Dynamic Panel Models

PS2701-2019

Longitudinal Analysis

Week 4

Professor Steven Finkel

Assumptions of Static Panel Models

- Our standard pooled panel regression model with U:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_k X_{kit} + \beta_m Z_{im} + U_i + \varepsilon_{it}$$

can be termed a "static" model – nothing carries over from one time period to the next aside from U_i carried

Assumptions of the Static Model:

- In FE models, all selection into treatment is due to U (with "common trends" assumption necessary for causal inference)
- In FE and RE models, all "temporal dependence" – i.e., Y at one time point being related to Y at the next or previous time point -- is due to the stable unobservables
- Effects of X are instantaneous and then "die out" until next period

- All of these assumptions may be incorrect or incomplete
  - There may be "endogenous selection" into treatment, such that prior Y may determine X (along with or instead of U)
  - There may be other forms of temporal dependence aside from U, i.e., the Ys may not be independent, controlling for U
  - The effects of X may persist across time periods or cumulate over time
- These conditions lead to models which can accommodate different forms of temporal dependence and selection processes
- These "dynamic models" are used extensively in multi-wave panels, especially with shorter time intervals between waves
- They present serious estimation challenges, however!!!

Dynamics and Temporal Dependence in Econometric Panel Models

- How can we incorporate temporal dependence, over and above the U_i term, into econometric-type panel analysis? There are three distinct, though not completely mutually exclusive, ways:
  - Include the lagged dependent (endogenous) variable Y_{t-1}
  - Model autocorrelation in the disturbance terms ε_{it}
  - Include time itself as an independent variable --- that is, Y_t is modeled as a function of the X, but it is also predicted to increase or decrease by some factor associated with time.
- We'll spend most of our time in this unit on lagged DV models, as they are most substantively distinct from the other two (and "robust" estimation handles some of the autocorrelation issues relatively easily to begin with)

Models with the Lagged Dependent (Endogenous) Variable

$$(1) \qquad Y_{it} = \alpha + \beta_1 Y_{it-1} + \beta_2 X_{1it} + \beta_3 X_{2it} + \ldots \beta_J X_{Jit} + U_i + \varepsilon_{it}$$

Why Include the Lagged DV in an "Autoregressive" Model?

- Substantive Justifications
  - Y_{t-1} may *cause* Y_t (budgets, wealth, e.g.) or Y may have inherent persistence (eg. "states" such as marriage, employment, or attitudes) or "inertia" (e.g., mood) – sometimes called **"state dependence"**
  - "Endogenous Selection" into a treatment based on Y_{t-1}, so need to control (see later slides for more elaborate endogenous regressor models)
- Good Statistical Reasons: Carry-Over and Cumulative Effects of X
  - In distributed lag models (e.g., the "Koyck" model), it can be shown that lagged Y summarizes all past inputs from X, which then carry forward to present period along with current input from X. Larger the effect of lagged Y, the bigger the "carry over" from the past

Why Include the Lagged DV? (continued)

- In “partial adjustment” models, it is assumed that there is a “target” or “equilibrium value” of Y that X leads Y to attain, but that, due to inertia in the DV, it does not reach its target immediately. The larger the effect from lagged Y, the slower it takes Y to reach the equilibrium value
- This also means that constant inputs from X will cumulate over time and become amplified, depending on magnitude of the lagged Y effect
  The equilibrium level of Y is given by $h^* = (\frac{\beta_2}{1-\beta_1})X_1$

  β_2 is the short-term effect, and $(\frac{\beta_2}{1-\beta_1})$ is the the “Long-Term Multiplier”. Compare large and small β_1 in these expressions.

