

Can't See the Forest for the Logs: On the Perils of Using Difference-in-Differences With a Log-Dependent Variable*

Brendon McConnell

February 14, 2024

Abstract

With a dependent variable in logs, the difference-in-differences (DD) term does not capture the outcome difference between treated and untreated groups over time. Rather it reflects an approximation of the proportional difference in growth rates across groups. As I show with both simulations and three published case studies, if the baseline outcome distributions are sufficiently different across groups, the DD parameter for a log-specification can be different in sign to that of a levels-specification. I provide a condition, based on (i) the aggregate time effect, and (ii) the difference in relative baseline outcome means, for when the sign-switch will occur.

Keywords— Difference-in-Differences, Functional Form.

JEL Codes— C01.

*I benefited a great deal from discussions about this work with Jaime Millán-Quijano. I am also very grateful to Marc F. Bellemare and James Rockey for detailed comments on an early draft, and to Giovanni Mastrobuoni and Richard Murphy for fruitful discussions and comments. McConnell: City, University of London, brendon.mcconnell@gmail.com.

1 Introduction

Difference-in-differences (DD) is almost certainly the most popular quasi-experimental research design currently used in a broad range of empirical settings. Its use extends beyond Economics into Political Science, Social Medicine and other fields. The seeming simplicity of the design has, at least until recently, played a large role in its popularity. A recent literature documenting underlying issues with difference-in-differences, particularly in the case of staggered roll-out of treatment implementation (Goodman-Bacon, 2021), has somewhat shattered the illusion of the simplicity of this research design.

In this paper, I highlight an additional complication that the applied researcher faces when operationalizing a DD design – the choice of functional form, and the subsequent consequences of this choice. As Ciani and Fisher (2019) note in related work, researchers may apply the log transformation, not because they believe the true underlying data generating process is multiplicative rather than additive, but rather due to concerns of skewness, or because the log transformation enables one to interpret the effect of controls in percentage terms, e.g., “I take the logarithm of the [outcome variable] to deal with this skewed distribution and to allow for an intuitive interpretation of results in percentage terms” (Fouka, 2019, p. 124).¹

In order to set the scene for this work, I survey all papers published in *The Quarterly Journal of Economics (QJE)* from 2018 to 2022. I present summary information on these articles in Table B1. Of the 49 *QJE* articles published in this five year span that use a DD design, almost all of these (46 articles) consider at least one continuous outcome variable. Roughly two thirds of these 46 articles (30 articles or 65%) impose a log transformation on at least one continuous variable, and three more impose an inverse hyperbolic sine transformation², which will lead to related issues.³ That over 65% of all DD papers in one of the top Economics journals use at least one log-dependent variable specification underscores the importance of better understanding what we recover using a DD design with a log transformed dependent variable.

I start by outlining the key differences between a DD model with level- and log-dependent variables. A levels specification implies an additive model for potential outcomes. In this case the common trend is additive – outcomes increase by a given level over time – as is the treatment effect. In the level-dependent variables case, the DD parameter returns the difference between the treated and untreated groups in changes in the outcome variable over time. In contrast, a log-specification implies a multiplicative model for potential outcomes, which in turn implies a multiplicative common trend – outcomes grow by a given proportion over time – and a multiplicative treatment effect. Accordingly, when one uses a log transformation of the dependent variable, the DD parameter approximates the *proportional difference in growth rates* between treated and untreated groups.⁴

¹Other papers are more explicit about the reasons and/or consequences of the choice between level- and log-specifications of the outcome variable when using difference-in-differences designs (Finkelstein, 2007; Powell and Leider, 2020; Park and Powell, 2021).

²See Bellemare and Wichman (2020) for an in-depth consideration of the implications of the inverse hyperbolic sine (IHS) transformation.

³The inverse hyperbolic sine of a random variable y , $\text{arsinh}(y) = \ln(y + \sqrt{y^2 + 1}) \approx \ln(2y)$ for large y .

⁴Denoting the DD parameter for the log specification as β_4 , the transformation $\exp(\beta_4) - 1$ yields the

I then shift my attention to the primary aim of this paper: to explicate the consequences of the functional form assumption when baseline means of group outcomes differ substantively. These consequences can be stark. For a given aggregate time effect, I show that with a sufficiently large difference between group outcome baselines means, it is possible for the level- and log-specifications of a DD design to yield DD parameter estimates of *different signs*. The choice of functional form in the DD context is by no means innocuous: one may conclude that a policy or intervention raised outcomes for the treatment group using one functional form, but may conclude the precise opposite using a different functional form, even with the same data, the same sample period, and the same control variables. To my knowledge, this is the first paper to methodically document this disparity. Given the wide use of DD designs for policy evaluation in areas that can give rise to such large differences in baseline outcome distributions – e.g., gender gaps in earnings, race gaps in the length of incarceration spells, house prices across regions or states, or school test scores across different education regimes – this point is likely to be of broad significance to applied researchers.

I provide a condition – based on (i) the common time effect experienced by both groups, and (ii) the difference in relative baseline outcome means – as to when we should expect a sign switch for the log-dependent variable case. This can straightforwardly be expressed in terms of the parameters of an additive DD model. Using simulations, I provide evidence of the sign disparity in estimated DD coefficients and verify the condition I propose via simulation.⁵

Next, I turn to three empirical case studies, all of which have been published in the *QJE* since 2020, and all of which feature a key regression specification involving a log-dependent variable and either a DD or event study design. Based on my subjective reading of these papers, the internally consistent underlying model for all three papers is an additive model, which implies the core specification should be in levels, not logs.

I note that there are two potential errors that researchers may make when specifying a log-dependent variable model in conjunction with a DD design. The first is that estimates based on a logarithmic specification can be seriously biased if estimated by OLS, as documented in the seminal work by Silva and Tenreyro (2006). I refer to this as *Jensen bias*, as due to Jensen's inequality, when one log-linearizes a multiplicative model, heteroskedasticity can lead to parameter bias, not just inefficiency (Silva and Tenreyro, 2006).

The second error, which I refer to as a *targeting error*, is that in many cases, researchers are interested in targeting an additive model – they care about the levels of the outcome, not the growth rate of the outcome – yet apply the log transformation to their dependent variable. This transformation implies a multiplicative model. This functional form assumption means that the DD parameter no longer returns a difference between treatment groups in differences in outcomes over time due to a policy, but rather a relative difference in growth rates of the outcome. Where baseline outcomes means across treatment are sufficiently different, this targeting error can lead to estimating a policy effect of a different sign to what one would estimate from an additive model.

precise proportional difference in growth rates.

⁵I focus my attention on the non-staggered timing, binary treatment case for this paper.

Based on these insights, I replicate the baseline log-dependent variable specifications of the three cases studies, and then compare these estimates to DD estimates based on both (i) a levels specification estimated by Poisson Pseudo Maximum Likelihood (PPML) and (ii) a levels specification estimated by OLS. The first extension assumes an underlying multiplicative model, but accounts for Jensen Bias. The second extension assumes an underlying additive model, and thereby estimates a difference over time in differences of outcomes across treatment groups, not a proportional difference in growth rates.⁶

Two of the articles that I consider as empirical case studies feature settings with disparate baseline outcome means. For these articles, I document a fragility in the findings one takes from estimating a DD specification with a log-dependent variable. This is primarily due to targeting error. The fragility of estimates that I document ranges from certain key estimates losing statistical significance at conventional levels, to documenting sign-switches in key DD estimates, to documenting completely different patterns of parameter estimates when addressing both Jensen bias and targeting error. Such findings underscore the empirical relevance of the issue at the core of this work.

The second key finding confirms the proposition I outline below, and highlights that researchers will only be confronted with the possibility of a sign-switch in DD estimates between an additive and a multiplicative model when baseline outcome means are sufficiently different. One of the case studies I consider implements a matched-DD design using panel data on workers, where the baseline outcomes are one of the variables targeted in the matching procedure. In this case there is a large degree of congruence between the DD estimates from the log- and level-dependent variable specifications.

The key conclusion I draw from both the simulation and empirical case study replication exercises is that *ambivalence* over the underlying model for the outcome variable is likely the cause of many researcher errors when using a DD design. A natural principle arises from the findings of this work: the applied researcher will ideally decide first on how to model parallel trends in potential outcomes in the setting one faces – in levels (i.e., an additive model) or in percentage terms (a multiplicative model). Finkelstein (2007) makes precisely such a reasoning when arguing for a multiplicative model: “a level specification would constrain the outcomes to grow by the same absolute amount each year, which would be inappropriate given the considerable variation in size across hospitals.” (Finkelstein, 2007, p. 9). If the researcher decides on an additive model, then the outcome variable should be specified in levels. If the researcher decides on a multiplicative model, then in order to avoid Jensen bias, it is advisable to use outcome in levels, but estimate the empirical model using PPML. In both cases, a DD design using a log-dependent variable and estimated by OLS is the inferior estimation strategy.

This principle stands in opposition to what often occurs in practice, where one first empirically examines whether or not parallel trends hold in logs or levels, and then proceeds with this functional form specification, without a clear, *a priori* consideration of how best to model parallel trends in potential outcomes in the specific setting. Starting by empirically checking the parallel trend for various functional forms is akin to putting the cart before the horse.

⁶I detail two further empirical examples of my own in the Appendix.

This paper contributes to the difference-in-differences literature by making clear the consequences of functional form assumptions in DD designs, particularly when working with groups with large baseline differences in outcome distributions. This builds on work by both Meyer (1995) and Kahn-Lang and Lang (2020) who noted that functional form assumptions would matter in such cases. This work also relates to the recent work by Roth and Sant'Anna (2023), who set out the conditions under which the parallel trend assumption is insensitive to functional form.

The secondary contribution of this paper is to serve as an episodic reminder of the impact of Jensen bias when estimating a log-linearized multiplicative model. In two of the three case studies I consider below, Jensen bias has a non-negligible impact on estimated coefficients. The initial warning regarding Jensen bias was made by Silva and Tenreyro (2006) in the context of estimating gravity models, but the issue extends far beyond empirical trade papers.

The remainder of the paper is organized as follows. Section 2 provides an overview of the DD model, makes precise what the level- and log-specifications are estimating and provides a condition for when we will find a sign-switch between the two specifications. Section 3 provides simulation evidence to both verify the sign-switch condition, and to highlight cases that generate disparate coefficient estimates across specifications. Section 4 presents three empirical case studies and Section 5 concludes.

2 The DD Model

2.1 The Nature of the Problem

Individual i can belong to treatment group ($D_i = 1$) or untreated comparison group ($D_i = 0$). We observe individuals in two periods – $T_t = 0$ and $T_t = 1$. Those in treatment group receive treatment in period 1. Using a potential outcomes framework, we can write down the realized outcome as $Y_{it} = (1 - D_i)Y_{it}(0) + D_iY_{it}(1)$, where $Y_{it}(0)$ and $Y_{it}(1)$ are the potential outcomes for individual i in absence of treatment and upon receipt of treatment respectively. We write the DD estimand as:

$$\begin{aligned}\alpha^{ATT} = & \{E[Y_{it} | D_i = 1, T_t = 1] - E[Y_{it} | D_i = 1, T_t = 0]\} \\ & - [E[Y_{it} | D_i = 0, T_t = 1] - E[Y_{it} | D_i = 0, T_t = 0]]\end{aligned}\quad (1)$$

The sample analog of (1) is the DD estimator

$$\hat{\alpha}^{DD} = [\bar{Y}_{T1} - \bar{Y}_{T0}] - [\bar{Y}_{C1} - \bar{Y}_{C0}] , \quad (2)$$

where the subscript T and C refer to treatment and control groups respectively. The subscripts 0 and 1 respectively refer to the pre- ($T_t = 0$) and post-policy ($T_t = 1$) periods. A simple regression specification we can use to estimate the ATT parameter is:

$$Y_{it} = \alpha_1 + \alpha_2 Treat_i + \alpha_3 Post_t + \alpha_4 (Treat_i \times Post_t) + \epsilon_{it} , \quad (3)$$

where $Treat_i$ is a treatment indicator, $Post_t$ the post-period indicator, and α_4 is the parameter of interest. Writing down an analogous specification with a log-transformed dependent variable of the form:

$$\ln Y_{it} = \beta_1 + \beta_2 Treat_i + \beta_3 Post_t + \beta_4 (Treat_i \times Post_t) + \mu_{it}, \quad (4)$$

implies an underlying model for Y_{it} that is multiplicative rather than additive:

$$Y_{it} = \exp(\beta_1 + \beta_2 Treat_i + \beta_3 Post_t + \beta_4 (Treat_i \times Post_t)) \eta_{it}, \quad (5)$$

where $\mu_{it} = \ln \eta_{it}$. The assumption of an underlying multiplicative model means that one is assuming that parallel trends hold multiplicatively, not additively. Put differently, in absence of the treatment, the outcome for both treatment and control groups would *grow by the same proportion*. This is very different to the analog for an additive model, where in absence of the treatment, both groups increase in levels by the same amount (Ciani and Fisher, 2019).

As noted by Mullahy (1999), we can write an expression for the exponentiated DD parameter, based on the multiplicative model, as:

$$\exp(\beta_4) = \frac{\frac{E[Y_{it} | D_i = 1, T_t = 1]}{E[Y_{it} | D_i = 1, T_t = 0]}}{\frac{E[Y_{it} | D_i = 0, T_t = 1]}{E[Y_{it} | D_i = 0, T_t = 0]}} = \frac{g_T}{g_C}, \quad (6)$$

where g_C and g_T are the respective growth rates in the outcome for control and treated groups.

