THIS WAS AN INTERESTING PAPER TO READ. I STRONGLY AGREE THAT THE “COMMON TRENDS” ASSUMPTION IS OFTEN NOT TENABLE; AND THAT IT CAN BE TESTED WHEN THERE IS SUFFICIENT PRETREATMENT DATA. I ALSO AGREE THAT WITH SUFFICIENT PRETREATMENT DATA ALTERNATIVE ASSUMPTIONS, SUCH AS “COMMON ACCELERATION” CAN IDENTIFY A TREATMENT EFFECT. ...[MORE]

...

THE PROBLEM IS THAT THIS IS ALL, I BELIEVE, PRETTY WELL KNOWN. SO I DO NOT SEE A METHODOLOGICAL CONTRIBUTION HERE (ALTHOUGH THE ISSUES ARE LAID OUT NICELY). THESE ISSUES ARE DISCUSSED IN STANDARD TEXTBOOKS. FOR EXAMPLE, ANGRIST AND PISCHKE’S WELL-KNOWN “MOSTLY HARMLESS ECONOMETRICS” FROM 2008 DISCUSSES THE COMMON TRENDS ASSUMPTION AND THE INCLUSION OF GROUP SPECIFIC TRENDS TO RELAX THAT ASSUMPTION (AND CITES EARLIER PAPERS THAT DO THIS). THE MORA & REGGIO PAPER THAT THE AUTHOR DOES CITE DOES NOT REALLY OFFER ANY METHODOLOGICAL INNOVATION, BUT DOES OFFER A NICE TAXONOMY AND NOMENCLATURE FOR ALTERNATIVES TO THE COMMON TRENDS ASSUMPTION IN STANDING DIFFERENCE-IN-DIFFERENCE DESIGNS (THE MORA & REGGIO PAPER HAS NEVER BEEN PUBLISHED ALTHOUGH THEY DID PUBLISH A STATA JOURNAL ARTICLE INTRODUCING A SOFTWARE PACKAGE THAT IMPLEMENTS THE VARIOUS ALTERNATIVES). SO THE CONTRIBUTION OF THIS PAPER WOULD NEED TO LIE IN THE EMPIRICAL APPLICATION, RATHER THAN THE METHODICAL DISCUSSION. I AM LESS QUALIFIED TO JUDGE THE CONTRIBUTION HERE.

Many thanks for drawing our attention to the fact that economists have for long tried to cope with the issue of non-parallel trends. In that sense, our WP does not innovate. Still, the way the WP copes with non-parallelism is original and deviates from standard practice. Conditional on appropriate redrafting, that suggests maintaining a methodological focus to the paper. When the parallel-trend assumption fails, most authors (e.g. Friedberg, 1999; Autor, 2003; Besley & Burgess, 2004) resort to a polynomial (linear,…) trend-augmented version of the canonical DD model (Angrist & Pischke, 2009).

$$Y_{it} = \alpha + \sum_{\tau=t2}^{T} \alpha^{\tau} I_{\tau,t} + \alpha^{D} D_i + \eta AFTER_t {*} D_i \; \left[+ \boldsymbol{\theta}\, \boldsymbol{t} {*} D_i\right] \qquad [1.]$$

