ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 725 March 2013

Dif-in-dif estimators of multiplicative treatment effects

Emanuele Ciani and Paul Fisher

Note : The Discussion Papers in this series are prepared by members of the Department of Economics, University of Essex, for private circulation to interested readers. They often represent preliminary reports on work in progress and should therefore be neither quoted nor referred to in published work without the written consent of the author.

Dif-in-dif estimators of multiplicative treatment effects*

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

[a]Department of Economics, University of Essex

[b]Centre for the Analysis of Public Policies, University of Modena and Reggio Emilia

March 19, 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-linearized 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.

JEL: C21, C51, I38

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

*We wish to thank João Santos Silva, Marco Francesconi, Jonathan James, Vincenzo Mariani, Juan Hernandez, Roberto Nisticò, Ben Etheridge and seminar participants at Essex for useful comments. Paul Fisher received financial support from the ESRC. Emanuele Ciani received financial support from the ESRC and from the Royal Economic Society Junior Fellowship. Data from the the Expenditure and Food Survey has been accessed through the UK Data Archive.

†Corresponding author: pfishe@essex.ac.uk

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, that are linearized 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. In total, 25 papers using a dif-in-dif estimator with continuous outcomes were found.[1]

Table 1 summarises the findings of the literaure review. In 11 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 monetarised quantities including expenditures, 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 control 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 generalized linear model (GLM) with log links.[2]

In this paper, we use a potential outcomes framework to discuss the connections between

[1]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 by the authors. 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).

[2]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 3, this is an empirical question and will depend on the particuar research question and data set under study.

log linearisation and causal inference in the dif-in-dif case. In section 2, we argue that the preference for a multiplicative or additive model is primarily related to the causal interpretation of the estimands. This choice should be taken before deciding whether or not to take logs, which should be rather understood as an estimation strategy. Indeed, a number of findings that are scattered in the literature (Mullahy, 1997; Gregg et al., 2006; Santos-Silva and Tenreyro, 2006; Blackburn, 2007) suggest that to estimate a multiplicative effect there is no need to log-linearize, because a simple and robust non-linear estimator (Poisson Pseudo-Maximum Likelihood) is available.[3] For the dif-in-dif case, we also point out that the popular log-linearized estimates of a multiplicative treatment effect might confound distributional changes with shifts in the mean.

Previous literature on non-linear dif-in-dif mostly focused on the interpretation of the interaction effect. Mullahy (1999) discussed the case of a log-linearized 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 different stream of research focused on estimation of exponential models (Manning, 1998; Manning and Mullahy, 2001; Ai and Norton, 2008). Santos-Silva and Tenreyro (2006) and Blackburn (2007) showed that the OLS estimator of the log-linearized model may not be consistent for the parameter of interest. Gregg et al. (2006) proposed a simple way to recover a percentage treatment effect from linear OLS.

Here we contribute to the literature reconciling the two streams of research for the dif-in-dif case. We point out 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 difference, at least from an $ex-post$ evaluation perspective. We then show that, as long as we are interested in the effect on the average, the assumptions needed for causal interpretation of the dif-in-dif exponential specification do not imply that the log-linearized model is equivalent.

[3] 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.

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. This was a conditional cash transfer for 16 to 18 year old students, meant to increase participation in post-16 full-time education. Section 4 concludes.

2 Model specification and inference

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 analyze the case of a continuous outcome y, such as earnings or consumption.

In order to identify a causal effect, we need to impose some structure, depending 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.

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 ($exp\left(\lambda_{post} - \lambda_{pre}\right) - 1$)

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

in both groups. Now we can assume a proportional treatment effect:

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

The quantity $exp\left(\delta\right)-1$ is not the average of a multiplicative effect, but rather a multiplicative effect on the average. We could also allow the proportional effect to differ across groups. Then the dif-in-dif strategy would identify the effect on the treated. Actually, we do not even need to assume a constant treatment effect across individuals in order to identify the quantity on the left hand side of (3).[5]

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 specified as in eq. 2, otherwise we would confound time and treatment effects. However, it does not matter whether we express the treatment effect as a percentage difference or as a level difference. Indeed, the former is the fraction on the left hand side of eq. (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 is more consistent to present it in percentage form. Otherwise, we should focus 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 a 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

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

[5]A similar discussion, related to IV estimation of an exponential model with treatment effects, can be found in Angrist (2001, pg. 9).