Recalling that $\ln(1+z) \approx z$ for small values of z , and returning to Equation (4), we see that the DD parameter we estimate (β_4) with a log-dependent variable can be expressed as:

$$\beta_4 \approx \exp(\beta_4) - 1 = \frac{\frac{E[Y_{it} | D_i = 1, T_t = 1]}{E[Y_{it} | D_i = 1, T_t = 0]} - \frac{E[Y_{it} | D_i = 0, T_t = 1]}{E[Y_{it} | D_i = 0, T_t = 0]}}{\frac{E[Y_{it} | D_i = 0, T_t = 1]}{E[Y_{it} | D_i = 0, T_t = 0]}} = \frac{g_T - g_C}{g_C}. \quad (7)$$

Equation (7) makes clear that when we estimate a DD specification with a log-dependent variable (as in Equation (4)), we are estimating an approximation of the *proportional difference in growth rates* of the outcome between the treated and untreated groups over the two periods. This is very different from what we measure with a level dependent variable – the difference between groups in changes over time.

In Proposition 1 below, I outline the conditions under which we will find a sign switch using a level and log specification.

Proposition 1. *For a given (non-zero) aggregate time effect ($\alpha_3 \neq 0$), if outcomes means at baseline are sufficiently different in relative terms, the functional form decision of specifying a level- or log-dependent variable can yield a sign difference for the DD parameter. More specifically:*

when $0 < |\Delta_T - \Delta_C| < \left| \Delta_C \frac{(E[Y_{T0}] - E[Y_{C0}])}{E[Y_{C0}]} \right|$, we will have $\text{sign}(\alpha_4) \neq \text{sign}(\beta_4)$.

This condition may alternatively be expressed in terms of the parameters of the additive DD model as:

when $|\alpha_4| < \left| \alpha_3 \frac{\alpha_2}{\alpha_1} \right|$, we will have $\text{sign}(\alpha_4) \neq \text{sign}(\beta_4)$.

Proof in Appendix A.

The Proposition above uses the notation $E[Y_{C0}] = E[Y_{it} | D_i = 0, T_t = 0]$, $E[Y_{T0}] = E[Y_{it} | D_i = 1, T_t = 0]$, $\Delta_C = E[Y_{it} | D_i = 0, T_t = 1] - E[Y_{it} | D_i = 0, T_t = 0]$, and $\Delta_T = E[Y_{it} | D_i = 1, T_t = 1] - E[Y_{it} | D_i = 1, T_t = 0]$. In the simulation approach I detail in Section 3, I verify Proposition 1.

2.1.1 The Parallel Trends Assumption

The key identifying assumption for the DD estimator to return the ATT is the parallel trends assumption. In recent work, Roth and Sant'Anna (2023) set out the conditions under which the parallel trend is insensitive to functional form. These conditions – either random assignment of treatment, stationarity of the distribution of potential outcomes for the untreated setting, $Y(0)$, or a combination of these two cases – are considerably stricter than is typically assumed by empirical researchers using DD approaches. When such conditions are not met, one must choose, and consequently justify, a functional form for the outcome variable – a parallel trend in levels obviates a parallel trend also holding in logs, and vice versa. This point was made by Meyer (1995), and has since been reiterated by several authors, including Angrist and Pischke (2009) and Kahn-Lang and Lang (2020).

In one sense, these findings regarding functional form and the parallel trend assumption limit the scope of cases that may benefit from the insights of this paper. If (i) a parallel trend in levels precludes a parallel trend in logs and (ii) empirical researchers typically assess the validity of using DD methods by some form of pre-trend inspection and/or testing⁷, then does the singular aim of this paper – to call attention to the possibility of a sign switch in DD estimates based on functional form assumptions – carry any weight? I argue that it does, both in cases that satisfy the conditions set out by Roth and Sant'Anna (2023), or in cases where limited pre-policy data limits the statistical power to detect divergent pre-trends. Having a clearer sense of when such issues may arise will hopefully be of use to applied researchers.

2.2 Two Solutions

The underlying source of the sign-switch issue highlighted above is a disparity in baseline outcome means across treatment groups. If one believes the underlying model is additive but wants the convenience of being able to interpret estimates in proportional terms, then the simplest

⁷See Roth (2022) for a discussion on the pitfalls of such pre-testing

solution to this issue is to estimate a DD model with a dependent variable in levels, and then express the DD parameter relative to the baseline mean for the control group.

For the researcher who is both wedded to the use of the log transformation – possibly due to skewness concerns – and ambivalent towards the underlying model they wish to assume for their outcome variable, our understanding of the source of the issue presents a potential solution: to ensure that there is no disparity in baseline outcome means. This can be done by matching methods, e.g., see the approach taken by Aneja and Xu (2022), which I discuss in Section 4.3. This could also be done by mechanically recentering the outcome distribution of the treated group in order to align baseline outcome means prior to applying the log transformation. I describe this approach in Appendix C.1. I am not an especial fan of this approach – I find it fiddly and inelegant, but I present it here for completeness.

3 Simulation Results

In this section, I provide simulation results to show that, using an additive and multiplicative model, one may estimate an ATT that differs not just in magnitude, but in sign. I take the additive model as the data generating process (DGP) in this case, and compare level- and log-dependent variable based DD specifications. The DGP is based on Equation (3), with different simulation specifications using different parameters values for α_1 , α_2 , α_3 , and α_4 .

The simulation results below illustrate the key insight of this paper –because the DD parameter from a level- and log-dependent variable specification respectively reflect a difference in differences in levels and a proportional difference in growth rates, holding fixed the time effect, one just needs to shift the distribution of outcomes for one of the groups to drive a wedge between the resulting parameter estimates.

Table 1 presents the first set of simulation results. All DD estimates for the level-dependent variables are positive, whereas the estimates for the log-dependent variable specifications are either precisely zero, or negative. For log specifications, I present both β_4 , and $\exp(\beta_4) - 1$, of which the latter equals the proportional difference in growth rates (see Equation (7)). At the base of the table, I present the cell means for the outcome variable across the four DD groups – thus one can easily see how the parameters are generated – as well as the proportional growth rates that I calculate from these cell means.

The key point of this table is to show that one can generate a positive DD estimate from a levels specification and, by merely shifting the baseline outcome distribution of the untreated group, also generate a precise zero or a negative DD estimate from a log specification. In Table C1 I present a complementary set of simulation results, where I fix the level specification to yield a DD estimate that is precisely zero (once again, the cells sample averages at the base of the table provide the key insight into how this is operationalized) for a levels specification, but which yields either a negative or positive DD estimate from a log specification.

Table 1: Simulating a Positive Levels DD Effect That Yields a Zero or Negative Log Effect

	Zero Log Effect			Negative Log Effect	
	(1)	(2)	(3)	(4)	(5)
a.) Level-Dependent Variable:					
DD Estimate $[\hat{\alpha}_4]$	0.40*** (0.00)	1.00*** (0.00)	0.40*** (0.00)	1.00*** (0.00)	0.80*** (0.00)
Recenter					
DD Estimate $[\hat{\alpha}_4^{RC}]$	0.40*** (0.00)	1.00*** (0.00)	0.40*** (0.00)	1.00*** (0.00)	0.80*** (0.00)
b.) Log-Dependent Variable:					
DD Estimate $[\hat{\beta}_4]$	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.043*** (0.000)	-0.017*** (0.001)
$\exp(\hat{\beta}_4) - 1$	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.042*** (0.000)	-0.017*** (0.001)
Recenter					
$\exp(\hat{\beta}_4^{RC}) - 1$	0.033*** (0.000)	0.067*** (0.000)	0.018*** (0.000)	0.083*** (0.000)	0.133*** (0.001)
Recenter-Then-Rescale					
$(\exp(\hat{\beta}_4^{RC}) - 1) \times (\bar{Y}_{C1} / \bar{Y}_{C0})$	0.040*** (0.000)	0.100*** (0.000)	0.020*** (0.000)	0.100*** (0.000)	0.160*** (0.001)
$\hat{\alpha}_4 / \bar{Y}_{C0}$	0.040	0.100	0.020	0.100	0.160
$(g_T - g_C) / g_C$	-0.000	0.000	-0.000	-0.042	-0.017
\bar{Y}_{C0}	10.0	10.0	20.0	10.0	5.0
\bar{Y}_{C1}	12.0	15.0	22.0	12.0	6.0
\bar{Y}_{T0}	12.0	12.0	24.0	20.0	10.0
\bar{Y}_{T1}	14.4	18.0	26.4	23.0	11.8

Notes: Results based on 1000 simulation runs. *** denotes significance at 1%, ** at 5%, and * at 10%. DD estimates for both a level- and log-dependent variable are presented in each column. The tables display the mean and (in parentheses) bootstrapped standard error of the DD estimates across all simulation runs. At the base of the table, the four elements of the DD are presented for reference. The sample size is 40000 in each simulation. Proportion treated and proportion in the post period are .5 and .5 respectively. The standard deviation of the error term is .2 in all simulations.

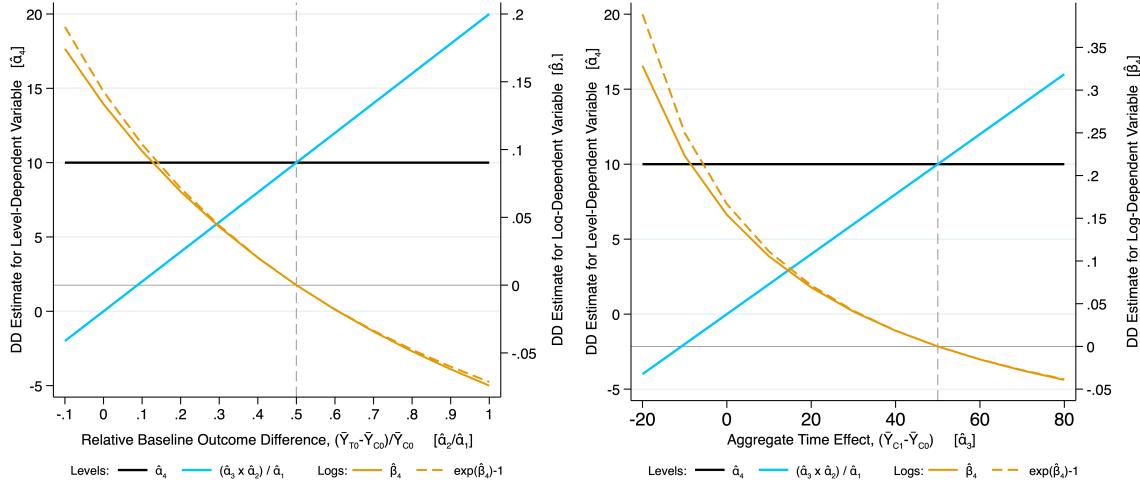
The secondary point of the table is to show that the RTR approach, clunky as it is, is a viable method to produce treatment effect estimates from a log-dependent variable specification that coincide with the analogous treatment effects from a level-dependent variable specification (as a proportion of the control baseline outcome mean). To see this note that for all five columns, the exponentiated RTR estimate equals the proportional treatment effect from the additive model $(\hat{\alpha}_4 / \bar{Y}_{C0})$.

In Figure 1, I present DD estimates from both a levels- and log-based specification for two separate simulation experiments.⁸ Based on the parameters I choose, the DD coefficient estimate is a constant equal to 10 for the levels specification. Keeping all parameters fixed except for either the relative baseline outcome difference (Figure 1a) or the aggregate time effect (Figure 1b), I verify Proposition 1. The figures make clear two key points. First, keeping the levels-based DD estimate fixed at a constant value and altering either the relative difference in baseline means (α_2 / α_1) or the aggregate time effect (α_3), it is possible to generate either a negative or a positive DD estimate for the log-dependent variable specification. Second, the sign switch occurs at precisely the point stated in Proposition 1 – when $\alpha_4 = \alpha_3 \alpha_2 / \alpha_1$.

The condition provided in Proposition 1 indicates that we require both a non-zero aggregate time effect ($\alpha_3 \neq 0$) and a difference in baseline outcome means ($\alpha_2 \neq 0$) to find a sign switch

⁸I present levels specifications using the left-hand y -axis and a black line. I present log specifications using the right-hand y -axis and orange lines.

Figure 1: The Sign Switch Occurs When $\alpha_4 = \alpha_3 \frac{\alpha_2}{\alpha_1}$



(a) Varying Baseline Outcome Differences

(b) Varying the Aggregate Time Effect

Notes: The underlying DGP that creates both graphs is an additive DD model specification, thus taking the form of Equation (3). In all cases, $E[\epsilon_{it} | D_i, T_t] = 0$ and $\sigma_\epsilon = .2$. In the left-hand graph, I set the following as fixed: $Y_{C0} = 50$, $Y_{T1} - Y_{T0} = 30$, $Y_{C1} - Y_{C0} = 20$. What varies is the relative baseline difference: $(Y_{T0} - Y_{C0})/Y_{C0} = \alpha_2/\alpha_1$. Once the relative baseline difference is fixed, this yields a specific value of Y_{T0} for the simulation exercise. In the right-hand graph, I set the following as fixed: $Y_{C0} = 50$, $Y_{T0} = 60$, $(Y_{T1} - Y_{T0}) = (Y_{C1} - Y_{C0}) + 10$. What varies is the aggregate time effect, $(Y_{C1} - Y_{C0})$. Once the aggregate time effect is fixed, this yields a specific value for both Y_{C1} and Y_{T1} for the simulation exercise. For each increment of relative baseline difference and aggregate time effect values respectively, 10,000 simulations are conducted. In both cases, the specification leads to a constant (levels) DD effect of 10 units, which is why the black line is horizontal. The sample size is 40,000 in each simulation. Proportion treated and proportion in the post period are .5 and .5 respectively.

across level- and log-specifications. In Figure C1 is to show that this is indeed the case. The underlying DGPs are identical to those underlying Figure 1, except for the key difference that in Figure C1a I set the aggregate time effect to zero, and in Figure C1b I set the baseline outcome mean wedge to zero. In both cases, the log-based DD estimate (shown in orange) never crosses the zero line i.e., there is no sign switch.

