Parametric estimation of treatment effects with time-varying qualification in panel data and repeated cross-sections

Kreshnik Xhangolli *

April 9, 2017

Abstract

We present an *unrestricted* linear model for identifying treatment effects in panel and repeated cross-sections with dichotomous time-varying qualification. The model is *unrestricted* in terms of (i) treatment effects for each of the subgroups introduced by the time-varying qualification and (ii) subgroups' time-differences in potential untreated outcomes (time-effects). The subgroup generalization is achieved by considering subgroup information *as given*. Instead time-effects generalization is achieved by combining subgroup indicator variables with time indicator variables. This panel data model enables estimation with fixed and random effects; the latter has not been considered before. This parametric approach allows for easier implementation of covariate control and for easier identification of sufficient conditions for treatment effects estimation in such setting. We employ the model to identify new scenarios and respective sufficient conditions for treatment effect estimation with time-varying qualification with panel data or repeated cross-sections.

JEL Classification Numbers: C14, C21, C33, I11.

Key Words: difference in differences, effect on in-stayers, effect on in-movers, untreated moving effect, panel data.

*Graduate School of Economics, Finance and Management, Goethe University Frankfurt, kreshnik.xhangolli@gmail.com

1 Introduction

In a Difference in Differences setting with time-varying qualification, treatment is applied to an eligible group of statistical units who fulfills some eligibility criteria after a certain point in time t_0. In this setting the group who receives the treatment after t_0 does not necessarily fulfill the eligibility criteria before t_0, allowing statistical units to move in or out the eligibility criteria before and after t_0. As showed by the example of [Lee and Kim, 2014], in this kind of setting *movers effect* are introduced which can be misleadingly considered as treatment effects.

To come around such problem, [Lee and Kim, 2014] consider the heterogeneous treatment effects for the subgroups introduced by a dichotomous time-varying qualification in a two-periods time frame. They suggest four non-parametric estimators (Section 2 on the following page) and show that these estimators can identify treatment effects for the two subgroups receiving treatment, if *"same time-effect conditions"* of appropriate subgroups is assumed. In addition the authors provide panel linear models that assume same time-effects, same group-effects and either (i) same treatment effect for all subgroups or (ii) pairwise differing treatment effects. They present, through first-differencing, parametric estimation of treatment effects in panels satisfying this kind of structure.

We propose a panel linear model that poses no restrictions on subgroups' time-effects or treatment-effects. Subgroup generalization is achieved by considering subgroup information *as given*. Time-invariant indicator variables are introduced for each subgroup. Instead unrestricted time-effects are measured by combining subgroup indicator variables with time indicator variables. Such approach builds a typical panel linear model that enables estimation with fixed and random effects. While a first difference estimator has been proposed on some restricted models by [Lee and Kim, 2014], to our knowledge this has not been the case with random effects. This is important since random effects allow for control on time-invariant covariates. We show that with this model it is possible to parametrically measure treatment effects without imposing the same time-effects and same group-effects structure proposed by [Lee and Kim, 2014]. The treatment we present can be considered as the parametric equivalent of the non-parametric estimators proposed by [Lee and Kim, 2014]. We believe that this parametric approach allows for easier implementation for covariate control and for identification of sufficient conditions for treatment effects estimation in such setting. We employ the model to identify new scenarios and respective sufficient conditions for treatment effects estimation with time-varying qualification with panel data or repeated cross-sections.

We succinctly specify the structure of the DD setting as specified in [Lee and Kim, 2014] in Section 2. In Section 3 we specify the unrestricted panel linear model, present first differ-

ence and random effects estimation and list scenarios and specific assumptions that identify treatment effects. In Section 4 we explore our model in a setting with repeated cross-sections.

2 Specifications of DD with time-varying qualification

The DD setting with time-varying qualification proposed by [Lee and Kim, 2014] has $\tilde{N}$ statistical units indexed by i which are observed at different time points indexed by t. Outcome of interest is denoted by Y_{it}, treatment and qualification indicators denoted respectively by D_{it} and Q_{it}, and covariates denoted by W_{it}. The controlling covariates can be time constant C_i or time-varying X_{it}, therefore $W_{it} = (C_i, X_{it})$. The causality at each period t consists of W_{it} being realized before D_{it} and both of them affecting outcome Y_{it} (see table 1).

Table 1: Diagram of causality

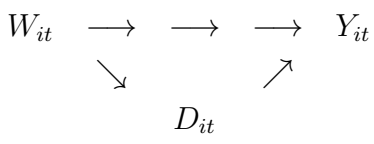

Table 2: Movers and stayers

	$Q_3 = 0$	$Q_3 = 1$
$Q_2 = 0$	out-stayers	in-movers
$Q_2 = 1$	out-movers	in-stayers

The time frame is restricted to two periods ($t = 2, 3$). Treatment is applied only to the group who is eligible at period $t = 3$, irrespectively of the group's eligibility at period $t = 2$, therefore $D_{it} = Q_{it}1[t = 3]$, where $1[.]$ is the indicator function. This setting introduces four groups (table 2). As a side note, we really appreciate this semantic introduced by [Lee and Kim, 2014] since allows for referencing of bigger subgroups such as *in-group (in-movers and in-stayers), out-group (out-movers and out-stayers), movers (in-movers and out-movers) and stayers (out-stayers and in-stayers).*

The potential outcome model identifies at every period t two potential responses, treated Y_{it}^1 and untreated Y_{it}^0. The relation between potential and observed outcomes is given by $Y_{it} = (1 - D_{it})Y_{it}^0 + D_{it}Y_{it}^1 = Y_{it}^0 + (Y_{it}^1 - Y_{it}^0)D_{it} = Y_{it}^0 + (Y_{it}^1 - Y_{it}^0)Q_{it}1[t = 3]$. The last equality holds since in this setting $D_{it} = Q_{it}1[t = 3]$. The observed stochastic process is $\{Y_{it}, Q_{it}, D_{it}, W_{it}, t = 2, 3\}, \quad i = 1, \ldots, \tilde{N}$ with *iid* sample properties. W_{it} may contain lagged variables but in that case 3 periods will be needed. Denoting $W_{i,t-1}^t \equiv (C_i, X_{i,t-1}', X_{it}')'$, [Lee and Kim, 2014] consider four W_2^3-conditional effects at the post-treatment period $t = 3$ (table 3a on the next page). The estimators proposed by [Lee and Kim, 2014] are presented in table 3b on the following page.

Table 3: Treatment effects and non-parametric estimators

(a) Treatment effects with time-varying qualification

$$\text{out-stayers} \equiv E(Y_3^1 - Y_3^0 | W_2^3, Q_2 = 0, Q_3 = 0) \quad \text{in-movers} \equiv E(Y_3^1 - Y_3^0 | W_2^3, Q_2 = 0, Q_3 = 1)$$
$$\text{out-movers} \equiv E(Y_3^1 - Y_3^0 | W_2^3, Q_2 = 1, Q_3 = 0) \quad \text{in-stayers} \equiv E(Y_3^1 - Y_3^0 | W_2^3, Q_2 = 1, Q_3 = 1)$$

(b) DD non-parametric estimators

Name	Definition
stayer DD :	$DD_{stay.} = E(\triangle Y_3 \| Q_2 = 1, Q_3 = 1) - E(\triangle Y_3 \| Q_2 = 0, Q_3 = 0)$
mover DD :	$DD_{move.} = E(\triangle Y_3 \| Q_2 = 0, Q_3 = 1) - E(\triangle Y_3 \| Q_2 = 1, Q_3 = 0)$
in-stayer out-mover DD :	$DD_{in-s,out-m} = E(\triangle Y_3 \| Q_2 = 1, Q_3 = 1) - E(\triangle Y_3 \| Q_2 = 1, Q_3 = 0)$
in-mover out-stayer DD :	$DD_{in-m,out-s} = E(\triangle Y_3 \| Q_2 = 0, Q_3 = 1) - E(\triangle Y_3 \| Q_2 = 0, Q_3 = 0)$

