HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 13/17

# Differences in Differences for Stayers and Movers with Time-Varying Qualification

Young-Sook Kim & Myoung-Jae Lee

August 2013

york.ac.uk/res/herc/hedgwp

# Difference in Differences for Stayers and Movers with a Time-Varying Qualification

(July 3, 2013)

Myoung-jae Lee*
Department of Economics
Korea University
Seoul 136-701, South Korea

Young-sook Kim
Korean Women's Development Institute
Seoul 122-707, South Korea
youngkim@kwdimail.re.kr

When a treatment in difference in differences (DD) is applied to a group of individuals with a time-constant qualification/eligibility such as gender or race, there is no issue of choosing to be (un-) treated. But if the qualification is time-varying, then the individuals may alter their qualifications to get (un-) treated. In this case, the conventional DD ignoring qualification changes is invalid in general, and it is interesting to know whether the newly qualified (the "in-movers") will experience the same treatment effect as those already qualified (the "in-stayers"). In this paper, *we explore identifying treatment effects with DD when the qualification is time-varying, and propose 'the effect on the in-stayers' as the main effect of interest.* The identification is implemented with a simple nonparametric estimator or with least squares estimator for panel linear models. An empirical analysis is provided using Korean data for the effects of the Basic Elder Pension on health-care expenditure, which shows that the conventional DD ignoring qualification changes is misleading by mixing the effect on the in-stayers with the effect on the in-movers.

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

Running Head: DD for Stayers

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

* Corresponding author: myoungjae@korea.ac.kr; 82-2-3290-2229 (phone); 82-2-926-3601 (fax).

## 1 Introduction

Difference in differences (DD) is a popular study design in finding the effects of a treatment. Many DD references can be found in Angrist and Krueger (1999), Heckman et al. (1999), Shadish et al. (2002), Besley and Case (2004), Bertrand et al. (2004), Athey and Imbens (2006) and Angrist and Pischke (2009). In DD with two periods, the treatment is applied to a group of qualified/eligible individuals. The qualification $Q$ can be time-constant as gender and race, or time-varying as having low income or no children.

Let the qualification dummy for person $i$ at time $t$ be $Q_{it}$, and the response variable be $Y_{it}$. When $Q_{it}$ is time-varying and the treatment is applied at $t = 3$, taking $Q_{i3} = 1$ as the treatment group and $Q_{i3} = 0$ as the control group can be misleading due to the *untreated moving effect of* $Q_{i2} \neq Q_{i3}$ *affecting* $Y_{i3}$ *regardless of the treatment.* For instance, consider an anti-smoking policy applied to big firms of 30 employees more, where $Y_{it}$ is the smoking amount of person $i$ at time $t$, and $Q_{it} = 1$ if person $i$ works in a big firm at time $t$. The atmosphere in big firms may be more conducive to anti-smoking, and consequently, the movers' *untreated* $Y_{it}$ many change simply due to the move. In the following, $i$ indexing individuals will be often omitted.

To see the point better, consider four groups of 'stayers' and 'movers' based on $(Q_2, Q_3)$:

| $Q_3 = 0$ group | $Q_3 = 1$ group |
|---|---|
| $Q_2 = 0, Q_3 = 0$ : out-stayers | $Q_2 = 0, Q_3 = 1$ : in-movers |
| $Q_2 = 1, Q_3 = 0$ : out-movers | $Q_2 = 1, Q_3 = 1$ : in-stayers |

The conventional DD based only on $Q_3$ would call $Q_3 = 0$ and $Q_3 = 1$ the "control" and "treatment" groups. Suppose there is no genuine treatment effect whereas there is an untreated moving effect that is 0 for the stayers, + for $(Q_2 = 0, Q_3 = 1)$ and − for $(Q_2 = 1, Q_3 = 0)$. Then, with $\Delta Y_3 \equiv Y_3 - Y_2$,

$$\begin{aligned} (i) &: E(\Delta Y_3|Q_3 = 1) > 0 > E(\Delta Y_3|Q_3 = 0) \implies E(\Delta Y_3|Q_3 = 1) - E(\Delta Y_3|Q_3 = 0) > 0 \\ (ii) &: E(\Delta Y_3|Q_2 = 1, Q_3 = 1) - E(\Delta Y_3|Q_2 = 0, Q_3 = 0) = 0 : \end{aligned} \quad (1.1)$$

the in-movers in the $Q_3 = 1$ group make $E(\Delta Y_3|Q_3 = 1) > 0$ whereas the out-movers in the $Q_3 = 0$ group makes $E(\Delta Y_3|Q_3 = 0) < 0$, which then gives a false positive DD effect in (1.1)(i), although comparing the in-stayers and the out-stayers gives zero effect in (1.1)(ii).

Even when there is a genuine treatment effect, the effect may differ between the in-movers ($Q_2 = 0, Q_3 = 1$) and the in-stayers ($Q_2 = 1, Q_3 = 1$)—both have $Q_3 = 1$. If the former is zero while the latter is not so that the combined effect on $Q_3 = 1$ is not zero, then changing $Q_3$ to get treated would be in vain. Although the probability of moving to seek/avoid the treatment may be slim in the anti-smoking example, people do migrate to a different state to get a higher income or better welfare benefit (e.g., Kennan and Walker 2010).

Our main proposal is that, to avoid the untreated moving effect, the 'stayer *DD' as in (1.1)(ii) should be used, instead of the conventional DD as in (1.1)(i). As it turns out, the stayer DD identifies the effect on 'the treated' where 'the treated' are the in-stayers (*$Q_2 = 1, Q_3 = 1$*), neither the in-movers* ($Q_2 = 0, Q_3 = 1$) *nor those with only* $Q_3 = 1$ as in the conventional DD.

We will use a simple nonparametric estimator, as well as least squares estimator (LSE) for panel linear models to implement our proposal. In the panel linear model, by imposing more structure, we can separately identify the effect on in-stayers, the effect on in-movers and the untreated moving effect. Often DD is estimated with repeated cross-sections, but since we need both $Q_2$ and $Q_3$, repeated cross-sections will not do. We will require balanced panel data of two (or three) periods.

The rest of this paper is organized as follows. Section 2 examines nonparametric identification and estimation of DD without specifying any model for $Y_{it}$. Section 3 uses panel linear models to easily implement the identification findings and to enhance understanding of DD. Section 4 makes various remarks. Section 5 presents an empirical analysis using Korean data for the effects of a subsistence elderly pension on health-care expenditure, which shows that the conventional DD ignoring $Q_t$ changes is misleading by mixing the effects on in-stayers and in-movers. Finally, Section 6 concludes.

## 2 Nonparametric DD

### 2.1 Basic Set-up and Notation

Let the treatment be $D_{it}$, and let $W_{it}$ be covariates including time-constant ones $C_i$ and time-varying ones $X_{it}$. In terms of timing within period $t$, $W_{it}$ is realized first and then $D_{it}$ in the beginning of period $t$, followed by $Y_{it}$ at the end of period $t$; both $W_{it}$ and $D_{it}$ can affect $Y_{it}$ as depicted in Diagram 1. Although $W_{it}$ may look like consisting only of the period-$t$

covariates, $W_{it}$ may include lagged time-varying covariates.