4 Empirical Case Studies

There are two potential errors that researchers may make when specifying a log-dependent variable model in conjunction with a DD design. The first is that since a logarithmic specification implies an underlying multiplicative model, and as documented in the seminal work by Silva and Tenreyro (2006), such estimates can be seriously biased if estimated by OLS. I refer to this as *Jensen bias*, as due to Jensen's inequality, when one log-linearizes a multiplicative model (e.g., Equation (5)) heteroskedasticity can lead to parameter bias, not just inefficiency (Silva and Tenreyro, 2006).

The second error, which I refer to as a *targeting error*, is that in many cases, researchers are

interested in targeting an additive model – they care about the levels of the outcome, not the growth rate of the outcome – yet apply the log transformation to their dependent variable. This transformation implies a multiplicative model for potential outcomes. Where baseline outcomes means across treatment are sufficiently different, this targeting error can lead to estimating a policy effect of a different sign to what one would estimate from an additive model.

In this section, I present three empirical case studies based on articles published in the *QJE* since 2020, and all of which feature a key regression specification involving a log-dependent variable and either a DD or event study design. Based on my subjective reading of these papers, the internally consistent underlying model for all three papers is an additive model, which implies the core specification should be in levels, not logs. I selected the first two papers – Derenoncourt and Montialoux (2021) and Hebllich et al. (2020) – as case studies based on large wedges in baseline outcome means. I chose the final paper – Aneja and Xu (2022) – as this paper employs a matched-DD model. The matching is based on coarsened exact matching (CEM), where the authors explicitly target the pre-treatment outcome mean along with other individual characteristics. Such targeting suggests that there should be minimal discrepancies across log- and level-dependent variable specifications for this final case.

For all three case studies I replicate the initial results (log-dependent variable, multiplicative model, estimated by OLS), and then provide two additional sets of results. First, to gauge the importance of Jensen bias, I estimate a level-dependent variable specification by PPML, which assumes a multiplicative model. I then estimate a level-dependent variable specification by OLS, assuming an additive model. This final set of results enables me to speak to the importance of targeting error.

4.1 Derenoncourt and Montialoux (2021)

Derenoncourt and Montialoux (2021) study the impact of a minimum wage expansion policy, comparing industries newly covered by a minimum wage policy (1967) with those previously treated (1938) as controls. The authors estimate a DD model of the form:

$$\log w_{ijst} = \alpha + \sum_{k=1}^2 \beta_k \text{Covered 1967}_j \times \mathbb{1}[t \in \text{period}_k] + \delta_j + \tau_t + \mathbb{X}'_{ijst} \Gamma + \varepsilon_{ijst}, \quad (8)$$

where w_{ijst} is the annual earnings for worker i in industry j , state s , and year t . The indicator variables period_1 and period_2 respectively cover the years 1967-1972 and 1973-1980.⁹

In panel (a) of Table 2 I start by replicating the specifications that yield column 1, Table I, and Table V.¹⁰ Based on the log-OLS specification, the authors conclude that the minimum

⁹The authors use CPS and BLS data. The replication evidence I provide here is based on the CPS data.

¹⁰Note that the estimation sample I use has one observation fewer than the corresponding sample of Derenoncourt and Montialoux (2021). Based on a top-coding procedure, there was an individual in 1961 with an annual earnings value of 107,571,072, whom I dropped.

Table 2: Derenoncourt and Montialoux (2021) Replication

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Baseline			Robustness			
	[1]		[2]		[3]	
	Black and White	Black	White	Black	White	Black
(a) $\ln(\text{Annual Earnings})$ – [Multiplicative Model] – OLS						
Covered in 1967 ×	.0658**	.0953***	.0542**	.0739***	.0515**	.074**
1967–1972 [$\hat{\beta}_1^a$]	(.0249)	(.0217)	(.0234)	(.0239)	(.0232)	(.0301)
Covered in 1967 ×	.0504	.078*	.0361	.0492	.0332	.0428
1973–1980 [$\hat{\beta}_1^a$]	(.042)	(.0371)	(.0417)	(.0391)	(.0412)	(.0429)
(b) Annual Earnings – [Multiplicative Model] – PPML						
Covered in 1967 ×	.056***	.0966***	.0486**	.0727***	.0465**	.0686***
1967–1972 [$\hat{\beta}_1^p$]	(.0205)	(.0212)	(.0202)	(.0211)	(.0199)	(.0208)
Covered in 1967 ×	.0291	.0465	.0214	.0151	.0189	.00931
1973–1980 [$\hat{\beta}_1^p$]	(.0453)	(.0311)	(.0471)	(.0355)	(.0466)	(.0371)
(c) Annual Earnings – [Additive Model] – OLS						
Covered in 1967 ×	563	1,503*	474	875	362	970*
1967–1972 [$\hat{\beta}_1^c$]	(707)	(754)	(708)	(591)	(688)	(540)
Covered in 1967 ×	-1,444	-1,025	-1,529	-1,820*	-1,641	-1,812*
1973–1980 [$\hat{\beta}_1^c$]	(1,499)	(1,006)	(1,640)	(869)	(1,617)	(892)
$\hat{\beta}_1^c / \bar{Y}_{C0}$.0132	.0537*	.0108	.0313	.00824	.0347*
	(.0166)	(.0269)	(.0161)	(.0211)	(.0157)	(.0193)
$\hat{\beta}_2^c / \bar{Y}_{C0}$	-.0339	-.0366	-.0349	-.065*	-.0374	-.0647*
	(.0351)	(.0359)	(.0374)	(.031)	(.0369)	(.0319)
\bar{Y}_{C0}	42,671	27,984	43,851	27,984	43,851	27,984
\bar{Y}_{T0}	29,932	20,179	31,626	20,179	31,626	20,179
Observations	407,822	37,770	370,052	37,770	370,052	36,895
State FE	N	N	N	Y	Y	N
State×Year FE	N	N	N	N	N	Y

Notes: Data used: March CPS, 1962–1981. Sample weights are used throughout the analysis. I impose the same sample restrictions as Derenoncourt and Montialoux (2021). *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the industry level. Baseline control variables include: a polynomial of order 3 in experience, dummies for: gender, years of schooling, working part-time, categories for weeks worked last year, hours worked last week, marital status, occupation, and industry. A dummy for Black is also included in Column 1.

wage policy lead to a statistically significant and economically meaningful short-run increase in earnings for those in treated industries, with Black workers experiencing a greater benefit. In panel (b), I account for Jensen bias, estimating an analogous specification for earnings in levels via PPML. The results are broadly to those from panel (a), hence I conclude that Jensen bias is not a significant concern in this setting.

In panel (c), Table 2 I present the results from a levels-OLS specification. It is at this point that it important to view the baseline outcome means, listed in the bottom panel of the table. There is a sizeable disparity in the baseline outcome means, which when coupled with a log-dependent variable can lead to sign differences in the treatment effect estimates from a level-and log-based specification.¹¹ The DD estimates from the levels-OLS specification, presented in panel (c), paint a very different picture to those from the analogous log-based specification in panel (a). The immediate short-run effects of the minimum wage policy are essentially null,

¹¹The respective treatment-control ratios of baseline outcome means for the Black and white, Black, and white samples is .70, .72, and .72.

although there does appear to be a small benefit to Black workers that is marginally statistically significant. The effect for Black workers is roughly half the size of the log-OLS estimate for the baseline model, present in column 2 (.054 instead of .095). This effect is transient however – the medium-term effects of the policy are negative in all seven sample-specification combinations, although only statistically significant at conventional levels for the two robustness specifications for Black workers (Columns 4 and 6).¹² Comparing specifications in panels (a) and (c), one can find evidence of the sign-switch that I discuss in Section 2.

4.2 Hebllich et al. (2020)

I next consider the reduced form evidence from Hebllich et al. (2020), specifically the results underlying Figure V of the article. I note that this reduced form evidence is not the core of the paper – rather it is used to motivate a structural model that allows the authors to trace the consequences of falling commuting costs on changing land use patterns within London, and the corresponding increase in size of the city. A key output from the authors' model is a counterfactual simulation of the impact of removing the London railway network, where population in levels is considered. The authors present event study evidence from dynamic DD and DDD specifications. The DD specification is:

$$\ln R_{jt} = \alpha_j + \sum_{\tau=-60}^{\tau=60} \beta_\tau (\mathbb{S}_j \times \mathbb{I}_{jt}) + (\mu_j \times \text{Year}_t) + d_t + v_{jt}, \quad (9)$$

where R_{jt} is parish population, \mathbb{S}_j an indicator if the parish has a railway station, and \mathbb{I}_{jt} a treatment year indicator.¹³ The DDD specification is

$$\begin{aligned} \ln R_{jt} = \alpha_j + \sum_{\tau=-60}^{\tau=60} \beta_\tau (\mathbb{S}_j \times \mathbb{I}_{jt}) + \sum_{\tau=-30}^{\tau=30} \gamma_\tau (\mathbb{S}_j \times \mathbb{I}_{jt} \times \mathbb{I}_j^{\text{Center}}) \\ + (\mu_j \times \text{Year}_t) + d_t + v_{jt}, \end{aligned} \quad (10)$$

where $\mathbb{I}_j^{\text{Center}}$ is an indicator for central London.

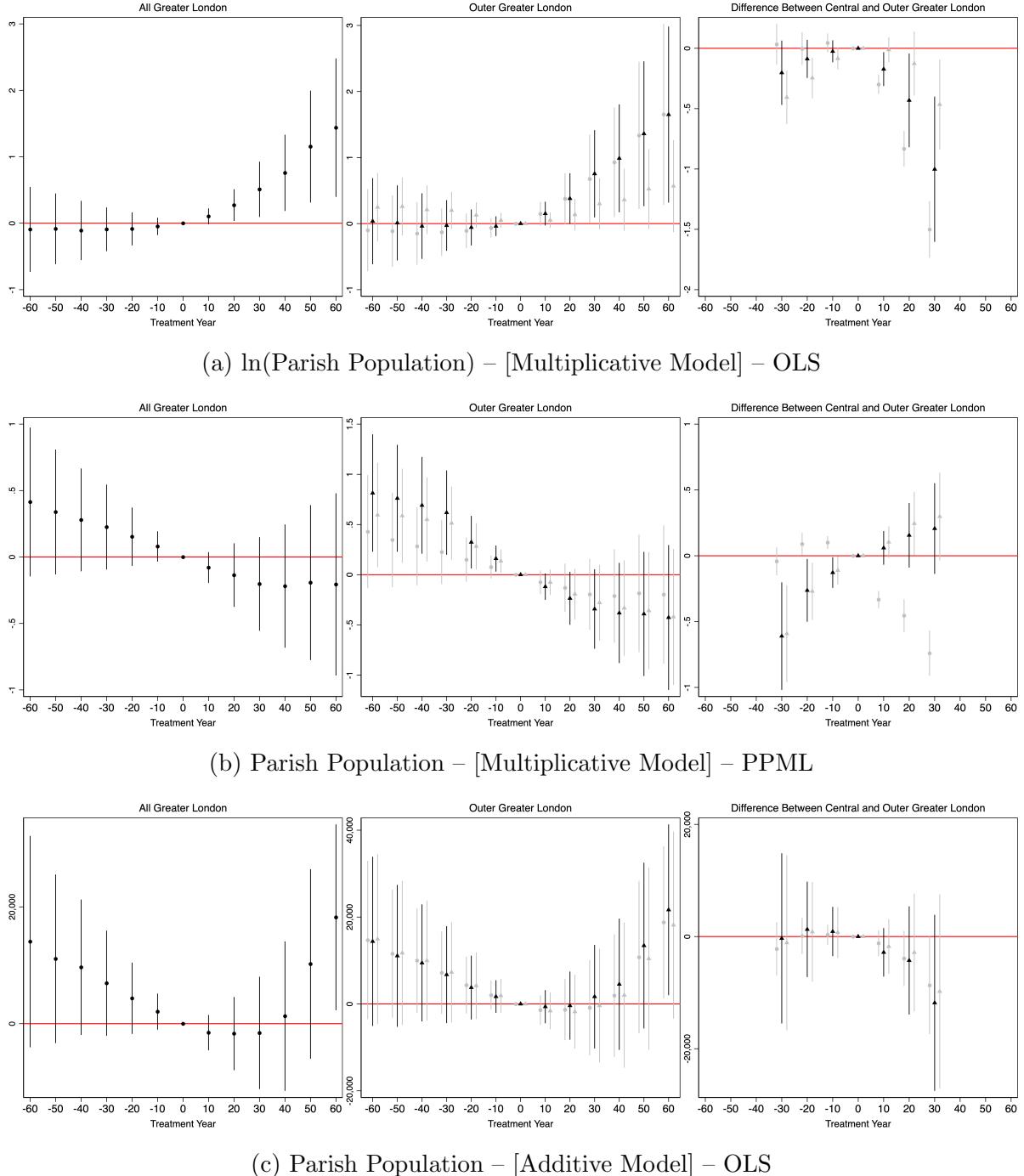
There is a sizeable disparity between never treated and ever treated parish population baseline means, which for the period before the first railway station (1801-1831) are respectively 2,199 and 9,107. I provide details of the parish population by treatment timing groups in Figure C3.