3 Treatment effects with panel data

3.1 Unrestricted Linear Panel Model

We consider the setting with two periods ($t = 2, 3$), in which treatment is applied only to the group who is eligible at period $t = 3$, irrespectively of the group's eligibility at period $t = 2$. In the unrestricted model we propose, subgroup information is considered *as given*, i.e. as a time-invariant covariate of the statistical unit. An alternative approach would be to create four dichotomous variables indicating unit's belonging to that specific subgroup (table 4). We will alternate between the two notations of subgroups depending on which presentation provides better clarity on the estimation method considered. We label them as the G-notation and Q-notation. Subgroups will be indexed by the pair (Q_2, Q_3) identifying that group. Therefore *out-stayers* will be indexed by 00 and so on. It is important to keep in mind that (i) subgroup variables are functions of the eligibility variables $Q_{i,2}, Q_{i,3}$ and (ii) subgroup information is considered as given and time-invariant; the same as gender would be considered.

Table 4: Alternative subgroups presentation

$$G_{00} = (1 - Q_{i,2})(1 - Q_{i,3}) \quad G_{01} = (1 - Q_{i,2})Q_{i,3}$$
$$G_{10} = Q_{i,2}(1 - Q_{i,3}) \quad G_{11} = Q_{i,2}Q_{i,3}$$

Using potential outcome framework in our setting means that each subgroup has potential *untreated* outcomes at $t = 2, 3$ and potential *treated* outcome at $t = 3$. In table 5 on the next page we present the potential outcomes, where we have used the same indexation logic

but we have substituted cumbersome potential outcome symbols, $Y^0_{i,00,t=2}$, with Greek letters, where γ refers to potential untreated outcome at $t = 2$, δ to potential untreated outcome at $t = 3$ and β to treatment effects. We employ this Greek letter notation through the rest of this paper. We would like to clarify two characteristics of the model presented in table 5.

Table 5: POM representation of the general case

Groups		Y^0_2		Y^0_3		Y^1_3	
out-stayers ($Q_2 = 0, Q_3 = 0$)	in-movers ($Q_2 = 0, Q_3 = 1$)	γ_{00}	γ_{01}	δ_{00}	δ_{01}	$\delta_{00} + \beta_{00}$	$\delta_{01} + \beta_{01}$
out-movers ($Q_2 = 1, Q_3 = 0$)	in-stayers ($Q_2 = 1, Q_3 = 1$)	γ_{10}	γ_{11}	δ_{10}	δ_{11}	$\delta_{10} + \beta_{10}$	$\delta_{11} + \beta_{11}$

First, the model implies heterogeneity of homogeneous subgroups, i.e. while statistical units among subgroups may differ in potential outcomes, statistical units belonging to the same subgroup are the same. Second, for ease of presentation the effects of observables (W_{it}) and unobservables (V_{it}) are not shown in table 5. The model of table 5 can be considered as the case when outcome is determined only by treatment and no other factors, therefore we have only constants. The complete model specification is given in equation M_1, with observables (W_{it}) and unobservables (V_{it}) as defined in [Lee and Kim, 2014].

$$\begin{aligned} Y^0_{i2} &= \gamma_{00}G_{00} + \gamma_{10}G_{10} + \gamma_{01}G_{01} + \gamma_{11}G_{11} + \beta'_w W_{i2} + V_{i2} \\ Y^0_{i3} &= \delta_{00}G_{00} + \delta_{10}G_{10} + \delta_{01}G_{01} + \delta_{11}G_{11} + \beta'_w W_{i3} + V_{i3} \\ Y^1_{i3} &= Y^0_{i3} + \underbrace{\beta_{00}G_{00} + \beta_{10}G_{10} + \beta_{01}G_{01} + \beta_{11}G_{11}}_{=\eta} \end{aligned} \tag{M_1}$$

The panel data representation of the model M_1 takes the form (appendix A)

$$\begin{aligned} Y_{it} =& \gamma_{00}G_{00}1[t=2] + \gamma_{10}G_{10}1[t=2] + \gamma_{01}G_{01}1[t=2] + \gamma_{11}G_{11}1[t=2] + \delta_{00}G_{00}1[t=3] \\ &+ \delta_{10}G_{10}1[t=3] + (\delta_{01}+\beta_{01})G_{01}1[t=3] + (\delta_{11}+\beta_{11})G_{11}1[t=3] + \beta'_w W_{it} + V_{it} \end{aligned} \tag{P_1}$$

The vector of covariates $(G_{00}1[t=2], G_{10}1[t=2], G_{01}1[t=2], G_{11}1[t=2], G_{00}1[t=3], G_{10}1[t=3], G_{01}1[t=3], G_{11}1[t=3], W_{it})'$ is known for each statistical unit at each time point $t = 2, 3$ and unobservables are included in $V_{it} = A_i + U_{it}$. Coefficients $(\gamma_{00}, \gamma_{10}, \gamma_{01}, \gamma_{11}, \delta_{00}, \delta_{10}, \delta_{01} + \beta_{01}, \delta_{11}+\beta_{11}, \beta'_w)'$ are estimated from panel model P_1. First note that it is possible to estimate only the sums $\delta_{.1} + \beta_{.1}$, instead of treatment coefficients $\beta_{.1}$. This means that (i) treatment effects can be measured only for *in-movers* and *in-stayers* and (ii) additional assumptions are needed to disentangle treatment effects. Both (i)-(ii) have already been pointed out by [Lee and Kim, 2014] through their non-parametric estimators, but we believe that in P_1 they are easier to understand. Point (i) is expected since potential treated outcomes are

observed only for these two subgroups [Lee and Kim, 2014]. With respect to (ii), in subsection 3.2, we identify new sufficient assumptions besides those previously described by [Lee and Kim, 2014].

Through panel model P_1, estimation of coefficients can be done with any fixed or random effects panel data estimation methods. This is a major contribution of this paper, since [Lee and Kim, 2014] suggest only a *first difference* estimator which is applied to a much more restricted model than P_1. In the end of this subsection we show how [Lee and Kim, 2014] models represent special cases of P_1.

While P_1 is a typical panel data representation of M_1, the latter can be presented also as a system of equations O_1. O_1 is a system of two equations $t = 2, 3$ with the same covariates but some differing coefficients.

$$\begin{aligned} Y_{i2} &= \gamma_{00}G_{00} + \gamma_{10}G_{10} + \gamma_{01}G_{01} + \gamma_{11}G_{11} + \beta_w' W_{i2} + V_{i2} \\ Y_{i3} &= \delta_{00}G_{00} + \delta_{10}G_{10} + (\delta_{01} + \beta_{01})G_{01} + (\delta_{11} + \beta_{11})G_{11} + \beta_w' W_{i3} + V_{i3} \end{aligned} \tag{O_1}$$

Methods for system OLS as specified in [Wooldridge, 2010] can be easily employed with O_1 also. Instead P_1, which can be derived either from M_1 or P_1, represents a typical panel data model, and therefore we use it as our workhorse. One can also consider P_1 as a surrogate of the least-square dummy variable approach. In this case dummies are added for combination of subgroups and time periods. Whichever approach one chooses, the main idea remains that of considering subgroup information as given and time-invariant. Differing coefficients for subgroup variables allow usage of fixed effects methods such as first differencing.

