Chapter 6 Stochastic Regressors

- 6.1 Stochastic regressors in non-longitudinal settings
- 6.2 Stochastic regressors in longitudinal settings
- 6.3 Longitudinal data models with heterogeneity terms and sequentially exogenous regressors
- 6.4 Multivariate responses
- 6.5 Simultaneous equation models with latent variables
- Appendix 6A – Linear projections

6.1 Stochastic regressors in non-longitudinal settings

- 6.1.1 *Endogenous stochastic regressors*
- 6.1.2 *Weak and strong exogeneity*
- 6.1.3 *Causal effects*
- 6.1.4 *Instrumental variable estimation*

- This section introduces stochastic regressors by focusing on purely cross-sectional and purely time series data.
- It reviews the non-longitudinal setting, to provide a platform for the longitudinal data discussion.

Non-stochastic explanatory variables

- Traditional in the statistics literature
- Motivated by designed experiments
  - **X** represents the amount of fertilizer applied to a plot of land.
- However, for survey data, it is natural to think of random regressors. Observational data ????
- On the one hand, the study of stochastic regressors subsumes that of non-stochastic regressors.
  - With stochastic regressors, we can always adopt the convention that a stochastic quantity with zero variance is simply a deterministic, or non-stochastic, quantity.
- On the other hand, we may make inferences about population relationships conditional on values of stochastic regressors, essentially treating them as fixed.

Endogenous stochastic regressors

- An *endogenous variable* is one that fails an exogeneity requirement – more later.
- It is customary in economics
  - to use the term endogenous to mean a variable that is determined within an economic system whereas
  - an exogenous variable is determined outside the system.
  - Thus, the accepted econometric/statistic usage differs from the general economic meaning.
- If $(\mathbf{x}_i, y_i)$ are i.i.d, then imposing the conditions

$$\mathrm{E}\,(y_i \mid \mathbf{x}_i\,) = \mathbf{x}_i' \,\boldsymbol{\beta} \text{ and } \mathrm{Var}\,(y_i \mid \mathbf{x}_i\,) = \sigma^2$$

are sufficient to estimate parameters.

- Define $\varepsilon_i = y_i - \mathbf{x}_i' \,\boldsymbol{\beta}$, and write the first condition as

$$\mathrm{E}\,(\varepsilon_i \mid \mathbf{x}_i\,) = 0.$$

- Interpret this to mean that ε_i and $\mathbf{x}_i$ are uncorrelated.

Assumptions of the Linear Regression Model with Strictly Exogenous Regressors

Wish to analyze the effect of all of the explanatory variables on the responses. Thus, define $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and require

- SE1. $\mathrm{E}\,(y_i \mid \mathbf{X}) = \mathbf{x}_i' \boldsymbol{\beta}$.
- SE2. $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ are stochastic variables.
- SE3. $\mathrm{Var}\,(y_i \mid \mathbf{X}) = \sigma^2$.
- SE4. $\{y_i \mid \mathbf{X}\}$ are independent random variables.
- SE5. $\{y_i\}$ is normally distributed, conditional on $\{\mathbf{X}\}$.

Usual Properties Hold

- Under SE1-SE4, we retain most of the desirable properties of our ordinary least square estimators of $\boldsymbol{\beta}$. These include:
  - the unbiasedness and
  - the Gauss-Markov property of ordinary least square estimators of $\boldsymbol{\beta}$.
- If, in addition, SE5 holds, then the usual t and F statistics have their customary distributions, regardless as to whether or not $\mathbf{X}$ is stochastic.
- Define the disturbance term to be $\varepsilon_i = y_i - \mathbf{x}_i' \boldsymbol{\beta}$ and
  - write SE1 as $\mathrm{E}(\varepsilon_i \mid \mathbf{X}) = 0$
  - is known as *strict exogeneity* in the econometrics literature.

