

Sanghyeok Lee and Myoung-jae Lee*

Difference in Differences, Ratio in Ratios, and Ratio in Odds Ratios for Limited Dependent Variables: A Review and More

<https://doi.org/10.1515/snde-2024-0125>

Received November 13, 2024; accepted July 31, 2025; published online August 19, 2025

Abstract: Difference in differences (DD) is widely used to find policy/treatment effects with observational data, but applying DD to limited dependent variables (LDV's) Y has been problematic. This paper addresses how to apply DD and related approaches, such as “ratio in ratios” and “ratio in odds ratios”, to LDV's under the unifying framework of ‘generalized linear models with link functions’. We review and evaluate DD and the related approaches with simulation and empirical studies, and recommend ‘Poisson Quasi-MLE’ for non-negative (such as count or zero-censored) Y and (multinomial) logit MLE for binary, fractional or multinomial Y .

Keywords: difference in differences; limited dependent variable; ratio in odds ratios; ratio in ratios

JEL Classification: C31; C34; C35

1 Introduction

Difference in Differences (DD) is one of the most popular research designs in social sciences. Not just in social sciences, DD has been gaining popularity also in natural sciences; see, e.g. Jena et al. (2015), Cataife and Pagano (2017), and McGrath et al. (2019). There are various references for DD: Lee (2005, 2016a), Angrist and Pischke (2009), Blundell and Dias (2009), Lechner (2011), Lee and Kim (2014), Kim and Lee (2017), Lee and Sawada (2020), Ariella and Lang (2020), Roth et al. (2023) and De Chaisemartin and d'Haultfoeuille (2020, 2023).

DD is basically for linear models with additive components, which makes applying DD to limited dependent variables (LDV's) problematic. This paper provides answers to this problem, using the unifying idea of ‘generalized linear models (GLM) with link functions’. An earlier version (arXiv: 2111.12948) of this paper presented the GLM-based approach first in 2021, which has been cited by Callaway (2023), McMichael (2023), Markowitz and Smith (2024) and Freeman et al. (2025), among others. The approach subsequently appeared also in Taddeo et al. (2022) and Wooldridge (2023).

The GLM approach specifies the conditional mean regression, but two approaches requiring weaker assumptions appeared recently. One is Tchetgen et al. (2024) who specify an “extended propensity score function”, instead of the conditional mean regression. The other is Kim and Lee (2025), who use a “causal-reduced-form” and apply the usual DD to LDV's to find a non-negatively weighted average of heterogeneous effects. The goal of this paper is to explain the GLM-based approach in detail so that it can be readily applied by practitioners, and review the literature.

*Corresponding author: Myoung-jae Lee, Department of Economics, Korea University, Seoul 02841, South Korea; and Department of Finance, Accounting & Economics, University of Nottingham Ningbo China, Ningbo 315100, China, E-mail: myoungjae@korea.ac.kr
Sanghyeok Lee, School of Economics and Trade, Kyungpook National University, Daegu 41566, South Korea, E-mail: sanghyeok.lee@knu.ac.kr

 Open Access. © 2025 the author(s), published by De Gruyter.  This work is licensed under the Creative Commons Attribution 4.0 International License.

Consider an outcome Y_{it} for unit i at time $t = 2, 3$, a time-constant treatment qualification dummy Q_i , and a binary treatment D_{it} ; we set $t = 2, 3$ to avoid the confusion with dummy variable values 0, 1. The hallmark of DD is that D_{it} is the interaction of Q_i and $1[t = 3]$: $D_{it} = Q_i 1[t = 3]$, where $1[A] \equiv 1$ if A holds and 0 otherwise; only the $Q_i = 1$ group is treated at $t = 3$, and untreated otherwise.

DD can be implemented with panel data or repeated cross-sections (RCS). We focus on RCS in this paper, while relating RCS to panel data whenever necessary, because RCS are easier to deal with as they can be handled as cross-section iid data; also, our empirical analysis uses RCS. If one desires to apply a RCS method in our paper to panel data, then the simplest way is to pool the panel data and regard the time-series data for unit i as a cluster/group with related observations. Inference with clustering in RCS can be done by block bootstrap or a clustering-robust variance.

With a sampling dummy S_i independent of the other random variables, RCS have

$$D_i = Q_i S_i \quad \text{where} \quad S_i \equiv 1[\text{unit } i \text{ is sampled at } t = 3].$$

In RCS, there is a huge reservoir of units, and random sampling is done each period. Hence, we can safely assume that each unit appears only in one period. Let Y_{it}^d be the potential version of Y_{it} for $D_{it} = d$, and $Y_i^d = (1 - S_i)Y_{i2}^d + S_i Y_{i3}^d$ be the RCS potential outcome. Let W_{it} denote covariates, and $W_i \equiv (1 - S_i)W_{i2} + S_i W_{i3}$ be the RCS covariates; RCS variables are derived from the underlying panel model variables. Henceforth, unless unclear, we often omit the subscript i indexing units.

As a preliminary, ignoring the covariates W for a while, define for RCS:

$$\mu_{QS} \equiv E(Y|Q, S) = \lambda^{-1}(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D), \quad \beta_\tau \equiv \beta_3 - \beta_2, \quad (1.1)$$

where $\lambda(\cdot)$ is a ‘link function’ as in GLM (Nelder and Wedderburn 1972), (β_2, β_3) are the period-(2,3) intercepts, β_τ is the time effect of $t = 3$ relative to $t = 2$, β_q is the group effect of $Q = 1$, and β_d is the desired treatment effect.

Since (Q, S) generates four cells for four parameters $(\beta_2, \beta_\tau, \beta_q, \beta_d)$ in (1.1), there seems no loss of generality in (1.1). However, (1.1) does include a restriction: QS should not appear separately from the treatment D . If the group effect of Q changes across time, then QS becomes relevant other than through D . This restriction – no change in the group effect over time – is the DD parallel/common trend assumption.

For continuous Y , $\lambda(\cdot)$ in (1.1) is the identity, in which case DD is defined as

$$\mu_{11} - \mu_{10} - (\mu_{01} - \mu_{00}) = \beta_2 + \beta_\tau + \beta_q + \beta_d - (\beta_2 + \beta_q) - \{(\beta_2 + \beta_\tau) - \beta_2\} = \beta_d;$$

DD removes $\beta_2 + \beta_\tau S + \beta_q Q$ to leave $\beta_d D$ that changes across both times and groups. In practice, to account for the covariates W , a linear model such as

$$E(Y|Q, S, W) = \beta_2 + \beta_\tau S + \beta_q Q + \beta_d D + \beta'_w W \quad (1.2)$$

is typically estimated to find the slope of D as the treatment effect.

For LDV's, the story changes much. E.g., consider $Y = 1[0 \leq Y^*]$ where Y^* is the latent continuous outcome. With the $N(0, 1)$ distribution function $\Phi(\cdot)$, the probit is

$$E(Y|Q, S) = P(Y = 1|Q, S) = \Phi(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D). \quad (1.3)$$

One way to stick to DD is estimating (1.3) to interpret β_d as the effect on Y^* , not on Y . E.g., if $\beta_d = 2$, then D shifts Y^* by twice the standard deviation (SD) of Y^* . However, many practitioners desire the effect as a change in $P(Y = 1|Q, S)$, not in Y^* .

The ‘marginal effect’ that is a change of $P(Y = 1|Q, S)$ in (1.3) due to D is

$$\Phi(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d) - \Phi(\beta_2 + \beta_\tau S + \beta_q Q). \quad (1.4)$$

Ai and Norton (2003) noted that this is not the correct effect, but their criticism applies to the case of an interaction treatment, where both Q and S are genuine treatments and the interest is in the effect of taking both treatments

(e.g. drugs) together. Differently from this, Q and S are not treatments per se in the usual DD; $D = QS$ just happened to be the way the treatment was implemented. Indeed, Puhani (2012, eq. (10)) showed that (1.4) with $S = Q = 1$ is a legitimate treatment effect of interest.

The complication involving (1.3) and (1.4) arises because DD is applied to a nonlinear model, although DD is designed basically for linear models. To drive home our point, consider the ‘log link’ $\lambda(\cdot) = \ln(\cdot) \Leftrightarrow \lambda^{-1}(\cdot) = \exp(\cdot)$, with which (1.1) becomes

$$\mu_{QS} \equiv E(Y|Q, S) = \exp(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D). \quad (1.5)$$

This is appropriate for non-negative Y due to the lower bound $\mu_{QS} > 0$. For (1.5), ‘ratio in ratios (RR)’ removes the time and group effects, and ‘RR minus one’ renders

$$\text{‘proportional effect’}: \left(\frac{\mu_{11}}{\mu_{10}} \right) / \left(\frac{\mu_{01}}{\mu_{00}} \right) - 1 = \frac{E(Y_3^1 - Y_3^0 | Q = 1)}{E(Y_3^0 | Q = 1)} = \exp(\beta_d) - 1, \quad (1.6)$$

which is proven below. Just as the linear model is estimated instead of the DD to find β_d , the exponential regression model (1.5) is usually estimated instead of the RR.

For positive outcomes, one might consider DD with $\ln Y$, which raises the issue of transforming Y before applying DD or RR. For this issue, Athey and Imbens (2006) proposed “change in changes” based on Y -transformation-invariant distributional effects, neither an additive (for DD) nor multiplicative (for RR) effect. Also, Roth and Sant’Anna (2023) considered distribution-function-based DD that is invariant/insensitive to any strictly monotonic transformation of Y to propose a moment inequality test for parallel trends. Athey and Imbens (2006) and Roth and Sant’Anna (2023) are not exactly practical, because W should be controlled conditionally.

For binary or multinomial outcomes, ‘logit link’ can be used:

$$\mu_{QS} \equiv E(Y|Q, S) = \frac{\exp(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D)}{1 + \exp(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D)}; \quad (1.7)$$

the logit link is $\lambda(\cdot) \equiv \ln\{\cdot/(1 - \cdot)\} \Leftrightarrow \lambda^{-1}(\cdot) = \exp(\cdot)/\{1 + \exp(\cdot)\}$. For (1.7), “ratio in odds-ratios (ROR)” similar to RR removes the time and group effects, and ROR minus one renders a “proportional odds effect” similar to (1.6). For binary outcomes, instead of estimating ROR directly, the logistic model in (1.7) is typically estimated, and for multinomial outcomes, multinomial logistic model is estimated.

In the remainder of this paper, Section 2 reviews the literature of DD with LDV, and lists this paper’s contributions near the end. Sections 3 and 4 examine RR and ROR, where W is explicitly controlled in addition to (Q, S) and estimation is done with Maximum Likelihood Estimator (MLE) or Quasi-MLE (QMLE). Section 5 presents an empirical analysis. Finally, Section 6 concludes this paper. The Appendix examines ROR with multinomial Y , and provides a simulation study showing that the usual linear-model DD can be misleading for LDV’s, and RR or ROR had better be used.

2 Literature for DD with LDV

2.1 Early Papers on DD with LDV

Despite the frequent appearance of LDV’s in reality, studies on DD with LDV are fairly scarce. The first highly cited paper is Ai and Norton (2003) who suggested to use the cross-derivative as the treatment effect: with our notation,

$$\partial^2 E(Y|Q = q, S = s, W) / \partial q \partial s \quad (\text{or to use the analogous “cross – difference”}). \quad (2.1)$$

Ai and Norton (2003) has been cited about 7,500 times in *Google Scholar*.

Ai and Norton (2003) is correct when the interest is in the interaction of two genuine treatments. However, QS itself is not a treatment in DD. Rather, QS just represents the way how the single, not double, treatment D is administered: only to the $Q = 1$ group at $t = 3$. Hence, as was noted in Section 1, Puhani (2012, eq. (10)) showed

that $\Phi(\beta_2 + \beta_\tau + \beta_q + \beta_d) - \Phi(\beta_2 + \beta_\tau + \beta_q)$ is an appropriate treatment effect. Puhani (2012) has been cited about 880 times in *Google Scholar*.

Wooldridge (2023, p. C39) concurred with Puhani: “Puhani’s definition is the correct one for identifying τ_2 when the PT assumption is stated in terms of the linear index”, where τ_2 is the average effect on the treated at the post-treatment period and ‘PT’ means parallel trends. Puhani (2012) addressed only the effect identification, whereas Ai and Norton (2003) discussed how to estimate (2.1) as well, and showed with an empirical example that (2.1) can differ from the slope of $D = QS$ in sign.

Unaware of the early version of our paper in 2021, Taddeo et al. (2022) applied GLM to DD with LDV, as they proposed to use RR for count outcomes, and ROR for binary outcomes. Differently from our paper, however, Taddeo et al. (2022) considered neither zero-censored nor fractional outcome, although they mentioned categorical outcomes briefly. Also, the word ‘QMLE’ never appeared, and the types of covariates were not discussed although they matter much for categorical outcomes.

Taddeo et al. (2022, p. 404) also suggested a sensitivity analysis for parallel trend assumption. E.g., as to be seen later, the multiplicative form of parallel trends requires a RR form with only Y_t^0 ’s (analogous to the RR in (1.6)) being equal to one. Replacing the constant one with another number can reveal how the treatment effect is affected when the number deviates from one. For instance, if an empirical finding gets reversed when the constant is 1.5, then it takes a 50 % violation of the parallel trends, and thus the initial finding may be deemed to be robust to violations of parallel trends.

2.2 DD and Staggered DD with LDV

The paper that overlaps most with our paper is Wooldridge (2023), which was written to generalize Wooldridge (2021) for linear DD models to nonlinear ones; for this reason, the working paper version of Wooldridge (2023) was available only in 2022. Wooldridge (2023) initially examined the canonical two-group two-period DD, and then moved onto staggered DD which has been gaining popularity these days; see, e.g. Callaway and Sant’Anna (2021), Goodman-Bacon (2021), Sun and Abraham (2021), Athey and Imbens (2022), Liu et al. (2024), and Borusyak et al. (2024).

Table 1 at the end of Section 2 of Wooldridge (2023) summarizes the recommendation for the appropriate estimator for each type of LDV: Poisson QMLE for count, zero-censored and non-negative outcomes; logit for binary and fractional outcomes; and multinomial logit for multinomial/categorical outcomes. These recommendations agree with our paper, which is why Wooldridge (2023) overlaps much with our paper. A word of caution for those trying to read our paper and Wooldridge’s: our (Q_i, D_{it}, W_{it}) is (D_i, W_{it}, X_{it}) in the notation of Wooldridge (2023), which can be confusing.

