SA-TIED
Southern Africa – Towards Inclusive Economic Development

Estimating employment responses to South Africa's Employment Tax Incentive

Joshua Budlender and Amina Ebrahim

SA-TIED Working Paper #187 | July 2021

UNITED NATIONS UNIVERSITY
UNU-WIDER
national treasury
Department: National Treasury
REPUBLIC OF SOUTH AFRICA
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI
planning, monitoring & evaluation
Department: Planning, Monitoring and Evaluation
REPUBLIC OF SOUTH AFRICA
the dtic
Department: Trade, Industry and Competition
REPUBLIC OF SOUTH AFRICA
SARS
TIPS

About the project

Southern Africa –Towards Inclusive Economic Development (SA-TIED)

SA-TIED is a unique collaboration between local and international research institutes and the government of South Africa. Its primary goal is to improve the interface between research and policy by producing cutting-edge research for inclusive growth and economic transformation in the southern African region. It is hoped that the SA-TIED programme will lead to greater institutional and individual capacities, improve database management and data analysis, and provide research outputs that assist in the formulation of evidence-based economic policy.

The collaboration is between the United Nations University World Institute for Development Economics Research (UNU-WIDER), the National Treasury of South Africa, the International Food Policy Research Institute (IFPRI), the Department of Monitoring, Planning, and Evaluation, the Department of Trade and Industry, South African Revenue Services, Trade and Industrial Policy Strategies, and other universities and institutes. It is funded by the National Treasury of South Africa, the Department of Trade and Industry of South Africa, the Delegation of the European Union to South Africa, IFPRI, and UNU-WIDER through the Institute's contributions from Finland, Sweden, and the United Kingdom to its research programme.

Corresponding author: jbudlender@umass.edu

WIDER Working Paper 2021/118

Estimating employment responses to South Africa's Employment Tax Incentive

Joshua Budlender[1] and Amina Ebrahim[2]

July 2021

Abstract: We present new evidence on the effects of South Africa's Employment Tax Incentive (ETI), a hiring and employment wage subsidy aimed at reducing youth unemployment. We show that attempts to estimate firm-level treatment effects via conditional difference-in-differences are likely to fail when comparing ETI to matched non-ETI firms. We show that even when event-study conditional pre-trends appear flat, the sensitivity of these estimates to the matching period means that pre-trends are not informative about counterfactual post-treatment parallel trends, and a broad array of matching approaches do not create credible post-treatment counterfactuals. We argue that this is likely due to mean reversion among matched non-ETI firms. A partial identification approach based on difference-in-differences with parametric time trends suggests that the ETI has increased firm-level youth employment, though some important caveats apply. Our results prompt a re-evaluation of the (sometimes contradictory) existing literature on the employment effects of the ETI: we judge that, in light of our findings, there is insufficient evidence to conclude on its impact either way.

Key words: difference-in-differences, employment, event-study, parallel trends, South Africa, wage subsidies

JEL classification: C23, J38, H25, H32

Acknowledgements: We would like to thank Murray Leibbrandt, Ihsaan Bassier, Arindrajit Dube, Jukka Pirttilä, and participants of the UMass Labor Economics Working Group for comments and suggestions at various stages of this project. We thank the SA-TIED team at the National Treasury and SARS, whose work producing the data made this research possible, and Marle Van Niekerk, who patiently explained various ETI institutional details to us. Joshua Budlender gratefully acknowledges support for this research from UNU-WIDER. Any errors remain our own.

[1] University of Massachusetts, Amherst, MA, USA; SALDRU, University of Cape Town, Cape Town, South Africa, corresponding author: jbudlender@umass.edu; [2] UNU-WIDER

This study has been prepared within the UNU-WIDER project Southern Africa—Towards Inclusive Economic Development (SA-TIED).

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-058-0

https://doi.org/10.35188/UNU-WIDER/2021/058-0

Typescript prepared by Gary Smith.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

In 2014 the South African government instituted a large, employer-borne payroll tax credit for newly hired low-wage young workers—the Employment Tax Incentive (ETI). The ETI is one of the state's most high-profile labour market interventions of the last decade, and at ZAR4.7 billion (US$312 million) claimed in tax credits in 2016/17, it cost 0.4 per cent of total tax revenue (National Treasury of South Africa 2019). The employment effects of the ETI are a matter of considerable public interest in South Africa. A small existing literature, discussed below, finds varying and sometimes contradictory results.

This paper presents new evidence on the employment effects of the ETI. In particular, it focuses on attempts to estimate firm-level impacts of the policy, using administrative tax data. We show that existing approaches to estimating firm-level impacts, which use conditional (matched) difference-in-differences (DiD) approaches to compare ETI to matched non-ETI firms, are unlikely to credibly identify ETI treatment effects, and indeed are likely to result in over-estimates of these effects. In documenting how these matching approaches fail, we show that even when conditional pre-trends appear flat, the sensitivity of these estimates to the matching period implies that pre-trends are not informative about post-treatment counterfactuals, and a broad array of matching approaches do not create credible post-treatment counterfactuals. We argue that this is likely due to mean reversion among matched non-ETI firms, an issue that is driven by, among other factors, substantial differences between ETI and non-ETI firms. ETI firms are much larger, have much faster employment growth, higher turnover, and younger workers, and pay lower wages than non-ETI firms. These differences make it very difficult to find non-ETI firms that can act as credible counterfactuals for ETI firms.

Due to the difficulties of finding credible controls, we attempt to parametrically control for differential time trends between ETI and non-ETI firms. We find that our point estimates are highly sensitive to the parametric trend used, and turn to the partial identification approach of Rambachan and Roth (2020a). This approach suggests that the ETI did increase firm-level youth employment, but is inconclusive when it comes to total and non-youth employment, and concerns about endogenous treatment adoption remain.

Methodologically, this paper contributes to a new literature that shows that 'pre-trend tests' can fail to reject invalid event-study designs, and that conditioning on flat pre-trends can exacerbate bias, as in Roth (2021). We also go into some detail when describing the assumptions needed for credible identification from unconditional and conditional DiD approaches, and we explain the advantages of event-study DiD designs over two-period DiD methods that are common in the South African literature. As part of this review, we also briefly discuss the new methodological literature that has identified problems with estimating staggered-adoption event studies by two-way fixed effects models, and carefully explain how to implement a 'stacked' regression to avoid these problems. It is hoped that this attention to identifying assumptions and implementation issues will be helpful both to other academics and to empirical research consumers such as policy-makers.

Substantively, our results prompt us to re-evaluate the South African literature on ETI treatment effects. With some new doubts about the credibility of the existing firm-level analyses that use conditional DiD approaches similar to those we discuss, we are left with worker-level analyses that do not find employment effects, and our partial identification result that does suggest a positive youth employment effect. There are reasons to be cautious about both sets of estimates, and we ultimately are agnostic about the likelihood of ETI employment effects.

The paper begins with an outline of the ETI policy details, and then moves to a literature review in Section 3, where we attempt to summarize the state of the existing ETI literature and make sense of apparently contradictory results. In Section 4 we discuss the administrative data we use, and in Section 5 present descriptive statistics. Section 6 carefully outlines the assumptions needed for credible identification of treatment effects when using unconditional and conditional DiD approaches, and then discusses estimation issues in a staggered-adoption event-study framework. Results are presented in Section 7, which starts with results from the unconditional DiD approach before moving to results from the conditional DiD. The discussion of the conditional DiD results in Section 7.2 largely focuses on why these results are not credible, and Section 7.3 suggests that mean reversion drives this issue. Section 7.4 tries to get around the problems with the conditional DiD methods by directly controlling for parametric differential trends, first to identify point estimates and then using a partial identification approach. Section 8 concludes by reiterating the main results, discusses how we evaluate the ETI literature in light of our findings, and briefly discusses avenues for our further research. This paper presents preliminary results and is a work-in-progress.

2 ETI policy details

The ETI is an optional employer-borne payroll tax credit, introduced on 1 January 2014, with the intention of mitigating extremely high youth unemployment in South Africa.[1] ETI-claiming firms receive refundable tax credits for *new hires* of South African workers between the ages of 18 and 29 who are paid at most ZAR6,000 per month and at least the prescribed sectoral minimum wage applicable to the firm. In cases where there is no sectoral minimum wage applicable to a given firm, an eligible worker must be paid at least ZAR2,000 per month in order for the subsidy to be claimed on their employment. (SARS nd, 2019).[2] The subsidy is only applicable to new hires, but employers can claim monthly tax credits for these workers for up to two years' employment—the policy is thus a hybrid of a hiring credit and general payroll tax credit. The value of the monthly credit is halved in the second year of employment.

The policy follows an unusual design in which the value of the credit relative to the wage is weakly decreasing in the wage, with two kinks, and a notch at the ZAR2,000 per month minimum for firms not bound by a sectoral minimum wage. The ETI subsidy is not a reduction in the payroll tax *rate*, but is a *lump sum* refundable tax credit determined according to the schedule in Table 1. Figure 1 shows how the credit affects the employment costs of eligible workers, depending on their wage, and taking into account the various minimum payment thresholds. A firm that is not subject to a legislated minimum wage must pay at least ZAR2,000 per month to claim the subsidy, which creates a notch at this point—and so a dominated region where firms have lower labour costs by increasing the wage of an eligible worker (Figure 1(a)). Firms bound by a minimum wage have a similar dominated region, just below the value of the minimum wage, but this region would only be relevant for potential minimum wage

[1] Workers hired after September 2013 were also eligible for the credit. The policy is currently due to end on 28 February 2029. It was initially set to run for three years, but was extended by two years in 2016 and then recently by another ten years. Its design was significantly altered in 2020 in response to COVID-19.

[2] These thresholds correspond to US$360 and 1,080 in 2014 PPP (purchasing power parity). The 2014 median monthly wage in South Africa was approximately ZAR3,100 (US$560 PPP). Monthly pay thresholds are adjusted to equivalent hourly thresholds for those working less than 160 hours per month. Subsidy thresholds increased in 2019 for inflation, and the introduction of a national minimum wage at the same time would likely have altered the effects of the policy, but this is outside the period of our data. Public entities may not claim the credits.

non-compliers.[3] Figure 1(b) shows how the value of the credit relative to the wage is weakly decreasing in the wage, with two kink points.[4]

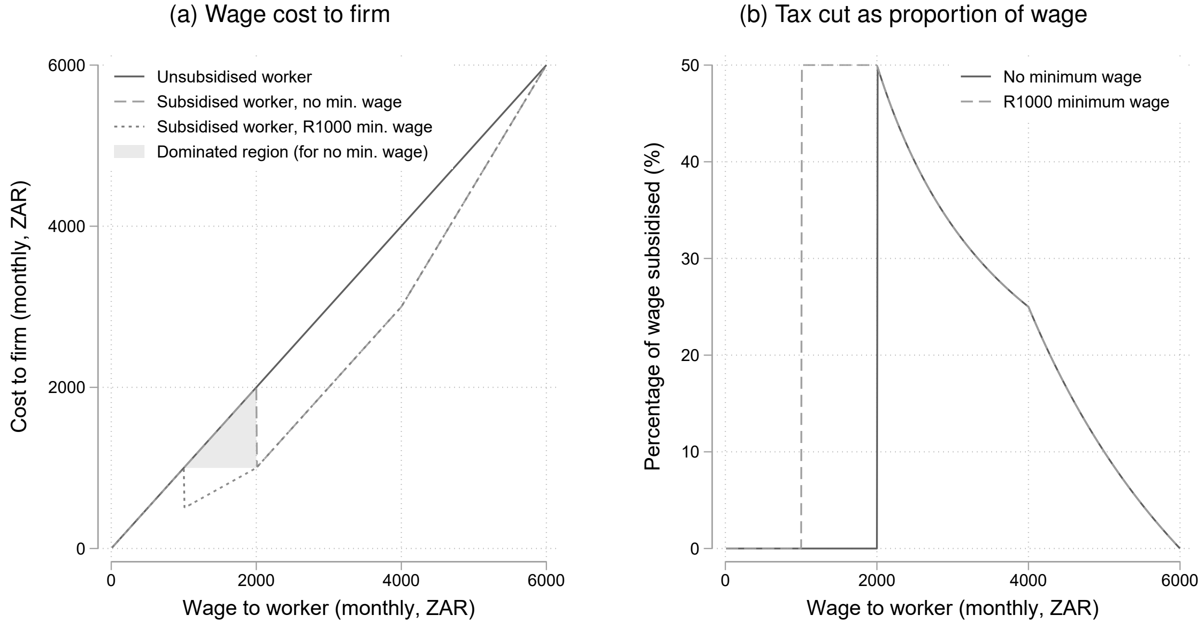
Table 1: Payroll tax credit determination, first 12 months of employment

Monthly wage (ZAR)	Tax credit determination (ZAR)	Monthly credit (ZAR)	Credit as percentage of wage
0–2,000	50% of monthly wage	0–1,000	50%
2,001–4,000	Fixed at 1,000	1,000	50 to 25%
4,001–6,000	1,000 – 0.5 × (wage – 4,000)	999–0	25 to 0%

Note: the table shows tax credit determination by monthly wage, ignoring minimum wage thresholds below which the subsidy cannot be claimed. These are figures for the first 12 months of claiming for a particular worker; the monthly credit value is halved in the second year. Threshold and credit determination formulas were adjusted in 2019, but this is outside the period of our data.

Source: authors' compilation.

Figure 1: Cost to firm of subsidized workers, first 12 months of employment



Note: the figure shows the cost to firms of employing subsidized and unsubsidized workers as per the ETI, abstracting from minimum wage thresholds. The subsidy is only applicable to new hires of South African citizens between the ages of 18 and 29. The subsidy lasts up to two years, and is halved in the second year.

Source: authors' illustration.

3 Existing literature

There is a small existing literature evaluating the effects of the ETI, which generally follows one of two broad approaches.

[3] It is possible that the ETI could improve minimum wage compliance—as far as we know, this has not yet been explored in the literature.

[4] In Section 8 we very briefly note that these unusual features of the ETI policy may allow firm-level analyses of the ETI based on firm treatment-intensity or bunching approaches, but in this paper we abstract from these issues and simply compare ETI-claiming and non-claiming firms. Investigating these alternative approaches constitutes an ongoing project for us.

3.1 Worker-eligibility approaches

First, Ranchhod and Finn (2015, 2016) and Ebrahim (2020a) leverage the *worker-specific* eligibility criteria to compare age- (and wage-) eligible *workers* against their ineligible counterparts in DiD settings.[5]

Using household survey data, Ranchhod and Finn (2015, 2016) find no employment effects for young workers (compared to a control group of older workers) 6 and 12 months after the implementation of the policy. They also do not find effects when using a 'before–after' design to test for time series breaks in labour market outcomes for young workers around the time of policy implementation. Ranchhod and Finn (2015, 2016) note that their analysis tests only for very short-run effects. In response to possible concerns about statistical power, they show that their null results are quite precise: only extremely small positive employment effects cannot be rejected.

Ebrahim (2020a) uses both household survey data and the job-level data of the NT-SARS (National Treasury–South African Revenue Service) tax administrative data to perform a similar style of worker-level analysis. However, she implements a *triple* DiD analysis, comparing age- and wage-eligible workers against age- and wage-*ineligible* workers. Her data cover the period 2011–18. She finds no statistically significant aggregate employment effects for eligible workers, though she does detect a small decrease in the unemployment rate of eligible youth, as well as very small increases and decreases in youth hiring and separation rates, respectively. She does, however, find small but notable wage increases for ETI-eligible workers as a result of the policy, which seem to be driven primarily by wage increases for workers earning less than ZAR2,000 per month before the policy, and more new jobs in the wage range of ZAR2,000 and ZAR4,000 per month.[6]

3.2 Firm-level approaches

The second broad approach is characterized by Ebrahim (2020b) and Bhorat et al. (2020), who use the NT-SARS employer–employee data to compare ETI-claiming and -non-claiming *firms* using DiD methods.[7] Because claiming and non-claiming firms are observably quite different, they implement *conditional* DiD methods, where various propensity score matching methods are used to construct comparable samples of ETI-claiming and -non-claiming firms for the DiD analysis.[8] We go into more detail when describing these papers as this is the approach we investigate in this paper.

