want to include $\hat{\epsilon}_{t-1}$, $\hat{\epsilon}_{t-2}$, $\ldots$, $\hat{\epsilon}_{t-p}$.
3. If the sample size is large, then the R^2 from this auxiliary regression will be distributed in the following manner:

$$(n-p)R^2 \sim \chi^2_p \tag{16}$$

   where p (the degrees of freedom) is the order of the error process or the number of included lags. If $(n-p)R^2$ exceeds the critical χ^2_p at the chosen level of significance, then we reject the null hypothesis of no serial correlation.

A drawback of the BG test is that the value of p i.e. the number of lags cannot be determined a priori. You should probably experiment with this to some extent. To estimate the BG test in STATA for a second order process we would type:

- regress Y X1 X2 etc.
- estat bgodfrey, lags(1/2)

2.2.3 Overview of Where Things Stand

If our tests find no serial correlation, then we are free to use OLS. The only issue becomes interpretation and deciding how to model any lags. If our tests find serial correlation AND we have no lagged dependent variables, then OLS is consistent but inefficient and the standard errors are wrong. As we'll see, we can solve this problem by employing generalized least squares (GLS-Prais Winsten) or the Cochrane-Orcutt procedure. If our tests find serial correlation AND we have a lagged dependent variable, then we have problems - we'll take a look at this in a while.

3 Generalized Least Squares

Let's imagine we are in a world where we have **serial correlation and no lagged dependent variable**. We'll assume that we have an AR1 error process so that we have:

$$Y_t = \beta_0 X_t + \epsilon_t \tag{17}$$

where

$$\epsilon_t = \nu_t + \rho\epsilon_{t-1} \tag{18}$$

If we knew ρ, we could use Generalized Least Squares. For example, if there were an AR1 error process then

$$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + \epsilon_{t-1} \tag{19}$$

If we multiply both sides by ρ, we have

$$\rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{t-1} + \rho\epsilon_{t-1} \tag{20}$$

Now if we subtract (116) from (121), then we have

$$(Y_t - \rho Y_{t-1}) = \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + (\epsilon_t - \rho\epsilon_{t-1}) \tag{21}$$

Since $\epsilon_t - \rho\epsilon_{t-1} = \nu_t$ from (18), we have

$$(Y_t - \rho Y_{t-1}) = \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + \nu_t \tag{22}$$

This can be expressed as

$$Y^* = \beta_0^* + \beta_1 X_t^* + \nu_t \tag{23}$$

Since ν_t is iid, we can now use OLS on our transformed variables.

Of course, we will not know ρ. As a result, we have to use our estimate of ρ from the OLS regression. This is referred to as Feasible Generalized Least Squares (FGLS). There are two procedures for doing this - Cochrane-Orcutt (CO) and Prais-Winsten (PW).[7] As you'll see, they are essentially identical except for how they treat the first observation.

3.1 Cochrane-Orcutt

1. Estimate the model by OLS and get residuals.[8]
2. Regress residuals on lagged residuals i.e. $\hat{\epsilon}_t = \hat{\epsilon}_{t-1} + \nu_t$. The coefficient on the lagged residual is your estimated $\hat{\rho}$.
3. Use $\hat{\rho}$ to generate $Y_t^* = Y_t - \hat{\rho}Y_{t-1}$ and $X_t^* = X_t - \hat{\rho}X_{t-1}$. Note that this means that we lose the first observation.

[7]Actually, only the Prais-Winsten method is really full FGLS.

[8]For the CO and PW procedures to work, it must be the case that OLS produces consistent estimates of β. As we'll see, this is why including a lagged dependent variable causes problems - it leads to inconsistent estimates of β since the lagged dependent variable is correlated with the error term if there is serial correlation.

4. Now regress Y_t^* on X_t^*. The constant in this regression will be $\beta_0(1-\hat{\rho})$. Generate new residuals from this regression.

5. Regress new residuals on lagged new residuals to get a new $\hat{\rho}$.

6. Go to step 3 and repeat until we get convergence.[9]

As I noted above, the Cochrane-Orcutt procedure means that we lose the first observation.[10] To estimate the CO procedure in STATA, type:

- prais y x_1 x_2 $\ldots$, corc

3.2 Prais-Winsten

The Prais-Winsten procedure is essentially the same except that you keep the first observation and it does not iterate. To keep the first observation, you have to use the following transformation: $\hat{Y_0} = Y_0(\sqrt{1-\hat{\rho}^2})$ and $\hat{X}_0 = X_0(\sqrt{1-\hat{\rho}^2})$. The CO and the PW procedures are asymptotically identical.[11] To estimate the PW procedure in STATA, type:

- prais y x_1 x_2 $\ldots$

While the PW procedure is obviously better in small samples where losing an observation can have a big effect on the results, it has become the procedure of choice for samples of all sizes. As far as I can tell, the implementation of these procedures in STATA requires assuming an AR1 error process.

3.3 Serial Correlation and a Lagged Dependent Variable

So far, we have looked at what happens when we do not have a lagged dependent variable. What happens if there is a lagged dependent variable? *If there is no serial correlation*, the inclusion of a lagged dependent variable will mean that OLS is biased but consistent (Keele & Kelly 2006). However, *if there is serial correlation*, then the inclusion of a lagged dependent variable model will mean that OLS is both biased and inconsistent. Note that you **cannot** solve this problem with the GLS procedure. Recall that for the GLS procedure to work, the first round OLS estimates must be consistent. However, this will not be the case if we include a lagged dependent variable since the lagged dependent variable will be correlated with the error term when there is serial correlation. Thus, we cannot use GLS to solve the problem of serial correlation when we have a lagged dependent variable.

Fortunately, in practice, models that do include a lagged dependent variable do not often show evidence of serial correlation - thus, the problem should be relatively rare. People have made a big deal about lagged

[9]Asymptotic theory actually says that the first round results are as good as the results from iterating. However, people still iterate.

[10]There are also some issues with dummy variables. See Gujarati (2003, 487-488).

[11]An alternative to these GLS based procedures is to recognize that the estimates from OLS are unbiased and consistent; it is just the standard errors that are wrong. We could simply correct these by using heteroskedasticity- and autocorrelation-consistent (HAC) standard errors - these are an extension of White's heteroskedasticty-consistent standard errors. These are often referred to as Newey-West standard errors. As Gujarati (2003, 485) notes, FGLS and HAC procedures are more efficient than OLS but they are asymptotic. It may be the case that OLS is better in really small samples. To estimate OLS with Newey-West standard errors type

- newey y x_1 x_2 $\ldots$, lag(#)

where the lag number is the order of the autoregressive process.

dependent variables in the context of serial correlation (Achen 2000). The reason is that may people seem to think that disturbances are God-given i.e. that if God gives a process serial correlation, then it always has serial correlation. However, in reality, adding a lagged dependent variable to a model that has serial correlation does not necessarily mean that this new model also has serial correlation. In many (most?) cases, the inclusion of a lagged dependent variable will clean up the serial correlation. A problem only arises if there is still some serial correlation left over once you've included the lagged dependent variable. It turns out that tests for serial correlation in models that include a lagged dependent variable do generally indicate whether there is any serial correlation left over (Keele & Kelly 2006). This is good since it means that we will know if we have a problem simply by testing for serial correlation - you should always conduct a BG test.

If you have serial correlation left over, then you might consider adding additional lags of the dependent variable or lags of the independent variables to clean it up. The other solution is to use instrumental variables. However, as Beck notes, this cure is nearly almost certainly worse that the disease because our instruments are usually pretty bad. Finally, we should note that OLS with a lagged dependent variable seems to do quite well so long as any remaining serial correlation is not too high (Keele & Kelly 2006).

3.4 Old Fashioned World: Conclusion

What we have looked at so far is really the old fashioned view of the world as it comes to serial correlation in stationary time series – serial correlation causes a technical violation of an OLS assumption, leading to incorrect estimates of the standard errors. We then tried to find ways to deal with this 'problem'. Here is an overview of what we found:

1. **Without lagged dependent variables**
   - No serial correlation i.e. $\rho = 0$: OLS is just fine.
   - With serial correlation: OLS leads to unbiased and consistent estimates of β. This is good. However, the standard errors are wrong. Thus, the OLS estimates are not efficient. A small amount of serial correlation only causes a small problem, but a large amount leads to large problems.
   - If there is serial correlation, you could use GLS (Prais-Winsten). Alternatively, you could decide to simply use Newey-West standard errors.
   - We need to be mindful that GLS and OLS with Newey-White standard errors can be quite biased if the correct model should include a lagged dependent variable (2006).
2. **With lagged dependent variables**
   - No serial correlation: OLS will be biased but consistent.
   - With serial correlation: OLS estimates are not consistent and GLS fails because lagged dependent variable is correlated with the error term. Fortunately, models with lagged dependent variables do not often show serial correlation.
   - OLS with lagged dependent variables still does relatively well when serial correlation is low.
   - If serial correlation remains after the inclusion of a lagged dependent variable, one solution would be to add additional lagged terms.
   - An alternative solution would be to use instrumental variables - but ...

4 An Aside: ARCH/GARCH Models

Before we get to the more modern approach of dealing with serial correlation, I want to have a brief aside on ARCH/GARCH models. Traditionally, we have been alert to the possibility of heteroskedasticity in cross-sectional data and to autocorrelation in time series data. However, Engle (1982) suggested that heteroskedasticity might also be a problem in time series contexts – he had noticed that large and small errors tended to occur in clusters in speculative financial markets such as exchange rates and stock market returns. Engle proposed the 'autoregressive conditional heteroskedasticity (ARCH)' model to look at heteroskedasticity in time series data. In other words, ARCH models help us when we are interested in the mean *and* variance of time series. The ARCH model has been expanded in a number of different ways and has been gradually adopted by some political scientists (Maestas & Preuhs 2000, Leblang & Bernhard 2006, Bernhard & Leblang 2006). These models can be useful if we are interested in understanding why a time series is more or less volatile.