Some Alternative Assumptions

- Regressors are said to be *predetermined* if
- SE1p. $\mathrm{E}\,(\varepsilon_i\ \mathbf{x}_i) = \mathrm{E}\,(\,(y_i - \mathbf{x}_i'\,\boldsymbol{\beta})\,\mathbf{x}_i) = \mathbf{0}$.
- The assumption SE1p is weaker than SE1.
  - SE1 does not work well with time-series data
  - SE1p is sufficient for consistent for consistency, not asymptotic normality.
- For asymptotic normality, we require a somewhat stronger assumption:
- SE1m. $\mathrm{E}\,(\,\varepsilon_i \mid \varepsilon_{i-1}, \ldots, \varepsilon_{1,}\ \mathbf{x}_i\,, \ldots, \mathbf{x}_1) = 0$ for all i .
- When SE1m holds, then $\{\varepsilon_i\}$ satisfies the requirements for a *martingale difference* sequence.
- Note that SE1m implies SE1p.

Weak and strong exogeneity

- For linear model exogeneity
  - We have considered *strict exogeneity* and *predeterminedness*.
  - Appropriately done in terms of conditional means.
  - It gives precisely the conditions needed for inference and is directly testable.
- Now we wish to generalize these concepts to assumptions regarding the entire distribution, not just the mean function.
  - Although stronger than the conditional mean versions, these assumptions are directly applicable to nonlinear models.
- We now introduce two new kinds of exogeneity, *weak* and *strong exogeneity*.

Weak exogeneity

- A set of variables are said to be *weakly exogenous* if, when we condition on them, there is no loss of information about the parameters of interest.
- Weak endogeneity is sufficient for efficient estimation.
- Suppose that we have random variables $(\mathbf{x}_1, y_1)$, …, $(\mathbf{x}_T, y_T)$ with joint probability density (or mass) function for $\mathrm{f}(y_1, \ldots, y_T, \mathbf{x}_1, \ldots, \mathbf{x}_T)$.
- By repeated conditioning, we write this as:

$$\mathrm{f}\left(y_1,\ldots,y_T,\mathbf{x}_1,\ldots,\mathbf{x}_T\right)=\prod_{t=1}^{T}\mathrm{f}\left(y_t,\mathbf{x}_t \mid y_1,\ldots,y_{t-1},\mathbf{x}_1,\ldots,\mathbf{x}_{t-1}\right)$$

$$=\prod_{t=1}^{T}\left\{\mathrm{f}\left(y_t \mid y_1,\ldots,y_{t-1},\mathbf{x}_1,\ldots,\mathbf{x}_t\right)\mathrm{f}\left(\mathbf{x}_t \mid y_1,\ldots,y_{t-1},\mathbf{x}_1,\ldots,\mathbf{x}_{t-1}\right)\right\}$$

Weak exogeneity

- Suppose that this joint distribution is characterized by vectors of parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ such that

$$\mathrm{f}\left(y_1,\ldots,y_T,\mathbf{x}_1,\ldots,\mathbf{x}_T\right)$$
$$=\left(\prod_{t=1}^{T}\mathrm{f}\left(y_t \mid y_1,\ldots,y_{t-1},\mathbf{x}_1,\ldots,\mathbf{x}_t,\boldsymbol{\theta}\right)\right)\left(\prod_{t=1}^{T}\mathrm{f}\left(\mathbf{x}_t \mid y_1,\ldots,y_{t-1},\mathbf{x}_1,\ldots,\mathbf{x}_{t-1},\boldsymbol{\psi}\right)\right)$$

- We can ignore the second term for inference about $\boldsymbol{\theta}$, treating the $\mathbf{x}$ variables as essentially fixed.
- If this relationship holds, then we say that the explanatory variables are *weakly exogenous*.

Strong Exogeneity

- Suppose, in addition, that

$$\mathrm{f}\left(\mathbf{x}_t \mid y_1,\ldots,y_{t-1},\mathbf{x}_1,\ldots,\mathbf{x}_{t-1},\boldsymbol{\psi}\right)=\mathrm{f}\left(\mathbf{x}_t \mid \mathbf{x}_1,\ldots,\mathbf{x}_{t-1},\boldsymbol{\psi}\right)$$

- that is, conditional on $\mathbf{x}_1, \ldots, \mathbf{x}_{t-1}$, that the distribution of $\mathbf{x}_t$ does not depend on past values of y, $y_1, \ldots, y_{t-1}$. Then, we say that $\{y_1, \ldots, y_{t-1}\}$ does not *Granger-cause* $\mathbf{x}_t$.
- This condition, together with weak exogeneity, suffices for *strong exogeneity*.
- This is helpful for prediction purposes.