- Mundane Statistical Reasons (controversial)
  - “Regression to the Mean”: random positive (negative) errors from t_1 will recede (increase) at t_2, leading to a **negative relationship between Y_{t-1} and change**
  - “Old school” method for controlling for unmeasured confounders

$$(1) \quad Y_{it} = \alpha + \beta_1 Y_{it-1} + \beta_2 X_{1it} + \beta_3 X_{2it} + \ldots \beta_J X_{Jit} + U_i + \varepsilon_{it}$$

- With all the assumptions we made earlier about U, ε, etc.
- This is the "**Dynamic Panel Model**" or "**State Dependence Model",** since the current value of Y depends on its prior state, and future states of Y depend on current ones. Y is also a function of stable unit-level unobservables (Ui) and an idiosyncratic error term
- It is straightforward to add different lags of X_j, depending on what theory says about the time that it should take *X* to exert causal influence on *Y*, and/or depending on how the lagged DV model was derived theoretically (Koyck versus Partial Adjustment, e.g.).
- One common model is the Autoregressive Distributed Lag Model (ADL), which includes the **lagged endogenous variable** Y_{it-1} and the **current and lagged** values of X. This incorporates the Koyck (X_t) and Partial Adjustment (X_{t-1}) specifications into a general model

Estimation of the Dynamic Panel Model

- Serious problem with estimation of the Dynamic Panel Model of (1): the presence of the lagged dependent variable Y_{t-1} has induced an intrinsic correlation between the independent variables and the equation's composite error term ($U_i+\varepsilon_{it}$). This means that we cannot estimate β_1 without bias.
- Write (1) with lagged Y as the DV:

$$(2) \quad Y_{it-1} = \alpha + \beta_1 Y_{it-2} + \beta_2 X_{1it-1} + \beta_3 X_{2it-1} + \ldots \beta_J X_{Jit-1} + U_i + \varepsilon_{it-1}$$

- Y_{it-1} is a direct function of U_i! So in (1) one of the independent variables (Y_{it-1}) is intrinsically related to the error term of that equation. This introduces *endogeneity bias* in estimates of the effects of (Y_{it-1}) and by extension, the other independent variables.

- Informally: Unless corrected, β_1 will absorb some of the effect that should rightly be attributed to the unit effect U_i, and, depending on the intercorrelations of the Xs with Y_{it-1}, it becomes harder to reject the null that these other variables are statistically insignificant.
- What should be done? First we need to get rid of the unit effects, and then see what we should do next
- One possibility: The fixed effects (FE) transformation

$$(3) \quad Y_{it} - \bar{Y}_{\mathrm{i}} = \beta_1 (Y_{it-1} - \bar{Y}_{\mathrm{i.t\text{-}1}}) + \beta_2 (X_{1it} - \bar{X}_{1\mathrm{i}}) + \ldots \beta_J (X_{Jit} - \bar{X}_{\mathrm{Ji}}) + (\mathrm{U_i} - \bar{\mathrm{U}}_{\mathrm{i}}) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

- Problem, though: $(\varepsilon_{it} - \bar{\varepsilon}_i)$ and $(Y_{it-1} - \bar{Y}_{\mathrm{i.t\text{-}1}})$ are still related!
- Why? Average ε contains ε_{it-1} , and ε_{it-1} is correlated with Y_{it-1}.

- So using FE, there is still a correlation between "demeaned" lagged Y and the demeaned error term.
- BUT: As $T \rightarrow \infty$, average error goes to 0, so in that case the error term in (3) would now be ε_{it}, and will therefore no longer be related to Y_{it-1}, so long as there is no autocorrelation in the ε_i.
- Thus, in long panels, the bias in dynamic panel models goes away!!! So FE models with the lagged DV are appropriate when T is (very) large. This is one of the reasons why it is recommended by Beck and Katz (1995) as one part of their "standard model" for Time Series Cross Section data (large T, small N).
- But in short panels, there is a bias in the FE estimation of the effect of lagged Y on the order of 1/T, which can be quite substantial.