4.1 ARCH Models

Consider the following model.

$$Y_t = \rho Y_{t-1} + X\beta + \epsilon_t \tag{24}$$

We typically treat the variance of $\epsilon_t = \sigma^2$ as a constant. However, we might think to allow the variance of the disturbance to change over time i.e. the conditional disturbance variance would be σ_t^2. Engle postulated that the conditional disturbance variance should be modeled as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2 \tag{25}$$

The lagged ϵ^2 terms are called ARCH terms and you can see why this is an 'autoregressive' or AR process. The Eq. (25) specifies an ARCH model of order p i.e. an ARCH(p) model. The conditional disturbance variance is the variance of ϵ_t, conditional on information available at time t-1 i.e.

$$\begin{aligned} \sigma_t^2 &= \text{var}(\epsilon_t | \epsilon_{t-1}, \ldots, \epsilon_{t-p}) \\ &= E(\epsilon_t^2 | \epsilon_{t-1}, \ldots, \epsilon_{t-p}) \\ &= E_{t-1}(\epsilon_t^2) \end{aligned} \tag{26}$$

where E_{t-1} indicates taking an expectation conditional on all information up to the end of period t-1. It is now easy to see that recent disturbances influence the variance of the current disturbance - the ARCH terms can be interpreted as news about volatility (or volatility shocks) from prior periods. A conditional disturbance variance like that shown in Eq. (25) can be obtained if the disturbance is defined as

$$\epsilon_t = \nu_t \sqrt{\sigma_t^2} \tag{27}$$

where ν_t is distributed as a standard normal (mean-zero, variance-one) white noise process and σ_t^2 is the conditional disturbance variance shown above.

This model is often simply written as

$$\begin{aligned} Y_t &= \rho Y_{t-1} + X\beta + \epsilon_t \\ \epsilon_t &\sim (0, \sigma_t^2) \end{aligned} \tag{28}$$

where

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \ldots + \alpha_p\epsilon_{t-p}^2 \tag{29}$$

We can test for ARCH effects fairly simply.

1. Run a regression of Y on X. Obtain the residuals $\hat{\epsilon}_t$.
2. Compute the OLS regression: $\hat{\epsilon}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1\hat{\epsilon}_{t-1}^2 + \ldots + \hat{\alpha}_p\hat{\epsilon}_{t-p}^2$
3. Test the join significance of $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ etc.

4.1.1 ARCH(1) Models

The simplest model is an ARCH(1) model. In other words, the conditional disturbance variance i.e. $\text{var}(\epsilon_t|\epsilon_{t-1})$ is

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 \tag{30}$$

and hence our model is

$$Y_t = \rho Y_{t-1} + X\beta + \nu_t\sqrt{\alpha_0 + \alpha_1\epsilon_{t-1}^2} \tag{31}$$

It follows that

$$\begin{aligned} E[\epsilon_t] &= E\left[\nu_t\sqrt{\alpha_0 + \alpha_1\epsilon_{t-1}^2}\right] \\ &= E[\nu_t]E\left[\sqrt{\alpha_0 + \alpha_1\epsilon_{t-1}^2}\right] = 0 \end{aligned} \tag{32}$$

since $E[\nu_t] = 0$. It also follows that $E[Y_t] = X\beta$. As a result, it is easy to see that this setup is a classical regression model.

While the unconditional disturbance variance (long-run variance) is constant i.e.[12]

$$\text{var}(\epsilon_t) = \frac{\alpha_0}{1-\alpha_1} \tag{33}$$

we already know that the conditional disturbance variance (short-run variance) varies over time i.e.

$$\text{var}(\epsilon_t|\epsilon_{t-1}) = \alpha_0 + \alpha_1\epsilon_{t-1}^2 = \sigma_t^2 \tag{34}$$

In other words, the disturbance ϵ_t is conditionally heteroskedastic with respect to ϵ_{t-1}.

The ARCH(1) model gets us the following features

- The short-run (conditional) variance (volatility) of the series is a function of the immediate past values of the (square of the) error term.
- This means that the effect of each new shock ϵ_t depends, in part, on the size of the shock in the previous period: A large shock in period t, increases the effect on Y of shocks in periods t+1, t+2 etc.
- This means that large shocks cluster together and the series goes through periods of large volatility and periods of low volatility.

[12]To see where this comes from, see Greene (2003, 238-239). This equation requires that we impose the constraints that $\alpha_0 > 0$ and that $0 < \alpha_1 < 1$ in order to keep the variance of the ϵ_ts positive (and stationary).

4.2 GARCH Models

As I noted above, it is possible to model higher order ARCH models. However, it turns out that these models are difficult to estimate since they often produce negative estimates of the αs. To solve this problem, people have turned to the GARCH model (Bollerslev 1986). Essentially, the GARCH model turns the AR process of the ARCH model into an ARMA process by adding in a moving average process. In the GARCH model, the conditional disturbance variance is now

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2 + \gamma_1 \sigma_{t-1}^2 + \gamma_2 \sigma_{t-2}^2 + \ldots + \gamma_q \sigma_{t-q}^2 \\ &= \alpha_0 + \sum_{j=1}^{p} \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^{q} \gamma_k \sigma_{t-k}^2 \end{aligned} \tag{35}$$

It is now easy to see that the value of the conditional disturbance variance depends on both the past values of the shocks and on the past values of itself. The simplest GARCH model is the GARCH(1,1) model i.e.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 \tag{36}$$

Successive substitution into the RHS of Eq. (36) leads to

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 \\ &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 (\alpha_0 + \alpha_1 \epsilon_{t-2}^2 + \gamma_1 \sigma_{t-2}^2) \\ &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \alpha_0 + \gamma_1 \alpha_1 \epsilon_{t-2}^2 + \gamma_1^2 \sigma_{t-2}^2) \\ &\vdots \\ &= \frac{\alpha_0}{1-\gamma_1} + \alpha_1 (\epsilon_{t-1}^2 + \gamma_1 \epsilon_{t-2}^2 + \gamma_1^2 \epsilon_{t-3}^2 \ldots) \end{aligned} \tag{37}$$

Thus, the current variance can be seen to depend on all previous squared disturbances; however, the effect of these disturbances declines exponentially over time. As in the ARCH model, we need to impose some parameter restrictions to ensure that the series is variance-stationary: in the GARCH(1,1) case, we require $\alpha_0 > 0$, $\alpha_1, \gamma_1 \geq 0$, and $\alpha_1 + \gamma_1 < 1$.

The nice thing about both the ARCH and GARCH setups is that they allow the conditional variance to be influenced by exogenous variables i.e. independent variables that we might be interested in. For example, we might be interested in how political events affect exchange rate volatility (Leblang & Bernhard 2006). If these exogenous variables are I_t, then the conditional variance in a GARCH(1,1) model is

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 + \delta I_t \tag{38}$$

This allows us to look at how independent variables affect the volatility of time series data.

4.3 Extensions

There are many, many extensions to these basic models. Two are shown below:

- **ARCH-in-mean (ARCH-M)**

  Basically, this model allows the ARCH effects to appear in the mean of Y as well as its variance. Thus, the model would be

$$Y_t = \beta X_t + \delta \sigma_t^2 + \epsilon_t \tag{39}$$

This sort of model might be appropriate where, say, returns to investment depend on risk as reflected in volatility.

- **Exponential ARCH/GARCH (E-(G)ARCH)**

  The ARCH/GARCH models treat errors as symmetric i.e. positive and negative shocks affect the conditional variance in the same way. However, you might think that actors respond to news asymmetrically i.e. bad news (negative shocks) might lead to greater volatility than good news (positive shocks). The E-(G)ARCH model allows shocks to have an asymmetric effect on the conditional variance.

There are a bunch of other models i.e. I-GARCH and FI-GARCH (Leblang & Bernhard 2006) and others.

5 Modern View of Serial Correlation

Rather than see serial correlation as a technical violation of an OLS assumption, the modern view is to think of time series data in the context of political dynamics. Instead of mechanistically worrying about the residuals, we try to develop theories and use specifications that capture the dynamic processes in question. From this perspective, we view serial correlation as a potential sign of improper theoretical specification rather than a technical violation of an OLS assumption. This view of serial correlation leads us to look at 'dynamic' regression models where 'dynamic' refers to the inclusion of lagged (dependent and independent) variables.

5.1 Lag Operator

Before we start looking at the various dynamic models, I want to introduce the lag operator. This will help us manipulate lagged variables. Some general rules are:

$$\begin{aligned} Ly_t &= y_{t-1} \\ L^2 y_t &= y_{t-2} \end{aligned} \tag{40}$$

We can thus write lag structures in terms of lag polynomials i.e.

$$(1 + \beta_1 L + \beta_2 L^2 + \ldots)y_t = y_t + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots \tag{41}$$

We could transform the following model

$$y_t = \rho y_{t-1} + \beta x_t + \delta x_{t-1} + \epsilon_t \tag{42}$$

in lag polynomial terms as

$$(1 - \rho L)y_t = (\beta + \delta L)x_t + \epsilon_t \tag{43}$$

which can be solved as

$$y_t = x_t \left(\frac{\beta + \delta L}{1 - \rho L} \right) + \frac{\epsilon_t}{1 - \rho L} \tag{44}$$

where

$$\frac{1}{1-\rho L} = 1 + \rho L + \rho^2 L^2 + \rho^3 L^3 \ldots \tag{45}$$

With this in hand, we're now going to look at a number of different models. We are going to pay special attention to the (i) **impulse response function** and (ii) **unit response function** associated with these models. As Greene (2003, 560) notes, looking at the impulse and unit response functions in time series models is the counterpart of looking at marginal effects in the cross-sectional setting. Imagine that our models are in equilibrium. An impulse response function is when the independent variable goes up one unit in one period and then back down to zero in the next period. The unit response function is when the independent variable goes up one unit and remains up one unit for all remaining periods.