Given the similarity between our paper and Wooldridge (2023), the reader may wonder what the differences are between the two papers. Wooldridge (2023) is more general than our paper, as it addresses staggered DD with LDV in multi-period settings. Wooldridge (2023) also compares ‘pooled estimators using all observations together’ to ‘imputation estimators using only the untreated observations to impute the counterfactual untreated outcomes of the $Q = 1$ group at $t = 3$ ’. In imputation estimators, the imputed entity is subtracted from the outcome average of the $Q = 1$ group at $t = 3$, which our paper using all observations together does not examine. Wooldridge (2023) further provides conditions to make the two types of estimators the same. In the following, we show what our paper has that Wooldridge (2023) does not.

First, when the adopted model is estimated for the LDV Y generated by its latent continuous version Y^* with a linear index $\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D + \beta'_w W$, our paper shows that β_d can be viewed either as the effect of D on Y^* , or as the proportional effect on Y for the log link as in (1.6) or the proportional-odds effect for the logit link. Even if Y^* is inconceivable, still the proportional effect interpretation holds. In contrast to this, Wooldridge (2023) takes one extra step to estimate $E(Y_3^1 - Y_3^0 | Q = 1, W)$ (or $E(Y_3^1 - Y_3^0 | W)$), which is not a necessity; e.g.,

$$\frac{1}{\sum_i Q_i} \sum_i Q_i \left\{ \exp(\hat{\beta}_2 + \hat{\beta}_\tau + \hat{\beta}_q + \hat{\beta}_d + \hat{\beta}'_w W_i) - \exp(\hat{\beta}_2 + \hat{\beta}_\tau + \hat{\beta}_q + \hat{\beta}'_w W_i) \right\}$$

is an estimator with the log link and QMLE estimates $(\hat{\beta}_2, \hat{\beta}_\tau, \hat{\beta}_q, \hat{\beta}_d, \hat{\beta}'_w)'$.

Second, we also review the more recent literature on DD with LDV in 2024 and 2025: Tchetgen et al. (2024) and Kim and Lee (2025), who tried to reduce the regression function specification errors. Kim and Lee (2025) is reviewed shortly below, whereas Tchetgen et al. (2024) is reviewed in Section 4 after ROR is introduced because their approach is applicable only with ROR.

Third, covariates are often “brushed aside” in the DD literature. That is, a theory is developed without W first, and when W appears, a remark such as “The theory developed without W holds with W , either by conditioning on W or by controlling W linearly” is made. Wooldridge (2023) also assumes time-constant W , and either condition on W or control W linearly. Handling W “lightly” in this way can be problematic, because the nature of W matters much in panel data (i.e. unit-constant/varying and time-constant/varying), and it is more so for categorical/multinomial outcomes, as the nature of W is three-dimensional (unit-constant/varying, time-constant/varying, and category-constant/varying), where normalizations matter much even for RCS as can be seen in Lee (2015) and our Appendix for multinomial outcomes.

Fourth, in addition to the above three main contributions, there are also some minor ones: (i) DD identification conditions are usually stated and verified for panel data models, but since we use QMLE/MLE with RCS, we verify them with RCS; (ii) “triple ratios” are used to deal with non-parallel multiplicative trends; (iii) we provide the details on justifying RR for zero-censored outcomes; (iv) for proportional odds effect which is difficult to grasp, we show a ‘rare event condition’ which turns the proportional odds effect into a proportional effect; and (v) we present the exact maximand of the QMLE/MLE, so that it can be easily understood and implemented by practitioners, going well beyond presenting a summary table as in Table 1 of Wooldridge (2023).

2.3 Causal Reduced Form for OLS to DD with LDV

Recently, Kim and Lee (2025) showed that, for any form of Y (binary, count, continuous, ...) with meaningful $Y^1 - Y^0$, a linear model holds always for $\Delta Y \equiv Y_3 - Y_2$, to which ordinary least squares (OLS) can be applied. To see how, recall $D_{it} = Q_i 1[t = 3]$ and observe:

$$Y_2 = Y_2^0, Y_3 = Y_3^0 + (Y_3^1 - Y_3^0)Q \Rightarrow \Delta Y \equiv Y_3 - Y_2 = \Delta Y^0 + (Y_3^1 - Y_3^0)Q.$$

Under PT $\Delta Y^0 \perp\!\!\!\perp Q|W$, take $E(\cdot|Q, W)$ on ΔY where ‘ $\perp\!\!\!\perp$ ’ stands for independence:

$$\begin{aligned} E(\Delta Y|Q, W) &= E(\Delta Y^0|W) + E(Y_3^1 - Y_3^0|Q = 1, W) \cdot Q \\ &\Rightarrow \Delta Y = E(\Delta Y^0|W) + E(Y_3^1 - Y_3^0|Q = 1, W)Q + U, \quad U \equiv \Delta Y - E(\Delta Y|Q, W). \end{aligned}$$

The ΔY equation linear in Q holds for any Y , and it is a “causal reduced form (CRF)” in the sense that it is a reduced/derived form with a causal parameter $\tau_1(W)$:

$$\Delta Y = E(\Delta Y^0|W) + \tau_1(W)Q + U, \quad \tau_1(W) \equiv E(Y_3^1 - Y_3^0|Q = 1, W), \quad E(U|Q, W) = 0. \quad (\text{CRF})$$

CRF may sound strange, but it has been used fruitfully in Lee (2018) and Lee et al. (2025) for binary exogenous D ; Lee (2021), Lee et al. (2023), Choi et al. (2023), and Kim and Lee (2024) for binary endogenous D ; Lee (2024) for binary exogenous D with a mediator; and Kim (2025) for network/spillover effect.

To remove $E(\Delta Y^0|W)$ from the CRF, take $E(\cdot|W)$ on the CRF:

$$E(\Delta Y|W) = E(\Delta Y^0|W) + \tau_1(W) \cdot p_W \quad \text{where} \quad p_W \equiv E(Q|W).$$

Subtract this from the CRF to get $\Delta Y - E(\Delta Y|W) = \tau_1(W) \cdot (Q - p_W) + U$. Now, OLS of $\Delta Y - E(\Delta Y|W)$ on $Q - p_W$ can be done, which is equivalent to OLS of ΔY on $Q - p_W$ because $Q - p_W$ is orthogonal to $E(\Delta Y|W)$. The propensity score p_W can be estimated with probit or logit.

Kim and Lee (2025) showed that OLS of ΔY on $Q - p_W$ is consistent to:

$$E\{w^{ow}(W)\tau_1(W)\} \quad \text{where} \quad w^{ow}(W) \equiv \frac{p_W(1-p_W)}{E\{p_W(1-p_W)\}} \geq 0 \quad \text{with} \quad E\{w^{ow}(W)\} = 1;$$

OW stands for “overlap weight”. *CRF is nonparametric as no parametric assumption is invoked, and OLS to the CRF is consistent to $E\{\omega^{ow}(W)\tau_1(W)\}$ for any LDV, as long as $Y^1 - Y^0$ makes sense.* Extending this finding to more general cases, such as multi-periods, non-binary D and endogenous Q , remains to be explored.

The fact that OLS estimates an OW-average of heterogeneous effects is well known (e.g. Angrist 1998; Angrist and Pischke 2009) under the “saturated model” assumption – i.e., p_W is equal to the linear projection of Q on W – which is, however, not necessary in the above OLS. See Choi and Lee (2023) for details on OW and its advantages. OW has been gaining popularity in statistics, epidemiology, and medical science; see, e.g. Li et al. (2018), Mao et al. (2019), Li (2019), Zhou et al. (2020), Thomas et al. (2020), Aminian et al. (2021), Cheng et al. (2022), Anderson et al. (2023), Wei et al. (2023), Matsouaka and Zhou (2024), and Xu et al. (2025).

3 Ratio in Ratios (RR) for Non-Negative Outcome

This section studies RR for non-negative outcomes including count and zero-censored outcomes. First, the identification aspect is examined. Second, although RR can be done by nonparametrically estimating the conditional means in RR, this is not how RR is done in practice; instead, a practical semiparametric estimator for RR using an exponential regression model is advocated. Third, several remarks are made.

3.1 Proportional Effect Identification with RR

To simplify notation when covariates W are allowed for, define

$$\mu_{QS}(w) \equiv E(Y|w, Q, S) \equiv E(Y|W = w, Q, S);$$

note $\mu_{Q0}(w) = E(Y_2|W_2 = w, Q)$ and $\mu_{Q1}(w) = E(Y_3|W_3 = w, Q)$. Analogously to (1.6), define RR or ‘proportional-effect +1 given $W = w$ ’:

$$RR(w) \equiv \frac{\mu_{11}(w)}{\mu_{10}(w)} / \frac{\mu_{01}(w)}{\mu_{00}(w)}.$$

The identification condition for $RR(w)$ is “unity multiplicative trends”:

$$\left(\frac{E(Y_3^0|w, Q=1)}{E(Y_2^0|w, Q=1)} \right) / \left(\frac{E(Y_3^0|w, Q=0)}{E(Y_2^0|w, Q=0)} \right) = 1; \quad (\text{ID}_{RR})$$

keep in mind that S is independent of the other random variables. In ID_{RR} , $E(Y_3^0|w, Q=1)$ is the counterfactual, because only Y_3^1 is realized for $Q=1$ at $t=3$.

ID_{RR} is analogous to the usual linear-model DD parallel trend condition:

$$\{E(Y_3^0|w, Q=1) - E(Y_2^0|w, Q=1)\} - \{E(Y_3^0|w, Q=0) - E(Y_2^0|w, Q=0)\} = 0, \quad (\text{ID}_{DD})$$

which dictates that the counterfactual $E(Y_3^0|w, Q=1)$ be constructed as $E(Y_2^0|w, Q=1) + E(Y_3^0|w, Q=0) - E(Y_2^0|w, Q=0)$. In the same vein, ID_{RR} dictates that the counterfactual be constructed multiplicatively as the baseline untreated outcome of the $Q=1$ group times the ratio of the control group means at $t=3$ to $t=2$:

$$E(Y_3^0|w, Q=1) = E(Y_2^0|w, Q=1) \times \frac{E(Y_3^0|w, Q=0)}{E(Y_2^0|w, Q=0)}. \quad (3.1)$$

The main point is that $RR(w) - 1$ is the ‘proportional effect on the treated at the post-treatment period’, as the first and last terms in the following show:

$$\begin{aligned}
RR(w) - 1 &= \left(\frac{E(Y|w, Q=1, S=1)}{E(Y|w, Q=1, S=0)} \right) / \left(\frac{E(Y|w, Q=0, S=1)}{E(Y|w, Q=0, S=0)} \right) - 1 \\
&= \left(\frac{E(Y_3^1|w, Q=1)}{E(Y_2^0|w, Q=1)} \right) / \left(\frac{E(Y_3^0|w, Q=0)}{E(Y_2^0|w, Q=0)} \right) - 1 \quad (\text{as } S \perp \text{the other variables}) \\
&= \frac{E(Y_3^1|w, Q=1)}{E(Y_3^0|w, Q=1)} \cdot \left(\frac{E(Y_3^0|w, Q=1)}{E(Y_2^0|w, Q=1)} \right) / \left(\frac{E(Y_3^0|w, Q=0)}{E(Y_2^0|w, Q=0)} \right) - 1 \\
&= \frac{E(Y_3^1|w, Q=1)}{E(Y_3^0|w, Q=1)} - 1 = \frac{E(Y_3^1 - Y_3^0|w, Q=1)}{E(Y_3^0|w, Q=1)} \quad (\text{under ID}_{RR}). \tag{3.2}
\end{aligned}$$

Instead of the difference effect $E(Y_3^1 - Y_3^0|w, Q=1)$, examining the proportional effect can be beneficial (Yadlowsky et al. 2021). E.g., suppose $E(Y_3^0|w, Q=1) = G(w)$ for a function $G(\cdot)$ and the proportional effect is a constant $\check{\beta}_d$. Then the difference effect $E(Y_3^1 - Y_3^0|w, Q=1) = \check{\beta}_d G(w)$ introduces effect heterogeneity unnecessarily, compared with the simple $\check{\beta}_d$. Proportional effects for exponential models have been advocated in many studies: Lee and Kobayashi (2001), Dukes and Vansteelandt (2018) and Ciani and Fisher (2019), among others.

If the dimension of W is low (or if W is discrete), $RR(w)$ can be estimated nonparametrically, but typically the dimension of W is high in practice, and thus we explore a simpler semiparametric exponential regression next – semiparametric because only $E(Y|W, Q, S)$ is specified, not the full distribution of $Y|(W, Q, S)$.

3.2 Poisson Quasi-MLE (QMLE)

In view of (1.5), suppose that a panel data exponential model holds for Y_{it}^d :

$$E(Y_{it}^d|W_2, W_3, Q) = E(Y_{it}^d|W_t, Q) = \exp(\beta_t + \beta_q Q + \beta_d d + \beta'_w W_t) \tag{3.3}$$

If periods $1, \dots, T$ were available, then $W_1, \dots, W_T$ would appear in the first term. Then ID_{RR} holds with (3.3):

$$\begin{aligned}
&\left(\frac{E(Y_3^0|w, Q=1)}{E(Y_2^0|w, Q=1)} \right) / \left(\frac{E(Y_3^0|w, Q=0)}{E(Y_2^0|w, Q=0)} \right) \\
&= \left(\frac{\exp(\beta_3 + \beta_q + \beta'_w w)}{\exp(\beta_2 + \beta_q + \beta'_w w)} \right) / \left(\frac{\exp(\beta_3 + \beta'_w w)}{\exp(\beta_2 + \beta'_w w)} \right) = 1. \tag{3.4}
\end{aligned}$$

For the corresponding RCS model, the observed Y in RCS is: due to $D = QS$,

$$\begin{aligned}
Y &= (1-S)Y_2 + SY_3 = (1-S)Y_2^0 + S \cdot \{(1-Q)Y_3^0 + QY_3^1\} \\
&= (1-S)Y_2^0 + S(1-Q)Y_3^0 + DY_3^1.
\end{aligned}$$

Take $E(\cdot|W_2, W_3, Q, S)$ on this Y and invoke (3.3): recalling $\beta_\tau \equiv \beta_3 - \beta_2$ in (1.1),

$$\begin{aligned}
E(Y|W_2, W_3, Q, S) &= (1-S)\exp(\beta_2 + \beta_q Q + \beta'_w W_2) \\
&\quad + S(1-Q)\exp(\beta_3 + \beta_q Q + \beta'_w W_3) + D\exp(\beta_3 + \beta_q Q + \beta_d + \beta'_w W_3) \\
&= \exp(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D + \beta'_w W), \quad W \equiv (1-S)W_2 + SW_3 \tag{3.5}
\end{aligned}$$

$$\Rightarrow RR(w) - 1 = \exp(\beta_d) - 1; \tag{3.6}$$

the second equality can be verified by substituting $Q, S = 0, 1$ into both sides. If β_d is near zero, then β_d is an approximate proportional effect. RCS data and the model (3.5) can be used to estimate β_d along with the other parameters.