Solutions with Smaller T: Adapt the First Difference (FD) Model

- Start with a FD version of the dynamic model:

(4) $$Y_{it} - Y_{\text{it-1}} = (\alpha - \alpha) + \beta_1(Y_{it-1} - Y_{it-2}) + \beta_2(X_{1it} - X_{1it-1}) + \ldots \beta_{\text{j}}(X_{jit} - X_{\text{jit-1}}) + (\text{U}_{\text{i}} - \text{U}_{\text{i}}) + (\varepsilon_{\text{it}} - \varepsilon_{it-1})$$

- This results in:
  - The U being differenced out of the model, so doesn't matter if we conceived of it as "fixed" or "random"
  - An error term that is now ($\varepsilon_{it} - \varepsilon_{it-1}$)
- Error term is *still* correlated with the lagged difference term" (Y_{it-1}-Y_{it-2}), so we are not quite out of the woods yet. (Why? Because the error term has e_{it-1} which is a cause of Y_{it-1}). But we are moving towards a solution.
- Can we find an instrumental variable or variables related to the lagged difference term (Y_{it-1}-Y_{it-2}) but unrelated to the error term ($\varepsilon_{it} - \varepsilon_{it-1}$)? If so, we can use that as a proxy in an IV or TSLS type estimation.
- If we can assume that there is no autocorrelation in the ε_{it}, PANEL DATA GIVES US MANY DIFFERENT INSTRUMENTAL VARIABLES!!!

The Anderson-Hsiao Method

- One possibility: Use the "twice lagged difference term" (Y_{it-2}-Y_{it-3}) as an instrument for the "lagged difference term" (Y_{it-1}-Y_{it-2}). None of the components of the twice-lagged difference term are correlated with *any* of the components of the error term ($\varepsilon_{it} - \varepsilon_{it-1}$).
- This is the original **"Anderson-Hsiao"** estimation method for the dynamic panel model. You can estimate this in STATA using difference scores
- Big Problem: We lose 3 waves of data to implement this procedure!!!!!
- To partially get around this, another possibility **(Anderson-Hsiao II)** would be to use the *level* variable Y_{it-2} instead of the twice-lagged difference term as an instrument. Y_{it-2} is uncorrelated with the error term ($\varepsilon_{it} - \varepsilon_{it-1}$). So is the lagged difference in X_{it-1} and X_{it-2}. This saves one wave of data and so you lose only 2 waves to implement this version in STATA

- Note: These procedures would not have worked with the fixed effects transformation because in small *T* panel studies, y_{it-2} (or any other lag of y) would still be related to $(\varepsilon_{it} - \bar{\varepsilon}_i)$!!
- (WHY?)
- This is why we use the FD model as the base for the dynamic panel estimation!

The Arellano-Bond "Generalized Method of Moments" Estimator

- Arellano and Bond (1991) showed that there are actually many more possible instruments for the lagged difference term in panel data sets. Their solution rests on the idea that "deeper" lags of *Y* may be used as instruments for the lagged difference term, with more and more lags being available as one moves forward in time in the panel. For example, in a six wave panel, for $(Y_{i6,}-Y_{i5})$, variables Y_{i4} , $Y_{i3,}$ $Y_{i2,}$ and Y_{i1} are all available as instruments for the lagged difference $(Y_{i5,}-Y_{i4})$; for $(Y_{i5,}-Y_{i4})$, $Y_{i3,}$ $Y_{i2,}$ and Y_{i1} are all available as instruments for the lagged difference $(Y_{i4,}-Y_{i3})$, etc.
- So as move through the panel you pick up more and more instruments to get better precision of the estimates. In addition, you can use both lagged levels and lagged differences of the exogenous X variables as other instruments.

