Chapter 6

Extension to Longitudinal Data

6.1 Introduction

So far, several cross-section experimental and non-experimental estimators have been reviewed. These estimators produce plausible estimates under specific assumptions outlined in the previous chapters. The experimental estimators seek to evaluate mainly the mean effect of Treatment on the Treated, given that one randomization takes place. The various non-experimental estimators account for estimation of structural models apart from the various mean impacts. Two kinds of estimation methods are dominant in the econometric literature of the latter estimators. These are the Maximum Likelihood and Heckman's 2-step procedures.

Here, we recast our discussion to cover the experimental and non-experimental approach in longitudinal studies. Verbeke and Molenberghs (1997) define longitudinal study as the analysis of data collected at different points of time for each unit of a population. Three different types of longitudinal studies occur in practice – *trend*, *cohort* and *panel* studies. In a *trend* study, data from samples of different groups of people at different points in time, but from the same sample population, are collected. In a *cohort* study, subjects who presently have a certain condition and/or receive a particular treatment are followed over time and compared with another group who are not affected by the condition under investigation.

Panel studies are different from the ones described above. They measure the same sample of respondents at different points in time. Unlike trend studies, panel can reveal both net change and cross change in the dependent variable. Hsiao (1990) mentions that these studies can reveal long-term shifting attitudes and patterns of behavior that might go unnoticed in one-shot case studies (e.g. cross-sectional studies). Depending on the purpose of the study, researchers can use either a continuous panel, consisting of members who report specific attitudes or behavior patterns on regular basis, or an interval panel, whose members agree to complete a certain number of measurement instruments only when the information is needed. In general panel studies provide data suitable for sophisticated statistical analysis and might enable researcher to predict cause-effect relationships. For specific examples the reader is referred to Verbeke and Molenberghs (1997, pages 155- 189).

The econometric literature on evaluation of social programs in a long-term framework is focused on panel studies. In the following paragraphs, we discuss methods of causal inference, first, for the experimental case and then in terms of the non-experimental, Maximum Likelihood and 2-step, procedures. The advantages and disadvantages of these methods are outlined and specification tests for model selection decisions are proposed.

6.2 Panel Data – The Experimental Case

An experimental evaluation procedure entails making certain comparisons between treated and untreated individuals. The most commonly used parameter that reflects these comparisons is the mean effect of treatment on the treated $E\left(Y_{1i}^{r} - Y_{0i}^{r} \middle| D_{i}^{r} = 1, X_{i}^{r}\right)$, where, as before, "*r*" denotes the observations under randomization. While data on $E\left(Y_{1i}^{r} \middle| D_{i}^{r} = 1, X_{i}^{r}\right)$ are available from program participants, the counterfactual $E\left(Y_{0i}^{r} \middle| D_{i}^{r} = 1, X_{i}^{r}\right)$ is not simple to be estimated.

Assuming access to cross-sectional data, the estimation of the counterfactual mean requires specific statistical methods. By randomization of participants in treatment and control groups, prior receiving the services of the program or at eligibility stage, a program trainee is paired with an "otherwise comparable" person who does not receive treatment. However, when the analyst has access to panel data, other estimators can be applied to evaluate a social program. These estimators do not require a straight comparison between a treatment and a control group at a specific time period. Instead, the data themselves exploit the intuitively appealing idea that persons can be in both states at different times, and that outcomes measured in one state at one time are good proxies for outcomes in the same state at other times, at least for the non-treatment state.

In the sequel, we describe two panel data estimators and adapt them in an experimental framework. These are the "before-after" and the "difference-in-differences" estimators. We discuss the assumptions that must be invoked to result in adequate estimates of the program effects to individuals as well as various applications on real data.

6.2.1 The Before-After estimator

To formulate this estimator we recall again the t script. Without loss of generality, let us denote by $Y_{1i}^r(t)$ the post-program earnings and by $Y_{1i}^r(t')$ the pre-program earnings for a treated person. The corresponding outcomes for the controls are $Y_{0i}^r(t)$ and $Y_{0i}^r(t')$, respectively. For simplicity, assume that program participation occurs only at a single time period. The before-after estimator uses preprogram earnings $Y_{0i}^r(t')$ to proxy the treatment state in post-program period. Under the assumption

$$E\left(Y_{0i}^r(t) - Y_{0i}^r(t') \middle| D_i^r(t) = 1, X_i^r(t)\right) = 0 \tag{6.1}$$

and that the trainees and controls have the same permanent components in earnings, that is their pre-training earnings are equal, the *before–after estimator* (BA) is given by:

$$\begin{aligned}
&E\left(\Delta_i^r(t) \middle| D_i^r(t) = 1, X_i^r(t)\right) \\
&= E\left(Y_{1i}^r(t) - Y_{0i}^r(t) \middle| D_i^r(t) = 1, X_i^r(t)\right) \\
&= E\left[\left(Y_{1i}^r(t) - Y_{0i}^r(t')\right) - \left(Y_{0i}^r(t) - Y_{0i}^r(t')\right) \middle| D_i^r(t) = 1, X_i^r(t)\right]
\end{aligned}$$

$$= E\left(Y_{1i}^r(t) - Y_{0i}^r(t') \middle| D_i^r(t) = 1, X_i^r(t)\right) - E\left(Y_{0i}^r(t) - Y_{0i}^r(t') \middle| D_i^r(t) = 1, X_i^r(t)\right) \tag{6.2}$$

$$= E\left(Y_{1i}^r(t) \middle| D_i^r(t) = 1, X_i^r(t)\right) - E\left(Y_{0i}^r(t') \middle| D_i^r(t) = 1, X_i^r(t)\right) \tag{6.3}$$

since the second term in equation (6.2) equals zero by assumption (6.1). This term denotes the approximation error and when (6.1) is satisfied difference (6.3) estimates adequately the BA parameter. Note that evaluation of BA does not require sampling the same persons in periods t and t′. One needs just persons from the same populations. In this way the major drawback of panel data, that is the inability to sample the same persons across subsequent years from many periods, is solved.

