What's Logs Got to do With it: On the Perils of log Dependent Variables and Difference-in-Differences*

Brendon McConnell

July 28, 2023

Abstract

The log transformation of the dependent variable is not innocuous when using a difference-in-differences (DD) model. With a dependent variable in logs, the DD term captures an approximation of the proportional difference in growth rates across groups. As I show with both simulations and two empirical examples, 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.

1 Introduction

The difference-in-differences (DD) model is almost certainly the most popular quasi-experimental method 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 its design has, at least until recently, played a large role in its popularity. A recent literature documenting underlying issues with DD models, particularly in the case of staggered roll-out of treatment implementation (Goodman-Bacon, 2021), has somewhat shattered the illusion of the simplicity of these models.

In this paper, I highlight an additional complication that the applied researcher faces when operationalizing a DD model – 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.[1]

I start by outlining the key differences between a DD model with level- and log-dependent variables. 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, 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. This difference in the two specifications can yield highly disparate DD estimates when the distributions of the treated and untreated groups are sufficiently different in the pre-policy period.[2]

I then shift my attention to the singular aim of this paper: to explicate the consequences of the functional form assumption when the distribution of group outcomes in the pre-policy period differ substantively. These consequences can be stark. For a given aggregate time effect, I show that with a sufficiently large difference between group outcome distributions in the pre-policy period, it is possible for the level- and log-specifications of a DD model to yield DD parameter estimates of *different signs*. Thus 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 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.

I start by providing 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. Next, using first simulations, and then two empirical examples, I provide evidence of the sign disparity in estimated DD coefficients.[3]

In both empirical settings, using the same data and the same set of controls and fixed effects, I document a statistically significant DD estimate for a levels specification that is the opposite sign that that from a log specification. Both of these empirical case studies share the feature that the outcome distributions of treated and untreated groups are sufficiently different – a necessary, but not sufficient condition to generate a sign difference in DD estimates for level and log specifications.

The first empirical example examines differences in total earnings for white and Black men before and after the Great Recession. In the aftermath of the Great Recession, average earnings for both groups fell. Due to considerably

---

[1]Some papers are more explicit about the consequences of the choice between level- and log-specifications of the outcome variable when using difference-in-differences models (Finkelstein, 2007; Powell and Leider, 2020; Park and Powell, 2021).

[2]Meyer (1995) warned of using DD specifications in cases where the distribution of outcomes for treated and untreated groups was sufficiently different in the pre-policy period, noting in particular the issue of non-linear transformations of the dependent variable when there was a non-zero time effect: "This problem occurs because nonlinear transformations of the dependent variable imply different marginal effects on the dependent variable at different levels of the dependent variable" (Meyer, 1995, p. 155). This warning was echoed more recently by Kahn-Lang and Lang (2020).

[3]In order to maintain a singular focus on the consequences of functional form assumptions for DD models, I restrict my attention to the non-staggered timing, binary treatment case for this paper.

lower baseline earnings for Black men, whilst I estimate a positive DD estimate in levels, I estimate a statistically significant negative DD coefficient in logs. The reason for this is that the smaller fall in earnings for Black men was larger in proportional terms. Thus, a study investigating the differential racial impact of the Great Recession on male earnings would yield an opposing conclusion depending on whether the researcher chose a levels- or log-dependent variable specification.

In the second empirical case study, I consider differential price responses in the London housing market in the period around the Brexit vote. The hypothesis in this case is that the higher exposure of Inner London properties to overseas investors may have left this area more exposed to Brexit-related uncertainty than Outer London. This is another setting with large differences in the baseline outcome distribution across groups – the average price of a house in Inner London is almost double that of Outer London. Consequently, whilst the levels of house prices rose significantly more in Inner London compared to Outer London, the growth rate of Outer London was larger, which again leads to a sign difference in DD parameter estimates in the level and log specifications.

This paper contributes to the difference-in-differences literature by making clear the consequences of functional form assumptions in DD models, 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.