- In other words: the instrument set changes for each differenced term that exists in the given panel. For a 6 wave panel, you would have, in addition to lagged levels and differences of the exogenous Xs, the following instruments in *levels* of Y:

Wave	1	2	3	4	5	6
Differenced DV	Out	$Y_2 - Y_1$	$Y_3 - Y_2$	$Y_4 - Y_3$	$Y_5 - Y_4$	$Y_6 - Y_5$
Differenced Lag DV	Out	Out	$Y_2 - Y_1$	$Y_3 - Y_2$	$Y_4 - Y_3$	$Y_5 - Y_4$
Instrument Set	–	–	–	–	–	Y_4
	–	–	–	–	Y_3	Y_3
	–	–	–	Y_2	Y_2	Y_2
	–	–	Y_1	Y_1	Y_1	Y_1

- Thus the Arellano-Bond estimator is generally said to be superior to Anderson-Hsiao, as it uses far more information and is thus more efficient. You can implement this in STATA as XTABOND.

Notes on Arellano-Bond

- The Arellano-Bond procedure depends on no autocorrelation in the idiosyncratic residuals ε_{it}. Recall that the FD (differenced) residuals have a first-order autocorrelation of -.5 under the null hypothesis that there is no autocorrelation in the ε_{it} levels. So within XTABOND, a test of whether there is autocorrelation in the levels of ε_{it} is based on whether there is "second-order" autocorrelation, or autocorrelation in the first differenced residuals, i.e. whether .

  $$Cov(\varepsilon_{it} - \varepsilon_{it-1}),(\varepsilon_{it-2} - \varepsilon_{it-3}) = 0$$

  This is reported under the model results for XTABOND with the post-estimation command "estat abond". If we reject this, then Arellano-Bond needs to be modified, the original equation transformed further to eliminate the autocorrelation, or we need to move to more complex error structures that we will not consider here, some of which are included in STATA's XTDPD and XTDPDSYS modules

- The Arellano-Bond procedure, as with all IV or TSLS procedures, also depends on the instrument set being *truly* exogenous, in that the predicted values generated by the instruments are independent of the equation's error term. When models are overidentified --- that is, when there are more instruments than endogenous variables --- we can test this assumption through what is known as the "Sargan test" of overidentifying restrictions

  Idea of the Sargan test: Regress the stage 2 residuals against **all** stage 1 exogenous variables, including those used to create the proxy endogenous variables and all of the exogenous X variables that are included in the stage 2 regression. $N*R^2$ of this equation follows a χ^2 distribution under the assumption that **all** instruments are orthogonal to the error term. If we reject the null, then we need to adjust the model in XTABOND – the exogenous variables used to create the proxy endogenous variables are *not* unrelated to the outcome equation's error term and hence are not good instruments.

- More on Sargan: it needs an *overidentified* model (with more instruments than potentially endogenous variables), because the way it works is really to use all but one of the IVs to produce the expected value of Y, and then test the relationship of the excluded IV with the error of that equation. If there is a relationship, then the excluded IV is not really a good IV. If you have only one excluded IV for each included endogeneous variable, Sargan test can't be used.
- Moreover, the test is theoretical in the sense that you need to assume that the included IVs are valid before you can test the effect of the given *excluded* IVs. So Sargan is better than nothing, but, as with all IV models, there really is no test that will tell you unequivocally if your IVs truly satisfy the assumptions of the IV method.
- "estat sargan" in STATA post-estimation produces the test

- You can also add your own instruments to the internally-generated set of panel instruments, and indeed, you should be encouraged to do so because the Sargan test often shows that the internal panel instruments are *not* truly exogenous.
- This test can be applied to *all* IV regression procedures, not only ABOND
- The Arellano-Bond procedure needs at least three waves of data. But with three waves of data you essentially have the analysis being done on one wave of data by the time you finish with the lags and differences. So while you will see some A-B analyses with three-wave data, usually there are more waves to make the analysis more convincing.
- In addition, you cannot test the no autocorrelation restriction even with three waves, because the test is based on second-order autocorrelation in the differences, and you only have second order differences when there are four waves of data. So you can do A-B with three waves, but you can only test one of the crucial assumptions of A-B with four waves.

