ISER Working Paper Series

Dif-in-dif estimators of multiplicative treatment effects

Paul Fisher
Institute for Social and Economic Research
University of Essex

Emanuele Ciani
Bank of Italy, Regional Economic Research Division, Florence
Centre for the Analysis of Public Policies, University of Modena and Reggio Emilia

No. 2014-14
March 2014

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

www.iser.essex.ac.uk

Non-technical summary

A researcher may be interested in analysing a policy intervention where the outcome of interest is continuous, such as wages or expenditure. Difference-in-differences (dif-in-dif) is a popular method applied to analyse such interventions. A common way for the researcher to proceed is to take the logarithm of the outcome under study (log-linearisation).

In this paper, we build on existing results scattered in the literature, to point out particular problems of log-linearisation when using a dif-in-dif approach. We argue for the use of an alternative strategy (pseudo maximum likelihood estimation) that works with the non-transformed outcome on the grounds that:

1. Log-linearisation complicates the application of the standard dif-in-dif assumptions, which refer to the non-transformed outcome and not its logarithm. We document that this point is typically not discussed in the dif-in-dif literature.
2. Even if the standard assumptions are correct, they may not be sufficient to identify the policy effect of interest by using the standard method (OLS) on the logarithm of the outcome. This occurs when there are policy effects that go beyond effects on the average of the outcome (eg. the policy impacts on the variance of the outcome).
3. We show that if the response to a policy intervention differs across affected individuals, then results from a log-transformed model do not generally have a meaningful interpretation.

We provide both a simulation exercise and evidence from an original real world policy intervention – the introduction of the Educational Maintenance Allowance in the UK - to illustrate our arguments.

Results from our simulation show that the commonly used methods can lead to misleading estimates of the effect of interest, even in the presence of small increases in variance associated with the policy intervention. An alternative approach based on a non-linear estimator, on the other hand, is capable of correctly estimating the average effect of the policy.

In our real world policy example, we study the effects of the Educational Maintenance Allowance (EMA) on household expenditure patterns. EMA was a conditional cash transfer introduced in the UK in 2004 and paid to young people staying on in post-16 full-time education. Results from log-linearisation suggest that EMA had no impact on households' expenditure, while our preferred method shows an increase in transport spending.

Dif-in-dif estimators of multiplicative treatment effects*

Emanuele Ciani[a,b] and Paul Fisher[†c]

[a]Bank of Italy, Regional Economic Research Division, Florence Branch
[b]Centre for the Analysis of Public Policies, University of Modena and Reggio Emilia
[c]Institute for Social and Economic Research, University of Essex, UK

November 14, 2013

Abstract

We consider a difference-in-differences setting with a continuous outcome, such as wages or expenditure. 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. Furthermore, it is preferable to use a non-linear estimator, because running OLS on the log-linearised model might confound distributional and mean changes. We illustrate the argument with an original empirical analysis of the impact of the UK Educational Maintenance Allowance on households' expenditure, and with a simulation exercise.

JEL: C21, C51, I38

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

*We wish to thank João Santos Silva, Marco Francesconi, Mike Brewer, Susan Harkness, Jonathan James, Iva Tasseva, Vincenzo Mariani, Juan Hernandez, Roberto Nisticò, Massimo Baldini, Ben Etheridge, Ludovica Giua and seminar participants at Essex for useful comments. Paul Fisher received financial support from the ESRC. Emanuele Ciani wrote part of this paper during his PhD at Essex, for which he received financial support from the ESRC and from the Royal Economic Society Junior Fellowship. 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 the Expenditure and Food Survey has been accessed through the UK Data Archive.

†Corresponding author: pfishe@essex.ac.uk; +44 (0)12068 73994; ISER, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, United Kingdom.

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 concerns, like skewness, and related estimation problems. 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.

To explore the attention received by this issue in the dif-in-dif literature, we reviewed papers published in one top journal with an empirical focus, the Quarterly Journal of Economics, between 2001-2011. A table with complete references is available in Appendix A. In total, 25 papers using a dif-in-dif estimator with continuous outcomes were found. 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 other 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 in the dependent variable, 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.[1]

[1] Whilst this literature review illustrates the possible extent of the problem in the dif-in-dif literature, we do not argue that estimates of treatment effects in the dif-in-dif setting will always be misleading. As we discuss in section 2, this is an empirical question and will depend on the

Previous theoretical literature on non-linear dif-in-dif mostly focused on the interpretation of the interaction effect. Mullahy (1999) discussed the case of a log-linearised exponential model. Ai and Norton (2003) showed that in non-linear models the marginal effect of the interaction term is not directly related to its coefficient in the linear index. However, Puhani (2012) recently argued that their way of calculating the marginal effect is not the correct one for the dif-in-dif case. 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) showed that the OLS estimator of the log-linearised model may not be consistent for the parameter of interest. Blackburn (2007) discussed how to estimate wage differentials without using logarithms.

In this paper we attempt to reconcile the two streams of research for the dif-in-dif case. Our main aim and contribution is to recollect in a unified setting a number of results that are scattered in the literature, in order to provide the practitioner with a clear guide on the choice of modelling and estimation. Using a potential outcome framework, we reinterpret 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.[2]

Specifically, we point out in section 2 that the choice between an exponential or a level model is essentially related to the common trends assumption. Differently, whether the treatment effect is multiplicative or additive does not make a large

particular research question and data set under study.

[2]Bertrand et al. (2004) discussed how serial correlation may severely bias inference in dif-in-dif, because conventional standard errors are likely to underestimate the true standard deviation. We do not discuss how to account for this problem in exponential models. However, the main example throughout their paper has $log\,(wage)$ as the dependent variable, so that the problems discussed here also apply in their context. Given that they proposed to collapse data over the pre-treatment and post-treatment period, further research might try to understand whether averaging logs introduces a different source of bias.

difference, at least from an $ex-post$ evaluation perspective. Although this follows from different results available in the literature, we could not find a reference that made this important point explicit. In terms of estimation, building on Santos Silva and Tenreyro (2006), but focusing on the dif-in-dif case, we set out the restrictive conditions under which treatment effects from the log-linear model are equivalent to treatment effects from an exponential specification. Our favourite interpretation of this problem is that the log-linearised estimates of a multiplicative treatment effect may confound distributional effects with shifts in the mean.

Fortunately, different authors (Mullahy, 1997; Santos Silva and Tenreyro, 2006; Blackburn, 2007) have pointed out that to estimate a multiplicative effect there is no need to log-linearise, because a simple and robust non-linear estimator (Poisson Pseudo-Maximum Likelihood) is available. Although Gregg et al. (2006) noted that it is possible to recover a percentage treatment effect from linear OLS in levels, we point out that one cannot give a causal interpretation to both the additive and multiplicative model. We also correct their calculation in order to properly account for a multiplicative time trend.

We finally show that, in the case of heterogeneous effects, the exponential dif-in-dif model with a conditional mean assumption does not identify the average multiplicative effect, but rather the multiplicative effect on the average. Moreover, the necessary conditions for the latter to be consistently estimated using log-linearisation are less likely to hold with heterogeneous effects. To the best of our knowledge, these two results were not discussed in the dif-in-dif literature, although they are partially related to a comment by Angrist (2001).

In section 3 we present an original applied example. We study the impact on households' expenditure of the introduction of the Educational Maintenance Allowance (EMA) in the UK. In section 4 we present a simulation to illustrate our main arguments. Section 5 concludes and summarises the discussion in terms of a guideline for practitioners.

2 Model specification and inference

A practitioner willing to estimate a dif-in-dif model with a continuous outcome, such as wages or expenditure, usually faces three main decisions:

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 model? Shall I take logs?

3. What kind of average effect is identified by my model?

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

2.1 Multiplicative or additive effects?

The simplest, though quite popular dif-in-dif setting involves two groups ($g \in$ *{control, treated}*) and two time periods ($t \in$ *{pre,post}*), with only one group actually receiving the treatment in the second period. In this paper, we analyse the case of a continuous outcome y, such as earnings or consumption.

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.

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}. In levels, we would state (Angrist and Pischke, 2009):

$$E\left[y_{1igt}|g,t\right] = E\left[y_{0igt}|g,t\right] + \delta^* = \mu_g^* + \lambda_t^* + \delta^*. \tag{1}$$

where we combine an additive common trends assumption with an additive treatment effect.[3]

Differently, one might specify an exponential model

$$E\left[y_{0igt}|g,t\right] = exp\left(\mu_g + \lambda_t\right) \tag{2}$$

where the assumption of common trends is in multiplicative form.[4] Over time, the outcome in the absence of treatment would increase by the same percentage ($exp\left(\lambda_{post} - \lambda_{pre}\right) - 1$) in both groups. Now we can assume a proportional treatment effect:

$$\frac{E\left[y_{1igt}|g,t\right] - E\left[y_{0igt}|g,t\right]}{E\left[y_{0igt}|g,t\right]} = exp\left(\delta\right) - 1. \tag{3}$$

where δ is a parameter on the linear index of the exponential model. We do not need a constant proportional treatment effect to identify the quantity on the left hand side of (3). However, if there are heterogeneous effects, $exp\left(\delta\right) - 1$ is not the average of the individual multiplicative effects, but rather a multiplicative effect on the average. More precisely, it identifies the multiplicative effect on the average for the treated group. A similar discussion, related to IV estimation of an exponential model with treatment effects, can be found in Angrist (2001, pg. 9). We return to this issue in section 2.3.

To be precise, the key difference between the exponential model and the linear one is in the common trends assumption. The choice of a multiplicative or additive treatment effect plays a less important role. If we are only interested in the *ex-post* evaluation problem, in the spirit of DiNardo and Lee (2011), we may just want to understand which share of the treated-control difference should be attributed to the treatment. With multiplicative time trends, we still need the counterfactual to be

[3]The superscript * is used to differentiate the model in levels from the multiplicative one. Note also that receiving the treatment, with a potential outcome y_{1igt}, does not coincide with being in the treated group (g=*treated*), because in the first period (t=*pre*) all individuals go untreated.