Diagram 1

$$\begin{array}{ccccc} W_{it} & \longrightarrow & \longrightarrow & \longrightarrow & Y_{it} \\ & \searrow & & \nearrow & \\ & & D_{it} & & \end{array}$$

Consider DD with a time-varying qualification $Q_{it}$ with the treatment applied at $t = 3$ to only the $Q_{i3} = 1$ group, whereas the $Q_{i3} = 0$ group is not treated at all during the entire time frame. One example is the aforementioned anti-smoking policy, and another example is a government-administered job-training at $t = 3$ targeted at low-income earners with $Y_{it}$ being income: $Q_{i3} = 1[Y_{i2} \text{ is low}]$ where $1[A] = 1$ if $A$ holds and 0 otherwise. That is,

$$D_{it} \equiv Q_{it} 1[t = 3].$$

We use $t = 3$ and $t = 2$ as the treatment and control periods to avoid confusion with 0 and 1 used for 'untreated' and 'treated'; also, if a lagged response is included in $W_{it}$, then we will need at least three periods ($t = 1, 2, 3$). Bear in mind the distinction between the *treatment indicator* $D_{it}$ and *the qualification indicator* $Q_{it}$.

Following the treatment effect literature as in Rosenbaum (2002), Lee (2005) and Imbens and Wooldridge (2009) among others, let $Y_{it}^1$ be the 'potential' treated response of person $i$ at period $t$, and $Y_{it}^0$ the potential untreated response, whereas the observed response is

$$Y_{it} = D_{it} Y_{it}^1 + (1 - D_{it}) Y_{it}^0 \implies Y_{it} = Y_{it}^0 \text{ for } t \leq 2 \quad \text{and} \quad Y_{i3} = D_{i3} Y_{i3}^1 + (1 - D_{i3}) Y_{i3}^0.$$

Before proceeding further, more words on notation are needed. Denote the conditional independence of $A$ and $B$ given $C$ as '$A \amalg B|C$'. In most parts of this paper, 'mean independence' rather than independence will be enough because DD is discussed using conditional means, but we will use independence because it makes "neater" presentations; e.g., a joint independence can be split into a conditional and its marginal versions, which however does not have its counter-part in mean independence. Conditional means such as $E(A|B = b)$ will be sometimes denoted just as $E(A|b)$. Define

$$W_{i,t-1}^t \equiv (C_i', X_{i,t-1}', X_{it}')' \quad \text{and} \quad Q_{i,t-1}^t \equiv Q_{i,t-1} Q_{it}.$$

In most cases, a realized value of a random vector $A$ will be denoted by its lower-case $a$. What is observed is

$$\{D_{it}, Q_{it}, W_{it}, Y_{it},\ t = 1, 2, 3\},\ i = 1, ..., \tilde{N}, \text{ that are iid (independent \& identically distributed).}$$

Define *four '$W_2^3$-conditional effects at the post-treatment period $t = 3$'*:

$$\begin{aligned}
\text{effect on out-stayers} &: E(Y_3^1 - Y_3^0|W_2^3, Q_2 = 0, Q_3 = 0); \\
\text{effect on in-movers} &: E(Y_3^1 - Y_3^0|W_2^3, Q_2 = 0, Q_3 = 1); \\
\text{effect on out-movers} &: E(Y_3^1 - Y_3^0|W_2^3, Q_2 = 1, Q_3 = 0); \\
\textit{effect on in-stayers} &: E(Y_3^1 - Y_3^0|W_2^3, Q_2 = 1, Q_3 = 1).
\end{aligned} \tag{2.1}$$

The last one, 'the effect on in-stayers', will be our main effect of interest.

## 2.2 Conventional DD and Stayer DD

The conventional (mean-based) nonparametric DD is

$$\begin{aligned}
&E(\Delta Y_3|W_2^3, Q_3 = 1) - E(\Delta Y_3|W_2^3, Q_3 = 0) \\
&= E(Y_3^1 - Y_2^0|W_2^3, Q_3 = 1) - E(Y_3^0 - Y_2^0|W_2^3, Q_3 = 0) \\
&= E(Y_3^1 - Y_2^0|W_2^3, Q_3 = 1) - E(Y_3^0 - Y_2^0|W_2^3, Q_3 = 1) \\
&\quad + E(Y_3^0 - Y_2^0|W_2^3, Q_3 = 1) - E(Y_3^0 - Y_2^0|W_2^3, Q_3 = 0)
\end{aligned}$$

subtracting and adding the 'counter-factual' $E(Y_3^0 - Y_2^0|W_2^3, Q_3 = 1)$ in the middle—counter-factual, as $Y_3^0$ is used despite $Q_3 = 1$.

The last two terms drop out under the '*conventional same time-effect condition*'

$$\Delta Y_3^0 \amalg Q_3|W_2^3 \implies E(\Delta Y_3^0|W_2^3, Q_3 = 1) - E(\Delta Y_3^0|W_2^3, Q_3 = 0) = 0. \tag{2.2}$$

Then the conventional DD becomes the $W_2^3$-conditional mean effect on the treated ($Q_3 = 1$) at the post-treatment period $t = 3$:

$$E(Y_3^1 - Y_2^0|W_2^3, Q_3 = 1) - E(Y_3^0 - Y_2^0|W_2^3, Q_3 = 1) = E(Y_3^1 - Y_3^0|W_2^3, Q_3 = 1). \tag{2.3}$$

This kind of DD identification was discussed by Lee and Kang (2006) for different types of data such as panels and repeated cross-sections. We may use only $W_3$ in the conditioning set, but $W_2^3$ makes the same time-effect condition more plausible as $Y_2$ appears in $\Delta Y_3$. Integrating out $W_2^3$ yields a marginal mean effect on the treated ($Q_3 = 1$).

Instead of the conventional DD, we propose to use the '*stayer DD*':

$$E(\Delta Y_3|W_2^3, Q_2 = 1, Q_3 = 1) - E(\Delta Y_3|W_2^3, Q_2 = 0, Q_3 = 0) \tag{2.4}$$
$$= E(Y_3^1 - Y_2^0|W_2^3, Q_2 = 1, Q_3 = 1) - E(Y_3^0 - Y_2^0|W_2^3, Q_2 = 0, Q_3 = 0)$$
$$= E(Y_3^1 - Y_2^0|W_2^3, Q_2 = 1, Q_3 = 1) - E(Y_3^0 - Y_2^0|W_2^3, Q_2 = 1, Q_3 = 1)$$
$$+E(\Delta Y_3^0|W_2^3, Q_2 = 1, Q_3 = 1) - E(\Delta Y_3^0|W_2^3, Q_2 = 0, Q_3 = 0).$$

*Under the stayer, not the conventional, same time-effect condition*

$$E(\Delta Y_3^0|W_2^3, Q_2 = 1, Q_3 = 1) - E(\Delta Y_3^0|W_2^3, Q_2 = 0, Q_3 = 0), \tag{2.5}$$