The major drawback of this estimator is its reliance to assumption (6.1). This assumption requires that among participants, the mean outcome in the non-treatment state is the same in t and t′. Violations of this assumption due to economic changes, affect negatively the estimator. An example of this problem is provided by a work of Ashenfelter (1978) who observes that prior to enrollment in a training program, participants experience a decline in their earnings. He called this phenomenon *Ashenfelter's dip* and it is a common feature of the pre-program earnings of participants (treated and non-treated) in training programs. If this decline in earnings is transitory and the dip is eventually restored, even in the absence of treatment, and if period t′ falls in the period of transitorily low earnings, then the approximation error will not average out and assumption (6.1) will be violated. On the other hand, if the decline is permanent, the before-after estimator is unbiased for the parameter of interest.

According to Heckman and Smith (1999), Ashenfelter's dip with transitory decline occurs very often in training program evaluations. A common way to solve this problem requires knowledge of many periods' pre-program data. If the program takes place in period t′ + 1 and Ashenfelter's dip occurs in period t′, an appealing solution is to extrapolate from the periods before t′ to generate the counterfactual state in period t. Given the absence of extrapolation errors and if it is safe to assume that such errors average out to zero across persons in period t, one can estimate by averaging and replace the biased data of period t′.

6.2.2 The Difference-in-Differences estimator

When the above problems cannot be solved by using past data, other estimators have to be considered. Another widely used approach to estimate the effect of Treatment on the Treated is the Difference-in-Differences estimator. This estimator is based on the previous parameter. Their difference is placed on that Difference-in-Differences approach subtracts not only participants' mean pre-program earnings from the mean of their post-program earnings but also non-participants' mean pre-program earnings from the mean of their post-program earnings. Analogously to the previous estimator, it must be satisfied:

$$E\left(Y_{0i}^{r}(t)-Y_{0i}^{r}(t')\middle|D_{i}^{r}(t)=1,X_{i}^{r}(t)\right)=E\left(Y_{0i}^{r}(t)-Y_{0i}^{r}(t')\middle|D_{i}^{r}(t)=0,X_{i}^{r}(t)\right) \qquad (6.4)$$

that is the pre-program earnings of treated and untreated persons should be equal in order to estimate the parameter without bias. The *Difference-in-Differences* estimator (DD) is given by:

$$\begin{aligned} &E\left(\Delta_{i}^{r}(t)\middle|D_{i}^{r}(t)=1,X_{i}^{r}(t)\right) \\ &=E\left(Y_{1i}^{r}(t)-Y_{0i}^{r}(t')\middle|D_{i}^{r}(t)=1,X_{i}^{r}(t)\right)-E\left(Y_{0i}^{r}(t)-Y_{0i}^{r}(t')\middle|D_{i}^{r}(t)=0,X_{i}^{r}(t)\right) \qquad (6.5)\end{aligned}$$

Assumption (6.4) is considered as weaker than assumption (6.1) since it does not require equality to zero for any term. Equality of the relative measurement errors, so that the second term in (6.5) accounts for the measurement errors of the first terms of (6.5), suffices. Again, here, the analyst needs just to sample persons from the same populations.

Ashenfelter's dip can also be considered as a problem in this case, but for completely different reasons from the ones described before. Heckman and Smith (1998) indicate that if $Y_i(t')$*'s* are measured at the time of transitory earnings dip and if non-participants *do not* experience a dip, then assumption (6.5) is violated because the time path of no-program earnings between t′ and t will be different between participants and non-participants.

Although panel data are widely regarded as panacea for selection and simultaneity problems in recent experimental studies, Heckman and Robb (1985a) and Heckman, LaLonde and Smith (1999) claim that there is no need to use panels to identify the impact of training on earnings if assumptions (6.1) and (6.4) hold. Similar obtained estimators, based on cost effective repeated cross-section data for unrelated persons, estimate the same parameters. An analogous statement also holds in the non-experimental case for the various estimators that provided and analyzed below on this chapter.

6.3 Panel Data – The Non-Experimental Case

Similar to the experimental case, program evaluation can be approached in panel studies by applying econometric procedures. Several estimators have been proposed. To be implemented, they require access in either panel data on outcomes measured before and after program implementation or in repeated cross-section data from the same population where at least one cross-section is from a period prior to the program. Again here assume that post-earnings are measured in period t and the pre-program earnings at period t′ while, again, program participation occurs only at a single time period.

In terms of the conventional model of Heckman (1979) and its assumption of normality for the error terms, the (mixed) selection bias model in a panel data framework is formulated as:

$$Y_i^{(1)}(t) = X_i(t)\beta + u_i(t) \tag{6.6}$$

$$D_i^{(l)}(t) = Z_i(t)\gamma + v_i(t) \tag{6.7}$$

$$D_i(t) = 1 \text{ if } D_i^{(l)}(t) \succ 0 \tag{6.8}$$

$$Y_i(t) = Y_i^{(l)}(t) \times D_i(t) \tag{6.9}$$

The new notation here is that the random errors $u_i(t)$ and $v_i(t)$ of the primary and selection equation respectively, are decomposed into:

$$\begin{aligned} u_i(t) &= \varepsilon_i + e\ (t) + \xi_i\ (t) \\ v_i(t) &= n_i + h\ (t) + \omega_i\ (t) \end{aligned} \tag{6.10}$$

where (ε_i, n_i) are mean zero error terms specific to individual *i* and constant over time (individual specific random effects)

$(e(t), h(t))$ are mean zero error terms specific to period *t*, constant across individuals and independent of all other values of $(e(t'), h(t'))$, respectively (time specific random effects)

$(\xi_i(t), \omega_i(t))$ are mean zero error terms specific to individual *i* at time *t*, independent of all other values of $(\xi_i(t'), \omega_i(t'))$, respectively.

