

Emanuele Ciani^{1,2} / Paul Fisher³

Dif-in-Dif Estimators of Multiplicative Treatment Effects

¹ Bank of Italy, Rome, Italy² Centre for the Analysis of Public Policies, University of Modena and Reggio Emilia, Modena, Italy³ Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK, E-mail: pfisher@essex.ac.uk

Abstract:

We consider a difference-in-differences setting with a continuous outcome. The standard practice is to take its logarithm and then interpret the results as an approximation of the multiplicative treatment effect on the original outcome. We argue that a researcher should rather focus on the non-transformed outcome when discussing causal inference. The first step should be to decide whether the time trend is more likely to hold in multiplicative or level form. If the former, it is preferable to estimate an exponential model by Poisson Pseudo Maximum Likelihood, which does not require statistical independence of the error term. Running OLS on the log-linearised model might instead lead to confounding distributional and mean changes. We illustrate the argument with a simulation exercise.

Keywords: difference-in-differences, log-linearisation, Poisson Pseudo Maximum Likelihood

JEL classification: C21, C51, I38

DOI: 10.1515/jem-2016-0011


1 Introduction

In applied empirical research, it is common to replace continuous outcomes, such as earnings or expenditure, with their logarithm. Often, the choice is motivated by distributional features, like skewness.¹ In the difference-in-differences (dif-in-dif) setting, the desire to give a causal interpretation to the estimates complicates the choice. The model the researcher has in mind is usually one with multiplicative effects, which are linearised taking logs. If this is the case, the assumptions needed for causal inference refer to the non-transformed model. In general, this is not explicitly discussed.

We reviewed papers published in the Quarterly Journal of Economics (QJE), between 2001 and 2011. For our main literature review, 25 papers using a dif-in-dif estimator with continuous outcomes were found. A table with complete references is available in Appendix A. In 9 cases, the outcome is not transformed and an additive model is estimated. We found 16 papers in which at least one outcome is expressed in logarithmic form. The variables most commonly log-transformed are earnings and productivity, followed by a group of monetary quantities including expenditure, land value, exports and loans. In only 5 out of 16 cases is an explicit reason for the log-transformation given. For example, Nunn and Qian (2011) refer to concerns about skewness, whereas DellaVigna and Kaplan (2007) state that they wish to account for percentage changes in the control variables. In general, no discussion of the impact of the log transformation on the causal interpretation is given. Only Finkelstein (2007) states that the OLS estimates for the log of the dependent variable relate to $E(\ln(y|x))$, and not $\ln(E(y|x))$. To provide estimates of $\ln(E(y|x))$, Finkelstein (2007) estimates a generalised linear model (GLM) with log links.

Previous theoretical literature on non-linear dif-in-dif mostly focused on the interpretation of the interaction effect (Mullahy 1999; Ai and Norton 2003; Puhani 2012). A separate stream of research, not directly related to dif-in-dif, focused on the estimation of exponential models (Mullahy 1997; Manning 1998; Manning and Mullahy 2001; Ai and Norton 2008; Santos Silva and Tenreyro 2006; Blackburn 2007). In this paper we reconcile the two streams of research for the dif-in-dif case. Using a potential outcome framework, we reinterpret and review previous findings to argue that the choice between a multiplicative and an additive model is fundamental to the causal interpretation of the estimands. This choice should be taken before deciding whether or not to take logs, which should be understood as an estimation strategy rather than a matter of model specification.

Paul Fisher is the corresponding author.

 ©2018 Paul Fisher et al., published by DeGruyter.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.

Specifically, in Section 2 we compare and contrast the additive and multiplicative models. We then point out that using OLS on the log-linearised model may give biased estimates of the true multiplicative effect. This problem can arise if the treatment causes not only a shift in the mean, but also other distributional changes, for instance an increase in the variance for the treated group. Fortunately, a simple and robust non-linear estimator (Poisson Pseudo Maximum Likelihood) is available. In Section 3 we present a simulation to illustrate our main arguments. Section 4 concludes with a guideline for practitioners.

2 Model Specification and Inference

A practitioner estimating a dif-in-dif model with a continuous outcome usually faces two main issues:

1. Shall I model the time trend in additive or in multiplicative form? And shall I report the treatment effect as a difference in levels, or as a percentage change?
2. How can I estimate the multiplicative model? Shall I take logs?

We argue that these points should be addressed independently, in order to correctly separate model specification from estimation. The next subsections are dedicated to these issues.

2.1 Multiplicative or Additive Effects?

In this section we compare multiplicative and additive dif-in-dif models. First, we highlight the key difference is the specification of the time trend. Second, we show that the two models are related, but crucially one cannot give a causal interpretation to both.

We start with the simplest, though quite popular dif-in-dif setting, involving two groups ($g \in \{\text{control}, \text{treated}\}$) and two time periods ($t \in \{\text{pre}, \text{post}\}$), with only one group actually receiving the treatment in the second period. We analyse the case of a continuous outcome y , such as earnings or consumption.² First, we specify a model for the expected value of y when non treated (y_{0igt}), conditional on g and t . The second step is to assume how the expected value of the potential outcome when treated (y_{1igt}) is related with the expected y_{0igt} .