*the stayer DD becomes the effect on in-stayers* $E(Y_3^1 - Y_3^0|W_2^3, Q_2 = 1, Q_3 = 1)$ *in (2.1).* Bear in mind the distinction between the stayer DD in (2.4) and the effect on in-stayers in (2.1): *the stayer DD is defined with the observed* $\Delta Y_3$ *and two stayer groups* $(Q_2 = 1, Q_3 = 1)$ *and* $(Q_2 = 0, Q_3 = 0)$*, whereas the effect on in-stayers is defined with the counter-factual* $Y_3^1 - Y_3^0$ *and the single in-stayer group* $(Q_2 = 1, Q_3 = 1)$*.*

Opposite to the stayer DD would be the 'mover DD':

$$E(\Delta Y_3|W_2^3, Q_2 = 0, Q_3 = 1) - E(\Delta Y_3|W_2^3, Q_2 = 1, Q_3 = 0)$$

where the treatment group is the in-movers and the control group is the out-movers; the former is treated at $t = 3$ whereas the latter is not. The mover DD and "mixed DD" are examined further in the appendix.

## 2.3 DD Implementation with Complete Pairing (CP)

The conventional and stayer DD's can be nonparametrically implemented using the usual matching method, but we will use a 'complete pairing (CP)' estimator in Lee (2009,2012) that is simpler. The CP estimator is matching of a sort, yet it holds a number of advantages over the usual matching as will be explained below.

Let the subscript 0 indicate the control group, and 1 the treatment group: $Q_3$ and $Q_2^3$ determine the group membership for the conventional DD and the stayer DD, respectively. For subject $i$ in the control group at time $t$, let $Y_{0,it}$ denote the response, and $M_{0,it}$ covariates of dimension $k \times 1$ to be controlled; analogously, $Y_{1,jt}$ and $M_{1,jt}$ are the response and covariates for subject $j$ in the treatment group at time $t$. $N_0$ and $N_1$ are the control and treatment group sizes with $N \equiv N_0 + N_1$; $N$ could be smaller than the original sample size $\tilde{N}$. Let $K$

be a kernel, e.g., the $k$-fold product of the $N(0,1)$ densities, and $h$ a bandwidth with $h \to 0^+$ as $N \to \infty$.

A CP kernel estimator for DD with continuous covariates $M$ is

$$\widehat{DD}_{23} \equiv \frac{1}{N_0 N_1}\sum_{i=1}^{N_0}\sum_{j=1}^{N_1}\frac{1}{h^k}K(\frac{M_{1,j3}-M_{0,i3}}{h})(\Delta Y_{1,j3}-\Delta Y_{0,i3})/\frac{1}{N_0 N_1}\sum_{i=1}^{N_0}\sum_{j=1}^{N_1}\frac{1}{h^k}K(\frac{M_{1,j3}-M_{0,i3}}{h}).$$

Cases of $M_3$ discrete or mixed (continuous/discrete) can be seen in Lee (2009,2012), who also presented the requisite assumptions and proofs for CP estimators. $\widehat{DD}_{23}$ shows from where the name CP comes: all available pairs are used. The appearance of $\Delta Y_{1,j3} - \Delta Y_{0,i3}$ makes it clear that $\widehat{DD}_{23}$ is indeed a DD estimator.

Suppose

$$0 < \lambda_\nu \equiv \lim_{N\to\infty} N_\nu/N < 1.$$

With '$\rightsquigarrow$' denoting convergence in law, it holds that, as $N \to \infty$,

$$\widehat{DD}_{23} \to^p \mu_{23} \equiv \int \{E(\Delta Y_{1,3}|M_{1,3}=m) - E(\Delta Y_{0,3}|M_{0,3}=m)\}\omega_3(m)dm,$$
$$\text{where} \quad \omega_3(m) \equiv \pi_{1,3}(m)\pi_{0,3}(m)/\int \pi_{1,3}(m)\pi_{0,3}(m)dm$$

and $\pi_{\nu,3}(m)$ is the $M_{\nu,3}$ density function, $\nu = 0,1$; $\mu_{23}$ is the marginal effect with the conditioning variable $M_{0,3}$ and $M_{1,3}$ integrated out by the *density-product-form weight* $\omega_3$.

As for the asymptotic distribution,

$$\sqrt{N}(\widehat{DD}_{23} - \mu_{23}) \rightsquigarrow N[0,\ (\lambda_0^{-1}\phi_0 + \lambda_1^{-1}\phi_1)/\{\int \pi_{1,3}(m)\pi_{0,3}(m)dm\}^2]$$

where $\phi_0$ and $\phi_1$ are consistently estimable with

$$\hat{\phi}_{N0} \equiv \frac{1}{N_0}\sum_{i=1}^{N_0}\{\frac{1}{N_1}\sum_{j=1}^{N_1}\frac{1}{h^k}K(\frac{M_{1,j3}-M_{0,i3}}{h})(\Delta Y_{1,j3}-\Delta Y_{0,i3})\}^2$$
$$\hat{\phi}_{N1} \equiv \frac{1}{N_1}\sum_{j=1}^{N_1}\{\frac{1}{N_0}\sum_{i=1}^{N_0}\frac{1}{h^k}K(\frac{M_{1,j3}-M_{0,i3}}{h})(\Delta Y_{1,j3}-\Delta Y_{0,i3})\}^2$$

and $\int \pi_{1,3}(m)\pi_{0,3}(m)dm$ is consistently estimable with the denominator of $\widehat{DD}_{23}$.

$\widehat{DD}_{23}$ has a number of advantages over the usual matching. First, using all available pairs obviates the need to pick the number of the matched subjects in the usual pair/multiple matching. Second, whereas the asymptotic distribution of matching estimators is difficult to derive and use (see Abadie and Imbens 2011,2012 and the references therein), that of CP is not. Third, CP performs better than the usual matching when the supports of $M_{0,3}$ and $M_{1,3}$

do not overlap well, because $\widehat{DD}_{23}$ has a built-in mechanism to take into account the problem: the product weight $\pi_{1,3}(m)\pi_{0,3}(m)$ is non-zero only when $\pi_{1,3}(m) \neq 0$ *and* $\pi_{0,3}(m) \neq 0$, and thus the support points of $M_{0,3}$ and $M_{1,3}$ in the non-overlapping region contribute neither to the effect estimate nor to the asymptotic variance. In contrast, the usual matching tackles the problem by discarding individuals whose closest match is farther away than a chosen 'caliper' value.

Being a nonparametric estimator, $\widehat{DD}_{23}$ has two practical problems shared by the usual matching. One is the dimension problem: if $k$ is high, $\widehat{DD}_{23}$ is unlikely to work well. We will avoid this problem by replacing $M_3$ in $\widehat{DD}_{23}$ with an estimated propensity score (Rosenbaum and Rubin 1983) for $P(Q_3 = 1|M_3)$ in the conventional DD or for $P(Q_2^3 = 1|M_2^3, Q_2 = Q_3)$ in the stayer DD; probit or its latent linear regression function will be used in our empirical analysis. The other problem is choosing $h$, for which Lee (2012) suggested the following two ad-hod schemes.