The heterogeneous effect can be easily allowed by replacing β_d with $\beta_d(W_t) = \beta_{d0} + \beta'_{dw} W_t$. Then we have $\beta_d(W) = \beta_{d0} + \beta'_{dw} W$ in RCS, and (3.5) becomes

$$E(Y|W, Q, S) = \exp\{\beta_2 + \beta_\tau S + \beta_q Q + \beta_d(W)D + \beta'_w W\} \Rightarrow RR(w) - 1 = \exp\{\beta_d(w)\} - 1.$$

It is often believed that if a treatment effect is random but a constant effect is assumed, then its mean effect would be estimated. However, this is false as the following explains.

The staggered DD literature revealed that, if a constant effect is assumed when the true effect is heterogeneous, then the popular two-way fixed effect estimator is consistent to a weighted average of heterogeneous effects with some weights negative. This paints a rather pessimistic picture, but positive findings also exist: if OLS is applied to a constant-effect model when the effect is actually heterogeneous in W , then the OLS is consistent to an OW average of heterogeneous effects; see Lee and Han (2024) and references therein. This kind of positive scenario was already noted when Kim and Lee (2025) was reviewed. It would take another full paper to investigate the analogous question for proportional effect in general settings, which we thus eschew.

The simplest estimator for (3.5) is Poisson QMLE, whose maximand is:

$$\sum_i \{Y_i(X'_i b) - \exp(X'_i b)\}, \quad X_i \equiv (1, S_i, Q_i, D_i, W'_i)', \quad b = (b_2, b_\tau, b_q, b_d, b'_w)'$$

as in Poisson MLE. The first order condition at $b = \beta$ is $\sum_i \{Y_i - \exp(X'_i \beta)\} X_i = 0$ for $\beta \equiv (\beta_2, \beta_\tau, \beta_q, \beta_d, \beta'_w)'$, and the second order derivative $-\sum_i X_i X'_i \exp(X'_i \beta)$ is n.d. for all b . Hence, just under $E(Y|X) = \exp(X' \beta)$, Poisson QMLE is consistent for β . For heterogeneous effects, we may use $\beta_d(W) = \beta_{d0} + \beta'_{dw} W$ in Poisson QMLE.

The difference between Poisson QMLE and Poisson MLE is that the first-order condition is taken just as a moment condition to estimate β in the former. Hence, the asymptotic variance of Poisson QMLE is estimated with a “sandwich-form” estimator. Since Poisson QMLE requires only $E(Y|X) = \exp(X' \beta)$, not the full distribution of $Y|X$, it is advocated by Santos Silva and Tenreiro (2006), which has been cited about 9,300 times in *Google Scholar* as of this writing.

3.3 Remarks

Here we make remarks on the applicability of the above RR identification and Poisson QMLE to count and zero-censored outcomes. The semiparametric exponential model (3.5) is appropriate for these, as it requires only the lower bound $0 \leq Y$.

3.3.1 Remarks on Count Outcomes

Suppose (3.5) holds under

$$Y = \exp(Y^*), \quad Y^* \equiv \beta_2 + \beta_\tau S + \beta_q Q + \beta_d D + \beta'_w W + U, \\ E\{\exp(U)|W_2, W_3, Q, S\} = E\{\exp(U)|W_2, W_3, Q\} = 1.$$

Then we can interpret β_d as the DD effect on Y^* , whereas $\exp(\beta_d) - 1$ is the proportional effect on Y . However, if Y is a count outcome with $P(Y = y|X) = \{\exp(X' \beta)\}^y \exp\{-\exp(X' \beta)\} / y!$ (i.e. Poisson distribution), then there is no Y^* , and the proportional effect interpretation on the observed Y with RR is the only way to interpret the slope β_d of D in the exponential model. This statement applies also to count outcomes based on other distributions such as Negative Binomial.

If $\beta_{q\tau}tQ$ with $\beta_{q\tau} \neq 0$ appears as a regressor in (3.3), then ID_{RR} fails due to $\beta_{q\tau}tQ$:

$$\left(\frac{\exp(\beta_3 + \beta_q + 3\beta_{q\tau} + w'\beta_w)}{\exp(\beta_2 + \beta_q + 2\beta_{q\tau} + w'\beta_w)} \right) / \left(\frac{\exp(\beta_3 + w'\beta_w)}{\exp(\beta_2 + w'\beta_w)} \right) = \exp(\beta_{q\tau}) \neq 1; \quad (3.7)$$

call this “non-unity multiplicative trends” – a RR analog for non-parallel (i.e. additive) trends in DD. Using tQ as a regressor is an easy way to test/allow for non-unity multiplicative trends, but tQ cannot be used if only two periods are available, because using tQ is equivalent to using $QS = D$.

One way to allow for non-unity multiplicative trends with more than two waves is using tQ as an extra regressor as just noted, and using tQ in the usual linear DD appeared in, e.g. Goodman-Bacon (2018), Dobkin et al. (2018), and Hwang and Lee (2020). For panel data, tQ can be used as such, but for RCS, $Q_i^t \equiv Q_i \sum_t S_{it} t$ should be used where $S_{it} = 1$ if i is sampled in period t and 0 otherwise. Then, the untreated group ratio/difference is allowed to change systematically with tQ over time, and any deviation from the systematic change is taken as the treatment effect. Using tQ can be generalized to using (tQ, t^2Q) as in Friedberg (1998) and Wolfers (2006), and going further, $(tQ, t^2Q, t^3Q, \dots)$ as in Lee (2016b).

Another way to deal with this kind of trend discrepancy is using triple ratios, or “ratio in ratios in ratios (RRR)” generalizing RR, analogously to triple differences (Lee 2016b) to allow for non-parallel trends in DD. With $t = 1, 2, 3$ available, let:

$$\left(\frac{E(Y_t^0|w, Q=1)}{E(Y_{t-1}^0|w, Q=1)} \right) / \left(\frac{E(Y_t^0|w, Q=0)}{E(Y_{t-1}^0|w, Q=0)} \right) = \gamma \quad \text{for } t = 2, 3, \quad (ID_{RRR})$$

which allows ID_{RR} to be violated when $\gamma \neq 1$ as follows.

With $m_{Qt}(w) \equiv E(Y|w, Q, \text{sampled at } t)$, observe

$$\begin{aligned} RRR(w) &\equiv \left[\frac{m_{13}(w)}{m_{12}(w)} / \left\{ \frac{m_{03}(w)}{m_{02}(w)} \right\} \right] / \left[\frac{m_{11}(w)}{m_{10}(w)} / \left\{ \frac{m_{01}(w)}{m_{00}(w)} \right\} \right] \\ &= \left[\left(\frac{E(Y_3^1|w, Q=1)}{E(Y_2^0|w, Q=1)} \right) / \left(\frac{E(Y_3^0|w, Q=0)}{E(Y_2^0|w, Q=0)} \right) \right] \\ &\quad / \left[\left(\frac{E(Y_2^0|w, Q=1)}{E(Y_1^0|w, Q=1)} \right) / \left(\frac{E(Y_2^0|w, Q=0)}{E(Y_1^0|w, Q=0)} \right) \right] \\ &= \frac{E(Y_3^1|w, Q=1)}{E(Y_3^0|w, Q=1)} \cdot \left[\left(\frac{E(Y_3^0|w, Q=1)}{E(Y_2^0|w, Q=1)} \right) / \left(\frac{E(Y_3^0|w, Q=0)}{E(Y_2^0|w, Q=0)} \right) \right] \\ &\quad / \left[\left(\frac{E(Y_2^0|w, Q=1)}{E(Y_1^0|w, Q=1)} \right) / \left(\frac{E(Y_2^0|w, Q=0)}{E(Y_1^0|w, Q=0)} \right) \right]. \end{aligned}$$

The last two terms in $[\cdot]$ are both equal to γ to cancel each other. Hence, under ID_{RRR} , RRR identifies the same effect as RR identifies, even when ID_{RR} fails.

3.3.2 Remarks on Zero-Censored Outcomes

Consider a RCS zero-censored model:

$$\begin{aligned} Y &= \max(0, Y^*) = Y^*1[0 < Y^*], \quad Y^* \equiv \beta_2 + \beta_\tau S + \beta_q Q + \beta_d D + \beta'_w W + U \\ &\Rightarrow E(Y|X) = E(Y^*1[0 < Y^*]|X). \end{aligned} \quad (3.8)$$

Since $E(Y|X) = E(Y^*1[0 < Y^*]|X)$ is non-negative without any upper bound, the exponential regression model (3.5) can be adopted, although it may not be as appealing as for count outcomes because the transformation $\max(0, \cdot)$ is not smooth.

Santos Silva and Tenreiro (2011) showed that the exponential regression holds for (3.8) if $Y_i = \sum_{j=1}^{M_i} Z_{ij}$, where M_i is a non-negative integer random variable such as Poisson count, and $(Z_{i1}, Z_{i2}, \dots)$ are iid positive random variables with $Z_{ij} \perp M_i | X_i$; $Y = 0$ occurs if $M = 0$. Then,

$$\begin{aligned} E(Y|X) &= E(M|X)E(Z_j|X) \\ &= \exp\{\alpha_2 + \beta_2 + (\alpha_\tau + \beta_\tau)S + (\alpha_q + \beta_q)Q + \beta_d D + (\alpha_w + \beta_w)'W\} \\ \text{if } E(M|X) &= \exp(\alpha_2 + \alpha_\tau S + \alpha_q Q + \alpha'_w W), \\ E(Z_j|X) &= \exp(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D + \beta'_w W). \end{aligned} \quad (3.9)$$

It is not clear what Y^* is here, but the interpretation of $\exp(\beta_d) - 1$ as a proportional effect on Y still holds regardless of what Y^* might be.

An example for $Y_i = \sum_{j=1}^{M_i} Z_{ij}$ is that Y_i is the person- i expenditure on unfrozen fish in a year, Z_{ij} is the expenditure on unfrozen fish in month j , M_i is the number of the unfrozen-fish-purchasing months, $Q_i = 1$ if living close enough to sea to access unfrozen fish, and $D_i = Q_i S_i$ is a policy to increase unfrozen fish prices. Here, M_i is how frequently unfrozen fish are purchased, which is unlikely to be affected by the policy because one cannot stock up on unfrozen fish, whereas Z_{ij} would be affected.

4 Ratio in Odds Ratios (ROR)

This section studies ROR. We examine the identification aspect first, followed by logit-based estimation for binary and fractional outcomes, and then by a review on Tchetgen et al. (2024). ROR is also applicable to multinomial outcome, but it is presented (along with a simulation study) in the Appendix due to the complexity involving multiple equations and additional notation.

4.1 Proportional Odds Effect Identification with ROR

For binary Y , define the ‘odds conditional on $(W = w, Q = q, S = s)$ ’ for RCS as

$$\begin{aligned} R_{qs}(Y; w) &\equiv \frac{P(Y = 1 | w, Q = q, S = s)}{P(Y = 0 | w, Q = q, S = s)} \quad \text{which leads to} \\ R_{11}(Y; w) &= R_{11}(Y_3^1; w), \quad R_{01}(Y; w) = R_{01}(Y_3^0; w), \\ R_{10}(Y; w) &= R_{10}(Y_2^0; w), \quad R_{00}(Y; w) = R_{00}(Y_2^0; w). \end{aligned} \quad (4.1)$$

Also define ‘Ratio in Odds Ratios (ROR) conditional on $W = w$ ’:

$$ROR(Y; w) \equiv \left(\frac{R_{11}(Y; w)}{R_{10}(Y; w)} \right) / \left(\frac{R_{01}(Y; w)}{R_{00}(Y; w)} \right).$$

The identification condition to be invoked for ROR is

$$ROR(Y^0; w) = \left(\frac{R_{11}(Y_3^0; w)}{R_{10}(Y_2^0; w)} \right) / \left(\frac{R_{01}(Y_3^0; w)}{R_{00}(Y_2^0; w)} \right) = 1, \quad (\text{ID}_{ROR})$$

where $R_{11}(Y_3^0; w)$ is a counterfactual. Just as in (3.1), ID_{ROR} dictates that the counterfactual $R_{11}(Y_3^0; w)$ be constructed multiplicatively:

$$R_{11}(Y_3^0; w) = R_{10}(Y_2^0; w) \times \frac{R_{01}(Y_3^0; w)}{R_{00}(Y_2^0; w)}.$$

Doing analogously to (3.2), $ROR(Y; w) - 1$ is the ‘proportional odds effect on the treated at the post-treatment period $t = 3$ ’:

$$\begin{aligned}
ROR(Y; w) - 1 &= \left(\frac{R_{11}(Y; w)}{R_{10}(Y; w)} \right) / \left(\frac{R_{01}(Y; w)}{R_{00}(Y; w)} \right) - 1 \\
&= \left(\frac{R_{11}(Y_3^1; w)}{R_{10}(Y_2^0; w)} \right) / \left(\frac{R_{01}(Y_3^0; w)}{R_{00}(Y_2^0; w)} \right) - 1 \\
&= \frac{R_{11}(Y_3^1; w)}{R_{11}(Y_3^0; w)} \cdot \left(\frac{R_{11}(Y_3^0; w)}{R_{10}(Y_2^0; w)} \right) / \left(\frac{R_{01}(Y_3^0; w)}{R_{00}(Y_2^0; w)} \right) - 1 \\
&= \frac{R_{11}(Y_3^1; w)}{R_{11}(Y_3^0; w)} - 1 = \frac{R_{11}(Y_3^1; w) - R_{11}(Y_3^0; w)}{R_{11}(Y_3^0; w)} \quad (\text{under ID}_{ROR}). \quad (4.2)
\end{aligned}$$

Although odds ratio is extensively used in the medical and epidemiological literature, still it is not necessarily easy to interpret the proportional odds effect. For this, suppose $Y = 1$ is a *rare event* in the sense

$$\frac{P(Y_3^1 = 0|w, Q = 1)}{P(Y_3^0 = 0|w, Q = 1)} \simeq 1 \quad \text{for all } w; \quad (4.3)$$

e.g. $Y = 1$ is a rare cancer occurrence such that $P(Y_3^d = 0|w, Q = 1) \simeq 1$ for all w and $d = 0, 1$. Under (4.3), $ROR(Y; w) = R_{11}(Y_3^1; w)/R_{11}(Y_3^0; w)$ in (4.2) becomes

$$\frac{P(Y_3^1 = 1|w, Q = 1)/P(Y_3^1 = 0|w, Q = 1)}{P(Y_3^0 = 1|w, Q = 1)/P(Y_3^0 = 0|w, Q = 1)} \simeq \frac{P(Y_3^1 = 1|w, Q = 1)}{P(Y_3^0 = 1|w, Q = 1)}.$$

Hence, the proportional odds effect (4.2) becomes the RR proportional effect (3.2):

$$ROR(Y; w) - 1 \simeq \frac{E(Y_3^1 - Y_3^0|w, Q = 1)}{E(Y_3^0|w, Q = 1)} \quad \text{under the rare event condition (4.3).}$$

$ROR(Y; w)$ can be estimated nonparametrically by substituting sample analogs into the components of $ROR(Y; w)$. However, as was the case for DD and $RR(w)$, this is not what practitioners would do. Instead, we apply logistic regression next.