[4]Mullahy (1997), reprised in Angrist (2001), proposed an exponential model for a multiplicative treatment effect, but focused on IV estimation.

specified as in eq. (2), otherwise we would confound time and treatment effects.

To clarify, figure 1 is generated with an exponential model as in eq. (2).[5] 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. However, it does not matter whether we express the effect as a percentage difference or as a level difference. Indeed, the former is the fraction on the left hand side of eq. (3), 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.

The situation is different if we are willing to predict how the policy will affect future outcomes. If we believe that the treatment is likely to have the same proportional effect in other time periods, then it should be presented in percentage form. Otherwise, the focus should be on the level difference $E\left[y_{1igt}|g,t\right]-E\left[y_{0igt}|g,t\right]$, again after accounting appropriately for the multiplicative time trend. Caution should be paid here, as it is not always clear how to perform such predictive analysis using dif-in-dif results.

With this caveat in mind, for the case of multiplicative effects the full structure for y_{1igt} is

$$E\left[y_{1igt}|g,t\right] = exp\left(\mu_g + \lambda_t + \delta\right). \tag{4}$$

Intuitively, the total percentage change in the expected outcome of the treated group is given by the composition 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)×(1+%effect)*. Differently, for the control group we have *(1+%change) = (1+%time)* .

Define the dummies $treated_{it}$ for the treatment group and $post_{it}$ for the second

[5] $\mu_{control} = -0.3; \mu_{treated} = -0.7; \lambda_{pre} = -0.2; \lambda_{post} = 0; \delta = 0.2.$

period. The particular data structure leads to an exponential model for observed outcomes

$$E\left[y_{it}|treated_{it}, post_{it}\right] = exp\left(\beta_0 + \beta_1 treated_{it} + \beta_2 post_{it} + \delta treated_{it} \times post_{it}\right) \quad (5)$$

$$\beta_0 \equiv \mu_{control} + \lambda_{pre};\ \beta_1 \equiv \mu_{treated} - \mu_{control};\ \beta_2 \equiv \lambda_{post} - \lambda_{pre}. \quad (6)$$

Although Ai and Norton (2003) showed that we should be careful when looking at the interaction term in non-linear models, here the coefficient on $treated_{it} \times post_{it}$ has a meaningful interpretation. Indeed, $exp\left(\delta\right)$ is a ratio of ratios (ROR), as highlighted by Mullahy (1999) in his discussion about the interpretation of the interaction term in log-linear dif-in-dif models:[6]

$$exp\left(\delta\right) = \frac{E\left[y_{it}|treated_{it} = 1, post_{it} = 1\right]}{E\left[y_{it}|treated_{it} = 1, post_{it} = 0\right]} \Big/ \frac{E\left[y_{it}|treated_{it} = 0, post_{it} = 1\right]}{E\left[y_{it}|treated_{it} = 0, post_{it} = 0\right]}. \quad (7)$$

Differently, the marginal effect of the interaction term would be the cross difference (Mullahy, 1999, pg. 7):

$$\frac{\Delta^2 E\left[y_{it}|treated_{it}, post_{it}\right]}{\Delta treated_{it} \Delta post_{it}} =$$
$$\left[exp\left(\beta_0 + \beta_1 + \beta_2 + \delta\right) - exp\left(\beta_0 + \beta_1\right)\right] - \left[exp\left(\beta_0 + \beta_2\right) - exp\left(\beta_0\right)\right]. \quad (8)$$

which is actually equal to the difference in difference estimand for the additive effects model.

Given the assumption of multiplicative effects, this cross-difference does not properly account for the time trend in the exponential model.[7] Therefore, in a multiplicative model the causal parameter of interest is the ROR. This point is related

[6]Similarly, Buis (2010) pointed out that in an exponential model the interaction term should be interpreted in a multiplicative scale.

[7]Mullahy (1999, pg. 12) warned the reader that the marginal effect and the ROR are related to two "different sense(s) of interaction". Here, we argue that the specification of the common trend assumption in a multiplicative or additive form is crucial in deciding which one to give a causal interpretation.

to the more general comment by Puhani (2012) that, 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.[8]

It should be noted that, when applied to the specific data structure, a linear model for the conditional expectation of y_{it} is also correctly specified, because it is saturated:

$$E\left[y_{it}|treated_{it}, post_{it}\right] = \gamma_0 + \gamma_1 treated_{it} + \gamma_2 post_{it} + \tau treated_{it} \times post_{it}. \qquad (9)$$

Indeed, the exponential model is just a reparametrisation of the linear one, with

$$exp\left(\delta\right) - 1 = \frac{\left(\gamma_0 + \gamma_1 + \gamma_2 + \tau\right) / \left(\gamma_0 + \gamma_1\right)}{\left(\gamma_0 + \gamma_2\right) / \gamma_0} - 1. \qquad (10)$$

This was noted by Gregg et al. (2006), who showed that we can estimate eq. (9) and then recover both the level and the percentage (multiplicative) effect. However, Gregg et al. (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 $exp\left(\delta\right) - 1$. The reason is that the percentage change in the treatment group is equal to *%effect+%time+%effect*× *%time*. If we subtract the percentage change in the control group, we are left with *%effect*× *(1+%time)*. The difference is likely to be negligible if *%time* is small.

In spite of the equivalence in (10), we cannot interpret both τ and δ as causal effects. If we believe that the common trends assumption holds in multiplicative terms, τ 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.

More generally, the equivalence (10) does not work if we are willing to condition on other covariates, such as demographic controls. The reason is that the equation

[8] Karaca-Mandic et al. (2012) agreed with Puhani (2012), mentioning that the interaction term is directly interpretable in non-linear dif-in-dif models when both *treated* and *post* are kept constant.

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 time effects are crucial for causal interpretation is related to Angrist and Pischke (2009, pg. 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

If one decides to focus on the multiplicative effect, in the simplest case we can recover it from linear estimates using eq. (10). However, this method does not work if we want to include other covariates or a time trend.

A popular alternative is to log-linearise the model. To understand the pros and cons of this strategy, we can follow the discussion in Santos Silva and Tenreyro (2006).[9] Define an error term η_{it}:

$$y_{it} = exp\left(\beta_0 + \beta_1 treated_{it} + \beta_2 post_{it} + \delta treated_{it} \times post_{it}\right)\eta_{it} \tag{11}$$

$$E\left[\eta_{it}|1, treated_{it}, post_{it}\right] = 1 \tag{12}$$

Consistently with previous section, the individual-transitory error term is mean independent of group and time. Similar to the standard linear dif-in-dif, we do not need full statistical independence to identify the treatment effect (Abadie, 2005; Athey and Imbens, 2006).

[9] Other approaches to this problem can be found in Mullahy (1998), Manning and Mullahy (2001), Blackburn (2007).

To estimate the model, we can log-linearise it

$$lny_{it} = \beta_0 + \beta_1 treated_{it} + \beta_2 post_{it} + \delta treated_{it} \times post_{it} + ln\eta_{it}. \tag{13}$$

However, as argued by Santos Silva and Tenreyro (2006) and Blackburn (2007), nothing ensures that $E\left[ln\eta_{it}|1, treated_{it}, post_{it}\right] = 0$. In general, this would be true if η_{it} is statistically independent from $x_{it} \equiv (1, treated_{it}, post_{it})$, which implies:[10]

$$Var\left[y_{it}|1, treated_{it}, post_{it}\right] = \sigma_{it}^2 exp\left(2\beta_0 + 2\beta_1 treated_{it} + 2\beta_2 post_{it} + 2\delta treated_{it} \times post_{it}\right) \tag{14}$$

where $\sigma_{it}^2 = Var\left(\eta_{it}\right)$. The ratio of variances between different groups or time periods should be directly related to the differences in the conditional mean. Furthermore, the treatment effect must not only shift the conditional mean, but also increase (or decrease) the conditional variance by a factor equal to the square of $exp\left(\delta\right)$.[11] This pattern of variance does not necessarily hold under the weaker condition of mean independence ($E\left[\eta_{it}|x_{it}\right] = 1$), which is sufficient to identify the multiplicative effect.

For instance, suppose that the condition $\eta_{it} \perp\!\!\!\perp x_{it}$, holds in the absence of the treatment, that is when $treated_{it} \times post_{it} \neq 1$. However, assume that the treatment has a distributional effect which differs from the simple increase in variance by $exp\left(2\delta\right)$. We can express this by stating that

$$\frac{Var\left[y_{it}|treated_{it} = 1, post_{it} = 1\right]}{Var\left[y_{it}|treated_{it} = 1, post_{it} = 0\right]} \neq exp\left(2\beta_2 + 2\delta\right). \tag{15}$$

[10]Whether statistical independence holds depends on the particular application. For instance, we replicated a study by Aguila et al. (2011) who used dif-in-dif on panel data to estimate the effect of the retirement of the household's head on total expenditure and expenditure on food. Results are only slightly affected by directly estimating the exponential model by PPML rather than log-linearising it. Full results are available on request. We thank Emma Aguila for providing us the data and the original do-files.

[11]A particular case when this pattern of variance would arise is one where the treatment effect multiplies each single individual outcome by $exp\left(\delta\right)$, that is if $y_{1igt} = y_{0igt} \times exp\left(\delta\right)$. Clearly, the pattern in (14) can arise in other specific cases.

Higher moments can be affected by the treatment as well. Even if the conditional expectation of lny_{it} is correctly specified because the model is saturated, the coefficient on the interaction $treated_{it} \times post_{it}$ would not be equal to the parameter of interest δ:

$$E\left[lny_{it}|1, treated_{it}, post_{it}\right] = \beta_0^* + \beta_1 treated_{it} + \beta_2 post_{it} + \delta^* treated_{it} \times post_{it} \tag{16}$$

$$\beta_0^* = \beta_0 + E\left[ln\eta_{it}|treated_{it} \times post_{it} \neq 1\right] \tag{17}$$

$$\delta^* = \delta + E\left[ln\eta_{it}|treated_{it} \times post_{it} = 1\right] - E\left[ln\eta_{it}|treated_{it} \times post_{it} \neq 1\right]. \tag{18}$$

OLS estimates for the log-linearised model would therefore be consistent for δ^*, which is confounding distributional with mean effects.[12] A similar bias would arise if the treatment had no effect at all on the outcome distribution, but in the second period there was some change in the variance of y within the treatment group that violates the assumption $\eta_{it} \perp\!\!\!\perp x_{it}$. Such a situation would be compatible with the multiplicative common trends assumption stated in terms of conditional mean (eq. 2), because it does not impose any restriction on higher moments.