2　The DD Model

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 $Y_{it} = (1 - D_i)Y_{it}(0) + D_i Y_{it}(1)$, 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:

$$\alpha^{ATT} = \{E[Y_{it} \mid D_i = 1, T_t = 1] - E[Y_{it} \mid D_i = 1, T_t = 0]]$$
$$- [E[Y_{it} \mid D_i = 0, T_t = 1] - E[Y_{it} \mid D_i = 0, T_t = 0]\} \tag{1}$$

The sample analog of (1) is the DD estimator

$$\hat{\alpha}^{DD} = \left[\overline{Y}_{T1} - \overline{Y}_{T0}\right] - \left[\overline{Y}_{C1} - \overline{Y}_{C0}\right] \tag{2}$$

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

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

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:

$$lnY_{it} = \beta_1 + \beta_2 Treat_i + \beta_3 Post_t + \beta_4 (Treat_i \times Post_t) + \mu_{it}, \tag{4}$$

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

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

where $\mu_{it} = ln\eta_i t$. As noted by Mullahy (1999), and again by Ciani and Fisher (2019), we can write an expression for

the exponentiated DD parameter, based on the multiplicative model, as:

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

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:

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

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

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

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

Proof in Appendix A.

The Proposition above uses the notation $E[Y_{C0}] = E[Y_{it} \mid D_i = 0, T_t = 0]$, $E[Y_{T0}] = E[Y_{it} \mid D_i = 1, T_t = 0]$, $\Delta_C = E[Y_{it} \mid D_i = 0, T_t = 1] - E[Y_{it} \mid D_i = 0, T_t = 0]$, and $\Delta_T = E[Y_{it} \mid D_i = 1, T_t = 1] - E[Y_{it} \mid D_i = 1, T_t = 0]$. In the simulation approach I detail in Section 3, I verify Proposition 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 untreated outcomes, $Y(0)$, 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 testing[4], then does the the singular aim of this paper – to call attention to the possibility of a sign switch in

---

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

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.

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 model specifications.[5]

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. In all cases, $E[\epsilon_{it} \mid D_i, T_t] = 0$ and $\sigma_\epsilon = .2$. The proportion of treated and the proportion in the post period are .5 in both cases, and the total sample size is 40,000. This means each of the 2×2 DD cells has 10,000 observations.

The simulation results below reflect 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: Simulating a Postive Levels DD Effect That Yields a Zero or Negative Log Effect

	Zero Log Effect			Negative Log Effect	
	(1)	(2)	(3)	(4)	(5)
Level-Dependent Variable:					
DD Estimate [$\hat{\alpha}_4$]	0.40***	1.00***	0.40***	1.00***	0.80***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log-Dependent Variable:					
DD Estimate [$\hat{\beta}_4$]	−0.000	−0.000	−0.000	−0.043***	−0.017***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
exp($\hat{\beta}_4$)–1	−0.000	−0.000	−0.000	−0.042***	−0.017***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
$\overline{Y}_{C0}$	10.00	10.00	20.00	10.00	5.00
$\overline{Y}_{C1}$	12.00	15.00	22.00	12.00	6.00
$\overline{Y}_{T0}$	12.00	12.00	24.00	20.00	10.00
$\overline{Y}_{T1}$	14.40	18.00	26.40	23.00	11.80
$(g_T - g_C)/g_C$	−0.000	0.000	−0.000	−0.042	−0.017

Notes: Results based on 10000 simulation runs. DD estimates for both a level- and log-dependent variable are presented in each column. The tables display the mean and (in parentheses) bootrapped 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.

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.[6]

---

[5]If one is interested in the complementary case where the data generating process is based on the multiplicative model, useful references include Silva and Tenreyro (2006) and Ciani and Fisher (2019).

[6]Given that I do not introduce additional control variables in the simulations, the underlying model is saturated – the four parameters of Equation (3) map directly to the four cells of the DD design. This means that the proportional growth rate I present in the final row of the table matches perfectly $exp(\beta_4) - 1$.

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.

Table 2: Simulating a Zero Levels DD Effect That Yields a Postive 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***	0.080***	0.154***	–0.087***	–0.154***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
$\exp(\hat{\beta}_4)$–1	0.167***	0.083***	0.167***	–0.083***	–0.143***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
$\overline{Y}_{C0}$	10.00	10.00	20.00	10.00	20.00
$\overline{Y}_{C1}$	12.00	12.00	24.00	12.00	28.00
$\overline{Y}_{T0}$	5.00	6.67	10.00	20.00	40.00
$\overline{Y}_{T1}$	7.00	8.67	14.00	22.00	48.00
$(g_T - g_C)/g_C$	0.167	0.083	0.167	–0.083	–0.143

Notes: Results based on 10000 simulation runs. DD estimates for both a level- and log-dependent variable are presented in each column. The tables display the mean and (in parentheses) bootrapped 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.