I present three sets of results based on this article. In Figure 2a I replicate the results presented in Figure V of Hebllich et al. (2020), based on a log-OLS specification. This provides extremely clear evidence of the impact on population growth of the arrival of a railway station in the parish. Figure 2b presents estimates of a multiplicative model for parish population using PPML, thereby circumventing the inherent Jensen bias present in a log-population specification. The results from this, correctly estimated, multiplicative model do not provide evidence of a

¹²I provide further replication results in the Appendix, replicating Figure V and Figure VI from Derenoncourt and Montialoux (2021) for both log-OLS and levels-OLS specifications in Figure C2.

¹³See pages 2080-2084 of Hebllich et al. (2020) for further details.

Figure 2: Heblich et al. (2020) Replication



Notes: I use the same parish-by-decade panel data sample as Heblich et al. (2020). The coefficient plot displays DD and DDD coefficients along with 95% confidence intervals. Standard errors are clustered at the borough level. Baseline control variables include: parish fixed effects, parish-specific trends, and year fixed effects. The figures in the left-hand column display the β_T coefficients from Equation (9). The figures in the central column display the β_T coefficients from Equation (10). The figures in the right-hand column display the γ_T coefficients from Equation (10). In figures in the central and right-hand columns, the three different color-shape combinations represent three different definitions of central London. See the notes to Figure V of Heblich et al. (2020) for more details.

positive effect of the arrival of a railway station.¹⁴

Finally, in Figure 2c I present results from a levels-OLS series of specifications. Given the authors target population in levels, not in growth rates, for their concluding counterfactual simulation, I take the levels specification to be the internally valid specification. The levels-OLS results suggest that the arrival of a railway station offset a negative deviation from local trend, yet it is only after 60 years that one finds a statistically significant DD estimate for this specification.

To conclude, if one maintains that a multiplicative model is correct and accounts for Jensen bias alone (Figure 2b), or if one decides that the additive model is the correct model instead (Figure 2c), one cannot replicate the pattern of event study parameter estimates documented in Figure V of Hebllich et al. (2020).

4.3 Aneja and Xu (2022)

Aneja and Xu (2022) study the consequences of the racial segregation of the civil service brought in with the Woodrow Wilson presidency. I selected this study as I suspected ex-ante that there would be no sign switches in the DD estimates when moving from a multiplicative to an additive model. The reason for this is that the authors using a matching procedure that explicitly targets the baseline outcome means. Without a disparity in baseline outcome means, there will be no sign-switch, as I show in Proposition 1.

The authors estimate a DD specification of the form:

$$\ln(w_{it}) = \beta \times Black_i \times Wilson_t + \theta_i + \tau_t + \varepsilon_{it}, \quad (11)$$

where w_{it} is annual earnings, $Black_i$ is an indicator for Black civil servants, and $Wilson_t$ is dummy for the period from 1913 onward.¹⁵ In Columns 2-4 of panel (a), Table 2 I replicate the baseline earnings results from Aneja and Xu (2022). I include an unconditional regression specification in Column 1, and I present additional evidence on the matched sample without using the CEM weights (Columns 5-8), in order to investigate the consequences for parameter estimates of re-introducing a baseline outcome disparity in this setting.

The purpose of the unconditional specification in Column 1 and Column 5 is that I can compare the exponentiated log-OLS coefficients ($\exp(\hat{\beta}_1) - 1$) to the proportional difference in growth rates of earnings ($(g_T - g_C)/g_C$), which I present in the lower panel of Table 3. In doing so I can assess the extent of Jensen bias in this setting, as these two numbers should be equal.

¹⁴I do not think that it is the case that railway stations were placed in areas with declining populations. Rather I think this reflects the fact that the authors specify a parish-specific trend in both their DD and DDD models, hence we are observing a pattern of deviations from long-term local trends. To explore this idea, I re-estimate all regression specifications, removing the parish-specific time trends. I present the corresponding estimates in Figure C4, and document a greater concordance in patterns of estimates across models in absence of parish-specific trends.

¹⁵See pages 924-926 of Aneja and Xu (2022) for more information on the matching procedure, and pages 926-928 for more information on the regression specification.

Table 3: Aneja and Xu (2022) Replication

	Matched, CEM-weighted				Matched, Unweighted			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(a) $\ln(\text{Annual Salary}) - [\text{Multiplicative Model}] - \text{OLS}$								
Black ×	-.0771***	-.0793***	-.0686***	-.0344***	-.0516***	-.054***	-.042***	-.00956
Wilson [$\hat{\beta}_1$]	(.0124)	(.0124)	(.00882)	(.00974)	(.00918)	(.00907)	(.00625)	(.00771)
$\exp(\hat{\beta}_1) - 1$	-.0742***	-.0763***	-.0663***	-.0338***	-.0503***	-.0526***	-.0411***	-.00952
	(.0115)	(.0114)	(.00823)	(.00941)	(.00871)	(.00859)	(.00599)	(.00764)
(b) Annual Salary – [Multiplicative Model] – PPML								
Black ×	-.115***	-.118***	-.106**	.0296	-.085**	-.0884**	-.0759*	.0483
Wilson [$\hat{\beta}_2$]	(.0445)	(.0453)	(.0458)	(.0527)	(.0384)	(.039)	(.0407)	(.047)
$\exp(\hat{\beta}_2) - 1$	-.109***	-.112***	-.1**	.03	-.0815**	-.0846**	-.0731*	.0495
	(.0397)	(.0403)	(.0412)	(.0542)	(.0352)	(.0357)	(.0378)	(.0493)
(c) Annual Salary – [Additive Model] – OLS								
Black ×	-.114***	-.117***	-.108**	45.3	-.136***	-.140***	-.123***	58.9
Wilson [$\hat{\beta}_3$]	(41.4)	(42.3)	(42.9)	(53.4)	(38.3)	(39.2)	(40.1)	(50.8)
$\hat{\beta}_3 / \bar{Y}_{C0}$	-.158***	-.163***	-.149**	.0627	-.155***	-.16***	-.141***	.0673
	(.0573)	(.0586)	(.0594)	(.074)	(.0437)	(.0448)	(.0458)	(.0581)
$(g_T - g_C) / g_C$	-.109				-.0815			
\bar{Y}_{C0}	722	722	722	722	875	875	875	875
\bar{Y}_{T0}	696	696	696	696	696	696	696	696
\bar{Y}_{C1}	994				1,169			
\bar{Y}_{T1}	854				854			
g_C	1.38				1.34			
g_T	1.23				1.23			
Year FE		Y	Y	Y		Y	Y	Y
Individual FE			Y	Y		Y	Y	Y
Race×Age FE				Y				Y
Observations	92,687	92,687	92,687	92,687	92,687	92,687	92,687	92,687

Notes: Data used: individual-by-year panel data following civil servants from Aneja and Xu (2022). *** denotes significance at 1%, ** at 5%, and * at 10%. Standard errors are clustered at the individual level. Black civil servants are matched to white counterparts using a CEM approach, as detailed in Aneja and Xu (2022). I use the CEM weights in the specifications that yield the estimates in columns 1–4, and present results from analogous unweighted regressions in columns 5–8. Columns 1 and 5 report coefficients from an unconditional DD regression. In columns 2 and 6, year fixed effects are included as additional controls. In columns 3 and 7, individual fixed effects are also included. Finally, in columns 4 and 8, race-by-age deciles are included as additional controls.

For both CEM-weighted (Column 1) and the unweighted (Column 5) results, the estimated DD coefficients from a log-OLS specification appear to be attenuated. I provide estimates from a level-PPML estimation approach in panel (b), which confirms this assessment, at least for the unconditional settings. Notice that the exponentiated PPML estimate is *exactly equal* to the proportional difference in growth rates that one estimates with a DD design and a log-dependent variable: $\exp(\hat{\beta}_2) - 1 = (g_T - g_C) / g_C$. To conclude, Jensen bias leads one to estimate too small a penalty to Black civil servants under the Woodrow Wilson presidency for the baseline specification (Column 3, Column 7).

The results from the levels-OLS approach confirm the evidence accrued from the levels-PPML approach. The sign-congruence of these results was expected. The size of the DD estimates from the multiplicative and additive model are broadly similar too – the proportional difference in growth rates from the PPML estimates ($\exp(\hat{\beta}_2) - 1$) are of a similar magnitude to the difference in differences of outcomes, normalized by the baseline control mean ($\hat{\beta}_3 / \bar{Y}_{C0}$). These similarities are only broad. For the baseline specification for the CEM-weighted data,

one would conclude, after accounting for Jensen bias, that Black civil servants experienced a 10% fall in earnings under the Wilson presidency assuming a multiplicative model, or a 14.9% decline based on an additive model, which would I argue is more appropriate here.

5 Conclusion

The aim of this paper is to make clear the consequence of functional form assumptions when one uses a DD model in an empirical setting where the baseline outcome distribution across groups differs substantively. I provide a condition, based on the aggregate time effect and the relative difference in baseline means, whereby a level- and log-specification will yield estimates of the DD term of opposing signs. The key reason that this sign-switch can occur is that using a DD model with a log-dependent variable leads to the estimation not of a difference-in-differences, but rather an approximation of the relative difference in growth rates across groups.

Using both simulations and empirical examples, I show that one can obtain DD estimates of different signs depending on whether one specifies the outcome variable in levels or in logs. Given the wide use of DD models for policy evaluation in areas that can give rise to such large differences in baseline outcome distributions – e.g., gender gaps in earnings, race gaps in the length of incarceration spells, house prices across regions or states, or school test scores across different education regimes – this point is likely to be of broad significance to applied researchers.

References

- Amanda Agan and Sonja Starr. Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment. *The Quarterly Journal of Economics*, 133(1):191–235, 2018.
- Ufuk Akcigit, John Grigsby, Tom Nicholas, and Stefanie Stantcheva. Taxation and Innovation in the Twentieth Century. *The Quarterly Journal of Economics*, 137(1):329–385, 2022.
- Alonso Alfaro-Ureña, Isabela Manelici, and Jose P Vasquez. The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages. *The Quarterly Journal of Economics*, 137(3):1495–1552, 2022.
- Abby Alpert, William N Evans, Ethan M J Lieber, and David Powell. Origins of the Opioid Crisis and its Enduring Impacts. *The Quarterly Journal of Economics*, 137(2):1139–1179, 2022.
- Marcella Alsan and Marianne Wanamaker. Tuskegee and the Health of Black Men. *The Quarterly Journal of Economics*, 133(1):407–455, 2018.
- Fernando Alvarez and David Argente. On the Effects of the Availability of Means of Payments: The Case of Uber. *The Quarterly Journal of Economics*, 137(3):1737–1789, 2022.
- Abhay Aneja and Guo Xu. The Costs of Employment Segregation: Evidence from the Federal Government Under Woodrow Wilson. *The Quarterly Journal of Economics*, 137(2):911–958, 2022. ISSN 0033-5533.
- Desmond Ang. The Effects of Police Violence on Inner-City Students. *The Quarterly Journal of Economics*, 136(1):115–168, 2021.
- Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2009.
- Clare Balboni, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil. Why Do People Stay Poor? *The Quarterly Journal of Economics*, 137(2):785–844, 2022.
- Oriana Bandiera, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat. The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats. *The Quarterly Journal of Economics*, 136(4):2195–2242, 2021.
- H. Spencer Banzhaf. Difference-in-differences hedonics. *Journal of Political Economy*, 129(8):2385–2414, 2021.
- Marc F. Bellemare and Casey J. Wichman. Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):50–61, 2020.
- Martin Beraja, Andreas Fuster, Erik Hurst, and Joseph Vavra. Regional Heterogeneity and the Refinancing Channel of Monetary Policy. *The Quarterly Journal of Economics*, 134(1):109–183, 2019.

Barbara Biasi and Heather Sarsons. Flexible Wages, Bargaining, and the Gender Gap. *The Quarterly Journal of Economics*, 137(1):215–266, 2022.

Kelly C. Bishop, Nicolai V. Kuminoff, H. Spencer Banzhaf, Kevin J. Boyle, Kathrine von Gravenitz, Jaren C. Pope, V. Kerry Smith, and Christopher D. Timmins. Best practices for using hedonic property value models to measure willingness to pay for environmental quality. *Review of Environmental Economics and Policy*, 14(2):260–281, 2020.

Zachary Bleemer. Affirmative Action, Mismatch, and Economic Mobility after California’s Proposition 209. *The Quarterly Journal of Economics*, 137(1):115–160, 2022.

Arthur Blouin and Rocco Macchiavello. Strategic Default in the International Coffee Market. *The Quarterly Journal of Economics*, 134(2):895–951, 2019.

Emily Breza and Cynthia Kinnan. Measuring the Equilibrium Impacts of Credit: Evidence from the Indian Microfinance Crisis. *The Quarterly Journal of Economics*, 136(3):1447–1497, 2021.

Emily Breza, Supreet Kaur, and Yogita Shamdasani. The Morale Effects of Pay Inequality. *The Quarterly Journal of Economics*, 133(2):611–663, 2018.

Paula Bustos, Gabriel Garber, and Jacopo Ponticelli. Capital Accumulation and Structural Transformation. *The Quarterly Journal of Economics*, 135(2):1037–1094, 2020.

Jing Cai and Shing-Yi Wang. Improving Management Through Worker Evaluations: Evidence from Auto Manufacturing. *The Quarterly Journal of Economics*, 137(4):2459–2497, 2022.

Lisa Cameron, Jennifer Seager, and Manisha Shah. Crimes Against Morality: Unintended Consequences of Criminalizing Sex Work. *The Quarterly Journal of Economics*, 136(1):427–469, 2021.

Davide Cantoni, Jeremiah Dittmar, and Noam Yuchtman. Religious Competition and Reallocation: the Political Economy of Secularization in the Protestant Reformation. *The Quarterly Journal of Economics*, 133(4):2037–2096, 2018.

Enrico Cantoni and Vincent Pons. Strict Id Laws Don’t Stop Voters: Evidence from a U.S. Nationwide Panel, 2008–2018. *The Quarterly Journal of Economics*, 136(4):2615–2660, 05 2021. ISSN 0033-5533.