6.3.1 Maximum Likelihood estimators

Given the distributional assumptions for the residuals:

$$\begin{pmatrix} u_i(t) \\ v_i(t) \end{pmatrix} \sim N\left[0, \begin{pmatrix} \sigma_u^2(t) & \rho\ \sigma_{uv}(t) \\ \rho\ \sigma_{uv}(t) & \sigma_v^2(t) \end{pmatrix}\right]$$

where $\sigma_u^2(t) = \sigma_v^2(t) = \sigma^2(t)$ and $\rho = Corr(u_i(t), v_i(t))$, selection bias exists if $\rho \neq 0$. Hausman and Wise (1979) propose a Maximum Likelihood estimator, which examines the impact of endogenous attrition, that is attrition occurred in the selection equation. The procedure yields efficient estimates of the structural parameters in the presence of attrition, as well as an estimate of a parameter that indicates the presence or absence of attrition bias. While the method was demonstrated using data from the *Gary income maintenance experiment*, Hausman and Wise (1979) support that their estimator is applicable in any kind of panels.

Later, Keane, Moffit and Runkle (1988) apply the Maximum Likelihood procedure to estimate the impact of heterogeneity on wages. Nijman and Verbeek (1992) analyze non-response bias in ML estimates of a life-cycle consumption model using data from the *Expenditure Index Panel* for the period of April 1984 to March 1987. Several ways are examined empirically to reveal the nature and the severity of the selection problem due to non-response, as well as a number of methods to estimate the resulting model.

Regarding the above approaches, Vella (1998) indicates the computational demands and the time-cost that ML estimation entails in evaluating multiple integrals. These limitations led analysts to develop simpler model-based procedures for panel data.

6.3.2 2-Step Estimators

Several econometric, 2-step approaches have been developed for program evaluations referring to panel studies. Here, we first describe some early procedures that are commonly used due to the simplicity in their application. Some proposed tests of model specification are then examined. However, because certain forms of selection bias are not eliminated by these procedures, other, modern, approaches are also discussed.

6.3.2.1 Some Early Approaches

1. *The Fixed-Effects (FE) estimator*

This method was developed by Ashenfelter (1978) who supposes that the earnings of the i^{th} individual in period t, $Y_i(t)$, can be modeled as shown by equation (6.6). The *Fixed - Effects* model assumes that the change in earnings between pre-program and post-program periods can be explained by changes in personal characteristics and environmental conditions during the intervening period and by program participation. Following Verbeek and Nijman (1992a), we treat *e(t)* and *h(t)* as fixed time effects, absorbed in $X_i(t)$ and $Z_i(t)$, respectively and consider the following transformation for the explanatory variable (and correspondingly for the dependent variable) of the primary equation:

$$X_i^d(t) = X_i(t) - \sum_{s=1}^{T} X_i(s) \times D_i(s); \;\; if \; D_i(s) \succ 0, \qquad (t = 1,...,T)$$

The FE estimator for *balanced* (B) and *unbalanced* (U) panels, respectively, is:

$$\beta(\mathrm{B}) = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_i^{d\,\prime}(t) \times X_i^d(t) \times c_i \right)^{-1} \times \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_i^{d\,\prime}(t) \times Y_i^d(t) \times c_i \right)$$

where $c_i = \left\{\left(\prod_{t=1}^{T} D_i(t)\right) = 1\right\}$, and

$$\beta(U) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_i^{d\prime}(t) \times X_i^d(t) \times D_i(t)\right)^{-1} \times \left(\sum_{i=1}^{N}\sum_{t=1}^{T} X_i^{d\prime}(t) \times Y_i^d(t) \times D_i(t)\right)$$

The above estimators are consistent provided that the following assumptions are met:

A1. $E\left(\xi_i^d(t) \middle| X_i(t), D_i(t)\right) = 0 \;\; or \;\; \sigma_{\xi\omega} = 0$ and sample selection operates purely through the individual specific terms ε_i.

A2. The earnings function that would prevail in the absence of training is the same for participants and non-participants.

Alternatively, Bassi (1984) formulates the FE estimator in terms of the *endogenous variable model* without imposing the restrictive assumption of fixed time effects. Under the consistency conditions A1 and A2, FE is presented as:

$$Y(t)_i - Y_i(t') = \left(X_i(t) - X_i(t')\right)\beta + \left(D_i(t)\theta_t - D_i(t')\theta_{t'}\right) + \left(\varepsilon_i - \varepsilon_i\right) + \left(e(t) - e(t')\right) + \left(\xi_i(t) - \xi_i(t')\right)$$

$$= \left(X_i(t) - X_i(t')\right)\beta + D_i(t)\theta_t + \left(e(t) - e(t')\right) + \left(\xi_i(t) - \xi_i(t')\right) \qquad (6.12)$$

where $D_i(t') \times \theta_{t'} = 0$ because t′ is chosen to be a period prior to training and thus the training effect $\theta_{t'}$ equals zero. Given that

$$E\left[\xi_i(t) - \xi_i(t') \middle| X_i(t), D_i(t)\right] = 0$$

OLS regression of the difference between post-program earnings in any year and earnings in any pre-program year on the change in regressor between those years and a dummy variable $D_i(t)$ for training status produce plausible estimates for β and θ_t.