Ebrahim (2020b) studies the effects of the ETI in the first full year after the policy, comparing 2013 (pre) to 2015 (post) outcomes, with matching on 2013 variables.[9] Data from 2011 and 2012 are used for lagged variables and she therefore uses a 2011–15 balanced panel. Using nearest-neighbour and kernel-weighted propensity score matching, her firm-level conditional DiD finds that the ETI notably increased employment of youths *and* non-youths at ETI-claiming firms compared to observably similar non-claiming firms (the effect size is an increase of about three youths and slightly less than six non-youths at ETI firms in 2015). In order to allay concerns about endogenous treatment selection, Ebrahim

[5] The earliest version of Ebrahim (2020a) was first published as Ebrahim and Pirttilä (2019). Ebrahim (2020a) is a PhD chapter, which we cite as the latest version of this work.

[6] This is plausibly a result of the notch at ZAR2,000 that we identify and discuss in Section 2, and is a subject of our continuing research.

[7] The earliest version of Ebrahim (2020b) was first published as Ebrahim et al. (2017). Ebrahim (2020b) is a PhD chapter, which we cite as the latest version of this work.

[8] Marcelin et al. (2019) use a very similar approach to examine effects of the ETI on firm investment, profits, and leverage, but it is excluded from our literature review, which focuses on employment and other labour market effects.

[9] Recall from Section 2 that the policy is in effect for only two months of the 2014 tax year.

(2020b) shows that there are positive employment effects for both positive-growth and negative-growth firms. As part of her common support analysis, Ebrahim (2020b) notes that the ETI take-up rate is very high for 'very large' firms (>1,200 employees), making it difficult to find similar non-claiming firms for the matching process, and these very large firms are therefore dropped from the sample. Her analysis is therefore restricted to firms with fewer than 1,200 employees. Ebrahim (2020b) notes that her agnosticism on very large firms is important for the overall policy evaluation, as these very large firms collectively employ over 50 per cent of ETI-eligible workers and claim more than half of the total subsidy value, despite being only 2 per cent of all ETI-claiming firms.

Bhorat et al. (2020) perform a very similar analysis to Ebrahim (2020b), with probably the three most-salient differences being that they do not exclude very large firms from their analysis, they include a wage restriction to focus on effects on subsidy-eligible youth rather than all youth, and that with an additional year of data they analyse effects in 2016 as well as 2015. They find very large percentage increases in firm-level youth-eligible employment associated with the ETI treatment (200 per cent in 2015 and 100 per cent in 2016), but interpret these as small magnitudes of actual employment-level changes by applying the estimated percentage changes to the level of changes of non-claiming firms over the period.[10] They similarly find quite large treatment effects in percentage terms for employment of wage-eligible individuals between the ages of 30 and 35 (about 30 per cent and 25 per cent in 2015 and 2016 respectively), and overall employment effects of about 11 per cent and 7 per cent. Like Ebrahim (2020b), Bhorat et al. (2020) also find that positive employment effects remain when restricting the sample to negative-growth firms. They do not discuss how they get around the large-firm common support issue noted by Ebrahim (2020b). While their estimated employment effects remain significant at the 5 per cent and 10 per cent levels for firms with more than 500 employees, precision does decrease for these large firms, and they do not separately discuss precision or matching credibility for 'very large' firms with more than 1,000 or 1,200 employees as in Ebrahim (2020b). As discussed, this is important for the overall policy evaluation because these firms claim the majority of the subsidy.

3.3 Discussion

The existing ETI literature thus seems to present conflicting results: the methods based on worker eligibility generally do not find employment effects, while papers comparing ETI-claiming to matched -non-claiming firms *do* find employment increases, and in fact find ETI employment effects for both eligible youths and non-youths. These estimates are not necessarily mutually incompatible. Because the worker-eligibility approaches use ineligible workers as the control group, they effectively estimate the *relative* increases in youth employment compared to the employment growth of ineligible workers. In cases where ineligible workers are not affected by the policy (and are a good counterfactual for eligible workers), this identifies the treatment effect. However, the firm-level results suggest that ineligible workers (specifically older workers) *are* affected by the policy, and may even experience *larger* employment effects than eligible youth. In this case it would not be surprising that the worker-eligibility approaches do not identify positive employment effects: the ETI may very well increase their employment for eligible workers, but not more than it increases the employment of ineligible workers.

This is in fact directly analogous to the case of Saez et al. (2019), who study wage pass-through from a payroll tax cut for young workers in Sweden. While they find no effect on the wages of eligible young workers when compared to slightly older ineligible workers, a firm-level treatment approach shows that

[10] It is not clear to us that this is the correct approach for converting their percentage change effects (from a logged outcome variable) into levels: this would seem to be applicable for an average treatment effect on the *non-treated*, but DiD analysis is more commonly used to estimate an average treatment effect on the treated, or perhaps an aggregate average treatment effect. The specific DiD estimates in Bhorat et al. (2020) may need to be interpreted cautiously in any case: we understand that at the time of writing the authors were reviewing their code for this part of their paper.

in fact workers' wages *did* increase in response to the policy, but that this occurred *across* the age distribution, thus invalidating the older workers as controls in the worker-eligibility approach.

However, just because spillover effects on ineligible workers could reconcile the South African ETI results, that does not mean that this is in fact what has happened. An alternative, plausible explanation is simply that the firm-level approaches have produced spurious results. Recall that differences between ETI and non-ETI firms are not disputed: firm-level approaches require matching methods to construct a valid control group for treated firms. These methods rely on a selection-on-observables assumption: that after matching, differences between ETI and non-ETI firms are as good as random. If confounding unobserved differences remain or if the matching process otherwise fails such that the DiD 'parallel trends' assumption does not hold, then treatment effects from the firm-level DiDs are not valid. As we discuss in Section 7, this issue is quite likely to cause over-estimates of ETI treatments if it is not addressed, because ETI firm employment growth is always much higher than that of non-ETI firms.

Another potential issue with the firm-level approaches is endogenous treatment selection: the idea that firms claim the ETI when they know that they are going to be hiring many workers anyway. This issue is difficult to deal with and is only partially addressed by looking at results separately for growing and shrinking firms, as in Ebrahim (2020b) and Bhorat et al. (2020). We briefly re-visit this issue in Section 8.

4 Data

This analysis uses the NT-SARS matched employer–employee firm-level panel, and covers the 2010–18 tax years (National Treasury and UNU-WIDER 2021a). The data are all at an annual tax year frequency. With the policy being implemented in 2014, we have an adequate number of pre- and post-periods in which to analyse the reform. However, tax years run from March to February rather than following the calendar year, so the 2014 tax year, for example, runs from 1 March 2013 to 28 February 2014. This means that the ETI could only be claimed for two months of the 2014 tax year. However, even if the 2014 tax year is excluded, we still have up to four pre- and four post-treatment tax years. The data cover the universe of tax-registered firms and employees. Informal firms and informally employed workers are not included.

The worker-level data include an indicator of whether the ETI credit is claimed for that worker, as well as the amount claimed (National Treasury and UNU-WIDER 2021b). They additionally have worker age and annual remuneration, with 'periods worked' and 'date employed' variables that can be used to approximate monthly wages.[11]

In general we define firms as PAYE entities (using the PAYE reference number), the payroll entity which very roughly approximates establishments. Occasionally (when indicated) we present variables that are defined per CIT (company income tax) entity, where the entity is the company at large. For example, when presenting statistics on sales per worker, we must use the number of workers at the CIT entity as the denominator, because sales are reported at the CIT level only.

The total subsidy value of ZAR4.07 billion in the 2017 tax year (see Table 3) reasonably closely matches the ZAR4.7 billion National Treasury number for the same year mentioned in Section 2. Divergences may be due to our efforts to clean the ETI-claiming variable of apparent data error (e.g. claims on age-

[11] See Ebrahim et al. (2021), Pieterse et al. (2018) and Kerr (2020) for further discussion of the data and especially the earnings variables.

ineligible workers) and our efforts to remove duplicate and pre-revision IRP5 job-level records, among other factors.

5 Descriptive statistics

ETI-claiming firms are very different from non-claiming firms, even before the ETI comes into effect. Table 2 shows the characteristics of a balanced panel of ETI and non-ETI firms for the sample period of 2010–18.[12] ETI firms are defined as firms that claim the ETI at least once in any time period, while non-ETI firms never claim the subsidy. Three key differences are particularly striking and important:

1. ETI firms are much larger than non-ETI firms, with substantially more employees and sales.
2. ETI firms have much faster employment growth than non-ETI firms.
3. ETI firms have a younger age structure of employees, with lower median age and a higher share of young employees.

ETI firms also have higher separation and hiring rates, suggesting greater worker turnover, and pay lower median wages. Perhaps surprisingly, there are not big differences in the value of sales per worker between ETI and non-ETI firms, and if anything sales per worker are higher for non-ETI firms. However, in general, the scale of the employment level and growth differences makes it clear that firms that claim the ETI are very different from those that do not. Even ETI firms that are observably similar to non-ETI firms in these dimensions are likely to be more professionalized than non-ETI firms, given that they claim the subsidy. While it is possible that the administrative costs of claiming the ETI would outweigh its benefits for some non-claiming firms, it is likely that many are foregoing a subsidy which would increase their after-tax profits due to lack of administrative capacity.

[12] Characteristics for the full unbalanced panel are shown in Appendix Table A1.

Table 2: Characteristic of ETI and non-ETI firm balanced panel, by tax year (means)

Tax year	Firms	Employment		Emp. growth (%)		Sep. rate (%)	Hire rate (%)	Sales (R, m)	Sales/emp. (R, m)	Age	Wage (R)
		All	Young	All	Young						
						Non-claiming firms					
2010	78,718	24	7			19		44	1.45	39	6,766
2011	78,718	26	8	47	26	19	25	43	1.29	39	6,946
2012	78,718	27	8	19	9	20	22	53	1.33	39	7,476
2013	78,718	28	8	13	7	19	20	59	1.41	40	7,877
2014	78,718	28	8	12	2	18	19	64	1.60	40	8,465
2015	78,718	28	8	15	2	17	19	65	1.67	41	9,011
2016	78,718	28	8	8	1	18	17	70	1.69	41	9,647
2017	78,718	28	7	9	0	17	16	76	1.79	41	10,489
2018	78,718	27	7	10	–2		16	47	1.78	42	10,999
						ETI-claiming firms					
2010	27,794	118	49			24		162	1.30	35	5,544
2011	27,794	132	55	108	63	27	32	167	1.13	35	5,290
2012	27,794	141	58	44	37	26	30	206	1.13	35	5,600
2013	27,794	150	62	44	42	26	30	229	1.19	35	5,844
2014	27,794	156	64	34	28	25	29	253	1.28	35	6,140
2015	27,794	162	65	30	27	25	29	273	1.29	35	6,383
2016	27,794	166	66	23	25	26	26	289	1.37	35	6,755
2017	27,794	167	65	24	13	26	26	313	2.05	36	7,546
2018	27,794	168	64	55	32		25	274	1.72	36	7,900

Note: the table shows characteristics of the balanced panel of firms in the study period, by tax year. All statistics are firm averages, although the age and (monthly) wage statistics are averages of firm medians. 'Sep. rate' is the separation rate for that year and 'Sales/emp.' the value of sales per worker at the CIT level. Firms are defined at the PAYE level. Rands (R) are in current prices, and sales and sales/emp. are annual and in millions of rands.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

The ETI directly and indirectly affects a large proportion of South Africa's formal firms and formally employed workers. By 2018, 14 per cent of firms in the NT-SARS panel claimed the ETI, and 12 per cent of workers had their wage subsidized by the ETI (Table 3).[13] However, because the ETI is disproportionately claimed by large firms, an even larger proportion of workers may be *indirectly* affected by the policy if there are within-firm spillover effects: about 50 per cent of formal workers are in firms that claim the ETI.

[13] The count of 'Total firms' in Table 3 only includes firms that have employees (per linked IRP5 records). While this dramatically reduces the number of firms in the data per year (there are otherwise around 800,000 distinct CIT entities per year), it is likely a decent measure of 'active' firms. ETI claiming is only recorded in the IRP5 data in any case.

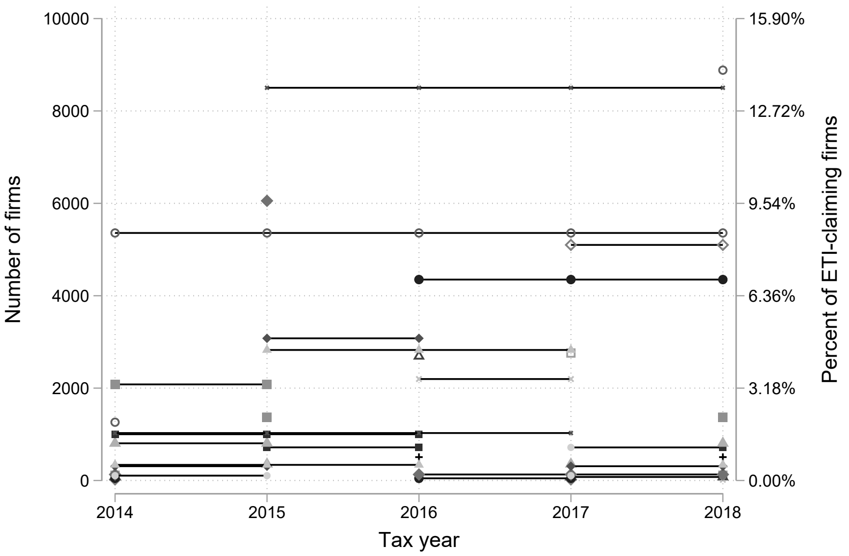
Table 3: ETI claiming in the NT-SARS data

Tax year	Firms			Workers (millions)			Subsidy
	All	ETI	ETI/all (%)	All	ETI	ETI/all (%)	(R, b)
2010	215,408	0	0	8.09	0.00	0	0.00
2011	226,529	0	0	8.98	0.00	0	0.00
2012	230,982	0	0	9.19	0.00	0	0.00
2013	232,832	0	0	9.41	0.00	0	0.00
2014	236,038	12,477	5	9.68	0.13	1	0.15
2015	243,225	33,866	14	10.04	0.80	8	2.94
2016	248,961	32,162	13	10.14	1.02	10	3.91
2017	253,428	33,233	13	10.24	1.14	11	4.07
2018	259,398	36,714	14	10.39	1.24	12	4.29

Note: the table shows the extent of ETI claiming by tax year. The ETI was operational for only a few months of the 2014 tax year. Firms are defined here at the CIT level. Rands are in current prices.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021b).

Figure 2: Prevalence of distinct firm ETI-claiming patterns



Note: the figure depicts the prevalence of particular firm-level ETI-claiming patterns, showing the staggered and sometimes intermittent nature of ETI adoption. Each horizontal combination of markers indicates a particular claiming pattern, with the markers indicating years when the ETI was claimed. When the markers are joined this indicates continuous ETI claiming for the joined period. The vertical height of the pattern shows how many firms fit the particular claiming pattern, with the absolute number of firms on the left axis and as a percentage of all ETI-claiming firms on the right axis. If a firm exits the panel it is counted as no longer claiming. Firms are defined at the PAYE level.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021b).