5.2 Static OLS Model with iid Errors

Let's start by looking at a static OLS model with iid errors. The simple static model with i.i.d errors is

$$Y_t = \beta_0 X_t + \nu_t \tag{46}$$

where ν_t is an i.i.d. or white noise error process.[13] This model cannot be written with lag operators since there are obviously no lagged terms. The impulse and unit response functions for this model are shown in Table 5.2. At time t=0, we are in equilibrium. At time t=1, we can think that there is a shock (X_t goes up by one unit for one period) - the impulse response is just β_0 for one period. In other words, the equilibrium goes up by β_0 for one period and then back down. One can think of an impulse response as an exogenous shock that occurs for one period but disappears i.e. winning the lottery. Alternatively, at time t=1, we can think that there is a shock (X_t goes up by one unit permanently) - the unit response is β_0 for all periods. In other words, the equilibrium goes up permanently by β_0. One can think of a unit response as a level change that occurs for all subsequent periods i.e. a wage increase.

Table 1: **Impulse and Unit Response Function for OLS**

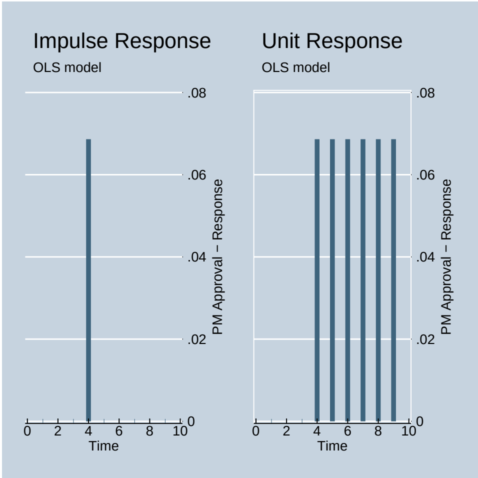
Horizon	Impulse	Response	Unit	Response
0	0	0	0	0
1	1	β_0	1	β_0
2	0	0	1	β_0
3	0	0	1	β_0

Note: The shock occurs in period 1

Imagine that β_0 was 0.07. Figure 5 indicates the impulse and unit response functions. As you can see, the impulse only has an effect of 0.07 (β_1) in the period in which it occurs, and zero effect in subsequent periods. The unit response leads to a level change where the equilibrium increases by 0.07 units and remains 0.07 units higher than it was originally.

[13]Throughout, iid errors will be denoted by ν_t.

Figure 5: Impulse and Unit Response Function for OLS



5.3 Static OLS model with AR(1) error process

Now consider a model with serial correlation but where the errors follow an AR1 process.

$$Y_t = \beta_0 X_t + \epsilon_t \tag{47}$$

where

$$\epsilon_t = \nu_t + \rho\epsilon_{t-1}$$

Recall that scholars often assume that their serial correlation takes this form and this leads them to use the Prais-Winsten or Cochrane-Orcutt procedure (or the Newey-White standard errors). This is the approach taken when we treat serial correlation just as a nuisance. However, we should really think about what this model means in terms of the dynamics of political processes. To see this more clearly, it helps to write the model in lag operator form.[14]Let's start with the error term:

$$\begin{aligned}\epsilon_t &= \nu_t + \rho\epsilon_{t-1}\\ \epsilon_t - \rho\epsilon_{t-1} &= \nu_t\\ (1-\rho L)\epsilon_t &= \nu_t\\ \epsilon_t &= \frac{\nu_t}{1-\rho L}\end{aligned}$$

[14]The static OLS model with an MA(1) error process is:

$$Y_t = \beta_0 X_t + \nu + \mu\nu_{t-1} \tag{48}$$

In lag operator terms this is:

$$Y_t = \beta_0 X_t + (1+\mu L)\nu_t \tag{49}$$

Substituting back into (47), we have:

$$Y_t = \beta_0 X_t + \frac{\nu_t}{1 - \rho L} \tag{50}$$

This is a short-hand for

$$Y_t = \beta_0 X_t + \nu_t + \rho\nu_{t-1} + \rho^2\nu_{t-2} + \rho^3\nu_{t-3}\ldots \tag{51}$$

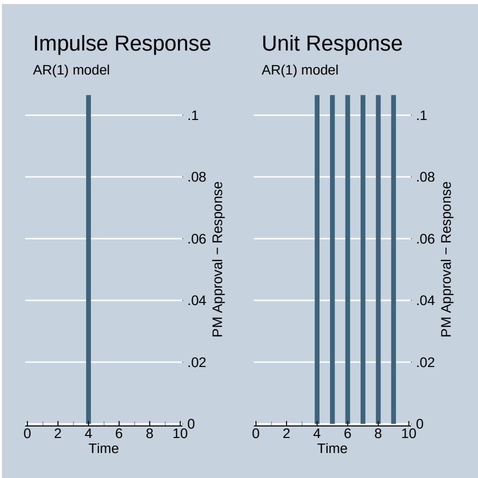
Written in this way, it becomes obvious that shocks in unmeasured variables have a (long-term) impact on Y_t that decline geometrically over time but that the actual measured variables X_t only have an instantaneous one period impact on Y_t. If you opt for this model, you are essentially saying that the effect of unmeasured variables persist (at a declining rate) over time but that the effect of measured variables do not. You have to ask yourself whether this makes sense. This is all easy to see once we examine the impulse and unit response functions. These are illustrated in Table 5.3 and in Figure 6.

Table 2: **Impulse and Unit Response Function for AR(1) error process**

Horizon	Impulse	Response	Unit	Response
0	0	0	0	0
1	1	β_0	1	β_0
2	0	0	1	β_0
3	0	0	1	β_0

Note: The shock occurs in period 1

Figure 6: Impulse and Unit Response Function for OLS with AR1 Error Process



The important thing to note is that this is exactly the same as for the OLS model examined earlier. An impulse increases Y for one period only, whereas a unit impact leads to a level change in the equilibrium. OLS and OLS with AR1 error process do differ, but not in terms of how the measured variables affect the equilibrium level of the dependent variable. Unmeasured variables in the OLS with AR1 error model have an impact that declines geometrically over multiple periods; the unmeasured variables in the normal OLS model with i.i.d. errors only have a one period impact. Thus, the OLS model and the OLS with AR1 error process differ in terms of the impact of the unmeasured variables but not the measured variables.

5.4 Distributed Lag Models

We might not like the way in which measured and unmeasured variables are treated differently in the *static* OLS model with AR1 error process. As a result, we might turn to *dynamic* models - but which one should we choose? There are two types of dynamic models: (i) **distributed lag models** and (ii) **autoregressive models**. Distributed lag models include lagged values of the independent variables, whereas autoregressive models include lagged values of the dependent variable. Let's start by looking at distributed lag models. One issue is whether to use a model with finite or infinite distributed lags.

5.4.1 Finite Distributed Lag Models

A finite distributed lag (FDL) model might be appropriate if we think that the impact of the measured (independent) variables lasts over a finite number of periods (and then stops) but that the impact of unmeasured variables do not. Below is an FDL model where the impact of the measured variables is thought to last two periods. Note that we are assuming iid errors - this would need to be tested with a Durbin-Watson statistic or a BG test.

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \nu_i \tag{52}$$

The impulse and unit response function for the FDL model are shown in Table 5.4.1. In this example, we

Table 3: **Impulse and Unit Response Function for FDL model**

Horizon	Impulse	Response	Unit	Response
0	0	0	0	0
1	1	β_0	1	β_0
2	0	β_1	1	β_0+β_1
3	0	0	1	β_0+β_1
4	0	0	1	β_0+β_1

Note: The shock occurs in period 1

assumed that the appropriate lag length was 2. However, it is rare that we ever actually know the right lag length or have a strong enough theory to inform us about it. This often means that we fall back on choosing in terms of what is most convenient. Other problems may arise with an FDL model. For example, (i) time series are often short and so the inclusion of the lagged variables may eat up a lot of degrees of freedom and (ii) the fact that the lagged Xs are likely to be highly correlated is likely to lead to severe multicollinearity. The FDL model also has the slightly unappealing feature that the finite nature of the lags mean that the effects of the measured variables cutoff sharply i.e. after two periods in the example above.

5.4.2 Infinite Distributed Lag Models

An infinitely distributed lag model is:

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots + \beta_k X_{t-k} + \nu_i \tag{53}$$

We have already noted some problems with such a model - multicollinearity, hard to know the appropriate lag length etc. Two slightly modified models have been proposed to help with these problems.

Koyck Model

The Koyck Model is

$$Y_t = \alpha + \beta X_t + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + \ldots + \beta\lambda^k X_{t-k} + \nu_t \tag{54}$$

with the restriction that $0 < \lambda < 1$. This model assumes that the influence of X on Y remains positive or negative over time and declines geometrically as k increases. We still don't know what the 'correct' lag length is. However, we can rewrite the model in the following way. First, lag both sides of the equation and multiple both sides by λ.

$$\lambda Y_{t-1} = \alpha\lambda + \beta\lambda X_{t-1} + \beta\lambda^2 X_{t-2} + \ldots + \beta\lambda^k X_{t-k} + \lambda\nu_{t-1} \tag{55}$$

Now subtract Eq. 55 from Eq. 54.

$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta X_t + (\nu_t - \lambda\nu_{t-1}) \tag{56}$$

Now add λY_{t-1} to both sides

$$Y_t = \alpha(1-\lambda) + \beta X_t + \lambda Y_{t-1} + \epsilon_t \tag{57}$$

where $\epsilon_t = \nu_t - \lambda\nu_{t-1}$. This transformation converts the Koyck model into a more standard AR1 model. Although we have limited the number of parameters to three α, β, λ, we can still get estimates of all the parameters in Eq. 54.