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 period. Given the particular data structure, we get an exponential model for observed outcomes

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

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

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

$$\frac{\Delta^2 E\left[y_{it}|treated_{it}, post_{it}\right]}{\Delta treated_{it} \Delta post_{it}} = \left[exp\left(\beta_0 + \beta_1 + \beta_2 + \delta\right) - exp\left(\beta_0 + \beta_1\right)\right] - \left[exp\left(\beta_0 + \beta_2\right) - \left(\beta_0\right)\right]. \tag{7}$$

This quantity is actually equal to the difference in difference estimand for the additive effects model, that is

$$\begin{aligned} \frac{\Delta^2 E\left[y_{it}|treated_{it}, post_{it}\right]}{\Delta treated_{it} \Delta post_{it}} &= \left\{E\left[y_{it}|treated_{it}=1, post_{it}=1\right] - E\left[y_{it}|treated_{it}=1, post_{it}=0\right]\right\} \\ &- \left\{E\left[y_{it}|treated_{it}=0, post_{it}=1\right] - E\left[y_{it}|treated_{it}=0, post_{it}=0\right]\right\}. \end{aligned} \tag{8}$$

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

[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.

the ROR. This point is related 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.

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:

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

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

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

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*$\times$ *%time*. If we subtract the percentage change in the control group, we are left with *%effect*$\times$*(1+%time)*. The difference is likely to be negligible if *%time* is small.

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.

The discussion about how the different specifications of time effects is 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-linearized 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 are willing to condition on other covariates, such as demographic controls, that enter as a component of the linear index in the exponential model. This is also true if we have more than two periods and we use them to include a time trend.

A popular alternative is to log-linearize the model. To understand the pros and cons of this strategy, we can follow the discussion in Santos-Silva and Tenreyro (2006).[8] 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 the specification of the dif-in-dif framework in terms of expected value, 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).

Alternatively, define $\epsilon_{it} = (\eta_{it} - 1)\, exp\,(\cdot)$:

$$y_{it} = exp\left(\beta_0 + \beta_1 treated_{it} + \beta_2 post_{it} + \delta treated_{it} \times post_{it}\right) + \epsilon_{it} \tag{13}$$

$$E\left[\epsilon_{it}|1, treated_{it}, post_{it}\right] = 0. \tag{14}$$

To estimate the model, we can log-linearize it

$$lny_{it} = \beta_0 + \beta_1 treated_{it} + \beta_2 post_{it} + \delta treated_{it} post_{it} + ln\eta_{it}. \tag{15}$$

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 $x_{it} \equiv (1, treated_{it}, post_{it})$, so that $\epsilon_{it} = exp\,(x_{it}\beta)\, v_{it}$, with $v_{it} \perp\!\!\!\perp x_{it}$.[9] In

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

[9] This would be the case if y_{it} is log-normally distributed. For instance, we replicated a study by Aguila

the dif-in-dif setting, statistical independence implies that:

$$Var\left[y_{it}|1, treated_{it}, post_{it}\right] = \sigma_v^2 exp\left(2\beta_0 + 2\beta_1 treated_{it} + 2\beta_2 post_{it} + 2\delta treated_{it} \times post_{it}\right) \tag{16}$$

where $\sigma_v^2 = Var\left(v_{it}\right)$. 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 $exp\left(\delta\right)$.[10] 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.[11]

For instance, suppose that the condition $\epsilon_{it} = exp\left(x_{it}\beta\right) v_{it}$, $v_{it} \perp\!\!\!\perp x_{it}$, 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

$$\frac{Var\left[y_{it}|treated_{it} = 1, post_{it} = 1\right]}{Var\left[y_{it}|treated_{it} = 1, post_{it} = 0\right]} \neq exp\left(2\beta_2 + 2\delta\right). \tag{17}$$

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:

$$E\left[lny_{it}|1, treated_{it}, post_{it}\right] = \beta_0^* + \beta_1 treated_{it} + \beta_2 post_{it} + \delta^* treated_{it} \times post_{it} \tag{18}$$

$$\beta_0^* = \beta_0 + E\left[ln\eta_{it}|treated_{it} \times post_{it} \neq 1\right] \tag{19}$$

$$\delta^* = \delta + E\left[ln\eta_{it}|treated_{it} \times post_{it} = 1\right] - E\left[ln\eta_{it}|treated_{it} \times post_{it} \neq 1\right]. \tag{20}$$

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 PPMLE rather than log-linearizing it. Indeed, Battistin et al. (2009) show that total expenditure tend to be log-normal. Full results are available on request. We thank Emma Aguila for providing us the data and the original do-files.

[10] 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 (16) can arise in other specific cases.

[11] Furthermore, if statistical independence holds, our model would be nested in the more general framework proposed by Athey and Imbens (2006).