2. The Random-Effects (RE) estimator

In many applications, it is implausible to assume that selection on unobservables occurs on the component ε_i. This is the case where the *random effects estimator* is appropriate. Let us define the vectors $Y_{iT} = (Y_{i1},...,Y_{it})'$, $X_{iT} = (X_{i1},...,X_{it})'$ and $\xi_{iT} = (\xi_{i1},...,\xi_{it})$. Also define the number of units for which $D_i(t) = 1$ as W_i and a $W_i \times T$ matrix R_i transforming $Y_i(t)$ into the W_i dimensional vector of observed values $Y_i^0(t)$. Matrix R_i is obtained by deleting rows of the T dimensional identity matrix corresponding to $D_i(t) = 0$. Under fixed time effects, by setting the following distributional assumption for the primary error term:

$$(\varepsilon_i + \xi_i(t)) \sim N[0, \quad \Omega]$$

where $\Omega = \sigma_\varepsilon^2 \iota\iota' + \sigma_\xi^2 I_T$ and I_T is the identity matrix of dimension T, $\iota_T = (1,...,1)'$, the random effects estimator is presented:

$$\beta(B) = \left(\sum_{i=1}^{N} X_i^{0'}(t) \times \Omega_i^{-1} \times X_i^0(t) \times c_i \right)^{-1} \times \left(\sum_{i=1}^{N} X_i^{0'}(t) \times \Omega_i^{-1} \times Y_i^0(t) \times c_i \right)$$

or

$$\beta(U) = \left(\sum_{i=1}^{N} X_i^{0'}(t) \times \Omega_i^{-1} \times X_i^0(t) \right)^{-1} \times \left(\sum_{i=1}^{N} X_i^{0'}(t) \times \Omega_i^{-1} \times Y_i^0(t) \right)$$

for the balanced and unbalanced panels, respectively.

Consistency of the RE estimator requires the stronger condition $E(\varepsilon_i + \xi_i(t) | X_i(t), D_i(t)) = 0$ or $\sigma_{\xi\omega} = 0$ *and* $\sigma_{\varepsilon\eta} = 0$. Thus it cannot produce consistent estimates if the selection is operating either through the individual and/or the idiosyncratic effects.

3. The First – Order Autoregressive estimator

In many applications of the FE estimator assumption A1 is not satisfied because participation is correlated with $\xi_i(t)$. Bassi (1984) refers to the case where individuals in areas experiencing unusually high unemployment are more likely to participate in the program. Similarly, for various reasons individuals may self-select into the program based on unobservables that are not constant over time, violating assumption (6.11) or alternatively program administrators may have an incentive to "cream", that is to choose which individuals will participate and thus turn out to work with *choice-based* samples of participants, analyzed extensively in Manski and Lerman (1977) and Manski and McFadden (1981). To allow for these types of correlation, it is necessary to generalize the FE estimator.

A useful generalization is obtained by decomposing the individual, time-specific error term $\xi_i(t)$ into two components –a systematic $\xi_i(t-1)$ and a random component $M_i(t)$. Then, the first-order autoregression model of $\xi_i(t)$ and $\xi_i(t')$, accounting for the above problems, may be written as:

$$\xi_i(t) = \xi_i(t-1) + M_i(t)$$
$$\xi_i(t') = \xi_i(t'-1) + M_i(t')$$

Intuitively, higher order autoregresion schemes may be applied to capture selection on unobservables successfully. Bassi (1984) discusses this approach and suggests an appealing recursive formula to estimate β and θ_t (the effect of participation in the endogenous variable model) in both self-selective and choice-based sampling designs.

Heckman and Robb (1985a) also discuss the case of choice-based sampling and propose an Instrumental Variable estimator to handle with this case. They also constitute that estimators, similar to FE, can be applied to repeated cross-section data of unrelated persons given a homogeneous environment (satisfaction of assumption (6.11)). Repeated cross-section data are cheaper to collect and do not suffer from problem of non-random attrition which plague panel data.

4. The Random – Growth (RG) estimator

Adopting the assumptions posed for the FE estimator, Heckman and Hotz (1989) discuss program evaluation when, in terms of the model (6.6) – (6.9), $u_i(t)$ has a factor structure of the form:

$$u_i(t) = \varepsilon_{1i} + t \times \varepsilon_{2i} + e(t) + \xi_i(t) \tag{6.14}$$

ε_{1i} is a person-specific, time-invariant component, ε_{2i} is a person-specific growth rate and ($\varepsilon_{1i}, \varepsilon_{2i}$) are assumed to have zero means, finite variances and independency of $v_i(t)$ for all i and t. In this specification, we formulate the random growth model as:

$$\begin{aligned}&\left(Y_i(t) - Y_i(t')\right) - \left(t - t'\right) \times \left(Y_i(t') - Y_i(t'-1)\right)\\ &\quad = D_i(t)\theta + \left[\left(X_i(t) - X_i(t')\right) - \left(t - t'\right) \times \left(X_i(t') - X_i(t'-1)\right]\times \beta\right.\\ &\quad\quad + \left(\xi_i(t) - v_i(t')\right) - \left(t - t'\right) \times \left(v_i(t') - v_i(t'-1)\right)\end{aligned} \tag{6.15}$$

Under the above error specification (6.14), the dependence between $u_i(t)$ and $v_i(t)$ is thought to arise because of dependence between $v_i(t)$ and ($\varepsilon_{1i}, \varepsilon_{2i}$) and is assumed to be eliminated by differencing out the outcomes of period t′ from those of t. The resulting estimators of β and θ_t are reported to be consistent under standard conditions. Furthermore, the asymptotic distribution of the estimator of β is invariant to the choice of earnings from other periods used to proxy for ε_{1i} and ε_{2i}, provided that all of the $\xi_i(t)$ have nonzero variances and the equation is estimated by Generalized Least Squares. Heckman and Hotz (1989) constitute that both the FE and RG models yield consistent estimators of the training effect when applied to choice-based samples, because it is based on conditional $\left(\text{on } D_i(t)\right)$ moment restrictions.