There are a number of problems with this model, though. The main problem is that we now have a lagged dependent variable with serial correlation - the problem we identified earlier. Thus, the model would produce biased and inconsistent results. If we were to estimate this model, we would have to use instrumental variables. Specifically, we would have to create an instrument for Y_{t-1} which is highly correlated with Y_{t-1} but uncorrelated with ϵ_t. One option might be to substitute X_{t-1} for Y_{t-1}. This model is not used that often in political science because we tend to have bad instruments.

Almon Model

The Almon model is more flexible than the Koyck model in that it allows the effect of X on Y to change over time. We can start with a general finite distributed lag model.

$$Y_t = \alpha + \sum_{i=1}^{k} \beta_i X_{t-i} + \nu_t \tag{58}$$

where i indexes the k lags. We might write:

$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \ldots + \alpha_m i^m \tag{59}$$

so that β_i is now an m^{th} order polynomial. We need to restrict the model such that m<k. We can stick Eq. 59 into Eq. 58 to get

$$\begin{aligned} Y_t &= \alpha + \alpha_0 \sum_{i=1}^{k} (\alpha_0 + \alpha_1 i + \alpha_2 i^2 + \ldots + \alpha_m i^m) X_{t-i} + \nu_t \\ &= \alpha + \alpha_0 \sum_{i=1}^{k} X_{t-i} + \alpha_1 \sum_{i=1}^{k} i X_{t-i} + \ldots + \alpha_m \sum_{i=1}^{k} i^m X_{t-i} + \nu_t \end{aligned} \tag{60}$$

Next, write

$$\begin{aligned} Z_{0t} &= \sum_{i=1}^{k} X_{t-i} \\ Z_{1t} &= \sum_{i=1}^{k} i X_{t-i} \\ Z_{2t} &= \sum_{i=1}^{k} i^2 X_{t-i} \\ &\vdots \\ Z_{mt} &= \sum_{i=1}^{k} i^m X_{t-i} \end{aligned} \tag{61}$$

This allows us to write

$$Y_t = \alpha + \alpha_0 Z_{0t} + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \ldots + \alpha_m Z_{mt} + \nu_t \tag{62}$$

This can be estimated with OLS. This model is nice in that it is very flexible with respect to how lags of X affect Y and it can be estimated with OLS since there's no autoregressive component. One disadvantage is that you have to specify k, the lag length, and the polynomial degree, m. The model can also be sensitive to specification. This model is not canned in STATA. The hardest part of doing this yourself is in creating the Zs. See Gujarati (2003) and notes by Zorn on how to do this.

5.5 Autoregressive Models

Having looked at distributed lag models, we now turn to autoregressive models. Recall that these models are ones that include lag(s) of the dependent variable.

5.5.1 Lagged Dependent Variable Model

The effect of a shock in the FDL model only lasts as many periods as there are lags of that independent variable. We might think that the effect of a shock lasts for many periods and declines at a geometric rate. As I mentioned earlier, it would be difficult to put in multiple lags of the independent variable because this would lead to high levels of multicollinearity etc. However, we can solve this problem by including a lagged dependent variable. The model is:

$$Y_t = \beta_0 X_t + \rho Y_{t-1} + \nu_t \tag{63}$$

In lag operator terms this is:

$$\begin{aligned} \beta_0 X_t + \nu_t &= Y_t - \rho Y_{t-1} \\ &= (1 - \rho L) Y_t \end{aligned} \tag{64}$$

This is solved as:

$$Y_t = \left(\frac{\beta_0}{1 - \rho L} \right) X_t + \frac{\nu_t}{1 - \rho L} \tag{65}$$

It is now easy to see that the impact of both measured and unmeasured variables persists over time but at a geometrically declining rate. Again, note that the lagged dependent variable model outlined here assumes that the errors are iid - you should check this as before. The impulse and unit response functions for the lagged dependent variable model are shown in Table 5.5.1.

Table 4: **Impulse and Unit Response Function for LDV model**

Horizon	Impulse	Response	Unit	Response
0	0	0	0	0
1	1	β_0	1	β_0
2	0	$\rho\beta_0$	1	$\beta_0 + \rho\beta_0$
3	0	$\rho^2\beta_0$	1	$\beta_0 + \rho\beta_0 + \rho^2\beta_0$
4	0	$\rho^3\beta_0$	1	$\beta_0 + \rho\beta_0 + \rho^2\beta_0 + \rho^3\beta_0$

Note: The shock occurs in period 1

The effect of the impulse is to increase Y by β_0 in the first period. The effect of the impulse persists into all future periods but declines at a geometric rate determined by ρ. Eventually, the effect of the impulse (shock) will disappear and we will return to our original equilibrium. If there is a unit response or level change, Y will initially go up by β_0 in period 1, then by $\beta_0 + \rho\beta_0$ in period 2, and then by $\beta_0 + \rho(\beta_0 + \rho\beta_0)$ in period 3 etc. The new equilibrium will be $\frac{\beta_0}{1-\rho}$ higher than the original equilibrium.

5.6 Autoregressive Distributed Lag Model

We have now looked at the two distinct types of dynamic models: (i) distributed lag models and (ii) autoregressive models. It is possible to combine these two models in a single model called the **autoregressive**

distributed lag (ADL) model. An ADL(1,1) model is shown below:[15]

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \rho Y_{t-1} + \nu_t \tag{67}$$

As before, this setup assumes iid errors - as always, you should test this. In lag operator form this is:

$$(1 - \rho L)Y_t = (\beta_0 + \beta_1 L)X_t + \nu_t$$

Solved, this is:

$$Y_t = \left(\frac{\beta_0 + \beta_1 L}{1 - \rho L}\right) X_t + \frac{\nu_t}{1 - \rho L} \tag{68}$$

The impulse and unit response functions from an ADL(1,1) model is given in Table 5.6. In many ways, the ADL model is similar to the FDL model. The effect of an impulse is to increase Y by β_0 in the first period. In the FDL model, the effect of the impulse in the second period is just β_1. However, the second period effect in the ADL model is $\beta_1 + \rho\beta_0$. In other words, the effect of having a lagged dependent variable is to make the effect from the previous period persist. Eventually, the effect of the impulse will disappear and we will return to our original equilibrium as in the FDL model. If we have a unit level change, then we initially increase Y by β_0 in the first period. In the next period, we increase Y by $\beta_0 + \beta_1 + \rho\beta_0$. The FDL model eventually led to a new equilibrium in two periods at $\beta_0 + \beta_1$. The ADL model is different because we reach a new equilibrium that is slightly higher than $\beta_0 + \beta_1$. This is again because the inclusion of a lagged dependent variable means that the effect of the unit level change from the previous period persists and is added to the unit level change in the current period.[16]

[15] Obviously, you can think of a more general ADL mode:

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{j=1}^{n}\sum_{i=0}^{q} \beta_{jp} X_{jt-i} + \nu_t \tag{66}$$

where ν_t is a white noise error process and $|\sum_{i=1}^{p} \alpha_i|$ is stationary. This is an ADL(p,q,n) model where p refers to the number of lags of Y_t, q the number of lags of X_t, and n the number of exogenous variables. Despite this more general model, the ADL(1,1) model has become the work horse for most modern time series analysis (for the reasons about to be explained).

[16] The ADL model can be generalized to allow each independent variable to have different speeds of adjustment. Belwo is a model with different speeds of adjustment on X_t and Z_t.

$$Y_t = \frac{\beta}{(1 - \rho L)} X_t + \frac{\delta}{(1 - rL)} Z_t + \nu_t \tag{69}$$

This can be written as an ADL model in the following way. First, multiply each side by $(1 - \rho L)(1 - rL)$:

$$Y_t(1 - \rho L)(1 - rL) = \beta X_t(1 - rL) + \delta Z_t(1 - \rho L) + \nu_t(1 - \rho L)(1 - rL)$$

$$Y_t(1 - \rho L - rL + \rho r L^2) = \beta X_t - r\beta X_{t-1} + \delta Z_t - \delta\rho Z_{t-1} + \nu_t(1 - (\rho + r)L + \rho r L^2)$$

$$Y_t - (\rho + r)Y_{t-1} + \rho r Y_{t-2} = \beta X_t - r\beta X_{t-1} + \delta Z_t - \delta\rho Z_{t-1} + \nu_t - (\rho + r)\nu_{t-1} + \rho r \nu_{t-2}$$

$$Y_t = \beta X_t - r\beta X_{t-1} + \delta Z_t - \delta\rho Z_{t-1} + (\rho + r)Y_{t-1} + \rho r Y_{t-2} + \nu_t - (\rho + r)\nu_{t-1} + \rho r \nu_{t-2} \tag{70}$$

Thus, you can estimate this as an ADL model. When you estimate it, you will automatically recover β and δ; however, you cannot get ρ and r directly. This approach can be generalized to more independent variables. For example, let n=number of independent variables. You will need n lags of Y and the error term, as well as n-1 lags of each independent variable. For example, if there were three independent variables then you would regress Y_t on Y_{t-1}, Y_{t-2}, Y_{t-3}, ν_t, ν_{t-1}, ν_{t-2}, ν_{t-3}, X_t^1, X_{t-1}^1, X_{t-2}^1, X_t^2, X_{t-1}^2, X_{t-2}^2, X_t^3, X_{t-1}^3, X_{t-2}^3.