The standard dif-in-dif in levels (Angrist and Pischke 2009) combines an additive common trends assumption with an additive treatment effect:

$$E[y_{1igt}|g, t] = E[y_{0igt}|g, t] + \delta^* = \mu_g^* + \lambda_t^* + \delta^*. \quad (1)$$

where μ_g^* are group specific effects, λ_t^* are time specific, while δ^* is the treatment effect. The superscript $*$ is used to differentiate the model in levels from the multiplicative one.³ The additive model for the counterfactual outcomes leads to a linear model for observed outcome y_{it} :

$$y_{it} = \beta_0^* + \beta_1^* \text{treated}_{it} + \beta_2^* \text{post}_{it} + \beta_3^* \text{treated}_{it} \times \text{post}_{it} + \epsilon_{it} \quad (2)$$

$$E[\epsilon_{it} | \text{treated}_{it}, \text{post}_{it}] = 0. \quad (3)$$

where treated_{it} is a dummy for the treatment group and post_{it} for the second period. If the correct model for the counterfactuals is additive, then $\beta_0^* = \mu_{\text{control}}^* + \lambda_{\text{pre}}^*$, $\beta_1^* = \mu_{\text{treated}}^* - \mu_{\text{control}}^*$ and $\beta_2^* = \lambda_{\text{post}}^* - \lambda_{\text{pre}}^*$. The coefficient on the interaction term captures the quantity of interest, because it is equal to the treatment effect: $\beta_3^* = \delta^*$.

Differently, one might specify an exponential model

$$E[y_{0igt}|g, t] = \exp(\mu_g + \lambda_t) \quad (4)$$

where the assumption of common trends is in multiplicative form (see Mullahy 1997, for a discussion of IV estimation of an exponential model). Over time, the outcome in the absence of treatment would increase by the same percentage in both groups. We can assume a proportional treatment effect:

$$\frac{E[y_{1igt}|g, t] - E[y_{0igt}|g, t]}{E[y_{0igt}|g, t]} = \exp(\delta) - 1, \quad (5)$$

which implies that the effect is expressed as a proportional change with respect to the counterfactual scenario in the absence of the treatment. This comes naturally in applications involving continuous variables such as consumption and wages, where changes are commonly expressed in proportional terms, and it is consistent with the (proportional) specification of the time trend and the group difference. The multiplicative model is therefore:

$$E[y_{1igt}|g, t] = \exp(\mu_g + \lambda_t + \delta) \quad (6)$$

Intuitively, the total percentage change in the expected outcome of the treated group is composed of a percentage change due to time (call it %time) and the percentage effect of the treatment (call it %effect), so that $(1 + \%change) = (1 + \%time) \times (1 + \%effect)$. Differently, for the control group $(1 + \%change) = (1 + \%time)$. In this case the counterfactual model leads to an exponential model for the observed outcomes:

$$y_{it} = \exp(\beta_0 + \beta_1 \text{treated}_{it} + \beta_2 \text{post}_{it} + \beta_3 \text{treated}_{it} \times \text{post}_{it}) \eta_{it} \quad (7)$$

where η_{it} is a multiplicative error term that satisfies a mean independence assumption:

$$E[\eta_{it} | \text{treated}_{it}, \text{post}_{it}] = 1. \quad (8)$$

If the correct model for the counterfactuals is multiplicative, then $\beta_0 = \mu_{\text{control}} + \lambda_{\text{pre}}$, $\beta_1 = \mu_{\text{treated}} - \mu_{\text{control}}$ and $\beta_2 = \lambda_{\text{post}} - \lambda_{\text{pre}}$. More importantly, the exponentiated coefficient on the interaction term, which is a ratio of ratios (ROR, see Mullahy 1999; Buis 2010), is the quantity of interest to the researcher because it is directly related to the proportional treatment effect:

$$\exp(\beta_3) = \frac{E[y_{it} | \text{treated}_{it} = 1, \text{post}_{it} = 1]}{E[y_{it} | \text{treated}_{it} = 1, \text{post}_{it} = 0]} / \frac{E[y_{it} | \text{treated}_{it} = 0, \text{post}_{it} = 1]}{E[y_{it} | \text{treated}_{it} = 0, \text{post}_{it} = 0]} = \exp(\delta). \quad (9)$$

The researcher can also calculate the impact of the treatment (on the treated) during the *post* period in levels:

$$\begin{aligned} & \exp(\beta_0 + \beta_1 + \beta_2 + \beta_3) - \exp(\beta_0 + \beta_1 + \beta_2) \\ &= \exp(\mu_{\text{treat}} + \lambda_{\text{post}} + \delta) - \exp(\mu_{\text{treat}} + \lambda_{\text{post}}) = E[y_{1i,g=\text{treated},t=\text{post}}] - E[y_{0i,g=\text{treated},t=\text{post}}]. \end{aligned} \quad (10)$$

The well-known suggestion by Ai and Norton (2003) for non-linear models would instead lead us to calculate the cross difference (Mullahy 1999), which is not equal to the treatment effect:

$$\begin{aligned} & [\exp(\beta_0 + \beta_1 + \beta_2 + \beta_3) - \exp(\beta_0 + \beta_1)] - [\exp(\beta_0 + \beta_2) - \exp(\beta_0)] = \\ & [\exp(\mu_{\text{treated}} + \lambda_{\text{post}} + \delta) - \exp(\mu_{\text{treated}} + \lambda_{\text{pre}})] - [\exp(\mu_{\text{control}} + \lambda_{\text{post}}) - \exp(\mu_{\text{control}} + \lambda_{\text{pre}})], \end{aligned} \quad (11)$$

Indeed, as argued by Puhani (2012) (reprinted in Karaca-Mandic, Norton, and Dowd 2012), in any non-linear dif-in-dif model with an index structure and a strictly monotonic transformation function, the treatment effect is not equal to the cross-difference of the observed outcome, but rather to the difference between two cross-differences. Differently from the general non-linear case of Puhani (2012), the exponentiated interaction coefficient is directly interpretable as the proportional effect.