4.2 Logit for Binary Outcome

Analogously to (3.3), suppose that a panel data logistic model holds for Y_{it}^d :

$$E(Y_{it}^d|W_2, W_3, Q) = E(Y_{it}^d|W_t, Q) = \frac{\exp(\beta_t + \beta_q Q + \beta_d d + \beta'_w W_t)}{1 + \exp(\beta_t + \beta_q Q + \beta_d d + \beta'_w W_t)}. \quad (4.4)$$

The logistic panel data model for Y_{it}^0 renders, with $\beta_\tau \equiv \beta_3 - \beta_2$,

$$\begin{aligned}
R_{11}(Y_3^0; w) &= \exp(\beta_2 + \beta_\tau + \beta_q + \beta'_w w), & R_{01}(Y_3^0; w) &= \exp(\beta_2 + \beta_\tau + \beta'_w w), \\
R_{10}(Y_2^0; w) &= \exp(\beta_2 + \beta_q + \beta'_w w), & R_{00}(Y_2^0; w) &= \exp(\beta_2 + \beta'_w w).
\end{aligned} \quad (4.5)$$

Hence, ID_{ROR} holds analogously to (3.4), because (4.5) yields

$$ROR(Y^0; w) = \left(\frac{R_{11}(Y_3^0; w)}{R_{10}(Y_2^0; w)} \right) / \left(\frac{R_{01}(Y_3^0; w)}{R_{00}(Y_2^0; w)} \right) = 1.$$

For the corresponding RCS, $Y = (1 - S)Y_2^0 + S(1 - Q)Y_3^0 + DY_3^1$ holds. Analogously to (3.5) and (3.6), take $E(\cdot|W_2, W_3, Q, S)$ on Y and invoke (4.4) to get

$$E(Y|W_2, W_3, Q, S) = \frac{\exp(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D + \beta'_w W)}{1 + \exp(\beta_2 + \beta_\tau S + \beta_q Q + \beta_d D + \beta'_w W)}, \quad W \equiv (1 - S)W_2 + SW_3. \quad (4.6)$$

Also, $R_{11}(Y_3^1; w) = \exp(\beta_2 + \beta_\tau + \beta_q + \beta_d + w'\beta_w)$ and $R_{11}(Y_3^0; w)$ in (4.5) yield

$$ROR(Y; w) - 1 = \{R_{11}(Y_3^1; w)/R_{11}(Y_3^0; w)\} - 1 = \exp(\beta_d) - 1.$$

Estimate β_d by the MLE with (4.6) to use $\exp(\beta_d) - 1$ as the proportional odds effect, which is also the RR proportional effect when $Y = 1$ is a rare event as in (4.3).

Suppose $\beta_{q\tau}tQ$ with $\beta_{q\tau} \neq 0$ appears as an extra regressor in (4.4). Then the parallel trends do not hold for the latent Y^* . The appearance of $\beta_{q\tau}tQ$ also ruins ID_{ROR} for binary Y because ID_{ROR} becomes (3.7), just as $\beta_{q\tau}tQ$ ruins ID_{RR} in (3.7). As in (3.7), using tQ is an easy way to test or allow for non-parallel trends in Y^* . Overall, the comments made for (3.7) hold more or less the same for ROR as well.

To allow for heterogeneous effect, suppose now that the slope of d in (4.4) is $\beta_d(W_t)$; e.g. $\beta_d(W_t) = \beta_{d0} + \beta'_{dw} W_t$. Then (4.6) and $ROR(Y; w) - 1$ become, respectively,

$$E(Y|W, Q, S) = \frac{\exp\{\beta_2 + \beta_\tau S + \beta_q Q + \beta_d(W)D + \beta'_w W\}}{1 + \exp\{\beta_2 + \beta_\tau S + \beta_q Q + \beta_d(W)D + \beta'_w W\}},$$

$$ROR(Y; w) - 1 = \exp\{\beta_d(w)\} - 1.$$

ROR is not applicable to ordinal outcomes, but they can be reduced to binary in multiple ways, and then the overlapping information in those ways can be combined with minimum distance estimation (see, e.g. Lee 2005, 2015).

4.3 Logit for Fractional Outcome

When Y takes on a value in $[0,1]$, Y is a fractional outcome; e.g. the share of asset invested in stocks. There are two types of fractional outcomes: (i) $P(Y = 0 \text{ or } Y = 1) = 0$ and (ii) $P(Y = 0 \text{ or } Y = 1) > 0$. Since logistic regression always gives a value in $(0,1)$, it can be adopted for type-(i) fractional outcome. As for type (ii), analogously to $\max(0, Y^*)$, we can use $Y = \max\{0, \min(Y^*, 1)\}$. Since $\max\{0, \min(\cdot, 1)\}$ is not smooth, one may object to adopting a logistic model for type (ii), but the following provides a justification under $0 < E(Y|X) < 1$ for all X .

For a function B_X of X with $0 < B_X < 1$ for all X , consider maximizing

$$E\{Y \ln B_X + (1 - Y) \ln(1 - B_X)\} = E[E(Y|X) \ln B_X + \{1 - E(Y|X)\} \ln(1 - B_X)] \quad (4.7)$$

with respect to (wrt) B_X . Differentiate (4.7) wrt B_X to obtain

$$E[E(Y|X)B_X^{-1} - \{1 - E(Y|X)\}(1 - B_X)^{-1}] = 0 \quad \text{when } B_X = E(Y|X).$$

Further differentiate this wrt B_X to see that $B_X = E(Y|X)$ is the unique maximizer:

$$E[-E(Y|X)B_X^{-2} - \{1 - E(Y|X)\}(1 - B_X)^{-2}] < 0.$$

Using this fact, Papke and Wooldridge (1996) maximized wrt b :

$$\sum_i \left\{ Y_i \ln \frac{\exp(X_i' b)}{1 + \exp(X_i' b)} + (1 - Y_i) \ln \frac{1}{1 + \exp(X_i' b)} \right\} \quad \text{under } E(Y|X) = \frac{\exp(X' \beta)}{1 + \exp(X' \beta)};$$

X_i and b were defined for the Poisson QMLE. The first-order condition is

$$\sum_i \left\{ Y_i - \frac{\exp(X_i' b)}{1 + \exp(X_i' b)} \right\} X_i = 0 \quad (\text{satisfied at } b = \beta).$$

As in Poisson QMLE, a “sandwich form” variance estimator should be used.

4.4 ROR and Extended Propensity Score

A disadvantage of proportional odds effect is its interpretation difficulty, unless the rare event condition (4.3) holds. For this, Tchetgen et al (2024; “TPR”) proposed to find $E(Y_3^1 - Y_3^0 | W, Q = 1)$ under a condition close to ID_{ROR} , as explained in this subsection. Although we assume discrete Y to ease exposition here, since each value of $y \neq 0$ is paired with $y = 0$, one might think of Y just as binary.

Define: for $y \neq 0$,

$$R_{qsy}(Y; w) \equiv \frac{P(Y = y | w, Q = q, S = s)}{P(Y = 0 | w, Q = q, S = s)} \quad \text{which leads to} \quad (4.8)$$

$$R_{11y}(Y; w) = R_{11y}(Y_3^1; w), \quad R_{01y}(Y; w) = R_{01y}(Y_3^0; w),$$

$$R_{10y}(Y; w) = R_{10y}(Y_2^0; w), \quad R_{00y}(Y; w) = R_{00y}(Y_2^0; w).$$

Also, assume “odds ratio equi-confounding” which generalizes ID_{ROR} :

$$\beta_2(w, y) = \beta_3(w, y) \text{ where } \beta_t(w, y) \equiv \ln \frac{P(Y_t^0 = y | w, Q = 1) / P(Y_t^0 = 0 | w, Q = 1)}{P(Y_t^0 = y | w, Q = 0) / P(Y_t^0 = 0 | w, Q = 0)},$$

with $y = 1$, $\beta_2(w, 1) = \beta_3(w, 1)$ for all w is equivalent to ID_{ROR} .

Using the Baye’s rule, rewrite $\beta_t(w, y)$ as:

$$\begin{aligned} \beta_t(w, y) &= \ln \frac{P(Q = 1 | w, Y_t^0 = y) / P(Q = 1 | w, Y_t^0 = 0)}{P(Q = 0 | w, Y_t^0 = y) / P(Q = 0 | w, Y_t^0 = 0)} \\ &= \ln \frac{P(Q = 1 | w, Y_t^0 = y)}{P(Q = 0 | w, Y_t^0 = y)} - \ln \frac{P(Q = 1 | w, Y_t^0 = 0)}{P(Q = 0 | w, Y_t^0 = 0)} \end{aligned} \quad (4.9)$$

$$\Leftrightarrow \ln \frac{P(Q = 1 | w, Y_t^0 = y)}{1 - P(Q = 1 | w, Y_t^0 = y)} = \ln \frac{P(Q = 1 | w, Y_t^0 = 0)}{P(Q = 0 | w, Y_t^0 = 0)} + \beta_t(w, y) \quad (4.10)$$

putting the last term of (4.9) on the opposite side. TPR calls $P(Q = 1 | w, Y_t^0 = y)$ “the extended propensity score function (given y)”.

TPR approximates the first term on the right-hand side of (4.10) linearly with $\eta_t' w$, and $\beta_t(w, y)$ with $\alpha_w' w + \alpha_y y$ (no t in α ’s, due to $\beta_2(w, y) = \beta_3(w, y)$), where η_t and α ’s are parameters; TPR also allows the interaction term $\alpha_{wy}' w y$, which is omitted here. The logit of $P(Q = 1 | w, Y_t^0 = y)$ in the left-hand side of (4.10) renders:

$$P(Q = 1 | W, Y_t^0) = \frac{\exp(\eta_t' W + \alpha_w' W + \alpha_y Y_t^0)}{1 + \exp(\eta_t' W + \alpha_w' W + \alpha_y Y_t^0)} = \frac{\exp\{(\eta_t + \alpha_w)' W + \alpha_y Y_t^0\}}{1 + \exp\{(\eta_t + \alpha_w)' W + \alpha_y Y_t^0\}}.$$

Now, $(\eta_t' + \alpha_w', \alpha_y)'$ can be estimated with the logistic MLE using only the $t = 2$ observations due to $Y_2 = Y_2^0$. However, $\eta_3 + \alpha_w$ cannot be estimated in the analogous way because the $Q = 1$ group does not have Y_3^0 , which is addressed next.

Observe:

$$P(Q = 1 | W, Y_3^0) = P(Q = 0 | W, Y_3^0) \cdot \exp\{(\eta_3 + \alpha_w)' W + \alpha_y Y_3^0\}$$

which implies

$$E\{(1 - Q) \exp\{(\eta_3 + \alpha_w)' W + \alpha_y Y_3\} - Q | W\} = 0 \quad (4.11)$$

$$\Rightarrow E\{E[(1 - Q) \exp\{(\eta_3 + \alpha_w)' W + \alpha_y Y_3^0\} - Q | W, Y_3^0] | W\} = 0$$

$$\Rightarrow E\{E[(1 - Q) \exp\{(\eta_3 - \eta_2)' W + (\eta_2 + \alpha_w)' W + \alpha_y Y_3^0\} - Q | W, Y_3^0] | W\} = 0$$

$$\Rightarrow E\{E[(1 - Q) \exp\{(\eta_3 - \eta_2)' W + (\eta_2 + \alpha_w)' W + \alpha_y Y_3^0\} - Q | W, Y_3^0] \cdot W\} = 0.$$

The last unconditional moment condition identifies $\eta_3 - \eta_2$ as the solution, if $\eta_2 + \alpha_w$ and α_y are known, for which $\beta_t(w, y)$ should be time-constant, so that $\eta_2 + \alpha_w$ and α_y found at $t = 2$ can be substituted into the moment condition at $t = 3$ to identify $\eta_3 - \eta_2$. Henceforth, write $(\eta_3 - \eta_2)'W + (\eta_2 + \alpha_w)'W$ as $(\eta_3 + \alpha_w)'W$ as in (4.11).