The first scheme is using a "rule-of-thumb" $h = p \cdot N^{-1/(k+4)}$ with the covariates (or the propensity score) standardized and $p$ ranging over, say $[1, 3]$; set $k = 1$ when a propensity score or its linear regression function is used. The second is applying cross-validation in predicting $\Delta Y_{1,3} - \Delta Y_{0,3}$ using the first $N_c \equiv \min(N_0, N_1)$ observations with a kernel regression of $\Delta Y_{1,3} - \Delta Y_{0,3}$ on $(M_{1,3}, M_{0,3})$. Here, the two groups are pooled, and the first $N_c$ observations in each group are handled as if they came from the same $N_c$ individuals. Lee (2009) in fact proposed a better scheme to select $h$, although it is omitted here as the scheme is more involved.

# 3 Panel Linear Models

As a starter, consider a simple 'intercept-shifting-treatment' panel linear model

$$Y_{it}^0 = \beta_t + \beta_q Q_{it} + \beta_w' W_{it} + V_{it} \quad \text{and} \quad Y_{it}^1 = \beta_d + Y_{it}^0$$

where $\beta$'s are parameters with $\beta_w \equiv (\beta_c', \beta_x')'$, and $V_{it} = \delta_i + U_{it}$ is an error term with a time-constant error $\delta_i$ and a time-varying error $U_{it}$. Since $Y_{i3}^1 - Y_{i3}^0 = \beta_d$, the four group effects in (2.1) are all $\beta_d$. Recalling $D_{it} = Q_{it}1[t = 3]$, we have

$$Y_{it} = \beta_t + \beta_d 1[t = 3]Q_{it} + \beta_q Q_{it} + \beta_w' W_{it} + V_{it}. \qquad (\mathrm{M}_0)$$

First-differencing $\text{M}_0$ with $t = 3$ yields, as $\beta_c' \Delta C_i = 0$ in $\beta_w' \Delta W_{i3}$,

$$\Delta Y_{i3} = \Delta\beta_3 + \beta_d Q_{i3} + \beta_q \Delta Q_{i3} + \beta_x' \Delta X_{i3} + \Delta U_{i3}. \tag{M$_0'$}$$

Estimating this with LSE or instrumental variable estimator would be the conventional DD.

To allow the in-mover and in-stayer difference in the treatment effect, keep the above $Y_{it}^0$, but replace the $Y_{it}^1$ equation with

$$Y_{it}^1 = \beta_d Q_{i,t-1} + \beta_m (1 - Q_{i,t-1}) + Y_{it}^0 :$$

with the treatment at $t$, the intercept shifts by $\beta_d$ if $Q_{i,t-1} = 1$ or by $\beta_m$ if $Q_{i,t-1} = 0$; $\beta_d = \beta_m$ takes us back to the above simple intercept-shift model. Since $Y_{i3}^1 - Y_{i3}^0 = \beta_d Q_{i2} + \beta_m (1 - Q_{i2})$, the four group effects in (2.1) are

$$\begin{aligned}
\text{effect on out-stayers} &: E(Y_3^1 - Y_3^0 | W_2^3, Q_2 = 0, Q_3 = 0) = \beta_m; \\
\text{effect on in-movers} &: E(Y_3^1 - Y_3^0 | W_2^3, Q_2 = 0, Q_3 = 1) = \beta_m; \\
\text{effect on out-movers} &: E(Y_3^1 - Y_3^0 | W_2^3, Q_2 = 1, Q_3 = 0) = \beta_d; \\
\text{effect on in-stayers} &: E(Y_3^1 - Y_3^0 | W_2^3, Q_2 = 1, Q_3 = 1) = \beta_d;
\end{aligned} \tag{3.1}$$

the effect on out-stayers and the effect on out-movers are counter-factuals as they cannot be treated at $t = 3$.

From $Y_{it} = (1 - D_{it}) Y_{it}^0 + D_{it} Y_{it}^1$,

$$\begin{aligned}
Y_{it} &= \beta_t + \{\beta_d Q_{i,t-1} + \beta_m (1 - Q_{i,t-1})\} D_{it} + \beta_q Q_{it} + \beta_w' W_{it} + V_{it} \\
&= \beta_t + \beta_d 1[t = 3] Q_{i,t-1} Q_{it} + \beta_m 1[t = 3](1 - Q_{i,t-1}) Q_{it} + \beta_q Q_{it} + \beta_w' W_{it} + V_{it}.
\end{aligned} \tag{M$_1$}$$

First-differencing $\text{M}_1$ with $t = 3$ yields

$$\Delta Y_{i3} = \Delta\beta_3 + \beta_d Q_{i2} Q_{i3} + \beta_m (1 - Q_{i2}) Q_{i3} + \beta_q \Delta Q_{i3} + \beta_x' \Delta X_{i3} + \Delta U_{i3}. \tag{M$_1'$}$$

LSE can be applied to this if $\Delta U_3$ is uncorrelated with the regressors; otherwise, instrumental variable estimator may be used with some instruments. Whereas $\text{M}_0'$ is not 'saturated' with respect to $(Q_2, Q_3)$, $\text{M}_1'$ is so with four parameters $(\Delta\beta_3, \beta_d, \beta_m, \beta_q)$ for the four groups determined by $(Q_2, Q_3)$. *Estimating $M_1'$, we can find the effect on in-stayers $\beta_d$ and the effect on in-movers $\beta_m$, along with the untreated moving effect $\beta_q$; this is our mover-stayer DD.*

It is interesting to see what the nonparametric DD leads to when $M_1$ is the data generating process (DGP), for which we ignore the part $\beta_x' \Delta X_{i3} + \Delta U_{i3}$ for a while. The four group means are, using $M_1$'

$$\begin{aligned} \text{mean for out-stayers} &: E(\Delta Y_3|Q_2=0,Q_3=0) = \Delta\beta_3; \\ \text{mean for in-movers} &: E(\Delta Y_3|Q_2=0,Q_3=1) = \Delta\beta_3 + \beta_m + \beta_q; \\ \text{mean for out-movers} &: E(\Delta Y_3|Q_2=1,Q_3=0) = \Delta\beta_3 - \beta_q; \\ \text{mean for in-stayers} &: E(\Delta Y_3|Q_2=1,Q_3=1) = \Delta\beta_3 + \beta_d. \end{aligned} \quad (3.2)$$

From this, we get four nonparametric DD's:

$$\begin{aligned} \text{stayer DD} &: E(\Delta Y_3|Q_2=1,Q_3=1) - E(\Delta Y_3|Q_2=0,Q_3=0) = \beta_d; \quad (3.3) \\ \text{mover DD} &: E(\Delta Y_3|Q_2=0,Q_3=1) - E(\Delta Y_3|Q_2=1,Q_3=0) = \beta_m + 2\beta_q; \\ \text{in-stayer out-mover DD} &: E(\Delta Y_3|Q_2=1,Q_3=1) - E(\Delta Y_3|Q_2=1,Q_3=0) = \beta_d + \beta_q; \\ \text{in-mover out-stayer DD} &: E(\Delta Y_3|Q_2=0,Q_3=1) - E(\Delta Y_3|Q_2=0,Q_3=0) = \beta_m + \beta_q. \end{aligned}$$

Notice $2\beta_q$ in the mover DD, as the in-movers gain the untreated moving effect $\beta_q$ while the out-movers lose $\beta_q$.