3.2 First Difference estimator

First differencing P_1 gives (appendix B)

$$\begin{aligned} \triangle Y_{i3} =& (\underbrace{\delta_{00} - \gamma_{00}}_{=\alpha_{00}})G_{00} + (\underbrace{\delta_{10} - \gamma_{10}}_{=\alpha_{10}})G_{10} + (\underbrace{\delta_{01} - \gamma_{01}}_{=\alpha_{01}} + \beta_{01})G_{01} + (\underbrace{\delta_{11} - \gamma_{11}}_{=\alpha_{11}} + \beta_{11})G_{11} \\ &+ \beta_x' \triangle X_{i3} + \triangle U_{i3} \end{aligned} \tag{FD_1}$$

where $\alpha_{..} = \delta_{..} - \gamma_{..}$ represents the time-difference in potential *untreated* outcomes for respective subgroups, $(Y^0_{3,..} - Y^0_{2,..})$, which is also the time-effects. As we show in the end of this subsection, the panel linear models of [Lee and Kim, 2014], assume same time-effects ($\triangle\beta_3$) for all subgroups. Clearly, FD_1 illustrates the unrestricted case. When considering the first difference estimator we employ also the alternative Q-notation FD_2 on the next page (appendix C). FD_2 rearranges the terms $((1 - Q_{i,2})(1 - Q_{i,3}), Q_{i,2}(1 - Q_{i,3}), (1 - Q_{i,2})Q_{i,3}, Q_{i,2}Q_{i,3})'$ of FD_1 into $(1, Q_{i,2}, Q_{i,3}, Q_{i,2}Q_{i,3})'$. As we will see, in some cases it is easier to identify sufficient

assumptions using this FD_2.

$$\begin{aligned}\triangle Y_{i3} =&\alpha_{00} + (\alpha_{10} - \alpha_{00})Q_{i,2} + (\alpha_{01} - \alpha_{00} + \beta_{01})Q_{i,3} \\ &+ (\alpha_{11} - \alpha_{10} - \alpha_{01} + \alpha_{00} + \beta_{11} - \beta_{01})Q_{i,2}Q_{i,3} + \beta_x' \triangle X_{i3} + \triangle U_{i3}\end{aligned} \tag{FD_2}$$

Before identifying sufficient assumptions by employing FD_1 and FD_2 we make three major comments.

First we re-iterate that since treatment is applied and observed only for $Q_3 = 1$, only the treatment effects for the *in-groups (in-stayers and in-movers)* can be identified. Therefore we have to assume treatment effects for out-stayers and out-movers. This said, from this point onward, we imply estimation of *in-groups* treatment effects whenever we refer to estimating treatment effects. In addition we will use the terms *time-effects* and *time-difference of potential untreated outcomes* interchangeably.

Second, estimation will be always biased with the time-difference of potential untreated outcomes (clearer in FD_1). Without additional assumptions it is not possible to disentangle treatment effects. One approach would be to assume zero or some constant time-effects, $\alpha_{11} = m, \alpha_{10} = n, m, n \in R$. To our knowledge this approach is not quite well accepted.

Third, time-effects for *out-movers* and *out-stayers* can be identified and estimated. Assuming some form of relationship between any combination of the *in-groups* with *out-groups* can make estimation of treatment effects possible. While we usually employ equal or opposite time-effects between the *in-groups* and *out-groups*, other forms can be used. Equal time-effects assumption is preferred since it provides the clear DD interpretation - compared subgroups are affected by same unobserved differences.

Case 1: $\alpha_{01} = \alpha_{00}$ is a sufficient condition to consistently estimate β_{01} using either FD_1 or FD_2. The assumption would mean equal time-difference in potential untreated outcomes between *in-movers* and *out-stayers*. Note that with this assumption only, it is not possible to estimate either β_{11} in FD_1 or its FD_2 equivalent, $\beta_{11} - \beta_{01}$.

Case 2: $\alpha_{01} = \alpha_{10}$ is a sufficient condition to consistently estimate β_{01} using either FD_1 or FD_2. This condition has been first identified by [Lee and Kim, 2013] as *"movers same time-effect condition"*[1]. Neither β_{11} in FD_1 or its FD_2 equivalent $\beta_{11} - \beta_{01}$, can be estimated in this case. $\alpha_{01} = \alpha_{10}$ would mean that *in-movers* and *out-movers* have the same *mover* effect. An alternative assumption that would still allow estimation would be $\alpha_{01} = -\alpha_{00}$ which was employed by [Lee and Kim, 2013] in their example that conventional DD could misleadingly assign treatment effects when there are none.

Case 3: $\alpha_{11} = \alpha_{00}$ is a sufficient condition to estimate β_{11} using either FD_1 or FD_2. This

[1]In subsection 3.3 on page 9 we show that under model M_1 the two assumptions are equivalent

is *"stayers same time-effect condition"* identified by [Lee and Kim, 2014][2]. The sufficiency is easier to see from FD_1. In FD_2 we note that the coefficient of $Q_{i,2}Q_{i,3}$ can be re-arranged as $((\alpha_{11} - \alpha_{10}) - (\alpha_{01} - \alpha_{00} + \beta_{01}) + \beta_{11})$. Since $\alpha_{11} = \alpha_{00}$, the first parenthesis equals the negative of $Q_{i,2}$ coefficient and the second parentheses equals $Q_{i,3}$ coefficient. Therefore β_{11} is estimated. Note that we cannot measure β_{01} employing only this assumption.

Case 4: $\alpha_{11} = \alpha_{10}$ is a sufficient condition to estimate β_{11} using either FD_1 or FD_2. This assumption would mean same time-effects for *in-stayers* and *out-movers*. Depending on the case under study other relationships may be more plausible; for example negative time-effects differences ($\alpha_{11} = \alpha_{10}$).The sufficiency is easier to see from FD_1. In FD_2, $Q_{i,2}Q_{i,3}$ coefficient becomes $(-(\alpha_{01} - \alpha_{00} + \beta_{01}) + \beta_{11})$, where the parenthesis equals $Q_{i,3}$ coefficient. Note that we cannot measure β_{01} by employing only this assumption.

Case 5: $\alpha_{11} - \alpha_{10} - \alpha_{01} + \alpha_{00} = 0$ is a sufficient condition to consistently estimate $\beta_{11} - \beta_{01}$ using either FD_1 or FD_2. This can be proved useful when the research question relates to the difference in treatment effects between the two in-groups. Note also that the terms of the assumption can be rearranged, for example $\alpha_{11} - \alpha_{10} = \alpha_{01} - \alpha_{00}$. The sufficiency is easily seen from FD_2. In FD_1, sufficiency is derived by appropriate combination of coefficients of measurable coefficients $\alpha_{00}, \alpha_{10}, \alpha_{01} + \beta_{01}, \alpha_{11} + \beta_{11}$. If in addition we assume same treatment effects for the in-groups ($\beta_{11} = \beta_{01}$) then this condition is sufficient to estimate the effect (case 10).

Case 6: $\alpha_{01} = \alpha_{11}$ is a sufficient condition to consistently estimate $\beta_{11} - \beta_{01}$ if the latter is of interest to the researcher. Preferably used with FD_2. If we are assuming same treatment effects for the in-groups ($\beta_{11} = \beta_{01}$) then this condition is sufficient to estimate the effect.

Case 7: Assumptions in Case 1 and 5 are jointly sufficient to estimate non-equal β_{11}, β_{01}. Assumptions in Case 1 and 6 can be considered a special case of Case 7.

Case 8: Assumptions in Case 2 and 5 are jointly sufficient to estimate non-equal β_{11}, β_{01}. Assumptions in Case 2 and 6 can be considered a special case of Case 8.

Case 9: Assumptions in Case 3 and 5 are jointly sufficient to estimate non-equal β_{11}, β_{01}. Assumptions in Case 3 and 6 can be considered a special case of Case 9.

Case 10: Assumptions in Case 4 and 5 are jointly sufficient to estimate non-equal β_{11}, β_{01}. Assumptions in Case 4 and 6 can be considered a special case of Case 10.