With all parameters identified, the DD effect $E(Y_3^1|Q = 1) - E(Y_3^0|Q = 1)$ is:

$$E(Y_3|Q = 1) - \frac{E[(1 - Q)Y_3 \exp\{(\eta_3 + \alpha_w)'W + \alpha_y Y_3\}]}{E[(1 - Q) \exp\{(\eta_3 + \alpha_w)'W + \alpha_y Y_3\}]} \quad (4.12)$$

To understand the second term, observe that its numerator is equal to:

$$\begin{aligned} & E[E\{(1 - Q)Y_3^0 \exp\{(\eta_3 + \alpha_w)'W + \alpha_y Y_3^0\} | W, Y_3^0\}] \\ &= E[Y_3^0 \exp\{(\eta_3 + \alpha_w)'W + \alpha_y Y_3^0\} \cdot E\{(1 - Q)|W, Y_3^0\}] \\ &= E\left[Y_3^0 \exp\{(\eta_3 + \alpha_w)'W + \alpha_y Y_3^0\} \cdot \frac{1}{1 + \exp\{(\eta_3 + \alpha_w)'W + \alpha_y Y_3^0\}}\right] \\ &= E\{Y_3^0 P(Q = 1|W, Y_3^0)\} = E\{E(Y_3^0 Q|W, Y_3^0)\} = E(Y_3^0 Q). \end{aligned}$$

Analogously, the denominator of the second term in (4.12) is $E(Q)$, which thus makes the second term of (4.12) equal to $E(Y_3^0|Q = 1)$.

Evaluating the contribution of TPR, they turn ROR into something that involves only $P(Q|\cdot)$, using the Bayes' rule. This way, one has to specify a binary model for Q , not for Y , which renders no advantage for binary Y though. For non-binary Y , such as ordinal, multinomial or even continuous, not specifying the outcome probability/density would be an advantage. However, it is not clear whether $\beta_2(w, y) = \beta_3(w, y)$ holds or not for a general Y , as ID_{ROR} and its generalization $\beta_2(w, y) = \beta_3(w, y)$ were motivated by the (multinomial) logit form of the Y_t^0 probability.

This problem notwithstanding, the main contribution of TPR is that, given the difficulty of interpreting proportional odds effect unless the rare event condition is invoked, TPR showed a way to find the usual DD effect under the ID_{ROR} -type condition $\beta_2(w, y) = \beta_3(w, y)$, avoiding the interpretation difficulty of proportional odds effect.

5 Empirical Analysis

In this section, we estimate the effects of the Affordable Care Act Dependent Coverage Provision ('DCP') on various health outcomes. Under the DCP that went into effect in September 2010, dependents can remain on the parent's private health plan until age 26. The treatment group is dependents aged 23–25, and the control group is dependents aged 27–29; 26 was excluded due to the treatment status ambiguity. Our data are from the Behavioral Risk Factor Surveillance System for 2007–2013, which is health-related telephone surveys in the U.S. Almost the same data were used in Barbaresco et al (2015) ('BCQ', henceforth), with small differences occurring due to updates, imputed values, data cleaning, etc.

BCQ considered 18 outcomes, of which we use 12. Each outcome variable has a different sample size, as we replaced "Don't Know" and "Refused" with missing values. With the sample size in brackets [...], the 12 outcome variables are: **(1)** 'any health insurance' [127,618], **(2)** 'any primary (care) doctor' [127,533], **(3)** needed medical care in the past year not taken due to cost ('cost blocked care') [108,433], **(4)** current smoker [126,557], **(5)** 'risky drinker (in the past 30 days)' [122,035], **(6)** 'obese (BMI ≥ 30)' [121,294], **(7)** 'pregnant (while) unmarried' [40,006], **(8)** '(alcoholic) drinks (in the past) 30 days' [121,845], **(9)** BMI [121,290], **(10)** days of the last 30 not in good mental health ('days poor mental') [125,681], **(11)** days of the last 30 not in good physical health ('days poor physical') [125,766], and **(12)** days of the last 30 with health-related limitations ('days health limits') [71,079]. The first seven outcomes are binary, and the remaining five are non-negative (counts or continuous).

Table 1 presents summary statistics on covariates: age, gender, race, marital status, education, state unemployment rate, ‘any DCP’ for any state mandate on DCP although the dependent may not be covered, household income, the number of children, ‘cell phones only’ (vs. cell phone plus landline), student, and unemployed. Because the treatment group is younger than the control group by 2–6 years, the treatment group has fewer married, fewer college degree, lower household income, fewer children, more students, and more unemployed. Also, the treatment group has the lower state unemployment rate, higher any DCP, and higher cell phone only.

Let ‘Lin-DD’ stand for the linear model DD using (1.2). Table 2 shows β_{qr} (non-parallel/multiplicative trends) and β_d (effect) estimates, although the effect of interest is the proportional effect $\exp(\beta_d) - 1$ due to Y being an LDV. Poisson QMLE estimates are $\tilde{\beta}_{qr}$ and $\tilde{\beta}_d$, and Lin-DD estimates ignoring the LDV nature are $\hat{\beta}_{qr}$ and $\hat{\beta}_d$. In Table 2, age-clustering-robust standard errors are computed using the survey weights.

Three main findings emerge from Table 2, which are also seen in the simulation part of the Appendix: (i) RR and ROR estimates differ much from Lin-DD estimates; (ii) ROR estimates are overall greater than Lin-DD whereas RR estimates are overall smaller than Lin-DD; and (iii) tests for $\beta_{qr} = 0$ in Lin-DD and in RR or ROR give the same conclusions at the 5 % significance level. Also, comparing RR, ROR and Lin-DD in their qualitative conclusions on testing for $\beta_d = 0$, they lead to the same conclusions at the 5 % significance level, except for ‘cost blocked care’.

Turning to interpreting effect magnitude, proportional odds effects for binary outcomes are a little difficult to interpret; e.g. DCP increases the odds of ‘any health insurance’ by about 41 %, an approximation to $\exp(0.41) - 1 \simeq 51$ %. This should not be taken as a drastic effect, because odds ratios can easily take on large values, which is, in fact, one of the reasons why some researchers prefer odds ratios. Compared with the overall large magnitudes in proportional odds effects for binary outcomes, the proportional effect magnitudes for non-negative outcomes are in a smaller scale and easier to interpret, ranging just over -0.004 to 0.263 ; e.g. DCP increases ‘drinks 30 days’ by about 26 %, an approximation to $\exp(0.26) - 1 \simeq 30$ %. As an example for proportional odds effects becoming proportional effects for rare events, unmarried pregnancies are fairly rare

Table 1: Summary statistics of covariates: mean & standard deviation (SD).

Covariates	Treated Mean (SD)	Control Mean (SD)	Covariates	Treated Mean (SD)	Control Mean (SD)
Age (age 23 omitted)			Household income (less than \$10 K omitted)		
Age 24	0.35 (0.48)	–	\$10 K–\$15 K	0.07 (0.26)	0.05 (0.22)
Age 25	0.32 (0.47)	–	\$15 K–\$20 K	0.10 (0.30)	0.08 (0.27)
Age 27	–	0.31 (0.46)	\$20 K–\$25 K	0.12 (0.32)	0.10 (0.30)
Age 28	–	0.34 (0.48)	\$25 K–\$35 K	0.14 (0.35)	0.13 (0.33)
Age 29	–	0.35 (0.48)	\$35 K–\$50 K	0.16 (0.37)	0.16 (0.37)
Female	0.51 (0.50)	0.51 (0.50)	\$50 K–\$75 K	0.14 (0.35)	0.18 (0.39)
Race (non-Hispanic whites omitted)			\$75 K \$ over	0.19 (0.39)	0.24 (0.43)
Black	0.11 (0.31)	0.11 (0.32)	Number of children		
Hispanic	0.23 (0.42)	0.22 (0.41)	1	0.23 (0.42)	0.23 (0.42)
Others	0.09 (0.28)	0.08 (0.27)	2	0.16 (0.36)	0.23 (0.42)
Married	0.30 (0.46)	0.56 (0.50)	3	0.05 (0.23)	0.11 (0.31)
Education (less than HS degree omitted)			4	0.02 (0.13)	0.04 (0.19)
High school (HS)	0.28 (0.45)	0.26 (0.44)	5 or more	0.01 (0.09)	0.02 (0.12)
Non-4-yr college	0.30 (0.46)	0.27 (0.44)	Cell phone only	0.70 (0.46)	0.67 (0.47)
College graduate	0.31 (0.46)	0.36 (0.48)	Student	0.11 (0.31)	0.05 (0.23)
State unemp. rate	7.23 (2.72)	7.37 (2.73)	Unemployed	0.13 (0.34)	0.12 (0.33)
Any DCP	0.26 (0.44)	0.04 (0.21)			

Any DCP: the state has any DCP mandate despite that the person is not covered.

Table 2: Non-parallel/multiplicative trend $\beta_{q\tau}$ & effect β_d : estimate (t-value).

Outcome variable	RR and ROR		Lin-DD (linear model DD)	
	$\tilde{\beta}_{q\tau}$ Estimate (tv)	$\tilde{\beta}_d$ Estimate (tv)	DD $\tilde{\beta}_{q\tau}$ Estimate (tv)	DD $\tilde{\beta}_d$ Estimate (tv)
<i>Binary outcome</i>				
Any health insurance	−0.009 (−0.27)	0.407 (3.06)	−0.002 (−0.34)	0.068 (2.88)
Any primary doctor	0.010 (0.39)	0.145 (1.28)	0.002 (0.34)	0.028 (1.18)
Cost blocked care	0.014 (2.78)	−0.167 (−1.82)	0.003 (3.54)	−0.029 (−1.97)
Current smoker	−0.053 (−2.70)	0.204 (2.89)	−0.009 (−2.86)	0.035 (3.15)
Risky drinker	0.027 (1.63)	−0.092 (−2.44)	0.005 (1.90)	−0.019 (−2.88)
Obese	0.032 (1.90)	−0.081 (−1.07)	0.007 (1.86)	−0.018 (−1.02)
Pregnant unmarried	0.008 (0.12)	−0.079 (−0.36)	0.000 (0.11)	−0.003 (−0.39)
<i>Non-negative outcome</i>				
Drinks 30 days	−0.085 (−3.48)	0.263 (3.05)	−1.346 (−4.03)	4.242 (3.04)
BMI	−0.001 (−0.83)	−0.004 (−0.73)	−0.035 (−0.86)	−0.104 (−0.66)
Days poor mental	−0.004 (−0.15)	0.050 (0.43)	−0.019 (−0.18)	0.203 (0.43)
Days poor physical	−0.037 (−0.94)	0.189 (1.18)	−0.087 (−0.96)	0.423 (1.13)
Days health limits	0.019 (0.47)	0.052 (0.30)	0.050 (0.40)	0.149 (0.28)

Table 3: Effects under parallel/multiplicative trends.

Outcome variable	RR and ROR Estimate (tv)	Lin-DD Estimate (tv)
<i>Binary outcome</i>		
Any health insurance	0.375 (4.57)	0.061 (4.41)
Any primary doctor	0.178 (5.26)	0.035 (4.76)
Cost blocked care	−0.113 (−1.34)	−0.019 (−1.38)
Current smoker	0.016 (0.49)	0.003 (0.49)
Risky drinker	0.006 (0.23)	0.001 (0.13)
Obese	0.033 (1.06)	0.008 (1.10)
Pregnant unmarried	−0.049 (−0.58)	−0.002 (−0.57)
<i>Non-negative outcome</i>		
Drinks 30 days	−0.037 (−0.50)	−0.569 (−0.39)
BMI	−0.009 (−3.21)	−0.227 (−3.19)
Days poor mental	0.036 (0.82)	0.136 (0.76)
Days poor physical	0.057 (2.32)	0.113 (1.90)
Days health limits	0.121 (3.05)	0.329 (2.54)

(4–5 %) in our data, and consequently, we can interpret the ROR estimate −0.078 for ‘pregnant unmarried’ as a 8 % decrease due to DCP.

BCQ checked out the parallel trend assumption with graphs plotting the pre-treatment trends across the treatment and control groups. BCQ also estimated their models using different time periods or using more aggregated data. Whereas these are informal ways of testing for parallel trends, our approach of using tQ as an extra regressor provides a formal yet simple way of testing parallel/unity-multiplicative trends. The $\tilde{\beta}_{q\tau}$ estimates reveal that parallel trend assumption in Y^* and the analogous ID_{RR}/ID_{ROR} assumptions do not hold at least for ‘cost blocked care’, ‘current smoker’, and ‘drinks 30 days’ while tests for $\beta_{q\tau} = 0$ in Lin-DD give almost the same conclusions.

To appreciate better how much difference allowing $\beta_{q\tau} \neq 0$ makes, Table 3 repeats Table 2 under the restriction $\beta_{q\tau} = 0$ (i.e. without using tQ as a regressor). The differences between Tables 2 and 3 are substantial both in terms of effect magnitude and t -value. In RR and ROR, only ‘any health insurance’ maintained their statistical significance, whereas ‘current smoker’, ‘risky drinker’ and ‘drinks 30 days’ become misleadingly insignificant by imposing $\beta_{q\tau} = 0$ falsely. Also, ‘any primary doctor’, BMI, ‘days poor physical’ and ‘days health limits’ become significant by imposing $\beta_{q\tau} = 0$ unnecessarily. In Lin-DD as well, only ‘any health insurance’ maintains its statistical significance in Tables 2 and 3, whereas the statistical significance of seven other outcomes is switched.

The main finding in BCQ is that DCP increases ‘any health insurance’, ‘any primary doctor’ and ‘risky drinker’, but decreases BMI. This finding is similar to that of the RR and ROR column in Table 3, except for ‘risky drinker’ that is insignificant in Table 3. This similarity is due to $\beta_{q\tau} = 0$ assumed in both BCQ and Table 3.

Since Table 3 imposes the unnecessary restriction $\beta_{q\tau} = 0$, it is interesting to compare the finding in BCQ to that in Table 2. The RR and ROR column of Table 2 reveals significantly increasing effects on ‘any health insurance’, ‘current smoker’ and ‘drinks 30 days’, and a significantly decreasing effect on ‘risky drinker’. Hence, only the increasing effect on ‘any health insurance’ is shared by BCQ and the RR and ROR column of Table 2; the sign of ‘risky drinker’ changes across BCQ and the RR and ROR column of Table 2. Overall, the differences due to allowing $\beta_{q\tau} \neq 0$ are large.