Table 5: **Impulse and Unit Response Function for ADL model**

Horizon	Impulse	Response	Unit	Response
0	0	0	0	0
1	1	β_0	1	β_0
2	0	$\beta_1 + \rho\beta_0$	1	$\beta_0 + \beta_1 + \rho\beta_0$
3	0	$\rho(\beta_1 + \rho\beta_0)$	1	$\beta_0 + \beta_1 + \rho(\beta_0 + \beta_1 + \rho\beta_0)$
4	0	$\rho^2(\beta_1 + \rho\beta_0)$	1	$\beta_0 + \beta_1 + \rho(\beta_0 + \beta_1 + \rho(\beta_0 + \beta_1 + \rho\beta_0))$

Note: The shock occurs in period 1

5.7 Error Correction Model

Another model is the error correction model (ECM).[17] While we will look at it in more detail next week when we deal with non-stationary time series, it is important to recognize that the ECM is perfectly appropriate for stationary time series.[18] In fact, we'll see in a moment that it is *exactly* equivalent to an ADL model but written slightly differently (Keele & DeBoeuf 2005). The term 'error correction models' applies to any model that directly estimates the rate at which changes in Y_t return to equilibrium after a change in X_t. The EC model has a nice behavioral justification in that it implies that the behavior of Y_t is tied to X_t in the long run and that short run changes in Y_t respond to deviations from that long run equilibrium. Below is an ECM:

$$\Delta Y_t = \beta_0 \Delta X_t + \gamma \left[Y_{t-1} - \beta_3 X_{t-1} \right] + \nu_t \tag{71}$$

where Δ refers to a first difference i.e. $\Delta Y_t = Y_t - Y_{t-1}$. As you can see, the model uses differences in both the dependent variables and the independent variables. The inclusion of the second term $Y_{t-1} - \beta_3 X_{t-1}$ is the explicit formulation of the fact that we assume that X and Y have a long-term equilibrium relationship.[19] More specifically, we know that any change in Y_t is a sum of two effects: (i) the short-run impact of the change in X_t on Y_t and (ii) the long-run impact of the deviation from the equilibrium value in period t adjusted at each period at the rate γ. Thus,

- β_0 captures the short-run relationship between X and Y. It indicates how Y and ΔY immediately change if X goes up one period. In other words, if X goes up by one unit in period 5, then Y and ΔY go up β_0 units in period 5.
- γ gives the rate at which the model re-equilibrates i.e. the speed at which it returns to its equilibrium level. Formally, γ tells us the proportion of the disequilibrium which is corrected with each passing period. This coefficient should be negative and less than the absolute value of one indicating its

[17] This model is sometimes referred to as a DHSY model after Davidson, Hendry, Srba and Yeo (1978).

[18] Many people think that the ECM is only appropriate for non-stationary time series models that are characterized by cointegration. However, this is wrong (Keele & DeBoeuf 2005).

[19] Lots of people estimate the following model:

$$\Delta Y_t = \beta_0 \Delta X_t \tag{72}$$

However, this is problematic since there is nothing that keeps X and Y in equilibrium i.e. there is no long-run relationship between X and Y. This is why the ECM model incorporates the second term $Y_{t-1} - \beta_3 X_{t-1}$. Note that if there is no long-run relationship, then $\gamma = 0$.

re-equilibrating properties. If $\gamma = 0$, then the process never re-equilibrates and if $\gamma = -1$, then re-equilibration occurs in one period.

5.7.1 ECM Equivalence with ADL(1,1) Model

It is easy to see that we can get an ECM by starting with an ADL(1,1) model.

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \rho Y_{t-1} + \nu_t \tag{73}$$

Subtract Y_{t-1} from both sides.

$$\Delta Y_t = \beta_0 X_t + \beta_1 X_{t-1} + (\rho - 1) Y_{t-1} + \nu_t \tag{74}$$

Let $(\rho - 1) = \gamma$ i.e.

$$\Delta Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \gamma Y_{t-1} + \nu_t \tag{75}$$

$\Delta X_t = X_t - X_{t-1}$. Thus, $X_t = \Delta X_t + X_{t-1}$. Substitute this in for X_t gives

$$\begin{aligned} \Delta Y_t &= \beta_0 \Delta X_t + \beta_0 X_{t-1} + \beta_1 X_{t-1} + \gamma Y_{t-1} + \nu_t \\ &= \beta_0 \Delta X_t + (\beta_0 + \beta_1) X_{t-1} + \gamma Y_{t-1} + \nu_t \end{aligned} \tag{76}$$

Let $\beta_2 = \beta_0 + \beta_1$. This gives:

$$\Delta Y_t = \beta_0 \Delta X_t + \beta_2 X_{t-1} + \gamma Y_{t-1} + \nu_t \tag{77}$$

We want a situation in which we have a $Y_{t-1} - X_{t-1}$, so rearrange (118) so that we have:

$$\Delta Y_t = \beta_0 \Delta X_t + \gamma \left[Y_{t-1} + \frac{\beta_2}{\gamma} X_{t-1} \right] + \nu_t \tag{78}$$

Thus, the error correction model can be written as:

$$\Delta Y_t = \beta_0 \Delta X_t + \gamma \left[Y_{t-1} - \beta_3 X_{t-1} \right] + \nu_t \tag{79}$$

where $\beta_3 = \frac{-\beta_2}{\gamma} = \frac{-(\beta_0+\beta_1)}{\gamma}$ and $\gamma = \rho - 1$. In other words, the following ADL(1,1) model

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \rho Y_{t-1} + \nu_t \tag{80}$$

can be rewritten as the following error correction model

$$\begin{aligned} \Delta Y_t &= \beta_0 \Delta X_t + (\rho - 1) \left[Y_{t-1} - \left(\frac{-(\beta_0 + \beta_1)}{\rho - 1} \right) X_{t-1} \right] + \nu_t \\ &= \beta_0 \Delta X_t + (\rho - 1) \left[Y_{t-1} + \left(\frac{\beta_0 + \beta_1}{\rho - 1} \right) X_{t-1} \right] + \nu_t \end{aligned} \tag{81}$$

5.7.2 ECM in Lag Operator Form

We can write the error correction model in lag operator form as well. Start with (9).

$$\Delta Y_t = \beta_0 \Delta X_t + \gamma \left[Y_{t-1} - \beta_3 X_{t-1} \right] + \nu_t$$

$$Y_t - Y_{t-1} = \beta_0 X_t - \beta_0 X_{t-1} + \gamma Y_{t-1} - \gamma \beta_3 X_{t-1} + \nu_t$$

$$Y_t - Y_{t-1} - \gamma Y_{t-1} = \beta_0 X_t - \beta_0 X_{t-1} + \gamma \beta_3 X_{t-1} + \nu_t$$

$$Y_t(1 - (1+\gamma)L) = (\beta_0 - (\beta_0 + \gamma\beta_3)L) X_t + \nu_t$$

Since $\gamma = \rho - 1$, $\beta_3 = \frac{-\beta_2}{\gamma}$ and $\beta_2 = \beta_0 + \beta_1$, this can be rewritten as:

$$\begin{aligned} Y_t(1 - (1 + \rho - 1)L) &= \left(\beta_0 - \left(\beta_0 + \gamma \left(\frac{-\beta_0 - \beta_1}{\gamma} \right) \right) L \right) X_t + \nu_t \\ Y_t(1 - \rho L) &= (\beta_0 + \beta_1 L) X_t + \nu_t \\ Y_t &= \left(\frac{\beta_0 + \beta_1 L}{1 - \rho L} \right) X_t + \frac{\nu_t}{(1 - \rho L)} \end{aligned} \tag{82}$$

Note that this is exactly the same as the ADL model written in lag operator form shown in Eq. (68). This is as one would expect given that we have just seen that the ADL model and error correction model are exactly equivalent.

5.7.3 Estimating an ECM Model

There are two ways to estimate an ECM model.

1. **Engle-Granger Two Step Procedure**

   - Estimate the following regression: $Y_t = \alpha + \gamma X_t + \epsilon_t$. In STATA, type

     ```
     regress Y X
     ```

   - From these estimates, generate the residuals i.e. $e_t = Y_t - \alpha - \gamma X_t$. This is how much the system is out of equilibrium. In STATA, type

     ```
     predict e,resid
     ```

   - Now include the lag of the residuals from the initial regression, so that we have $\Delta Y_t = \beta_0 + \beta_1 \Delta X_{t-1} + e_{t-1}$. In STATA, type

     ```
     regress d.Y L.d.X L.e
     ```

2. **One Step Procedure**

   - Simply estimate the following model:

$$\begin{aligned}\Delta Y_t &= \beta_0 + \beta_1 \Delta X_t + \gamma\left[Y_{t-1} - \alpha - \gamma X_{t-1}\right] + u_t \\ &= (\beta_0 - \rho\alpha) + \beta_1 \Delta X_t + \rho Y_{t-1} - \rho\gamma X_{t-1} + u_t \end{aligned} \tag{83}$$

   In STATA, type

```
regress d.Y d.X L.Y L.X
```

   - In this setup, the coefficient on the lagged dependent variable (ρ) is the coefficient on the error correction mechanism and β_1 is the short-run effect of X_t on Y_t.
   - Thus, you can estimate the ECM as a single equation, where changes in Y_t are a function of changes in X_t, the one period lagged value of X, and the one period lagged value of Y. Note that when we estimate this model we will be unable to determine what β_0 is or what α is since all we will get is the constant $\beta_0 - \rho\alpha$. This is not a problem if we simply want to look at impulse and unit response functions.

These two procedures are asymptotically equivalent and will be useful when we discuss issues of cointegration next week.

5.7.4 Impulse and Unit Response Functions

As with the other models, we might be interested in looking at the short-run and long-run effects of the independent variables as seen through the impulse and unit response functions. In the case of the ECM model, we might be interested in these response functions for both Y_t and ΔY_t. These are shown in Tables 5.7.4 and 5.7.4.

Table 6: **Impulse Response Function for ECM Model**

Horizon	X_t	ΔX_t	Y_t	ΔY_t
1	0	0	0	0
2	1	1	β_1	β_1
3	0	-1	$\rho\beta_1 - \rho\gamma$	$-\beta_1 + \rho\beta_1 - \rho\gamma$
4	0	0	$(1+\rho)(\rho\beta_1 - \rho\gamma)$	$\rho(\rho\beta_1 - \rho\gamma)$
5	0	0	$(1+\rho)^2(\rho\beta_1 - \rho\gamma)$	$\rho(1+\rho)(\rho\beta_1 - \rho\gamma)$
6	0	0	$(1+\rho)^3(\rho\beta_1 - \rho\gamma)$	$\rho(1+\rho)^2(\rho\beta_1 - \rho\gamma)$

Note: The shock occurs in period 2

In terms of the impulse response functions, the initial shock in X_t that causes Y_t and ΔY_t to change is gradually eroded over time back to the original equilibrium.