Case 11: For Cases 1-4 in addition we could also use the same treatment effect assumption ($\beta_{11} = \beta_{01}$) to identify the other group.

Here we show that [Lee and Kim, 2014] proposed models are special cases of FD_1 and FD_2.

[2]Supra note

[Lee and Kim, 2014] consider panel linear models L_1 and L_2. In both models, time-differences in potential untreated outcomes are a combination of time-effects[3] ($\triangle\beta_3$) and group-effects (β_q). While these effects are constant they are not applied equally to all subgroups. From table 6a it is clear that model L_1 assumes $\alpha_{00} = \alpha_{11} = \triangle\beta_3$ (Case 3), $\alpha_{01} = \alpha_{10} = \triangle\beta_3 + \beta_q$ (Case 3) and $\beta_{01} = \beta_{11}$ (Case 11). Equal coefficient assumption makes one of the other two assumptions redundant. Instead Model L_2 assumes $\alpha_{00} = \alpha_{11} = \triangle\beta_3$, $\alpha_{01} = \alpha_{10} = \triangle\beta_3 + \beta_q$ and different treatment effects (table 6b).

$$(L_1)\begin{cases} Y_{it}^0 &= \beta_t + \beta_q Q_{i,t} + \beta_w' W_{it} + V_{it} \\ Y_{it}^0 &= Y_{it}^0 + \beta_d \end{cases} \qquad (L_2)\begin{cases} Y_{it}^0 &= \beta_t + \beta_q Q_{i,t} + \beta_w' W_{it} + V_{it} \\ Y_{it}^0 &= Y_{it}^0 + \beta_d Q_{i,t-1} + \beta_m (1 - Q_{i,t-1}) \end{cases}$$

Table 6: POM representation

(a) L1 model

Groups		Y_2^0		Y_3^0		Y_3^1	
out-stayers ($Q_2 = 0, Q_3 = 0$)	in-movers ($Q_2 = 0, Q_3 = 1$)	β_2	β_2	β_3	$\beta_3 + \beta_q$	$\beta_3 + \beta_d$	$\beta_3 + \beta_q + \beta_d$
out-movers ($Q_2 = 1, Q_3 = 0$)	in-stayers ($Q_2 = 1, Q_3 = 1$)	$\beta_2 + \beta_q$	$\beta_2 + \beta_q$	β_3	$\beta_3 + \beta_q$	$\beta_3 + \beta_d$	$\beta_3 + \beta_q + \beta_d$

(b) L2 model

Groups		Y_2^0		Y_3^0		Y_3^1	
out-stayers ($Q_2 = 0, Q_3 = 0$)	in-movers ($Q_2 = 0, Q_3 = 1$)	β_2	β_2	β_3	$\beta_3 + \beta_q$	$\beta_3 + \beta_d$	$\beta_3 + \beta_q + \beta_d$
out-movers ($Q_2 = 1, Q_3 = 0$)	in-stayers ($Q_2 = 1, Q_3 = 1$)	$\beta_2 + \beta_q$	$\beta_2 + \beta_q$	β_3	$\beta_3 + \beta_q$	$\beta_3 + \beta_m$	$\beta_3 + \beta_q + \beta_m$

3.3 Parametric vs. non-parametric assumptions

In the previous subsection we claimed that assumptions made in Case 2 and 3 are equivalent respectively to *"movers-"* and *"stayers-" "same time-effect"* identified by [Lee and Kim, 2014]. This is true if we add to the structural model M_1, the assumption of zero conditional mean of the error for each subgroup $E(V_{it}|Q_{i2} = i, Q_{i3} = j, W_{it}) = 0, \quad t = 2, 3 \quad i, j = 0, 1$. This assumption is sufficient (also stronger) than the usual $E(V_{it}|Q_{i2}, Q_{i3}, W_{it}) = 0, \quad t = 2, 3$. With the added assumption the following conditional means are derived

$$\begin{aligned}
&E(Y_{i2}^0|Q_{i2} = 0, Q_{i3} = 0, W_{i2}) = \gamma_{00} + \beta_w' W_{i2} \quad E(Y_{i3}^0|Q_{i2} = 0, Q_{i3} = 0, W_{i3}) = \delta_{00} + \beta_w' W_{i3} \\
&E(Y_{i2}^0|Q_{i2} = 1, Q_{i3} = 0, W_{i2}) = \gamma_{10} + \beta_w' W_{i2} \quad E(Y_{i3}^0|Q_{i2} = 1, Q_{i3} = 0, W_{i3}) = \delta_{10} + \beta_w' W_{i3} \\
&E(Y_{i2}^0|Q_{i2} = 0, Q_{i3} = 1, W_{i2}) = \gamma_{01} + \beta_w' W_{i2} \quad E(Y_{i3}^1|Q_{i2} = 0, Q_{i3} = 1, W_{i3}) = \delta_{01} + \beta_w' W_{i3} \\
&E(Y_{i2}^0|Q_{i2} = 1, Q_{i3} = 1, W_{i2}) = \gamma_{11} + \beta_w' W_{i2} \quad E(Y_{i3}^1|Q_{i2} = 1, Q_{i3} = 1, W_{i3}) = \delta_{11} + \beta_w' W_{i3}
\end{aligned} \tag{CE_1}$$

[3] The definition of time-effects in these model is different from that of time-difference in potential untreated outcomes

"Stayers same time-effect" has been defined as

$$E(Y_{i3}^0|Q_{i2}=1,Q_{i3}=1,W_{i3}) - E(Y_{i2}^0|Q_{i2}=1,Q_{i3}=1,W_{i2}) = \\ E(Y_{i3}^0|Q_{i2}=0,Q_{i3}=0,W_{i3}) - E(Y_{i2}^0|Q_{i2}=0,Q_{i3}=0,W_{i2})$$

and from conditional expectations in CE_1 it is easy to see that the condition is equivalent to

$$\alpha_{11} = \delta_{11} - \gamma_{11} = \alpha_{00} = \delta_{00} - \gamma_{00}$$

In the same way it can be shown that *"movers same time-effect"* defined as

$$E(Y_{i3}^0|Q_{i2}=0,Q_{i3}=1,W_{i3}) - E(Y_{i2}^0|Q_{i2}=0,Q_{i3}=1,W_{i2}) = \\ E(Y_{i3}^0|Q_{i2}=1,Q_{i3}=0,W_{i3}) - E(Y_{i2}^0|Q_{i2}=1,Q_{i3}=0,W_{i2})$$

is equivalent to $\alpha_{01} = \alpha_{10}$. In addition, we can see from CE_1, that non-parametric assumptions linear on $E(\triangle Y_{i3}^0|W_i)$ are equivalent to linear combinations of $\alpha_{..}$.

Lastly we emphasize the meaning of $\gamma_{..}$ and $\delta_{..}$. As already pointed out in table 4, these coefficients can be considered as the case when outcome is determined only by treatment and no other covariates or as conditional mean evaluated at $W_{it} = 0$. For example $E(Y_{i2}^0|Q_{i2} = 0, Q_{i3} = 0, W_{i2} = 0) = \gamma_{00}$. Note that usually $E(Y_{i2}^0|Q_{i2} = 0, Q_{i3} = 0) = E(E(Y_{i2}^0|Q_{i2} = 0, Q_{i3} = 0, W_i)) = \gamma_{00} + \beta_w' E(W_{i2}) \neq \gamma_{00} = E(Y_{i2}^0|Q_{i2} = 0, Q_{i3} = 0, W_{i2} = 0)$.