6 Conclusions

Difference in Differences (DD) is one of the most popular approaches in finding the effect of a binary treatment D on an outcome Y . However, DD is suitable for linear models, and consequently, applying DD to limited dependent variables (LDV's), or more generally to nonlinear models, has been problematic. Many researchers with LDV's simply ignore the LDV nature to use a linear model.

The goal of this paper is to explore what can be done in this case, adopting the framework of generalized linear models (GLM) with link functions. Because several papers have been published on GLM for DD with LDV since an early version of our paper in 2021, we reviewed the literature including recent studies not covered by the other reviews. The main recommendation is using Poisson QMLE for non-negative (such as count or zero-censored) Y and (multinomial) logit MLE for binary, fractional or multinomial Y . This agrees with the other studies adopting GLM. Despite the overlap, however, this paper differs from the other reviews as follows.

First, we focused on proportional (odds) effect using ‘ratio in ratios (RR)’ or ‘ratio in odds ratios (ROR)’, instead of trying to obtain the usual average effect on the treated (or on the population) which requires an unnecessary extra step. Second, as was noted just above, we reviewed the recent literature on DD with LDV that has not been covered by the other reviews. Third, we paid more attention to covariate (W) types, instead of simply assuming a time-constant W . Fourth, we made further minor contributions, such as verifying the identification condition for repeated cross sections (RCS) and proposing triple ratios to deal with non-parallel multiplicative trends.

Our simulation study revealed that RR and ROR can give much different findings from DD, although our empirical study showed that the DD findings may not differ much from RR and ROR as well. Also, using a power function of t times the policy qualification dummy to account for non-parallel/multiplicative trends made big differences in our empirical findings, compared to imposing the parallel/multiplicative trend restriction unnecessarily from the outset.

Acknowledgments: The authors are grateful to the Editor and a reviewer for their helpful comments and directing the authors’ attention to the relevant literature.

Data and program availability: The data and programs used in this paper will be made publicly available on a public depository, once the paper is accepted.

Compliance with ethical standard, no conflict of interest, and no AI usage: No human/animal subject is involved in this research, and there is no conflict of interest to disclose. Also, no generative AI-related technology has been used for this paper.

Funding information: The research of Myoung-jae Lee has been supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00337766).

Appendix

A.1 Simulation Study

Our simulation study addresses four types of LDV's: (i) positive continuous, (ii) count, (iii) zero-censored, and (iv) binary. Poisson QMLE is applied to (i), (ii) and (iii), and logistic MLE to (iv); their estimates are compared with the linear model DD ('Lin-DD'). Fractional Y is not tried because it is not yet clear how to generate fractional Y subject to the exponential regression, and multinomial Y is addressed separately below because it is inconceivable to apply Lin-DD to multinomial outcome.

In the following, we explain (i) and Table A1 in detail, from which it will be clear how (ii), (iii) and (iv) are dealt with and how to interpret the other tables. In all cases, the effect of interest is $\exp(\beta_d) - 1$, which is the proportional (odds) effect, but we examine β_d mainly, because knowing β_d is equivalent to knowing $\exp(\beta_d) - 1$.

For (i) positive continuous outcome, we generate Y_{it} for $t = 0, 1, 2, 3$:

$$\begin{aligned} Y_{it} &= \exp(\beta_t + \beta_q Q_i + \beta_{q\tau} t Q_i + \beta_d D_{it} + U_{it}) \quad \text{where} \\ P(Q_i = 0) &= P(Q_i = 1) = 0.5, D_{it} = Q_i 1[t = 3], U_{i0}, U_{i1}, U_{i2}, U_{i3} \text{ iid } N(0, 1) \perp Q_i, \\ \beta_0 &= -2, \beta_1 = -2, \beta_2 = -1, \beta_3 = -1, \beta_q = 0.5, \beta_{q\tau} = 0, 0.5, \beta_d = 0, 0.5; \end{aligned} \quad (\text{A.1})$$

$\beta_{q\tau} = 0$ makes the parallel trends hold in Y^* and ID_{RR} hold in Y , but not $\beta_{q\tau} = 0.5$.

From the Y_{it} 's in (A.1), the RCS outcome Y_i and its regressor X_i are obtained:

$$\begin{aligned} S_i \text{ is the sampled period for } i, S_{it} &\equiv 1[S_i = t], P(S_{it} = 1) = 0.25 \quad \text{for all } t, \\ Y_i &= \sum_{t=0}^3 S_{it} Y_{it}, \quad Q_i^\tau \equiv Q_i \sum_{t=1}^3 S_{it} t, \quad X_i \equiv (1, S_{i1}, S_{i2}, S_{i3}, Q_i, Q_i^\tau, D_i)' \\ &\text{for parameters } \{\beta_0 + \ln(1.64), \beta_1 - \beta_0, \beta_2 - \beta_0, \beta_3 - \beta_0, \beta_q, \beta_{q\tau}, \beta_d\}; \end{aligned} \quad (\text{A.2})$$

1.64 comes from $E\{\exp(U_{it})\} = 1.64$ with $U_{it} \sim N(0, 1)$, which appears due to

$$E(Y_{it} | Q_i, S_i) = E\{\exp(\beta_t + \beta_q Q_i + \beta_{q\tau} t Q_i + \beta_d D_{it}) \cdot E\{\exp(U_{it})\}.$$

That is, having $E\{\exp(U_{it})\} = 1.64$ and β_0 is equivalent to having $E\{\exp(U_{it})\} = 1$ and $\beta_0 + \ln(1.64)$. For the other LDV models, the RCS data are generated analogously.

Table A1 presents the results from 5,000 repetitions with $N = 250$ and 10,000; each entry shows the absolute bias ($|\text{Bias}|$), SD, and Root Mean Squared Error (RMSE). With $N = 250$, Lin-DD estimates $\hat{\beta}_{q\tau}$ and $\hat{\beta}_d$ do sometimes better than the Poisson QMLE $\tilde{\beta}_{q\tau}$ and $\tilde{\beta}_d$, but this is due to the low SD's; the $|\text{Bias}|$ of Lin-DD $\hat{\beta}_d$ is huge in several cases. With $N = 10,000$, the $|\text{Bias}|$'s for the Lin-DD estimates remain almost the same as those with $N = 250$ whereas the gaps in SD between Lin-DD and Poisson QMLE are reduced, and consequently, Poisson QMLE does better than Lin-DD. Using tQ solves the problem of ID_{RR} violation for Poisson QMLE, but not for Lin-DD; $\hat{\beta}_{q\tau}$ in Lin-DD is biased much even when $\beta_{q\tau} = 0$. In short, Table A1 demonstrates that Lin-DD is highly biased when the true model is exponential for positive Y .

For (ii) count outcome, similarly to (A.1), Y_{it} is generated from the Poisson distribution with parameter $\exp(\beta_t + \beta_q Q_i + \beta_{q\tau} t Q_i + \beta_d D_{it})$ for $t = 0, 1, 2, 3$. Then Y_i and X_i are generated as in (A.2), and Poisson

Table A1: Positive Y : |Bias|, SD and (RMSE).

$N = 250$	$\beta_{q\tau}, \beta_d: 0, 0$	$\beta_{q\tau}, \beta_d: 0.5, 0$	$\beta_{q\tau}, \beta_d: 0, 0.5$	$\beta_{q\tau}, \beta_d: 0.5, 0.5$
$\tilde{\beta}_{q\tau}$	0.00 0.23 (0.23)	0.00 0.24 (0.24)	0.00 0.23 (0.23)	0.00 0.24 (0.24)
$\tilde{\beta}_d$	0.01 0.60 (0.60)	0.01 0.60 (0.60)	0.01 0.60 (0.60)	0.01 0.60 (0.60)
DD $\hat{\beta}_{q\tau}$	0.12 0.15 (0.19)	0.48 0.33 (0.59)	0.12 0.15 (0.19)	0.48 0.33 (0.59)
DD $\hat{\beta}_d$	0.08 0.46 (0.47)	1.05 1.38 (1.73)	0.08 0.56 (0.57)	3.48 1.96 (4.00)
$N = 10,000$				
$\tilde{\beta}_{q\tau}$	0.00 0.04 (0.04)	0.00 0.04 (0.04)	0.00 0.04 (0.04)	0.00 0.04 (0.04)
$\tilde{\beta}_d$	0.00 0.10 (0.10)	0.00 0.10 (0.10)	0.00 0.10 (0.10)	0.00 0.10 (0.10)
DD $\hat{\beta}_{q\tau}$	0.12 0.02 (0.13)	0.48 0.05 (0.49)	0.12 0.02 (0.13)	0.48 0.05 (0.49)
DD $\hat{\beta}_d$	0.08 0.07 (0.11)	1.03 0.22 (1.05)	0.07 0.09 (0.11)	3.44 0.30 (3.45)

$\beta_{q\tau} = 0$ for parallel trends in Y^* & ID_{RR} in Y ; $\beta_d (\exp(\beta_d) - 1)$ is the desired effect; $\tilde{\beta}_{q\tau}, \tilde{\beta}_d$: Poisson QMLE; $\hat{\beta}_{q\tau}, \hat{\beta}_d$: linear-model DD.

Table A2: Poisson count Y : |Bias|, SD and (RMSE).

$N = 250$	$\beta_{q\tau}, \beta_d: 0, 0$	$\beta_{q\tau}, \beta_d: 0.5, 0$	$\beta_{q\tau}, \beta_d: 0, 0.5$	$\beta_{q\tau}, \beta_d: 0.5, 0.5$
$\tilde{\beta}_{q\tau}$	0.01 0.40 (0.40)	0.00 0.38 (0.38)	0.01 0.40 (0.40)	0.00 0.38 (0.38)
$\tilde{\beta}_d$	0.00 0.85 (0.85)	0.00 0.78 (0.78)	0.01 0.84 (0.84)	0.00 0.78 (0.78)
DD $\hat{\beta}_{q\tau}$	0.08 0.10 (0.13)	0.10 0.14 (0.17)	0.08 0.10 (0.13)	0.10 0.14 (0.17)
DD $\hat{\beta}_d$	0.06 0.31 (0.31)	0.61 0.47 (0.77)	0.17 0.33 (0.37)	1.88 0.53 (1.95)
$N = 10,000$				
$\tilde{\beta}_{q\tau}$	0.00 0.06 (0.06)	0.00 0.05 (0.05)	0.00 0.06 (0.06)	0.00 0.05 (0.05)
$\tilde{\beta}_d$	0.00 0.12 (0.12)	0.00 0.10 (0.10)	0.00 0.11 (0.11)	0.00 0.10 (0.10)
DD $\hat{\beta}_{q\tau}$	0.08 0.02 (0.08)	0.10 0.02 (0.10)	0.08 0.02 (0.08)	0.10 0.02 (0.10)
DD $\hat{\beta}_d$	0.05 0.05 (0.07)	0.62 0.07 (0.63)	0.16 0.05 (0.16)	1.89 0.08 (1.89)

$\beta_{q\tau} = 0$ for parallel trends in Y^* & ID_{RR} in Y ; $\beta_d (\exp(\beta_d) - 1)$ is the desired effect; $\tilde{\beta}_{q\tau}, \tilde{\beta}_d$: Poisson QMLE; $\hat{\beta}_{q\tau}, \hat{\beta}_d$: linear-model DD.

QMLE is implemented. The same parameters as in (A.2) are estimated except for the intercept because $\ln(1.64)$ is no more present. Table A2 presents the simulation results, and what was mentioned for Table A1 applies to Table A2 almost word to word.

For (iii) zero-censored outcome, we use (3.9) where $M_i \sim \text{Poisson}(1)$ with $P(M_i = 0) = 0.37$ and $Y_{it} = \sum_{j=0}^{M_i} Z_{ijt}$ with $Z_{ijt} = \exp\{\beta_t + \beta_q Q_i + \beta_{q\tau} t Q_i + \beta_d D_{it} + N(0, 1)\}$. The same parameters as in (A.2) are estimated except for the intercept because $\exp(1)$ from $E(M)$ is added to β_0 in view of (3.9). Despite the big difference in the data generating processes, Table A3 differ little from Tables A1 and A2, and all comments made for Tables A1 and A2 apply to Table A3 as well. The similarities in the findings from Tables A1–A3 seem to stem from the common exponential regression specification.

For (iv) binary outcome, Y_{it} is generated with Logistic U_{it} :

$$Y_{it} = 1[0 < \beta_t + \beta_q Q_i + \beta_{q\tau} t Q_i + \beta_d D_{it} + U_{it}], \quad U_{i0}, U_{i1}, U_{i2}, U_{i3} \text{ are iid Logistic.}$$

$\beta_{q\tau} = 0$ makes the parallel trends hold for Y^* , and makes ID_{ROR} hold for Y .

Since the logistic regression is used in Table A4, the results in Table A4 differ much from those in Tables A1–A3. First, the overall magnitude of |Bias| is much smaller than in Tables A1–A3. Second, surprisingly, when $\beta_{q\tau} = \beta_d = 0$, the Lin-DD estimates with almost zero bias do several times better than the logistic MLE estimates. Third, biases in Lin-DD are persistent even when N increases to 10,000, which implies that Lin-DD will be eventually dominated by logistic MLE for a large enough N . Nevertheless, less harm is seen in using Lin-DD for binary Y , compared with the other LDV's.

Table A3: Zero-censored Y : |Bias|, SD and (RMSE).