Not all ETI-claiming firms begin claiming the ETI in 2014 and continue claiming until the end of the data period. In fact, a key feature of the ETI policy is *staggered take-up*, meaning firms begin claiming at different times. Some firms also decide to *stop* claiming the ETI, though just over 50 per cent of firms continue claiming until the end of the panel even when counting panel-exit as non-claiming. Figure 2 shows this staggered and sometimes intermittent nature of ETI claiming. About 8,900 firms only claim the ETI in 2018 (about 14 per cent of all ETI-claiming firms), with similar but slightly fewer firms claiming continuously from 2015 through to 2018. About 5,400 firms claim the ETI throughout our panel period, constituting 8.5 per cent of ETI-claiming firms. As an example of intermittent claiming, the squares in the bottom fifth of the panel show that around 1,400 firms (2 per cent) claim the subsidy only in 2015 and 2018.

6 Empirical approach

6.1 Unconditional and conditional event-study DiD

The primary identification strategy we use to estimate ETI treatment effects is matched event-study DiD, based on firm-level ETI treatment. Our DiD approach compares ETI-claiming firms to -non-claiming firms, before and after the policy. The advantage of this approach over strategies that leverage worker-eligibility criteria (as in Ranchhod and Finn (2015, 2016) and Ebrahim (2020a)) is that it allows for spillover effects of the policy on ineligible workers—which are potentially relevant for credible identification but are also interesting to study in their own right. The primary cost of this approach is that we must impose a (conditional) parallel trends assumption for ETI and non-ETI firms, which will be reasonable only under some specific circumstances. Below we outline the technical requirements for unconditional and conditional DiD methods, which we refer back to throughout this paper.

Unconditional DiD

Consider an observed outcome Y_{it} for firm i in period t.[14] For the purposes of explanation we consider three time periods, $t = -2, -1, 0$, where treatment (claiming the ETI) begins between period -1 and 0. Period -1 will be our 'reference period', while period 0 is the first post-treatment period: we will quantify the magnitude of the ETI effect by looking at changes between periods -1 and 0. The set of firms that receive treatment have $D_i = 1$, while non-treated firms have $D_i = 0$. However, for every firm we consider its 'potential outcome' *if it had not been treated*, where $Y_{it}(1)$ and $Y_{it}(0)$ are the outcomes for firm i if it does and does not receive treatment in time t, respectively. We never observe both $Y_{it}(1)$ and $Y_{it}(0)$ for a given i and t; instead we observe $Y_{it} = D_i Y_{it}(1) + (1 - D_i)Y_{it}(0)$. That is, we observe a firm's treated or non-treated outcome depending on whether it is actually part of the treatment ($D_i = 1$) or control ($D_i = 0$) group. Further, we assume that the ETI treatment has no causal effect on the outcome prior to a firm actually receiving the treatment, so that $Y_{it}(1) = Y_{it}(0)$ when $t < 0$. The causal effect we are interested in is the *average treatment effect on the treated* (ATT), which is $\tau_{\text{ATT}} = \mathbb{E}[Y_{i0}(1) - Y_{i0}(0) | D_i = 1]$. In words, this is the differences in outcome due to treatment for the treated group. The difficulty is that we do not observe $Y_{i0}(0)$ for $D_i = 1$, the counterfactual outcome for a treated firm if it had not received treatment.

This ATT is typically estimated with a dynamic event-study regression with a form something like:[15]

$$Y_{it} = \lambda_i + \varphi_t + \sum_{s \neq -1} \beta_s \times \mathbb{1}[t = s] \times D_i + \varepsilon_{it} \tag{1}$$

where λ_i and φ_t are unit and time fixed effects, respectively. The estimated coefficient $\hat{\beta}_0$ is the difference-in-differences between the (sample means of) the outcomes between the treated and untreated groups, between periods -1 and 0: $\hat{\beta}_0 = (\bar{Y}_{1,0} - \bar{Y}_{1,-1}) - (\bar{Y}_{0,0} - \bar{Y}_{0,-1})$. Note that now $\bar{Y}_{d,t}$ denotes the sample mean $Y_{i,t}$ for firms in treatment group d in period t. No coefficient is estimated for $\hat{\beta}_{-1}$ because it must be omitted as the reference period against which treatment effects are defined. Pre-period coefficients are therefore similarly defined relative to this omitted category, so that $\hat{\beta}_{-2} = (\bar{Y}_{1,-2} - \bar{Y}_{1,-1}) - (\bar{Y}_{0,-2} - \bar{Y}_{0,-1})$.

[14] This exposition draws substantially from Rambachan and Roth (2020a).

[15] See Section 6.2 for discussion of how best to actually estimate this type of model in a 'staggered take-up' context like we have with the ETI.

Taking expectations, reorganizing terms, and recalling the definition of the ATT above, we can show that:

$$\mathbb{E}[\hat{\beta}_0] = \tau_{\text{ATT}} + \underbrace{\mathbb{E}[Y_{i,0}(0) - Y_{i,-1}(0)|D_i = 1] - \mathbb{E}[Y_{i,0}(0) - Y_{i,-1}(0)|D_i = 0]}_{\delta_0} \tag{2}$$

The two terms on the right of Equation (2) are what the growth would have been in the outcome, *in the absence of treatment*, for the treated (ETI) and non-treated (non-ETI) firms, respectively. Collectively, as δ_0, they represent the post-period differential growth trend between the treated and non-treated groups. If growth in the outcome had been the same for these two groups $\mathbb{E}[Y_{i,0}(0) - Y_{i,-1}(0)|D_i = 1] = \mathbb{E}[Y_{i,0}(0) - Y_{i,-1}(0)|D_i = 0]$, the two terms would cancel out, $\delta_0 = 0$, and the estimated $\hat{\beta}_0$ identifies the treatment effect τ_{ATT}. This is the 'parallel trends' assumption, the requirement that the treated and non-treated groups would have had parallel growth paths in the absence of treatment. Note that this assumption does *not* require that the treated and non-treated firms have similar *levels* of the outcome, but that they have the same growth. If the parallel trends assumption is violated, the bias in the estimated treatment effect is equal to the extent of the parallel trends violation δ_0, which can be positive or negative. In the case of the ETI, where we saw from Section 5 that treated (ETI) firms seem to secularly grow faster than non-treated (non-ETI) firms, even before the policy is implemented, this suggests that perhaps $\delta_0 > 0$, in which case $\hat{\beta}_0$ will be upwardly biased as an estimate of τ_{ATT}.

The parallel trends assumption cannot be tested, because we do not observe $Y_{i0}(0)$ for $D_i = 1$. However, *pre-period* trends are often used to judge how plausible the post-period parallel trend assumption is, following much the same logic as we used above with reference to the descriptive statistics. The implicit assumption here is that differential growth rates in the absence of treatment would be stable over the pre- and post-periods. Under our assumption that the treatment does not affect outcomes before it is implemented, we can directly identify the pre-period differential trend δ_{-2}:

$$\mathbb{E}[\hat{\beta}_{-2}] = \underbrace{\mathbb{E}[Y_{i,-2}(0) - Y_{i,-1}(0)|D_i = 1] - \mathbb{E}[Y_{i,-2}(0) - Y_{i,-1}(0)|D_i = 0]}_{\delta_{-2}} \tag{3}$$

so that $\hat{\beta}_{-2} = \delta_{-2}$. One of the primary advantages of the event-study DiD, compared to a two-period 'before–after' DiD, is that in estimating these pre-period coefficients we get some idea of the plausibility of the parallel trends assumption. Roth (2021) has demonstrated that pre-trend tests can have low power to reject differential trends, and that conditioning on non-rejection of parallel trends can have pathological effects (we discuss a related issue below), but there is little doubt that estimating and comparing pre-trends is superior to ignoring them when it comes to evaluating the credibility of a DiD analysis.

Conditional DiD

However, as discussed already, the above parallel trends assumption is unlikely to hold when comparing employment between ETI-claiming and -non-claiming firms. ETI firms generally have much faster employment growth than non-ETI firms. The approach used by Ebrahim (2020b) and Bhorat et al. (2020) to get around this issue is to use *conditional* DiD methods. The core idea of this approach is to use matching methods to construct a subset of non-ETI firms that are suitably similar to ETI firms, so that the parallel trends assumption holds between these two groups, and then to perform the DiD analysis comparing the ETI sample to this matched non-ETI sample. Canonical papers explicating versions of this approach are Heckman et al. (1997) and Abadie (2005), with Callaway and Sant'Anna (2020) presenting a modern and integrated approach. A set of covariates X_i that predict treatment group status D_i are used for the matching procedure, so that after controlling for characteristics X_i, the treatment status is as good as random between the groups. Concretely, this transforms the implicit *unconditional* parallel trends assumption implicit in Equation (2) into a *conditional* parallel trends assumption that uses some method to condition on X_i:

$$\mathbb{E}[Y_{i,0}(0) - Y_{i,-1}(0)|D_i = 1, X_i] = \mathbb{E}[Y_{i,0}(0) - Y_{i,-1}(0)|D_i = 0, X_i] \tag{4}$$

Analogously to the $\hat{\beta}_s$ coefficients in Equations (2) and (3), coefficient estimates $\hat{\gamma}_s$ from this conditional approach are then

$$\mathbb{E}[\hat{\gamma}_0] = \tau_{\text{ATT}} + \mathbb{E}[Y_{i,0}(0) - Y_{i,-1}(0)|D_i = 1, X_i] - \mathbb{E}[Y_{i,0}(0) - Y_{i,-1}(0)|D_i = 0, X_i] \tag{5}$$
$$\mathbb{E}[\hat{\gamma}_{-2}] = \mathbb{E}[Y_{i,-2}(0) - Y_{i,-1}(0)|D_i = 1, X_i] - \mathbb{E}[Y_{i,-2}(0) - Y_{i,-1}(0)|D_i = 0, X_i] \tag{6}$$

There are a variety of ways to adjust the estimating Equation (1) and estimation procedure so that it allows conditioning on characteristics X_i, and that the estimated coefficients reflect the forms in Equations (5) and (6). These include coarsened exact matching, regression adjustment, nearest-neighbour matching, inverse probability weighting, kernel weighting, and doubly robust methods. The latter four methods usually rely on estimating propensity scores. The mechanics of these different methods are not the focus of this paper, and the reader is referred to Heckman et al. (1997), Abadie (2005), Goldschmidt and Schmieder (2017), Callaway and Sant'Anna (2020), and Ebrahim (2020b) for discussion and implementation of these methods.[16]

One note about implementation is worth discussing, however. The reader will notice that the X_i terms above do not have a time period index t. This is because these matching characteristics should generally be time-invariant or set constant according to their level in some pre-treatment period. Time-varying and post-treatment covariates can be 'bad controls' when they are correlated with the outcome of interest, obscuring the true treatment effect. This means, for example, that if one wanted to match firms on firm median wages, one needs to decide which pre-period to use for these wages: firm wages fluctuate over time, so a particular period (or group of periods) must be chosen, and that level of wages used for the matching procedure in all of the other periods.

Once these issues are resolved, the conditional estimates are interpreted analogously to the unconditional case. Similarly to Equations (2) and (3), one can use the coefficient $\hat{\gamma}_{-2}$ from Equation (6) to see whether there are differential pre-trends; and if not, this is taken as evidence that there are unlikely to be differential post-trends, so that the post-treatment parallel trends assumption (Equation (4)) holds, and the $\hat{\gamma}_0$ of Equation (5) identifies the ATT. We emphasize again that the identifying assumption here requires that *trends* be similar between the treated and non-treated groups after matching, not levels. It is common practice to show that a matching procedure 'balances' the *levels* of covariates between treated and control groups in a conditional DiD, but this is only relevant in so far as this matched group also has balanced growth rates of the outcome variable. In fact, matching on levels that does not match trends can cause *additional* bias in a DiD analysis, compared to no matching at all (see Daw and Hatfield (2018), which we return to below).

A contribution of this paper (discussed below) is to show that conditional DiD estimates can be highly sensitive to the pre-treatment period chosen for matching characteristics. A consequence is that the pre-treatment coefficient $\hat{\gamma}_{-2}$ from Equation (6) likely will *not* be informative about differential trends in the post-period. In the context of the ETI, this is likely to lead to severely upward-biased estimates of the treatment effect.

[16] The matching issues we document in Section 7.2 are robust to all of the above methods, though coarsened exact matching (CEM) performs less poorly. The better performance of the CEM approach is interesting, especially in light of King and Nielsen (2019), and is a topic of our further investigations. We also abstract from common support issues in this paper, as specifications with common support restrictions do not materially affect the matching issues we document.

6.2 DiD event-study estimation

Until recently, unconditional DiD event-study models were frequently estimated according to a specification like Equation (1), using two-way fixed effects. However, in the last few years a significant literature has emerged regarding event-study estimation in contexts with staggered treatment adoption and more than two time periods—exactly our setting. Papers such as Goodman-Bacon (2018) and Sun and Abraham (2020) show that the common practice of estimating these models with a simple two-way fixed effects specification can result in significantly biased treatment effect estimates. In simulations, Baker et al. (2021) show that this bias can in some cases even lead to wrongly signed estimates.

The main cause of this bias is relatively straightforward. Goodman-Bacon (2018) and Sun and Abraham (2020) show that in an event-study two-way fixed effects specification with staggered treatment adoption, the resulting treatment effect estimate can be decomposed into a *weighted average of all the possible two-period DiD comparisons*. The first problem is that the two-way fixed effect mechanism is therefore not judicious about restricting itself to 'clean comparisons'. For example, as part of the set of two-period DiD comparisons, outcomes for *late* adopters (e.g. a firm that first claims the ETI in 2017) are compared to outcomes of *early* adopters (e.g. a firm that first claims the policy in 2015) in a two-period DiD framework. But early adopters are not valid 'non-treated' counterfactuals for late adopters when there are dynamic treatment effects. The second problem is that the weighting of the different two-period DiD estimates into an aggregate effect can be undesirable for policy purposes. It is in fact possible for some comparisons to be given *negative* weights.

Not all event studies with staggered treatment adoption will have biased estimates when treatment effects are estimated via a two-way fixed effects model, but the conditions are quite general: staggered-adoption and heterogeneous treatment effects over time or across units (Baker et al. 2021). The resolutions of these issues are not, however, very complicated: they involve being careful and explicit about control and treatment groups for each event, and deliberately choosing a method to aggregate the various simple DiD comparisons. Baker et al. (2021) perform simulations for three methods, which all work well: Sun and Abraham (2020), Callaway and Sant'Anna (2020), and the 'stacked' approach of Cengiz et al. (2019). The latter two methods allow conditional DiD estimates with matching, and were used at various points in the course of this analysis. The results reported in this paper all use the 'stacked' approach.[17]

Our 'stacked' specification

The stacked approach requires identifying 'events' in the data when groups of firms begin claiming the ETI. For example, firms that first claim the ETI in 2014 constitute the treated firms in one event, while those that first claim the ETI in 2015 constitute the treated firms in another event. These firms are then 'stacked' such that they are centred in 'event time'. For firms that first claim the ETI in 2014, their 2014 year is event-time 0, 2015 is event-time 1, and 2013 is event-time –1. For firms that first claim the ETI in 2015, their 2015 year is event-time 0, 2016 is event-time 1, and 2014 is event-time –1. DiD regressions are then run with respect to the event-time variable, so that –1 (for example) refers to the immediate pre-treatment period for all firms, regardless of which calendar year that is for them.

The researcher has great flexibility in defining events and choosing which firms to include. In our main specification, we make sure that treated firms are selected only if they have at least three consecutive periods of ETI claiming after they first claim the ETI, and four periods of no-claiming before. This ensures that we have adequate pre- and post-periods to diagnose pre-trend issues and observe dynamic effects. It also means we have only three events. Only firms that first claim the ETI in 2014, 2015,

[17] The matching issues documented in this paper were also evident when using the Callaway and Sant'Anna (2020) estimator and specifying which periods to use for pre-treatment matching.

and 2016 have sufficient post-periods to be included in our analysis (recall that our sample ends in 2018).