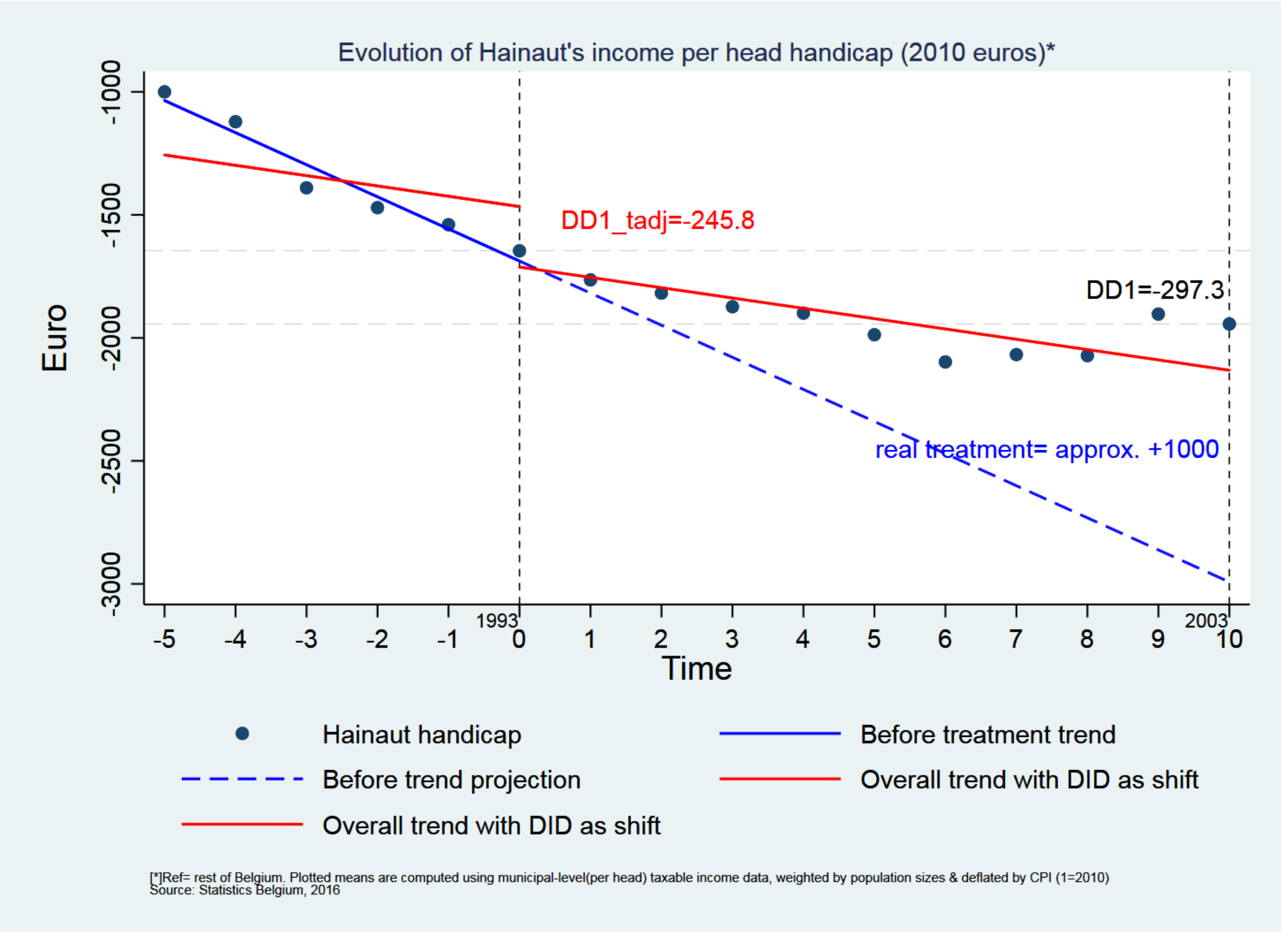
with $I_{\tau,t}$=1 if $t=\tau$ and 0 otherwise, and where Y_{it} is entity i's outcome in time t, D the treatment dummy, *AFTER* the after treatment dummy, and here t is a continuous variable.

Coefficient $\boldsymbol{\theta}$ captures the linear trend characterizing the treated entities. And η - a trend shift around $t=0$ – measures the treatment effect. As suggested by Wolfers (2006), the problem with this strategy is that it uses post-treatment observations, and that the treatment outcome takes the form of a once-in-a-time trend shift. A case in point is visible on Figure 1. The latter describes the evolution of income per head in the Belgian province of Hainaut (in deviation to the rest of Belgium), before and after it benefited from EU money.[1] That treatment began in 1994 and lasted until 2000. The trend is clearly negative prior to treatment, and still so after. The estimation of η, using the canonical DD model 10 years after treatment, delivers a negative value, in the range of -300€. A 'placebo' estimation of that model evidently reveals that there was no parallelism before the treatment started. So, the - 300€ figure is not trustworthy. This justifies estimating the trend-augmented eq.[1]. The red line on Figure 1 depicts the result. After treatment, the income handicap tends to stabilize, and this explains the moderately negative estimated trend ($\boldsymbol{\theta} < 0$). By construction, this trend applies to the pre-treatment period. Being negative, it delivers "corrected" DD estimates that are less negative than the traditional ones (-245.8€>-297.3€). Also, η corresponds to the trend shift just after $t=0$.[2] And as income handicap after treatment is