Causal effects

- Researchers are interested in causal effects, often more so than measures of association among variables.
- Statistics has contributed to making causal statements primarily through randomization.
  - Data that arise from this random assignment mechanism are known as *experimental*.
  - In contrast, most data from the social sciences are *observational*, where it is not possible to use random mechanisms to randomly allocate observations according to variables of interest.
- Regression function measures relationships developed through the data gathering mechanism, not necessarily the relationships of interest to researchers.

Structural Models

- A *structural model* is a stochastic model representing a causal relationship, as opposed to a relationship that simply captures statistical associations.
- A sampling based model is derived from our knowledge of the mechanisms used to gather the data.
  - The sampling based model directly generates statistics that can be used to estimate quantities of interest
  - It is also known as an *estimable* model.

Causal Statements

- Causal statements are based primarily on substantive hypotheses in which the researcher carefully develops.
- Causal inference is theoretically driven.
- Causal processes cannot be demonstrated directly from the data; the data can only present relevant empirical evidence serving as a link in a chain of reasoning about causal mechanisms.
- Longitudinal data are much more useful in establishing causal relationships than (cross-sectional) regression data because, for most disciplines, the "causal" variable must precede the "effect" variables in time.
- Lazarsfeld and Fiske (1938) considered the effect of radio advertising on product sales.
  - Traditionally, hearing radio advertisements was thought to increase the likelihood of purchasing a product.
  - Lazarsfeld and Fiske considered whether those that bought the product would be more likely to hear the advertisement, thus positing a reverse in the direction of causality.
  - They proposed repeatedly interviewing a set of people (the 'panel') to clarify the issue.

Instrumental variable estimation

- Instrumental variable estimation is a general technique to handle problems associated with the disconnect between the structural model and a sampling based model.
- To illustrate, consider the linear model

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i ,$$

yet not all of the regressors are predetermined, $\mathrm{E}\,(\varepsilon_i\, \mathbf{x}_i) \neq \mathbf{0}$.

- Assume there a set of predetermined variables, $\mathbf{w}_i$, where
  - $\mathrm{E}\,(\varepsilon_i\ \mathbf{w}_i) = \mathbf{0}$ (predetermined)
  - $\mathrm{E}\,(\mathbf{w}_i\, \mathbf{w}_i')$ is invertible.
- An instrumental variable estimator of $\boldsymbol{\beta}$ is

$$\mathbf{b}_{IV} = (\mathbf{X}'\, \mathbf{P_W}\, \mathbf{X})^{-1}\, \mathbf{X}'\, \mathbf{P_W}\, \mathbf{y},$$

where $\mathbf{P_W} = \mathbf{W}\,(\mathbf{W}'\mathbf{W}\,)^{-1}\,\mathbf{W}'$ is a projection matrix and

$\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)'$ is the matrix of instrumental variables.

- Within $\mathbf{X}'\, \mathbf{P_W}$ is $\mathbf{X}'\, \mathbf{W} = \sum_i \mathbf{x}_i \mathbf{w}_i'$
- this sum of cross-products drives the calculation fo the correlation between **x** and **w**.

Omitted Variables Application

- The structural regression function as E $(y_i \mid \mathbf{x}_i, \mathbf{u}_i) =$ $\mathbf{x}_i' \boldsymbol{\beta} + \boldsymbol{\gamma}' \mathbf{u}_i$, where $\mathbf{u}_i$ represents unobserved variables.
- Example- Card (1995) wages in relation to years of education.
  - Additional control variables include years of experience (and its square), regional indicators, racial indicators and so forth.
  - The concern is that the structural model omits an important variable, the man's "ability" ($\mathbf{u}$), that is correlated with years of education.
  - Card introduces a variable to indicate whether a man grew up in the vicinity of a four-year college as an instrument for years of education.
  - Motivation - this variable should be correlated with education yet uncorrelated with ability.
  - Define $\mathbf{w}_i$ to be the same set of explanatory variables used in the structural equation model but with the vicinity

Instrumental Variables

- Additional applications include:
  - Measurement error problems
  - Endogeneity induced by systems of equations (Section 6.5).
- The choice of instruments is the most difficult decision faced by empirical researchers using instrumental variable estimation.
- Try to choose instruments that are highly correlated with the endogeneous explanatory variables.
- Higher correlation means that the bias as well as standard error of $\mathbf{b}_{IV}$ will be lower.