3.4 RE estimator

Fixed effects estimators do not allow for time-invariant covariates, unless we apply a dummy-variable approach as with subgroup variables. Sometimes it is preferred to control for the effects of time-invariant covariates such as gender, therefore random effects methods are used. Model P_1 can be employed for random-effects estimation depending on the structure we impose on the unobservables V_{it}. [Wooldridge, 2010] provides a very good treatment of random effects estimation in linear panel models.

$$\begin{aligned} Y_{it} =& \gamma_{00}G_{00}1[t=2] + \gamma_{10}G_{10}1[t=2] + \gamma_{01}G_{01}1[t=2] + \gamma_{11}G_{11}1[t=2] + \delta_{00}G_{00}1[t=3] \\ &+ \delta_{10}G_{10}1[t=3] + (\delta_{01}+\beta_{01})G_{01}1[t=3] + (\delta_{11}+\beta_{11})G_{11}1[t=3] + \beta_w' W_{it} + V_{it} \end{aligned} \tag{P_1}$$

The three major comments mentioned in first difference estimation are valid also for random effects estimation. First, only β_{01} and β_{11} can be identified. Second, estimation will be always biased with the potential untreated outcomes at $t = 3$ (or $\delta_{.1}$). Third, potential untreated outcomes at $t = 2$ for all subgroups and at $t = 3$ for *out-groups* can be identified

and estimated. Assuming some form of relationship between any combination of $\delta_{\cdot 1}$ with the other estimable coefficient makes it possible to disentangle treatment effects. Assumptions outlined in first difference estimation could be employed.

4 Treatment effects with repeated cross-section data

We follow the approach of [Lee and Kang, 2006] when considering the case with cross-sectional data. In this setting, the timing for statistical unit i to be observed is random and is denoted by $S_i = 1[\text{stat. unit } i \text{ sampled at } t = 3]$. Then the observed outcome takes the form

$$Y_i = (1 - S_i)Y_{i2}^0 + S_i Y_{i3} = (1 - S_i)Y_{i2}^0 + S_i Y_{i3}^0 + S_i Q_{i3}(Y_{i3}^1 - Y_{i3}^0)$$

Considering structural model M_1 in a repeated cross-section setting gives observed outcome equation C_1 (appendix D)

$$\begin{aligned} Y_i =& (1 - S_i)(1 - Q_{i2})[\gamma_{00}(1 - Q_{i3}) + \gamma_{01}Q_{i3}] + (1 - S_i)Q_{i2}[\gamma_{10}(1 - Q_{i3}) + \gamma_{11}Q_{i3}] + \\ &+ S_i(1 - Q_{i3})[\delta_{00}(1 - Q_{i2}) + \delta_{10}Q_{i2}] + S_i Q_{i3}[(\delta_{01} + \beta_{01})(1 - Q_{i2}) + (\delta_{11} + \beta_{11})Q_{i2}] \\ &+ \beta_w' W_i + V_i \end{aligned} \tag{C_1}$$

Equation C_1 is particularly useful since in repeated cross-sections a statistical unit is usually observed only once. This means that we can recover information only for the terms $S_i Q_{i3} = S_i Q_i$, $(1-S_i)Q_{i2} = (1-S_i)Q_i$, $S_i(1-Q_{i3}) = S_i(1-Q_i)$ and $(1-S_i)(1-Q_{i2}) = (1-S_i)(1-Q_i)$. Assuming zero conditional mean of the error for each subgroup $E(V_i|S_i = i, Q_i = j, W_i) = 0, \quad i, j = 0, 1$, which is sufficient (also stronger) than the usual $E(V_i|S_i, Q_i, W_i) = 0$, gives structural conditional mean C_2 (appendix E)

$$\begin{aligned} &E(Y_i|S_i, Q_i, W_i) = \\ &\quad (1 - S_i)(1 - Q_i)[\gamma_{00}P(Q_{i3} = 0|S_i = 0, Q_i = 0, W_i) + \gamma_{01}P(Q_{i3} = 1|S_i = 0, Q_i = 0, W_i)] \\ &+(1 - S_i)Q_i[\gamma_{01}P(Q_{i3} = 0|S_i = 0, Q_i = 1, W_i) + \gamma_{11}P(Q_{i3} = 1|S_i = 0, Q_i = 1, W_i)] \\ &+S_i(1 - Q_i)[\delta_{00}P(Q_{i2} = 0|S_i = 1, Q_i = 0, W_i) + \delta_{10}P(Q_{i2} = 1|S_i = 1, Q_i = 0, W_i)] \\ &+S_i Q_i[(\delta_{01} + \beta_{01})P(Q_{i2} = 0|S_i = 1, Q_i = 1, W_i) + (\delta_{11} + \beta_{11})P(Q_{i2} = 1|S_i = 1, Q_i = 1, W_i)] \\ &+\beta_w' W_i \end{aligned} \tag{C_2}$$

Terms in [.] in equation C_2 are unobservable and depend on W_i, therefore regression on covariates $((1 - S_i)(1 - Q_i), (1 - S_i)Q_i, S_i(1 - Q_i), S_i Q_i, W_i)'$ will produce biased estimates. To eliminate bias we could either assume conditional independence of Q_{i2}, Q_{i3} (i.e. subgroups) on covariates W_i, or assume equal $\gamma_{\cdot\cdot}$ and $\delta_{\cdot\cdot}$ coefficients within [.] terms. Note that both approaches render the terms in [.] into constants. One may choose to employ the approach

that best suits the case under study. For example assuming conditional independence may not be reasonable in the case where poor people cannot switch treatments and are more likely to be *out-stayers*. On other cases, assuming equal coefficients may be too strong of an assumption. We consider both cases in detail.

When considering the case of conditional independence assumption, we analyze only the first [.] term. The same logic is applied to the remaining [.] terms. Combining conditional independence with the interpretation of $\gamma_{..}$ and $\delta_{..}$ coefficients pointed out in subsection 3.3, transforms first [.] term into [4]

$$\begin{aligned} E(Y^0_{i2}|Q_{i2}=0,&W_{i2}=0) = E(Y^0_{i2}|Q_{i2}=0,Q_{i3}=0,W_{i2}=0)P(Q_{i3}=0|S_i=0,Q_i=0) \\ &+ E(Y^0_{i2}|Q_{i2}=0,Q_{i3}=1,W_{i2}=0)P(Q_{i3}=1|S_i=0,Q_i=0) \qquad (C_3) \end{aligned}$$

C_3 can be considered either as the weighted average of γ_{00} and γ_{01}, or as potential untreated outcome at $t=2$ for the subgroup defined by $Q_{i2}=1$ measured at $W_{it}=0$. From the other [.] terms we would get $E(Y^0_{i2}|Q_{i2}=1,W_{i2}=0), E(Y^0_{i3}|Q_{i3}=0,W_{i2}=0)$ and $E(Y^1_{i3}|Q_{i3}=1,W_{i2}=0)$. For ease of presentation, we use $\bar{\gamma}_{0.}, \bar{\gamma}_{1.}, \bar{\delta}_{.0}, \bar{\delta}_{.1}+\bar{\beta}_{.1}$ notation for the coefficients, where the bar identifies the mean, the dot the unobserved (averaged upon) qualification variable and the Greek letter notation as previously mentioned.

Measuring treatment effects means identifying $\bar{\beta}_{.1}$, which as a result of subgroup treatment homogeneity and conditional independence, represents average treatment effects for *stayer* group ($Q_{i3}=1$). Since we can measure only the sum $\bar{\delta}_{.1}+\bar{\beta}_{.1}$ we need additional assumptions that relate $\bar{\delta}_{.1}$ to the other measurable coefficients $\bar{\gamma}_{0.}, \bar{\gamma}_{1.}, \bar{\delta}_{.0}$. The most common functional form with a clear DD interpretation is $\bar{\delta}_{.1}-\bar{\gamma}_{1.}=\bar{\delta}_{.0}-\bar{\gamma}_{0.}$, which is equivalent to the *identification* assumption for cross-section data identified by [Lee and Kang, 2006].