Endogenous Regressor Models

- Discussion of panel data from econometric tradition thus far has focused on models that attempt to take into account unit heterogeneity (i.e., stable unobservables, or the U_i term), and to estimate parameters that avoid inconsistency produced by U's potential correlation with *included* independent variables

$$(5) \qquad Y_{it} = \alpha + (\beta_1 Y_{it-1}) + \beta_2 X_{1it} + \beta_3 Z_{1i} + \ldots \beta_j X_{jit} + \beta_k Z_{ki} \ldots U_i + \varepsilon_{it}$$

- FE/FD models used b/c of possible correlation between time-varying independent variables X_{it} and U_i
- Arellano-Bond dynamic models used b/c of intrinsic correlation between Y_{it-1} and U_i in models with lagged dependent variable
- Hausman-Taylor models proposed to handle possible correlation between time-invariant independent variables Z_{it} and U_i

- All these models attempt to deal with problem of *endogeneity bias* caused by presence of U_i term:
  - OLS cannot distinguish the effects that properly belong to X, Z and/or Y_{t-1} and the effects that properly belong to U, as OLS procedures produce a regression line where the independent variables and the error term are uncorrelated by construction. Hence any joint correlation between X, Z and/or Y_{t-1} and U is "thought" by OLS to belong to the independent variables
- We know that the error term in pooled panel models is composed of two parts: U_i and ε_{it}. Can there also be endogeneity bias produced by correlation between the independent variables and ε_{it}, the idiosyncratic, time-specific component of the error?
  - Yes! Big potential problem!
  - How? Unstable omitted variables; reciprocal causality, measurement error
  - We know the solution by now! (IV methods!)

Exogenous and Endogenous Variables

- To frame discussion, useful to extend our earlier terminology about different kinds of variables, each with different relationships with their equation's error terms
- Variables may be:
  - "strictly exogenous"
  - "predetermined" or in a "sequentially exogenous" relationship; or
  - "endogenous"

- **"Strict exogeneity"** is when ε_{it}, is unrelated to **all** of the X_{it}, that is X_{it-1}, X_{it-2}, X_{it+1}, X_{it+2}, etc. This is mainly the sense of exogeneity that we have been using so far in the course. We can say under *strict exogeneity* that X_{it} at any *t* is unrelated to ε_{it} at any *t* – past, present, or future.
- **"Sequentially exogeneity"** is weaker. It is when ε_{it} is unrelated to the **current** X_{it}, and unrelated to all of the **past** X_{it-1}, X_{it-2}, X_{it-3}, etc. However, ε_{it} may be related to **future** values of X.
- Halaby (p.533) gives an example from a model predicting children's mental health (Y) from parents' marital status (X). If, as seems likely, some idiosyncratic aspect of children's mental health at time *t* (which is contained in ε_{it}) affects the parents' divorce decision at time *t+1*, then we have *sequential exogeneity*, and we say that X is a **predetermined** variable, but not a **strictly exogenous** variable.
- This is a case of **"endogenous selection"** into current "treatment" (parent's marital status) based on prior values of the dependent variable (Y_{t-1} – children's mental health)

- In the predetermined case, X at time *t* is unrelated to current ε and to all future ε, but it is related to the ε in the immediate past. This may result from a lagged effect of Y_{t-1} on X_t, because ε_{t-1} is one of the causes of Y_{t-1}, which then causes X_t. In these cases, there is an *intrinsic* relationship between ε_{t-1} and X_t, and X_t would then be considered to be a "predetermined" (not "strictly exogenous") variable.
- A fully **"endogenous"** variable is when ε_t is related to *current* X_t. By extension this means that ε_{t-1} is related to X_{t-1}, ε_{t-2} is related to X_{t-2}, etc.
- Simultaneity and measurement error lead variables to be *fully endogenous*, in this terminology, while omitted variables or unobserved shocks may lead to full endogeneity, or, depending on how you conceive the problem, the likely lag effect of the omitted factors, selection into treatment, etc., may leave a situation of "sequential exogeneity"