6.2. Stochastic regressors in longitudinal settings

- This section covers
  - No heterogeneity terms
  - Strictly exogeneous variables
- Both of these settings are relatively straightforward
  - Without heterogeneity terms, we can use standard (cross-sectional) methods
  - With strictly exogeneous variables, we can directly use the techniques described in Chapters 1-5

Longitudinal data models without heterogeneity terms

- *Assumptions of the Longitudinal Data Model with Strictly Exogenous Regressors*
- SE1. E $(y_{it} \mid \mathbf{X}) = \mathbf{x}_{it}' \boldsymbol{\beta}$.
- SE2. $\{\mathbf{x}_{it}\}$ are stochastic variables.
- SE3. Var $(\mathbf{y}_i \mid \mathbf{X}) = \mathbf{R}_i$.
- SE4. $\{\mathbf{y}_i \mid \mathbf{X}\}$ are independent random vectors.
- SE5. $\{\mathbf{y}_i\}$ is normally distributed, conditional on $\{\mathbf{X}\}$.

- Recall that $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ is the complete set of regressors over all subjects and time periods.

Longitudinal data models without heterogeneity terms

- No heterogeneity terms, but one can incorporate dependence among observations from the same subject with the $\mathbf{R}_i$ matrix (such as an autoregressive model or compound symmetry).
- These strict exogeneity assumptions do not permit lagged dependent variables, a popular approach for incorporating intra-subject relationships among observations.
  - However, one can weaken this to a pre-determined condition such as:
  - SE1p. E $(\varepsilon_{it}\ \mathbf{x}_{it}) = \mathrm{E}\left((y_{it} - \mathbf{x}_{it}'\ \boldsymbol{\beta})\ \mathbf{x}_{it}\right) = \mathbf{0}$.
- Without heterogeneity, longitudinal and panel data models have the same endogeneity concerns as the cross-sectional models.

Longitudinal data models with heterogeneity terms and strictly exogenous regressors

- From customary usage or a structural modeling viewpoint, it is often important to understand the effects of endogenous regressors when a heterogeneity term $\boldsymbol{\alpha}_i$ is present in the model.
- We consider the linear mixed effects model of the form

$$y_{it} = \mathbf{z}_{it}' \, \boldsymbol{\alpha}_i + \mathbf{x}_{it}' \, \boldsymbol{\beta} + \varepsilon_{it}$$

- and its vector version

$$\mathbf{y}_i = \mathbf{Z}_i \, \boldsymbol{\alpha}_i + \mathbf{X}_i \, \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \, .$$

- Define $\mathbf{X}^* = \{\mathbf{X}_1, \mathbf{Z}_1, \ldots, \mathbf{X}_n, \mathbf{Z}_n\}$ to be the collection of all observed explanatory variables and
- $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \ldots, \boldsymbol{\alpha}_n')'$ to be the collection of all subject-specific terms.

Assumptions of the Linear Mixed Effects Model with Strictly Exogenous Regressors Conditional on the Unobserved Effect

- SEC1. E $(\mathbf{y}_i \mid \boldsymbol{\alpha}, \mathbf{X}^*) = \mathbf{Z}_i\, \boldsymbol{\alpha}_i + \mathbf{X}_i\, \boldsymbol{\beta}$.
- SEC2. $\{\mathbf{X}^*\}$ are stochastic variables.
- SEC3. Var $(\mathbf{y}_i \mid \boldsymbol{\alpha}, \mathbf{X}^*) = \mathbf{R}_i$.
- SEC4. $\{\mathbf{y}_i\}$ are independent random vectors, conditional on $\{\boldsymbol{\alpha}\}$ and $\{\mathbf{X}^*\}$.
- SEC5. $\{\mathbf{y}_i\}$ is normally distributed, conditional on $\{\boldsymbol{\alpha}\}$ and $\{\mathbf{X}^*\}$.
- SEC6. E $(\boldsymbol{\alpha}_i \mid \mathbf{X}^*) = \mathbf{0}$ and Var $(\boldsymbol{\alpha}_i \mid \mathbf{X}^*) = \mathbf{D}$.
  Further, $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n\}$ are mutually independent, conditional on $\{\mathbf{X}^*\}$.
- SEC7. $\{\boldsymbol{\alpha}_i\}$ is normally distributed, conditional on $\{\mathbf{X}^*\}$.