[1] See Vandenberghe (2016) for more details about EU-Objective 1-Hainaut.
[2] Defining $t=year-1993$

larger, that shift is still negative; suggesting that the EU policy failed (it "caused" approx: - 245€ of additional income handicap). Yet, $\boldsymbol{\theta}$ underestimates the actual pre-treatment trend (in blue on Figure 1). Before treatment, the handicap was growing faster than after. Prolonging the initial trend up to t=10 suggests that, *ceteris paribus*, the income handicap might have reached -3,000€, while it ended being less than -2,000€. The tentative conclusion is that the real treatment outcome was positive (in the range of +1,000€). What we propose hereafter is an alternative way of correcting DD estimation, that solely uses pre-treatment observations.

Figure 1 – The limitations of trend-augmented DD[$]



[$] Plotted values are (municipal)- population-weighted mean differences between Hainaut and rest-of-Belgium. These are used to estimate a linear trend-adjusted DD model.

Mora & Reggio (2012) suggest that DD analysis can be done by estimating a generalize fully-flexible equation, where the right-hand part only consists of time, treatment and time*X*treatment dummies:

$$Y_{it} = \gamma + \sum_{\tau=t_2}^{T} \gamma_\tau I_{\tau,t} + \gamma^D D_i + \sum_{\tau=t_2}^{T} \gamma_\tau^D I_{\tau,t} D_i \qquad [2.]$$

with $t=t_1, \ldots .\ T$ and $I_{\tau,t}$=1 if $t=\tau$ and 0 otherwise, covering before and after treatment periods.

The advantages to this equation are manyfold. First, conditional on the availability of many pre-treatement periods in the data, the OLS-estimated coefficients can be used to compute a whole family of difference-in-difference estimators $DD_{[p]}$, where $p=1, 2...q$ is the degree of parallelism underpinning identification. The canonical DD model is noted $DD_{[1]}$, and rests on parallelism of degree 1 ($Parallel_{[1]}$ hereafter).[3] Without $Parallel_{[1]}$, one should estimate $DD_{[2]}$ that rests on $Parallel_{[2]}$, i.e. outcome growth rate parallelism.[4] If $Parallel_{[2]}$ fails, one should turn to $DD_{[3]}$ with requires $Parallel_{[3]}$ or outcome acceleration[5] parallelism…and so on up to degree $p=q$, if data permit. Second, eq. [2], unlike eq.[1] can capture dynamic (ie. lagged) responses to treatment.[6] Third, – and this is something we particularly stress in the contex to this paper as it brings a solution to Wolfer's trend & shift problem – corrections for the violation of $Parallel_{[p]}$ rests solely on pre-treatment observations.

Consider the canonical $DD_{[1]}/Parallel_{[1]}$ estimator, with just before-and-after observations t^* and t^*+1. Treatment effect writes[7],[8]

$$DD_{[p=1]}^{t^*+1;t^*}=(\gamma^D_{t^*+1}+\gamma^D)-(\gamma^D_{t^*}+\gamma^D)=\gamma^D_{t^*+1}-\gamma^D_{t^*}. \qquad [3.]$$

Also, Eq. [2] can be used to assess $Parallel_{[1]}$ prior to treatment. Using pre-treatment periods t^*-2, t^*-1, one can compute 'placebo' $DD_{[1]}$ capturing the deviation from $Parallel_{[1]}$ prior to treatment. For instance, $DD_{[1]}^{t^*;t^*-1}=\gamma^D_{t^*}-\gamma^D_{t^*-1}$. should not be statistically different from zero. It not, then treated and control trends diverge before treatment (as illustrated on Figure 1 or its stylised equivalent Figure 2). And identification should rests on $Parallel_{[2]}$. The point is this can be easly achieved by computing

$$DD_{[p=2]}^{t^*+1;\,t^*-1}=DD_{[1]}^{t^*+1;\,t^*}-DD_{[1]}^{t^*;t^*-1}=(\gamma^D_{t^*+1}-\gamma^D_{t^*})-(\gamma^D_{t^*}-\gamma^D_{t^*-1}) = \gamma^D_{t^*+1}-2\gamma^D_{t^*}+\gamma^D_{t^*-1} \qquad [4.]$$