The estimator of interest might not be affected by a situation as in Blackburn (2007), where the conditional variance across groups does not follow the pattern in eq. (14), but the condition is respected over time within the same group. Suppose that $E\left[\eta_{it}|x_{it}\right] = 1$. However, assume that the variance and higher moments in the distribution of η_{it} depend on the group, though neither on the time period, nor on the treatment. In general, we would have that

$$\begin{aligned} E\left[ln\eta_{it}|treated_{it} = 0, post_{it} = 0\right] &= E\left[ln\eta_{it}|treated_{it} = 0, post_{it} = 1\right] \\ \neq E\left[ln\eta_{it}|treated_{it} = 1, post_{it} = 0\right] &= E\left[ln\eta_{it}|treated_{it} = 1, post_{it} = 1\right]. \end{aligned} \tag{19}$$

[12]This would hold even if the true treatment effect on the mean was zero, and there was no difference across groups or time ($\beta_1 = \beta_2 = 0$).

Therefore, both the intercept and the coefficient on the group dummy ($treated_{it}$) will be different from β_0 and β_1, but the coefficient on the interaction would be the true treatment effect.[13]

Nevertheless, we know from the literature that there is an alternative estimation strategy which is consistent in both cases, because it only requires η_{it} to be mean independent from x_{it}, and not necessarily statistically independent. Santos Silva and Tenreyro (2006) and Blackburn (2007) proposed to directly estimate the non-linear model.[14] In practice, one can use both Non Linear Least Squares (NLS) and Poisson Quasi Maximum Likelihood (PPML), which are both consistent as long as the conditional mean is correctly specified. Santos Silva and Tenreyro (2006) argued in favour of the latter, because NLS is likely to be less efficient. PPML can be implemented in the most popular statistical packages and results can be easily interpreted.[15] In Stata™, one can simply run the *poisson* command, with all variables in levels. Although we do not need $Var\left[y_{it}|x_{it}\right]$ to be as in eq. (14) for PPML to be consistent, a different pattern of heteroskedasticity would make standard inference invalid. Hence, the robust covariance matrix should be used.

One important point to highlight is that, in the potential outcomes model, we imposed assumptions only on the conditional expectations. This can be justified by the fact that we are often interested only on the average. Athey and Imbens (2006) proposed instead a generalised dif-in-dif model that gives a structural interpretation

[13]An example comes from the study by Meyer (1995). They used dif-in-dif to estimate how workers' compensation affect time out of work, exploiting the fact that Kentucky introduced a change in the benefit for the high earning group. We replicated their analysis by directly estimating the exponential model by PPML, instead of log-linearising. The point estimate of the change in Kentucky is basically unchanged (although it loses statistical significance). However, we observe a large difference in the estimate for the "high earnings" dummy. Full results are available on request. The original microdata were obtained from Wooldridge's dataset (http://ideas.repec.org/p/boc/bocins/injury.html).

[14]For the cross-sectional case, Mullahy (1997) proposed a GMM estimator for an exponential model when an instrument for treatment status is available.

[15]An applied example can be found in Santos Silva and Tenreyro (2010), who estimated the effect of the introduction of the Euro on trade by using PPML. Blackburn (2010) also estimated an exponential model for the effect of migration on earnings growth using panel data, which can be seen an application of the dif-in-dif setup with longitudinal data. Both articles are focused only on the empirical exercise and do not discuss the issues that we cover in this paper.

to all differential changes in the distribution of the outcome y over time. Their assumptions on the model of y would therefore be valid for any $f(y)$, where $f(\cdot)$ is a strictly monotone transformation (such as log). Differently, in this paper we give a structural interpretation only to changes in the expected value. We ignore higher moments of the distribution of y, which are allowed to change either as a consequence of treatment or time. As noted by Athey and Imbens (2006, pg. 435-436), this 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.

2.3 Heterogeneous effects

One important question is what the level and multiplicative model are effectively identifying when treatment effects are heterogeneous. In the level model, given the additive nature of the effects, the well know result is that the dif-in-dif estimand identifies

$$E\left[y_{1igt} - y_{0igt} | g = treated\right] \tag{20}$$

To discuss the multiplicative model, we follow the IV-exponential model in Angrist (2001) and we include an individual effect ω_i and a heterogeneous treatment effect δ_i:[16]

$$y_{0igt} = exp\left(\lambda_t + \omega_i\right)\eta^*_{it}, \tag{21}$$

$$y_{1igt} = y_{0igt} exp\left(\delta_i\right) = exp\left(\lambda_t + \omega_i + \delta_i\right)\eta^*_{it}, \tag{22}$$

$$E\left[\eta^*_{it} | t, \omega_i, \delta_i\right] = 1; \tag{23}$$

Similarly, Blackburn (2007, pg. 91-92) discussed how to include individual-fixed effects in an exponential conditional mean, in order to correctly estimate wage dif-

[16]Here the superscript * is used to differentiate the error η^*_{it} in the unobservable model from the error η_{it} in the observable model.

ferentials. Differently from his paper, we also allow individual heterogeneity in the treatment effects and we mainly focus on the estimation of their average.

As in the standard dif-in-dif, we need the composition of both groups not to change over time, so that the expected values of the individual effects are stable (see Blundell and Macurdy, 1998). The (unobservable) model generating the outcome becomes

$$y_{it} = exp\left(\lambda_t + \omega_i + \delta_i \mathbf{1}\left[g = treated, t = post\right]\right)\eta^*_{it}. \tag{24}$$

If we use PPML to estimate the observable model

$$y_{it} = exp\left(\beta_0 + \beta_1 treated_{it} + \beta_2 post_{it} + \delta treated_{it} \times post_{it}\right)\eta_{it}. \tag{25}$$

we know that $exp\left(\delta\right)$ identifies the ROR:

$$exp\left(\delta\right) = \frac{E\left[y_{it}|treated_{it} = 1, post_{it} = 1\right]}{E\left[y_{it}|treated_{it} = 1, post_{it} = 0\right]} / \frac{E\left[y_{it}|treated_{it} = 0, post_{it} = 1\right]}{E\left[y_{it}|treated_{it} = 0, post_{it} = 0\right]} \tag{26}$$

Given the model in (24), it follows that

$$\begin{aligned} & exp\left(\delta\right) - 1 = \\ & \frac{exp\left(\lambda_{post}\right) E\left[exp\left(\omega_i\right) exp\left(\delta_i\right)|g = treated\right]}{exp\left(\lambda_{pre}\right) E\left[exp\left(\omega_i\right)|g = treated\right]} / \frac{exp\left(\lambda_{post}\right) E\left[exp\left(\omega_i\right)|g = control\right]}{exp\left(\lambda_{pre}\right) E\left[exp\left(\omega_i\right)|g = control\right]} - 1 \\ & \qquad = \frac{E\left[y_{1igt}|g = treated\right] - E\left[y_{0igt}|g = treated\right]}{E\left[y_{0igt}|g = treated\right]} \end{aligned} \tag{27}$$

so that only the multiplicative effect on the average is identified, and not the average of the multiplicative effect. This was already noted by Angrist (2001) for IV estimation of an exponential model.

If we assume the stronger condition that $\eta^*_{it} \perp\!\!\!\perp (t, \omega_i, \delta_i)$, we can log-linearise the fixed effect model in (24):

$$lny_{it} = \lambda_t + \omega_i + \delta_i \mathbf{1}\left[g = treated, t = post\right] + ln\eta^*_{it}. \tag{28}$$

where statistical independence ensures that $ln\eta^*_{it}$ is mean-independent from t, ω_i and δ_i, so that (28) represents a conditional expectation. In this case, the standard dif-in-dif level regression applied to lny_{it} identifies

$$\begin{aligned}&\left\{E\left[lny_{it}|treated_{it} = 1, post_{it} = 1\right] - E\left[lny_{it}|treated_{it} = 1, post_{it} = 0\right]\right\}\\ &\quad - \left\{E\left[lny_{it}|treated_{it} = 0, post_{it} = 1\right] - E\left[lny_{it}|treated_{it} = 0, post_{it} = 0\right]\right\}.\end{aligned} \tag{29}$$

which from model (28) is equal to

$$\begin{aligned}&\left\{\lambda_{post} + E\left[\omega_i|g = treated\right] + E\left[\delta_i|g = treated\right] - \lambda_{pre} - E\left[\omega_i|g = treated\right]\right\}\\ &\quad - \left\{\lambda_{post} + E\left[\omega_i|g = control\right] - \lambda_{pre} - E\left[\omega_i|g = control\right]\right\} =\\ &\quad = E\left[\delta_i|g = treated\right]\end{aligned} \tag{30}$$

This quantity, although related to the original parameters, is not of direct interest. The reason is that the individual multiplicative effect is equal to $exp\left(\delta_i\right) - 1$, but in general $E\left[\delta_i|g = treated\right] \neq E\left[exp\left(\delta_i\right) - 1|g = treated\right]$.[17]

The problem is that statistical independence of the error term holds in the *unobserved* model (24). Differently, the error term η_{it} in the *observed* model is likely to be heteroskedastic, as it depends on the distribution of the individual effects δ_i within each group. Using log-linearisation we can recover the causal effect on the logs $E\left[\delta_i|g = treated\right] = E\left[lny_{1igt} - lny_{0igt}|g = treated\right]$ but we cannot use it to go back to the multiplicative effect on the average in the original scale (27).