Observables Representation of the Linear Mixed Effects Model with Strictly Exogenous Regressors Conditional on the Unobserved Effect

- SE1. $\mathrm{E}\,(\mathbf{y}_i \mid \mathbf{X}^*) = \mathbf{X}_i\,\boldsymbol{\beta}$.
- SE2. $\{\mathbf{X}^*\}$ are stochastic variables.
- SE3a. $\mathrm{Var}\,(\mathbf{y}_i \mid \mathbf{X}^*) = \mathbf{Z}_i\,\mathbf{D}\,\mathbf{Z}_i' + \mathbf{R}_i$.
- SE4. $\{\mathbf{y}_i\}$ are independent random vectors, conditional on $\{\mathbf{X}^*\}$.
- SE5. $\{\mathbf{y}_i\}$ is normally distributed, conditional on $\{\mathbf{X}^*\}$.

Strictly Exogenous Regressors Conditional on the Unobserved Effect

- These assumptions are stronger than strict exogeneity.
- For example, note that $\mathrm{E}\,(\mathbf{y}_i \mid \boldsymbol{\alpha}, \mathbf{X}^*) = \mathbf{Z}_i\, \boldsymbol{\alpha}_i + \mathbf{X}_i\, \boldsymbol{\beta}$ and $\mathrm{E}\,(\boldsymbol{\alpha}_i \mid \mathbf{X}^*) = \mathbf{0}$ together imply that

$$\mathrm{E}\,(\mathbf{y}_i \mid \mathbf{X}^*) = \mathrm{E}\,(\mathrm{E}\,(\mathbf{y}_i \mid \boldsymbol{\alpha}, \mathbf{X}^*) \mid \mathbf{X}^*)$$
$$= \mathrm{E}\,(\mathbf{Z}_i\, \boldsymbol{\alpha}_i + \mathbf{X}_i\, \boldsymbol{\beta} \mid \mathbf{X}^*) = \mathbf{X}_i\, \boldsymbol{\beta}\ .$$

- That is, we require strict exogeneity of the disturbances ($\mathrm{E}\,(\boldsymbol{\varepsilon}_i \mid \mathbf{X}^*) = \mathbf{0}$) and
- that the unobserved effects ($\boldsymbol{\alpha}$) are uncorrelated with the disturbance terms ($\mathrm{E}\,(\varepsilon_i\, \boldsymbol{\alpha}') = \mathbf{0}$).

Example - Taxpayers

Demographic Characteristics

- MS - taxpayer's marital status.
- HH - head of household
- DEPEND - number of dependents claimed by the taxpayer.
- AGE - age 65 or over.

Economic Characteristics

- LNTPI - natural logarithm of the sum of all positive income line items on the return, in 1983 dollars..
- MR - marginal tax rate. It is computed on total personal income less exemptions and the standard deduction.
- EMP - Self-employed binary variable.
- PREP - indicates the presence of a paid preparer.
- LNTAX - natural logarithm of the tax liability, in 1983 dollars. This is the response variable of interest.

Example - Taxpayers

- Because the data was gathered using a random sampling mechanism, we can interpret the regressors as stochastic.
- Demographics, and probably EMP, can be safely argued as strictly exogenous.
- $LNTAX_t$ should not affect $LNTPI_t$, because LNTPI is the sum of positive income items, not deductions.
- Tax preparer variable (PREP)
  - it may be reasonable to assume that the tax preparer variable is predetermined, although not strictly exogenous.
  - That is, we may be willing to assume that this year's tax liability does not affect our decision to use a tax preparer because we do not know the tax liability prior to this choice, making the variable predetermined.
  - However, it seems plausible that the prior year's tax liability will affect our decision to retain a tax preparer, thus failing the strict exogeneity test.