$N = 250$	$\beta_{qr}, \beta_d: 0, 0$	$\beta_{qr}, \beta_d: 0.5, 0$	$\beta_{qr}, \beta_d: 0, 0.5$	$\beta_{qr}, \beta_d: 0.5, 0.5$
$\tilde{\beta}_{qr}$	0.01 0.31 (0.31)	0.01 0.32 (0.32)	0.01 0.31 (0.31)	0.01 0.32 (0.32)
$\tilde{\beta}_d$	0.01 0.80 (0.80)	0.02 0.80 (0.80)	0.01 0.80 (0.80)	0.02 0.80 (0.80)
DD $\hat{\beta}_{qr}$	0.12 0.19 (0.22)	0.47 0.42 (0.63)	0.12 0.19 (0.22)	0.47 0.42 (0.63)
DD $\hat{\beta}_d$	0.07 0.59 (0.60)	1.04 1.76 (2.04)	0.07 0.71 (0.72)	3.44 2.48 (4.24)
$N = 10,000$				
$\tilde{\beta}_{qr}$	0.00 0.05 (0.05)	0.00 0.05 (0.05)	0.00 0.05 (0.05)	0.00 0.05 (0.05)
$\tilde{\beta}_d$	0.00 0.12 (0.12)	0.00 0.13 (0.13)	0.00 0.12 (0.12)	0.00 0.13 (0.13)
DD $\hat{\beta}_{qr}$	0.12 0.03 (0.13)	0.48 0.07 (0.49)	0.12 0.03 (0.13)	0.48 0.07 (0.49)
DD $\hat{\beta}_d$	0.08 0.09 (0.12)	1.03 0.28 (1.06)	0.07 0.11 (0.13)	3.43 0.39 (3.46)

$\beta_{qr} = 0$ for parallel trends in Y^* & ID_{RR} in Y ; $\beta_d (\exp(\beta_d) - 1)$ is the desired effect; $\tilde{\beta}_{qr}, \tilde{\beta}_d$: Poisson QMLE; $\hat{\beta}_{qr}, \hat{\beta}_d$: linear model DD.

Table A4: Binary Y : |Bias|, SD and (RMSE).

$N = 250$	$\beta_{qr}, \beta_d: 0, 0$	$\beta_{qr}, \beta_d: 0.5, 0$	$\beta_{qr}, \beta_d: 0, 0.5$	$\beta_{qr}, \beta_d: 0.5, 0.5$
$\tilde{\beta}_{qr}$	0.01 0.50 (0.50)	0.01 0.51 (0.51)	0.01 0.50 (0.50)	0.01 0.51 (0.51)
$\tilde{\beta}_d$	0.02 1.16 (1.16)	0.04 1.18 (1.19)	0.04 1.15 (1.15)	0.13 1.54 (1.55)
DD $\hat{\beta}_{qr}$	0.02 0.07 (0.08)	0.35 0.08 (0.36)	0.02 0.07 (0.08)	0.35 0.08 (0.36)
DD $\hat{\beta}_d$	0.02 0.21 (0.21)	0.02 0.21 (0.21)	0.39 0.21 (0.45)	0.43 0.20 (0.48)
$N = 10,000$				
$\tilde{\beta}_{qr}$	0.00 0.07 (0.07)	0.00 0.07 (0.07)	0.00 0.07 (0.07)	0.00 0.07 (0.07)
$\tilde{\beta}_d$	0.00 0.17 (0.17)	0.00 0.17 (0.17)	0.00 0.17 (0.17)	0.00 0.17 (0.17)
DD $\hat{\beta}_{qr}$	0.02 0.01 (0.03)	0.35 0.01 (0.35)	0.02 0.01 (0.03)	0.35 0.01 (0.35)
DD $\hat{\beta}_d$	0.02 0.03 (0.04)	0.02 0.03 (0.04)	0.39 0.03 (0.39)	0.43 0.03 (0.43)

$\beta_{qr} = 0$ for parallel trends in Y^* & ID_{ROR} in Y ; $\beta_d (\exp(\beta_d) - 1)$ is the desired effect; $\tilde{\beta}_{qr}, \tilde{\beta}_d$: logit estimates; $\hat{\beta}_{qr}, \hat{\beta}_d$: linear-model DD.

A.2 Multinomial Logit for DD with Multinomial Outcome

A.2.1 Identification

For multinomial outcome Y taking on a value among $0, 1, \dots, C$ classes, define the ‘class- c odds’ (with the base class 0) conditional on $(W = w, Q = q, S = s)$ as

$$R_{qs}^c(Y; w) \equiv \frac{P(Y = c | w, Q = q, S = s)}{P(Y = 0 | w, Q = q, S = s)} \quad \text{which implies}$$

$$R_{11}^c(Y; w) = R_{11}^c(Y_3^1; w), \quad R_{01}^c(Y; w) = R_{01}^c(Y_3^0; w),$$

$$R_{10}^c(Y; w) = R_{10}^c(Y_2^0; w), \quad R_{00}^c(Y; w) = R_{00}^c(Y_2^0; w),$$

analogously to (4.1). Also define ‘class- c ROR conditional on $W = w$ ’:

$$ROR^c(Y; w) \equiv \left(\frac{R_{11}^c(Y; w)}{R_{10}^c(Y; w)} \right) / \left(\frac{R_{01}^c(Y; w)}{R_{00}^c(Y; w)} \right).$$

The identification condition for ROR^c with multinomial outcome is

$$ROR^c(Y^0; w) = \left(\frac{R_{11}^c(Y_3^0; w)}{R_{10}^c(Y_2^0; w)} \right) / \left(\frac{R_{01}^c(Y_3^0; w)}{R_{00}^c(Y_2^0; w)} \right) = 1. \quad (ID_{ROR^c})$$

As in (4.2), $ROR^c(Y; w) - 1$ is equal to the ‘class- c proportional odds effect on the treated at the post-treatment period $t = 3$ ’:

$$ROR^c(Y; w) - 1 = \frac{R_{11}^c(Y_3^1; w) - R_{11}^c(Y_3^0; w)}{R_{11}^c(Y_3^0; w)} \quad \text{under ID}_{ROR^c}.$$

Also, as in (4.3), if $Y = c \neq 0$ is a rare event in the sense of (4.3), then

$$ROR^c(Y; w) - 1 \simeq \frac{P(Y_3^1 = c | w, Q = 1) - P(Y_3^0 = c | w, Q = 1)}{P(Y_3^0 = c | w, Q = 1)} \quad (\text{A.3})$$

which is the class- c proportional effect on the treated at the post-treatment period.

A.2.2 Estimation

In panel multinomial choice with classes $c = 0, 1, \dots, C$, there are a few possibilities for regressors, depending on whether they vary across units, classes or times. Here, we consider three types of regressors: A_i varying only across units (e.g. race), H_{it} varying only across units and times (e.g. income), and W_{ict} varying across units, classes and times (e.g. expense from choosing class c). Let the ‘latent utility from class c ’ of unit i at period $t = 2, 3$ be

$$L_{itc}^d \equiv \beta_{tc} + \beta_{qc}Q_i + \beta_{dc}d + \beta'_{ac}A_i + \beta'_{hc}H_{it} + \beta'_{wc}W_{itc} + U_{itc}, \quad c = 0, 1, \dots, C \quad (\text{A.4})$$

where the error terms $(U_{i20}, \dots, U_{i2C}, U_{i30}, \dots, U_{i3C})$ are iid with the type-I extreme value distribution, and independent of all regressors at all times (‘strict exogeneity’).

The potential choice Y_{it}^d with $D = d$ is

$$Y_{it}^d = \sum_{j=0}^C \left(j \times 1 \left[L_{itj}^d > L_{itk}^d \quad \text{for all } k \neq j \right] \right);$$

Y_{it}^d takes on $0, 1, \dots, C$, depending on which class gives the maximum utility. Using (A.4), the choice probabilities for the untreated $Y_{it}^0 = c \in \{0, 1, \dots, C\}$ are:

$$\begin{aligned} P(Y_{it}^0 = c | Q_i, A_i, H_{it}, W_{i20}, \dots, W_{i2C}, W_{i30}, \dots, W_{i3C}) \\ &= P(Y_{it}^0 = c | Q_i, A_i, H_{it}, W_{it0}, \dots, W_{itC}) \\ &= \frac{\exp(\beta_{tc} + \beta_{qc}Q_i + \beta'_{ac}A_i + \beta'_{hc}H_{it} + \beta'_{wc}W_{itc})}{\sum_{j=0}^C \exp(\beta_{tj} + \beta_{qj}Q_i + \beta'_{aj}A_i + \beta'_{hj}H_{it} + \beta'_{wj}W_{itj})} \\ &= \frac{\exp(\Delta\beta_{tc} + \Delta\beta_{qc}Q_i + \Delta\beta'_{ac}A_i + \Delta\beta'_{hc}H_{it} - \beta'_{w0}W_{it0} + \beta'_{wc}W_{itc})}{1 + \sum_{j=1}^C \exp(\Delta\beta_{tj} + \Delta\beta_{qj}Q_i + \Delta\beta'_{aj}A_i + \Delta\beta'_{hj}H_{it} - \beta'_{w0}W_{it0} + \beta'_{wj}W_{ijt})}, \\ \Delta\beta_{tj} &\equiv \beta_{tj} - \beta_{t0}, \quad \Delta\beta_{qj} \equiv \beta_{qj} - \beta_{q0}, \quad \Delta\beta_{aj} \equiv \beta_{aj} - \beta_{a0}, \quad \Delta\beta_{hj} \equiv \beta_{hj} - \beta_{h0}; \end{aligned}$$

the last equality holds, dividing through by $\exp(\beta_{t0} + \beta_{q0}Q_i + \beta'_{a0}A_i + \beta'_{h0}H_{it} + \beta'_{w0}W_{it0})$ for the base class $c = 0$. The numerator of the last ratio is one for the base class 0. ID_{ROR^c} holds for $P(Y_{it}^0 = c | \cdot)$, whose proof is similar to the proof for ID_{ROR} .

Analogously derive the model for $P(Y_{it}^1 = c | \cdot)$, which then gives (i omitted)

$$\begin{aligned} R_{11}^c(Y^d; w) &= \exp(\Delta\beta_{3c} + \Delta\beta_{qc} + \Delta\beta_{dc}d + \Delta\beta'_{ac}A + \Delta\beta'_{hc}H_3 - \beta'_{w0}W_{30} + \beta'_{wc}W_{3c}) \\ \text{where } \Delta\beta_{dc} &\equiv \beta_{dc} - \beta_{d0}. \end{aligned}$$

Since $P(Y_{it}^1 = c | \cdot)$ differs from $P(Y_{it}^0 = c | \cdot)$ only in the extra term $\Delta\beta_{dc}$, we get the class- c proportional effect, if the rare event condition (4.3) holds:

$$\{R_{11}^c(Y_3^1; w)/R_{11}^c(Y_3^0; w)\} - 1 = \exp(\Delta\beta_{dc}) - 1.$$

In RCS, we observe $Y \equiv (1 - S)Y_2 + SY_3$ where $Y_t = 0, \dots, C$, along with Q, S and

$$A, H \equiv (1 - S)H_2 + SH_3, \quad W_0 \equiv (1 - S)W_{20} + SW_{30}, \dots, \quad W_C \equiv (1 - S)W_{2C} + SW_{3C};$$

with i omitted. The choice probabilities are, with $\Delta^2\beta_{3j} \equiv \Delta\beta_{3j} - \Delta\beta_{2j}$, $j = 0, 1 \dots C$,

$$P(Y = c | Q, S, A, H, W_0, \dots, W_C) = \frac{\exp\{\Delta\beta_{2c} + \Delta^2\beta_{3c}S + \Delta\beta_{qc}Q + \Delta\beta_{dc}D + \Delta\beta'_{ac}A + \Delta\beta'_{hc}H - \beta'_{w0}W_0 + \beta'_{wc}W_c\}}{1 + \sum_{j=1}^C \exp(\Delta\beta_{2j} + \Delta^2\beta_{3j}S + \Delta\beta_{qj}Q + \Delta\beta_{dj}D + \Delta\beta'_{aj}A + \Delta\beta'_{hj}H - \beta'_{w0}W_0 + \beta'_{wj}W_j)}.$$

The numerator is 1 for $c = 0$, because all differences are relative to $c = 0$.

Because D alters the choice probability for class c by β_{dc} , the “net increase” in the propensity to choose class c relative to the base class 0 is $\Delta\beta_{dc} \equiv \beta_{dc} - \beta_{d0}$, not β_{dc} . Estimate $\Delta\beta_{d1}, \dots, \Delta\beta_{dC}$ with cross-section multinomial logit using the last display. Then, $\exp(\Delta\beta_{dc}) - 1$ is the class- c proportional odds effect relative to the class 0, and the class- c proportional effect as well when $Y = c$ is a rare event in the sense of (4.3).

A.2.3 Simple Simulation Study for Multinomial Outcome

Table A5 presents the results from a simulation study for multinomial outcome.

Our simulation study using the above $P(Y = c | Q, S, A, H, W_0, \dots, W_C)$ with $C = 2$ has the following design (the error terms generated as in (A.4), and H_{it} excluded):

$$\begin{aligned} A &\sim \text{Uniform}(-1, 1), \quad W_{2c}, W_{3c} \quad \text{for } c = 0, 1, 2 \text{ are iid } N(0, 1), \quad P(Q = 1) = 0.5, \\ P(S = 1) &= 0.5, \quad \beta_{20} = \beta_{30} = \beta_{q0} = \beta_{d0} = \beta_{a0} = \beta_{w0} = 0 \quad (\text{for class 0}), \\ \beta_{21} = \beta_{22} &= -4, \quad \beta_{31} = \beta_{32} = -5, \quad \beta_{q1} = \beta_{q2} = -0.5 \quad (\text{for classes 1, 2}), \\ \beta_{d1} = \beta_{d2} &= 0.5, \quad \beta_{a1} = \beta_{a2} = 0.5, \quad \beta_{w1} = \beta_{w2} = 0.5 \quad (\text{for classes 1, 2}). \end{aligned}$$

That is, the class-0 parameters are all zero, and the parameters of classes 1 and 2 are the same. Due to $\beta_{20} = \beta_{30} = 0$ but $\beta_{21} = \beta_{22} = -4$ and $\beta_{31} = \beta_{32} = -5$ (much smaller intercepts for classes 1 and 2 relative to class 0), the events $Y = 1, 2$ are rare.

Table A5: Multinomial Y : 3 classes, $N = 10,000$, 5,000 repetitions.