Solutions to Endogeneity Problems

- Depend on the kinds of variables we have in the model (strict versus sequential versus fully endogenous)

$$(6) \qquad Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 Z_{1i} + U_i + \varepsilon_{it}$$

- Let's assume that time-varying variable number 1 (X_{1it}) and ε_{it} are related for one or more of the reasons we just mentioned, but that time-varying variable number 2 (X_{2it}) is not related to ε_i at any time. Thus X_1 is a fully endogenous variable, and X_2 is a strictly exogenous variable. (Time invariant variable Z's relationship to ε_{it} is irrelevant, as we will see). How do we estimate this model?

- First, we have to deal with the U_i problem, and then we turn to the ε_{it} problem. So one possibility is our old friend the Fixed Effects (FE) transformation:

$$(7) \quad (Y_{it}-\bar{Y}_i) = \alpha + \beta_1(X_{1it}-\bar{X}_{1i}) + \beta_2(X_{2it}-\bar{X}_{2i}) + (\varepsilon_{it}-\bar{\varepsilon}_i)$$

with both the time-invariant Z and the time-invariant U_i swept out of the equation. But we cannot estimate this model using the normal FE procedures we used earlier, because the de-meaned X_1 term is related to the de-meaned error term (because X_1 is related to ε_{it}).

- We need an exogenous instrumental variable(s) for de-meaned X_1 that is **not** related to the de-meaned error term. Where to find this?

- In the dynamic panel models considered earlier, we used lagged values of X (or Y_{t-1} under some conditions). Can we apply the same logic and use lagged X_{1t-1} or the twice-lagged X_{1t-2} as instruments for the de-meaned endogenous X_1 in equation (7)?
- **ANSWER: NO!** Why not? Because X_{1t-1} and X_{1t-2} are related to ε_{it-1} and ε_{it-2}, respectively, and hence related to $\bar{\varepsilon}$ in (7). Hence *the lagged Xs are not appropriate instruments because they too are related to the error term of the FE equation.*
- Would the situation be any different if we relaxed the problem and assumed that X_1 was not fully endogenous, but rather "predetermined"? In that case we would think that lagged Y (and hence ε_{it-1}) was related to X_1, but that X_{1t} was not related to the contemporaneous error ε_{it}. Would this allow us to use lagged values of X_1 to estimate the FE equation?
- **ANSWER: NO!** Because all the values of lagged X_1 are *still* related to *some* past value of ε_{it}, and hence *still* related to $\bar{\varepsilon}$ in (7).

Morals of the Story (So Far)

- If you use the FE transformation to eliminate the U_i, then *all* instrumental variables must be **strictly exogenous**, unrelated to all past, current, or future values of ε_{it} so that they are unrelated to the average error term in (7). *This precludes the use of lagged values of endogenous or predetermined variables as instruments.*
- So IVs must be found from "outside" the system, just like in the cross-sectional case. This is hard (but not impossible) to do, just like in the cross-sectional case.
- This strategy is okay, but fails to fully exploit the advantage of panel data in providing supplemental (or primary) instruments under some conditions

Alternative IV Approach: Extend the FD Model

- First, use FD to "sweep out" U:

$$(8) \quad (Y_{it} - Y_{\text{it-1}}) = \beta_1 (X_{1it} - X_{1it-1}) + \beta_2 (X_{2it} - X_{2it-1}) + (\varepsilon_{\text{it}} - \varepsilon_{it-1})$$

- What about using lags of X_1 as instrumental variables now?
- **ANSWER: IT DEPENDS.**
- If X_1 is "predetermined," then X_{t-1} is unrelated to both components of the FD error term, and hence it may be used as an IV in (8).
- X_{t-1} is, however, related to ε_{t-2}, but the beauty of the FD transformation for this purpose is that we have only a finite number of lags of ε_t to deal with. So, in situations of *sequential exogeneity*, we may use the lagged value of X_1, the twice-lagged value of X_1, and even the lagged first difference $(X_{1t-1} - X_{1t-2})$ as instruments for $(X_{1t} - X_{1t-1})$ in the FD model.