Alternatively, one could be faced up with the assumptions posed for the RE estimator where differencing method cannot produce consistent estimates of β and θ_t. However, this case is not analyzed in the literature. Heckman and Hotz (1989) provide three model specification tests for the FE and RG models. These tests are discussed in the following paragraph along with the specification tests of Bassi (1984) and Verbeek and Nijman (1992a).

6.3.2.2 Specification Tests

✓ *Tests proposed by Bassi (1984)*

Bassi (1984) provides two model specification tests to select between the fixed effects and the random-effects models in a particular application. The null hypotheses of the tests are:

$H_0(1)$: The coefficient on the dummy variable measuring program participation in a pre-program RE estimator is not significally different from zero.

$H_0(2)$: The coefficients in a pre-program RE estimator are the same between participants and non-participants.

He declares that non-rejection of neither $H_0(1)$ nor $H_0(2)$ in terms of specification (6.12), denotes the existence of correlation between the random error $u_i(t) - u_i(t')$ and the selection variable $D_i(t)$, since in pre-program data the coefficients that are tested should not be statistically significant different to zero. Thus assumption A1 is violated and a simple random-effects estimator is sufficient for unbiased estimation of training effects. If, however, either null hypothesis is rejected, then a fixed effects estimator should be chosen as more efficient than the alternative.

✓ *Tests proposed by Heckman and Hotz (1989)*

Heckman and Hotz (1989) propose three model specification tests for the fixed-effects and the random-growth models. The first two tests are close to these proposed by Bassi (1984). The first (T1) is based on access to data on pre-program earnings and regressor variables for future program participants and non-participants. For the fixed effect and random growth versions of this test, one has to modify equations (6.12) and (6.14) and use pre-program information for periods t and t΄. Then, the estimated value of θ_t should not be statistically significant different from zero for any correctly specified selection-correction model.

The second specification test (T2) is based on access to data on post-program earnings. However, the definition of the dichotomous variable $D_i(t)$ is modified to

$D_i(t) = 1$ if an observation is a member of an experimental control group and $D_i(t) = 0$ if an observation is a member of a comparison group (as in method of matching). Since neither group receives training, estimated values of θ_t under fixed effects or random growth models, should not be statistically significantly different from zero for a valid non-experimental, selection-correction estimator.

Finally, Heckman and Hotz (1989) provide a test for model restrictions (T3). They observe that under the assumptions justify the FE and RG estimators, values of $Y_i(t)$ from periods other than those specified by equations (6.12) and (6.13) should not appear as regressors in those equations. A test that the coefficients on these extraneous $Y_i(t)$ values are equal to zero is a test of the restrictions implied by these models. This test requires access only on post-program data. If none of the models passes (T3) then one has to reject all models and look for alternatives.

Nevertheless, Heckman and Hotz (1989, rejoinder) claim that T3 is unlikely to be of general practical use because data on experimental control groups are rarely available, due to the prohibitive cost of experimentation. On the other hand, even if they were available, the authors comment that such tests may be a useful only as a way of using data from pilot experimental studies to narrow the class of non-experimental estimators.

Finally, Heckman and Hotz (1989) refer to the difficulties posed on the development of alternative model-selection criteria based on predictive criteria such as Bayesian. Although consider such an approach very attractive, they comment that is suffers from practical problems. Bayesian information criteria assume that a "correct" model is included among the possible models considered. This is not true in many evaluation situations. We already mentioned above that if none of the models examined passes the tests then the analyst has to look for alternative models.

Moreover, predictive criteria assume known prior probabilities of participation and specification of the full joint distribution of outcomes. The aspect of assuming a specific distribution for participation probabilities prior to the program is extremely complex. The analyst usually recognizes the selective nature of participation decision but very rarely can explain the reasons of this selectivity, since it occurs due to unobserved on the analyst factors. Thus, the respective distribution is very difficult to be determined.

✓ *Tests proposed by Verbeek and Nijman (1992)*

Verbeek and Nijman (1992a) discuss several tests to check for the presence of selectivity bias in estimators based on panel data. They considered the selectivity bias of the FE and RG estimators for linear models, although they support that most of the tests could be straightforwardly generalized to nonlinear models. The work of these two authors indicated that the FE is more robust to non-response biases than the RE estimator. A few simple tests have been conducted to infer on the validity of this result. Neither of them required estimation of the model under selectivity nor a specification of the response mechanism.

A proposed test for selection bias refers to the important model selection discussion of Heckman and Hotz (1989) earlier. Verbeek and Nijman (1992a) test the null hypotheses:

$VN_0(1)$: $H_0^{FE} : E\left(\xi_i^d(t) \middle| D_i(t)\right) = 0$ (the FE estimator is consistent)

where $\xi_i^d(t)$ is a transformed error term of the FE estimator.

$VN_0(2)$: $H_0^{RE} : E\left(\varepsilon_i + \xi_{it} \middle| D_{it}\right) = 0$ (the FE and RE estimators are consistent)

Non-rejection of each hypothesis means consistency for the corresponding estimator. The second hypothesis implies the first in the sense that whenever the random effects estimator is consistent, the FE is consistent as well. However, if both hypotheses cannot be rejected in a significant level, then FE estimator must be preferred to the RE as more efficient. This result agrees with the findings of Bassi (1984). Finally, if both hypotheses are rejected then other models have to be considered.

The authors apply this theory on a model with one explanatory variable. They note that including an additional variable that is uncorrelated with $X_i(t)$ essentially would not change the results, while inclusion of a correlated variable would result in biases that depend heavily on the sign and the magnitude of the correlation coefficient.

Wooldridge (1995) do not agree with the above results. He supports that rejection in the Verbeek and Nijman's (1992a) comparison of RE versus FE procedures could simply be due to the relative robustness of these procedures to serial dependence and heteroscedacity rather than their tendency to deviate in the presence of selectivity bias. Then, he suggests other tests procedures, under different assumptions for a model with FE structure. Specifically, he considers an assumption that puts no restrictions on how $D_i(t)$ relates to $(\varepsilon_i, X_i(t))$. This model is described later.