The treatment effect can therefore be easily expressed in levels in both models (using eq. 10). This highlights that the key difference between them is how the common trends assumption is specified. The reason is that the counterfactual y_{0igt} is identified by looking at the change in the control group over time. Once this counterfactual is correctly modeled, we can use it to understand which share of the average change in the treated group has to be attributed to the treatment. To clarify, Figure 1 is generated with an exponential model as in eq. (6). In this case the treated group starts from a lower position. Given the multiplicative trend, in the absence of the treatment the increase in this group over time would be smaller in absolute value. Therefore a standard dif-in-dif in levels would underestimate the share of the change that has to be attributed to the treatment. The bias

will be larger the larger the pre-treatment difference between the groups and the larger the proportional time effect. Differently, once we account correctly for the multiplicative time trend, it does not matter whether we express the treatment effect as a percentage difference or as a level difference. Indeed, the former is the fraction on the left hand side of eq. (5), while the latter is simply its numerator. Nevertheless, once the time trend is in multiplicative form, having a multiplicative treatment effect leads to an exponential model, which is clearer and easier to estimate. Furthermore, the effect in levels as measured in eq. (10) is specific to the *post* period. If we want to predict how the policy will affect future outcomes, and we believe that the true treatment effect is multiplicative, then it is more appropriate to focus on the percentage change.

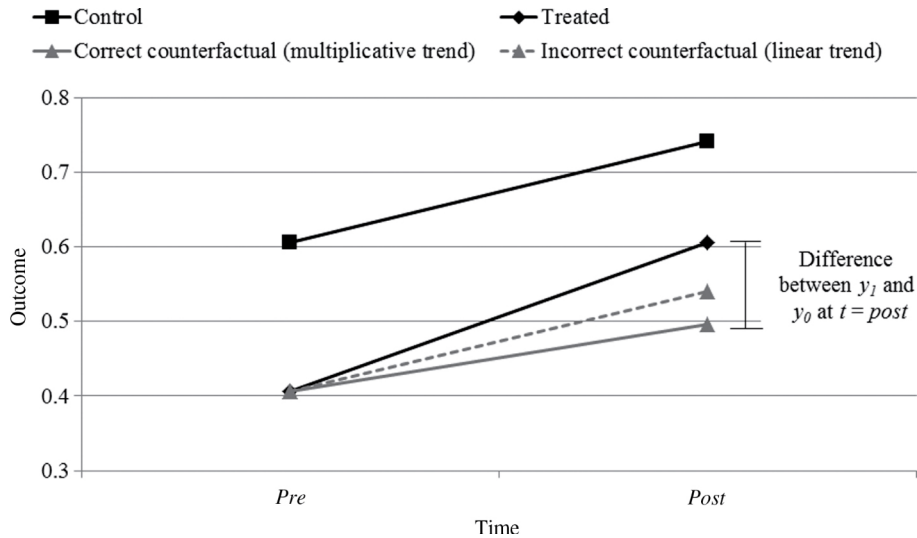


Figure 1: An Example of a Dif-in-Dif Setting with Multiplicative Time Trend. The figure shows the expected value of the actual and counterfactual outcome for the different groups according to the exponential model in eq. (4) and (5) with $\mu_{control} = -0.3$; $\mu_{treated} = -0.7$; $\lambda_{pre} = -0.2$; $\lambda_{post} = 0$; $\delta = 0.2$.

In a standard dif-in-dif setting with two periods and two groups, the practitioner may still be induced to treat the two models as fully equivalent. This is actually true with respect to the model for *observed* outcomes, because both the exponential model (7) and the linear one (2) are saturated: the four parameters fit perfectly the four averages given by the combination of $treated_{it}$ and $post_{it}$. Indeed, the exponential model is just a reparametrisation of the linear one, with

$$\exp(\beta_3) - 1 = \frac{(\beta_0^* + \beta_1^* + \beta_2^* + \beta_3^*) / (\beta_0^* + \beta_1^*)}{(\beta_0^* + \beta_2^*) / \beta_0^*} - 1. \tag{12}$$

This was noted by Gregg, Waldfoegel, and Washbrook (2006), who showed that we can estimate eq. (2) and then recover both the level and the percentage (multiplicative) effect.⁴

In spite of (12), if the true model for counterfactuals is multiplicative, β_3^* cannot be interpreted as a causal parameter, because it includes not only the level change due to the treatment, but also the difference between the time change in levels for the treatment and control groups [in fact, β_3^* is equal to the cross-difference from eq. (11)]. More generally, the equivalence (12) does not hold if we are willing to condition on other covariates, such as demographic controls. The reason is that the equation for the observed outcome is no longer saturated. Therefore it must be that either the linear model is correctly specified, or the exponential one, but not both. This is also true if we have more than two periods and a time trend is included.

The discussion of how the different specifications of the counterfactual are crucial for causal interpretation is related to Angrist and Pischke (2009, p. 230) comment that the assumption of common trends can hold either in logs or in levels, but not in both. We find it more natural to look at the choice between multiplicative or additive effects, rather than focusing on whether taking logs or not. This perspective has the advantage of stressing the distinction between specification and estimation. More importantly, in the next section we show that the multiplicative model and the log-linearised one are equivalent only under a strong restriction.