Table 7: **Unit Response Function for ECM Model**

Horizon	X_t	ΔX_t	Y_t	ΔY_t
1	0	0	0	0
2	1	1	β_1	β_1
3	1	0	$\beta_1 + \rho\beta_1 - \rho\gamma$	$\rho\beta_1 - \rho\gamma$
4	1	0	$(1+\rho)(\beta_1 + \rho\beta_1 - \rho\gamma) - \rho\gamma$	$\rho(\beta_1 + \rho\beta_1 - \rho\gamma) - \rho\gamma$
5	1	0	$(1+\rho)^2(\beta_1 + \rho\beta_1 - \rho\gamma) - \rho\gamma$	$\rho[(1+\rho)(\beta_1 + \rho\beta_1 - \rho\gamma) - \rho\gamma] - \rho\gamma$
6	1	0	$(1+\rho)^3(\beta_1 + \rho\beta_1 - \rho\gamma) - \rho\gamma$	$\rho[(1+\rho)^2(\beta_1 + \rho\beta_1 - \rho\gamma) - \rho\gamma] - \rho\gamma$

Note: The shock occurs in period 2

In terms of the unit response function, the initial shock in presidential approval causes ΔY_t to change. It then gradually approaches its new equilibrium where ΔY_t is changing by a given amount each period. A similar story can be told with the unit response function on Y_t.

5.8 Comparing Models: Tests of Nested Model Specifications

The models that we have considered differ in (i) their inclusion/exclusion of lagged values of the dependent and independent variables, and (ii) their assumptions about the error process. Table 8 gives an overview. The nice thing here is that all of the models outlined in the table are a subset of an ADL model with AR1 errors i.e.

$$Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \rho Y_{t-1} + \epsilon_t \tag{84}$$

where

$$\epsilon_t = \phi\epsilon_{t-1} + \nu_t \tag{85}$$

For example, we have already seen that the ECM can be derived from the ADL model. The lagged dependent variable model is simply the ADL model with $\beta_1 = 0$. The same is true for the simple static model if $\beta_1 = 0$ and $\rho = 0$. The models with iid error processes are a subset of ADL with AR1 error process because they occur if $\phi = 0$.

Table 8: Connection between Models

Model	X_t	X_{t-1}	Y_{t-1}	Error Process
OLS	Yes			i.i.d.
AR1 error process	Yes			AR1
Finite distributed lag	Yes	Yes		i.i.d.
Lagged dependent variable	Yes		Yes	i.i.d.
Error correction	Yes	Yes	Yes	i.i.d.
ADL	Yes	Yes	Yes	i.i.d.

One question you might have is which model to use. Obviously, theory should be our guide. However, we rarely have a strong enough theory to distinguish between the different models. As a result, Keele and

DeBoeuf (2005), along with many others, argue that you should start with a general model like the ADL (or ECM) and test down to a more specific model. Since the models are all 'nested' in the ADL model, this is quite easy. For example, if one wanted to see whether the ADL model shown in Eq. 84 reduces to an LDV model, you could do a simply test whether $\beta_1 = 0$. Similarly, you could test whether the ADL model reduces to an FDL model by testing whether $\rho = 0$ or not. When considering which model to choose, you should also take account of whether you have remaining serial correlation with the tests outlined earlier.

5.9 Interpretation Issues

In cross-sectional analysis, all estimated effects are necessarily contemporaneous and, therefore, static. In contrast, we can have two types of effects in time series models.

- An exogenous variable may have only short term effects on the outcome variable. These may occur at any lag, but the effect does not persist into the future. Here the effect of X_t on Y_t has 'no memory'.
- An exogenous variable may have both short and long run effects. In this case, the changes in X_{t-s} affect Y_t but that effect is distributed across several future time periods. Often this occurs because the adjustment process necessary to maintain long run equilibrium is distributed over some number of time points.

Dynamic specifications like the ones that we have looked at allow us to estimate and test for both short and long run effects and to compute a variety of quantities of interest. Short run effects are readily available in both the ADL and ECM models.

5.9.1 Long Run Effects

The long-run equilibrium defines the state to which the time series converges to over time. Consider the ADL(1,1) model:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \nu_t \tag{86}$$

The long run equilibrium for this model is:

$$\begin{aligned} Y^* &= \frac{\alpha_0}{1-\alpha_1} + \frac{\beta_0 + \beta_1}{1-\alpha_1} X^* \\ &= k_0 + k_1 X^* \end{aligned} \tag{87}$$

where $k_0 = \frac{\alpha_0}{1-\alpha_1}$ and $k_1 = \frac{\beta_0+\beta_1}{1-\alpha_1}$, and k_1 gives the 'long-run multiplier effect' of X_t with respect to Y_t. We can think of the long run multiplier as the total effect X_t has on Y_t distributed over future time periods. The connection between the long run equilibrium and multiplier with the response functions we looked at earlier should be obvious. Graphing these response functions can be a useful way of showing how the effect of X_t is distributed over time. In some cases, the long run equilibria and the long run multiplier are of greater interest than short run effects.

When the equilibrium relationship between Y_t and X_t is disturbed at some point in time, then $Y^* - (k_0 + k_1 X^*) \neq 0$. In this case, we would expect Y_t to start moving back towards equilibrium in the next period. Interest in the rate of return to equilibrium (the error correction process) is often motivated by the desire to understand just how responsive a process is. The ADL(1,1) model provides us with information about the speed of the error correction. Based on Eq. 86, the speed of adjustment is given by $(1 - \alpha_1)$. Obviously, increases in α_1 produces slower rates of error correction.

Assume that we have the following ECM:

$$\Delta Y_t = \alpha_0 + \alpha_1^* Y_{t-1} + \beta_0^* \Delta X_t + \beta_1^* X_{t-1} + \nu_t \tag{88}$$

In an ECM, we directly estimate the error correction rate, α_1^*, the short run effect of X_t, and their standard errors. The long run multiplier, k_1 is also more readily calculated in the ECM than in the ADL:

$$k_1 = \frac{\beta_1^*}{-\alpha_1^*} = \frac{\beta_1 + \beta_0}{1 - \alpha_1} \tag{89}$$

Neither the ECM or ADL model provide a direct estimate of the standard error for the long run multiplier. However, since the long run multiplier is the ratio of two coefficients in the ECM $\frac{\beta_1^*}{\alpha_1^*}$, the standard error can be derived from the ECM. This can be derived from the following equation:

$$\text{var}\left(\frac{a}{b}\right) = \frac{1}{b^2}\text{var}(a) + \frac{a^2}{b^4}\text{var}(b) - 2\left(\frac{a}{b^3}\right)\text{cov}(a, b) \tag{90}$$

In addition to knowing the magnitude of the total effect of a shock as measured by the long run multiplier, we might also want to know how many periods it takes for some portion of the total effect of a shock to dissipate. The median lag length tells us the first lag, r, at which at least half of the adjustment toward long run equilibrium has occurred following a shock in X_t. It is calculated by listing the effect of a unit change in X_t at each successive lag, standardizing it as a portion of the cumulative effect, and then noting at which lag the sum of these individual effects exceed half of the long run effect.

5.9.2 Example

Let's look at an example from Keele and DeBoeuf (2005) where they use an ECM and ADL model on the same (simulated) data. The results are shown in Table 9.

Table 9: ADL and ECM Estimates

Model	ADL	ECM
Y_{t-1}	0.75	-0.25
	(0.02)	
X_t	0.53	
	(0.06)	
X_{t-1}	0.25	0.77
	(0.07)	
ΔX_t		0.53
ΔY_t		
k_1	3.12	3.08
N	249	249
R^2	0.94	0.45

First, you might be interested in the short run effects of X_t on Y_t. For the ADL model, these are given explicitly by the estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ - the coefficients on X_t and X_{t-1}. These are 0.53 and 0.25. Thus, an increase in X_t leads to an increase in Y_t of 0.53 in the same period as the increase and a further 0.25 in the next period. For the ECM, the coefficient on ΔX_t i.e $\hat{\beta}_1^*$ should equal $\hat{\beta}_0$. For the second short run effect in the ECM model, things are slightly less clear. In the ADL model, $\hat{\beta}_1$; in the ECM model, we use $\hat{\beta}_1^* - \hat{\beta}_0^* = 0.77 - 0.53 = 0.24$.[20] Second, you might be interested in the long-run multiplier, k_1. For the ADL model, the long run multiplier is $\frac{0.53+0.25}{0.75} = 3.12$. For the ECM model, the long run multiplier is $\frac{\hat{\beta}_1^*}{-\hat{\alpha}_1^*} = \frac{0.77}{-(-0.25)} = 3.08$. You could calculate the standard error around these long run multipliers using the formula given in Eq. 90. Third, we can interpret the error correction rate - this is given directly in the ECM model i.e. $\hat{\alpha}_1^* = -0.25$. It can be obtained from the $(1 - \alpha_1) = 0.25$. Thus, the relationship between X_t and Y_t, following the shock to X_t, will return to equilibrium at a rate of 25% each period. Fourth, we could calculate the median lag length. The easiest way to do this is to look at the ADL model and assume that we are in equilibrium except for a shock in X_t. Thus, in the first period, we would have:

$$Y_t = 0.53X_t \tag{91}$$

Normalizing as a proportion of the estimated long run effect, we have 0.53/3.12=0.17 or 17%. The effect of the shock one period later is

$$\begin{aligned} Y_{t+1} &= 0.75Y_t + 0.25X_t \\ &= 0.75(0.53) + 0.25 = 0.648 \end{aligned} \tag{92}$$

Normalizing as a proportion of the estimated long run effect, we have 0.648/3.12=0.208 or 20.8%. We are now up to 37.8% of the total effect. In the next period, we have

$$\begin{aligned} Y_{t+2} &= 0.75Y_t \\ &= 0.75(0.648) = 0.486 \end{aligned} \tag{93}$$

Normalizing as a proportion of the estimated long run effect, we have 0.486/3.12=0.156 or 15.6%. We have now exceeded one half of the long run effect i.e. 17%+20.8%+15.6%=53.4%. Thus, the median lag length is 2 i.e. it takes two periods after the initial shock in time t for half the long run effect to be done.