Table 2 presents 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.

In Figure 1, I present DD estimates from both a levels- and log-based specification for two separate simulation experiments. 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.

What is concerning for applied researchers is that, if one has identified (i) a particular treatment and control group in order to evaluate a policy and (ii) a given sample period, then both the baseline outcome distributions of the respective groups and the resulting time effect is given. Thus, with a given sample, one may find oneself at the equivalent of either the left-hand side or the right-hand side of the x-axis in the graphs shown in Figure 1. Unlike the simulation exercise presented here, this will not be a choice, and one will not be able to shift the ratio of relative baseline outcome means or the aggregate time effect to explore the sensitivity of the parameter estimates to functional form assumptions. I build on this point with the two empirical case studies below.

4   Empirical Case Studies

I now present two empirical case studies. 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.

6

Figure 1: The Sign Switch Occurs When $\alpha_4 = \alpha_3 \dfrac{\alpha_2}{\alpha_1}$
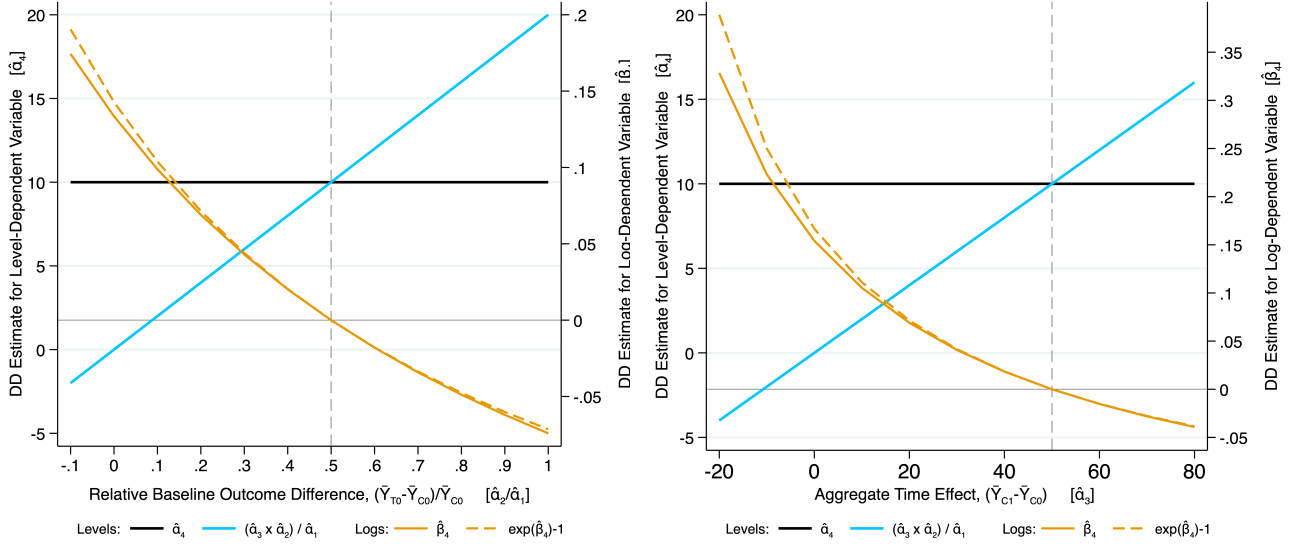
(a) Varying the Relative Baseline Outcome Difference   (b) Varying the Aggregate Time Effect

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: $Y_{C0} = 50$, $Y_{T1} - Y_{T0} = 30$, $Y_{C1} - Y_{C0} = 20$. 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. The standard deviation of the error term is .2 in all simulations.

4.1 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:

$$Income_{it} = \delta Black_i + \gamma(Black_i \times Post_t) + X_i'\pi + \theta_{s\times t} + \epsilon_{it} \tag{8}$$

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 period[7]. 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. $\theta_{s\times t}$ 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 B.1.