We must also choose control firms for each event. Control firms must have at least seven consecutive periods of no-claiming (to match the three post- and four pre-periods), and we additionally impose that they cannot have received treatment any time before these 'clean' seven periods (to ensure our estimates are not contaminated by dynamic treatment effects among control firms). It is possible for both never-treated and not-yet-treated firms to be used as a control. For example, a firm that first claims the ETI in 2017 will have seven periods of non-claiming (2010–16), which means it can be used as a control firm. Firms can be re-used as controls in multiple events. We enforce a balanced panel in the stacked dataset.

Summary statistics of our stacked dataset are given in Table 4. As in our overall balanced panel in Table 2, ETI firms are substantially larger than non-ETI firms and have much faster growth rates of employment. The ETI firms in our stacked panel are slightly larger than those in the overall balanced panel, and do not grow as fast. The percentage differential between the growth rate of ETI and non-ETI firms is, however, much greater in the stacked panel than the overall balanced panel. As in the overall balanced panel, ETI firms have younger and lower-paid employees, and higher separation rates, than non-ETI firms. There are 15,654 treated and 299,746 control firms per year in the stacked panel. The total value of ETI claims made in the three post-periods of the stacked panel is ZAR6.259 billion.

Table 4: Characteristics of ETI and non-ETI firm stacked panel, by event time (means)

Event time	Emp. (all)	Emp. all (growth) (%)	Sep. rate (%)	Youth/hires (%)	Sales (R, m)	Age	Wage (R)
			Non-claiming firms				
–4	24		20		44	39	6,940
–3	25	7	20	42	49	39	7,364
–2	26	1	19	43	55	39	7,853
–1	27	0	18	42	59	40	8,366
0	27	–1	18	41	63	40	8,947
1	27	–1	18	40	66	41	9,597
2	27	–3		40	63	41	10,125
			ETI-claiming firms				
–4	174		27		209	34	4,531
–3	194	13	28	55	240	34	4,751
–2	208	8	28	55	274	34	4,955
–1	214	7	26	54	303	34	5,234
0	233	16	26	59	330	34	5,098
1	238	3	27	56	353	34	5,522
2	243	–1		55	366	34	5,918

Note: the table shows characteristics of firms in the main balanced event-study panel. All statistics are firm averages, though the age and (monthly) wage statistics are averages of firm medians. 'Sep. rate' is the separation rate for that year and 'Youth/hires' is the proportion of hires in that year which are youth. Firms are defined at the PAYE level. Rands (R) are in current prices, and sales are annual and in millions of rands.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

Once the stacked dataset is created, we estimate the (unconditional) DiD event-study according to the specification

$$Y_{ite} = \lambda_{ie} + \varphi_{te} + \sum_{s=-4, s\neq -1}^{2} \beta_s \times \mathbb{1}[t=s] \times D_i + \varepsilon_{ite} \tag{7}$$

The main difference between Equations (1) and (7) is the e indices which indicate 'events'. The unit and time fixed effects λ_{ie} and φ_{te} above are the same as in Equation (1), except that they are now interacted with event indicators, so that there are separate unit and time fixed effects for each event. Otherwise the interpretation of β_s is the same, but we have now ensured that we are only making 'clean' comparisons to identify our treatment effects.[18]

For the conditional event-study DiD results reported in this paper, we use inverse probability weighting (IPW) based on propensity score estimation Abadie (2005). As noted above, the matching issues we identify are robust to a variety of different matching methods, and we only report IPW results for brevity. Propensity scores are estimated using logit models separately for each event, for a given pre-treatment time period t. The outcome is the firm treatment status, and we use a variety of sets of covariates depending on the exercise at hand (e.g. Figure 4 uses different treatment predictors than Figure 5, as noted in their discussion). The estimated score for each firm in that time period is then assigned to the firm for all of its time periods, by event. The propensity score is then converted into an inverse probability weight. We estimate ATT effects, and therefore the weights for treated firms are set to 1. With the inverse probability weights in hand, we estimate our conditional event-study DiD estimates just as above per Equation (7), but weight the observations by their inverse probability weights.

7 Results

7.1 Unconditional baseline DiD

Figure 3 shows the results from unconditional event-study DiD regressions along the lines of Equation (7), separately for outcomes of all employment, young employment, and old employment. 'Old' employees are defined as those older than the ETI cut-off age of 29 (we sometimes use 'non-youth', which is more accurate but also more clumsy). All outcome variables, here and throughout the results sections, are natural logarithms of the counts of employment.[19] The panels on the left-hand side show estimated ETI treatment effects, while those on the right-hand side separately track the conditional evolution (having controlled for firm effects) of the treated and control firms. Differences between the treated and control estimates in the right-hand panels are equivalent to the treatment effects in the left-hand panels. In Figure 3 and throughout the results section, period –1 is used as the omitted reference category, centred at 0.

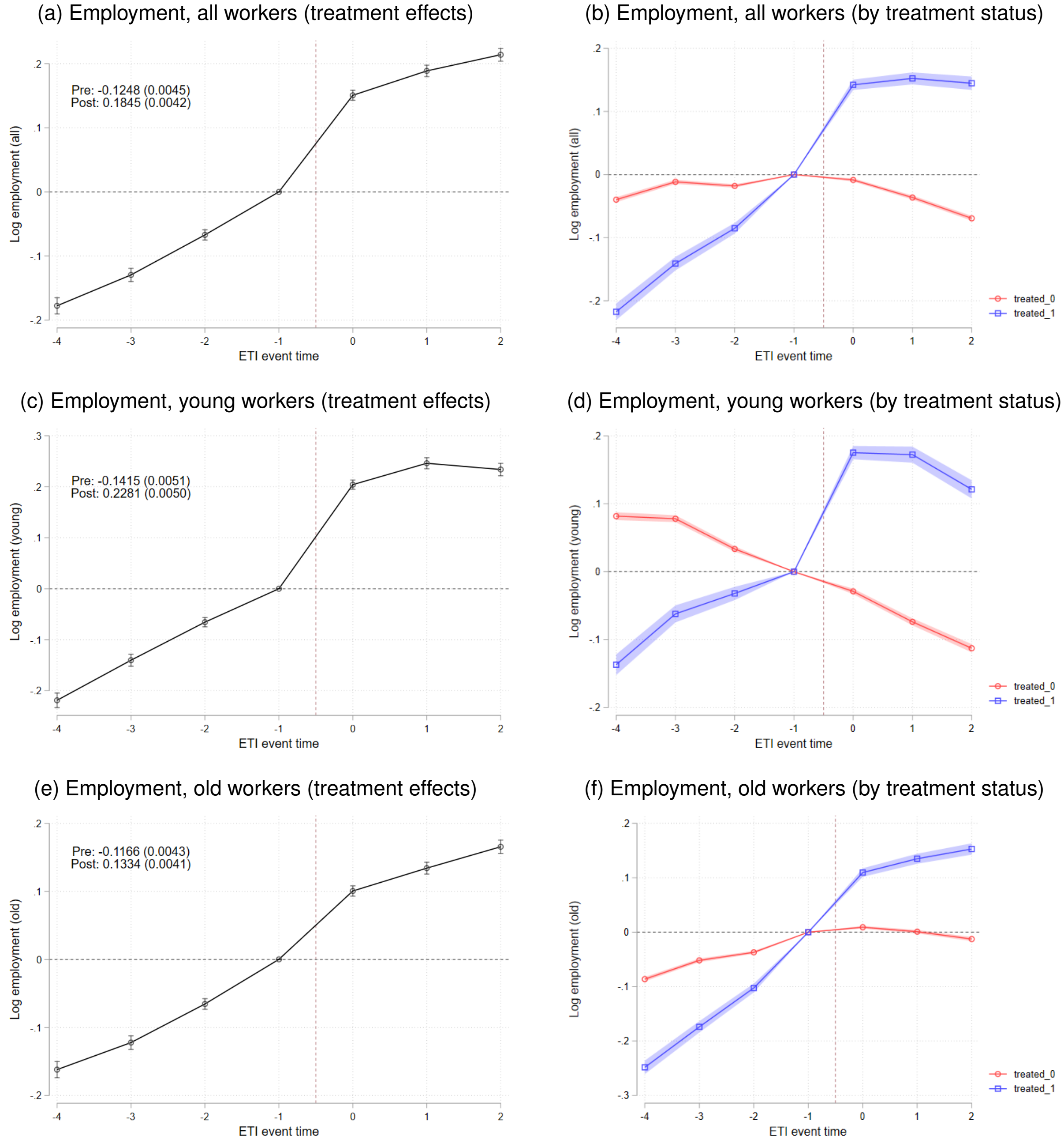
It is immediately evident from Figure 3 that the unconditional parallel trends assumption is highly unlikely to hold. For all outcomes there are strong differential trends between the treatment and control groups in the pre-treatment period, which we expect to continue in the post-period. The aggregate post-period treatment coefficient reported in panel (a) is 0.1845 log points, which converted into a percentage ($(e^0.1845 - 1) \times 100$) is approximately 20 per cent. If the design was credible, this would suggest that the ETI increased employment of ETI firms by one-fifth. However, it clearly is not a credible estimate due to the dramatic pre-trends, which suggest that divergences between ETI and non-ETI firms are not only due to the causal effect of the policy, but also because these are groups of firms on already-different growth paths. The left-hand panels show clearly where these pre-trends come from, with strong employment growth over time for ETI firms, while there is anaemic to negative growth for control firms.

The clear implausibility of the unconditional parallel trends assumption is what motivates the conditional, matched DiD approach (as also used by Ebrahim (2020b) and Bhorat et al. (2020)), which we examine next.

[18] An additional small difference is that our summation operator is now from period –4 to 2, per our choice of four pre-periods and three post-periods. Another issue to note is that standard errors should be clustered by the event, interacted with whatever other level of clustering is used.

[19] We use natural logarithms rather than counts of employment because we expect ETI effects in terms of jobs created to be somewhat proportional to firm size.

Figure 3: Employment event studies, baselines (no matching or parametric trends)



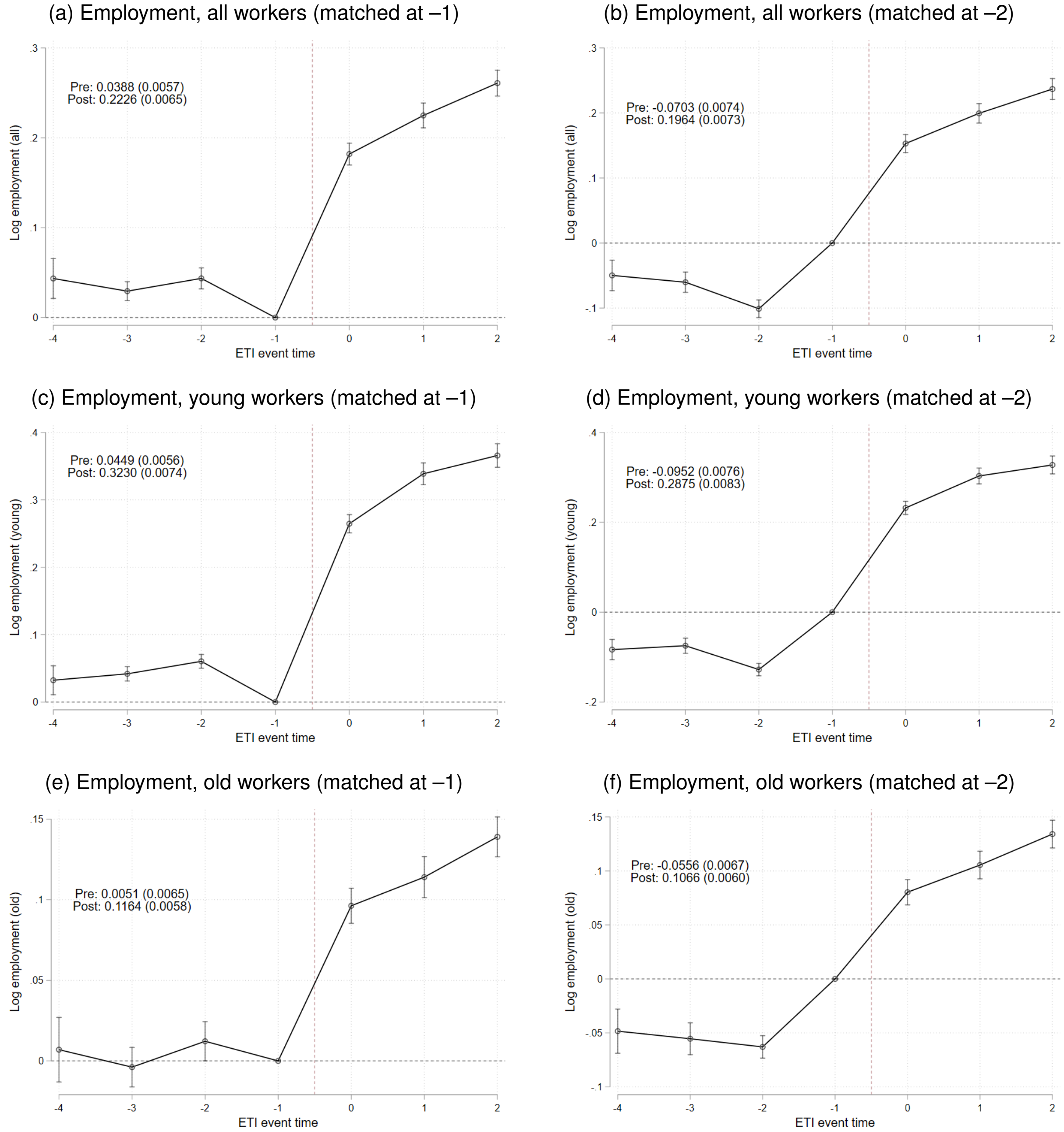
Note: the figure depicts stacked event-study estimates for the 'baseline' DiD specification discussed above, with no matching or parametric trends. Treatment effects are the impact of claiming the ETI. Blue lines and red lines respectively show the conditional evolution of treated and untreated firms' outcomes; differences between these coefficients give the treatment effects. Average pre- and post-period aggregate treatment effects with standard errors in parentheses are shown for the treatment effect figures; 95 per cent confidence intervals are shown with capped spikes and shaded areas, and standard errors are clustered at the firm level.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

7.2 Conditional, matched DiD

Figure 4 presents analogous estimates to Figure 3, but now from a *conditional* DiD event study, using an IPW matching procedure as outlined in Section 6.2. The variables used for the propensity score estimation are the firm median age, the separation rate, firm (CIT) sales, and the share of hires which are youths. Propensity scores are estimated in period –1 for the left-hand panels and period –2 for the right-hand panels. Only the estimated treatment effects are shown.

Figure 4: Employment event studies, IPW matching



Note: the figure depicts stacked event-study treatment effect estimates, with IPW matching. Propensity scores for the IPW matching are estimated using pre-treatment characteristics in period –1 (the immediate pre-treatment period) and period –2 in the left- and right-hand panels respectively. The explanatory variables used for the propensity score estimation are the firm median age, the separation rate, firm (CIT) sales, and the share of hires which are youths. Average pre- and post-period aggregate treatment effects with standard errors in parentheses are shown; 95 per cent confidence intervals are shown with capped spikes and standard errors are clustered at the firm level.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

At first glance, the left-hand panels seem to suggest that the matching procedure works relatively well. In panel (e), showing effects on employment of old workers, the aggregate pre-treatment coefficient is not statistically significantly different from zero, and there is no clear trend. While the pre-period estimates are a bit more messy in panels (a) and (b), they are close to being flat, and if anything the negatively sloping pre-trend they show seems to go in the *opposite* direction to what we would be worried about based on the unconditional estimates in Figure 3. In these circumstances, where the pre-trend is in the opposite direction to the post-treatment effect, researchers are often less worried about the bias from differential trends, reasoning that any bias is likely to cause estimate attenuation rather than more worrying exaggeration (Rambachan and Roth 2020a).