In the four nonparametric DD's, only the stayer DD identifies the genuine treatment effect $\beta_d$ (the effect on in-stayers) whereas the other three identify a genuine effect ($\beta_d$ or $\beta_m$) contaminated by the untreated moving effect $\beta_q$. That is, when $M_1$ is the DGP, only the stayer DD is consistent for a treatment effect. This problem does not disappear even when $\beta_d = \beta_m$, under which there are, however, ways other than the stayer DD to find $\beta_d$; e.g., add up the last two DD's to get $\beta_d + \beta_m + 2\beta_q$ and subtract the mover DD from this to obtain $\beta_d$.

In contrast to the four DD's in (3.3), *the conventional DD is (recall $M_1'$ without $\beta_x'\Delta X_{i3} + \Delta U_{i3}$)*

$$\begin{aligned} & E(\Delta Y_3|Q_3=1) - E(\Delta Y_3|Q_3=0) \\ &= E\{\Delta\beta_3 + \beta_d Q_2 + \beta_m(1-Q_2) + \beta_q(1-Q_2)\} - E\{\Delta\beta_3 + \beta_q(-Q_2)\} \\ &= \beta_d E(Q_2) + (\beta_m + \beta_q)\{1 - E(Q_2)\} + \beta_q E(Q_2) \\ &= \beta_d E(Q_2) + \beta_m\{1 - E(Q_2)\} + \beta_q \end{aligned} \quad (3.4)$$

*that is 'the mixture of the effect on in-stayers $\beta_d$, the effect on in-movers $\beta_m$ and the untreated moving effect $\beta_q$.* If $\beta_d = \beta_m$ holds (i.e., $M_0$ holds), then the conventional DD becomes

*$\beta_d + \beta_q$; the conventional nonparametric DD does not work even when $\beta_d = \beta_m$ so long as $Q_t$ is time-varying.*

## 4 Further Remarks

This section makes a few remarks for the nonparametric and panel linear model DD's under $M_0$, not $M_1$, for simplification. Firstly, the base panel linear model $M_0$ is generalized by allowing $Y_{i,t-1}$ as a regressor. Secondly, time-varying parameters are considered. Thirdly, what the stayer-DD nonparametric same time-effect condition amounts to for $M_0$ is discussed. Fourthly, the so-called 'Ashenfelter (1978) dip' problem is shown to disappear in the stayer DD.

**Lagged Response as a Regressor**

Suppose the lagged response $Y_{i,t-1}$ is used as a regressor in $M_0$. Then we get

$$Y_{it} = \beta_t + \beta_y Y_{i,t-1} + \beta_d 1[t=3]Q_{it} + \beta_q Q_{it} + \beta_w' W_{it} + V_{it}$$
$$\implies \Delta Y_{i3} = \Delta\beta_3 + \beta_y \Delta Y_{i2} + \beta_d Q_{i3} + \beta_q \Delta Q_{i3} + \beta_x' \Delta X_{i3} + \Delta U_{i3}.$$

This results in a couple of changes. First, LSE would not work, because the regressor $\Delta Y_2 = Y_2 - Y_1$ is related to $\Delta U_3 = \Delta V_3 = V_3 - V_2$ through $V_2$. Second, at least three waves $(1,2,3)$ are needed so that at least $Y_1$ can be used as an instrument for $\Delta Y_2$; if four waves $(0,1,2,3)$ are available, then $(Y_0, Y_1)$ may be used as instruments for $\Delta Y_2$. Third, whereas the panel linear model can cope with the endogeneity of $\Delta Y_2$ with relative ease, it is not clear how the nonparametric estimation of DD can be done in this case.

**Time-Varying Parameters**

Suppose, not just the intercept $\beta_t$, but also all slopes are time-varying in $M_0$ to yield

$$Y_{it} = \beta_t + \beta_{dt} 1[t=3]Q_{it} + \beta_{qt} Q_{it} + \beta_{wt}' W_{it} + V_{it}$$
$$\implies \Delta Y_{i3} = \Delta\beta_3 + \beta_{d3} Q_{i3} + \beta_{q3} Q_{i3} - \beta_{q2} Q_{i2} + \beta_{w3}' W_{i3} - \beta_{w2}' W_{i2} + \Delta U_{i3}$$
$$= \Delta\beta_3 + (\beta_{d3} + \beta_{q3}) Q_{i3} - \beta_{q2} Q_{i2} + \beta_{w3}' W_{i3} - \beta_{w2}' W_{i2} + \Delta U_{i3}.$$

The genuine treatment effect is $\beta_{d3}$ which can be simply redefined as $\beta_d$, but $\beta_{d3}$ is not identified as $\beta_{q3}$ is added to $\beta_{d3}$. That is, if the qualification group effect (or the untreated moving effect) $\beta_{qt}$ is time-varying, then the genuine treatment effect is not identified, although the time-varying feature in the other parameters ($\beta_t$, $\beta_{dt}$ and $\beta_{wt}$) does not pose any problem.

**Same Time-Effect Condition**

Under $M_0'$, the $W_2^3$-conditional four group means are (recall (3.2) with $\beta_d = \beta_m$ and $\Delta X_3$ and $\Delta U_3$ added back in):

$$\begin{aligned}
E(\Delta Y_3|W_2^3, Q_2 &= 0, Q_3 = 0) = \Delta\beta_3 + \beta_x'\Delta X_3 + E(\Delta U_3|W_2^3, Q_2 = 0, Q_3 = 0);\\
E(\Delta Y_3|W_2^3, Q_2 &= 0, Q_3 = 1) = \Delta\beta_3 + \beta_d + \beta_q + \beta_x'\Delta X_3 + E(\Delta U_3|W_2^3, Q_2 = 0, Q_3 = 1);\\
E(\Delta Y_3|W_2^3, Q_2 &= 1, Q_3 = 0) = \Delta\beta_3 - \beta_q + \beta_x'\Delta X_3 + E(\Delta U_3|W_2^3, Q_2 = 1, Q_3 = 0);\\
E(\Delta Y_3|W_2^3, Q_2 &= 1, Q_3 = 1) = \Delta\beta_3 + \beta_d + \beta_x'\Delta X_3 + E(\Delta U_3|W_2^3, Q_2 = 1, Q_3 = 1). \quad (4.1)
\end{aligned}$$