2.2 Estimation of a Multiplicative Dif-in-Dif Model

A popular solution to estimate a multiplicative model is to log-linearise it. As is well known, this practice may lead to biased estimates. However, this issue is often neglected in the dif-in-dif case (see Section 1). To

understand how it applies in this context, we follow the discussion in Santos Silva and Tenreyro (2006), but adapt it to the dif-in-dif setting. Log-linearising the model for the observed outcome (eq. 7) we obtain:

$$\ln y_{it} = \beta_0 + \beta_1 \text{treated}_{it} + \beta_2 \text{post}_{it} + \delta \text{treated}_{it} \times \text{post}_{it} + \ln \eta_{it}. \quad (13)$$

where we used the fact that $\beta_3 = \delta$ if the counterfactual model is multiplicative. The OLS estimator for the coefficient on the interaction term consistently estimates a quantity, say $\tilde{\delta}$, equal to

$$\begin{aligned} \tilde{\delta} = & \delta + \{E[\ln \eta_{it} | \text{treated}_{it} = 1, \text{post}_{it} = 1] - E[\ln \eta_{it} | \text{treated}_{it} = 1, \text{post}_{it} = 0]\} \\ & - \{E[\ln \eta_{it} | \text{treated}_{it} = 0, \text{post}_{it} = 1] - E[\ln \eta_{it} | \text{treated}_{it} = 0, \text{post}_{it} = 0]\} \end{aligned} \quad (14)$$

The mean-independence assumption $E[\eta_{it} | \text{treated}_{it}, \text{post}_{it}] = 1$, which has been imposed on the exponential model (eq. 8), does not ensure that $\ln \eta_{it}$ is mean-independent, as a consequence of Jensen's inequality. In general, the conditional expectation of $\ln \eta_{it}$ depends on higher moments of the conditional distribution of η_{it} .

Three interesting cases can be discussed. The first and most restrictive is when η_{it} is statistically independent from $x_{it} \equiv (1, \text{treated}_{it}, \text{post}_{it})$. Under this condition $E[\ln \eta_{it} | x_{it}]$ is a constant and therefore log-OLS is consistent for δ .⁵ Consider the restrictive condition this places on the conditional variance of the outcome:

$$\text{Var}[y_{it} | 1, x_{it}] = \text{Var}(\eta_{it}) \exp(2\beta_0 + 2\beta_1 \text{treated}_{it} + 2\beta_2 \text{post}_{it} + 2\delta \text{treated}_{it} \times \text{post}_{it}) \quad (15)$$

In short, the ratio of variances between different groups or time periods should be directly related to the differences in the conditional mean. Hence the time trend must not only shift the conditional mean, but also multiply the conditional variance by a factor equal to the square of $\exp(\beta_2)$. Similarly, the treatment effect must also shift the variance by a factor equal to the square of $\exp(\delta)$. This pattern of variance does not necessarily hold under the weaker condition of mean independence ($E[\eta_{it} | x_{it}] = 1$) introduced in eq. (8).

In the second case of interest statistical independence holds within groups, but not across groups (this is the cross-sectional 2 group case in Blackburn 2007). It follows that $E[\ln \eta_{it} | x_{it}]$ is constant over time within groups. Although both the intercept and the coefficient on the group dummy will be different from β_0 and β_1 , the coefficient on the interaction would be equal to δ , because the two terms in curly brackets in eq. (14) are both equal to zero. In terms of the conditional variance, we are allowing the distribution of the outcome to be arbitrarily different across the two groups, but we have reasons to believe that the dispersion over time respects condition (15) within each group.

In the third case of interest, only mean-independence holds, hence the expectation of $\ln \eta_{it}$ may change over time within the same group. Both terms in curly brackets in eq. (14) would be different from zero and log-OLS would give a biased estimate of δ . Focusing only on the second moment, this implies that there is an additional change in the variance within groups apart from the shifts induced by the time trend and treatment. This can happen, for instance, if a shock to the control group increases the dispersion, but not the average, in a way that violates (15). This mean preserving shock would not violate the parallel trends assumption, which only refers to the expected values, but would still lead to a bias for Log-OLS. A similar situation arises if there is an additional shock to the dispersion in the treated group that violates (15).⁶

This last case is likely to occur if treatment effects are heterogeneous across individuals. If all treated individuals show a response to the treatment equal to δ , then condition (15) is likely to be respected (in the absence of other distributional changes), because all values are shifted by a proportional factor. But if treatment effects are heterogeneous, then this is not necessarily true, because the distribution may change in different directions. A similar situation arises in those settings in which we define the treatment group in terms of eligibility for a policy, but not all the individuals in it effectively receive the treatment. Usually in these cases we are willing to estimate the intention-to-treat (that is the effect of being eligible). By construction, in the eligible group only some individuals (those actually treated) experience a change due to the policy, hence there may be other distributional changes apart from the mean.