[17]Clearly, if there is little variation in δ_i, and its average is quite small in absolute value, then the two are likely to be quite similar. But this would be a case in which there is little heterogeneity in the treatment effect, so that there is not a big difference between the individual and the average δ.

In a nutshell, in the presence of heterogeneous effects the standard mean independence assumptions behind the exponential dif-in-dif model only allow us to identify the multiplicative effect on the average for the treated group, and not an average multiplicative effect.[18] The presence of heterogeneous treatment effects is likely to induce a dependence between the error term η_{it} in the observed model and the covariates. Therefore the statistical independence assumption is not likely to hold, and the OLS estimates of the log-linearised model would not recover the quantity of interest.

3 An applied example

To provide an example, we apply the PPML estimator in a dif-in-dif setting to assess the effects of the recent introduction of the Educational Maintenance Allowance (EMA) in the United Kingdom on household expenditures. EMA provided up to £30 per week to 16-18 year olds in low income households, conditional on them attending a full time educational course. Two further bonus payments of up to £100 were available if additional educational targets were met. The policy was introduced nationwide in September 2004, although a pilot took place in 1999-2000. Dearden et al. (2009) provide evidence suggesting that EMA pilots increased post 16 schooling by 5-7%. Further details of the reform can be found in Appendix B.

Here we study how families targeted by the scheme spent the available resources. In line with the theory of the previous section, we specify a multiplicative model for expenditure and present dif-in-dif estimates of the effect of EMA on 7 major expenditure categories, using both OLS on the log-linearised values and PPML.

[18]Further restrictions can be imposed. For instance, if one assumes that ω_i and δ_i are statistically independent, then eq. (27) reduces to the average multiplicative effect. But we find this assumption quite implausible. Furthermore, any interpretation in terms of individual multiplicative effect should be carefully evaluated: if for some individuals $y_{0igt} = 0$, then y_{1igt}/y_{0igt} is not defined and therefore the individual semi-elasticity does not exist.

3.1 Data and identification strategy

We take advantage of expenditure data from the first five years of the Expenditure and Food Survey (EFS).[19] Interviews took place across a year. All income and expenditure figures are in weekly equivalents and are expressed in December 2005 terms using the retail price index, available from the Office for National Statistics.

The estimation sample consists of all households with at least one child aged either 14, 15 or 17 and responding to the EFS in one of the first five years (2001-2005) of the survey. The EFS operates on the basis of a financial year (April – March), so that to be precise our sample includes periods from April 2001 until March 2006. The reform coincides with the start of the school year in 2004 (September). Given that we have more than two time periods, we include a full set of year dummies, plus a set for the month of interview to account for seasonality. The departure from the simple 2×2 setting implies that the exponential model is not simply a reparametrisation of the level one.

Treatment and control groups are not defined according to information on education status, which may be endogenous to the reform. Rather, information is used on exogenous age at interview of the household members. The treated group of households is defined to be those where at least one 17 year old is residing, because conditional on having low income they will be eligible to receive EMA.[20] The control group is formed of households where at least one 14 or 15 year old resides, excluding households defined to be in the treated group as above. Table B1 in appendix B demonstrates that, under this definition, treatment and control groups are similar in terms of observable characteristics.

In line with the spirit of this paper, we focus on the issues related to functional form specification and estimation of a multiplicative model. Clearly, accounting

[19] Further details are available in Appendix B.

[20] We exclude 16 and 18 year olds to avoid misclassification, because full information on date of birth would be required to determine EMA eligibility status. The EFS only contains information on age at interview.

properly for the time trend does not imply that estimates can be interpreted as causal effects. Other problems, such as the presence of an Ashenfelter's dip or of anticipation effects, can bias the results. The interested reader can found a more detailed discussion in Appendix B.

Before the UK wide rollout in September 2004, a pilot took place in 41 English Local Education Authorities (LEAs) in years 1999 and 2000 (see Dearden et al., 2009, for an assessment). Therefore, EMA was already in operation in 41 of the 150 English LEAs before the start of our sample period. Pilot areas cannot be removed from the treated group as the EFS does not record information on LEA status. This implies that treatment is less than 100 percent for the treated group and that the presented estimates of the effect of the policy on household expenditure patterns, therefore, represent a lower bound of the effect on those actually receiving EMA. On the one hand, it may be enough for a policy maker to know this intent to treat, on the other, the interest may be in the effect for those that actually received EMA. In Appendix B we comment on the possibility of rescaling the estimated treatment effects to reflect this fact. We show that in the case of a multiplicative model, rescaling may be problematic.

We present estimates for 7 major areas of spending: food and non-alcoholic drinks; alcoholic beverages and tobacco; clothing and footwear; furnishings, household equipment and carpets; transport; communication; and recreation. Following on from the earlier discussion, it is natural to specify the common trends assumption in multiplicative form. That is expenditures, following the growth of the economy, increase by a constant percentage in the absence of treatment.

In general, we know that total household expenditure tends to be log-normally distributed (see, for instance, Battistin et al., 2009). One might claim that, in this case, log-linearisation is harmless. This is not necessarily the case. First of all, the required assumption refer to the conditional distribution of y_{it}, that is within group and time period. Secondly, even if the error η_{it} was log-normal, we would need it to be statistically independent from the covariates. Indeed, in the simulation in section 4 we show that a log-normally distributed disturbance is not enough for log-OLS consistency. Finally, nothing ensures log-normality of each category of spending. This is particularly true if there is a non-negligible proportion of zero, which may be due to measurement error related to the recording of small amounts, as was pointed out for trade data by Santos Silva and Tenreyro (2006).

All estimates are obtained using Stata 11. For OLS we used the standard command `regress`, while PPML estimates are obtained using the command `poisson`. We chose the `robust` option for standard errors.[21]

3.2 Results

Table 1 presents dif-in-dif estimates of the effect of the national roll out of the EMA scheme on each of the 7 major spending categories for the treated group of households. The results in columns 1 correspond to estimates of the multiplicative effect using OLS on the log-linearised model, while in column 2 the reform effect is estimated directly using the PPML estimator. For completeness, we also report OLS estimates for a level model of expenditure in column 3. It is important to stress that, as usual, observations with zero expenditure are dropped from OLS log estimates, while PPML allows us to keep them. Nevertheless, results are quite similar when excluding these cases for all estimators or setting the logarithm equal to zero in the case of zero expenditure (results available on request).

Following the national roll out of EMA in September 2004, we expect the treated group of households to increase expenditures in some of these areas. For the OLS

[21]Procedures to correct for the fact that regular standard errors may overstate the precision of estimates of a treatment effect in dif-in-dif regressions have been the subject of much debate in recent years (see Bertrand et al., 2004; Donald and Lang, 2007; Wooldridge, 2003, 2006). If shocks are common to observations in a given group and year, then the error terms within a group are not independent but correlated. Moreover, in the case of multiple time periods, errors are also likely to be serially correlated. Under the definition of the treatment and control groups in this paper, we see no reason to believe that there are shocks that occur at the group level. Moreover, sample sizes are relatively small and adjustments to standard errors will be conservative if observations are indeed independent.

log-linearised estimates in column 1 we see positive dif-in-dif estimates for food, non-alcoholic drinks; alcoholic beverages and tobacco; transport; recreation and negative effects for clothing footwear; furnishings household equipment and carpets; communication. None of the estimated effects are, however, statistically significant.

Turning to the reform effects in column 2, EMA might also have distributional effects that make the multiplicative error term statistically dependent on the time and group dummies. In this case, we expect the previous OLS log results to suffer from bias, while PPML results should be consistent. For most of the categories, coefficients are in line with the OLS log results, but for transport spending the estimated coefficient has increased in magnitude. Moreover, it is now statistically different from zero at the 5% level. The result implies an increase around 23 percent in transport spending due to the reform, calculated as $exp\left(\widehat{\delta}\right)-1$. This finding is in line with evidence from the EMA piloting, in which EMA recipients were more likely to be contributing to transport expenditures compared to non-recipients and EMA eligibles residing in control areas (see Ashworth et al., 2001, p. 59). In comparison to the standard log expenditure estimates of column 1, the PPML coefficient implies an EMA effect of 10.8 percentage points bigger, which is more precisely estimated. For the remaining spending categories, we observe statistically insignificant coefficients, which are also generally smaller than the effect on transport.[22]

On the OLS level results of column 3, the estimated signs and significance of the interaction terms match well with the PPML results. For transport spending, the $treated \times post$ interaction is a statistically significant £16.46, which corresponds to 18.3 percent of the pre-reform mean. However, the dif-in-dif coefficient presented only corresponds to the causal effect of EMA on the level of expenditure if we are willing to impose common trends in expenditure levels. If the common multiplicative

[22] One criticism might be that the result for transport is accidental, because in a full set of regressions it is not unlikely to find at least one statistically significant estimate, e.g. we have a multiple hypothesis testing problem. However, here we focus on the difference between OLS (logs) and PPML results. Moreover, given the relatively small sample sizes, rather than making adjustments to standard errors, which can be conservative and computationally intensive (Duflo et al., 2008), we draw on the evidence from the EMA trials to give further support to our conclusions.

trend is the correct one, then no meaningful interpretation can be given to the coefficient of the level model.

Columns 4-6 try to better target the groups affected by the reform by repeating the previous analysis for a sample of low income households. Results give further strength to the main finding with PPML estimates in column 5 suggesting that households devoted the additional resources from EMA primarily to transport spending. The PPML estimate increases in magnitude with little reduction in the precision (comparing to column 2). This is once again in contrast to the OLS log result which remains smaller and statistically insignificant.

We compare how the models perform on Ramsey's RESET test (Ramsey, 1969) for misspecification of the conditional mean. It involves calculating the square of the fitted values and including them as an additional regressor. P-values for the significance of this coefficient are reported alongside the main results in table 1. The OLS log, PPML, and OLS level specifications all typically pass the test. For the full sample, no evidence is found of misspecification of the conditional mean whether the estimates are for OLS log, PPML or OLS level. For the low income sample the same picture emerges; however, the test is only marginally passed in the case of the PPML and level estimates for alcoholic beverages and tobacco; and recreation; and the OLS log results for clothing and footwear.