which is the difference between the observed t^*+1 outcome level handicap[9] $\gamma^D_{t^*+1}$ and its prediction $\gamma^D_{t^*}+DD_{[1]}^{t^*;t^*-1}$ given the handicap in t^* and its expected rise due to growth-rate difference between t^* and t^*-1. This prediction uses only regression coefficients driven by pre-treatment observations; a major difference with the trend-augmented method of eq.[1]. Note finally that the above logic can be generalized in many ways: to the case of lagged/dynamic treatment effects, or to $DD_{[p=q]}/Parallel_{[p=q]}$.where $q>2$ (Vandenberghe, 2016).

---

3 If outcome level change by unit of time (i.e 1st derivate) is "speed", then $Parallel_{[1]}$ means stable level differences due to identifical speeds.

4 If outcome growth rate change by unit of time (2nd derivative) is "acceleration" , then $Parallel_{[2]}$ means stable growth rate differences due to same accelerations.

5 If outcome acceleration change by unit of time (3rd derivative) is "surge", then $Parallel_{[3]}$ corresponds to a situation where acceleration differences remain stable due to identical surges.

6 The pattern of lagged effects is usually of substantive interest, e.g. if treatment effect should grow or fade as time passes.

7 When estimating eq. [2] with only 2 periods, $\gamma^D_{t^*}$ is subsumed into the constant $\gamma^D$and $DD_{[1]}$ is directly captured by the timeXtreatment coefficient.

8 Treatment effect' standard error must account for the fact that it consists of a linear combination of estimated coefficients, and thus of the covariance between variables. That is automatically done by STATA test or lincom commands used hereafter, that exploit the variance-covariance matrix of the estimated coefficients.

9 Net of the initial handicap in t^*-1 : γ^D

HOWEVER, I DO HAVE A FURTHER CONCERN. IT IS NOW WIDELY APPRECIATED THAT STANDARD INFERENCE METHODS ARE NOT APPROPRIATE IN DIFFERENCE-IN-DIFFERENCE (AND RELATED) SETTINGS WHERE THE NUMBER OF GROUPS ARE SMALL. THERE ARE PARTICULAR CHALLENGES WHEN, AS IN THIS CASE, THERE IS A SINGLE TREATED UNIT. METHODS TO DEAL WITH THESE SITUATIONS HAVE BEEN DEVELOPED BY, EG., CONLEY AND TABER (RESTAT,2011) AND BY ABADIE ET AL. (JASA, 2010). APPROPRIATE METHODS DO NOT SEEM TO HAVE BEEN APPLIED HERE.

Thanks again for telling us about this important aspect of DID estimation. But contrary to what you suggest, we are not using a single entity to identify the treatment effect. The province of Hainaut corresponds to 69 municipalities. As explained in the data section, we analyse the evolution of (population weighted) average taxable income at municipal level. Each of the 69 municipalities in Hainaut are considered as "treated" and are compared to either 84 municipalies forming the Liège province, or 193 forming the rest of Wallonie or even 520 for the rest of Belgium.

Table 1- Municipality count. Hainaut, Liège, rest of Belgium or rest of Wallonia

Rest of Belgium	520
Rest of Wallonia	193
Liège	84
Hainaut	**69**
Total	589

References

Angrist, J. D. and S. Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Autor, D. (2003), Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing, *Journal of Labor Economics*, 21(1), pp. 1-42.

Besley, T. & R. Burgess (2004), Can Labor Regulation Hinder Economic Performance? Evidence from India, *The Quarterly Journal of Economics*, Oxford University Press, 119(1), pp. 91-134.

Friedberg, L. (1998), Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data, *American Economic Review*, 88(3), pp. 608-627.

Mora, R. & I. Reggio (2012), Treatment effect identification using alternative parallel assumptions, *UC3M Working papers- Economics*, Universidad Carlos III de Madrid. Departamento de Economía.

Vandenberghe, V. (2016), Treatment-Effect Identification Without Parallel paths An illustration in the case of Objective 1-Hainaut/Belgium, 1994-2006, *IRES-WP*- 2016-031, IRES-UCL.

Wolfers, J. (2006), Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results, *American Economic Review*, 96(5), pp. 1802-1820.