The same time-effect condition for the stayer DD is thus

$$\begin{aligned}
&E(\Delta Y_3^0|W_2^3, Q_2 = 0, Q_3 = 0) = E(\Delta Y_3^0|W_2^3, Q_2 = 1, Q_3 = 1)\\
&\iff \Delta\beta_3 + \beta_x'\Delta X_3 + E(\Delta U_3|W_2^3, Q_2 = 0, Q_3 = 0)\\
&\quad = \Delta\beta_3 + \beta_x'\Delta X_3 + E(\Delta U_3|W_2^3, Q_2 = 1, Q_3 = 1)
\end{aligned}$$

which is implies by $\Delta U_3 \amalg Q_2^3|W_2^3$ that rules out confounding due to $\Delta U_3$. Under this, the stayer DD becomes $\beta_d$ because, using (4.1),

$$\begin{aligned}
&E(\Delta Y_3|W_2^3, Q_2 = 1, Q_3 = 1) - E(\Delta Y_3|W_2^3, Q_2 = 0, Q_3 = 0)\\
&= \beta_d + E(\Delta U_3|W_2^3, Q_2 = 1, Q_3 = 1) - E(\Delta U_3|W_2^3, Q_2 = 0, Q_3 = 0) = \beta_d.
\end{aligned}$$

**Ashenfelter Dip Problem**

The so-called 'Ashenfelter (1978) dip' for job trainings is that the treatment group experience a dip (i.e., a low $Y_2$ in earnings) just before getting treated. Since the 'dip' is transitory by definition, the treatment group is bound to have a higher $Y_3$ even without the treatment—an untreated moving effect. Considering the effect on in-stayers takes care of the Ashenfelter dip problem as follows.

Suppose that a training is administered to the unemployed so that $Q_t = 1[Y_{t-1} \leq 0]$. There are two types in the unemployed: the persistently unemployed ($Q_2 = 1[Y_1 \leq 0] = 1$ and $Q_3 = 1[Y_2 \leq 0] = 1$), and the temporarily unemployed ($Q_2 = 1[Y_1 \leq 0] = 0$ and $Q_3 = 1[Y_2 \leq 0] = 1$). Differently from the conventional DD, stayer DD would not be "fooled" by the Ashenfelter dip problem, as the movers are either not used at all (in the nonparametric stayer DD), or the three effects (the effect on in-stayers $\beta_d$, the effect on in-movers $\beta_m$ and the untreated moving effect $\beta_q$) are separated from one another (in the panel linear model DD). Separating stayers and movers is different from trying various control periods ($t = 2$, $t = 1$ or $t = 0$) as was done in Ashenfelter (1978).

## 5 Empirical Example

In January 2008, South Korea started the 'Basic Elder Pension (BEP)' for persons of age 65 or higher. BEP is to provide a minimal support to the elderly. The eligibility condition other than age is based on income and assets: assets are presumed to make 5% interest per annum, and if the 'total income' (the actual income plus the derived income from the assets) was below about \$740 per month for a single person or \$1184 for a married couple in 2008, then the person was eligible for BEP in 2008. The BEP amount varied across the recipients; the monthly average was \$84 for a single and \$134 for a couple in 2008. We will find the effects of BEP on monthly health expenditure, using household panel data collected by the Korea Institute for Health and Social Affairs for 2007 ($t = 2$; just before BEP) and 2008 ($t = 3$; just after BEP).

One problem with the panel data is that $Q_2$ is not observed although $Q_3$ is, because the eligibility threshold for the total income varied year to year since 2008, and the threshold did not exist in 2007. Another problem is that, although income and assets variables are available, the actual $Q_3$ is based on an 'adjusted total income' depending on various factors. For instance, labor income is subtracted from the total income, and the real-estate values are adjusted for BEP purpose depending on where one lives. We tried to duplicate $Q_3$ using the 2008 threshold and 2008 income and asset variables in the data, but the proportion of $Q_3$ being the same as the duplicate was only about 50%, as our data set is not detailed enough to give the adjusted total income for BEP. Fortunately, there is a variable $L_t$ in the data: $L_t = 1$ means being below 0.6 times the median of 'the household income divided by (the number of household members)$^{1/2}$'. For 2008, 75% of the elderly of age 65 or above in the data have $L_3 = Q_3$. Hence we used $L_2$ as $Q_2$ for 2007. This results in a measurement error problem for $Q_2$, but addressing this issue for DD is beyond the scope of this paper.

With $N = 2201$ with the average age 73 in 2007, the proportions of the stayers and movers based on $Q_2$ and $Q_3$ are

$$\text{out-stayers } (Q_2 = 0, Q_3 = 0) : 0.182; \qquad \text{in-movers } (Q_2 = 0, Q_3 = 1) : 0.087;$$
$$\text{out-movers } (Q_2 = 1, Q_3 = 0) : 0.169; \qquad \text{in-stayers } (Q_2 = 1, Q_3 = 1) : 0.563.$$

As in most panel data, maintaining the current status is more likely than changing it; thus the proportion of the stayers $0.745 = 0.182 + 0.563$ is about three times the proportion of the movers 0.255—one more reason to use the stayers than the movers.

For the covariates, we use monthly disposable income excluding BEP, whether self-assessed health status is good or not ('good health'), having a chronic disease or not of 6 months or longer ('chronic disease'), having health problems or not ('health problems'), unemployed or not ('unemployed'), owning a house/condominium or not ('own home'), the number of household members ('# members'), and spouse age $\geq 65$ or not ('spouse age65'; 0 if no spouse). There are (almost) time-constant variables available such as schooling, the gender of household head and living in Seoul or not, but they are not shown in Table 1 below; they drop out in the first-differenced panel linear models, and they are not used either in CP.

Table 1: Avg and SD (Min, Max) of Variables ($N = 2201$)

| | 2007 | 2008 |
|---|---|---|
| monthly health expense ($10) | 13.0, 22.2 (0, 420) | 14.9, 24.0 (0, 262) |
| monthly income ($10) | 118, 108 (0, 1277) | 122, 124 (0, 1156) |
| good health | 0.515, 0.500 (0, 1) | 0.505, 0.500 (0, 1) |
| chronic disease | 0.797, 0.402 (0, 1) | 0.799, 0.401 (0, 1) |
| health problems | 0.433, 0.496 (0, 1) | 0.478, 0.500 (0, 1) |
| unemployed | 0.622, 0.485 (0, 1) | 0.618, 0.486 (0, 1) |
| own home | 0.606, 0.489 (0, 1) | 0.596, 0.491 (0, 1) |
| # members | 1.833, 0.922 (1, 8) | 1.851, 0.987 (1, 8) |
| spouse age65 | 0.379, 0.485 (0, 1) | 0.387, 0.487 (0, 1) |

Table 1 shows the descriptive statistics for the $Y$ and $X$ variables; expenditure and income variables were divided by the price indices to obtain the real values. Although $\ln Y$ (subject to a transformation when $Y = 0$) is to be used for estimation below, Table 1 shows $Y$ in 1000 Korean Won ($\simeq$\$10); the same thing can be said for the monthly income. Although BEP is fairly small (about $100 per month), it is not so small compared with the average monthly health expenditure and income.

Table 2 presents the LSE to the first-difference model $M_0'$ with only $Q_3$ and $\Delta Q_3$ along with $\Delta X_3$ as the regressors; this is the conventional DD with only $Q_3$ capturing the treatment effect. In each column, the estimates are shown with t-values in $(\cdot)$. In the first column, logarithm is taken on $Y$ and income if they are greater than one; otherwise, 1 is added before logarithm is taken. In the second column, one is added to both $Y$ and income before logarithm is taken. In the third column, only the observations with positive $Y$ and income are used; $N = 2046$ with 7% of the observations lost.