From C_2 we get the following conditional means

$$\begin{aligned} E(Y^0_{i2}|Q_i=0,S_i=0) &= \bar{\gamma}_{0.}+\beta'_w W_{i2} & E(Y^0_{i3}|Q_i=0,S_i=1) &= \bar{\delta}_{.0}+\beta'_w W_{i3} \\ E(Y^0_{i2}|Q_i=1,S_i=0) &= \bar{\gamma}_{1.}+\beta'_w W_{i2} & E(Y^0_{i3}|Q_i=1,S_i=1) &= \bar{\delta}_{.1}+\beta'_w W_{i3} \\ & & E(Y^1_{i3}|Q_i=1,S_i=1) &= \bar{\delta}_{.1}+\bar{\beta}_{.1}+\beta'_w W_{i3} \qquad (CE_2) \end{aligned}$$

The *"Stayers same time-effect"* has been defined as $E(Y^0_3|Q_3=1)-E(Y^0_2|Q_2=1) = E(Y^0_3|Q_3=0)-E(Y^0_2|Q_2=0))$ and from conditional expectations in CE_2 it is easy to see that the condition is equivalent to $\bar{\delta}_{.1}-\bar{\gamma}_{1.}=\bar{\delta}_{.0}-\bar{\gamma}_{0.}$.

Using conditional means in CE_2 the relation between parametric and non-parametric methods is outlined in DD. In terms of assumptions, the *parametric* method is equivalent to the

[4]Note that $P(Q_{i3}=1|S_i=0,Q_i=0)=P(Q_{i3}=1|Q_{i2}=0)=P(Q_{i3}=1|Q_{i2}=0,W_{i2}=0)$. Last equality derives from conditional independence.

non-parametric DD estimator, but with the inherent advantages of the parametric methods for control on covariates and ease in understanding.

$$E(Y_i|Q_3=1) - E(Y_i|Q_2=1) - [E(Y_i|Q_3=0) - E(Y_i|Q_2=0)] = \bar{\delta}_{.1} + \bar{\beta}_{.1} - \bar{\gamma}_{1.} - (\bar{\delta}_{.0} - \bar{\gamma}_{0.}) \tag{DD}$$

When considering equal coefficients approach, we have presented sufficient assumptions in EQ. Assumptions (iii) and (iv) impose the same Y_2^0 for sub-groups with equal Q_{i2}, (ii) imposes the same Y_3^0 for sub-groups with $Q_{i3} = 0$ and (i) imposes equality on Y_3^1 on the treated sub-groups. It is important to note that (i)-(iv) are sufficient to estimate τ, which also includes bias term. In order to estimate treatment effects we need additional assumptions between $\delta_{.1}$ and the other estimable coefficients.

$$\begin{aligned} &(i) \quad \delta_{01} + \beta_{01} = \delta_{11} + \beta_{11} = \tau && (iii) \quad \gamma_{10} = \gamma_{11} = \gamma_{1.} \\ &(ii) \quad \delta_{00} = \delta_{10} = \delta_{.0} && (iv) \quad \gamma_{00} = \gamma_{01} = \gamma_{0.} \end{aligned} \tag{EQ}$$

In addition to (i)-(iv), assumptions (v) $\delta_{01} = \delta_{11} = \delta_{.1}$[5] and (vi) $\delta_{.1} - \gamma_{1.} = \delta_{.0} - \gamma_{0.}$ are sufficient to estimate treatment effects. Assumption (vi) is only one functional form relating $\delta_{.1}$ to the other measurable potential outcomes $\gamma_{1.}, \gamma_{0.}, \delta_{.0}$. This form is preferred since it provides the clear DD interpretation - groups (not necessary sub-groups) are affected by same unobserved differences. Other forms, such as opposite effects $\delta_{.1} - \gamma_{1.} = -(\delta_{.0} - \gamma_{0.})$, could be employed if supported by the specifics of the case under study [6].

Differing treatment effects, $\beta_{01} \neq \beta_{11}$, can also be measured. Sufficient assumptions would be (i)-(iv) and (vi) together with assumptions that relate δ_{01} and δ_{11} to the other measurable potential outcomes $\gamma_{1.}, \gamma_{0.}, \delta_{.0}$. Same caveats apply.

We conclude with two sid notes. First, note that $(i) - (vi) \Rightarrow E(Y_3^0|Q_3 = 1) - E(Y_2^0|Q_2 = 1) = E(Y_3^0|Q_3 = 0) - E(Y_2^0|Q_2 = 0)$, which is *identification* assumption described by [Lee and Kang, 2006]. Second, note that models of [Lee and Kim, 2014] are special cases of model C_1. For example, model L_1 (also table 6a) fulfills assumptions (i)-(vi); (i) $\tau = \beta_3 + \beta_q + \beta_d$, (ii) $\delta_{.0} = \beta_3$, (iii) $\gamma_{1.} = \beta_2 + \beta_q$, (iv) $\gamma_{0.} = \beta_2$, (v) $\delta_{.1} = \beta_3 + \beta_q$ and (vi) $\delta_{.1} - \gamma_{1.} = \delta_{.0} - \gamma_{0.} = \triangle\beta_3$.

[5] (i) and (v) are equivalent to (i) and $\beta_{01} = \beta_{11}$

[6] There is no limit to the functional form. Same percent change in untreated potential outcomes can also be considered $\frac{\delta_{.1} - \gamma_{1.}}{\gamma_{1.}} = \frac{\delta_{.0} - \gamma_{0.}}{\gamma_{0.}}$

5 Discussions

We have considered in full the case for dichotomous time-varying qualification and 2 time periods, which we believe is the most frequent case for treatment effects with time-varying qualification. Our approach of identifying and considering *as given* all the possible subgroups introduced by the time-varying qualification can be extended to the cases where the qualifying variable represents more than one treatment and the time frame extends to 3 or more periods. Also allowing switching of treatments would affect only the set of subgroups considered, but not the general approach. The above outlined cases may require bigger datasets because of the higher number of parameters to be estimated.

References

[Lee and Kang, 2006] Lee, M.-j. and Kang, C. (2006). Identification for difference in differences with cross-section and panel data. *Economics letters*, 92(2):270–276.

[Lee and Kim, 2013] Lee, M.-J. and Kim, Y.-S. (2013). Difference in differences for stayers and movers with a time-varying qualification. Technical report, HEDG, c/o Department of Economics, University of York.

[Lee and Kim, 2014] Lee, M.-J. and Kim, Y.-S. (2014). Difference in differences for stayers with a time-varying qualification: Health expenditure elasticity of the elderly. *Health economics*, 23(9):1134–1145.