Table 3 presents the key parameter estimates from both a level and log version of Equation (8). The DD estimate presented in Column (1) suggests that Black men fared slightly better in the Great Recession than their white counterparts. Figure B1 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

---

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

Table 3: 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**	1045	-.0231	346	-.0511**
	(467)	(.0142)	(762)	(.0191)	(487)	(.0213)
$\exp(\hat{\beta}_4)-1$		-.0354***		-.0228		-.0498**
		(.0137)		(.0187)		(.0202)
$\overline{Y}_{C0}$	48088	48088	58498	58498	31485	31485
$\overline{Y}_{C1}$	45905	45905	55160	55160	29806	29806
$\overline{Y}_{T0}$	30020	30020	38644	38644	22197	22197
$\overline{Y}_{T1}$	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.

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 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 3 present the DD parameters for both level and log specification, splitting the sample by education levels.

4.2 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:

$$Price_{it} = \delta(Inner_i \times Post_t) + \sum_{m=1}^{M}(Market_m \times Post_t \times X_i'\gamma_m) + \pi_{m \times t} + \theta_b + \epsilon_{it} , \qquad (9)$$

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.[8] $Post_t$ is a dummy for properties sold post-Brexit vote, and δ is the parameter of interest.

---

[8]To 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.

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.[9] 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 4. 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. 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 B2, to understand why. Once again, the disparate baseline outcome distributions play a key role. Whilst 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 B2), in panel B of Table 4 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.

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.

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.

---

[9] https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23

Table 4: 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 × Inner London $[\hat{\alpha}_4]$	10,045	16,861**	12,125*	10,639*	4,588
	(8,285)	(7,031)	(6,416)	(6,035)	(5,713)
Aii.) log(Price) Specification					
Post × Inner London $[\hat{\beta}_4]$	-.039***	-.0421***	-.0502***	-.0574***	-.0694***
	(.00379)	(.00334)	(.00307)	(.00286)	(.00267)
$\exp(\hat{\beta}_4)-1$	-.0383***	-.0412***	-.049***	-.0558***	-.0671***
	(.00365)	(.0032)	(.00292)	(.0027)	(.00249)
$\overline{Y}_{C0}$	457,259	451,339	445,855	439,377	437,491
$\overline{Y}_{C1}$	482,338	488,284	490,084	490,799	491,442
$\overline{Y}_{T0}$	791,380	779,916	776,787	773,525	775,656
$\overline{Y}_{T1}$	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 × Inner London $[\hat{\alpha}_4]$	14,069**	12,321**	14,122***	11,360**	8,373*
	(6,310)	(5,640)	(5,205)	(4,966)	(4,741)
Bii.) log(Price) Specification					
Post × Inner London $[\hat{\beta}_4]$	-.0355***	-.0416***	-.0483***	-.0565***	-.0665***
	(.00473)	(.00418)	(.00383)	(.00355)	(.00331)
$\exp(\hat{\beta}_4)-1$	-.0349***	-.0408***	-.0472***	-.055***	-.0643***
	(.00457)	(.00401)	(.00365)	(.00336)	(.0031)
$\overline{Y}_{C0}$	340,331	336,618	333,686	328,396	325,193
$\overline{Y}_{C1}$	371,619	373,860	376,044	376,540	377,647
$\overline{Y}_{T0}$	674,757	665,064	659,124	656,169	652,712
$\overline{Y}_{T1}$	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.

References

Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press, 2009.

H. Spencer Banzhaf. Difference-in-differences hedonics. *Journal of Political Economy*, 129(8):2385–2414, 2021.

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.

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.

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.

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

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.

John Mullahy. 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.

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.