Given that using OLS on the log-linearised model may lead to biased estimates, which are the available alternatives? If the error term is log-normal, one can work out the analytical formula for the bias and estimate the model using Maximum Likelihood Estimation (in Appendix B we derive the analytical formulas). However, practitioners using dif-in-dif for the estimation of treatment effects usually avoid introducing distributional assumptions other than mean-independence, so that MLE cannot be employed. Fortunately, we know from the literature that we can directly estimate the exponential model assuming only mean-independence (Santos Silva and Tenreyro 2006; Blackburn 2007), by using either Non Linear Least Squares (NLS) or Poisson Pseudo Maximum Likelihood (PPML). Santos Silva and Tenreyro (2006) argued in favour of the latter, because NLS is likely to be less efficient. The PPML simply estimates the model as if it was a Poisson, by maximizing the relative

likelihood.⁷ The fact that the actual variable is continuous rather than count does not hinder the consistency of the estimator, because PPML is consistent as long as the mean is correctly specified (as exponential).⁸ As the other properties of the Poisson distribution are not respected, a robust covariance matrix should be used. More generally, there are known difficulties of getting standard errors right in dif-in-dif designs. We discuss these in the context of the multiplicative setting in Section 3.2.

Instead of directly going for PPML, one could prefer to test whether statistical independence of the error term is likely to hold, in order to use OLS on the log-linearized model. One could use a Park test (Manning and Mullahy 2001; Santos Silva and Tenreiro 2006) for whether the conditional variance of y is proportional to the conditional mean squared, using consistent estimates from PPML. A more standard alternative is the Breusch-Pagan (BP) test for whether the estimated variance of the residuals from log-OLS is statistically dependent on the value of the *treated* \times *post* variable.

A researcher may also want to perform tests of whether the multiplicative specification is appropriate. This is not possible in the standard 2-periods 2-groups, because the additive and multiplicative model are observationally equivalent. When there are more time periods/groups (provided the model is not saturated), one could use Ramsey's RESET test (Ramsey 1969) for misspecification of the conditional mean or the robust Lagrange Multiplier (LM) test proposed by Wooldridge (1992). However, as discussed, the log-linearised model can be correctly specified as linear even if its coefficients do not identify the true treatment effect on the untransformed outcome, hence the test may not be informative if it fails to reject both specifications. Specification testing is further discussed in Appendix B.

Finally, we argued that the presence of heterogeneous treatment effects in the multiplicative model is a reason to avoid log-OLS and prefer alternatives, such as PPML, that do not require statistical independence. However, one important question is what the multiplicative model actually identifies in this circumstance. In the level model, given the additive nature of the effects, the well known result is that the dif-in-dif estimand identifies the average treatment effect on the treated. One may expect this interpretation to apply to the multiplicative case as well. Going back to the counterfactual model (eq. 5), we know that the quantity identified by the empirical model, $\exp(\delta) - 1$, captures the ratio between the average difference $y_{1i} - y_{0i}$ and the average outcome in the absence of treatment. Similarly to the linear case, we do not need to impose the constraint that the treatment effect is constant across all individuals. However, if it is heterogeneous in the population, $\exp(\delta) - 1$ captures the multiplicative treatment effect on the average, and not the average multiplicative effect (Angrist 2001, makes a similar point for the identification of treatment effects in IV estimation of exponential models). Further details can be found in Appendix B.

In the next section we illustrate our arguments and the use of PPML using a simulation. We also show that, for the dif-in-dif case, the BP test, relative to the Park test, seems to have more power in detecting deviations from homoskedasticity. In Appendix C, we further illustrate our arguments with an original empirical analysis of a UK educational grant on households' expenditure (the Educational Maintenance Allowance). In our analysis, treatment is less than 100 percent for the treated group and so treatment effects are heterogeneous. It also covers more than two time periods, and allows us to illustrate the use of RESET and LM tests for misspecification.

3 Simulations

3.1 Simulation of a Standard Dif-in-Dif Setting with Two Groups and Two Periods

We consider a similar setting to Figure 1 (see Appendix D for a graph using the simulation parameters). The outcome of interest is generated according to eq. (7), $i = 1, \dots, 2631$.

Each replication is generated according to $\beta_0 = 3.5$, $\beta_1 = -0.4$, $\beta_2 = 0.03$ and with the hypothetical reform having a constant multiplicative treatment effect equal to $\delta = 0.2$. The group sizes are 1090 for the treated and 1541 for the control (and so match with the applied example in Appendix C). Simulations with a negative time trend and a positive mean difference between groups lead to the same conclusions (available on request).

Each individual observation is generated according to $y_{it} = \exp(x_{it}\beta)\eta_{it}$ where η_{it} is log-normally distributed with $E[\eta_{it}|x_{it}] = 1$. The variance of η_{it} is specified as:

$$\text{Var}[\eta_{it}|x_{it}] = \exp(\alpha \times \mathbf{1}(\textit{treated}_{it} \times \textit{post}_{it} = 1)) \quad (16)$$

where $\mathbf{1}$ is an indicator function and α a parameter that determines the degree of heteroskedasticity in η_{it} . It follows that:

$$\text{Var}[\ln\eta_{it}|x_{it}] = \sigma_{\textit{treated}_{it}, \textit{post}_{it}}^2 = \ln[1 + \exp(\alpha \times \mathbf{1}(\textit{treated}_{it} \times \textit{post}_{it} = 1))] \quad (17)$$

$$E[\ln\eta_{it}|x_{it}] = -\frac{\sigma_{treated_{it},post_{it}}^2}{2} = -\frac{\ln[1 + \exp(\alpha \times \mathbf{1}(treated_{it} \times post_{it} = 1))]}{2} \quad (18)$$

To assess the performance of the three estimation strategies outlined above, simulations are reported for five key values of α . Table 1 reports results from 1000 replications of the simulation procedure.⁹

Table 1: Simulation Results.