Verbeek and Nijman (1992a) also suggest some tests for variable addition into either model, based on the use of variables denoting the number of times a particular observation appears in the sample, and whether or not the observation appears in all periods. They conclude that when response is partly determined by an individual effect, which is correlated with the regressor, the power of the variable addition tests is also equal to the power of the LaGrange Multiplier test derived when the model is estimated by ML.

6.3.2.3 Modern Approaches

An advantage of FE and RE estimators is that they do not formulate a selection equation and do not specify a response mechanism. They only define a form for the disturbances of the primary and the selection equation, and without imposing any distributional assumptions, FE and RE estimator result in adequate estimates of the parameters of the selection model and, in extend, of the mean outcomes. However, these estimators do not eliminate specific form of selection biases. Recent papers have extended the work upon estimation in panel data studies. Some alternative estimators are presented here.

1. E.M. Algorithm approach

The rationality of E.M. algorithm has already been outlined in paragraph 5.13. We add that this approach can certainly applied in a panel data study not only to replace the missing values due to the evaluation problem $(Y_{1i}(t) | D_i(t) = 0, X_i(t))$ and $(Y_{0i}(t) | D_i(t) = 1, X_i(t))$ but also to replace missing information due to attrition from panel studies.

Despite the wide use of E.M. algorithm in various applications (see for example Molenberghs, Bijnens and Shaw, 1997), the literature barely considers this method in order to estimate the missing data that constitute the evaluation problem (see Heckman, 1990a). It seems that the disadvantages of this method lead the analysts to develop econometric estimation methods (see paragraph 6.4).

2. Replacement Functions

Replacement functions are also referred in Heckman (1990a) as a tool *to solve the missing data problem by invoking functional form assumptions on outcome and participation equations*. In particular additive separability of the outcome functions in terms of functions of observables and unobservables is usually invoked although this assumption is not essential to this approach. The idea is to solve out some unobservables in terms of other unobservables and observables. To see this clearer, let us introduce the following longitudinal selection model:

$$Y_i(t) = g\left(X_i(t), D_i(t); \theta\right) + u_i(t) \quad \textit{with} \;\; E\left(u_i(t)\right) = 0 \text{ and } t = 1,\ldots,T \tag{6.16}$$

Under the assumptions imposed for the fixed effects model, suppose the simple case where $u_i(t) = \alpha_i(t)Q + M_i(t), M_i(t) \perp Q$ for all t where Q is a factor element with associated known factor loading $\alpha_i(t) = 1$. Then, if:

$$\left(D_i(1),\ldots,D_i(T)\right) \perp \left(M_i(1),\ldots,M_i(T)\right)$$

or

$$\left(D_i(1),\ldots,D_i(T)\right) \perp \left(M_i(1),\ldots,M_i(T)\right) \left| \left(X_i(1),\ldots,X_i(T)\right)\right.$$

eliminating Q from the model makes D exogenous with respect to disturbances in the transformed primary model since $E\left(u_i(t)\right) = E\left(M_i(t)\right) \neq 0$. A sensible replacement function for Q is:

$$Q = Y_i(t') - g\left(X_i(t'), D_i(t'); \theta\right) - M_i(t')$$

that yields the conventional FE model as

$$
\begin{aligned}
Y_i(t) &= g\big(X_i(t), D_i(t);\theta\big) + u_i(t) \Leftrightarrow \\
Y_i(t) &= g\big(X_i(t), D_i(t);\theta\big) + Y_i(t') - g\big(X_i(t'), D_i(t');\theta\big) - M_i(t') + M_i(t) \Leftrightarrow \\
Y_i(t) - Y_i(t') &= g\big(X_i(t), D_i(t);\theta\big) - g\big(X_i(t'), D_i(t');\theta\big) + M_i(t) - M_i(t') \qquad (6.17)
\end{aligned}
$$

From (6.17) θ can be consistently estimated by a simple Least Squares method.

When Q represents a vector J × 1 of factors (J ≠ 1) and $\alpha_i(t) \neq 1$, a more general replacement function form can be the below:

$$Q = \frac{1}{T_t} \sum_{t'=1}^{T} \big(Y_{t'} - g\big(X_{it'}, D_{it'};\theta\big) - M_{it'}\big) \qquad (6.18)$$

Heckman (1990a) generalizes once more this approach by discussing for replacement functions when $1 \leq J \prec T$ and $\alpha_i(t)$ be a function of $X_i(t)$.

3. Wooldridge's (1995) estimator

Wooldridge (1995) derives a conditional expectation for the primary equation of model (6.6)-(6.10) that leads to a selection bias test. He assumes that FE estimator produce plausible estimates if:

$$E\big(\xi_i(t) \big| \varepsilon_i, X_i(t), D_i(t)\big) = 0 \quad as\ N \to \infty \qquad (6.19)$$

provided that all periods $t = 1,...,T$ are available for any cross-section drawn from the population. In practice, this assumption is more restrictive than the one of Verbeek and Nijman (1992a), posed for the consistency of the FE estimator. However, it is simpler to be stated and less cumbersome to be applied to cases where selection is based on the exogenous variables $X_i(t)$. In fact, because assumption (6.19) puts no restrictions on how $D_i(t)$ relates to $(\varepsilon_i, X_i(t))$, it follows that selection can depend on $(\varepsilon_i, X_i(t))$ in an arbitrary, and thus more general, fashion.