David Card, Stefano DellaVigna, Patricia Funk, and Nagore Iribarri. Are Referees and Editors in Economics Gender Neutral? *The Quarterly Journal of Economics*, 135(1):269–327, 2020.

Vasco M Carvalho, Makoto Nirei, Yukiko U Saito, and Alireza Tahbaz-Salehi. Supply Chain Disruptions: Evidence from the Great East Japan Earthquake. *The Quarterly Journal of Economics*, 136(2):1255–1321, 2021.

Doruk Cengiz, Arindrajit Dube, Attila Lindner, and Ben Zipperer. The Effect of Minimum Wages on Low-Wage Jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454, 2019.

Gabriel Chodorow-Reich, Gita Gopinath, Prachi Mishra, and Abhinav Narayanan. Cash and the Economy: Evidence from India's Demonetization. *The Quarterly Journal of Economics*, 135(1):57–103, 2020.

Emanuele Ciani and Paul Fisher. Dif-in-dif estimators of multiplicative treatment effects. *Journal of Econometric Methods*, 8(1):20160011, 2019. doi: doi:10.1515/jem-2016-0011.

Zack Cooper, Stuart V Craig, Martin Gaynor, and John Van Reenen. The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured. *The Quarterly Journal of Economics*, 134(1):51–107, 2019.

Joppe de Ree, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers. Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia. *The Quarterly Journal of Economics*, 133(2):993–1039, 2018.

Ellora Derenoncourt and Claire Montialoux. Minimum Wages and Racial Inequality. *The Quarterly Journal of Economics*, 136(1):169–228, 2021.

Christian Dustmann, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp vom Berge. Reallocation Effects of the Minimum Wage. *The Quarterly Journal of Economics*, 137 (1):267–328, 2022.

Paul J Eliason, Benjamin Heebsh, Ryan C McDevitt, and James W Roberts. How Acquisitions Affect Firm Behavior and Performance: Evidence from the Dialysis Industry. *The Quarterly Journal of Economics*, 135(1):221–267, 2020.

Amy Finkelstein. The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare. *The Quarterly Journal of Economics*, 122(1):1–37, 02 2007.

Olle Folke and Johanna Rickne. Sexual Harassment and Gender Inequality in the Labor Market. *The Quarterly Journal of Economics*, 137(4):2163–2212, 2022.

Vasiliki Fouka. Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I. *The Review of Economic Studies*, 87(1):204–239, 05 2019.

Meredith Fowlie, Michael Greenstone, and Catherine Wolfram. Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program. *The Quarterly Journal of Economics*, 133(3):1597–1644, 2018.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277, 2021.

Anna Gumpert, Henrike Steimer, and Manfred Antoni. Firm Organization with Multiple Establishments. *The Quarterly Journal of Economics*, 137(2):1091–1138, 2022.

Sergei Guriev, Nikita Melnikov, and Ekaterina Zhuravskaya. 3G Internet and Confidence in Government. *The Quarterly Journal of Economics*, 136(4):2533–2613, 2021.

Stephen Hansen, Michael McMahon, and Andrea Prat. Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach. *The Quarterly Journal of Economics*, 133(2):801–870, 2018.

Guojun He, Shaoda Wang, and Bing Zhang. Watering Down Environmental Regulation in China. *The Quarterly Journal of Economics*, 135(4):2135–2185, 2020.

Stephan Heßlich, Stephen J Redding, and Daniel M Sturm. The Making of the Modern Metropolis: Evidence from London. *The Quarterly Journal of Economics*, 135(4):2059–2133, 05 2020.

Alessandro Iaria, Carlo Schwarz, and Fabian Waldinger. Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science. *The Quarterly Journal of Economics*, 133(2):927–991, 2018.

Katrine Jakobsen, Kristian Jakobsen, Henrik Kleven, and Gabriel Zucman. Wealth Taxation and Wealth Accumulation: Theory and Evidence From Denmark. *The Quarterly Journal of Economics*, 135(1):329–388, 2020.

Simon Jäger, Benjamin Schoefer, Samuel Young, and Josef Zweimüller. Wages and the Value of Nonemployment. *The Quarterly Journal of Economics*, 135(4):1905–1963, 2020.

Simon Jäger, Benjamin Schoefer, and Jörg Heining. Labor in the Boardroom. *The Quarterly Journal of Economics*, 136(2):669–725, 2021.

Ariella Kahn-Lang and Kevin Lang. The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics*, 38(3):613–620, 2020.

David A Keiser and Joseph S Shapiro. Consequences of the Clean Water Act and the Demand for Water Quality. *The Quarterly Journal of Economics*, 134(1):349–396, 2019.

Patrick Kline, Neviana Petkova, Heidi Williams, and Owen Zidar. Who Profits from Patents? Rent-Sharing at Innovative Firms. *The Quarterly Journal of Economics*, 134(3):1343–1404, 2019.

Nicolai V. Kuminoff and Jaren C. Pope. Do ‘capitalization effects’ for public goods reveal the public’s willingness to pay? *International Economic Review*, 55(4):1227–1250, 2014.

Bruce D. Meyer. Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics*, 13(2):151–161, 1995.

Sarah Miller, Norman Johnson, and Laura R Wherry. Medicaid and Mortality: New Evidence From Linked Survey and Administrative Data. *The Quarterly Journal of Economics*, 136(3):1783–1829, 2021.

John Mullaly. Interaction effects and difference-in-difference estimation in loglinear models. Working Paper 245, National Bureau of Economic Research, November 1999.

Sujeong Park and David Powell. Is the rise in illicit opioids affecting labor supply and disability claiming rates? *Journal of Health Economics*, 76:102430, 2021.

Lisa M. Powell and Julien Leider. The impact of seattle's sweetened beverage tax on beverage prices and volume sold. *Economics & Human Biology*, 37:100856, 2020.

Evan K Rose. Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders. *The Quarterly Journal of Economics*, 136(2):1199–1253, 2021.

Jonathan Roth. Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–322, 2022.

Jonathan Roth and Pedro H. C. Sant'Anna. When is parallel trends sensitive to functional form? *Econometrica*, 91(2):737–747, 2023.

Jason J Sandvik, Richard E Saouma, Nathan T Seegert, and Christopher T Stanton. Workplace Knowledge Flows. *The Quarterly Journal of Economics*, 135(3):1635–1680, 2020.

Bradley Setzler and Felix Tintelnot. The Effects of Foreign Multinationals on Workers and Firms in the United States. *The Quarterly Journal of Economics*, 136(3):1943–1991, 2021.

J. M. C. Santos Silva and Silvana Tenreyro. The Log of Gravity. *The Review of Economics and Statistics*, 88(4):641–658, 11 2006.

Matthew Smith, Danny Yagan, Owen Zidar, and Eric Zwick. Capitalists in the Twenty-First Century. *The Quarterly Journal of Economics*, 134(4):1675–1745, 2019.

Chenzi Xu. Reshaping Global Trade: The Immediate and Long-Run Effects of Bank Failures. *The Quarterly Journal of Economics*, 137(4):2107–2161, 2022.

Appendix

A Proofs

Repeating elements of Section 2 to keep this section self-contained, we can write the DD estimand as:

$$\begin{aligned} \alpha^{ATT} &= \{E[Y_{it} | D_i = 1, T_t = 1] - E[Y_{it} | D_i = 1, T_t = 0]\} \\ &\quad - \{E[Y_{it} | D_i = 0, T_t = 1] - E[Y_{it} | D_i = 0, T_t = 0]\} \end{aligned} \quad (12)$$

The sample analog of (1) is the DD estimator

$$\hat{\alpha}^{DD} = [\bar{Y}_{T1} - \bar{Y}_{T0}] - [\bar{Y}_{C1} - \bar{Y}_{C0}] \quad (13)$$

A simple regression specification we can use to estimate the ATT parameter is:

$$Y_{it} = \alpha_1 + \alpha_2 Treat_i + \alpha_3 Post_t + \alpha_4 (Treat_i \times Post_t) + \epsilon_{it}, \quad (14)$$

where $Treat_i$ is a treatment indicator, $Post_t$ the post-period indicator, and α_4 is the parameter of interest.

In Section 2, I note that the DD parameter we estimate (β_4) with a log-dependent variable can be expressed as:

$$\beta_4 \approx \exp(\alpha_4) - 1 = \frac{\frac{E[Y_{it} | D_i = 1, T_t = 1]}{E[Y_{it} | D_i = 1, T_t = 0]} - \frac{E[Y_{it} | D_i = 0, T_t = 1]}{E[Y_{it} | D_i = 0, T_t = 0]}}{\frac{E[Y_{it} | D_i = 0, T_t = 1]}{E[Y_{it} | D_i = 0, T_t = 0]}} = \frac{g_T - g_C}{g_C}, \quad (15)$$

where g_C and g_T are the respective growth rates in the outcome for control and treated groups.

A.1 Proof for Proposition 1

First, when the level-DD is positive i.e., $\alpha_4 > 0$:

$$\begin{aligned} \alpha_4 > 0 &\Rightarrow (E[Y_{11}] - E[Y_{10}]) - (E[Y_{01}] - E[Y_{00}]) > 0 \\ &\Rightarrow \Delta_1 - \Delta_0 > 0 \end{aligned} \quad (16)$$

Now consider the case where log-DD is zero i.e., $\beta_4 = 0$. This occurs when $g_T = g_C$ (see

Equation (15))

$$\begin{aligned}
\beta_4 = \exp(\beta_4) - 1 = 0 &\Rightarrow \frac{E[Y_{11}]}{E[Y_{10}]} \left/ \right. \frac{E[Y_{01}]}{E[Y_{00}]} - 1 = 0 \\
&\Rightarrow \frac{E[Y_{11}]}{E[Y_{10}]} \left/ \right. \frac{E[Y_{01}]}{E[Y_{00}]} = 1 \\
&\Rightarrow \frac{E[Y_{11}]}{E[Y_{10}]} = \frac{E[Y_{01}]}{E[Y_{00}]} \\
&\text{(minus 1 from both sides)} \quad \Rightarrow \frac{E[Y_{11}]}{E[Y_{10}]} - 1 = \frac{E[Y_{01}]}{E[Y_{00}]} - 1 \\
&\Rightarrow \frac{E[Y_{11}] - E[Y_{10}]}{E[Y_{10}]} = \frac{E[Y_{01}] - E[Y_{00}]}{E[Y_{00}]} \\
&\Rightarrow \frac{\Delta_1}{E[Y_{10}]} = \frac{\Delta_0}{E[Y_{00}]} \tag{17} \\
&\Rightarrow \Delta_1 E[Y_{00}] = \Delta_0 E[Y_{10}] \\
&\text{(minus } \Delta_0 E[Y_{00}] \text{ from both sides)} \quad \Rightarrow (\Delta_1 - \Delta_0) E[Y_{00}] = \Delta_0 (E[Y_{10}] - E[Y_{00}]) \\
&\Rightarrow (\Delta_1 - \Delta_0) = \Delta_0 \frac{(E[Y_{10}] - E[Y_{00}])}{E[Y_{00}]} \tag{18}
\end{aligned}$$

$$[\text{Express in terms of (additive) DD model parameters}] \quad \Rightarrow \alpha_4 = \alpha_3 \frac{\alpha_2}{\alpha_1} \tag{19}$$

Combining Equation (16) and Equation (18) yields a condition for $\text{sign}(\alpha_4) \neq \text{sign}(\beta_4)$:

when $0 < (\Delta_1 - \Delta_0) < \Delta_0 \frac{(E[Y_{10}] - E[Y_{00}])}{E[Y_{00}]}$ we will have $\text{sign}(\alpha_4) \neq \text{sign}(\beta_4)$.

Next, consider the case when the level-DD coefficient is negative i.e., $\alpha_4 < 0$. In this case, there will be discordance in signs when $\beta_4 > 0$ or $g_T > g_C$, which means the left-hand side of Equation (18) is greater than the right-hand side. This leads to a sign-switch condition for the case where $\alpha_4 < 0$:

when $\Delta_0 \frac{(E[Y_{10}] - E[Y_{00}])}{E[Y_{00}]} < (\Delta_1 - \Delta_0) < 0$ we will have $\text{sign}(\alpha_4) \neq \text{sign}(\beta_4)$.

The two conditions can be summarized simply as the following condition, which completes the proof:

when $0 < |\Delta_T - \Delta_C| < \left| \Delta_C \frac{(E[Y_{T0}] - E[Y_{C0}])}{E[Y_{C0}]} \right|$, we will have $\text{sign}(\alpha_4) \neq \text{sign}(\beta_4)$.