	(1) Log-OLS	(2) PPML	(3) Level-OLS
$\alpha = 0$			
<i>Treated</i> × <i>Post</i>	0.1999924 (0.0711943)	0.1984138 (0.0842905)	4.689477 (2.510064)
<i>Treated</i>	-0.399073 (0.0398905)	-0.3976382 (0.0474284)	-10.84626 (1.290385)
Park Test	0.052	0.014	-
Breusch-Pagan	0.049	-	-
$\alpha = 0.1$			
<i>Treated</i> × <i>Post</i>	0.1710568 (0.0731679)	0.1951523 (0.0882464)	4.585499 (2.62745)
<i>Treated</i>	-0.4002561 (0.0404467)	-0.4000494 (0.0470304)	-10.90821 (1.277552)
Park Test	0.04	0.005	-
Breusch-Pagan	0.154	-	-
$\alpha = 0.2$			
<i>Treated</i> × <i>Post</i>	0.1506154 (0.0727646)	0.2026701 (0.0887447)	4.814591 (2.625155)
<i>Treated</i>	-0.4010721 (0.0406465)	-0.401937 (0.0496038)	-10.96927 (1.345574)
Park Test	0.046	0.017	-
Breusch-Pagan	0.441	-	-
$\alpha = 0.4$			
<i>Treated</i> × <i>Post</i>	0.0861068 (0.0776181)	0.1960358 (0.096877)	4.632431 (2.870258)
<i>Treated</i>	-0.399795 (0.0404998)	-0.3997392 (0.0484243)	-10.90645 (1.310166)
Park Test	0.064	0.022	-
Breusch-Pagan	0.945	-	-

Results from 10,000 replications. Mean of the estimated coefficients reported with standard deviations in parentheses.

The first special case of interest is where $\alpha = 0$ implying η_{it} is statistically independent of the treatment and other regressors. Here, OLS estimates from the log-linear model will provide consistent estimates of the multiplicative treatment effect. As expected both the log-OLS (column 1) and PPML (column 2) estimates are close to the true multiplicative treatment effect of 0.2. Whilst the difference between the two estimates is negligible, the log-OLS estimates are less dispersed, confirming the greater efficiency of the OLS estimator under statistical independence of the error term.

With $\alpha = 0.1$ we introduce heteroskedasticity, as in the third case of interest of Section 2.2, where in the *post* period there is change in the dispersion of the treated group that violates condition (15). The analytical bias can be calculated as (see equation 14 and Appendix B)

$$E[\ln\eta_{it}|treated_{it} \times post_{it} = 1] - E[\ln\eta_{it}|treated_{it} \times post_{it} \neq 1] = -0.5[\ln(1 + \exp(\alpha))] + 0.5[\ln(2)]. \quad (19)$$

Therefore, we expect that an increase in variance in the post-treatment period for the treated group (due to $\alpha > 0$) should induce a negative bias. Accordingly, we observe that now the log-OLS procedure performs less well. The distance from the true effect is around 2.9 percentage points, similar to the 2.6 point difference that can be calculated using formula (19). As expected, the bias is increasing with α , even though the variance of the estimated effects remains small. For example, the mean of the estimated treatment effects being only 43% of

the true effect when $\alpha = 0.4$. On the other hand, the PPML estimator performs well under all values of α , giving estimates close to the true treatment effect in all cases.

It is worth pointing out that the parameter values considered above imply an independent effect of treatment on the conditional variance of y that deviates only slightly from statistical independence. For example, under the strongest pattern of heteroskedasticity considered ($\alpha = 0.4$), the independent effect of treatment is to increase the conditional standard deviation of y by only 22%, whereas when $\alpha = 0.1$ the increase in standard deviation is just 5%. Even when very small distributional effects of treatment are introduced, the estimates from the log-linearised model are strongly biased. Moreover, equation (19) is independent of δ , so the bias as a proportion of the treatment effect will be larger if the treatment effect were smaller (indeed in much applied work treatment effects are much more modest than we consider here).

One could test whether log-linearisation is likely to give consistent estimates. Table 1 reports rejection rates at the 5% level for both the Park and the BP tests. The Park test is conducted both on the log-linearised estimates and directly on the PPML ones. Results from the simulation are not promising. In all cases the Park test fails to detect the mild pattern of heteroskedasticity that treatment introduces into the model. The rejection rates are around 5% for all values of α .¹⁰ For the BP test, the heteroskedasticity introduced into η_i is detected with reasonable power. For example, in the case where $\alpha = 0.4$ the test detects the inadequacy of the log-linearised specification 94.5% of the time.

Another important question is what would happen if, ignoring the multiplicative structure, we estimate a standard additive dif-in-dif regression. Column 3 of Table 1 presents such estimates. We observe that the estimated treatment effects repeatedly underestimate the true reform impact. The estimate is £4.69 in the baseline case, in contrast to the change in levels implied by the multiplicative model (£5.06), which can be calculated from eq. (10) using the true parameters.¹¹ The bias is independent from α . So although the regression for y_i is saturated and therefore correctly specified, the level estimates confound the treatment and trend effects.

The discussion above has mainly focussed on the consistency properties of PPML but we know that it may not be the most efficient estimator. If we do indeed have statistical independence of the error term, then log-OLS will be both consistent and efficient. If the error is truly log-normal, even when $\alpha \neq 0$ one could also use the MLE estimator. In Appendix B we show that, for this simulation, the efficiency loss using PPML is quite modest.