In all three columns of Table 2, BEP has a significantly positive effect of 17-19% on health expenditure, and the untreated moving effect is -15 to -18% (significant): those who become newly eligible reduces their health expenditure by 15-18%, which is plausible as they become poorer. The regressors have similar slopes across the three columns, except ln(income), unemployed and # members. Good health has a positive effect, which might be due to the reverse causality that would make good health endogenous. The unemployed spends more, which could happen if they have less opportunity cost of using health care.

Table 2: LSE (tv) for Conventional Panel DD Model

| | $\ln(\cdot + 1[\cdot < 1])$ | $\ln(\cdot + 1)$ | + only ($N = 2046$) |
|---|---|---|---|
| 1 | -0.039 (-0.68) | -0.035 (-0.69) | -0.029 (-0.49) |
| $Q_3$ | 0.187 (2.76) | 0.165 (2.71) | 0.182 (2.44) |
| $\Delta Q_3$ | -0.179 (-2.69) | -0.165 (-2.77) | -0.150 (-2.11) |
| ln(income) | 0.395 (7.29) | 0.363 (7.17) | 0.525 (7.99) |
| good health | 0.219 (4.95) | 0.207 (5.28) | 0.239 (4.69) |
| chronic disease | 0.218 (3.23) | 0.192 (3.21) | 0.203 (2.66) |
| health problems | 0.134 (3.19) | 0.116 (3.10) | 0.144 (3.05) |
| unemployed | 0.210 (2.70) | 0.185 (2.65) | 0.132 (1.49) |
| own home | -0.111 (-1.42) | -0.080 (-1.13) | -0.078 (-0.85) |
| # members | 0.114 (2.07) | 0.098 (1.99) | 0.056 (0.93) |
| spouse age65 | 0.560 (4.50) | 0.520 (4.58) | 0.525 (3.92) |

Table 3 presents the LSE to the first-difference model $M_1'$ with $Q_2Q_3$, $(1 - Q_2)Q_3$ and $\Delta Q_3$ along with $\Delta X_3$ as the regressors; this is our mover-stayer DD with $Q_2Q_3$ and $(1 - Q_2)Q_3$ having different slopes to represent the effect on in-stayers and the effect on in-movers, respectively. The effect on in-stayers is significantly positive with magnitude a little smaller than the slope of $Q_3$ in Table 2, whereas the effect on in-movers is near zero; the untreated moving effect is insignificantly negative with magnitude somewhat smaller than that in Table 2. In words, when one stays poor ($Q_2Q_3 = 1$) but receives BEP, the health expenditure increases by about 16% (significant); when one becomes poorer ($(1-Q_2)Q_3 = 1$) but receives BEP, hardly no change in health expenditure; when one becomes poorer ($\Delta Q_3 = 1$) without receiving BEP, the health expenditure decreases by about 10% (insignificant).

Clearly, the effect on in-stayers differs from the effect on in-movers in Table 3, and thus the LSE for Table 2 is misspecified with the false restriction $\beta_d = \beta_m$; the outcome of this misspecification is also apparent in the difference between the untreated moving effects in Tables 2 and 3. The estimates for the other regressors differ little between Tables 2 and 3.

Table 3: LSE (tv) for New Panel DD Model

| | $\ln(\cdot + 1[y < 1])$ | $\ln(\cdot + 1)$ | + only ($N = 2046$) |
|---|---|---|---|
| 1 | -0.010 (-0.16) | -0.006 (-0.11) | 0.001 (0.01) |
| $Q_2Q_3$ | 0.168 (2.34) | 0.146 (2.25) | 0.162 (2.10) |
| $(1 - Q_2)Q_3$ | 0.041 (0.24) | 0.016 (0.10) | 0.025 (0.14) |
| $\Delta Q_3$ | -0.120 (-1.32) | -0.104 (-1.29) | -0.088 (-0.91) |
| ln(income) | 0.395 (7.28) | 0.364 (7.15) | 0.525 (7.99) |
| good health | 0.219 (4.94) | 0.207 (5.27) | 0.239 (4.69) |
| chronic disease | 0.219 (3.26) | 0.193 (3.23) | 0.204 (2.69) |
| health problems | 0.136 (3.22) | 0.118 (3.13) | 0.145 (3.08) |
| unemployed | 0.210 (2.72) | 0.186 (2.66) | 0.133 (1.50) |
| own home | -0.110 (-1.41) | -0.079 (-1.12) | -0.078 (-0.85) |
| # members | 0.114 (2.07) | 0.098 (1.99) | 0.056 (0.93) |
| spouse age65 | 0.561 (4.51) | 0.520 (4.58) | 0.525 (3.92) |

One concern in the above DD's with panel linear models is that some part of the model might be misspecified. To dissipate this concern, we apply the CP approach in Table 4. Table 4 shows two CP's: the conventional DD in the first three rows uses the propensity score for $Q_3$ on $\Delta X_3$ ignoring the untreated moving effect, and the stayer DD in the last three rows uses the propensity score for $Q_2^3 = 1$ on $\Delta X_3$ given $Q_2 = Q_3$. In each CP, the propensity score using probit, say $\Phi(\alpha_x' \Delta X)$, is obtained, and then the probit regression function $\alpha_x' \Delta X$ is used as one-dimensional covariate to control. In each CP, the middle row uses the cross-validation scheme mentioned in the main text to select $p$ in $h = pN^{-1/5}$. To demonstrate that the CP estimates and t-values are not sensitive to $h$, the remaining two rows in each CP use a much smaller bandwidth 0.1 and a much larger bandwidth 4.0.

Table 4 shows that the conventional DD gives insignificant effect (about 6-7% in the middle row), whereas the stayer DD gives a significantly positive effect on in-stayers (about 15-19% in the middle row)—recall that the stayer DD identifies the effect on in-stayers. The conventional DD yields a mixed effect for the in-stayers and in-movers, which is why the conventional DD effect is much smaller and insignificant. The effect on in-stayers shown by the stayer DD is remarkably similar to that (15-17%) in Table 3 given by the slope of $Q_2Q_3$.