Taxpayer Model -With heterogeneity terms

- Consider the error components model
  - We interpret the heterogeneity terms to be unobserved subject-specific (taxpayer) characteristics, such as "ability," that would influence the expected tax liability.
  - One needs to argue that the disturbances, representing "unexpected" tax liabilities, are uncorrelated with the unobserved effects.
- Moreover, Assumption SEC6 employs the condition that the unobserved effects are uncorrelated with the observed regressor variables.
  - One may be concerned that individuals with high earnings potential who have historically high levels of tax liability (relative to their control variables) may be more likely to use a tax preparer, thus violating this assumption.

Fixed effects estimation

- If one is concerned with Assumption SEC6, then a solution may be fixed effects estimation (even when we believe in a random effects model formulation).
- Intuitively, this is because the fixed effects estimation procedures “sweep out” the heterogeneity terms
  - they do not rely on the assumption that they are uncorrelated with observed regressors.
- Some analysts prefer to test the assumption of correlation between unobserved and observed effects by examining the difference between these two estimators – “Hausman test” Section 7.2.

6.3 Longitudinal data models with heterogeneity terms and sequentially exogenous regressors

- The assumption of strict exogeneity, even when conditioning on unobserved heterogeneity terms, is limiting.
  - Strict exogeneity rules out current values of the response (y_{it}) feeding back and influencing future values of the explanatory variables (such as $\mathbf{x}_{i,t+1}$).
- An alternative assumption introduced by Chamberlain (1992) allows for this feedback.
- We say that the regressors are *sequentially exogenous conditional on the unobserved effects* if

$$\mathrm{E}\,(\,\varepsilon_{it} \mid \alpha_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}\,) = 0.$$

or (in the error components model)

$$\mathrm{E}\,(\,y_{it} \mid \alpha_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}\,) = \alpha_i + \mathbf{x}_{it}'\,\boldsymbol{\beta} \text{ for all } i, t.$$

- After controlling for α_i and $\mathbf{x}_{it}$, no past values of regressors affect the expected value of y_{it}.

Lagged dependent variable model

- This formulation allows us to consider lagged dependent variables as regressors

$$y_{it} = \alpha_i + \gamma\, y_{i,t\text{-}1} + \mathbf{x}_{it}'\,\boldsymbol{\beta} + \varepsilon_{it}\ ,$$

- This is sequentially exogenous conditional on the unobserved effects
  - To see this, use the set of regressors $\mathbf{o}_{it} = (1,\ y_{i,t\text{-}1},\ \mathbf{x}_{it}')'$ and $\mathrm{E}\ (\varepsilon_{it} \mid \alpha_i, y_{i,1}, \ldots, y_{i,t\text{-}1}, \mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,t}) = 0$.
  - The explanatory variable $y_{i,t\text{-}1}$ is not strictly exogenous so that the Section 6.2.2 discussion does not apply.

Estimation difficulties of lagged dependent variable model

- Estimation of the lagged dependent variable model is difficult because the parameter γ appears in both the mean and variance structure.

$$\mathrm{Cov}\,(y_{it}, y_{i,t\text{-}1}) = \mathrm{Cov}\,(\alpha_i + \gamma y_{i,t\text{-}1} + \mathbf{x}_{it}'\,\boldsymbol{\beta} + \varepsilon_{it}\,, y_{i,t\text{-}1})$$
$$= \mathrm{Cov}\,(\alpha_i, y_{i,t\text{-}1}) + \gamma\,\mathrm{Var}\,(y_{i,t\text{-}1}).$$

and

$$\mathrm{E}\,y_{it} = \gamma\ \mathrm{E}\,y_{i,t\text{-}1} + \mathbf{x}_{it}'\,\boldsymbol{\beta} = \gamma\ (\gamma\ \mathrm{E}\,y_{i,t\text{-}2} + \mathbf{x}_{i,t\text{-}1}'\,\boldsymbol{\beta}\,) + \mathbf{x}_{it}'\,\boldsymbol{\beta}$$
$$= \ldots = (\mathbf{x}_{it}' + \gamma\,\mathbf{x}_{i,t\text{-}1}' + \ldots + \gamma^{t\text{-}2}\,\mathbf{x}_{i,2}')\boldsymbol{\beta} + \gamma^{t\text{-}1}\,\mathrm{E}\,y_{i,1}\,.$$

- Thus, $\mathrm{E}\,y_{it}$ clearly depends on γ.
- Moreover, special estimation techniques are required.