In Appendix D we analyse the case in which the treatment has no distributional effect, but (as in the second case of interest of Section 2.2) the pattern of variance across the treated and control groups does not respect the proportional structure from eq. (15). As discussed above, the log-OLS estimator for the treatment effect is consistent, while the treated-control difference is biased. We also analysed the case with a constant variance of y . Again here log-OLS performs poorly. Results are available from the authors.

3.2 Standard Errors in DID Designs: Simulation with Autocorrelated Errors

Procedures to correct for the fact that regular standard errors may overstate the precision of estimates of a treatment effect in DID regressions have been the subject of much debate and the literature is still unsettled (Bertrand, Duflo, and Mullainathan 2004; Donald and Lang 2007; Wooldridge 2003; 2006). Regular standard errors are derived under an iid assumption. This is violated when many years of (grouped) data is analysed in the presence of serially correlated outcomes. By means of a set of simulations, Bertrand, Duflo, and Mullainathan (2004) show that, in this context, conventional standard errors tend to over-reject the null of a zero reform effect.

To examine the relative performance of PPML and log-OLS standard errors, we performed a different set of monte-carlo simulations where the data generating processes are multiplicative AR(1) models with log normal errors and varying degrees of autocorrelation. Further details and results are provided in Appendix D. We found that rejection rates of a zero reform effect are increasing in the amount of serial correlation (as in Bertrand, Duflo, and Mullainathan 2004) but importantly, that they are comparable for both estimators. There was an indication that log-OLS performs slightly better under the most extreme pattern of autocorrelation we considered ($\rho = 0.8$), but marginally so. Put together, we conclude PPML standard errors appear to be no more biased than the log-OLS ones.

We also examined the performance of two common solutions (originally proposed by Bertrand, Duflo, and Mullainathan 2004): (1) restrict analysis to a short panel and (2) cluster standard errors at the group level (for large g).¹² Both solutions worked well in our simulations. When the sample size is very small (either small g or t) PPML performed relatively worse, but once again the differences were fairly small.

We conclude that the issues due to autocorrelation in DID designs are generally no more of a problem for the PPML estimator compared to the log-OLS one. Furthermore, while clustering standard errors may solve this issue, it does not address the inconsistency of log-OLS when the multiplicative error violates statistical

independence, that we discussed in Section 2.2. If, apart from autocorrelation, we also introduce patterns of heteroskedasticity as in Section 3.1 (Table 1), log-OLS rejection rates for the placebo reform are well above PPML ones, irrespective of whether we correct standard errors or not.

4 Conclusion

We critically assessed the standard practice of log-linearising in a dif-in-dif setting. We argued that a researcher should first decide whether a multiplicative or additive effect model is appropriate for the non-transformed outcome, because we cannot give a causal interpretation to both. If the multiplicative model is chosen and the researcher makes only a standard mean independence assumption, using PPML on the non-transformed variable can be preferable to using OLS on the log-linearisation. The reason is that the latter might give biased estimates of the multiplicative effect if there are changes in the higher moments of the outcome distribution that make the log-linearised error not mean independent. In particular, this bias may cause the OLS estimator to confound other distributional effects with the treatment effect on the mean.

As a summary, we think that the best practice for an applied researcher willing to estimate a dif-in-dif model with continuous outcome should be (a summary table can be found at the end of Appendix D):

1. Decide whether the time trend is more likely to hold in multiplicative or in level form.
2. If in levels, the best solution would be to use the standard level model and estimate it through OLS. The coefficient on the interaction term could be interpreted as an average treatment effect for the treated.
3. If in multiplicative form, the most coherent solution is to use and estimate an exponential model, with a multiplicative treatment effect.
 - a. In the presence of heterogeneous effects we can identify the multiplicative effect on the average for the treated group, and not an average multiplicative effect.
 - b. Without covariates, the multiplicative treatment effect can be recovered from OLS estimates of the standard dif-in-dif regression in levels (eq. 12).
 - c. Estimating the exponential model with PPML allows for covariates and for the presence of zeros in the dependent variable, and does not require statistical independence of the error term.
 - d. The researcher can test for heteroskedasticity using a BP test for the presence of heteroskedasticity with respect to the *treated* \times *post* variable. If they fail to reject the null of homoskedasticity, and the researcher is willing to assume statistical independence, OLS on the log-linearised model would be unbiased and efficient. This method also requires to eliminate or censor the zeros, which may introduce another source of bias.

Acknowledgments

We wish to thank João Santos Silva, Marco Francesconi, Mike Brewer, Susan Harkness, Jonathan James, Iva Tasseva, Vincenzo Mariani, Juan Hernandez, Roberto Nisticó, Massimo Baldini, Ben Etheridge, Ludovica Giua, seminar participants at Essex, two anonymous referees and the editor for useful comments. Financial support from the ESRC (Fisher and Ciani) and from the Royal Economic Society Junior Fellowship (Ciani) are gratefully acknowledged. The views expressed in this paper are those of the author and do not necessarily reflect those of the Bank of Italy. Data from the Expenditure and Food Survey has been accessed through the UK Data Archive.

Notes

- 1 Manning (1998) lays out common reasons for taking the logarithm of a dependent variable.
- 2 Several assumptions are required in order to identify the causal effect of the treatment. We draw attention on those related to the functional form. These depend on which feature of the distribution of y we are interested in. Here we focus on the expected value, which is usually the target in program evaluation using dif-in-dif. Athey and Imbens (2006) proposed instead a generalised dif-in-dif model that gives a structural interpretation to all differential changes in the distribution of the outcome y over time. As noted by Athey and Imbens (2006, pp. 435–436), our approach, focused on the conditional mean, is not nested in their model, unless one assumes that all individual shocks are statistically independent from group and time.
- 3 Note also that receiving the treatment, with a potential outcome y_{1igt} , does not coincide with being in the treated group, because in the first period all individuals go untreated.