Note that virtually all of these quantities of interest are captured to some extent in a graphical representation of the impulse and unit response functions. These figures can be a good way of showing how the effect of an increase in X_t feeds through on Y_t.

[20] Differences in these estimates occur due to rounding - they will be the same.

6 Stationarity

Stationarity is essentially a restriction on the data generating process over time. More specifically, stationarity means that the fundamental form of the data generating process remains the same over time.[21] This is indicated in the moments of the process. For example, mean stationarity means that the expected value of the process is constant over time i.e.

$$E[Y_t] = \mu \; \forall \; t \tag{94}$$

Variance stationarity means that the variance is temporally stable:

$$\text{var}[Y_t] = E[(Y_t - \mu)^2] \equiv \sigma_Y^2 \; \forall \; t \tag{95}$$

and covariance stationarity is the same i.e.

$$\text{cov}[Y_t, Y_{t-s}] = E[(Y_t - \mu)(Y_{t-s} - \mu)] = \gamma_s \; \forall \; s \tag{96}$$

In this last case, this means that the serial correlation of two observations $\{Y_t, Y_{t-s}\}$ depends only on the lag s and not on 'where' in the series they fall. What I have described is called weak stationarity (stationarity in the moments). A stricter form of stationarity requires that the joint probability distribution (all the moments) of series of observations $\{Y_1, Y_2, \ldots, Y_t\}$ is the same as that for $\{Y_{1+s}, Y_{2+s}, \ldots, Y_{t+s}\}$ for all t and s. In general, we tend to be only interested in weak stationarity.

6.1 Integrated Processes

Time series can be categorized according to the nature of the data generating process that underlies them. There are three general characteristics: autoregressive series, moving average series, and integrated series. You can also get combinations of these series i.e. ARIMA series. We have already seen autoregressive and moving average (ARMA) series earlier. Now, I'll describe an integrated series. An integrated series is one in which the value of Y_t is simply the sum of random 'shocks'.

6.1.1 I(1) Series

An example is the integrated of order one (I(1)) series:

$$Y_t = Y_{t-1} + \nu_t \tag{97}$$

where ν_t is a white noise (i.i.d.) error process as before. As you'll recall, this is equivalent to the more general model

$$Y_t = \rho Y_{t-1} + \nu_t \tag{98}$$

[21] Notes based on Zorn and other readings.

where $\rho = 1$. This is the random walk model from earlier. By recursively substituting Y_{t-k-1} for Y_{t-k}, we can see that

$$
\begin{aligned}
Y_t &= Y_{t-2} + \nu_{t-1} + \nu_t \\
&= Y_{t-3} + \nu_{t-2} + \nu_{t-1} + \nu_t \\
&= \sum_{t=0}^{T} \nu_t
\end{aligned} \tag{99}
$$

In other words, the random walk process is simply a sum of all past random shocks. This means that the effect of any one shock persists – a large increase or decrease in Y at some point t will cause the series to shift up or down until a countervailing shock comes along. As a result of this persistence, integrated series tend to drift – this explains the slow up and down movements in Figure 3 we saw earlier. This figure makes it look as if the series is trending – in fact it isn't, it is just that the persistence in this series makes it look that way. In reality, we say that this type of series is drifting and that it can drift away from its mean for long periods of time. It is this tendency to drift that means that I(1) series are non-stationary. In particular, the variance of an I(1) process is equal to

$$
\text{var}[Y_t] = E[(Y_t)^2] = t\sigma^2 \tag{100}
$$

and the covariance is

$$
\text{cov}[Y_t, Y_{t-s}] = |t - s|\sigma^2 \tag{101}
$$

As you can see, these equations are different from Equations 95 and 96. It is easy to see that both values depend on t, which is why the I(1) series is non-stationary.

Because an I(1) series is simply the sum of all previous shocks, it is easy to model it – difference it. In other words, we can rearrange Equation 97

$$
Y_t - Y_{t-1} = \nu_t = \triangle Y_t \tag{102}
$$

As you can see, the differenced series is just a white noise process ν_t. If you plot the differenced series, it will look like the earlier figure of our stationary series.

In general, the order of integration can be thought of as the number of differencings a series requires to be made stationary. For example, a non-stationary I(1) series, once it is differenced once, is stationary. Similarly, an I(d) series is one which, when differenced d times becomes stationary. For example, an I(2) series is

$$
Y_t = \nu_t + 2\nu_{t-1} + 3\nu_{t-2} + \ldots \tag{103}
$$

If we difference this equation we have

$$
\begin{aligned}
\triangle Y_t &= [\nu_t + 2\nu_{t-1} + 3\nu_{t-2} + \ldots] - [\nu_{t-1} + 2\nu_{t-2} + 3\nu_{t-3} \ldots] \\
&= \sum_{j=0}^{T} \nu_{t-j}
\end{aligned} \tag{104}
$$

If we difference this again, we have

$$\begin{aligned}\triangle^2 Y_t &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = [\nu_t + \nu_{t+1} + \ldots] - [\nu_{t-1} + \nu_{t-2} + \ldots] \\ &= \nu_t \end{aligned} \tag{105}$$

which is stationary.

Another kind of series that produces a higher-order integrated process is one with a polynomial trend. Consider the following model.

$$Y_t = t^2 \tag{106}$$

Differencing this gets

$$\begin{aligned}\triangle Y_t &= t^2 - (t-1)^2 \\ &= t^2 - t^2 + 2t - 1 \\ &= 2t - 1 \end{aligned} \tag{107}$$

and further differencing gives

$$\triangle^2 Y_t = (2t-1) - (2t-3) = 2 \tag{108}$$

As Zorn notes, there are not many practical applications involving orders of integration higher than I(1).

6.2 Different Types of Non-Stationarity

Typically, when we want to look at the stationarity of a series we might have the following equation

$$Y_t = \rho Y_{t-1} + \nu_t \tag{109}$$

As we saw at the very beginning of these notes, Y_t is stationary if $\rho < 1$ and non-stationary if $\rho \geq 1$. However, it turns out that there are many reasons why, or ways in which, a series may be non-stationary. For example, recall the random walk model from earlier

$$Y_t = Y_{t-1} + \nu_t \tag{110}$$

This is an example of a non-stationary series that can be made stationary by first differencing. As a result, it is called a first-order non-stationary series or an integrated series of order one i.e. I(1). It is an I(1) series since first differencing produces a series that is I(0) i.e.

$$Y_t - Y_{t-1} = \nu_t = \triangle Y_t \tag{111}$$

since ν_t is stationary.

Another model that is non-stationary is shown below

$$Y_t = \beta t + \nu_t \tag{112}$$

where t is a time counter. It is easy to see that this model is non-stationary since $\bar{Y}$ is increasing over time. This series is also an I(1) series since first differencing produces a series that is I(0):

$$\begin{aligned}\Delta Y_t = Y_t - Y_{t-1} &= \beta t + \nu_t - (\beta(t-1) + \nu_{t-1}) \\ &= \beta t + \nu_t - \beta t + \beta - \nu_{t-1} \\ &= \nu_t - \nu_{t-1} + \beta \end{aligned} \tag{113}$$

because $\nu_t - \nu_{t-1} + \beta$ is stationary.

One problem with these two models is that the differencing by itself does not tell you why the initial series of non-stationary. We could address this problem by examining

$$Y_t = \rho Y_{t-1} + \beta t + \nu_t \tag{114}$$

and test for H_0 : $\hat{\beta}$=0.

- If we cannot reject $\hat{\beta} = 0$, then this is evidence in favor of a series being a random walk without a trend.
- If we can reject $\hat{\beta} = 0$, then this suggests that the series has a deterministic time trend.

The bottom line is that non-stationarity can be caused by several different things.

6.3 Testing for Non-Stationarity

So, how do we test for non-stationarity? Let's start with the following model.

$$Y_t = \rho Y_{t-1} + \nu_t \tag{115}$$

If you recall from earlier, we have the following three possibilities.

1. $|\rho| > 1$
   Series is non-stationary and explosive. Past shocks have a greater impact than current ones. These situations are uncommon and we have not looked at how to deal with them.
2. $|\rho| < 1$
   Series is stationary and the effects of shocks die out exponentially according to ρ. The series reverts to its mean. We have already seen what to do with these.
3. $|\rho| = 1$
   Series is non-stationary, shocks persist at full force, and the series is not mean-reverting. This is the random walk model and as we saw, the variance increases with t and so we have the infinite variance problem.

Typically, we are interested in determining between the last two scenarios i.e. $|\rho| < 1$ or $|\rho| = 1$. The question is whether we have a **unit root** or not (also known as a random walk) i.e. is $|\rho| = 1$?

6.3.1 Dickey-Fuller Test

You might think that it would be easy to test for stationarity by simply looking to see whether $\hat{\rho}$=1. However, things are not quite so simple. The problem is that the distribution of $\hat{\rho}$ is not standard. Although $\hat{\rho}$ is a consistent estimate of ρ, it does not follow a standard t-distribution. This is where the Dickey-Fuller distribution comes in. This distribution is right-skewed meaning that t-statistics will tend to be large and negative. It also means that we will tend to over-reject the null hypothesis of non-stationarity i.e. $\rho = 1$. In other words, we may conclude that our series is stationary when it really isn't if we use a t-test. In effect, the Dickey-Fuller test for a unit root amounts to estimating $\hat{\rho}$ and doing a standard-looking t-test for $H_0 : \hat{\rho} = 1$ but using a non-standard set of critical values.