A first point of concern, however, might be noting that the aggregate post-period treatment effects in the left-hand panels of Figure 4 are usually quite similar to or are larger than those in the unconditional regressions in Figure 3. This is by no means an impossible result, but in general one might expect the treatment effect to attenuate as we address the pre-existing differential trends which we expect to cause an upward bias in the unconditional post-treatment effect, per Equation (2).

The problem is definitively revealed by looking at the right-hand panels, which have the same IPW matching procedure but match on period –2 rather than period –1 characteristics. It is instructive to look at panel (f). Pre-trends are flat up until period –2, but then increase and revert to a similar pattern as seen in the unconditional Figure 3(e). The fundamental identifying logic of the matched approach is that by controlling for pre-period characteristics, we remove the *post*-treatment differential trend. Recall that it is *this* post-treatment trend differential that must be 0 for the parallel trends assumption to hold, not the pre-treatment differential trend. The pre-treatment differential trend is only important in so far as it informs us about the likely shape of the post-treatment trend. The break in the trend at period –2 (the period used for matching) in Figure 4(f) suggests that, in this case:

1. The conditional pre-trend is not informative about the trend after the matching period.

2. Matching on pre-period characteristics does not remove post-matching differential trends.

This is evident also from panels (b) and (d), where there is a sharp break after the matching period of –2. We have no reason to think that a similar dynamic is not at play after period –1 in the left-hand panels, which would mean that the treatment effect at period 0 (and after) is contaminated by bias due to differential trends that have not been successfully controlled.

Figure 5 gives some idea of what is happening here, by separately showing the conditional evolution of outcomes for the treated and non-treated groups. For simplicity, we only show the 'all employment' outcome, and propensity scores for the matching procedure in this case are estimated using only the pre-treatment employment level as a treatment predictor.

First, panels (a) and (b) show that this issue persists even when using a different set of matching variables (in this case, just one variable). We tried a variety of different approaches for selecting matching variables, including selection via machine-learning approaches such as LASSO and gradient-boosting, and the issue persists in some form under all specifications. Appendix Figures A1 and A2 show that this issue remains when using roughly the same matching variables as in Ebrahim (2020b) ('ELR') and Bhorat et al. (2020) ('DPRU').[20]

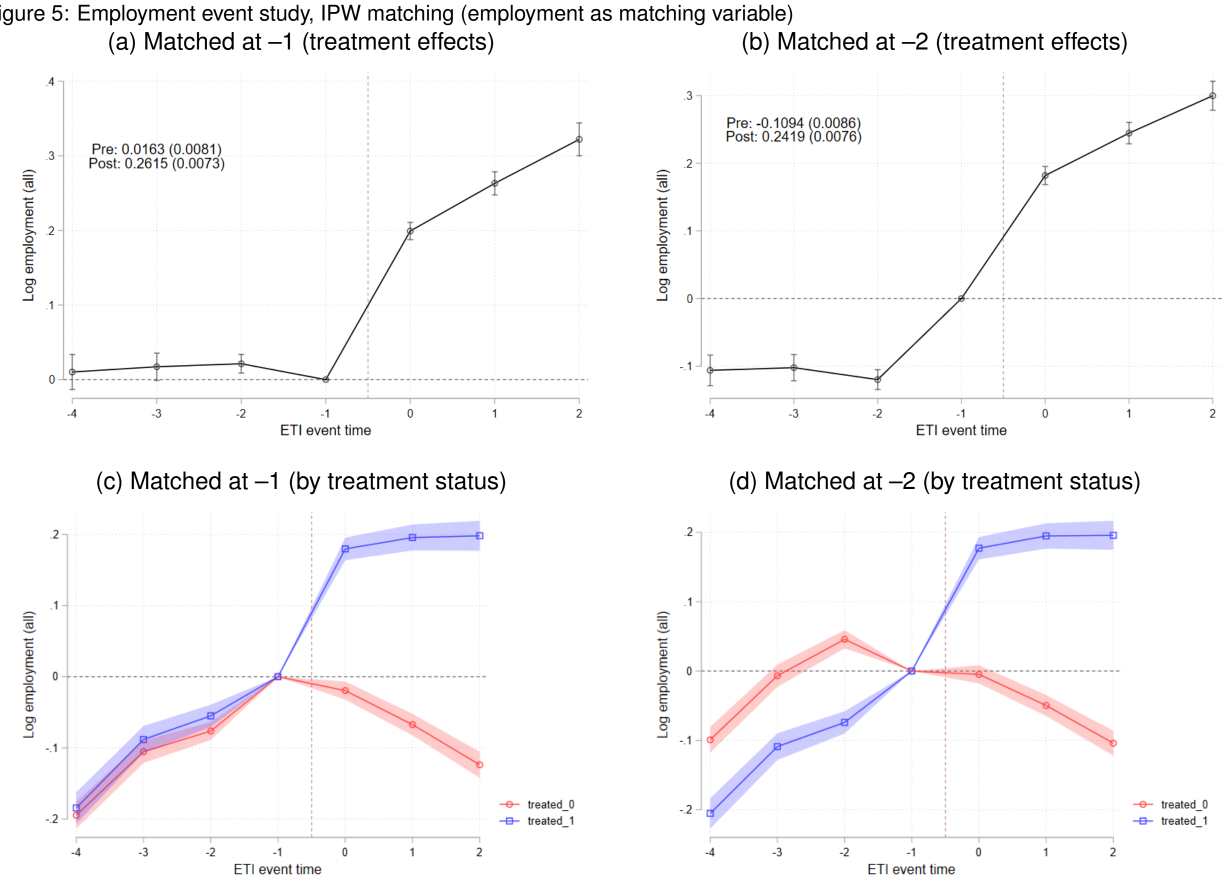
It is, however, more useful to look at panels (c) and (d) of Figure 5. Compared to panel (b) of Figure 3, panel (c) shows that the matching procedure only affects the shape of the *non-treated* line. This is to be expected, as for an ATT the IPW only affects non-treated units, to make them comparable to the treated units. However, what is quite peculiar is that the matching procedure only strongly shifts the non-treated conditional outcomes *in the pre-period*. In the pre-period, the red non-treated line shifts such that it reflects a similar rate of employment growth as the treated firms, but there is little difference in the

[20] Readers may notice that our estimates of treatment effects in these graphs are somewhat larger than those in Ebrahim (2020b) and Bhorat et al. (2020). However, only a few of our outcome variables are directly comparable to those in Ebrahim (2020b) and Bhorat et al. (2020), and in any case our stacked approach with time periods up until 2018 means that we have a very different sample. Our own examination of event-by-event estimates tentatively suggests that estimated ETI 'treatment effects' were smaller in 2015 and 2016 (the periods Ebrahim (2020b) and Bhorat et al. (2020) consider) than the later years that are included in our analysis.

post-period between Figure 3(b) and Figure 5(c). If anything, the post-period decline in employment for non-treated firms is *greater* in the conditional specification in Figure 5(c). Figure 5(d) shows that this is all driven by the matching period. In this graph, the rate of growth of conditional employment is similar for treated and non-treated firms up until period –2 (the period used for matching), after which there is a reversion to negative growth for non-treated firms (and in fact the growth is more negative than in the unconditional case).

This is a serious problem that represents a failure of the matching procedure. The purpose of matching in a conditional DiD setup is to remove post-treatment differential trends; to make the conditional parallel trends assumption hold. However, Figures 4 and 5 show that the matching procedure only resolves divergent trends *up until the time period used for matching*. In periods *after* the one used for matching, trends are either very similar to what we see in the unconditional baseline case, or reflect spurious trend breaks as non-treated firms' outcomes revert to their unconditional levels. This means that post-period coefficients do not represent unbiased estimates of treatment effects. Instead, as per Equation (2), they will be (probably severely) biased by differential post-treatment trends, which remain as divergent as they were in the baseline unconditional cases discussed in Section 7.1. Because ETI firms grow faster than non-ETI firms, the differential trend term in Equation 5 is likely to be positive, so that estimated treatment effects are *upward*-biased.

Figure 5: Employment event study, IPW matching (employment as matching variable)



Note: the figure depicts stacked event-study estimates for the outcome of log employment, with IPW matching. Pre-period log employment is the only variable used to estimate the propensity scores for the IPW matching. Figures on the left use log employment in period –1, while figures on the right use log employment in period –2. Average pre- and post-period aggregate treatment effects with standard errors in parentheses are shown for the treatment effect figures; 95 per cent confidence intervals are shown with capped spikes and shaded areas, and standard errors are clustered at the firm level.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

7.3 Mean reversion and matched DiD

We believe that this failure of the matching process is caused by mean reversion among the non-treated firms, which we explain below. The potential problem of mean reversion for matching DiD estimation does not seem to be well-remarked upon in economics, but has been outlined by Daw and Hatfield (2018) in the field of empirical public health. Daw and Hatfield (2018) consider a two-period before-and-after DiD setting, so they do not discuss the implications of mean reversion for pre-trends. What they show is that in cases where there are:

1. large differences in the *levels* of the outcome, between treated and control group, and
2. low *serial correlation* in the outcome across units

then conditioning on levels of the outcome (or covariates correlated with levels of the outcome) may not resolve a differential trends issue, and in fact can create spurious 'effects'. Low serial correlation in outcomes means that a firm's employment level in one period is not especially informative about its employment level in an adjacent period. Conditioning on levels means finding firms that have similar pre-treatment levels of employment, and restricting the analysis to these firms.[21] The intuition of this issue is fairly straightforward: the matched treatment and control firms are generally very different and only happened to have similar employment levels in the pre-period by random chance (perhaps due to random shocks), and in the next period, because of their low serial correlation, they will tend to revert to somewhere around their usual (very different) mean levels of employment. This mean reversion from their abnormal pre-period levels to their more normal levels will cause differential trends between the matched treated and control firms. With this divergence happening between the pre- and post-periods, this can easily be mistaken for (or bias) an estimated treatment effect. Daw and Hatfield (2018) present simulations showing how this issue can create apparent treatment effects even where none exists.

Is this plausibly the issue in our case of the ETI? It is demonstrably the case that there is a large difference in employment levels between ETI and non-ETI firms. An immediate objection is that it is unlikely that ETI firms have particularly low serial correlation in employment: these are very large, fast-growing firms, which by virtue of claiming the ETI are likely to be fairly professionally run. However, in our case, examining ATT estimates, the serial correlation of employment for *ETI firms* is not the issue: rather, our matching procedure selects (and up-weights) *non-ETI firms* that have similar employment levels (or covariates correlated with employment levels) to ETI firms in the matching period. Non-ETI firms are small, grow slowly if they grow at all, and their non-claiming of the ETI may suggest that they are not particularly professionally run. It is plausible that these firms have relatively low serial correlation in employment.

Daw and Hatfield (2018) do not discuss what mean reversion would do to estimated pre-trends, because they analyse a two-period DiD setting. In general, we might expect mean reversion among non-treated firms to result in an event study that looks something like Appendix Figure A2(a) or (less obviously) Figure 4(a). In these cases, there is a downward-sloping trend divergence before the matching period, and then an upward-sloping trend divergence after the period. This would reflect a process where non-ETI firms matched to ETI firms in period –1 have necessarily had exceptionally high growth before period –1 (as a result of consecutive positive shocks), which is the process by which they reached levels of employment comparable to ETI firms. By conditioning on employment levels which are far out of the ordinary for non-ETI firms, we implicitly select firms which, by chance, have had abnormally high growth before the matching period. This growth may indeed be faster than normal ETI firm growth rates, leading to the downward-sloping pre-trend. However, after period –1, these firms will tend to revert to their usual employment levels, as we are no longer conditioning on employment levels after

[21] Alternatively, in the case of IPW matching, it means weighting these firms more highly.

the matching period. This will lead to sharp negative employment growth after the matching, as seen in Figure 5(d).

We highlight the cases of Appendix Figure A2(a) and Figure 4(a) because the 'V-shaped' pre- and post-trends may immediately be suggestive of mean reversion to some researchers. However Panels (e) and (f) of Figure 4, and Figure 5, suggest that even apparently flat pre-trends may hide mean reversion: it all depends on the growth rate of the non-treated firms needed to match outcomes with treated firms. As far as we are aware, this is a new point when it comes to problems with pre-trend tests, which complements Daw and Hatfield (2018) and Roth (2021). We are currently in the process of creating simulations, calibrated to the moments of the ETI data, which investigate the plausibility of mean reversion for explaining our observed trends.

Lastly, we should note that there is potentially an obvious solution to this issue, discussed by Daw and Hatfield (2018): match on outcome *growth rates* rather than *levels*. However, we found that in the case of the ETI, this performed poorly outside the matching period. This can be seen in Figure A2, because Ebrahim (2020b) matches on both the employment growth rate and its lag. This leads to estimates which look highly plausible, with extremely flat pre-trends. However these trends do not remain flat outside the matching period (see period –4 when matching on period –1, and period –1 when matching on period –2). This accords with the intuition above: if non-treated firms have low serial correlation and only by chance have similar growth rates as treated firms, we should not expect matching on trends to address differential trends outside the matching period (such as post-treatment differential trends, crucially).

7.4 Parametric trends

Point estimates

The difficulty of implementing a credible matched DiD analysis is a substantial impediment to estimating ETI firm-level treatment effects. This is especially frustrating because the unconditional results of Figure 3 do seem to suggest that the ETI may indeed have some effect on employment: there seems to be a trend-break in the estimated treatment effects just as firms first claim the ETI (period 0). This intuition can be formalized by parametrically controlling for differential time trends, so that Equation (7) is adjusted such that:

$$Y_{ite} = \lambda_{ie} + \varphi_{te} + \sum_{s=-4, s \notin \{a,b\}}^{2} \beta_s \times \mathbb{1}[t = s] \times D_i + \rho t_{de}^{\alpha} + \varepsilon_{ite} \tag{8}$$

where t_e^{α} is a continuous time variable t raised to the power α, and interacted with indicators for treatment status d and events e. Rather than trying to restrict comparisons to observably similar firms, this approach seeks to directly model the differential trend between the treated and control firms. Wolfers (2006) shows that one needs to be careful not to control for *post*-treatment dynamics when using this approach, but this issue can be avoided by including sufficient lagged treatment dummies in the model, so that estimation of the time trends comes from omitted pre-periods. One needs at least two omitted pre-periods in order to estimate parametric trends; these are denoted by a and b in Equation (8), and we impose the restrictions $a, b < 0$ and $a \neq b$.

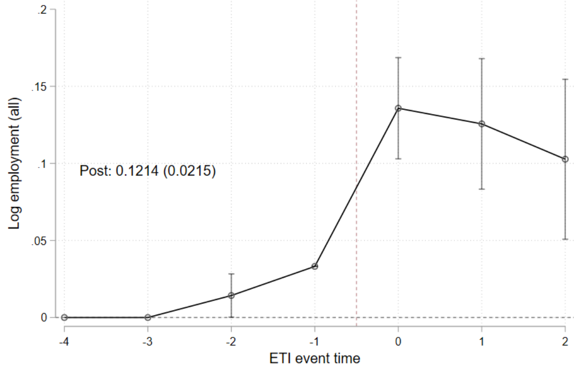
The primary difficulty of this approach is that the researcher must choose a particular parametric trend to impose, which needs to be close to the underlying dynamic for inference to be credible. While linear trends ($\alpha = 1$) are a common approach, there is no reason *a priori* to expect the trend divergence to be linear in the case of ETI-claiming versus -non-claiming firms. We therefore examined how pre-trends looked when imposing linear and quadratic trends for a variety of combinations of omitted periods a and b. It quickly became apparent that while linear trends were sometimes insufficiently convex, quadratic trends were too aggressive. In Figures 6 and 7 we present ‘diagnostic’ event-study estimates for the linear case ($\alpha = 1$) and an intermediate case with $\alpha = 3/2$.