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

Appendix

A Proofs

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

$$\alpha^{ATT} = \{E[Y_{it} \mid D_i = 1, T_t = 1] - E[Y_{it} \mid D_i = 1, T_t = 0]]$$
$$- [E[Y_{it} \mid D_i = 0, T_t = 1] - E[Y_{it} \mid D_i = 0, T_t = 0]\} \tag{10}$$

The sample analog of (1) is the DD estimator

$$\hat{\alpha}^{DD} = \left[\overline{Y}_{T1} - \overline{Y}_{T0}\right] - \left[\overline{Y}_{C1} - \overline{Y}_{C0}\right] \tag{11}$$

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

$$Y_{it} = \alpha_1 + \alpha_2 Treat_i + \alpha_3 Post_t + \alpha_4 (Treat_i \times Post_t) + \epsilon_{it}, \tag{12}$$

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:

$$\beta_4 \approx exp(\beta_4) - 1 = \frac{\dfrac{E[Y_{it} \mid D_i = 1, T_t = 1]}{E[Y_{it} \mid D_i = 1, T_t = 0]} - \dfrac{E[Y_{it} \mid D_i = 0, T_t = 1]}{E[Y_{it} \mid D_i = 0, T_t = 0]}}{\dfrac{E[Y_{it} \mid D_i = 0, T_t = 1]}{E[Y_{it} \mid D_i = 0, T_t = 0]}} = \frac{g_T - g_C}{g_C}, \tag{13}$$

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., $\alpha_4 > 0$:

$$\alpha_4 > 0 \Rightarrow (E[Y_{11}] - E[Y_{10}]) - (E[Y_{01}] - E[Y_{00}]) > 0$$
$$\Rightarrow \Delta_1 - \Delta_0 > 0 \tag{14}$$

Now consider the case where log-DD is zero i.e., $\beta_4 = 0$. This occurs when $g_T = g_C$ (see Equation (13))

$$\beta_4 = exp(\beta_4) - 1 = 0 \Rightarrow \frac{E[Y_{11}]}{E[Y_{10}]} \Big/ \frac{E[Y_{01}]}{E[Y_{00}]} - 1 = 0$$

$$\Rightarrow \frac{E[Y_{11}]}{E[Y_{10}]} \Big/ \frac{E[Y_{01}]}{E[Y_{00}]} = 1$$

$$\Rightarrow \frac{E[Y_{11}]}{E[Y_{10}]} = \frac{E[Y_{01}]}{E[Y_{00}]}$$

(minus 1 from both sides) $\quad \Rightarrow \dfrac{E[Y_{11}]}{E[Y_{10}]} - 1 = \dfrac{E[Y_{01}]}{E[Y_{00}]} - 1$

$$\Rightarrow \frac{E[Y_{11}] - E[Y_{10}]}{E[Y_{10}]} = \frac{E[Y_{01}] - E[Y_{00}]}{E[Y_{00}]}$$

$$\Rightarrow \frac{\Delta_1}{E[Y_{10}]} = \frac{\Delta_0}{E[Y_{00}]} \tag{15}$$

$$\Rightarrow \Delta_1 E[Y_{00}] = \Delta_0 E[Y_{10}]$$

(minus $\Delta_0 E[Y_{00}]$ from both sides) $\quad \Rightarrow (\Delta_1 - \Delta_0)E[Y_{00}] = \Delta_0(E[Y_{10}] - E[Y_{00}])$

$$\Rightarrow (\Delta_1 - \Delta_0) = \Delta_0 \frac{(E[Y_{10}] - E[Y_{00}])}{E[Y_{00}]} \tag{16}$$

[Express in terms of (additive) DD model parameters] $\quad \Rightarrow \alpha_4 = \alpha_3 \dfrac{\alpha_2}{\alpha_1} \tag{17}$

Combining Equation (14) and Equation (16) yields a condition for $sign(\alpha_4) \neq sign(\beta_4)$:

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

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 (16) is greater than the right-hand side. This leads to a sign-switch condition for the case where $\alpha_4 < 0$:

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

The two conditions can be summarized simply as the following condition, which completes the proof:
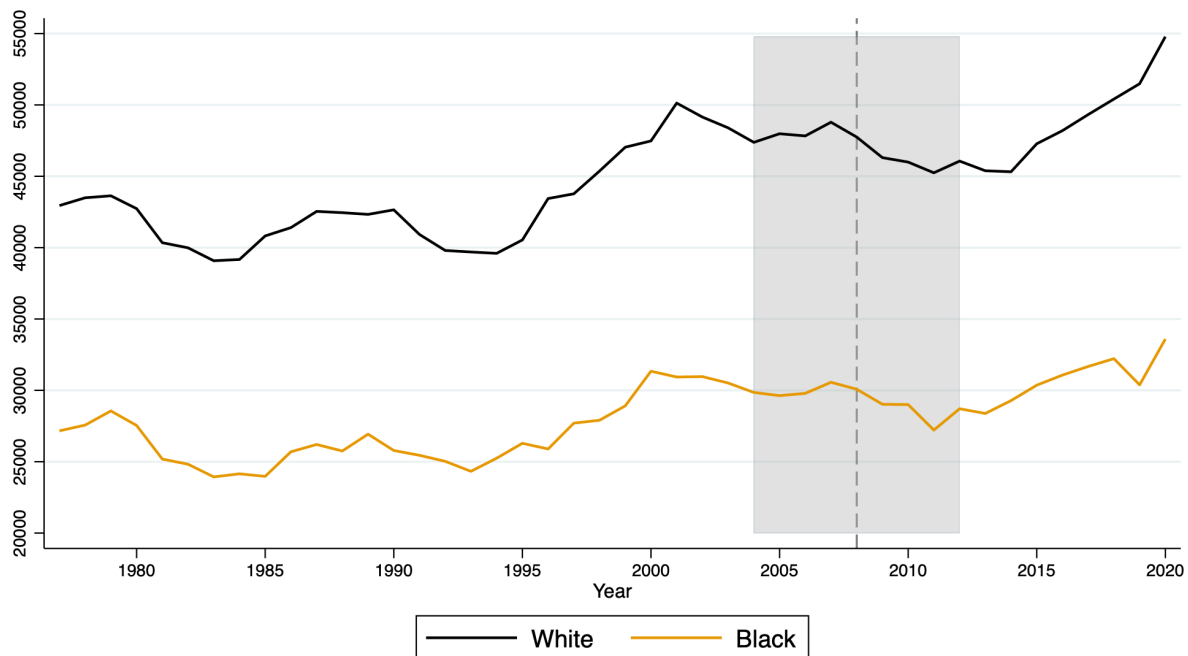
when $0 < |\Delta_T - \Delta_C| < \left| \Delta_C \dfrac{(E[Y_{T0}] - E[Y_{C0}])}{E[Y_{C0}]} \right|$, we will have $sign(\alpha_4) \neq sign(\beta_4)$.

B Summary Statistics

B.1 Male Earnings and The Great Recession

B.1.1 Long-Run Trends in Outcomes

Figure B1: Mean Annual Earnings By Race

B.1.2 Balance Tables

Table B1: 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	23.7	[.000]	22.5	23.3	[.000]	[.043]
	(11.3)	(11.7)		(11.2)	(11.5)		
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.

B.1.3 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.