[Wooldridge, 2010] Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

A Deriving the panel model

We derive the results using the Q-notation. Going to the G-notation is straightforward.

$$\begin{aligned}
Y_{it} =& Y_{i2}1[t=2] + Y_{i3}1[t=3] = Y^0_{i2}1[t=2] + Y^0_{i3}1[t=3] + (Y^1_{i3} - Y^0_{i3})Q_{i,3}1[t=3] \\
=& \gamma_{00}(1-Q_{i,2})(1-Q_{i,3})1[t=2] + \gamma_{10}Q_{i,2}(1-Q_{i,3})1[t=2] + \gamma_{01}(1-Q_{i,2})Q_{i,3}1[t=2] \\
&+ \gamma_{11}Q_{i,2}Q_{i,3}1[t=2] + \beta_w' W_{i2}1[t=2] + V_{i2}1[t=2] + \delta_{00}(1-Q_{i,2})(1-Q_{i,3})1[t=3] \\
&+ \delta_{10}Q_{i,2}(1-Q_{i,3})1[t=3] + \delta_{01}(1-Q_{i,2})Q_{i,3}1[t=3] + \delta_{11}Q_{i,2}Q_{i,3}1[t=3] \\
&+ \beta_w' W_{i3}1[t=3] + V_{i3}1[t=3] + \eta Q_{i,3}1[t=3]
\end{aligned}$$

where $\eta = \beta_{00}(1-Q_{i,2})(1-Q_{i,3}) + \beta_{10}Q_{i,2}(1-Q_{i,3}) + \beta_{01}(1-Q_{i,2})Q_{i,3} + \beta_{11}Q_{i,2}Q_{i,3}$. Noting that $Q^2_{it} = Q_{it}$ and $(1-Q_{it})Q_{it} = 0$, the terms measuring treatment effects becomes $\eta Q_{i,3} = \beta_{01}(1-Q_{i,2})Q_{i,3} + \beta_{11}Q_{i,2}Q_{i,3}$. Also $\beta_w'(W_{i2}1[t=2] + W_{i3}1[t=3]) = \beta_w' W_{it}$ and $V_{i2}1[t=2] + V_{i3}1[t=3] = V_{it}$. Then observed outcome is expressed as

$$\begin{aligned}
Y_{it} =& \gamma_{00}(1-Q_{i,2})(1-Q_{i,3})1[t=2] + \gamma_{10}Q_{i,2}(1-Q_{i,3})1[t=2] + \gamma_{01}(1-Q_{i,2})Q_{i,3}1[t=2] \\
&+ \gamma_{11}Q_{i,2}Q_{i,3}1[t=2] + \delta_{00}(1-Q_{i,2})(1-Q_{i,3})1[t=3] + \delta_{10}Q_{i,2}(1-Q_{i,3})1[t=3] \\
&+ (\delta_{01}+\beta_{01})(1-Q_{i,2})Q_{i,3}1[t=3] + (\delta_{11}+\beta_{11})Q_{i,2}Q_{i,3}1[t=3] + \beta_w' W_{it} + V_{it}
\end{aligned}$$

Substituting $G_{00} = (1-Q_{i,2})(1-Q_{i,3}), G_{01} = (1-Q_{i,2})Q_{i,3}, G_{10} = Q_{i,2}(1-Q_{i,3}), G_{11} = Q_{i,2}Q_{i,3}$, gives G-notation.

B Deriving first difference equation

Deriving the first difference equation is straightforward if we employ model O_1. For completeness we illustrate most of the steps how to derive first differences from model P_1. We use G-notation for shortness of presentation. Also note that when we are considering specific observations, $1[t=2]$ becomes either $1[3=2]$ or $1[3=2]$, and so on.

$$\begin{aligned}
Y_{i2} =& \gamma_{00}G_{00}1[2=2] + \gamma_{10}G_{10}1[2=2] + \gamma_{01}G_{01}1[2=2] + \gamma_{11}G_{11}1[2=2] + \delta_{00}G_{00}1[2=3] \\
&+ \delta_{10}G_{10}1[2=3] + (\delta_{01}+\beta_{01})G_{01}1[2=3] + (\delta_{11}+\beta_{11})G_{11}1[2=3] + \beta_w' W_{i3} + V_{i3}
\end{aligned}$$

$$\begin{aligned}
Y_{i3} =& \gamma_{00}G_{00}1[3=2] + \gamma_{10}G_{10}1[3=2] + \gamma_{01}G_{01}1[3=2] + \gamma_{11}G_{11}1[3=2] + \delta_{00}G_{00}1[3=3] \\
&+ \delta_{10}G_{10}1[3=3] + (\delta_{01}+\beta_{01})G_{01}1[3=3] + (\delta_{11}+\beta_{11})G_{11}1[3=3] + \beta_w' W_{i3} + V_{i3}
\end{aligned}$$

$$
\begin{aligned}
\triangle Y_{i3} =& Y_{i3} - Y_{i2} \\
=& \gamma_{00} G_{00} (\underbrace{1[3=2] - 1[2=2]}_{=-1}) + \gamma_{01} G_{01} (\underbrace{1[3=2] - 1[2=2]}_{=-1}) + \gamma_{10} G_{10} (\underbrace{1[3=2] - 1[2=2]}_{=-1}) \\
&+ \gamma_{11} G_{11} (\underbrace{1[3=2] - 1[2=2]}_{=-1}) + \delta_{00} G_{00} (\underbrace{1[3=3] - 1[2=3]}_{=1}) + \delta_{10} G_{10} (\underbrace{1[3=3] - 1[2=3]}_{=1}) \\
&+ (\delta_{01} + \beta_{01}) G_{01} (\underbrace{1[3=3] - 1[2=3]}_{=1}) + (\delta_{11} + \beta_{11}) G_{11} (\underbrace{1[3=3] - 1[2=3]}_{=1}) + \beta_x' \triangle X_{i3} + \triangle U_{i3} \\
=& (\underbrace{\delta_{00} - \gamma_{00}}_{=\alpha_{00}}) G_{00} + (\underbrace{\delta_{10} - \gamma_{10}}_{=\alpha_{10}}) G_{10} + (\underbrace{\delta_{01} - \gamma_{01}}_{=\alpha_{01}} + \beta_{01}) G_{01} + (\underbrace{\delta_{11} - \gamma_{11}}_{=\alpha_{11}} + \beta_{11}) G_{11} + \beta_x' \triangle X_{i3} + \triangle U_{i3}
\end{aligned}
$$

C Alternative representation in Q-notation

The alternative Q-representation uses the terms $1, Q_{i,2}, Q_{i,3}, Q_{i,2}Q_{i,3}$ instead of $(1-Q_{i,2})(1-Q_{i,3}), (1-Q_{i,2})Q_{i,3}, Q_{i,2}(1-Q_{i,3}), Q_{i,2}Q_{i,3}$. Therefore potential outcomes could also be represented as

$$
\begin{aligned}
&\gamma_{00}(1-Q_{i,2})(1-Q_{i,3}) + \gamma_{10}Q_{i,2}(1-Q_{i,3}) + \gamma_{01}(1-Q_{i,2})Q_{i,3} + \gamma_{11}Q_{i,2}Q_{i,3} = \\
&\qquad \gamma_{00} + (\gamma_{10}-\gamma_{00})Q_{i,2} + (\gamma_{01}-\gamma_{00})Q_{i,3} + (\gamma_{11}-\gamma_{10}-\gamma_{01}+\gamma_{00})Q_{i,2}Q_{i,3}
\end{aligned}
$$

Then model M_1 can be presented as model M_1'.

$$
\begin{aligned}
Y_{i2}^0 &= \gamma_{00} + (\gamma_{10}-\gamma_{00})Q_{i,2} + (\gamma_{01}-\gamma_{00})Q_{i,3} + (\gamma_{11}-\gamma_{10}-\gamma_{01}+\gamma_{00})Q_{i,2}Q_{i,3} + \beta_w' W_{i2} + V_{i2} \\
Y_{i3}^0 &= \delta_{00} + (\delta_{10}-\delta_{00})Q_{i,2} + (\delta_{01}-\delta_{00})Q_{i,3} + (\delta_{11}-\delta_{10}-\delta_{01}+\delta_{00})Q_{i,2}Q_{i,3} + \beta_w' W_{i3} + V_{i3} \\
Y_{i3}^1 &= Y_{i3}^0 + \underbrace{\beta_{00} + (\beta_{10}-\beta_{00})Q_{i,2} + (\beta_{01}-\beta_{00})Q_{i,3} + (\beta_{11}-\beta_{10}-\beta_{01}+\beta_{00})Q_{i,2}Q_{i,3}}_{=\theta} \qquad (M_1')
\end{aligned}
$$