As the simplest selectivity pattern can be thought the following:

$$E\left(\xi_i(t)\middle|\varepsilon_i, \eta_i, Z_i(t), \omega_i(t), D_i(t)\right) = \rho \times \omega_i(t) \tag{6.20}$$

for some unknown scalar ρ. Under equation (6.20), Wooldridge (1995) shows that a conditional expectation of the primary equation can be represented as:

$$E\left(Y_i(t)\middle|\varepsilon_i, \eta_i, Z_i(t), \omega_i(t), D_i(t)\right) = X_i(t)\beta + \varepsilon_i + \rho \times \omega_i(t) \tag{6.21}$$

A test for H_0: ρ = 0, is a test for the FE estimator. The author describes a procedure to test this null hypothesis using the t-statistic for an estimate of ρ, $\hat{\rho}$. Under some conditions, a serial correlation and heteroscedacity-robust standard error can be computed for $\hat{\rho}$.

Finally, the author extends his method in several ways. First, he allows for serial dependence and heterogeneity in the selection equation. Then, he recasts the selection mechanism as follows:

Alternative Selection Mechanism

$$h \equiv \max\left(0, h_i^{(l)}(t)\right)$$

$$D_i(t) \equiv 1\left[h_i^{(l)}(t) \succ 0\right], \quad t = 1,\ldots,T$$

This specification is considered under a partially observable selection variable and under observation of the selection indicator only.

4. Kyriazidou's (1997) estimator

Kyriazidou's (1997) approach to the panel data selection model is based on the assumption that the time effects *e(t)* and *h(t)* of model (6.6) – (6.10) are absorbed into the conditional mean of the primary equation and thus they are excluded from her analysis.

Therefore, she develops a fixed-effects model for two time periods t′ and t over the sample that satisfies $D_i(t') = D_i(t) = 1$. Observe that the individual effects ε_i and n_i are also eliminated in the FE model since they are constant over the time periods t and t′. However, the error terms $\xi_i(t)$ and $\omega_i(t)$ remain in the analysis and their correlation with the selection variable $D_i(t)$ operates for the selection bias. Kyriazidou applies a 2-step approach to correct for the selection bias and estimate the primary equation without making any distributional assumptions for the error terms.

The correction term that is used here is given by:

$$
\begin{aligned}
& E\left(\xi_i(t) - \xi_i(t') \middle| D_i(t) = D_i(t') = 1, \zeta_i\right) \\
&= E\left(\xi_i(t) \middle| \omega_i(t) \prec z_i'(t)\gamma + \eta_i, \xi_i(t') \prec z_i'(t')\gamma + n_i, \zeta_i\right) \\
&\qquad - E\left(\xi_i(t') \middle| v_i(t) \prec z_i'(t)\gamma + n_i, v_i(t') \prec z_i'(t')\gamma + \eta_i, \zeta_i\right) \\
&= \Lambda_i(t)\left[z_i'(t)\gamma + n_i, z_i'(t')\gamma + n_i, \zeta_i\right] - \Lambda_i(t')\left[z_i'(t)\gamma + n_i, z_i'(t')\gamma + n_i, \zeta_i\right] \\
&= \lambda_i(t) - \lambda_i(t') \qquad (6.22)
\end{aligned}
$$

where $\zeta = [X_i(t'), X_i(t), Z_i(t'), Z_i(t), n_i, \varepsilon_i]$ and $\Lambda_i(t)$ is an unknown function determining the value of the correction term. When $Z_i'(t)\gamma = Z_i'(t')\gamma$ and $D_i(t) = D_i(t') = 1$ the sample selection effect $\lambda_i(t)$ and $\lambda_i(t')$ is the same in both periods t and t′. Thus, for the particular individual *i,* applying the first differences:

$$
\begin{aligned}
E\left(Y_i(t) - Y_i(t')\right) &= \left(X_i(t) - X_i(t')\right)\beta + E\left(\left(\varepsilon_i - \varepsilon_i\right)\right) + E\left(\left(\xi_i(t) - \xi_i(t')\right)\right) \\
&= \left(X_i(t) - X_i(t')\right)\beta + E\left(\left(\xi_i(t) - \xi_i(t')\right)\right) \\
&= \left(X_i(t) - X_i(t')\right)\beta
\end{aligned}
$$

eliminates n_i and $\lambda_i(t) - \lambda_i(t')$ and yields consistent estimates of β. This result is similar to the one described for the FE estimator of Chamberlain (1982).

Kyriazidou (1997) generalizes her approach by indicating that in general there is no reason to expect that $Z_i'(t)\gamma = Z_i'(t')\gamma$ and thus $\lambda_i(t) - \lambda_i(t') = 0$ since the sample selection effect $\lambda_i(t)$ depends not only on the partially unobservable vector ζ_i, but also on the generally unknown joint conditional distribution of ($\xi_i(t)$, $\omega_i(t)$, $\omega_i(t')$), which may differ across individuals as well as over time for the same individual. The primary equation "of difference" can be written then:

$$Y_i(t) - Y_i(t') = (X_i(t) - X_i(t')\beta + E\left(\xi_i(t) - \xi_i(t') \middle| D_i(t) = D_i(t') = 1, \zeta_i\right) + c_i$$

The author suggests estimating γ by some procedure that does not impose distributional assumptions on the $\omega_i(t)$. An appealing one is the method of the Maximum score estimator of Manski (1987), who developed a semi-parametric method for estimating the parameters of a discrete choice model in a panel data setting without making any distributional assumptions for the disturbances. He assumes only that:

a) The disturbances are time-stationary, that is $F_{\xi|(x_i(t),\varepsilon_i)} = F_{\xi|(x_i(t'),\varepsilon_i)}$, where $F_{\xi|(x_i(t),\varepsilon_i)}$ and $F_{\xi|(x_i(t'),\varepsilon_i)}$ are the distributions of the disturbances conditional on $X_i(t)$ and $X_i(t')$, respectively and an unobserved time invariant person – specific effect ε_i.
b) The support of $F_{\xi|(x_i(t'),\varepsilon_i)}$ is R^1, for all ($X_i(t')$, ε_i) (unbounded support).
c) The explanatory variables $X_i(t)$ vary enough over time.