The results of the test illustrate an important point that correct specification is not sufficient for causal interpretation. For example, in the simple 2-period-2-group case, the conditional expectations of both lny and y are correctly specified as linear because the model is saturated. Our choice about which estimates to interpret as causal effects critically depends on our belief about the nature of common trends. Furthermore, under heteroskedasticity the effect estimated with logs might confound mean with distributional effects, even though the model for lny is correctly specified.

4 A simulation

In order to illustrate the theoretical arguments, this section reports results from monte-carlo simulations. A group of hypothetical reforms are considered that have common multiplicative time and treatment effects on the mean, but that differ in terms of their distributional impacts.

We consider a setting similar to that reported in figure 1, so that the reader can follow the difference between the estimates in levels and the ones in multiplicative form. The outcome of interest is generated according to an exponential dif-in-dif equation:

$$E\left(y_{it}|x\right) = exp\left(\beta_0 + \beta_1 treated_{it} + \beta_2 post_{it} + \beta_3 treated_{it} \times post_{it}\right) \qquad (31)$$

i=1,..., 2631. Here, the conditional mean is made up of an intercept β_0, a multiplicative treated group effect β_1, a common multiplicative post-reform effect β_2, and a multiplicative treatment effect equal to the ratio-of-ratios β_3. The sample size and size of the groups are selected to be as in the applied example of section 3, in order to be able to detect similar differences between estimators.

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 $\beta_3 = 0.2$.[23] This set of parameters implies that for the treated group the difference in levels between y_{1i} and y_{0i} (the counterfactual) in the post treatment period would be $exp\left(3.5 - 0.4 + 0.03 + 0.2\right) - exp\left(3.5 - 0.4 + 0.03\right) = £5.06$. This is the part of the change that can be attributed to the reform after accounting properly for the multiplicative trend. It is different from the estimand of a standard dif-in-dif in levels, which would be: $[exp(3.5 - 0.4 + 0.03 + 0.2) - exp(3.5 - 0.4)] - [exp(3.5 + 0.03) - exp(3.5)] = £4.73]$.

A random error term is introduced, so that each individual observation is gener-

[23]Simulations with a negative time trend and a positive difference between groups lead to the same conclusions and are available on request.

ated according to $y_{it} = exp(x_{it}\beta)\eta_{it}$ where η_{it} is a log normal random variable with mean 1, and $var(\eta_{it}|x) = \sigma^2_{it}$. The variance σ^2_{it} is specified as:

$$\sigma^2_{it} = exp(\alpha \times \mathbf{1}(treated_{it} \times post_{it} = 1)) \tag{32}$$

where $\mathbf{1}$ is an indicator function and α a parameter that determines the degree of heteroskedasticity in η_{it}. $\alpha = 0$ implies a multiplicative error term with constant variance equal to 1 while $\alpha \neq 0$ implies a multiplicative error term that is heteroskedastic with respect to treatment. That is, for $\alpha \neq 0$ the treatment operates not only on the conditional mean, but also having an additional independent effect on the conditional error variance.[24] Note that any real value of α is in line with the standard dif-in-dif identifying assumption of common trends (in this case multiplicative) expressed in terms of the conditional mean of y_{it}, which places no restriction on the conditional variance. Furthermore, even if here we are interpreting the additional variance introduced by $\alpha \neq 0$ as a distributional effect of the reform, this may not be the case. If the heteroskedasticity is due to other changes in higher moments over time, the conditional mean independence assumption would still be enough for PPML to be consistent, but would not be sufficient for log-OLS.

To assess the performance of the three estimation strategies outlined above, simulations are reported for five key values of α. The first special case of interest is the scenario where $\alpha = 0$ implying η_{it} is statistically independent of the treatment and other regressors, so that OLS estimates from the log-linear model will provide consistent estimates of the multiplicative treatment effect. In contrast, the other four specifications for the conditional variance consider increasing strengths of heteroskedasticity with $\alpha = 0.1$, $\alpha = 0.2$, $\alpha = 0.3$, and $\alpha = 0.4$. Here, OLS estimates from the log-linear model may confound mean effects with distributional effects, whereas PPML estimates should be consistent for all values of α.

[24]Here, receiving the treatment has an additional effect on the error variance, but being in the treated group not.

For the specified values of α we also consider the case where the researcher wrongly estimates the model in levels, not accounting for the multiplicative time trend. The OLS level models are expected to be biased for the true reform effect, although the degree of heteroskedasticity is not expected to affect the magnitude of the bias but only the usual efficiency properties.

Table 2 reports results from 1000 replications of the simulation procedure. For each value of α, reported are the mean and standard deviations of the estimated treatment effects, in addition to estimates of the treated coefficients (β_1's), which while not the main focus, may also be of interest in a dif-in-dif study. Whilst columns 1 and 2 report estimates of the multiplicative treatment effect, column 3 gives estimates of the level treatment effect from a model not correctly accounting for the multiplicative time trend.

Starting with the case where α=0, as expected both the OLS log in column 1 and PPML estimates in column 2 are close to the true multiplicative treatment effect of 0.2. Whilst the difference between the two estimates is negligible, the OLS log estimates are slightly less dispersed, confirming the greater efficiency of the OLS estimator under statistical independence of the error term.

Moving to the case where $\alpha = 0.1$, where the treatment now has a distributional effect above that due to the simple increase in the conditional mean, we observe that the OLS log procedure performs less well. Here, the OLS log estimates confound the distributional effect of treatment with the mean effect and are biased for the true multiplicative effect. As expected, the value of this bias increases with the value of α, even though the variance of the estimated effects remains small. For example, the mean of the estimated treatment effects being only 43 percent of the true effect when $\alpha = 0.4$. On the other hand, the PPML estimates perform 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 percent, whereas when $\alpha = 0.1$ the increase in standard deviation is just 5 percent. Even when these very small distributional effects of treatment are introduced, the estimates in the table from the log-linearised model are strongly biased.

Column 3 of table 2 presents estimates from standard OLS estimation of the model in levels. From the table, we observe that the estimated treatment effects repeatedly underestimate the true reform impact. The effect is £4.69 in the baseline case, in contrast to the change in levels that is implied by the multiplicative model (£5.06). The bias is independent from the value of α. So although the regression for y_i is saturated and therefore correctly specified, the estimated effect in levels confounds the treatment and trend effects. Note, that the estimate does match well with the actual value of the estimand for the dif-in-dif in levels (4.73), implied by the parameters.

Given an exponential model for the conditional mean of y, a researcher may wish to test whether estimation by log-linearisation will be consistent. Table 2 also presents evidence on the performance of two tests. The first is a Park test (Manning and Mullahy, 2001; Santos Silva and Tenreyro, 2006) for whether the conditional variance of y is proportional to the conditional mean squared. This involves testing whether γ_1 is statistically different from 2 in the equation:

$$(y_i - \widetilde{y}_i)^2 = \gamma_0(\widetilde{y}_i)^{\gamma_1} + \epsilon_i \tag{33}$$

where $\widetilde{y}$ is a consistent estimate of $E[y_i|x]$, obtained using PPML. Inference from equation (33) uses the Eicker-White robust covariance matrix estimator. A rejection of $\gamma_1 = 2$ implies a rejection of the log-linear model. The second test is the standard Breusch-Pagan (BP) test for heteroskedasticity, which is the optimal test for heteroskedasticity when errors are normally distributed. It examines whether the estimated variance of the residuals from the log model is statistically dependent on the value of the *treated* $\times$ *post* variable. Evidence of heteroskedasticity in the model for the log of expenditure is interpreted as a rejection of log-linearisation as an estimation strategy. For both tests, rejection rates at the 5 percent level are reported in the table.

For the Park test, equation (33) is estimated by both log-linerisation (column 1) and directly by PPML (column 2). 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 and the rejection rates are around 5 percent for all values of α.[25] For the BP test, moving away from the baseline scenario where the log-linear model gives consistent estimates of the treatment effect, the test detects the heteroskedasticity introduced into η_i and has reasonable power. For example, in the case where α=0.4 the test detects the inadequacy of the log-linearised specification 94.5 percent of the time. These results suggest that testing for heteroskedasticity in the model for *lny* with a BP test can be informative when deciding upon an estimation strategy with a multiplicative model in the dif-in-dif setting.

In appendix C we also analyse the case in which the treatment has no distributional effect, but the pattern of variance across the treated and control groups does not respect the proportional structure from equation 14. As discussed in the theoretical section, in this case the log-OLS estimator for the treatment effect is consistent, while the treated-control difference is biased.[26]

[25] 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 percent.

[26] We also analysed the case with a constant variance of y. Again here log-OLS for the treatment effect performs poorly. Results are available from the authors.

5 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, using Poisson Pseudo Maximum Likelihood can be preferable to log-linearisation. The reason is that the latter might confound changes in higher moments of the outcome distribution with the treatment effect on the mean.

As a summary, we think that the best practice for a practitioner willing to estimate a dif-in-dif model with continuous outcome should be:

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) Without covariates, the multiplicative treatment effect can be recovered from OLS estimates of the standard dif-in-dif regression in levels (eq. 10).

   (b) 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.

   (c) 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.

(d) In the case of heterogeneous effects, the exponentiated coefficient on the interaction term $(exp\,(\delta)-1)$, can be interpreted as a multiplicative effect on the average for the treated group.

References

Abadie, A., 2005. Semiparametric difference-in-differences estimators. Review of Economic Studies 72(1), 1–19.

Aguila, E., Attanasio, O., Meghir, C., 2011. Changes in consumption at retirement: Evidence from panel data. The Review of Economics and Statistics 93(3), 1094–1099.

Ai, C., Norton, E.C., 2003. Interaction terms in logit and probit models. Economics Letters 80(1), 123–129.

Ai, C., Norton, E.C., 2008. A semiparametric derivative estimator in log transformation models. Econometrics Journal 11(3), 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(1), 2–16.

Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics. Princeton University Press.

Ashworth, K., Hardman, J., Liu, W., Maguire, S., Middleton, S., Dearden, L., Emmerson, C., Frayne, C., Goodman, A., Ichimura, H., Meghir, C., 2001. Education

Maintenance Allowance: The First Year - A Quantitative Evaluation. Department for Education and Employment (DfEE) Research Report No. RR257.

Athey, S., Imbens, G.W., 2006. Identification and inference in nonlinear difference-in-differences models. Econometrica 74(2), 431–497.

Battistin, E., Blundell, R., Lewbel, A., 2009. Why is consumption more log normal than income? Gibrat's law revisited. Journal of Political Economy 117(6), 1140–1154.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? The Quarterly Journal of Economics 119(1), 249–275.

Blackburn, M.L., 2007. Estimating wage differentials without logarithms. Labour Economics 14(1), 73–98.

Blackburn, M.L., 2010. The impact of internal migration on married couples' earnings in Britain. Economica 77(307), 584–603.

Blundell, R., Macurdy, T.E., 1998. Labor supply: a review of alternative approaches. IFS Working Paper Series No. W98/18.

Buis, M.L., 2010. Stata tip 87: Interpretation of interactions in non-linear models. The Stata Journal 10(2), 305–308.

Dearden, L., Emmerson, C., Meghir, C., 2009. Conditional cash transfers and school dropout rates. Journal of Human Resources 44(4).

DellaVigna, S., Kaplan, E., 2007. The fox news effect: Media bias and voting. The Quarterly Journal of Economics 122(3), 1187–1234.

DiNardo, J., Lee, D.S., 2011. Program evaluation and research designs. in Ashenfelter, O., and Card, D., (eds), Handbook of Labor Economics, Elsevier 4A, 463–536.

Donald, S.G., Lang, K., 2007. Inference with Difference-in-Differences and Other Panel Data. The Review of Economics and Statistics 89, 221–233.

Duflo, E., Glennerster, R., Kremer, M., 2008. Using Randomization in Development Economics Research: A Toolkit. Elsevier. volume 4 of *Handbook of Development Economics*. chapter 61. pp. 3895–3962.

Finkelstein, A., 2007. The aggregate effects of health insurance: Evidence from the introduction of medicare. The Quarterly Journal of Economics 122(1), 1–37.

Gregg, P., Waldfogel, J., Washbrook, E., 2006. Family expenditures post-welfare reform in the UK: Are low-income families starting to catch up? Labour Economics 13(6), 721–746.

Karaca-Mandic, P., Norton, E.C., Dowd, B., 2012. Interaction Terms in Nonlinear Models. Health Services Research 47(1pt1), 255–274.

Manning, W.G., 1998. The logged dependent variable, heteroscedasticity, and the retransformation problem. Journal of Health Economics 17(3), 2983–295.

Manning, W.G., Mullahy, J., 2001. Estimating log models: to transform or not to transform? Journal of Health Economics 20(4), 461–494.

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

Mullahy, J., 1997. Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. The Review of Economics and Statistics 79(4), 586–593.

Mullahy, J., 1998. Much ado about two: Reconsidering retransformation and the two-part model in health econometrics. Journal of Health Economics 17(3), 247–281.

Mullahy, J., 1999. Interaction effects and difference-in-difference estimation in log-linear models. NBER Technical Working Paper 245.

Nunn, N., Qian, N., 2011. The potato's contribution to population and urbanization: Evidence from a historical experiment. The Quarterly Journal of Economics 126(2), 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(1), 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(2), 350–371.

Santos Silva, J., Tenreyro, S., 2010. Currency unions in prospect and retrospect. Annual Review of Economics 2(1), 51–74.

Santos Silva, J.M.C., Tenreyro, S., 2006. The log of gravity. The Review of Economics and Statistics 88(4), 641–658.

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 .

Tables and Figures

Figure 1: An example of a dif-in-dif setting with multiplicative time trend

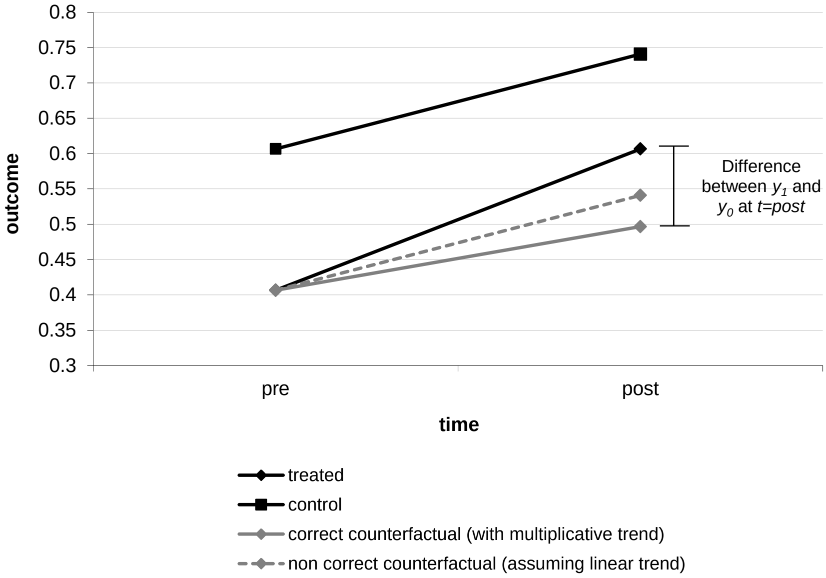

Table 1: Estimates of the EMA Effect on 6 Major Expenditure Categories

	Full Sample			Low Income Sample		
	(1) OLS Log	(2) PPML	(3) OLS Level	(4) OLS Log	(5) PPML	(6) OLS Level
Food and Non-alcoholic drinks	0.069	0.026	1.777	0.039	0.010	0.584
	(0.051)	(0.041)	(2.838)	(0.080)	(0.065)	(3.868)
Observations	2626	2631	2631	1314	1317	1317
Reset(p-value)	0.0927	0.7315	0.8800	0.5150	0.5410	0.5940
Alcoholic Beverages and Tobacco	0.026	0.104	1.276	-0.022	0.178	2.458
	(0.104)	(0.101)	(1.659)	(0.156)	(0.152)	(2.320)
Observations	1968	2631	2631	935	1317	1317
Reset(p-value)	0.8720	0.2829	0.1083	0.4835	0.0107	0.0417
Clothing and Footwear	-0.137	0.012	-0.477	-0.136	-0.024	-1.638
	(0.100)	(0.091)	(4.060)	(0.152)	(0.142)	(5.350)
Observations	2342	2631	2631	1128	1317	1317
Reset(p-value)	0.8158	0.9232	0.3009	0.0337	0.8239	0.3901
Furnishings, HH Equipment, Carpets	-0.126	-0.092	-3.511	-0.150	-0.211	-6.015
	(0.115)	(0.177)	(7.108)	(0.170)	(0.219)	(6.976)
Observations	2575	2631	2631	1282	1317	1317
Reset(p-value)	0.2364	0.5347	0.3370	0.8011	0.1806	0.1282
Transport	0.116	0.208**	16.458**	0.127	0.328**	18.292*
	(0.098)	(0.095)	(8.276)	(0.153)	(0.163)	(9.851)
Observations	2505	2631	2631	1204	1317	1317
Reset(p-value)	0.7219	0.9051	0.6783	0.0936	0.7100	0.3995
Communication	-0.083	-0.032	-0.553	-0.011	0.029	0.559
	(0.064)	(0.074)	(1.406)	(0.095)	(0.118)	(2.027)
Observations	2548	2631	2631	1248	1317	1317
Reset(p-value)	0.4642	0.3287	0.2247	0.6068	0.1991	0.1732
Recreation	0.014	-0.010	-0.633	0.083	0.204	12.676
	(0.081)	(0.092)	(7.745)	(0.118)	(0.150)	(9.522)
Observations	2629	2631	2631	1315	1317	1317
Reset(p-value)	0.8858	0.1458	0.0923	0.4734	0.0194	0.0131

Notes: * p<.10, ** p<.05, *** p<.01. Standard errors (robust) in parentheses. Treated group formed of households with at least one individual aged 17. Control group formed of households with at least one individual aged 15 (excluding households with a 16-18 year old). Columns 1-3 present estimates of the reform effect for the full sample of households, Columns 4-6 present estimates of the reform effect for the subsample of households in the bottom half of the earnings distribution (excluding earnings from 16-18 year olds). Models include a full set of year and month of interview dummies, a treatment status indicator and a post reform indicator interacted with the treatment status dummy (coefficient presented). All expenditure categories are in weekly equivalent and are expressed in December 2005 terms.

Table 2: Simulation Results

	(1) OLS Log	(2) PPML	(3) OLS Level
$\alpha = 0$			
Treated x Post	.1999924	.1984138	4.689477
	(0.0711943)	(0.0842905)	(2.510064)
Treated	-.399073	-.3976382	-10.84626
	(0.0398905)	(0.0474284)	(1.290385)
Park Test	0.052	0.014	-
Breusch-Pagan	0.049	-	-
$\alpha = 0.1$			
Treated x Post	.1710568	.1951523	4.585499
	(0.0731679)	(0.0882464)	(2.62745)
Treated	-.4002561	-.4000494	-10.90821
	(0.0404467)	(0.0470304)	(1.277552)
Park Test	0.04	0.005	-
Breusch-Pagan	0.154	-	-
$\alpha = 0.2$			
Treated x Post	.1506154	.2026701	4.814591
	(0.0727646)	(0.0887447)	(2.625155)
Treated	-.4010721	-.401937	-10.96927
	(0.0406465)	(0.0496038)	(1.345574)
Park Test	0.046	0.017	-
Breusch-Pagan	0.441	-	-
$\alpha = 0.3$			
Treated x Post	.1203693	.1993501	4.727764
	(0.0726566)	(0.0908685)	(2.672351)
Treated	-.3996677	-.3990473	-10.88376
	(0.0386935)	(0.0457816)	(1.232807)
Park Test	0.052	0.014	-
Breusch-Pagan	0.756	-	-
$\alpha = 0.4$			
Treated x Post	.0861068	.1960358	4.632431
	(0.0776181)	(0.096877)	(2.870258)
Treated	-.399795	-.3997392	-10.90645
	(0.0404998)	(0.0484243)	(1.310166)
Park Test	0.064	0.022	-
Breusch-Pagan	0.945	-	-

Notes: Results from 10,000 replications of the simulation procedure described in section 4. Mean of the estimated coefficients reported with standard deviations in parentheses.