B.2 London House Prices and the Brexit Vote

B.2.1 Long-Run Trends in Outcomes

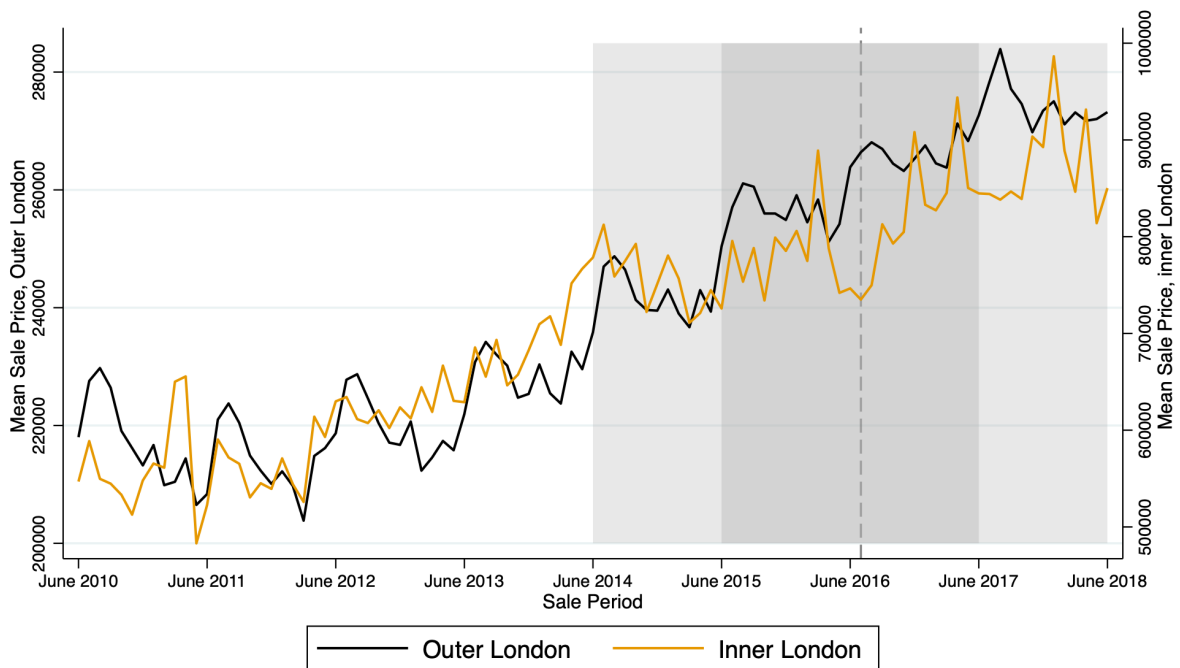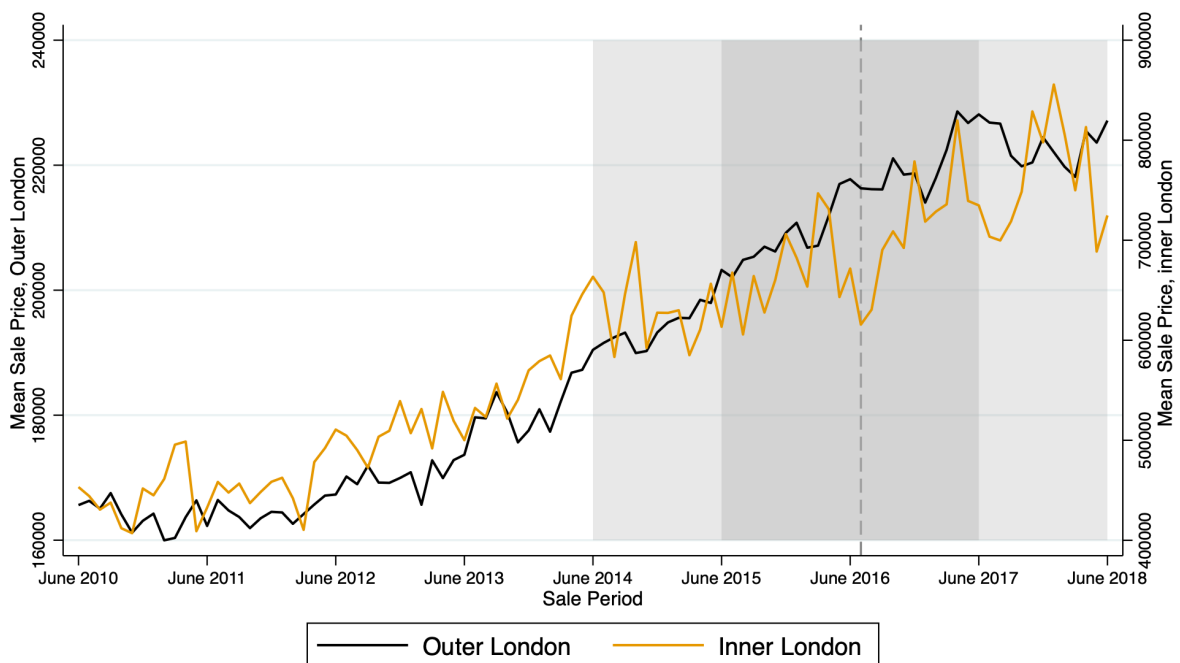Figure B2: Mean House Prices in Inner and Outer London – All Properties

Figure B3: Mean House Prices in Inner and Outer London – Apartments Only

B.2.2 Balance Tables

Table B2: 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.

B.2.3 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.[10][11] 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.[12] .

---

[10]Source: `https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads`, HM Land Registry.

[11]The 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`.

[12]`https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23`