Figures 6 and 7, which calibrate on periods other than –1 to allow for a ‘test’ in the –1 period, show that while linear trends seem to work better for youth employment, $\alpha = 3/2$ trends work better for the all-employment specification. We therefore present both types of trends in Figure 8, where estimates are shown for all, youth, and old employment, with time trends calibrated on periods –4 and –1. Using $\alpha = 3/2$ for all and old employment, and linear trends for youth employment, these estimates would suggest that on average the ETI increased all employment by about 1 per cent (not statistically significant), youth employment by about 8.5 per cent, and decreased old employment by about 2.5 per cent in the three periods after firms first claimed the ETI.

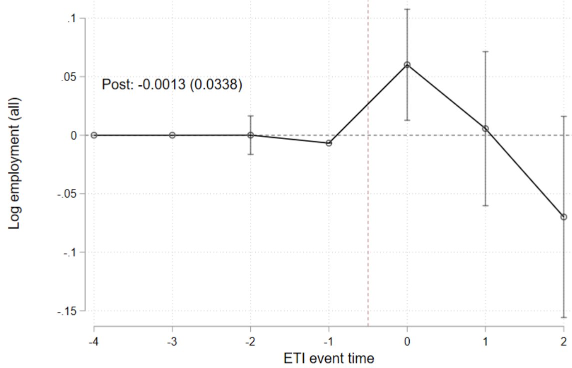
However, there are obvious problems with these point estimates. Most worryingly, they are highly sensitive to the type of trend used: the ‘all employment’ estimate is six times higher with linear trends, while the ‘old employment’ estimate swaps sign. With only four pre-treatment periods, two of which are omitted, we also have very few periods to assess pre-treatment fit. The convex trends (including quadratic trends not shown here) also tend to lead to possibly implausibly large negative treatment effects in the later post-periods: it is hard to imagine why the ETI would decrease employment in this way, and is likely a mechanical consequence of imposing implausible convex growth divergence, which becomes particularly dramatic in later periods. We therefore do not put much stock in the particular point estimates.

Figure 6: Employment (all) event study, testing parametric time trends

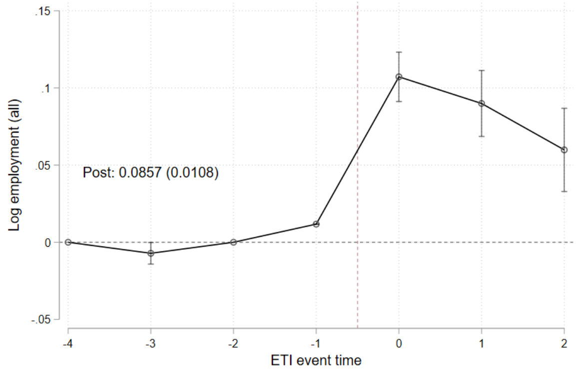
(a) Linear trend (fit on –4 and –3)



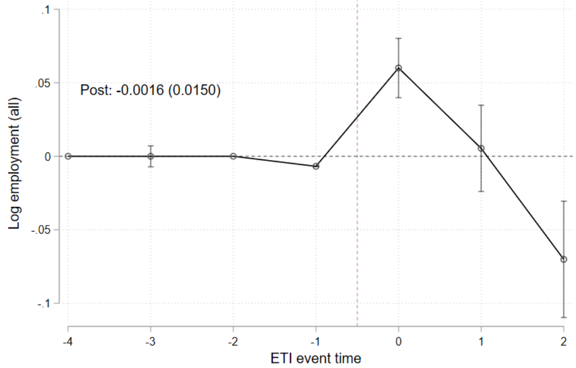
(b) Non-linear trend $t^{3/2}$ (fit on –4 and –3)



(c) Linear trend (fit on –4 and –2)



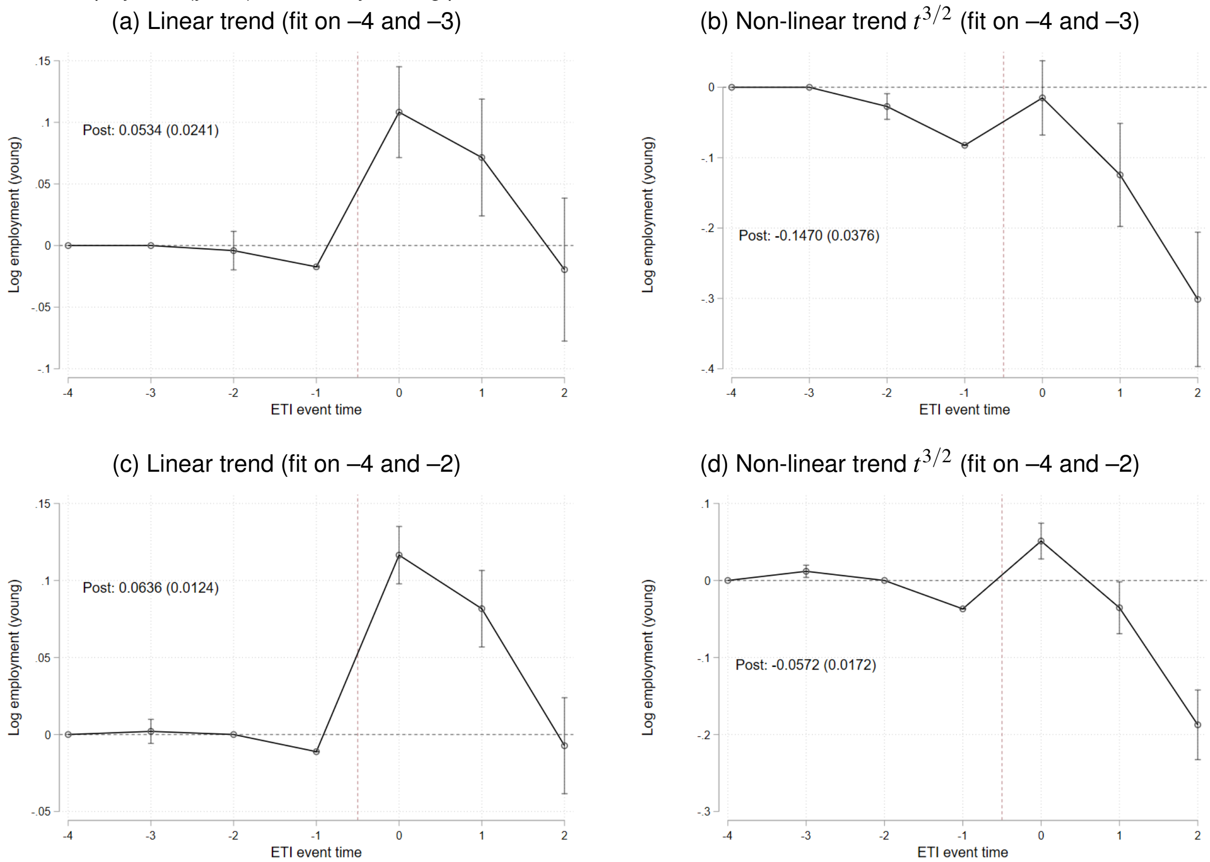
(d) Non-linear trend $t^{3/2}$ (fit on –4 and –2)



Note: the figure depicts stacked event-study estimates for the unconditional DiD specification with parametric trends. The outcome is the natural logarithm of total firm employment. Panels on the left include event-by-treatment linear time trends, while panels on the right include similar time trends but of the form $t^{3/2}$. Time trends in the top panel are fitted to periods –4 and –3. Time trends in the top panel are fitted to periods –4 and –2. Average post-period aggregate treatment effects with standard errors in parentheses are shown; 95 per cent confidence intervals are shown and standard errors are clustered at the firm level.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

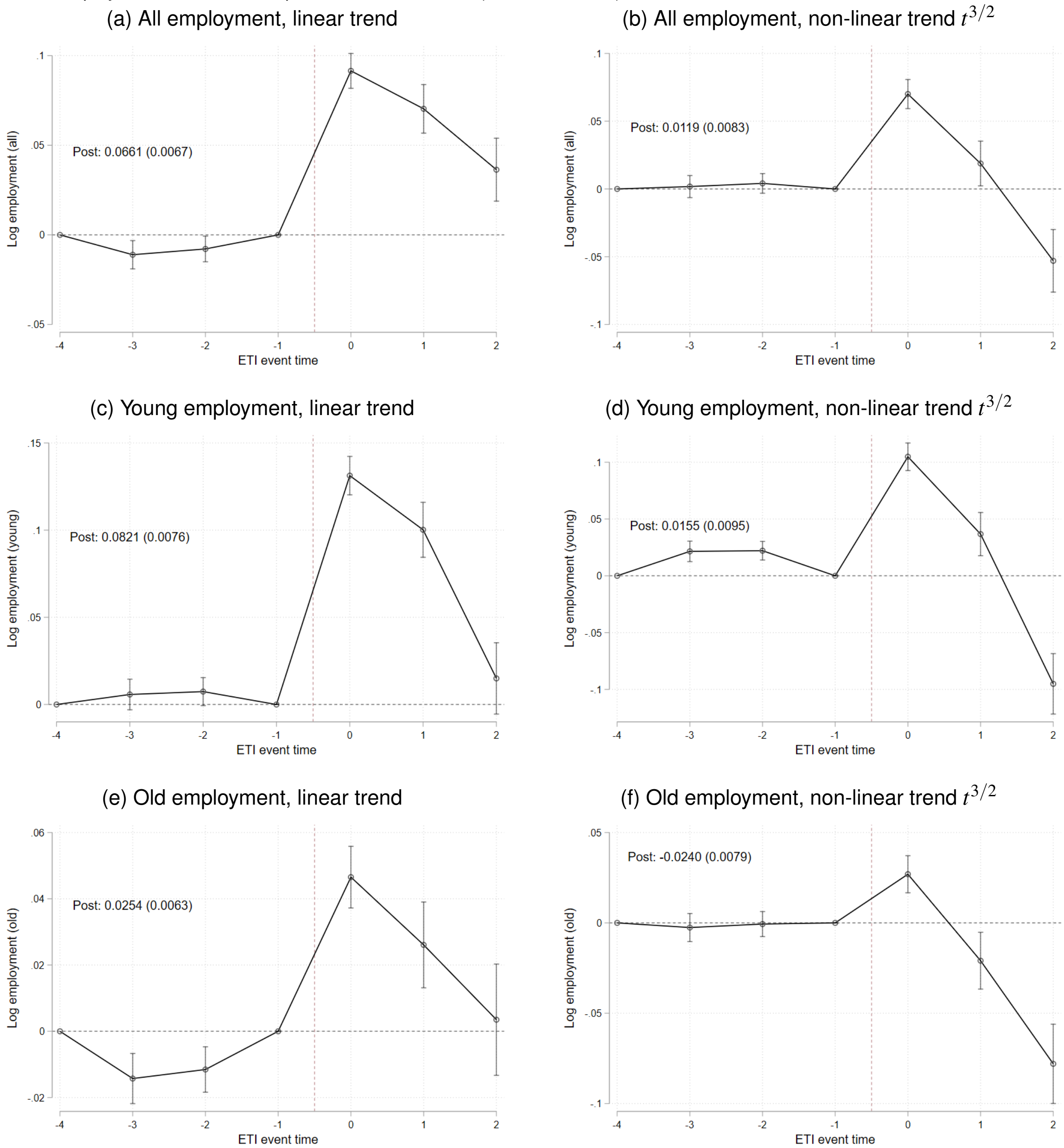
Figure 7: Employment (youth) event study, testing parametric time trends



Note: the figure depicts stacked event-study estimates for the unconditional DiD specification with parametric trends. The outcome is the natural logarithm of firm youth employment. Panels on the left include event-by-treatment linear time trends, while panels on the right include similar time trends but of the form $t^{3/2}$. Time trends in the top panel are fitted to periods –4 and –3. Time trends in the top panel are fitted to periods –4 and –2. Average post-period aggregate treatment effects with standard errors in parentheses are shown; 95 per cent confidence intervals are shown and standard errors are clustered at the firm level.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

Figure 8: Employment event studies, parametric time trends (fit on –4 and –1)



Note: the figure depicts stacked event-study estimates for the unconditional DiD specification with parametric trends. The outcome is the natural logarithm of firm youth, non-youth, and total employment, as indicated. Panels on the left include event-by-treatment linear time trends, while panels on the right include similar time trends but of the form $t^{3/2}$. Time trends are fitted to periods –4 and –1. Time trends in the top panel are fitted to periods –4 and –2. Average post-period aggregate treatment effects with standard errors in parentheses are shown; 95 per cent confidence intervals are shown and standard errors are clustered at the firm level.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

Partial identification

Given the failure of matching approaches, the idea of directly modelling the post-treatment trend divergence nonetheless remains appealing. We conclude with *partial identification* results using the approach of Rambachan and Roth (2020a). Partial identification allows us to identify a *set* of estimates that are consistent with a class of restrictions. We do not identity a particular point estimate, but a range of values and associated confidence intervals. If the restrictions are weak and/or the underlying data are not particularly suggestive of an effect, we may not be able to conclude much. However with strong restrictions and/or data that are highly suggestive of effects, we can sometimes make meaningful conclusions (e.g. 'there is a positive effect'), even if we cannot identify the particular point estimate.

To make full use of the Rambachan and Roth (2020a) method, one needs to impose restrictions informed by the particular context and economic knowledge. This is a topic of further investigation for us, and for now we simply demonstrate the method using their $\Delta^{SD}(M)$ restriction, which is based on smoothness of the trend and a limit on how far the trend can diverge from linearity. Specifically, the restriction is defined for each value M, where M quantifies the allowable divergence from linearity according to:

$$\Delta^{SD}(M) = \{\delta : |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leq M, \forall t\} \tag{9}$$

This expression uses the same notation as in Equations (2) and (3), so that δ_t is the differential trend between the (non-treated) potential outcomes of the treated group versus those of the control group in period t. The vector δ contains the differential trends δ_t for all t. The parameter $M \geq 0$ restricts the degree to which the slope of δ can change between consecutive periods. When M is large, we allow significant departures from linear trends. When $M = 0$, we require that the differential trend be linear (the change in the slope over time is constant). Rambachan and Roth (2020a) develop methods to conduct valid inference given a set of restrictions such as $\Delta^{SD}(M)$. Following their recommendations for restrictions of the form $\Delta^{SD}(M)$, we use their fixed length confidence interval (FLCI) approach.