B DD Research Design Literature Review

Table B1: Articles Using a DD Design Published 2018-2022 in The *QJE*

Article	Any Con- tinuous Out- come?	Ever Log Out- come?	Article	Any Con- tinuous Out- come?	Ever Log Out- come?
Agan and Starr (2018)	0	–	Derenoncourt and Montialoux (2021)	1	1
Alsan and Wanamaker (2018)	1	1	Cameron et al. (2021)	1	0†
Breza et al. (2018)	1	0	Jäger et al. (2021)	1	1
Hansen et al. (2018)	1	0	Carvalho et al. (2021)	1	1
de Ree et al. (2018)	1	0	Guriev et al. (2021)	1	0
Iaria et al. (2018)	1	0	Rose (2021)	1	0
Fowlie et al. (2018)	1	1	Miller et al. (2021)	1	0
Cantoni et al. (2018)	1	0	Setzler and Tintelnot (2021)	1	1
Beraja et al. (2019)	1	1	Breza and Kinnan (2021)	1	1
Cooper et al. (2019)	1	1	Cantoni and Pons (2021)	1	1
Keiser and Shapiro (2019)	1	1	Akcigit et al. (2022)	1	1
Blouin and Macchiavello (2019)	0	–	Biasi and Sarsons (2022)	1	1
Kline et al. (2019)	1	1	Bandiera et al. (2021)	1	1
Cengiz et al. (2019)	1	1	Dustmann et al. (2022)	1	1
Smith et al. (2019)	1	0	Bleemer (2022)	1	1
Chodorow-Reich et al. (2020)	1	1	Aneja and Xu (2022)	1	1
Jakobsen et al. (2020)	1	1	Alpert et al. (2022)	1	0
Card et al. (2020)	1	0†	Balboni et al. (2022)	1	0
Eliason et al. (2020)	1	1	Gumpert et al. (2022)	1	1
Bustos et al. (2020)	1	1	Alfaro-Ureña et al. (2022)	1	1
Sandvik et al. (2020)	1	1	Alvarez and Argente (2022)	1	1
Heblich et al. (2020)	1	1	Xu (2022)	1	1
Jäger et al. (2020)	1	0	Cai and Wang (2022)	1	0
He et al. (2020)	1	1	Folke and Rickne (2022)	0	–
Ang (2021)	1	0			

Notes: This table displays summary information on all articles published in *The Quarterly Journal of Economics (QJE)* over the five year period 2018-2022 that used a difference-in-difference research design. Any Continuous Outcome? is 0 if all outcomes are binary, and 1 otherwise. For articles that included at least one continuous outcome, Ever Log Outcome? is zero if the continuous outcome is never presented in log form, and 1 otherwise. † denotes cases where a continuous outcome is present, and although the outcome is not logged, it is presented in inverse hyperbolic sine form.

C Additional Results

C.1 The Recenter-Then-Rescale (RTR) Approach

A simple way to operationalize this rescaling is the following:

$$Y_{it}^{RC} = \begin{cases} Y_{it} & \text{if } D_i = 0 \\ Y_{it} - (\bar{Y}_{T0} - \bar{Y}_{C0}) & \text{if } D_i = 1, \end{cases} \quad (20)$$

where \bar{Y}_{C0} and \bar{Y}_{T0} are the respective baseline outcome means for the control and treated groups. This rescaling is a mean shift of the treatment group distribution – it does not change the levels-based DD estimate, as I confirm below in the simulation exercise.

What does the DD estimate based on $\ln Y_{it}^{RC}$ return? Plugging the rescaled outcome into Equation (7), and noting that as a direct consequence of the rescaling we have $E[Y_{it}^{RC} | D_i = 1, T_t = 0] = E[Y_{it}^{RC} | D_i = 0, T_t = 0]$, we see that we estimate the difference in post-treatment outcomes, scaled by the post-treatment control mean:

$$\beta_4^{RC} \approx \exp(\beta_4^{RC}) - 1 = \frac{E[Y_{it}^{RC} | D_i = 1, T_t = 1] - E[Y_{it}^{RC} | D_i = 0, T_t = 1]}{E[Y_{it}^{RC} | D_i = 0, T_t = 1]}. \quad (21)$$

Note the denominator in Equation (21) – it is the control outcome mean in the post-period. I don't find this particularly appealing as it is mixing treatment effects with time effects. With a rescaling, we can express this in terms of the pre-treatment control mean, which aligns this approach with the first solution I propose: to rescale the levels DD estimate by the control baseline mean. I show in the simulation results below that in all cases, this recenter-then-rescale (RTR) approach returns the same estimate as the levels DD estimate normalized by the baseline control mean. We can express the estimate treatment effect from the RTR approach as:

$$(\exp(\beta_4^{RC}) - 1) \times \frac{E[Y_{it}^{RC} | D_i = 0, T_t = 1]}{E[Y_{it}^{RC} | D_i = 0, T_t = 0]} = \frac{E[Y_{it}^{RC} | D_i = 1, T_t = 1] - E[Y_{it}^{RC} | D_i = 0, T_t = 1]}{E[Y_{it}^{RC} | D_i = 0, T_t = 0]}. \quad (22)$$

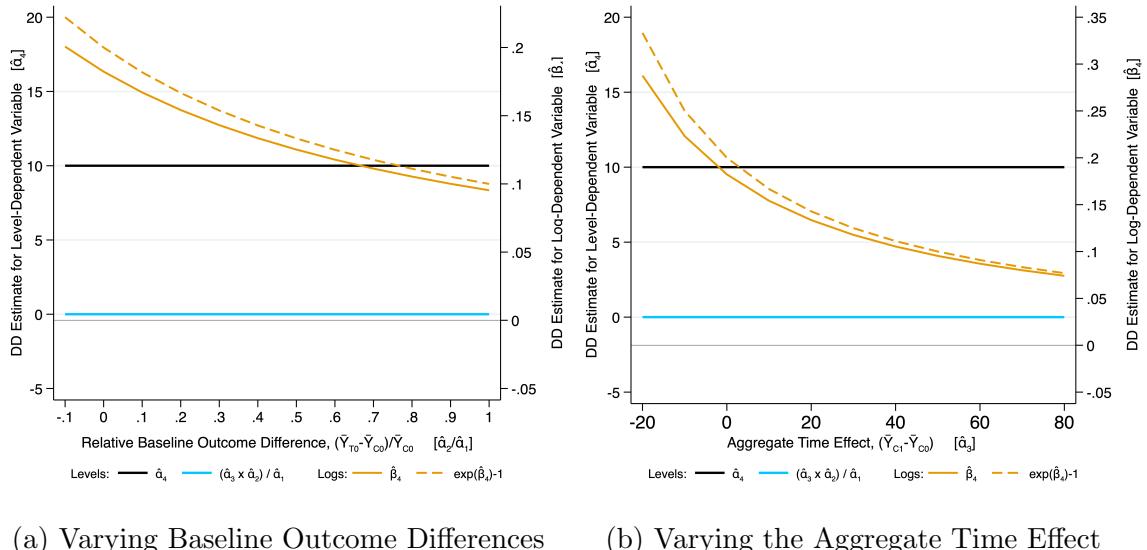
C.2 Additional Simulation Results

Table C1: Simulating a Zero Levels DD Effect That Yields a Positive or Negative Log Effect

	Positive Log Effect			Negative Log Effect	
	(1)	(2)	(3)	(4)	(5)
Level-Dependent Variable:					
DD Estimate $[\hat{\alpha}_4]$	-0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)
Log-Dependent Variable:					
DD Estimate $[\hat{\beta}_4]$	0.154*** (0.001)	0.080*** (0.000)	0.154*** (0.000)	-0.087*** (0.000)	-0.154*** (0.000)
$\exp(\hat{\beta}_4) - 1$	0.167*** (0.001)	0.083*** (0.000)	0.167*** (0.000)	-0.083*** (0.000)	-0.143*** (0.000)
$(g_T - g_C)/g_C$	0.167	0.083	0.167	-0.083	-0.143
\bar{Y}_{C0}	10.00	10.00	20.00	10.00	20.00
\bar{Y}_{C1}	12.00	12.00	24.00	12.00	28.00
\bar{Y}_{T0}	5.00	6.67	10.00	20.00	40.00
\bar{Y}_{T1}	7.00	8.67	14.00	22.00	48.00

Notes: Results based on 100 simulation runs. *** denotes significance at 1%, ** at 5%, and * at 10%. DD estimates for both a level- and log-dependent variable are presented in each column. The tables display the mean and (in parentheses) bootstrapped standard error of the DD estimates across all simulation runs. At the base of the table, the four elements of the DD are presented for reference. The sample size is 40000 in each simulation. Proportion treated and proportion in the post period are .5 and .5 respectively. The standard deviation of the error term is .2 in all simulations.

Figure C1: There is no Sign Switch Without a Baseline Outcome Difference or a non-Zero Aggregate Time Effect



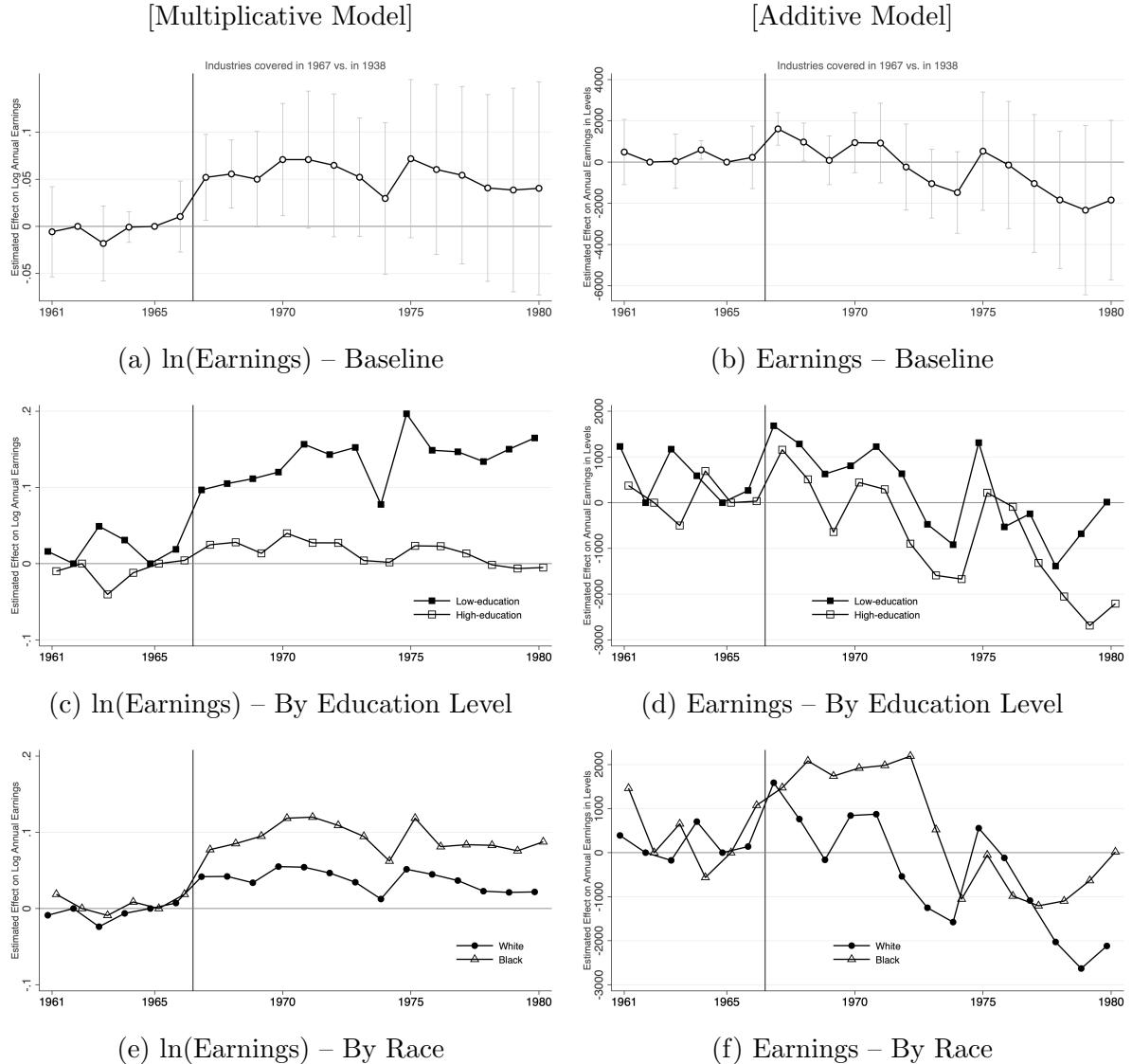
(a) Varying Baseline Outcome Differences

(b) Varying the Aggregate Time Effect

Notes: The underlying DGP that creates both graphs is an additive DD model specification, thus taking the form of Equation (3). In the left-hand graph, I set the following as fixed: $Y_{C0} = 50$, $Y_{T1} - Y_{T0} = 30$, $Y_{C1} - Y_{C0} = 0$ (this last choice means the aggregate time effect is zero). What varies is the relative baseline difference: $(Y_{T0} - Y_{C0})/Y_{C0} = \alpha_2/\alpha_1$. Once the relative baseline difference is fixed, this yields a specific value of Y_{T0} for the simulation exercise. In the right-hand graph, I set the following as fixed: $Y_{C0} = 50$, $Y_{T0} = 50$, $(Y_{T1} - Y_{T0}) = (Y_{C1} - Y_{C0}) + 10$. Note, there is no longer a baseline difference in outcome means. What varies is the aggregate time effect, $(Y_{C1} - Y_{C0})$. Once the aggregate time effect is fixed, this yields a specific value for both Y_{C1} and Y_{T1} for the simulation exercise. For each increment of relative baseline difference and aggregate time effect values respectively, 10,000 simulations are conducted. In both cases, the specification leads to a constant (levels) DD effect of 10 units, which is why the black line is horizontal. The sample size is 40,000 in each simulation. Proportion treated and proportion in the post period are .5 and .5 respectively. The standard deviation of the error term is .2 in all simulations.