Table 4: $p$ in $h = pN^{-1/5}$, Treatment Effect (tv) for Complete Pairing

| | $\ln(\cdot + 1[\cdot < 1])$ | $\ln(\cdot + 1)$ | + only |
|---|---|---|---|
| Conventional DD | 0.1: 0.097 (1.65) | 0.1: 0.077 (1.47) | 0.1: 0.103 (1.58) |
| | *0.5: 0.073 (1.29)* | *0.5: 0.056 (1.11)* | *1.2: 0.060 (0.97)* |
| | 4.0: 0.016 (0.28) | 4.0: 0.007 (0.14) | 4.0: 0.019 (0.31) |
| Stayer DD | 0.1: 0.223 (2.67) | 0.1: 0.184 (2.47) | 0.1: 0.248 (2.68) |
| | *0.7: 0.173 (2.36)* | *0.6: 0.146 (2.21)* | *0.6: 0.185 (2.29)* |
| | 4.0: 0.169 (2.38) | 4.0: 0.146 (2.27) | 4.0: 0.169 (2.21) |

## 6 Conclusions

This paper examined difference in differences (DD) with a time-varying qualification $Q_t$. Differently from a time-constant qualification, when $Q_t$ is time-varying, taking $Q_t = 1$ as the treatment group and $Q_t = 0$ as the control group is misleading as there are movers ($Q_{t-1} \neq Q_t$): if $Q_{t-1} \neq Q_t$ affects the response variable regardless of the treatment, then there occurs an 'untreated moving effect'. Also, when there is a genuine treatment effect, the effect on the 'in-movers' ($Q_{t-1} = 0, Q_t = 1$) may be different from the effect on the 'in-stayers' ($Q_{t-1} = 1, Q_t = 1$).

*Our main suggestion for nonparametric DD was to use only the in-stayers as the treatment group and the out-stayers as the control group; this nonparametric 'stayer DD' identifies the effect on in-stayers. For DD based on panel linear models, our suggestion was to use $Q_{t-1}Q_t$, $(1 - Q_{t-1})Q_t$ and $\Delta Q_t$ in the first differenced model, not just $Q_t$ and $\Delta Q_t$ as typically done in practice. This allows finding the effect on in-stayers (the slope of $Q_{t-1}Q_t$),* *the effect on in-movers (the slope of* $(1-Q_{t-1})Q_t$*) and the untreated moving effect (the slope of* $\Delta Q_t$*)*; using only $Q_t$ and $\Delta Q_t$ mixes up the effects on in-stayers and in-movers.

We presented an empirical analysis for the effects of the Basic Elder Pension (BEP) in Korea on health-care expenditure. Both the nonparametric estimator and the linear panel model estimator showed about *15-19% significant increase in health-care expenditure due to BEP for the effect on in-stayers, whereas the effect on in-movers and the untreated moving effect seem near zero or insignificant.* Apart from this example, in general, knowing that the effect on in-movers is zero (which is impossible in the conventional DD approach of using only $Q_t$ and $\Delta Q_t$ in the first-differenced linear model) is informative, as this would discourage individuals from changing their $Q_t$ from 0 to 1 to get newly treated.

# APPENDIX: Mover DD and Mixed DD's

The nonparametric mover DD is

$$\begin{aligned}
&E(\Delta Y_3|W_2^3, Q_2=0, Q_3=1) - E(\Delta Y_3|W_2^3, Q_2=1, Q_3=0)\\
&= E(Y_3^1 - Y_2^0|W_2^3, Q_2=0, Q_3=1) - E(Y_3^0 - Y_2^0|W_2^3, Q_2=1, Q_3=0)\\
&= E(Y_3^1 - Y_2^0|W_2^3, Q_2=0, Q_3=1) - E(Y_3^0 - Y_2^0|W_2^3, Q_2=0, Q_3=1)\\
&\quad + E(Y_3^0 - Y_2^0|W_2^3, Q_2=0, Q_3=1) - E(Y_3^0 - Y_2^0|W_2^3, Q_2=1, Q_3=0).
\end{aligned}$$

Under the 'mover same time-effect condition'

$$E(Y_3^0 - Y_2^0|W_2^3, Q_2=0, Q_3=1) - E(Y_3^0 - Y_2^0|W_2^3, Q_2=1, Q_3=0),$$

the mover DD becomes $E(Y_3^1 - Y_3^0|W_2^3, Q_2=0, Q_3=1)$ that is 'the $W_2^3$-conditional mean effect on the in-movers at the post-treatment period $t=3$'.

It is not clear which is more plausible between the stayer and mover same time-effect conditions. It is possible that the similarity between the in-movers and out-movers is greater than the similarity between the in-stayers and out-stayers. This possibility notwithstanding, as mentioned in the main text, the mover DD is not good because it may include an untreated moving effect.

Going further, instead of the stayer DD and mover DD, we may look at "mixed DD's"

$$\begin{aligned} E(\Delta Y_3 | W_2^3, Q_2 &= 1, Q_3 = 1) - E(\Delta Y_3 | W_2^3 = w, Q_2 = 1, Q_3 = 0), \\ E(\Delta Y_3 | W_2^3, Q_2 &= 0, Q_3 = 1) - E(\Delta Y_3 | W_2^3 = w, Q_2 = 0, Q_3 = 0) \end{aligned}$$

so long as the treatment group has $Q_3 = 1$ and the control group has $Q_3 = 0$.

# REFERENCES


Abadie, A. and G. Imbens, 2011, Bias-corrected matching estimators for average treatment effects, Journal of Business and Economic Statistics 29, 1-11.

Abadie, A. and G. Imbens, 2012, A martingale representation for matching estimators, Journal of the American Statistical Association 107, 833-843.

Angrist, J. and A. Krueger, 1999, Empirical strategies in labor economics, in the Handbook of Labor Economics, Vol.3A, edited by O. Ashenfelter and D. Card, New York, Elsevier.

Angrist, J. and J.S. Pischke, 2009, Mostly harmless econometrics, Princeton University Press.

Ashenfelter, O., 1978, Estimating the effect of training program on earnings. Review of Economics and Statistics 60, 47-57.

Athey, S. and G.W. Imbens, 2006, Identification and inference in nonlinear difference-in-differences models, Econometrica 74, 431-497.

Bertrand, M., E. Duflo and S. Mullainathan, 2004, How much should we trust differences-in-differences estimates, Quarterly Journal of Economics 119, 249-275.

Besley, T. and A. Case, 2004, Unnatural experiments? Estimating the incidence of endogenous policies, Economic Journal 110, F672-F694.

Heckman, J.J., R. LaLonde and J.A. Smith, 1999, The economics and econometrics of active labor market programs, in Handbook of Labor Economics III, edited by O. Ashenfelter and D. Card, North-Holland.

Imbens, G.W. and J.M. Wooldridge, 2009, Recent developments in the econometrics of program evaluation, Journal of Economic Literature 47, 5-86.

Kennan, J. and J.R. Walker, 2010, Wages, welfare benefits and migration, Journal of Econometrics 156, 229-238.

Lee, M.J., 2005, Micro-econometrics for policy, program, and treatment effects, Oxford University Press.

Lee, M.J., 2009, Nonparametric tests for distributional treatment effects for censored responses, Journal of the Royal Statistical Society (Series B) 71, 243-264.

Lee, M.J., 2012, Treatment effects in sample selection models and their nonparametric estimation, Journal of Econometrics 167, 317-329.

Lee, M.J. and C.H. Kang, 2006, Identification for difference in differences with cross-section and panel data, Economics Letters 92, 270-276.

Rosenbaum, P., 2002, Observational studies, 2nd ed., Springer.

Rosenbaum, P.R. and D.B. Rubin, 1983, The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 41-55.

Shadish, W.R., T.D. Cook and D.T. Campbell, 2002, Experimental and quasi-experimental designs for generalized causal inference, Houghton Mifflin Company.