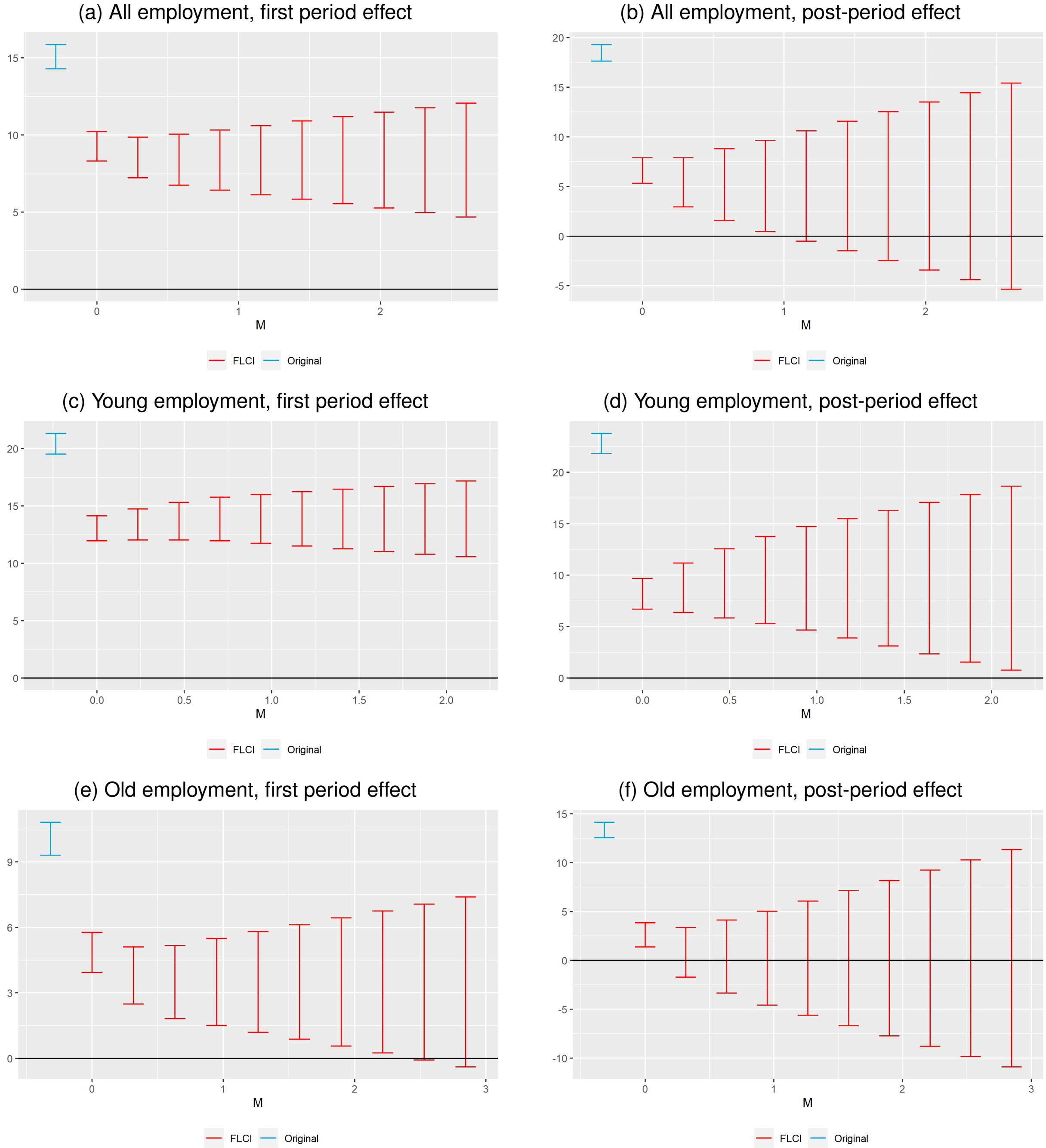
Figure 9 presents confidence intervals for ranges of treatment effects that are consistent with a given M, from our unconditional results estimated by Equation (7). The left-hand panels of Figure 9 show treatment effects in the immediate post-treatment period, while the right-hand panels show average treatment effects across the post-periods. Note that treatment effects are scaled by being multiplied by 100, compared to the estimates in Figure 3. The blue confidence intervals in Figure 9 are those of the unconditional estimates shown in Figure 3.

The range of Ms considered come from the Rambachan and Roth (2020a) default procedure, based on the standard error of the first post-treatment coefficient. As noted above, ideally we would select our own range of plausible restrictions based on the particular context of the ETI. For now, we simply note that given our particular effect scaling, $M = 2$ allows quadratic trend differentials, which we view as an implausibly large divergence from linearity based on the pre-period trends.[22] This means that the ranges of M shown in Figure 9 probably (over-)cover what we would consider plausible deviations from linearity.

If we take $M = 2$ as the smallest value of M we would consider implausible, the left-hand panels of Figure 9 suggest that the immediate post-treatment effect is robustly positive, for all employment outcomes. That is, we can reject a zero or negative immediate post-treatment effect for the plausible range of differential trends. However, when looking at the aggregate effect of the ETI over the post-treatment period, a different picture emerges. We certainly cannot reject a zero or negative effect of the ETI on non-youth employment: even at mild deviations from linearity the confidence set includes zero and negative values (panel (f)). The case of all employment is more difficult: zero and negative effects cannot be rejected around $M = 1$ (panel (a)). Does $M = 1$ represent an implausibly large deviation from linearity? This would allow $\alpha \approx 1.7$, per Equation (8). We make no judgement either way. However, what *is* striking is that a positive aggregate youth employment effect is distinguishable from 0 at all plausible values of M. While this partial identification approach cannot give us a point estimate or the number of jobs created by the ETI, this latter result is perhaps useful, given the existing conflicting evidence on youth employment effects from Ranchhod and Finn (2015, 2016) and Ebrahim (2020a).

[22] We say 'allows' rather than 'equals' because any given value of M allows concave as well as convex deviations from linearity.

Figure 9: Employment event studies, partial identification using the Rambachan and Roth (2020a) method



Note: the figure depicts confidence intervals for ETI treatment effects, per the partial identification approach of Rambachan and Roth (2020a). The blue confidence interval shows the confidence interval associated with the estimate from the baseline unconditional event-study regression. The orange confidence intervals show the confidence interval of the range of possible treatment effect estimates associated with the given allowable deviation from linear differential trends, denoted by M, for a constraint of the form $\Delta^{SD}(M)$ (using the FLCI method). See the text for a discussion. Left-hand panels show confidence intervals for the treatment effect in the first period after claiming the ETI, while right-hand panels show the same but for the total average effect over the three post-periods. Nintey-five per cent confidence intervals are shown and standard errors are clustered at the firm level.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

8 Conclusion

The primary contribution of this paper has been to demonstrate the difficulties associated with estimating credible firm-level employment effects of the South African ETI. We first characterize ETI and non-ETI firms, and show that these are very different types of firms. ETI firms are larger, have faster employment growth, higher turnover, younger workers, and pay lower wages than non-ETI firms. This means that unconditional DiD approaches, comparing ETI to non-ETI firms, do not result in credible estimates of treatment effects, as the parallel trends assumption is certainly violated. We demonstrate the improbability of the parallel trends assumption in this context by presenting unconditional event-study estimates that show pre-treatment trend divergences.

We then investigate the credibility of conditional DiD methods, which match ETI firms to comparable non-ETI firms, and which have been used by existing studies to estimate firm-level ETI treatment effects (Bhorat et al. 2020; Ebrahim 2020b). We show that these approaches are highly sensitive to the period used for matching, that pre-trends under these circumstances are not informative about post-treatment differential trends, and that treatment effects estimated using these approaches are very likely to be significantly upward-biased, as the conditional parallel trends assumption is very unlikely to hold. We explain the failure of the matched DiD approach as likely being a consequence of mean reversion among matched non-ETI firms. Methodologically, our discussion of mean reversion identifies a seemingly new case where 'flat' event-study pre-trends may be deceiving when it comes to design validity, which contributes to an emerging literature on this topic (see Roth 2021). Substantively, the issue we identify is driven by ETI and non-ETI firms being so fundamentally different that credible matches are just very rare.

We end by trying to parametrically control for differential trends between ETI and non-ETI firms. While the unconditional DiD event-study estimates do seem to suggest a trend-break in estimated treatment effects when the subsidy is claimed, point estimates with parametric trends are ultimately highly sensitive to the particular trends imposed. We therefore implement a partial identification approach from Rambachan and Roth (2020a). While little can be said about firm-level total employment effects without further assumptions, and even less about non-youth employment effects, we find fairly robust evidence that the ETI did increase firm-level youth employment.

How should one evaluate the existing ETI literature, given our findings here? Our view is that little weight should be put on employment estimates from the firm-level matched DiD analyses. It is possible that the estimates in Ebrahim (2020b) are more credibly identified than the matched DiD we implement here, because Ebrahim (2020b) drops large firms from her sample, which will reduce employment-level differences between ETI and non-ETI firms. However, common support restrictions do not substantively alter our results here. Given the issues we identify, our view is that going forward, firm-level matched DiD analysis of the ETI needs to show pre-trends and sensitivity of trends to the period used for matching, in addition to the usual checks, in order for estimates to be credible. Excluding the existing firm-level approaches, the remaining (worker-level) ETI analyses find zero or very small effects of the ETI on employment (Ebrahim 2020a; Ranchhod and Finn 2015, 2016). On this basis, one could conclude that the ETI has not increased employment. However these estimates rely on there being no spillover effects on ineligible workers, an assumption that may not hold. Additionally, our partial identification results do suggest some positive impact of the ETI on youth employment. Ultimately, we remain agnostic about whether the ETI has had positive employment effects.

One issue that we have not discussed is the possibility of endogenous treatment adoption. That is, firms may choose to enrol in the ETI when they know that they are about to increase employment anyway. This would be consistent with our partial identification results, which suggest that ETI firms differentially increased their employment after claiming the ETI.[23] In this case, our partial identification results would *not* identify treatment effects. Given the difficulties associated with firm-level approaches that compare ETI and non-ETI firms, there may be scope for firm-level approaches that leverage other kinds of firm-level variation in ETI treatment. One could perhaps leverage the peculiar design of the ETI, where low minimum wage or high youth share firms get more subsidy, and dominated wage regions create bunching in the firm-wage distribution. Investigating the feasibility of approaches that use this kind of variation constitutes part of our ongoing research on this topic.

References

Abadie, A. (2005). 'Semiparametric Difference-In-Differences Estimators'. *Review of Economic Studies*, 72(1): 1–19. https://doi.org/10.1111/0034-6527.00321

Baker, A., D.F. Larcker, and C.C. Wang (2021). 'How Much Should We Trust Staggered Difference-In-Differences Estimates?' Finance Working Paper 736/2021. Brussels: European Corporate Governance Institute. https://doi.org/10.2139/ssrn.3794018

Bhorat, H., R. Hill, S. Khan, K. Lilenstein, and B. Stanwix (2020). 'The Employment Tax Incentive in South Africa: An Impact Assessment'. DPRU Working Paper 202007. Cape Town: Development Policy Research Unit, University of Cape Town.

Bischof, D. (2017). 'New Graphic Schemes for Stata: Plotplain and Plottig'. *Stata Journal*, 17(3): 748–59. https://doi.org/10.1177/1536867X1701700313

Bravo, M.C. (2018). 'GTOOLS: Stata Module to Provide a Fast Implementation of Common Group Commands'. Statistical Software Components. Boston, MA: Boston College Department of Economics.

Callaway, B., and P.H. Sant'Anna (2020). 'did: Difference in Differences'. R package version 2.0.1.907.

Callaway, B., and P.H. Sant'Anna (2020). 'Difference-In-Differences with Multiple Time Periods'. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.12.001

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). 'The Effect of Minimum Wages on Low-Wage Jobs'. *Quarterly Journal of Economics*, 134(3): 1405–54. https://doi.org/10.1093/qje/qjz014

Correia, S. (2014). 'REGHDFE: Stata Module to Perform Linear or Instrumental-Variable Regression Absorbing Any Number of High-Dimensional Fixed Effects'. Statistical Software Components. Boston, MA: Boston College Department of Economics.

Daw, J.R., and L.A. Hatfield (2018). 'Matching and Regression to the Mean in Difference-In-Differences Analysis'. *Health Services Research*, 53(6): 4138–56. https://doi.org/10.1111/1475-6773.12993

Ebrahim, A. (2020a). 'Individual-Level Responses to a Firm-Side Subsidy'. In: 'A Policy for the Jobless Youth: The Employment Tax Incentive'. Unpublished PhD Thesis. Cape Town: University of Cape Town.

Ebrahim, A. (2020b). 'Estimating Firm-Level Impacts of the ETI'. In: 'A Policy for the Jobless Youth: The Employment Tax Incentive'. Unpublished PhD Thesis. Cape Town: University of Cape Town.

Ebrahim, A., and J. Pirttilä (2019). 'Can a Wage Subsidy System Help Reduce 50 Per Cent Youth Unemployment: Evidence from South Africa'. WIDER Working Paper 2019/28. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2019/662-3

Ebrahim, A., C.F. Kreuser, and M. Kilumelume (2021). 'The Guide to the CIT-IRP5 Panel Version 4.0'. WIDER Working Paper [forthcoming]. Helsinki: UNU-WIDER.

[23] It would also be consistent with the findings in Ebrahim (2020b) and Bhorat et al. (2020), where employment of non-young workers increases after the ETI is claimed, though we expect that the issue of differential post-treatment trends discussed in this paper may be more important in explaining these results.

Ebrahim, A., M. Leibbrandt, and V. Ranchhod (2017). 'The Effects of the Employment Tax Incentive on South African Employment'. WIDER Working Paper 2017/5. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2017/229-8

Goldschmidt, D., and J.F. Schmieder (2017). 'The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure'. *Quarterly Journal of Economics*, 132(3): 1165–217. https://doi.org/10.1093/qje/qjx008

Goodman-Bacon, A. (2018). 'Difference-in-Differences with Variation in Treatment Timing'. NBER Working Paper 25018. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w25018

Heckman, J.J., H. Ichimura, and P.E. Todd (1997). 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme'. *Review of Economic Studies*, 64(4): 605–54. https://doi.org/10.2307/2971733

Jann, B. (2017). 'KMATCH: Stata Module for Multivariate-Distance and Propensity-Score Matching, Including Entropy Balancing, Inverse Probability Weighting, (Coarsened) Exact Matching, and Regression Adjustment'. Statistical Software Components. Boston, MA: Boston College Department of Economics.

Kerr, A. (2020). 'Earnings in the South African Revenue Service IRP5 Data'. WIDER Working Paper 2020/62. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/819-1

King, G., and R. Nielsen (2019). 'Why Propensity Scores Should Not be Used for Matching'. *Political Analysis*, 27(4): 435–54. https://doi.org/10.1017/pan.2019.11

Marcelin, I., D. Brink, D.O. Fadiran, and H.A. Amusa (2019). 'Subsidized Labour and Firms: Investment, Profitability, and Leverage'. WIDER Working Paper 2019/50. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2019/684-5

National Treasury of South Africa (2019). *Budget Review 2019*. Pretoria: National Treasury of South Africa.

National Treasury and UNU-WIDER (2021a). 'CIT-IRP5 Firm Panel 2008–2018' [dataset]. Version 4.0. Pretoria: South African Revenue Service [producer of the original data], 2020. Pretoria: National Treasury and UNU-WIDER [producer and distributor of the harmonized dataset], 2021.

National Treasury and UNU-WIDER (2021b). 'IRP5 Worker-Level Data 2011–2018' [dataset]. Version 4.0. Pretoria: South African Revenue Service [producer of the original data], 2020. Pretoria: National Treasury and UNU-WIDER [producer and distributor of the harmonized dataset], 2021.

Pieterse, D., E. Gavin, and C.F. Kreuser (2018). 'Introduction to the South African Revenue Service and National Treasury Firm-Level Panel'. *South African Journal of Economics*, 86: 6–39. https://doi.org/10.1111/saje.12156

Rambachan, A. and J. Roth (2020a). 'An Honest Approach to Parallel Trends'. Unpublished, 12 November version.

Rambachan, A., and J. Roth (2020b). Honestdid. R package version 0.1.0.

Ranchhod, V., and A. Finn (2015). 'Estimating the Effects of South Africa's Youth Employment Tax Incentive: an Update'. SALDRU Working Paper 152. Cape Town: South Africa Labour and Development Research Unit.