Supplementary material (online)

Appendix A: Literature review

We consider as dif-in-dif papers those where the authors explicitly describe their estimation strategy as dif-in-dif or where a policy intervention affects differently periods/groups and a dif-in-dif estimator is implicitly exploited. A paper is recorded as having a continuous outcome if at least one dependent variable is continuous (or discrete but with many mass points, such as hours worked). In cases where multiple outcomes were analysed, we report only the continuous ones or the one with several mass points. A paper is recorded as having a logged outcome if at least one dependent variable undergoes a log transformation.

Table A1: Literature review: dif-in-dif papers with continuous outcome from the QJE 2001-2011

Author and Year	Continuous Outcome Summary	Outcome Logged	Explicit comment on why log transformation
Autor (2001)	hourly wage	✓	
Donohue and Levitt (2001)	crime rate, arrest rate, arrest level	✓	✓
Donohue et al. (2002)	black/white ratio: teacher salaries, pupil/teacher ratio, term length		
Bertrand et al. (2004)	weekly earnings	✓	
Finkelstein (2004)	new vaccine clinical trials		
Morgan et al. (2004)	absolute deviation from conditional mean growth in: gross state product, employment, personal income, state GNP		
Khwaja and Mian (2005)	default rate of firm, loan amount, export value, export value/total loans	✓	
Bailey (2006)	hours/weeks worked		
Stevenson and Wolfers (2006)	homicide rate		
Bandiera et al. (2007)	workers' productivity	✓	
Bleakley (2007)	schooling and earnings	✓	
DellaVigna and Kaplan (2007)	vote share and total vote cast	✓	✓
Field (2007)	weekly hours in labour force		
Finkelstein (2007)	admissions, patient days, beds, payroll expenditures, and total expenditures	✓	✓
Donohue and Levitt (2008)	arrests	✓	
Foote and Goetz (2008)	arrests	✓	
Gruber and Hungerman (2008)	charitable giving	✓	
Verhoogen (2008)	difference over time of an estimated coefficient from OLS and IV regressions of log outcome on log domestic sales	✓	
Björkman and Svensson (2009)	equipment and waiting times		
Jayachandran and Lleras-Muney (2009)	life expectancy, literacy, fertility (birth rate, # births, female pop. Aged 15-45)	✓	
Jensen and Oster (2009)	children enrolment rate		
Hong and Kacperczyk (2010)	coverage and forecast bias		
Hornbeck (2010)	acres of land, productivity, land value	✓	✓
Dittmar (2011)	city growth	✓	
Nunn and Qian (2011)	total population	✓	✓
Number of dif-in-dif papers:	Total	25	
	Logging outcome	16	
	Comment on why logs	5	
	Addressing estimation issues	1	

Appendix B: Further details for the applied example

The policy

EMA was rolled out nationally in the UK from September 2004. This followed on from evidence documenting the poor international performance of the UK in terms of post 16 education take-up (see for example Blanden, Gregg and Machin, 2005). It provided resources to low income young people staying in post compulsory education, with the objective of increasing take-up rates of post 16 schooling. The programme was far reaching - by December of the 2010/11 academic year, 44 percent of 16-18 year olds in full-time education had received an EMA payment or 31 percent of all 16-18 year olds.[1]

EMA required attendance of a full time educational course at a school or college; or a course leading to an apprenticeship or foundation learning programme.[2] The policy was rolled out nationwide in September 2004. Therefore, students in low income households turning age 16 before that date and entering non-advanced education would be eligible for the allowance. The maximum value of the award was worth a considerable £30 per week, which was paid to the bank account of the eligible young person. Further bonus payments were available for meeting educational targets on course attendance, exam attendance and the completion of coursework assignments. These were paid twice annually and worth £100. It was usually claimed for two years of study, ending in the academic year a claimant turned 19. The final value of the award depended on household income and the meeting of educational targets, with the greatest amounts going to the poorest students.[3]

Prior to the national rollout, EMA was piloted in 15 Local Education Authorities

[1]Based on authors calculations from "Participation in Education, Training and Employment by 16-18 Year Olds in England , End 2011", Department for Education; and "Educational Maintenance Allowance Take-up", Young People's Learning Agency.

[2]A course is defined as one with at least 12 teaching hours per week, lasting for at least 10 weeks and be at an institution that is inspected by a public body, to assure quality.

[3]Income of the claimant was not included in the calculation of household income, however, claimants were restricted to engaging in part-time work of at most 24 hours per week. EMA did not interact with other UK benefits.

(LEAs), followed by a further 41 in the year 2000.[4] Therefore, for the national rollout, the pool of recipients consisted of those aged 16 at the start of the 2004/05 academic year (september), 16-17 year olds in 2005/06 academic year and 16-18 year olds in 2006/07 academic year; whilst in the pilot regions, EMA was rolled out from the 1999/2000 academic year. Dearden, Emmerson and Meghir (2009) provide an excellent description of the policy environment and piloting of the programme. Their evidence suggests that the piloting of the scheme had substantial policy effects, with estimates implying a 4.5 percentage point increase in the first year of post 16 schooling and 6.7 percentage points in the second year. Further results indicate that the increased school participation largely comes from those who are not otherwise working and have low prior ability levels.

Data details

The EFS is managed by the Office for National Statistics. The data is available online through the Economic and Social Data Service. The survey changed in 2008 to become the Living Costs and Food Survey. The primary purpose of the EFS is to provide expenditure weights for the consumer and retail price indexes. The survey records all expenditure items for a random sample of UK households. Expenditure items for all individuals aged over 7 in a household are recorded through a detailed expenditure diary over a two week period. Expenditures are then aggregated to the household level and into broad expenditure categories, and finally converted in weekly equivalents. The survey thus provides household level expenditure information for broad expenditure categories and disaggregated expenditures on specific consumption items.

[4] LEAs are local authorities responsible for education. There are 152 LEAs in England and Wales.

Table B1: Pre-Reform Summary Statistics

	Treated	Control	Mean Diff
Household Characteristics			
Number aged 16-18	1.15	0.00	1.15***
HH Labour Income (Less Income from 14-18 year olds)	362.90	353.35	9.55
HH Size	4.09	4.00	0.10
HH Owned	0.69	0.69	0.00
Social Housing	0.22	0.25	-0.03
North East	0.05	0.04	0.01
North West	0.09	0.09	0.00
Merseyside	0.03	0.03	-0.00
Yorkshire and the Humber	0.08	0.08	-0.00
East Midlands	0.06	0.07	-0.00
West Midlands	0.08	0.10	-0.02
Eastern	0.08	0.10	-0.01
London	0.09	0.09	0.01
South East	0.11	0.12	-0.01
South West	0.08	0.07	0.01
Wales	0.06	0.05	0.01
Scotland	0.08	0.08	0.00
Northern Ireland	0.11	0.10	0.01
Expenditures			
Food non-alcoholic drinks and Clothing	69.31	66.32	2.99
Alcoholic Beverages and Tobacco	18.47	15.83	2.65*
Clothing and Footwear	50.55	38.54	12.01***
Furnishings, HH Equipment, Carpets	40.51	40.26	0.25
Transport	90.11	78.83	11.28*
Communication	19.88	15.81	4.08***
Recreation	86.97	87.95	-0.98
Observations	1799		

Notes: * p<.10, ** p<.05, *** p<.01. Control group formed of households with at least one individual aged 14-15 (excluding households with 16-18 year olds). All expenditure categories are in weekly equivalent and are expressed in December 2005 terms.

Possible violations of dif-in-dif assumptions

One of the possible problems in giving a causal interpretation to dif-in-dif estimates is that the treatment group assignment could reflect short-term idiosyncratic shocks, causing an Ashenfelter's dip (see Ashenfelter (1978) and Blundell and Dias (2009)). In such a scenario units assigned to the treatment group may recover more quickly in terms of the outcome of interest, than those in the control group. This is not a likely problem for the identification strategy outlined above, where treatment group assignment is allocated according to information on age.

Another possible source of bias is the presence of "anticipation effects", where households changed their spending behaviour prior to the reform. Pre-reform control households with younger children could potentially adjust their spending behaviour in anticipation of becoming EMA eligible in the post-reform period. This would lead to a downward bias in the estimated reform effects, assuming control households anticipating eligibility increased current spending. For the pre-reform treatment group (i.e. those with a 17 year old member), there is no problem of anticipation. Given that the roll-out of the policy applied only to new entrants in post 16 education, these individuals were ineligible.[5] A final important issue that could affect the causal interpretation of the dif-in-dif estimates is the presence of general equilibrium effects that influenced the spending behaviour of the control group. EMA is a large programme and any increase in post 16 participation rates implies increased competition for post 16 schooling places. This may further affect the spending behaviour of the control households in the event that it caused a change in their expected post 16 schooling plans or future expected wage rates, which in turn lead them to adjust their current spending behaviour.[6]

A note on rescaling

As discussed in the methodology section, the initial piloting of the scheme means that in 41 of the 150 English LEAs, EMA was in operation before the start of the sample period. The above estimates therefore reflect a lower bound for the effect of EMA on the treated. Whilst the EFS data does not record information on LEA status, Government Office Region (GOR) information is available with each GOR being made up of multiple LEAs. Given information on EMA receipt by LEA, one may wonder whether it is possible to rescale the estimated EMA effects to reflect the

[5] However, if there are younger siblings in the household then anticipation effects are theoretically possible.