- The null hypothesis is that the series has a unit root.
- If t is greater than the critical values, then the series is stationary (no unit root).
- If t is less than the critical values, then the series is non-stationary (unit root)

An important point to note when using the Dickey-Fuller test is that the errors are assumed to be white noise.

Another way of estimating the Dickey-Fuller test is by doing the following:

$$\begin{aligned} \Delta Y_t &= (\rho - 1)Y_{t-1} + u_t \\ &= \delta Y_{t-1} + u_t \end{aligned} \tag{116}$$

where we now need to test to see if $\delta = 0$. If $\delta = 0$, we have non-stationarity. If $\delta \neq 0$ then we have stationarity. Note that this is exactly how STATA does its Dickey-Fuller test. So, the important thing to note is that you are testing to see if $\delta = 0$ and not whether $\rho = 1$ when using STATA. This transformation is nice because it means that we do not have to calculate the test statistic $\frac{\hat{\rho}-1}{\sigma_{\hat{\rho}}}$.

If you recall, there are different sources of non-stationarity. To test for these different sources, you can estimate slightly different D-F tests.

1. No Constant (no drift, no trend)

If we estimate a model without a constant, we are assuming that there is no drift or deterministic time trend. Such a model is shown below. This allows us to test for a unit root.

$$Y_t = \rho Y_{t-1} + \nu_t \tag{117}$$

You can estimate this D-F test using STATA by typing

- dfuller Y, noconstant

You can also do the D-F test manually by typing:

- regress D.Y L.Y, noconstant

where $D.Y$ means the change in Y ($\triangle Y$) and $L.Y$ is the lag of Y (Y_{t-1}).

2. CONSTANT (DRIFT, NO TREND)

As I said earlier, we may believe that there is non-stationarity for other reasons than a unit root. One reason is that there may be drift. It is possible to test for both a unit root and drift with the following model.

$$Y_t = \alpha + \rho Y_{t-1} + \nu_t \tag{118}$$

Note the introduction of a constant now (α). It turns out that this model can be rewritten (see Equation 99) as

$$Y_t = Y_0 + \alpha t + \sum_{t=1}^{T} u_t \tag{119}$$

It is now easy to see that the series has a drift which manifests itself as the trend αt; over time the drift will come to dominate the series. Note that the presence of drift makes the series non-stationary on its own irrespective of whether there is a unit root. This means that if we think there might be drift and are worried about non-stationarity, we need to test both $\hat{\rho} = 1$ and $\hat{\alpha} = 0$. We might also want to do a joint test. Note that the critical values for testing this model are different to the critical values used to test the model with no constant shown earlier.

As before, we can rewrite (118) as

$$\begin{aligned} \Delta Y_t &= \alpha + (\rho - 1)Y_{t-1} + u_t \\ &= \alpha + \delta Y_{t-1} + u_t \end{aligned} \tag{120}$$

You can estimate this version of the D-F test using STATA by typing:

- dfuller Y, regress

You can also do it manually by typing

- regress D.Y L.Y[22]

3. CONSTANT AND TIME TREND

Instead of there being a drift, we may have a series with a deterministic trend. Such a model is shown below

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + u_t \tag{121}$$

Now α really is just a constant. The deterministic time trend is captured by βt. Again, the deterministic time trend can lead to non-stationarity on its own. Thus, to test for non-stationarity we need to test both $\hat{\rho} = 1$ and $\beta = 0$. We can do a joint F-test as well. Again, we need to use a slightly different set of critical values to those used in the models with drift and the no constant.

As before we can reformulate (121) as

$$\begin{aligned} \Delta Y_t &= \alpha + \beta t + (\rho - 1)Y_{t-1} + u_t \\ &= \alpha + \beta t + \delta Y_{t-1} + u_t \end{aligned} \tag{122}$$

[22]Note again that we could just as easily type 'regress prezapproval L.prezapproval' etc. and test that $\hat{\rho} = 1$ rather than that $\hat{\delta} = 0$.

You can estimate this model in STATA by typing

- dfuller Y L.Y, trend regress

Or you can do it manually by typing:

- generate trend = _n
- regress D.Y L.Y trend

Zorn notes that it is quite common to check for time trends and that it is worse to omit a trend from a model when the data generating process has one than to include one when it does not. In practice, people nearly always include a constant and they often include a trend.

CRITICAL VALUES

The critical values for rejecting the null hypotheses with the D-F test depend on whether we have a no constant model, a drift model, or a trend model. Below is a summary of the critical values for each model.

Table 10: Critical Values for Dickey-Fuller Tests

Model	p-values	Critical Value
No Constant	p<0.05	-1.939
	p<0.01	-2.566
Drift, No Trend	p<0.05	-2.862
	p<0.01	-3.434
Trend	p<0.05	-3.413
	p<0.01	-3.964

t-statistics in parentheses.

6.3.2 Augmented Dickey-Fuller and Phillips-Perron Test

As I stated earlier, the Dickey-Fuller test requires that the errors be uncorrelated i.e. no serial correlation. Why is serial correlation a problem? Well imagine that we had a model where the first difference of Y is a stationary AR(p) process:

$$\triangle Y_t = \sum_{i=1}^{p} d_i \triangle Y_{t-i} + \nu_t \qquad (123)$$

This model says that the model for Y_t is

$$\triangle Y_t = Y_{t-1} + \sum_{i=1}^{p} d_i \triangle Y_{t-i} + \nu_t \qquad (124)$$

If this is really our model and we estimate a standard D-F test i.e.

$$Y_t = \hat{\rho} Y_{t-1} + \nu_t \qquad (125)$$

then the term $\sum_{i=1}^{p} d_i \triangle Y_{t-i}$ gets put in the error term ν_t. This induces an AR(p) structure in the νs and the standard D-F statistic will be wrong.

So, what can we do? Well, we have two possibilities.

AUGMENTED DICKEY-FULLER TEST

Instead of estimating

$$Y_t = \rho Y_{t-1} + \nu_t \tag{126}$$

we estimate

$$\Delta Y_t = \delta Y_{t-1} + \alpha_i \sum_{i=1}^{p} \Delta Y_{t-i} + \nu_t \tag{127}$$

As before, we will test whether $\hat{\hat{\delta}} = 0$ to see if we have a unit root.[23]In effect, we are changing the model that we use in the test with the augmented D-F test. We can estimate this test in STATA by typing

- dfuller Y, noconstant lags(1)

Just as with the standard D-F test, we can still include a constant/trend term i.e.

$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \alpha_i \sum_{i=1}^{p} \Delta Y_{t-i} + \nu_t \tag{128}$$

The test that I have shown is assuming that we have an AR(1) process. Essentially what we have done is to include a lag of δY_{t-1} so as to ensure that the νs are white noise. However, one lag may not be enough i.e. we may not have an AR(1) process. This will require adding in additional lags of δY_{t-1}. There are some decisions to make when determining how many lags to put in. If you don't put enough lags in then there will be serial correlation in the errors and the test will not work or you can put too many lags in and the power of the test statistic goes down. Zorn suggests that you start with a large value of p i.e. include a lot of lags and reduce it if the values of $\hat{\hat{\delta}}$ are insignificant. An alternative strategy is to estimate the model with different values of p and then use an AIC/BIC/F-test to determine which is the best option.

PHILLIPS-PERRON TEST

The Phillips-Perron test does not change the model that it uses in the test. In other words, it keeps the model we used originally:

$$\Delta Y_t = \alpha + \rho Y_{t-1} + \nu_t \tag{129}$$

Without going into any detail, the P-P test adjusts the original D-F test statistic so that it more closely conforms to the 'standard' D-F distribution. If you wanted to estimate a P-P test in STATA, you would type one of the following depending on what you think the model is:

- pperron Y, lags(1)

[23]Remember that this is the same as having $Y_t = \rho Y_{t-1} + d_i \sum_{i=1}^{p} Y_{t-i} + \nu_t$ and testing to see if ρ=1.

- pperron Y, lags(1) trend

Note that the 'lags' don't refer to lags of δY_{t-1} any more. As before, though, you will need to play around with how many lags to include.

TBA

- Granger Causality
- Spurious Regression
- Cointegration
- Error Correction Models again

References

Achen, Christopher H. 2000. "Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables." Paper presented at the Annual Meeting of Political Methodology, Los Angeles.

Beck, Nathaniel, Kristian Skrede Gleditsch & Kyle Beardsley. 2006. "Space is More than Geography: Using Spatial Econometrics in the Study of Political Economy." *International Studies Quarterly* 50:27–44.

Bernhard, William & David Leblang. 2006. *Democratic Processes and Financial Markets*. New York: Cambridge University Press.

Bollerslev, Tim. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31:307–327.

Davidson, J. E. H., D. F. Hendry, F. Srba & S. Yeo. 1978. "Econometric Modelling of the Aggregate Time Series Relationship Between Consumers' Expenditure and Income in the United Kingdom." *Economic Journal* 88:661–692.

Engle, Robert. 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation." *Econometrika* 50:987–1008.

Greene, William. 2003. *Econometric Analysis*. New Jersey: Prentice Hall.

Gujarati, Damodar. 2003. *Basic Econometrics*. New York: McGraw Hill, Inc.

Johnston, Jack & John DiNardo. 1997. *Econometric Methods*. New York: McGraw Hill Companies, Inc.

Keele, Luke & Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14:186–205.

Keele, Luke & Suzanna DeBoeuf. 2005. "Taking Time Seriously: Dynamic Regression." Paper presented at the 2005 Annual Meeting of Political Methodology, Florida State University.

Leblang, David & William Bernhard. 2006. "Parliamentary Politics and Foreign Exchange Markets: The World According to GARCH." *International Studies Quarterly* 50:69–92.

Maestas, Cherie & Robert R. Preuhs. 2000. "Modeling Volatility in Political Time Series." *Electoral Studies* 19:95–110.