OLS estimates for the log-linearized 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 statistical independence 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. (16), but the condition is respected over time within the same group. Suppose that $\epsilon_{it} = exp\left(x_{it}\beta\right) v_{igt}$ holds, where $E\left[v_{igt}|x_{it}\right] = 0$. However, assume that the variance and higher moments in the distribution of v_{igt} depend on the group, though neither on the time period, nor on the treatment. In general, we would have that

$$
\begin{aligned}
E\left[ln\eta_{it}|treated_{it}=0, post_{it}=0\right] &= E\left[ln\eta_{it}|treated_{it}=0, post_{it}=1\right] \\
\neq E\left[ln\eta_{it}|treated_{it}=1, post_{it}=0\right] &= E\left[ln\eta_{it}|treated_{it}=1, post_{it}=1\right]. \quad (21)
\end{aligned}
$$

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 would be consistent in both cases, because it does not require η_{it} to be statistically independent from x_{it}. 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

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

[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 PPMLE, instead of log-linearizing. The point estimates of the change in Kentucky is basically unchanged (although it looses in statistical significant). However, we observe a large difference in 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.

Squares (NLS) and Poisson Quasi Maximum Likelihood (PPML), which are both consistent as far as the conditional mean is correctly specified. Santos-Silva and Tenreyro (2006) argued in favor of the latter, because NLS is likely to be more inefficient. PPML can be implemented in the most popular statistical packages and results can be easily interpreted. For instance, Santos-Silva and Tenreyro (2010) used it in a dif-in-dif setting to estimate the effect of the introduction of the Euro on trade. Blackburn (2010) estimated an exponential model for the effect of migration on earnings growth using panel data, which is an application of the dif-in-dif setup with longitudinal data. 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. (16) for PPML to be consistent, different pattern of heteroskedasticity make standard inference not valid. 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 generalized dif-in-dif model that gives a structural interpretation 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.

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 substantial resources to low income families with a child staying on in post compulsory education. As such, the policy is a candidate for a scenario in which treatment increases not only the mean of the outcome but also its variance. This is the first study to explore quantitatively how households allocate the additional resources provided by the allowance. 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 spending categories. Firstly, OLS log-linear models are estimated; these are then compared to results from models which estimate the treatment effect directly using the Poisson-pseudo maximum likelihood estimator. For completeness, results are additionally presented from level models of expenditure.

Importantly, key differences are revealed between the OLS log and PPML estimates which are both economically and statistically meaningful. PPML estimates are in line with earlier findings from the piloting of the reform suggesting that households primarily allocated the extra income to transport spending. The results support the central tenet of this paper, that is when estimating multiplicative models in a dif-in-dif setting, as a robustness check, results from the usual log-linear model should be compared to estimates from the Poisson pseudo-maximum likelihood approach.

3.1 The Policy

EMA was introduced in an attempt to tackle low take-up rates of post compulsory education in the UK.[15] EMA was means-tested on parental income and worth up to £30 per week for 16-18 year olds staying on in post compulsory schooling (less than university). It was usually claimed for two years of study. Additionally, further bonus payments were available for meeting educational targets. These were paid twice annually and worth up to £100. The policy was initially piloted in 15 local education authorities (LEAs) followed by a further 41 in the year 2000. In September 2004, the policy was rolled out nationwide to 150 LEAs. Students in low income households aged 16 before September 2004 and entering non-advanced education would be eligible for the allowance. Dearden et al. (2009) provide an excellent description of the policy environment and piloting of the programme. This paper considers the national rollout of the policy and asks how families targeted spent the available resources.

[15] Blanden et al. (2005) document the low take up rates of post 16 education in the UK

3.2 Data and Identification Strategy

We take advantage of expenditure data from the first five years of the Expenditure and Food Survey (EFS).[16] 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. The survey thus provides household level expenditure information for broad expenditure categories and disaggregated expenditures on specific consumption items. EFS interviews took place across a year and all income and expenditure figures 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) and the reform coincides with the start of the school year in 2004 (September). The first three years plus the first five months of the fourth year therefore correspond to the pre-reform period whilst the remaining data forms the post-reform period. 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 reparametrization of the level one.

To avoid concerns about the potential endogeneity of education status, we avoid defining treatment/control groups by this variable. We define our treated group of households to be those where at least one 17 year old is residing.[17] The control group is formed of households where at least one 14 or 15 year old resides, excluding those in the treated group. Table 2 demonstrates that these households are similar in observable characteristics to the treated group of households.

[16] 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. All calculations are made using Stata 11, and do-files producing the final results starting from raw data are available from the authors.

[17] To focus more precisely on the treated group information on date of birth would be required to determine EMA eligibility status. The EFS only contains information on age at interview.

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, increases by a constant percentage in the absence of treatment.

All estimates are obtained using Stata. 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.