- 4 However, Gregg, Waldfogel, and Washbrook (2006) defined the dif-in-dif “percentage method” as the percentage change in the treatment group minus the percentage change for the controls. This differs from $\exp(\delta) - 1$. The reason is that the percentage change in the treatment group is equal to $\%effect + \%time + \%effect \times \%time$. If we subtract the percentage change in the control group, we are left with $\%effect \times (1 + \%time)$. The difference is likely to be negligible if $\%time$ is small.
- 5 Note that even under statistical independence the OLS estimator on the log-linearised model could still be biased for the intercept term.
- 6 It is worth noticing that such a shock in the treated group during the *post* period would induce a bias even if the true treatment effect on the mean was zero.
- 7 Practically, PPML can be implemented in the most popular statistical packages and results can be easily interpreted. In Stata™, one can simply run the *poisson* command, with the dependent variable in levels.
- 8 An alternative and equivalent way to fit the model is by GLM with a log link function.
- 9 Table 1 reports results for the main coefficients of interest only (treatedxpost and treated). Results for the full model are included in Appendix D.
- 10 When stronger patterns of heteroskedasticity were introduced to the model, for example with a simulation with a constant variance of y , the performance of the Park test improves with rejection rates reaching 72%.
- 11 Notice, instead, that the estimate is close to the cross-difference from eq. (11), £4.73, which is not the treatment effect.
- 12 Another simple and commonly used correction is to collapse the data into a pre and post period. This works because averaging a linear model gives the sum of the average of each term. This is not always appropriate for a non-linear model and so cannot be used in the multiplicative setting.

References

- Ai, C., and E. C. Norton. 2003. “Interaction Terms in Logit and Probit Models.” *Economics Letters* 80: 123–129.
- Ai, C., and E. C. Norton. 2008. “A Semiparametric Derivative Estimator in Log Transformation Models.” *Econometrics Journal* 11: 538–553.
- Angrist, J. D. 2001. “Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice.” *Journal of Business and Economic Statistics* 19: 2–16.
- Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.
- Athey, S., and G. W. Imbens. 2006. “Identification and Inference in Nonlinear Difference-in-Differences Models.” *Econometrica* 74: 431–497.
- Bertrand, M., E. Duflo, and S. Mullainathan. 2004. “How Much Should We Trust Differences-in-Differences Estimates?” *The Quarterly Journal of Economics* 119: 249–275.
- Blackburn, M. L. 2007. “Estimating Wage Differentials Without Logarithms.” *Labour Economics* 14: 73–98.
- Buis, M. L. 2010. “Stata Tip 87: Interpretation of Interactions in Non-Linear Models.” *The Stata Journal* 10: 305–308.
- DellaVigna, S., and E. Kaplan. 2007. “The Fox News Effect: Media Bias and Voting.” *The Quarterly Journal of Economics* 122: 1187–1234.
- Donald, S. G., and K. Lang. 2007. “Inference with Difference-in-Differences and Other Panel Data.” *The Review of Economics and Statistics* 89: 221–233.
- Finkelstein, A. 2007. “The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare.” *The Quarterly Journal of Economics* 122: 1–37.
- Gregg, P., J. Waldfogel, and E. Washbrook. 2006. “Family Expenditures Post-Welfare Reform in the UK: Are Low-Income Families Starting to Catch Up?” *Labour Economics* 13: 721–746.
- Karaca-Mandic, P., E. C. Norton, and B. Dowd. 2012. “Interaction Terms in Nonlinear Models.” *Health Services Research* 47: 255–274.
- Manning, W. G. 1998. “The Logged Dependent Variable, Heteroscedasticity, and the Retransformation Problem.” *Journal of Health Economics* 17: 283–295.
- Manning, W. G., and J. Mullahy. 2001. “Estimating Log Models: To Transform or Not to Transform?” *Journal of Health Economics* 20: 461–494.
- Mullahy, J. 1997. “Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior.” *The Review of Economics and Statistics* 79: 586–593.
- Mullahy, J. 1999. “Interaction Effects and Difference-in-Difference Estimation in Log-Linear Models.” NBER Technical Working Paper 245.
- Nunn, N., and N. Qian. 2011. “The Potato’s Contribution to Population and Urbanization: Evidence from a Historical Experiment.” *The Quarterly Journal of Economics* 126: 593–650.
- Puhani, P. A. 2012. “The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear “Difference-in-Differences” Models.” *Economics Letters* 115: 85–87.
- Ramsey, J. B. 1969. “Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis.” *Journal of the Royal Statistical Society. Series B (Methodological)* 31: 350–371.
- Santos Silva, J. M. C., and S. Tenreiro. 2006. “The Log of Gravity.” *The Review of Economics and Statistics* 88: 641–658.
- Wooldridge, J. 1992. “Some Alternatives to the Box-Cox Regression Model.” *International Economic Review* 33: 935–955.
- Wooldridge, J. M. 2003. “Cluster-Sample Methods in Applied Econometrics.” *American Economic Review* 93: 133–138.
- Wooldridge, J. M. 2006. “Cluster-Sample Methods in Applied Econometrics: An Extended Analysis.” Working Paper, Michigan State University, Department of Economics.

Supplemental Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/jem-2016-0011>).