- If X_1 is fully endogenous, then we cannot use X_{1t-1} as an instrument any longer. Why not? Because the error term in the FD model has ε_{t-1}, and endogeneity means that X_{1t-1} and ε_{t-1} are related. But what about X_{1t-2}? Sure! Twice-lagged X_1 is unrelated to both components of the FD error term, so we can use it as an instrument in an IV-TSLS estimation of (8).

$$(8) \qquad (Y_{it} - Y_{\text{it-1}}) = \beta_1 (X_{1it} - X_{1it-1}) + \beta_2 (X_{2it} - X_{2it-1}) + (\varepsilon_{\text{it}} - \varepsilon_{it-1})$$

- See? No relationship with the instruments to this equation's error term!
- So: lagged X (and farther back lags and differences) can be used as IV in situations of sequential exogeneity;

  twice lagged X (and farther back lags and differences) can be used as IVs in situations of full endogeneity
- All of this depends, again, on FD as the base model, not FE!!!

Notes

- The FD-IV estimators can therefore make use of lagged values of endogenous and predetermined Xs, so it is generally preferred over FE-IV. Under some conditions (e.g., large T) however, the FE bias in (7) declines considerably and you may be better off with FE-IV than with the FD-IV procedure.
- You need at least three waves of data to estimate models with fully endogenous variables.
- All of these procedures depend on the assumption of *no autocorrelation* in the ε_t.
- It is recommended to use the Sargan test we discussed earlier to examine whether the IVs that you have used are truly independent of the error term in the outcome equation. This is implemented in XTIVREG

- These principles apply to the dynamic panel models such as Arellano-Bond that we considered earlier as well. In XTABOND, you declare whether variables are strictly exogeneous, predetermined, or endogenous, and the appropriate lags are used as instruments. That means: lags of 0 for strictly exogenous variables (i.e. variables instrument for themselves whenever they are strictly exogenous), lags of 1 and back for predetermined variables, and lags of 2 and back for endogenous variables.
- It is always a good idea to have outside exogenous (or predetermined) variables available to add to the instrument set to get better precision of the proxy variable for the second stage regression. If you can find such variables, so much the better.

Autocorrelated Errors Model

- Another way to incorporate temporal dynamics: model autocorrelation in the idiosyncratic error term ε_{it}. Assumes that some other factors or forces are pushing the idiosyncratic unit-time error terms up or down over time, independent of the stable unobserved variables represented by U_i.
- What are they? Many possibilities, among them:
  - omitted exogenous shocks that linger for several time periods
  - omitted variables that may change over time
  - correlated measurement errors
- Example: if the DV is crime rate for US cities, such factors could be weather-related disturbances, idiosyncratic patterns of gang warfare, temporary spikes in drug prices, etc., all of which may affect crime rates over time, independently from both the main theoretical measured variables as well as a unmeasured stable unit effect for the given city.

The AR(1) Model

- Consider the original panel model with unit effects:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \ldots \beta_j X_{jit} + U_i + \varepsilon_{it}$$

- Previous assumptions:
  - U stable for each unit *i*
  - ε_{it} normally distributed with no autocorrelation. Value at one point in time is independent of values from the past or future
- If can't maintain this last assumption, many options for covariance structure of the errors (most of which we won't have time to talk about). One very popular one is where a given error term ε_{it} is dependent on its previous time period's value, as in:

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + v_{it}$$

- where v_{it} is assumed to be normally distributed, mean of 0, constant variance σ_v^2 and independent (non-autocorrelated)
- This is known as the AR(1) model for the disturbances: autoregressive with a time lag of 1
- This is the same model we have been considering, but this time with an idiosyncratic disturbance term for case *i* at time *t* that is a function of its value at time *t-1*, with a regression coefficient of ϱ.
- There is an easy test available (from Woolridge 2002; Drukker 2003) to see if there may be autocorrelated disturbances in the data: compare the correlation of the residuals from the first difference (FD) model to its expected value under the null hypothesis (no autocorrelation) of -.5. This test is implemented in a STATA add-on module called XTSERIAL. If it is significant, you can reject the null. However, it does not mean that the form of the autocorrelation necessarily **is** first-order, as in equation 2. It simply means it is "not no autocorrelation." (It does not necessarily rule out a lagged DV model either).
- Note: As noted above, the difference model implies that

$$Corr(\varepsilon_{it} - \varepsilon_{it-1}),(\varepsilon_{it-1} - \varepsilon_{it-2}) = -.5$$

under the hypothesis of no autocorrelation (and equal variance) in the ε_{it}

Estimation of the AR(1) Model

- The first-order autocorrelated disburbance model is estimated in non-panel applications via (feasible) GLS procedures: we get an estimate of ϱ, then we weight the equation for Y at time *t-1* by ϱ, then subtract the time *t-1* equation from the time *t* equation:

$$(a)\ \ Y_t = \alpha + \beta_1 X_{1t} + \ldots \beta_J X_{Jt} + \varepsilon_t \qquad \text{with } \varepsilon_t = \rho\varepsilon_{t-1} + v_{it}$$

$$(b)\ \ \rho Y_{t-1} = \alpha\rho + \rho\beta_1 X_{1t-1} + \ldots \rho\beta_J X_{Jt-1} + \rho\varepsilon_{t-1}$$

$$(c)\ \ Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta_1(X_{1t} - \rho X_{1t-1}) + \beta_1(X_{Jt} - \rho X_{Jt-1}) + \varepsilon_t - \rho\varepsilon_{t-1}$$

- Only difference for panels: For FE, we first subtract unit-level means to get rid of "U", then estimate ϱ with this "fixed effects version" via "Prais-Winsten" regression, then transform de-meaned Y and de-meaned X with ϱ as above and re-estimate
- For RE: use more complex "Baltagi-Wu GLS Estimator": see Stata Longitudinal Modeling Manual

- **Note**: we lose the first wave of observations with fixed effects AR(1), so need at least three waves of data for this.
- **Note**: AR(1) is a plausible structure for the idiosyncratic disturbances, but it is not the only possible one. It predicts that the correlation between errors at one wave apart will be ϱ, two waves apart will be ϱ^2, three waves apart will be ϱ^3, etc. When errors do not decline geometrically like this, alternative patterns may be modeled instead. There may be an additional effect from previous lags, e.g., as in an AR(2) model, or there may be a "moving average" pattern to the residuals, whereby the error at one time is a function of some parameter (say, δ_t), and, in an MA(1) process, some additional fraction of δ at t-1, in an MA(2) process some additional fraction of δ at t-2, etc., with the fraction of δ at each time being an additional parameter to estimate. Or, the pattern of residual correlations may be different.
- See Singer and Willett, chapter 7 for some additional alternatives as applied to the RE model; we will discuss some of these later.

Time as an IV

- Assume either that there is a general effect of time on all units OR that individual units are modeled as having distinct trajectories in *Y* over time.
- Temporal dependence in *Y* is due not to the presence of omitted variables, nor to the previous time period's response, but rather due to the fact that units are *growing* or *changing* in possibly predictable ways on the dependent variable over time.
- Here we are getting into "growth modeling". In this tradition, which is common in the hierarchical or multilevel modeling approach to longitudinal analysis, we try to model the impact of time on *Y* and explain why some cases grow at faster or slower rates than others. We will take up these models in a few weeks.