3.3 Results

Table 3 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-linearized 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 log-linearized 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 percentage point 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 (Ashworth et al., 2002, see). In comparison to the standard log expenditure estimates of column 1, the PPML coefficient implies an EMA effect of 10.7 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.[18]

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 x post interaction is a statistically significant £16.51. 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 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 4 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 (38 percent of the PPML estimate) and statistically insignificant. For the subsample of low income households, we therefore see that considering the OLS log results alone would again lead to misleading inference. It is only once the PPML results are examined that we can separate the mean effect of the reform from the distributional effect. Finally, if we alternatively prefer the assumption of common trends in expenditure levels, the estimate has increased in magnitude

[18]One criticism might be that the result for transport is accidental, because in a full rest of regressions it is not unlikely to find at least one statistically significant estimate. However, here we focus on the difference between OLS (logs) and PPML results. Moreover, the result is justified by the separate evidence from the pilot.

and implies a significant increase in expenditure of £18.21 per week.

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 2. 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, 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 on 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 Conclusion

We critically assessed the standard practice of log-linearizing 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-linearization. The reason is that the latter might confound changes in higher moments of the outcome distribution with the treatment effect on the mean. For instance, evidence from log-linearised estimates suggests that the Educational Maintenance Allowance had no impact on households' expenditure, while PPML results show an increase in transport spending.

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., 2002. Education Maintenance Allowance: The first year - a quantitative evaluation. Department for Education and Employment Research (DfEE) Report No. 257.

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

Table 1: 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 growth rate		
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 labor 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 enrollment 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	

Note: in cases where other binary 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 2: 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	354.34	8.56
HH Size	4.09	4.00	0.09
Household Owned	0.69	0.69	0.00
Social Housing	0.22	0.24	-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	66.36	2.94
Alcoholic Beverages and Tobacco	18.47	15.81	2.66*
Clothing and Footwear	50.55	38.59	11.96***
Furnishings, HH Equipment, Carpets	40.51	40.37	0.14
Transport	90.11	78.96	11.15*
Communication	19.88	15.80	4.08***
Recreation	86.97	88.05	-1.08
Observations	1796		

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)

Table 3: Estimates of the EMA Effect on 6 Major Expenditure Categories

	Full Sample			Low Income Sample		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS Log	PPML	OLS Level	OLS Log	PPML	OLS Level
Food and	0.069	0.026	1.741	0.039	0.007	0.441
Non-alcoholic drinks	(0.051)	(0.041)	(2.837)	(0.080)	(0.065)	(3.863)
Observations	2623	2628	2628	1312	1315	1315
Reset(p-value)	0.0860	0.7180	0.8631	0.3021	0.3417	0.3888
Alcoholic Beverages	0.024	0.102	1.250	-0.027	0.173	2.380
and Tobacco	(0.104)	(0.101)	(1.659)	(0.156)	(0.152)	(2.321)
Observations	1966	2628	2628	933	1315	1315
Reset(p-value)	0.8783	0.2859	0.1091	0.5257	0.0123	0.0470
Clothing and	-0.138	0.012	-0.457	-0.137	-0.025	-1.658
Footwear	(0.100)	(0.091)	(4.061)	(0.152)	(0.142)	(5.350)
Observations	2340	2628	2628	1127	1315	1315
Reset(p-value)	0.8230	0.9425	0.2930	0.0329	0.7993	0.3780
Furnishings, HH	-0.122	-0.090	-3.423	-0.144	-0.207	-5.896
Equipment, Carpets	(0.115)	(0.177)	(7.110)	(0.170)	(0.219)	(6.970)
Observations	2572	2628	2628	1280	1315	1315
Reset(p-value)	0.2778	0.5370	0.3391	0.7201	0.1901	0.1321
Transport	0.117	0.208**	16.504**	0.124	0.327**	18.209*
	(0.098)	(0.095)	(8.279)	(0.153)	(0.163)	(9.848)
Observations	2502	2628	2628	1202	1315	1315
Reset(p-value)	0.6391	0.9168	0.6689	0.0925	0.7055	0.3957
Communication	-0.083	-0.032	-0.551	-0.011	0.029	0.564
	(0.064)	(0.074)	(1.406)	(0.095)	(0.118)	(2.027)
Observations	2545	2628	2628	1246	1315	1315
Reset(p-value)	0.4623	0.3296	0.2243	0.5828	0.1914	0.1680
Recreation	0.015	-0.010	-0.621	0.089	0.210	13.067
	(0.081)	(0.092)	(7.748)	(0.118)	(0.150)	(9.540)
Observations	2626	2628	2628	1313	1315	1315
Reset(p-value)	0.9062	0.1368	0.0864	0.5044	0.0158	0.0112

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).