Always by considering information on $Q_{i,2}, Q_{i,3}$ as given and noting that $\theta Q_{i,3} = \beta_{01}Q_{i,3} + (\beta_{11}-\beta_{01})Q_{i,2}Q_{i,3}$ we could either derive either P_1' or O_1' from M_1'.

$$
\begin{aligned}
Y_{it} =& \gamma_{00}1[t=2] + (\gamma_{10}-\gamma_{00})Q_{i,2}1[t=2] + (\gamma_{01}-\gamma_{00})Q_{i,3}1[t=2] \\
&+ (\gamma_{11}-\gamma_{10}-\gamma_{01}+\gamma_{00})Q_{i,2}Q_{i,3}1[t=2] + \delta_{00}1[t=3] + (\delta_{10}-\delta_{00})Q_{i,2}1[t=3] \\
&+ (\delta_{01}-\delta_{00}+\beta_{01})Q_{i,3}1[t=3] + (\delta_{11}-\delta_{10}-\delta_{01}+\delta_{00}+\beta_{11}-\beta_{01})Q_{i,2}Q_{i,3}1[t=3] \\
&+ \beta_w' W_{it} + V_{it} \qquad (P_1')
\end{aligned}
$$

$$
\begin{aligned}
Y_{i2} &= \gamma_{00} + (\gamma_{10}-\gamma_{00})Q_{i,2} + (\gamma_{01}-\gamma_{00})Q_{i,3} + (\gamma_{11}-\gamma_{10}-\gamma_{01}+\gamma_{00})Q_{i,2}Q_{i,3} + \beta_w' W_{i2} + V_{i2} \\
Y_{i3} &= \delta_{00} + (\delta_{10}-\delta_{00})Q_{i,2} + (\delta_{01}-\delta_{00}+\beta_{01})Q_{i,3} + (\delta_{11}-\delta_{10}-\delta_{01}+\delta_{00}+\beta_{11}-\beta_{01})Q_{i,2}Q_{i,3} \\
&\quad + \beta_w' W_{i3} + V_{i3} \qquad (O_1')
\end{aligned}
$$

Using the same approach as in appendix B we arrive at the alternative Q-notation first difference equation

$$\begin{aligned}\triangle Y_{i3} =&\alpha_{00} + (\alpha_{10} - \alpha_{00})Q_{i,2} + (\alpha_{01} - \alpha_{00} + \beta_{01})Q_{i,3} \\ &+ (\alpha_{11} - \alpha_{10} - \alpha_{01} + \alpha_{00} + \beta_{11} - \beta_{01})Q_{i,2}Q_{i,3} + \beta_x' \triangle X_{i3} + \triangle U_{i3}\end{aligned}$$

D Deriving the repeated cross-sections model

We use model M_1 and the Q-representation.

$$\begin{aligned}Y_i =&(1 - S_i)Y_{i2}^0 + S_i Y_{i3} = (1 - S_i)Y_{i2}^0 + S_i Y_{i3}^0 + S_i Q_{i3}(Y_{i3}^1 - Y_{i3}^0) \\ =&(1 - S_i)(1 - Q_{i,2})[\gamma_{00}(1 - Q_{i,3}) + \gamma_{01}Q_{i,3}] + (1 - S_i)Q_{i2}[\gamma_{10}(1 - Q_{i3}) + \gamma_{11}Q_{i3}] + \\ &+ (1 - S_i)\beta_w' W_{i2} + (1 - S_i)V_{i2} + S_i(1 - Q_{i3})[\delta_{00}(1 - Q_{i2}) + \delta_{10}Q_{i2}] \\ &+ S_i Q_{i3}[\delta_{01}(1 - Q_{i2}) + \delta_{11}Q_{i2}] + S_i\beta_w' W_{i3} + S_i V_{i3} \\ &+ S_i Q_{i3}[\beta_{01}(1 - Q_{i2}) + \beta_{11}Q_{i2}] + \underbrace{Q_{i3}(1 - Q_{i3})}_{=0} S_i[\beta_{10}(1 - Q_{i2}) + \beta_{00}Q_{i2}] \\ =&(1 - S_i)(1 - Q_{i2})[\gamma_{00}(1 - Q_{i3}) + \gamma_{01}Q_{i3}] + (1 - S_i)Q_{i2}[\gamma_{10}(1 - Q_{i3}) + \gamma_{11}Q_{i3}] + \\ &+ S_i(1 - Q_{i3})[\delta_{00}(1 - Q_{i2}) + \delta_{10}Q_{i2}] + S_i Q_{i3}[(\delta_{01} + \beta_{01})(1 - Q_{i2}) + (\delta_{11} + \beta_{11})Q_{i2}] \\ &+ \beta_w' \underbrace{[(1 - S_i)W_{i2} + S_i W_{i3}]}_{=W_i} + \underbrace{[(1 - S_i)V_{i2} + S_i V_{i3}]}_{=V_i}\end{aligned}$$

E Structural conditional mean for repeated cross-sections

In addition to the structural model C_1 we need the assumption of zero conditional mean of the error for each subgroup $E(V_i|S_i = i, Q_i = j, W_i) = 0, \quad i, j = 0, 1$. This assumption is sufficient (also stronger) than the usual $E(V_i|S_i, Q_i, W_i) = 0$.

$$\begin{aligned}E(Y_i|S_i, Q_i, W_i) =& (1 - S_i)(1 - Q_i)[\gamma_{00}(1 - E(Q_{i3}|S_i, Q_i, W_i)) + \gamma_{01}E(Q_{i3}|S_i, Q_i, W_i)] \\ &+ (1 - S_i)Q_i[\gamma_{01}(1 - E(Q_{i3}|S_i, Q_i, W_i) + \gamma_{11}E(Q_{i3}|S_i, Q_i, W_i)] \\ &+ S_i(1 - Q_i)[\delta_{00}(1 - E(Q_{i2}|S_i, Q_i, W_i) + \delta_{10}E(Q_{i2}|S_i, Q_i, W_i)] \\ &+ S_i Q_i[(\delta_{01} + \beta_{01})(1 - E(Q_{i2}|S_i, Q_i, W_i) + (\delta_{11} + \beta_{11})E(Q_{i2}|S_i, Q_i, W_i)] + \beta_w' W_i \\ =&(1 - S_i)(1 - Q_i)[\gamma_{00}P(Q_{i3} = 0|S_i, Q_i, W_i) + \gamma_{01}P(Q_{i3} = 1|S_i, Q_i, W_i)] \\ &+ (1 - S_i)Q_i[\gamma_{01}P(Q_{i3} = 0|S_i, Q_i, W_i) + \gamma_{11}P(Q_{i3} = 1|S_i, Q_i, W_i)] \\ &+ S_i(1 - Q_i)[\delta_{00}P(Q_{i2} = 0|S_i, Q_i, W_i) + \delta_{10}P(Q_{i2} = 1|S_i, Q_i, W_i)] \\ &+ S_i Q_i[(\delta_{01} + \beta_{01})P(Q_{i2} = 0|S_i, Q_i, W_i) + (\delta_{11} + \beta_{11})P(Q_{i2} = 1|S_i, Q_i, W_i)] + \beta_w' W_i\end{aligned}$$

We get C_2 noting that the product (.)(.)[.] is nonzero only when the respective (.)(.) combination is non-zero. For example

$$\underbrace{(1-S_i)}_{=(.)}\underbrace{(1-Q_i)}_{=(.)}\underbrace{[\gamma_{00}P(Q_{i3}=0|S_i,Q_i,W_i)+\gamma_{01}P(Q_{i3}=1|S_i,Q_i,W_i)]}_{=[.]} =$$

$$(1-S_i)(1-Q_i)[\gamma_{00}P(Q_{i3}=0|S_i=0,Q_i=0,W_i)+\gamma_{01}P(Q_{i3}=1|S_i=0,Q_i=0,W_i)]$$