First differencing technique

- First differencing proves to be a suitable device for handling certain types of endogenous regressors.
- Taking first differences of the lagged dependent variable model yields

$$y_{it} - y_{i,t\text{-}1} = \gamma\,(y_{i,t\text{-}1} - y_{i,t\text{-}2}) + \varepsilon_{it} - \varepsilon_{i,t\text{-}1}\ ,$$

eliminating the heterogeneity term.

- Ordinary least squares estimation using first differences (without an intercept term) yields an unbiased and consistent estimator of γ.
- First differencing can also fail - see the “feedback” example.

Example – *Feedback*

- Consider the error components $y_{it} = \alpha_i + \mathbf{x}_{it}' \boldsymbol{\beta} + \varepsilon_{it}$ where $\{\varepsilon_{it}\}$ are i.i.d.
- Suppose that the current regressors are influenced by the "feedback" from the prior period's disturbance through the relation $\mathbf{x}_{it} = \mathbf{x}_{i,t\text{-}1} + \upsilon_i \, \varepsilon_{i,t\text{-}1}$, where $\{\upsilon_i\}$ is an i.i.d.
- Taking differences of the model, we have

$$\Delta y_{it} = y_{it} - y_{i,t-1} = \Delta \mathbf{x}_{it}' \boldsymbol{\beta} + \Delta\varepsilon_{it}$$

where $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$ and $\Delta\mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{i,t-1} = \upsilon_i \, \varepsilon_{i,t-1}$.

- The ordinary least squares estimator of $\boldsymbol{\beta}$ are asymptotically biased.
  – Due to the correlation between $\Delta\mathbf{x}_{it}$ and $\Delta\varepsilon_{it}$.

Transform + instrumental variable estimation

- By a transform, we mean first differencing or fixed effects, to sweep out the heterogeneity.
- Assume balanced data and that the responses follow the model equation

$$y_{it} = \alpha_i + \mathbf{x}_{it}' \boldsymbol{\beta} + \varepsilon_{it} \,,$$

yet the regressors are potentially endogenous.
- Also assume that the current disturbances are uncorrelated with current as well as past instruments.
- Time-constant heterogeneity parameters are handled via sweeping out their effects,
  - let $\mathbf{K}$ be a $(T-1) \times T$ upper triangular matrix such that $\mathbf{K}\,\mathbf{1} = \mathbf{0}$.

- Thus, the transformed system is

$$\mathbf{K}\,\mathbf{y}_i = \mathbf{K}\,\mathbf{X}_i'\,\boldsymbol{\beta} + \mathbf{K}\,\boldsymbol{\varepsilon}_i\ ,$$

- Could use first differences

$$\mathbf{K}_{FD} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

- So that

$$\mathbf{K}_{FD}\mathbf{y} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{T-1} \\ y_T \end{pmatrix} = \begin{pmatrix} y_2 - y_1 \\ y_3 - y_2 \\ \vdots \\ y_T - y_{T-1} \end{pmatrix}$$

- Arrellano and Bover (1995) recommend

$$\mathbf{K}_{FOD} = \text{diag}\begin{pmatrix}\sqrt{\frac{T}{T-1}} & \cdots & \sqrt{\frac{1}{2}}\end{pmatrix}\begin{pmatrix} 1 & -\frac{1}{T-1} & -\frac{1}{T-1} & \cdots & -\frac{1}{T-1} & -\frac{1}{T-1} & -\frac{1}{T-1} \\ 0 & 1 & -\frac{1}{T-2} & \cdots & -\frac{1}{T-2} & -\frac{1}{T-2} & -\frac{1}{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}$$

- Defining $\boldsymbol{\varepsilon}_{i,FOD} = \mathbf{K}_{FOD}\, \boldsymbol{\varepsilon}_i$, the tth row is

$$\varepsilon_{it,FOD} = \sqrt{\frac{T-t}{T-t+1}}\left(\varepsilon_{it} - \frac{1}{T-t}\left(\varepsilon_{i,t+1} + \ldots + \varepsilon_{i,T}\right)\right)$$

- These are known as “forward orthogonal deviations.” They are used in time series – have slightly better properties.