C.3 Additional Results for the Replication of Derenoncourt and Montialoux (2021)

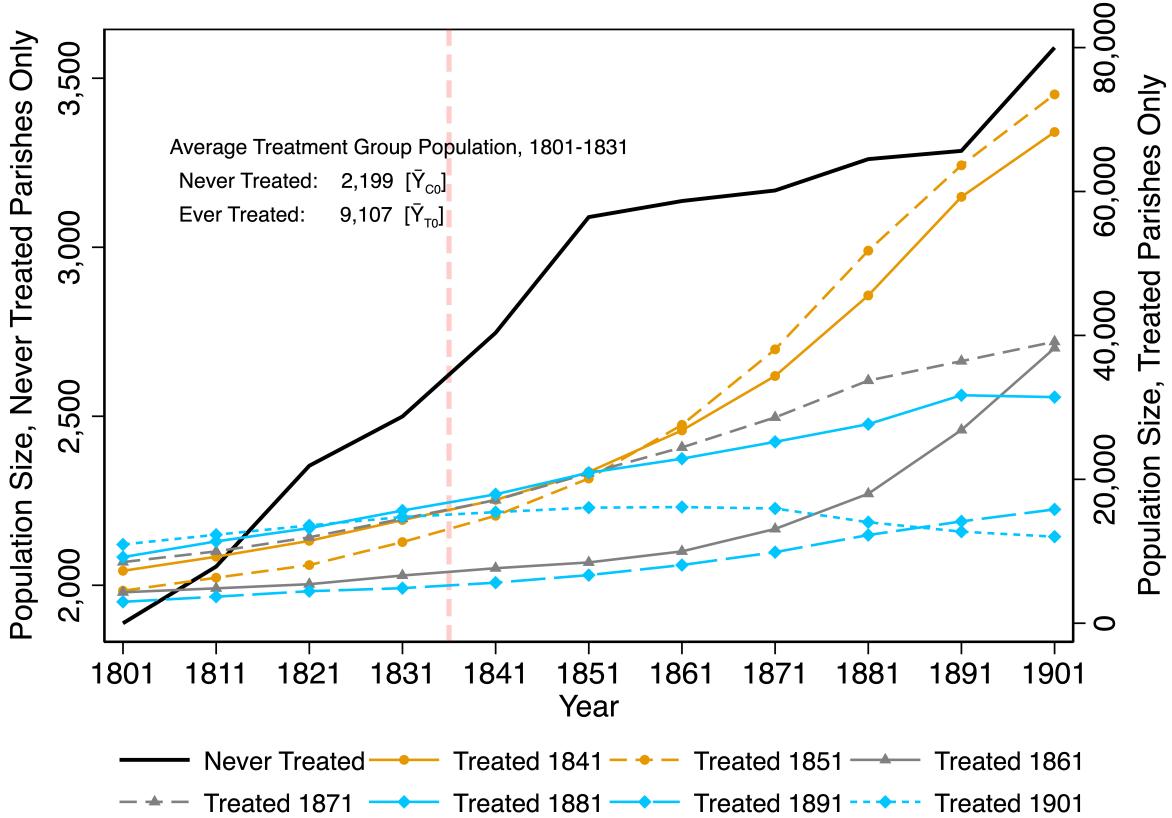
Figure C2: The Pattern of Estimates From Derenoncourt and Montialoux (2021) is Substantively Different With a Level-Dependent Variable Specification



Notes: Data used: March CPS, 1962-1981. Sample weights are used throughout the analysis. I impose the same sample restrictions as Derenoncourt and Montialoux (2021). The coefficient plot displays DD coefficients along with 95% confidence intervals. Standard errors are clustered at the industry level. Baseline control variables include: a polynomial of order 3 in experience, dummies for: gender, years of schooling, working part-time, categories for weeks worked last year, hours worked last week, marital status, occupation, and industry. A dummy for Black is also included in all specifications except for the final rows, as these figures reflect estimates by race. Figure C2c and Figure C2d are based on splitting the sample based on highest level of education (low = strictly less than high school graduation, high = high school graduation and above).

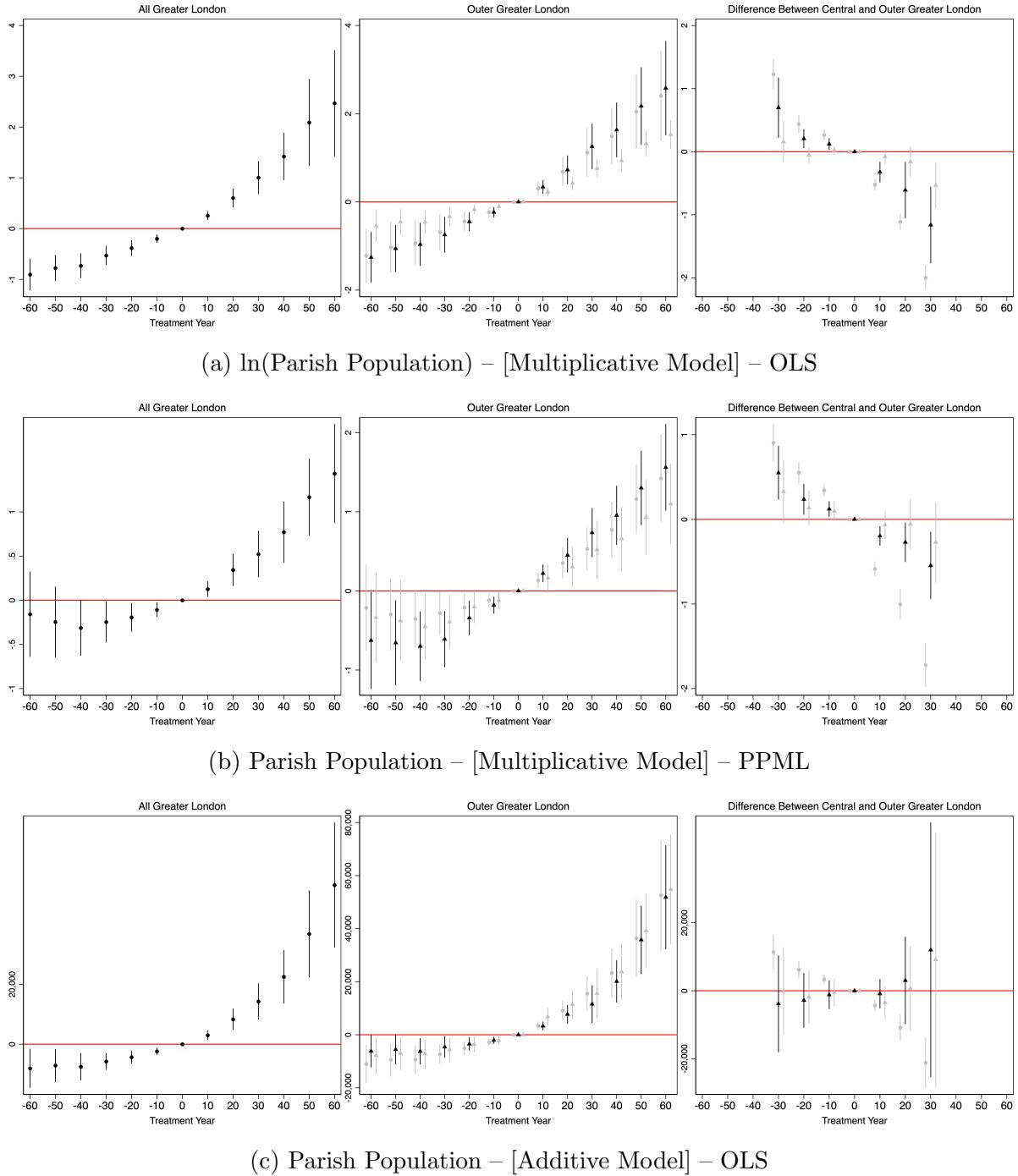
C.4 Additional Results for the Replication of Hebllich et al. (2020)

Figure C3: Parish Population Over Time by Treatment Groups



Notes: I use the same parish-by-decade panel data sample as Hebllich et al. (2020). The dashed, vertical line represents the arrival of the first railway opening.

Figure C4: Heblich et al. (2020) Baseline Model Without Parish-Specific Time Trends



Notes: I use the same parish-by-decade panel data sample as Heblich et al. (2020). The coefficient plot displays DD and DDD coefficients along with 95% confidence intervals. Standard errors are clustered at the borough level. Baseline control variables include: parish fixed effects and year fixed effects, but exclude parish-specific trends – this exclusion is what differentiates Figure C4 from Figure 2. The figures in the left-hand column display the β_τ coefficients from Equation (9). The figures in the central column display the β_τ coefficients from Equation (10). The figures in the right-hand column display the γ_τ coefficients from Equation (10). In figures in the central and right-hand columns, the three different color-shape combinations represent three different definitions of central London. See the notes to Figure V of Heblich et al. (2020) for more details.

D Empirical Examples

I now present two empirical examples to supplement the three published case studies that I present in Section 4. Both of the settings were chosen with an eye on the disparate baseline outcome distribution across the groups of interest, which as I note above is a necessary, but not sufficient, condition to generate DD estimates of opposing signs for the level- and log-dependent variable specifications. I first examine the differential racial impact of the great recession on working age males in the US. I then investigate the differential impact of the Brexit vote on Inner vs Outer London property prices. Note, these two examples are not based on previously published work.

D.1 Male Earnings and The Great Recession

The empirical specification I use to investigate the differential racial impact of the Great Recession on male earnings is:

$$Income_{it} = \delta Black_i + \gamma(Black_i \times Post_t) + X_i' \pi + \theta_{sxt} + \epsilon_{it} \quad (23)$$

where $Income_{it}$ is the total income earned in the previous year, and is specified in either levels or logs, for individual i in year t . $Black_i$ takes a value 0 for non-Hispanic white males, and a value of 1 for Black males. $Post_t$ is a dummy for the post-Great Recession period¹⁶. X_i is a vector of individual characteristics that includes dummies for highest level of educational attainment, dummies for potential experience in years, an indicator for being married, and dummies for metro area classification. θ_{sxt} is a set of state-by-year fixed effects. The error term is ϵ_{it} . I specify Eicker-Huber-White standard errors throughout. I present a set of summary statistics for the setting, and an overview of the data and sample selection decisions, in Appendix D.2.

Table D1 presents the key parameter estimates from both a level and log version of Equation (23). The DD estimate presented in Column (1) suggests that Black men fared slightly better in the Great Recession than their white counterparts. Figure D1 shows that both groups suffered absolute falls in incomes during this period, so the DD estimate in Column (1) reflects that the income drop for Black men was smaller than that for white men. The results in Column (2) present a starkly different conclusion, with a statistically significant, negative DD estimate. Using the raw means for the four DD cells, we can reconcile these two estimates – although Black men experience a lower absolute income drop, relatively it was larger than for white men as the drop occurred from a lower baseline level of income, which is why the coefficient in Column (1) is positive, and in Column (2) is negative.

With the benefit of having worked through what one recovers from a level and log specification in Section 2, and seen the consequences for these different functional form specifications of disparate baseline outcome distributions in Section 3, with the information at hand it is a

¹⁶Given that income in year t reflects income from the previous year, I code $Post_t = 1[Year >= 2009]$ in order to capture the Great Recession kicking in 2008.

Table D1: Level- and Log-Specifications Yield Estimates of Opposing Signs for the Impact of the Great Recession on the Racial Income Gap

	(1)	(2)	(3)	(4)	(5)	(6)
	All Education Levels		No College		College	
	Income in Levels	Income in Logs	Income in Levels	Income in Logs	Income in Levels	Income in Logs
Post × Black	775*	-.0361** (.0142)	1045 (762)	-.0231 (.0191)	346 (487)	-.0511** (.0213)
$\exp(\hat{\beta}_4)-1$		-.0354*** (.0137)		-.0228 (.0187)		-.0498** (.0202)
\bar{Y}_{C0}	48088	48088	58498	58498	31485	31485
\bar{Y}_{C1}	45905	45905	55160	55160	29806	29806
\bar{Y}_{T0}	30020	30020	38644	38644	22197	22197
\bar{Y}_{T1}	28495	28495	36080	36080	20866	20866
Adjusted R^2	.201	.202	.159	.152	.0846	.123
Observations	253,491	253,491	155,947	155,947	97,544	97,544

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The dependent variable is total income in 1999 Dollars, measured in levels in odd-numbered columns, and in logs in even-numbered columns. The following control variables are included in all regressions: state-by-year fixed effects, dummies for highest level of educational attainment, dummies for potential experience in years, an indicator for being married, and dummies for metro area classification. Data used: CPS 2005-2012.

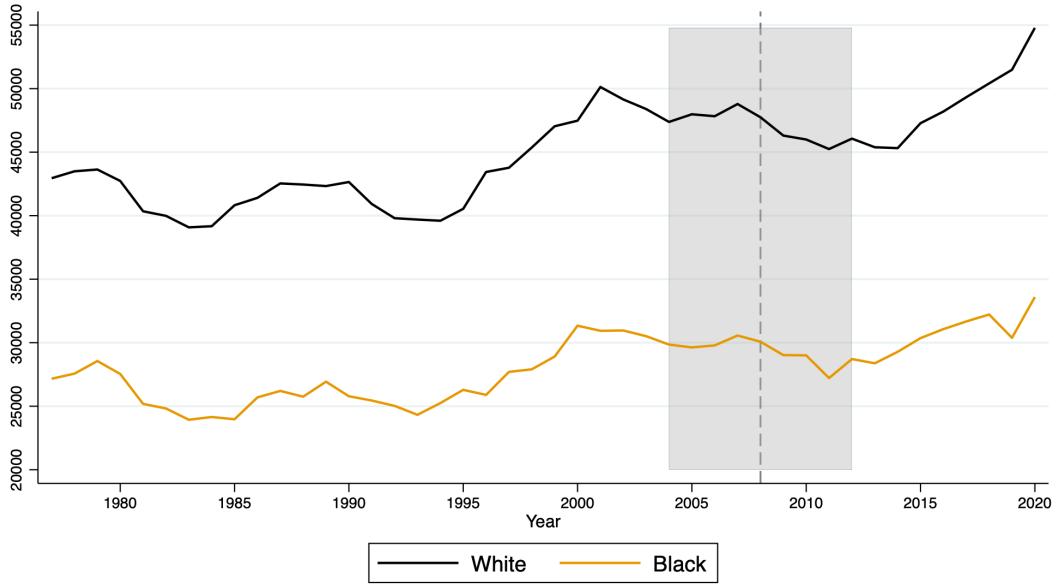
straightforward task to dissect the reason for the sign difference documented in Columns (1) and (2). However, in a typical applied setting, where a researcher is aiming to document the impact of a new policy using a DD model, the source of such disparate results may be less clear. It is the hope that this paper will aide in such settings.