[6] The models estimated in this section do not include household level covariates. In a model with covariates, consistent estimation of the treatment effect further requires that the covariates are exogenous to the reform.

fact that treatment on the treated group is less than 100 percent, but by a known number. Here, we point out that rescaling a multiplicative treatment effect may not always make sense.[7]

We know that when effects are heterogeneous, PPML returns the multiplicative effect on the average. For the case of EMA, we know that the effect of the treatment is zero for a known share p of recipients. Following the main text, $exp(\delta)-1$ identifies

$$\frac{E\left[y_{1igt}|g = treated\right] - E\left[y_{0igt}|g = treated\right]}{E\left[y_{0igt}|g = treated\right]}. \tag{1}$$

However, both the numerator and denominator are going to be a weighted average of the two groups counterfactuals, so that $exp(\delta) - 1$ is equal to

$$\frac{\left(E\left[y_{1it}|g = treated, LEA = notpilot\right] - E\left[y_{0igt}|g = treated, LEA = notpilot\right]\right) \times (1-p)}{E\left[y_{0igt}|g = treated, LEA = pilot\right] \times p + E\left[y_{0igt}|g = treated, LEA = notpilot\right] \times (1-p)}. \tag{2}$$

Under the assumption that

$$E\left[y_{0igt}|g = treated, LEA = pilot\right] = E\left[y_{0igt}|g = treated, LEA = notpilot\right] \tag{3}$$

it is fairly trivial to rescale the estimated multiplicative effects of the results section. The scale factor $(1-p)$ could be calculated from publicly available data on EMA receipt by GOR and by then appropriately weighting for regional shares from the main estimation sample. However, eq. 3 is unlikely to hold in this example, where the pilot regions were on average much poorer than the national roll out areas. For this reason, we argue that rescaling may not make sense and caution against making such adjustments to the intent to treat estimates.

[7] On the one hand, one might argue that as a policy maker it is enough to know the intent to treat.

Appendix C: Simulation of a reform with only mean effects, in the presence of a different variance across groups

We consider a separate scenario corresponding to the case where $\alpha = 0$, so that the multiplicative error term is statistically independent of the treatment, but where its properties depend on the group. Specifically, y_{it} is heteroskedastic with respect to the group status but not to the treatment itself. To illustrate this scenario, consider the simulation procedure above but now σ_{it} is generated according to:

$$\sigma_{it}^2 = exp(\gamma \times \mathbf{1}(treated_{it} = 1))$$

where γ is now the parameter determining the degree of heteroskedasticity with respect to group status.

Simulations were made for five values of $\gamma = \{0, 0.1, 0.2, 0.3, 0.4\}$. Estimates of the multiplicative treatment effects along with the multiplicative treated group effects from OLS Log and PPML estimators are presented in table C1, along with the OLS level models.

The table confirms that, under the particular form of heteroskedasticity considered, log linerization works well when estimating the multiplicative treatment effect, but it performs poorly in terms of estimating the treated group effect. In contrast, PPML performs well for both the treated group and $treated \times post$ coefficients and for all values of γ. This illustrates the point made in section 2 that, in the dif-in-dif setting, heteroskedasticity is only a problem for consistently estimating the treatment effect when the treatment itself (and not the group) has an independent distributional effect; although estimates of the treated group effect can be misleading.

Finally, for the wrongly specified OLS level model in column 3, estimates of the multiplicative treatment effect again confound treatment and trend effects. Additionally, here we see that the form of heteroskedasticity means that the estimates of the treated group effect become more dispersed for higher values of gamma.

Table C1: Simulation

	(1) OLS Log	(2) PPML	(3) OLS Level
$\gamma = 0$			
Treated x Post	.1976176	.19624	4.618121
	(0.0698541)	(0.084088)	(2.510334)
Treated	-.3995088	-.3996061	-10.8962
	(0.0404981)	(0.047648)	(1.302451)
Park Test	0.053	0.008	-
Breusch-Pagan	0.061	-	-
$\gamma = 0.1$			
Treated x Post	.1976049	.1960462	4.614949
	(0.0712667)	(0.0866465)	(2.570692)
Treated	-.4251419	-.3996628	-10.89587
	(0.0413711)	(0.0490935)	(1.328722)
Park Test	0.096	0.023	-
Breusch-Pagan	0.255	-	-
$\gamma = 0.2$			
Treated x Post	.197592	.1958383	4.61164
	(0.0727196)	(0.0893964)	(2.636176)
Treated	-.4520214	-.3997227	-10.89545
	(0.0422676)	(0.0506476)	(1.357337)
Park Test	0.242	0.063	-
Breusch-Pagan	0.705	-	-
$\gamma = 0.3$			
Treated x Post	.1975789	.1956156	4.608198
	(0.0742098)	(0.0923475)	(2.707136)
Treated	-.4801382	-.3997863	-10.89492
	(0.0431861)	(0.0523165)	(1.388475)
Park Test	0.488	0.157	-
Breusch-Pagan	0.955	-	-
$\gamma = 0.4$			
Treated x Post	.1975658	.1953774	4.604632
	(0.0757344)	(0.0955096)	(2.783938)
Treated	-.5094768	-.3998539	-10.89427
	(0.0441245)	(0.0541067)	(1.422332)
Park Test	0.715	0.304	-
Breusch-Pagan	0.996	-	-

Notes: Results from 10,000 replications of the simulation procedure described in section 4. Mean of the estimated coefficients reported with standard deviations in parentheses.

Additional references

Ashenfelter, O., 1978. Estimating the Effect of Training Programs on Earnings. The Review of Economics and Statistics 60, 47–57.

Autor, D.H., 2001. Why do temporary help firms provide free general skills training? The Quarterly Journal of Economics 116(4), 1409–1448.

Bailey, M.J., 2006. More power to the pill: The impact of contraceptive freedom on women's life cycle labor supply. The Quarterly Journal of Economics 121(1), 289–320.

Bandiera, O., Barankay, I., Rasul, I., 2007. Incentives for managers and inequality among workers: Evidence from a firm-level experiment. The Quarterly Journal of Economics 122(2), 729–773.

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? The Quarterly Journal of Economics 119(1), 249–275.

Björkman, M., Svensson, J., 2009. Power to the people: Evidence from a randomized field experiment on community-based monitoring in Uganda. The Quarterly Journal of Economics 124(2), 735–769.

Blanden, J., Gregg, P., Machin, S., 2005. What's the good of education? The Economics of education in the United Kingdom. in Machin S., Vignoles A., What's the good of education? The economics of education in the United Kingdom, Princeton University Press .

Bleakley, H., 2007. Disease and development: Evidence from hookworm eradication in the American South. The Quarterly Journal of Economics 122(1), 73–117.

Blundell, R., Dias, M.C., 2009. Alternative Approaches to Evaluation in Empirical Microeconomics. Journal of Human Resources 44, 565–640.

Dearden, L., Emmerson, C., Meghir, C., 2009. Conditional cash transfers and school dropout rates. Journal of Human Resources 44(4).

DellaVigna, S., Kaplan, E., 2007. The fox news effect: Media bias and voting. The Quarterly Journal of Economics 122(3), 1187–1234.

Dittmar, J.E., 2011. Information technology and economic change: The impact of the printing press. The Quarterly Journal of Economics 126(3), 1133–1172.

Donohue, J.J., Heckman, J.J., Todd, P.E., 2002. The schooling of southern blacks: The roles of legal activism and private philanthropy. The Quarterly Journal of Economics 117(1), 225–268.

Donohue, J.J., Levitt, S.D., 2001. The impact of legalised abortion on crime. The Quarterly Journal of Economics 116(2), 379–420.

Donohue, J.J., Levitt, S.D., 2008. Measurement error, legalized abortion, and the decline in crime: A response to Foote and Goetz. The Quarterly Journal of Economics 123(1), 425–440.

Field, E., 2007. Entitled to work: Urban property rights and labor supply in Peru. The Quarterly Journal of Economics 122(4), 1561–1602.

Finkelstein, A., 2004. Static and dynamic effects of health policy: Evidence from the vaccine industry. The Quarterly Journal of Economics 119(2), 527–564.

Finkelstein, A., 2007. The aggregate effects of health insurance: Evidence from the introduction of medicare. The Quarterly Journal of Economics 122(1), 1–37.

Foote, C.L., Goetz, C.F., 2008. The impact of legalized abortion on crime: Comment. The Quarterly Journal of Economics 123(1), 407–423.

Gruber, J., Hungerman, D.M., 2008. The church versus the mall: What happens when religion faces increased secular competition? The Quarterly Journal of Economics 123(2), 831–862.

Hong, H., Kacperczyk, M., 2010. Competition and bias. The Quarterly Journal of Economics 125(4), 1683–1725.

Hornbeck, R., 2010. Barbed wire: Property rights and agricultural development. The Quarterly Journal of Economics 125(2), 767–810.

Jayachandran, S., Lleras-Muney, A., 2009. Life expectancy and human capital investments: Evidence from maternal mortality declines. The Quarterly Journal of Economics 124(1), 349–397.

Jensen, R., Oster, E., 2009. The power of tv: Cable television and women's status in India. The Quarterly Journal of Economics 124(3), 1057–1094.

Khwaja, A.I., Mian, A., 2005. Do lenders favor politically connected firms? Rent provision in an emerging financial market. The Quarterly Journal of Economics 120(4), 1371–1411.

Morgan, D.P., Rime, B., Strahan, P.E., 2004. Bank integration and state business cycles. The Quarterly Journal of Economics 119(4), 1555–1584.

Nunn, N., Qian, N., 2011. The potato's contribution to population and urbanization: Evidence from a historical experiment. The Quarterly Journal of Economics 126(2), 593–650.

Stevenson, B., Wolfers, J., 2006. Bargaining in the shadow of the law: Divorce laws and family distress. The Quarterly Journal of Economics 121(1), 267–288.

Verhoogen, E.A., 2008. Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector. The Quarterly Journal of Economics 123(2), 489–530.