The estimates of γ are then used to estimate the regression equation of interest. A plausible estimator of β is given by:

$$\hat{\beta} = \left[\sum_{i=1}^{N} \hat{\psi}_{iN} \times \Delta\left[X_i'(t)\right] \times \Delta\left[X_i'(t)\phi_i\right]\right]^{-1} \times \left[\sum_{i=1}^{N} \hat{\psi}_{iN} \times \Delta\left[X_i'(t)\right] \times \Delta\left[Y_i'(t)\phi_i\right]\right]^{-1}$$

where $\Delta[\cdot]$ denotes the first differences of the respective term in brackets, $\hat{\psi}_{iN}$ is a weight that declines to zero as the magnitude of the difference $\left|Z_i'(t)\gamma - Z_i'(t')\gamma\right|$ increases and $\varphi_i \equiv 1\{D_i(t) = D_i(t') = 1\}$. Most usually kernel weights of the form:

$$\hat{\psi}_{iN} = \frac{1}{h_N} \times K\left\{\frac{\Delta\left[Z_i(t)\gamma\right]}{h_N}\right\}$$

are applied, where K is a kernel density function and h_N is a sequence of "bandwidths" which tend to zero as $N \to \infty$. Thus, for non-zero $|\Delta[Z_i'(t)\gamma]|$ the weight ψ_{iN} shrinks as the sample size N increases, while for fixed N, a larger $|\Delta[Z_i'(t)\gamma]|$ corresponds to a smaller weight. It is showed that the estimator of β consistent and asymptotically normal.

5. A Non-parametric approach

To estimate parametrically the parameters of the (mixed) selection bias model (6.6)-(6.9), one should assume that the individual and time specific random effects are normally distributed with mean zero. Deviation from normality of the random effects can have an important effect on inferences involving these terms (see Verbeke and Lesaffre, 1996). For this reason, an alternative, more relaxed approach may be needed.

Ghidey, Lesaffre, Eilers and Verbeke (2002) describe an alternative linear mixed model with a P-spline smoothed density function for the random effects. This model relaxes the usual normality assumption in a classical linear mixed model to a more general distribution function for the random effects. It also has the advantage that the true underlying distribution will be better estimated when the normality assumption fails. Although, the authors develop this approach in a panel framework without the presence of selectivity bias, a selection bias model estimation procedure may be developed based on these findings.

6.4 Evaluation Under Dropouts

One aspect of self-selection that is often ignored is the bias created by ignoring missing information from subjects that either refuse or are unable to participate after a number of sequential measurements. This problem, known as dropping out in panel studies, is endemic to all large-scale social programs. Social researchers may restrict their analysis to those subjects who are observed at the cases where missingness occurs Completely at Random (MCAR), or at least at Random (MAR). Nevertheless, in Non-Ignorable missing situations (NIM), restriction of the analysis only to observed individuals leads in severe biases and inappropriate standard errors for the model estimates. The extend and the direction of any bias will depend upon the magnitude of the NIM mechanisms that affect the data.

Crouchley, Oskrochi and Bradley (2002) study the case of NIM data in terms of an evaluation study where selection bias also appears. They construct a mixed parametric model for a bivariate selection mechanism, which includes one component for missing subjects and another component for selection into the state of interest. After testing for ignorable missingness, the authors compare the parameter estimates found with the ones obtained from simpler models that do not take into account the NIM mechanisms.

6.5 Discussion on the Panel Data Estimators

A significant amount of literature involves discussion on evaluation social programs within a panel data setting. As in the cross-section case, either an experimental design or econometric models can be applied for evaluation purposes. Matching has not been met as an option in the present literature although, a generalization of matching methods for panel analysis is straightforward.

In experimental methods, participants have been randomized in treatment and control group before the program starts and the pre-program data on these individuals are collected. Post-program data (outcomes) are those collected after the completion of the program. A direct comparison of post-program and pre-program (if needed) outcomes can produce plausible estimates of the participation impact. In this framework, panel data estimators are easy to be implemented.

In a structural model framework, the various estimators discussed above can be applied to produce plausible estimates under specific distributional assumptions and under specific forms for the error terms. Fixed-Effects and Random-Effects estimators, as well as their extensions, have been successfully implemented in the past. However, as Vella (1998) indicates these estimators cannot eliminate all kinds of selection bias. Thus other estimators have to be considered. Wooldrigde’s (1995) method have been successfully implemented in real data. Kyriazidou’s (1997) estimator has been also applied in practice but it is restricted for the special case where the time effects e(t) and h(t) are absorbed into the conditional mean of the primary equation excluded from the analysis.

The major limitation of the panel data studies is the phenomenon of attrition from the samples of participants. The large number of participants at the first time the data are collected is usually followed by a much less number of persons in collecting the post-program data. Various reasons such as inability to find respondents that have moved in other places, refusal of response in post-program period and other natural reasons (e.g. deaths) are common sources of attrition in panel studies. Note that sample attrition is not the same as dropping out of the program. Both control and treatment group members can attrit from the sample causing much serious problem in the analysis.

Sample attrition poses a problem for experimental evaluations when it is correlated with individuals' characteristics or with the impact of treatment conditional on characteristics. Heckman, LaLonde and Smith (1999) support that persons with poorer labor market characteristics tend to have higher attrition rates. Even if attrition affects both experimental and control group members in the same way, the experiment estimates the mean impact of the program only for those who remain in the sample. In this case, the experimental estimate of training is biased because individuals' status *r* is correlated with their likelihood of being in the sample. In this setting, experimental evaluations become non-experimental evaluations since some assumptions have to be made to deal with selection bias. Determination of the missing mechanism contributes to overcome these problems. Possibly, a useful alternative to panels would be the repeated cross-section data that are studied extensively in the modern econometric literature.