- To define the instrumental variable estimator, let $\mathbf{W}_i$ be a block diagonal matrix with the *t*th block given by

$$(\mathbf{w}_{1,i1}{}' \ \mathbf{w}_{2,i1}{}' \ \ldots \ \mathbf{w}_{2,it}{}').$$

- That is, define

$$\mathbf{W}_i^* = \begin{pmatrix} \mathbf{w}'_{1,i1} \ \mathbf{w}'_{2,i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{w}'_{1,i1} \ \mathbf{w}'_{2,i1} \quad \mathbf{w}'_{2,i2}\right) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \left(\mathbf{w}'_{1,i1} \ \ \mathbf{w}'_{2,i1} \quad \mathbf{w}'_{2,i2} \quad \cdots \quad \mathbf{w}'_{2,i,T_i-1}\right) \end{pmatrix}$$

- This implies E $\mathbf{W}_i{}' \ \mathbf{K} \ \boldsymbol{\varepsilon}_i = \mathbf{0}$, our sequentially exogeneity assumption.

The estimator

- We define the instrumental variable estimator as

$$\mathbf{b}_{IV} = \left(\mathbf{M}'_{WX}\boldsymbol{\Sigma}_{IV}^{-1}\mathbf{M}_{WX}\right)^{-1}\mathbf{M}'_{WX}\boldsymbol{\Sigma}_{IV}^{-1}\mathbf{M}_{Wy}$$

- where $\mathbf{M}_{WX} = \sum_{i=1}^{n}\mathbf{W}_i'\mathbf{K}\mathbf{X}_i$ $\quad \mathbf{M}_{Wy} = \sum_{i=1}^{n}\mathbf{W}_i'\mathbf{K}\mathbf{y}_i$

- And $\boldsymbol{\Sigma}_{IV} = \mathrm{E}\left(\mathbf{W}_i'\,\mathbf{K}\,\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{K}'\,\mathbf{W}_i\right)$

- Estimate $\boldsymbol{\Sigma}_{IV}$ via two-stage least squares

Feedback Example

- Recall the relation $\mathbf{x}_{it} = \mathbf{x}_{i,t\text{-}1} + \upsilon_i\, \varepsilon_{i,t\text{-}1}$, .
- A natural set of instruments is to choose $\mathbf{w}_{it} = \mathbf{x}_{it}$.
- For simplicity, use the first difference transform.
- With these choices, the *t*th block of $\mathrm{E}\, \mathbf{W}_i{}'\, \mathbf{K}_{FD}\, \boldsymbol{\varepsilon}_i$ is

$$\mathrm{E}\begin{pmatrix} \mathbf{x}_{i1} \\ \vdots \\ \mathbf{x}_{it} \end{pmatrix}\left(\varepsilon_{i,t+1} - \varepsilon_{i,t}\right) = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

- so the sequentially exogeneity assumption is satisfied.

Taxpayer Example

- We suggested that a heterogeneity term may be due to an individual's earning potential
  - this may be correlated with the variable that indicates use of a professional tax preparer.
- Moreover, there was concern that tax liabilities from one year may influence the choice in subsequent tax year's choice of whether or not to use a professional tax preparer.
- If this is the case, then the instrumental variable estimator provides protection against this sequential endogeneity concern.

6.4 Multivariate responses

- 6.4.1 Multivariate regressions
- 6.4.2 Seemingly unrelated regressions
- 6.4.3 Simultaneous equations models
- 6.4.4 Systems of equations with error components

6.5 Simultaneous-Equations Models with Latent Variables

- 6.5.1 Cross-Sectional Models
- 6.5.2 Longitudinal Data Applications