The remaining columns of Table D1 present the DD parameters for both level and log specification, splitting the sample by education levels.

D.2 Male Earnings and The Great Recession – Supplemental Information

D.2.1 Long-Run Trends in Outcomes

Figure D1: Mean Annual Earnings By Race



Notes: Data used: CPS 1977-2020. The graph plots mean annual earnings, in 1999 Dollars, separately for Black and white men. Sample weights are used throughout.

D.2.2 Balance Tables

D.2.3 Data and Sample Selection

I use the ASEC March supplement of the CPS for the years 2005-2012, and restrict the sample to males, aged 25-65, who are either white non-Hispanic or Black. I omit those housed in group quarters. I additional remove those in the army, agricultural workers, those in the private household sector, and the self-employed. In order to maintain the same sample for both level and log specifications, I restrict total income to be strictly positive.

Table D2: Balance Tests for CPS Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	White			Black			
	Pre	Post	p-value: Difference	Pre	Post	p-value: Difference	p-value: DD
Sample Size	115,525	113,577		17,846	19,244		
Potential Experience	23.2 (11.3)	23.7 (11.7)	[.000]	22.5 (11.2)	23.3 (11.5)	[.000]	[.043]
Education:							
≤ High School	.0757	.0645	[.000]	.147	.129	[.000]	[.106]
High School	.318	.311	[.004]	.4	.394	[.297]	[.941]
Some College	.269	.274	[.024]	.27	.29	[.000]	[.008]
College	.221	.231	[.000]	.13	.128	[.685]	[.010]
Postgraduate	.117	.12	[.067]	.053	.0588	[.028]	[.357]
Married	.65	.632	[.000]	.443	.419	[.000]	[.362]
Metro Status:							
≤ Non-Metro	.184	.177	[.000]	.106	.0987	[.030]	[.904]
Central City	.196	.204	[.000]	.441	.433	[.188]	[.016]
Outside Central City	.452	.453	[.818]	.35	.36	[.077]	[.125]

Notes: Means and standard deviations (in parentheses for continuous covariates) are shown. *p*-values are based on OLS regressions with Eicker-Huber-White standard errors. Data used: CPS 2005-2012.

D.3 London House Prices and the Brexit Vote

I next turn to a different setting – the London housing market in the period around the Brexit vote. To understand the differential impact on house prices in Inner and Outer London, I specify a hedonic house price model of the form:

$$Price_{it} = \delta(Inner_i \times Post_t) + \sum_{m=1}^M (Market_m \times Post_t \times X'_i \gamma_m) + \pi_{m \times t} + \theta_b + \epsilon_{it}, \quad (24)$$

where $Price_{it}$ is the house price of house i (specified in either levels or logs), sold in period t (measured at the month-by-year level). $Inner$ takes the value of 0 for Outer London boroughs, and the value of 1 for Inner London boroughs.¹⁷ $Post_t$ is a dummy for properties sold post-Brexit vote, and δ is the parameter of interest.

X_i is a vector of property characteristics, specifically interactions between dummies for property type categories and new build status, and interactions between dummies for leasehold and new build status. I include interaction between the vector of housing characteristics, X_i , and market dummies in order to respect the “law of one price function” (Bishop et al., 2020). This allows the valuation of key property characteristics to vary across local housing markets. I allow the coefficients on all housing characteristics to differ in the pre and post periods, thereby allowing the hedonic price function to shift post-policy. I do so in order to avoid conflation bias (Kuminoff and Pope, 2014; Banzhaf, 2021).

$\pi_{m \times t}$ captures month-by-year housing market shocks to house prices. Housing markets are Travel To Work Areas – similar to Commuting Zones in the US. θ_b is a spatial fixed effect at the level of Output Area, akin to a census block in the US. Output Areas (OA) are the smallest census-based geographical unit – there are 181,408 of these in England and Wales, with an average population of 309 at the 2011 census.¹⁸. The Output Area fixed effect will capture all time-invariant local amenities – green spaces, transport links, shops, proximity to busy roads or motorways, as well as many slow-moving time-varying area characteristics (I am considering a minimum of 2 years, and a maximum of 4 years for these estimations), such as access to good schools or proximity to sources of pollution. The error term is ϵ_{it} . I specify Eicker-Huber-White standard errors throughout.

I first provide results for all property types in panel A of Table D3. The DD estimates for the level specification goes against my initial hypothesis that post-Brexit, Inner London property prices would suffer. For a variety of time windows around the Brexit vote ranging from 12 to 24 months, I document positive increases in Inner London prices relative to Outer London.

¹⁷To confuse matters there are two definitions of Inner London – the statutory definition, and the statistical version. In order to be consistent with both measures, I apply the strictest definition – in order for me to classify a borough as Inner London, a borough must be classified as Inner London by both definitions. This leads me to code the following boroughs as belonging to Inner London: Camden, Hackney, Hammersmith and Fulham, Islington, Kensington and Chelsea, Lambeth, Lewisham, Southwark, Tower Hamlets, Wandsworth, and Westminster.

¹⁸<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23>

As before, the log specification results are of the opposite sign, documenting a 4-7% decline in relative growth rates of Inner London properties. As in the previous section, we are sufficiently informed with the cell sample means, and the trends documented in Figure D2, to understand why. Once again, the disparate baseline outcome distributions play a key role. While Inner London properties experience a slight increase in levels compared to Outer London properties, Inner London properties (which started off at a much higher baseline level) grew less, leading to a negative proportional difference in growth rates – what the DD parameter approximates in a log specification.

Given that apartments account for almost 80% of property transactions in Inner London (see Table D4), in panel B of Table D3 I restrict the sample to only apartment transactions, and repeat the analysis. The reason to do so is to get a fairer sense of the house price impact of Brexit. The unintended consequence of this sample restriction was to create a larger wedge between the baseline outcome distributions. Looking at Column (3), the ratio of sample means for treatment to control in the baseline period is 1.74 in panel A, but 1.98 in panel B. This increased disparity in baseline outcome sample means explains at least part of why the difference between the level and log specifications is even more pronounced in panel B.

Table D3: Level- and Log-Specifications Yield Estimates of Opposing Signs for the Impact of the Brexit Vote on Inner London House Prices

	Time Window Around Brexit Vote				
	12 Months	15 Months	18 Months	21 Months	24 Months
A.) All Property Types					
Ai.) Price Specification					
Post \times Inner London [$\hat{\alpha}_4$]	10,045 (8,285)	16,861** (7,031)	12,125* (6,416)	10,639* (6,035)	4,588 (5,713)
Aii.) log(Price) Specification					
Post \times Inner London [$\hat{\beta}_4$]	-.039*** (.00379)	-.0421*** (.00334)	-.0502*** (.00307)	-.0574*** (.00286)	-.0694*** (.00267)
$\exp(\hat{\beta}_4)-1$	-.0383*** (.00365)	-.0412*** (.0032)	-.049*** (.00292)	-.0558*** (.0027)	-.0671*** (.00249)
\bar{Y}_{C0}	457,259	451,339	445,855	439,377	437,491
\bar{Y}_{C1}	482,338	488,284	490,084	490,799	491,442
\bar{Y}_{T0}	791,380	779,916	776,787	773,525	775,656
\bar{Y}_{T1}	825,855	829,664	836,659	845,105	847,116
Observations	210,626	262,937	309,348	358,806	411,928
B.) Apartments Only					
Bi.) Price Specification					
Post \times Inner London [$\hat{\alpha}_4$]	14,069** (6,310)	12,321** (5,640)	14,122*** (5,205)	11,360** (4,966)	8,373* (4,741)
Bii.) log(Price) Specification					
Post \times Inner London [$\hat{\beta}_4$]	-.0355*** (.00473)	-.0416*** (.00418)	-.0483*** (.00383)	-.0565*** (.00355)	-.0665*** (.00331)
$\exp(\hat{\beta}_4)-1$	-.0349*** (.00457)	-.0408*** (.00401)	-.0472*** (.00365)	-.055*** (.00336)	-.0643*** (.0031)
\bar{Y}_{C0}	340,331	336,618	333,686	328,396	325,193
\bar{Y}_{C1}	371,619	373,860	376,044	376,540	377,647
\bar{Y}_{T0}	674,757	665,064	659,124	656,169	652,712
\bar{Y}_{T1}	714,726	712,594	724,406	734,085	734,700
Observations	118,272	147,786	173,710	201,350	231,266

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The dependent variable house price in specification i.) and log house price in specification ii.). The following control variables are included in all regressions: block fixed effects, post-by-TTWA-by-property-type-by-new-build dummies, post-by-TTWA-by-leasehold-by-new-build dummies, and TTWA-by-year-by-month fixed effects. Data used: Land Registry Price Paid data 2014-2018.

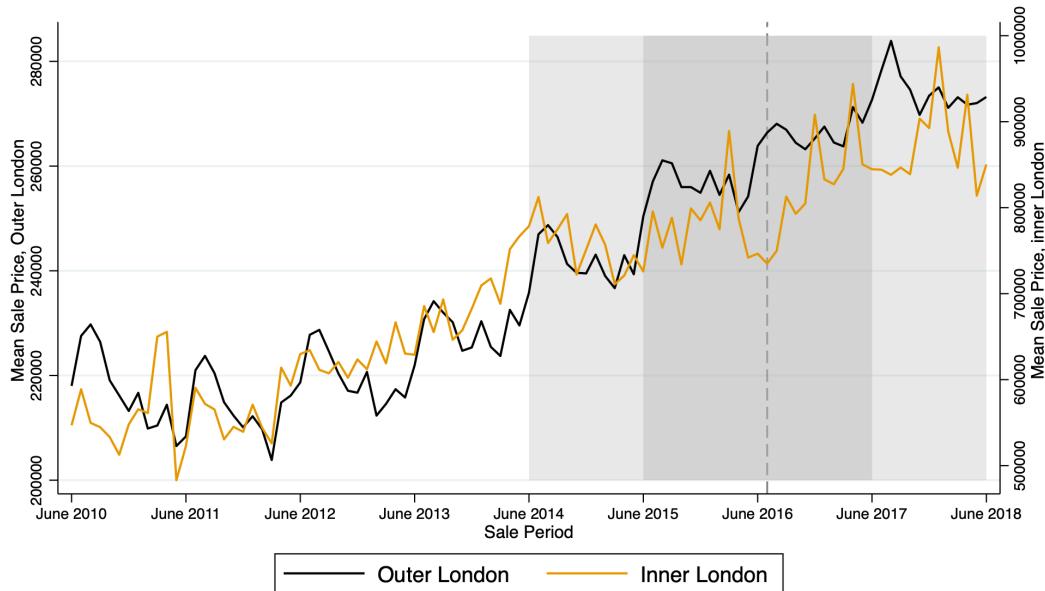
D.4 London House Prices and the Brexit Vote – Supplemental Information

D.4.1 Data Sample Selection

I use the Land Registry Price Paid data for a 2 year window around the Brexit vote, which occurred on 23 June 2016. The Price Paid data covers almost every house sale in England and Wales.^{19²⁰} The data is limited in terms of the household characteristics it contains (indicators for whether the property is a leasehold, if the property is a new-build, as well as property type). In order to compensate for the paucity of household characteristic controls, I specify an extremely low-level spatial fixed effect at the level of Output Area, akin to a census block in the US. Output Areas (OA) are the smallest census-based geographical unit – there are 181,408 of these in England and Wales, with an average population of 309 at the 2011 census.²¹

D.4.2 Long-Run Trends in Outcomes

Figure D2: Mean House Prices in Inner and Outer London – All Properties



Notes: Data used: HM Land Registry, 24 June 2010 - 23 June 2018. The graph plots mean monthly house prices for all properties sold during this period, separately for Inner and Outer London. I augment months to finish on the 23rd of the month, in order to align the data with the Brexit vote.

¹⁹Source: <https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads>, HM Land Registry.

²⁰The Land Registry list reasons for the minority of sales that are not registered at <https://www.gov.uk/guidance/about-the-price-paid-#data-excluded-from-price-paid-data>.

²¹<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23>

Figure D3: Mean House Prices in Inner and Outer London – Apartments Only



Notes: Data used: HM Land Registry, 24 June 2010 - 23 June 2018. The graph plots mean monthly house prices for all apartments sold during this period, separately for Inner and Outer London. I augment months to finish on the 23rd of the month, in order to align the data with the Brexit vote.

D.4.3 Balance Tables

Table D4: Balance Tests for House Sales Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Outer London			Inner London			
	Pre	Post	p-value: Difference	Pre	Post	p-value: Difference	p-value: DD
Sample Size	108,268	96,396		56,386	48,912		
New Build	.105	.139	[.000]	.169	.227	[.000]	[.000]
Property Type:							
Detached House	.0662	.0625	[.001]	.00727	.00634	[.065]	[.021]
Semi-Detached House	.184	.189	[.004]	.0271	.0281	[.355]	[.044]
Townhouse/Terraced	.302	.295	[.000]	.169	.167	[.319]	[.112]
Apartments	.448	.454	[.007]	.797	.799	[.354]	[.275]
Leasehold	.458	.466	[.000]	.806	.808	[.419]	[.078]

Notes: Means are shown. *p*-values are based on OLS regressions with Eicker-Huber-White standard errors. Data used: HM Land Registry, December 24, 2014 - December 23, 2017.