	Class $c = 1$			Class $c = 2$		
	True, Bias	SD, RMSE	AvgSD	True, Bias	SD, RMSE	AvgSD
$\Delta\beta_{2c}$	-4.0, 0.019	0.15, 0.023	0.15	-4.0, 0.030	0.15, 0.024	0.15
$\Delta^2\beta_{3c}$	-1.0, 0.017	0.28, 0.078	0.28	-1.0, 0.011	0.28, 0.077	0.28
$\Delta\beta_{qc}$	-0.5, 0.004	0.23, 0.055	0.23	-0.5, 0.006	0.23, 0.054	0.23
$\Delta\beta_{dc}$	0.5, 0.004	0.41, 0.166	0.41	0.5, 0.007	0.41, 0.168	0.41
$\Delta\beta_{ac}$	0.5, 0.003	0.16, 0.026	0.17	0.5, 0.005	0.17, 0.027	0.17
β_{w0}	0.0, 0.001	0.07, 0.004	0.07			
β_{wc}	0.5, 0.000	0.09, 0.008	0.10	0.5, 0.001	0.09, 0.008	0.10

AvgSD is the average of the asymptotic SD estimates.

Table A5 presents the simulation results, where each entry consists of true values (True), |Bias|, SD, RMSE, and the average of the asymptotic SD estimates (AvgSD). Overall, biases are very small, and AvgSD's are almost the same as the (simulation) SD's. The multinomial logit with RCS works well even for rare events $Y = 1, 2$.

References

- Ai, Chunrong, and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters* 80 (1): 123–9.
- Aminian, Ali, Abbas Al-Kurd, Rickesha Wilson, James Bena, Hana Fayazzadeh, Tavankit Singh, et al. 2021. "Association of Bariatric Surgery with Major Adverse Liver and Cardiovascular Outcomes in Patients with Biopsy-Proven Nonalcoholic Steatohepatitis." *JAMA* 326 (20): 2031–42.
- Anderson, Timothy S., Shoshana J. Herzig, Bocheng Jing, W. John Boscardin, Kathy Fung, Edward R. Marcantonio, et al. 2023. "Clinical Outcomes of Intensive Inpatient Blood Pressure Management in Hospitalized Older Adults." *JAMA Internal Medicine* 183 (7): 715–23.
- Angrist, Joshua D. 1998. "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants." *Econometrica* 66 (2): 249–88.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Ariella, Kahn-Lang, and Kevin Lang. 2020. "The Promise and Pitfalls of Differences-in-Differences: Reflections on *16 and Pregnant* and Other Applications." *Journal of Business & Economic Statistics* 38 (3): 613–20.
- Athey, Susan, and Guido W. Imbens. 2006. "Identification and Inference in Nonlinear Difference-in-Differences Models." *Econometrica* 74 (2): 431–97.
- Athey, Susan, and Guido W. Imbens. 2022. "Design-Based Analysis in Difference-in-Differences Settings with Staggered Adoption." *Journal of Econometrics* 226 (1): 62–79.
- Barbaresco, Silvia, Charles J. Courtemanche, and Yanling Qi. 2015. "Impacts of the Affordable Care Act Dependent Coverage Provision on Health-Related Outcomes of Young Adults." *Journal of Health Economics* 40: 54–68.
- Blundell, Richard, and Monica Costa Dias. 2009. "Alternative Approaches to Evaluation in Empirical Microeconomics." *Journal of Human Resources* 44 (3): 565–640.
- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *The Review of Economic Studies* 91 (6): 3253–85.
- Callaway, Brantly. 2023. "Difference-in-Differences for Policy Evaluation." In *Handbook of Labor, Human Resources and Population Economics*, edited by K. F. Zimmermann, 1–61. Springer-Nature.
- Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–30.
- Cataife, Guido, and Monica B. Pagano. 2017. "Difference in Difference: Simple Tool, Accurate Results, Causal Effects." *Transfusion* 57 (5): 1113–4.
- Cheng, Chao, Fan Li, Laine E. Thomas, and Fan Li. 2022. "Addressing Extreme Propensity Scores in Estimating Counterfactual Survival Functions via the Overlap Weights." *American Journal of Epidemiology* 191 (6): 1140–51.
- Choi, Jin-Young, and Myoung-Jae Lee. 2023. "Overlap Weight and Propensity Score Residual for Heterogeneous Effects: A Review with Extensions." *Journal of Statistical Planning and Inference* 222: 22–37.
- Choi, Jin-Young, Goeun Lee, and Myoung-Jae Lee. 2023. "Endogenous Treatment Effect for Any Response Conditional on Control Propensity Score." *Statistics & Probability Letters* 196: 109747.
- Ciani, Emanuele, and Paul Fisher. 2019. "Dif-in-Dif Estimators of Multiplicative Treatment Effects." *Journal of Econometric Methods* 8 (1): 20160011.
- De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *The American Economic Review* 110 (9): 2964–96.
- De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2023. "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey." *The Econometrics Journal* 26 (3): C1–C30.
- Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo. 2018. "The Economic Consequences of Hospital Admissions." *The American Economic Review* 108 (2): 308–52.
- Dukes, Oliver, and Stijn Vansteelandt. 2018. "A Note on G-estimation of Causal Risk Ratios." *American Journal of Epidemiology* 187 (5): 1079–84.
- Freeman, Rain E., Quynh A. Nguyen, Sandy Haaf, Jason L. Salemi, and Catherine M. Bulka. 2025. "Florida Red Tides and Adverse Birth Outcomes: Investigating Harmful Algae as a Novel Environmental Exposure." *Chemosphere* 385: 144536.
- Friedberg, L. 1998. "Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data." *The American Economic Review* 88 (3): 608–27.

- Goodman-Bacon, Andrew. 2018. "Public Insurance and Mortality: Evidence from Medicaid Implementation." *Journal of Political Economy* 126 (1): 216–62.
- Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–77.
- Hwang, Hyeonjun, and Myoung-Jae Lee. 2020. "A Simple Makeover Can Increase Bus Ridership: The Story of Tayo Bus." *Transport Policy* 97: 103–12.
- Jena, Anupam B., Dana P. Goldman, and Seth A. Seabury. 2015. "Incidence of Sexually Transmitted Infections After Human Papillomavirus Vaccination Among Adolescent Females." *JAMA Internal Medicine* 175 (4): 617–23.
- Kim, Bora. 2025. Forthcoming. "Estimating Spillover Effects in the Presence of Isolated Nodes." *Spatial Economic Analysis*. <https://doi.org/10.1080/17421772.2025.2523982>.
- Kim, Young-Sook, and Myoung-Jae Lee. 2017. "Ordinal Response Generalized Difference-in-Differences with Varying Categories: The Health Effect of a Disability Program in Korea." *Health Economics* 26 (9): 1121–31.
- Kim, Bora, and Myoung-Jae Lee. 2024. "Instrument-Residual Estimator for Multi-Valued Instruments under Full Monotonicity." *Statistics & Probability Letters* 213: 110187.
- Kim, Bora, and Myoung-Jae Lee. 2025. "Overlap-Weighted Difference-in-Differences: A Simple Way to Overcome Poor Propensity Score Overlap." *Economics Letters* 250: 112301.
- Lechner, Michael. 2011. "The Estimation of Causal Effects by Difference-in-Difference Methods." *Foundations and Trends in Econometrics* 4 (3): 165–224.
- Lee, Myoung-Jae. 2005. *Micro-Econometrics for Policy, Program, and Treatment Effects*. Oxford University Press.
- Lee, Myoung-Jae. 2015. "Panel Conditional and Multinomial Logit Estimators." In *The Oxford Handbook of Panel Data*, edited by B. Baltagi, 202–32. Oxford University Press.
- Lee, Myoung-Jae. 2016a. *Matching, Regression Discontinuity, Difference in Differences, and Beyond*. Oxford University Press.
- Lee, Myoung-Jae. 2016b. "Generalized Difference in Differences with Panel Data and Least Squares Estimator." *Sociological Methods & Research* 45 (1): 134–57.
- Lee, Myoung-Jae. 2018. "Simple Least Squares Estimator for Treatment Effects Using Propensity Score Residuals." *Biometrika* 105 (1): 149–64.
- Lee, Myoung-Jae. 2021. "Instrument Residual Estimator for Any Response Variable with Endogenous Binary Treatment." *Journal of the Royal Statistical Society (Series B)* 83 (3): 612–35.
- Lee, Myoung-Jae. 2024. "Direct, Indirect and Interaction Effects Based on Principal Stratification with a Binary Mediator." *Journal of Causal Inference* 12 (1): 20230025.
- Lee, Myoung-Jae, and Chirok Han. 2024. "Ordinary Least Squares and Instrumental Variable Estimators for Any Outcome and Heterogeneity." *STATA Journal* 24 (1): 72–92.
- Lee, Myoung-Jae, and Young-Sook Kim. 2014. "Difference in Differences for Stayers with a Time-Varying Qualification: Health Expenditure Elasticity of the Elderly." *Health Economics* 23 (9): 1134–45.
- Lee, Myoung-Jae, and S. Kobayashi. 2001. "Proportional Treatment Effects for Count Response Panel Data: Effects of Binary Exercise on Health Care Demand." *Health Economics* 10 (5): 411–28.
- Lee, Myoung-Jae, and Yasuyuki Sawada. 2020. "Review on Difference in Differences." *Korean Economic Review* 36 (1): 135–73.
- Lee, Goeun., Jin-Young Choi, and Myoung-Jae Lee. 2023. "Minimally Capturing Heterogeneous Complier Effect of Endogenous Treatment for Any Outcome Variable." *Journal of Causal Inference* 11 (1): 20220036.
- Lee, Myoung-Jae, Goeun Lee, and Jin-Young Choi. 2025. "Linear Probability Model Revisited: Why it Works and How it Should be Specified." *Sociological Methods & Research* 54 (1): 173–86.
- Li, Fan. 2019. "Propensity Score Weighting for Causal Inference with Multiple Treatments." *Annals of Applied Statistics* 13 (4): 2389–415.
- Li, Fan, Kari Lock Morgan, and Alan M. Zaslavsky. 2018. "Balancing Covariates via Propensity Score Weighting." *Journal of the American Statistical Association* 113 (521): 390–400.
- Liu, Licheng, Ye Wang, and Yiqing Xu. 2024. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." *American Journal of Political Science* 68 (1): 160–76.
- Mao, Huzhang, Liang Li, and Tom Greene. 2019. "Propensity Score Weighting Analysis and Treatment Effect Discovery." *Statistical Methods in Medical Research* 28 (8): 2439–54.
- Markowitz, Sara, and Andrew J. D. Smith. 2024. "Nurse Practitioner Scope of Practice and Patient Harm: Evidence from Medical Malpractice Payouts and Adverse Action Reports." *Journal of Policy Analysis and Management* 43 (2): 420–45.
- Matsouaka, Roland A., and Yunji Zhou. 2024. "Causal Inference in the Absence of Positivity: The Role of Overlap Weights." *Biometrical Journal* 66 (4): 2300156.
- McGrath, Susan P., Irina M. Perreard, Melissa D. Garland, Kelli A. Converse, and Todd A. Mackenzie. 2019. "Improving Patient Safety and Clinician Workflow in the General Care Setting with Enhanced Surveillance Monitoring." *IEEE Journal of Biomedical and Health Informatics* 23 (2): 857–66.
- McMichael, Benjamin J. 2023. "Supply-Side Health Policy: The Impact of Scope-of-Practice Laws on Mortality." *Journal of Public Economics* 222: 104901.
- Nelder, John Ashworth, and R. W. M. Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society (Series A)* 135 (3): 370–84.

- Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables with an Application to 401 (k) Plan Participation Rates." *Journal of Applied Econometrics* 11 (6): 619–32.
- Puhani, Patrick A. 2012. "The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear 'Difference in Differences' Models." *Economics Letters* 115 (1): 85–7.
- Roth, Jonathan, and Pedro H. C. Sant'Anna. 2023. "When is Parallel Trends Sensitive to Functional Form?" *Econometrica* 91 (2): 737–47.
- Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics* 235 (2): 2218–44.
- Santos Silva, J. M. C., and Silvana Tenreiro. 2006. "The Log of Gravity." *The Review of Economics and Statistics* 88 (4): 641–58.
- Santos Silva, J. M. C., and Silvana Tenreiro. 2011. "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator." *Economics Letters* 112 (2): 220–2.
- Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (2): 175–99.
- Taddeo, Marcelo M., Leila D. Amorim, and Rosana Aquino. 2022. "Causal Measures Using Generalized Difference-in-Difference Approach with Nonlinear Models." *Statistics and Its Interface* 15 (4): 399–413.
- Tchetgen, Eric J. Tchetgen, Chan Park, and David B. Richardson. 2024. "Universal Difference-in-Differences for Causal Inference in Epidemiology." *Epidemiology* 35 (1): 16–22.
- Thomas, Laine E., Fan Li, and Michael J. Pencina. 2020. "Overlap Weighting: A Propensity Score Method that Mimics Attributes of a Randomized Clinical Trial." *Journal of the American Medical Association* 323 (23): 2417–8.
- Wei, Jie, Hyon K. Choi, Nicola Dalbeth, Xiaoxiao Li, Changjun Li, Chao Zeng, et al. 2023. "Gout Flares and Mortality After Sodium-Glucose Cotransporter-2 Inhibitor Treatment for Gout and Type 2 Diabetes." *JAMA Network Open* 6 (8): e2330885.
- Wolfers, Justin. 2006. "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results." *The American Economic Review* 96 (5): 1802–20.
- Wooldridge, Jeffrey M. 2021. *Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators*. Available at SSRN 3906345.
- Wooldridge, Jeffrey M. 2023. "Simple Approaches to Nonlinear Difference-in-Differences with Panel Data." *The Econometrics Journal* 26 (3): C31–C66.
- Xu, Shenbo, Bang Zheng, Bowen Su, Stan Neil Finkelsten, Roy Welsch, Kenney Ng, et al. 2025. "Can Metformin Prevent Cancer Relative to Sulfonyleureas? A Target Trial Emulation Accounting for Competing Risks and Poor Overlap via Double/Debiased Machine Learning Estimators." *American Journal of Epidemiology* 194 (2): 512–23.
- Yadlowsky, Steve, Fabio Pellegrini, Federica Lionetto, Stefan Braune, and Lu Tian. 2021. "Estimation and Validation of Ratio-based Conditional Average Treatment Effects Using Observational Data." *Journal of the American Statistical Association* 116 (533): 335–52.
- Zhou, Yunji, Roland A. Matsouaka, and Laine Thomas. 2020. "Propensity Score Weighting Under Limited Overlap and Model Misspecification." *Statistical Methods in Medical Research* 29 (12): 3721–56.