Ranchhod, V., and A. Finn (2016). 'Estimating the Short Run Effects of South Africa's Employment Tax Incentive on Youth Employment Probabilities Using a Difference-in-Differences Approach'. *South African Journal of Economics*, 84(2): 199–216. https://doi.org/10.1111/saje.12121

Roth, J. (2021). 'Pre-Test with Caution: Event-Study Estimates after Testing for Parallel Trends'. Unpublished, 20 April version.

Saez, E., B. Schoefer, and D. Seim (2019). 'Payroll Taxes, Firm Behavior, and Rent Sharing: Evidence from a Young Workers' Tax Cut in Sweden'. *American Economic Review*, 109(5): 1717–63. https://doi.org/10.1257/aer.20171937

SARS (n.d.). 'Guide for Employers in Respect of Employment Tax Incentive'. External Guide PAYE-GEN-01-G05 (Rev. 8). Pretoria: South African Revenue Services.

SARS (2019). 'How Does the Employment Tax Incentive (ETI) Work?'. Available at: https://www.sars.gov.za/TaxTypes/PAYE/ETI/Pages/How-does-the-Employment-Tax-Incentive-(ETI)-work.aspx. (accessed 27 September 2019).

Sun, L., and S. Abraham (2020). 'Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects'. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.09.006 https://doi.org/10.1016/j.jeconom.2020.09.006

Wickham, H., M. Averick, J. Bryan, W. Chang, L.D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T.L. Pedersen, E. Miller, S.M. Bache, K. Müller, J. Ooms, D. Robinson, D.P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani (2019). 'Welcome to the tidyverse'. *Journal of Open Source Software*, 4(43): 1686. https://doi.org/10.21105/joss.01686

Wolfers, J. (2006). 'Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results'. *American Economic Review*, 96(5): 1802–20. https://doi.org/10.1257/aer.96.5.1802

Appendix A: Figures and tables

Table A1: Characteristic of ETI and non-ETI firm unbalanced panel, by tax year (means)

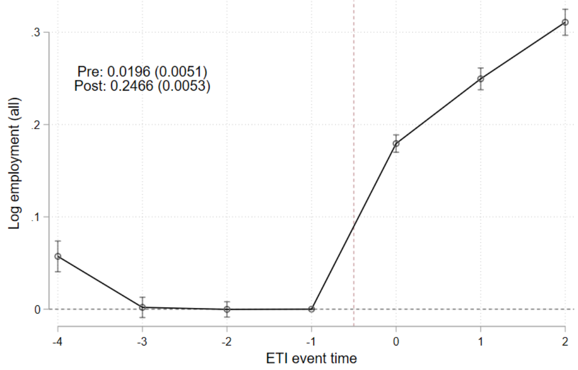
Tax year	Firms	Employment		Emp. growth (%)		Sep. rate (%)	Hire rate (%)	Sales (R, m)	Sales/emp. (R, m)	Age	Wage (R)
		All	Young	All	Young						
Non-claiming firms											
2010	184,838	23	7			32		57	2.05	38	6,335
2011	192,728	23	7	51	25	31	38	52	1.18	39	6,407
2012	194,646	22	7	24	9	31	33	55	1.25	39	6,895
2013	192,704	21	7	18	8	30	30	50	1.44	39	7,300
2014	192,050	20	6	16	3	29	30	51	1.59	39	8,056
2015	194,835	19	6	19	3	28	30	44	1.65	40	8,587
2016	198,013	19	5	16	3	29	29	43	1.70	40	9,242
2017	201,268	18	5	17	3	28	30	46	1.89	40	9,996
2018	207,412	18	5	20	4		31	31	1.83	40	10,611
ETI-claiming firms											
2010	34,029	114	47			26		173	1.27	35	5,350
2011	37,409	120	50	111	66	29	40	173	1.14	35	5,155
2012	39,848	124	52	63	47	28	37	205	1.18	35	5,333
2013	43,544	125	53	62	52	28	39	217	1.30	35	5,462
2014	47,512	125	52	64	43	27	38	225	1.22	35	5,625
2015	51,954	123	51	61	50	30	39	226	1.26	34	5,708
2016	54,359	119	49	38	29	32	35	224	1.38	35	6,032
2017	55,330	119	48	44	20	32	34	233	1.65	35	6,616
2018	54,835	121	48	56	35		32	191	1.45	35	6,957

Note: the table shows characteristics of the unbalanced panel of firms in the study period, by tax year. All statistics are firm averages, though the age and (monthly) wage statistics are averages of firm medians. 'Sep. rate' is the separation rate for that year and 'Sales/emp.' the value of sales per worker at the CIT level. Firms are defined at the PAYE level. Rands (R) are in current prices, and sales and sales/emp. are annual and in millions of rands.

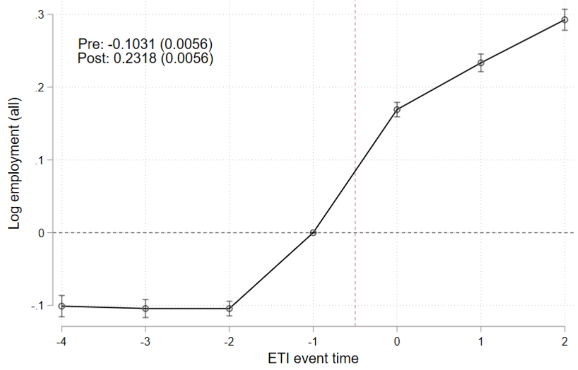
Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

Figure A1: Employment event studies, IPW matching (approximate ELR spec.)

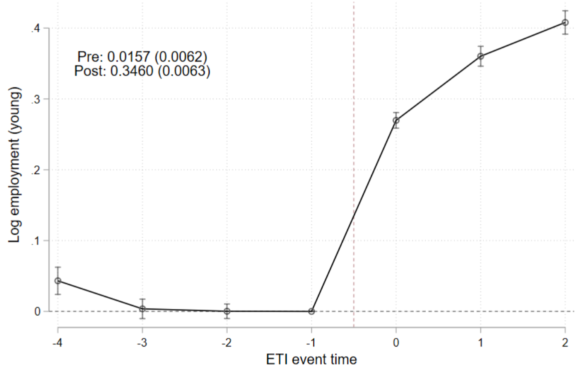
(a) Employment, all workers (matched at –1)



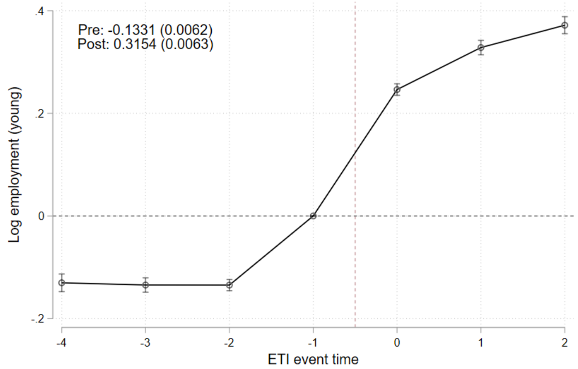
(b) Employment, all workers (matched at –2)



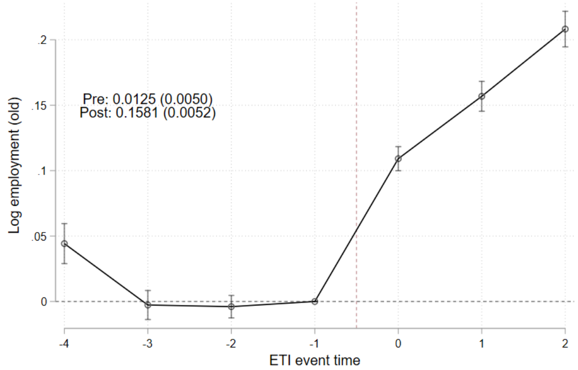
(c) Employment, young workers (matched at –1)



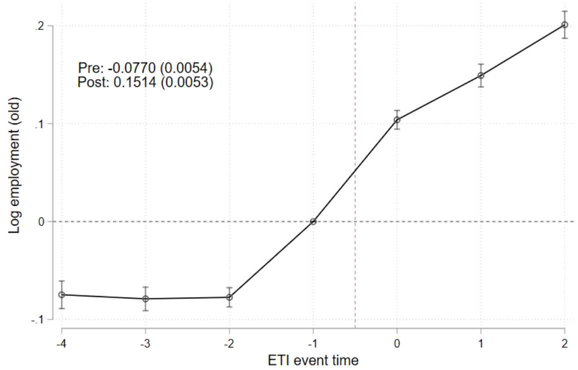
(d) Employment, young workers (matched at –2)



(e) Employment, old workers (matched at –1)



(f) Employment, old workers (matched at –2)

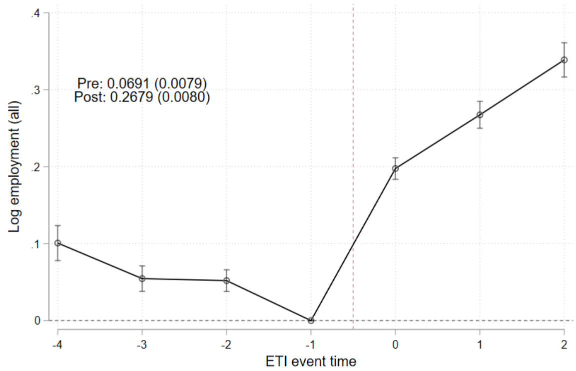


Note: the figure depicts stacked event-study treatment effect estimates, with IPW matching. Propensity scores for the IPW matching are estimated using pre-treatment characteristics in period –1 (the immediate pre-treatment period) and period –2 in the left- and right-hand panels, respectively. The explanatory variables used for the propensity score estimation are roughly equivalent to those in Ebrahim (2020b): a categorical firm employment variable, firm (CIT) fixed assets, firm (CIT) sales, firm (CIT) one-digit industry, firm (PAYE) province, average firm wage, average firm age, employment growth, youth employment growth, lagged employment growth, and lagged youth employment growth. Average pre- and post-period aggregate treatment effects with standard errors in parentheses are shown; 95 per cent confidence intervals are shown with capped spikes and standard errors are clustered at the firm level.

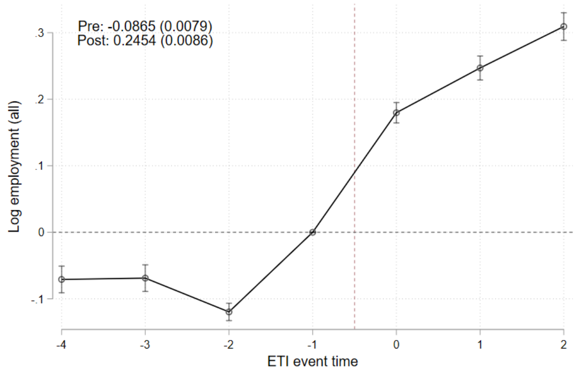
Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

Figure A2: Employment event studies, IPW matching (approximate DPRU spec.)

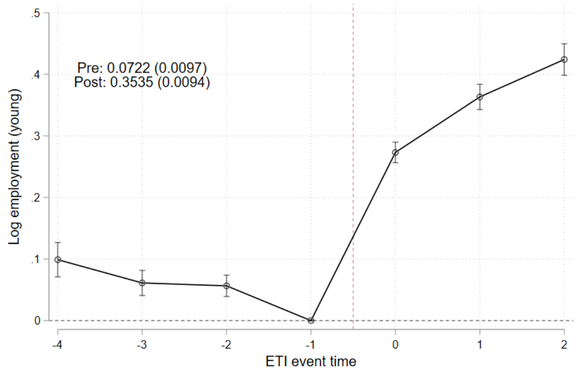
(a) Employment, all workers (matched at –1)



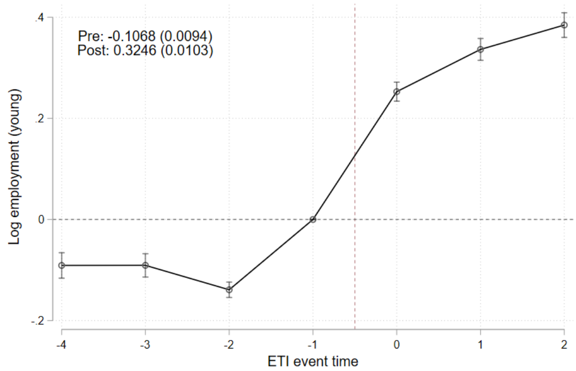
(b) Employment, all workers (matched at –2)



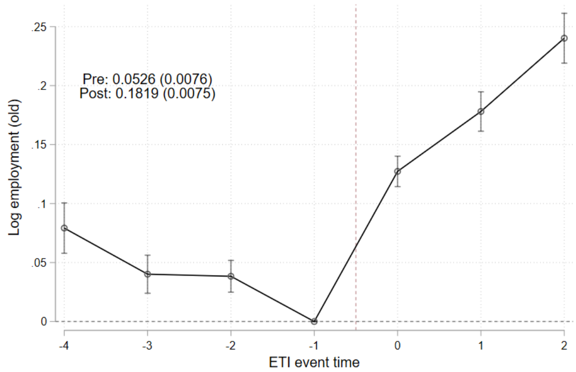
(c) Employment, young workers (matched at –1)



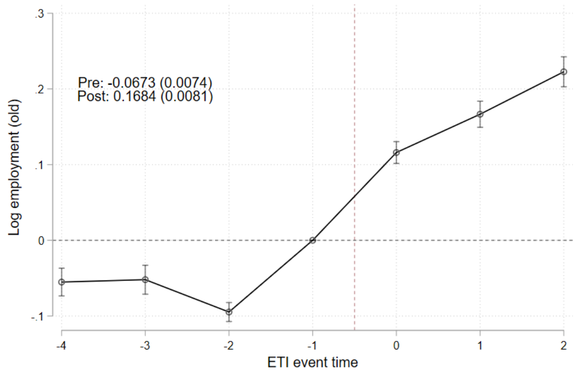
(d) Employment, young workers (matched at –2)



(e) Employment, old workers (matched at –1)



(f) Employment, old workers (matched at –2)



Note: the figure depicts stacked event-study treatment effect estimates, with IPW matching. Propensity scores for the IPW matching are estimated using pre-treatment characteristics in period –1 (the immediate pre-treatment period) and period –2 in the left- and right-hand panels, respectively. The explanatory variables used for the propensity score estimation are roughly equivalent to those in Bhorat et al. (2020): log employment, firm (CIT) fixed assets, firm (CIT) sales, firm (CIT) one-digit industry, firm (CIT) import–export status, average firm wage, average firm age, employment growth, and youth employment growth. Average pre- and post-period aggregate treatment effects with standard errors in parentheses are shown; 95 per cent confidence intervals are shown with capped spikes and standard errors are clustered at the firm level.

Source: authors' compilation based on National Treasury and UNU-WIDER (2021a,b).

Appendix B: Data

This data appendix is created as per UNU-WIDER requirements for users of the National Treasury Secure Data Facility (NT-SDF). It reports on data directly used for the results presented in this paper, and does not include other variables and programs used in our ongoing research on this topic.

Data access

The data used for this research was accessed from the NT-SDF. Access was provided under a non-disclosure agreement, and our output was checked so that the anonymity of no firm or individual would be compromised. Our results do not represent any official statistics (NT or SARS). Similarly, the views expressed in our research are not necessarily the views of the NT or SARS.

Data used: CIT-IPR5 panel (`citirp5_v4`) and year-by-year IRP5 job-level data (`v4`). Date of first access for this project: 19 October 2020. Last accessed: 10 June 2021.

Software

Our analysis was primarily conducted using Stata 16. User-written programs and schemes used include `reghdfe` (Correia 2014), `gtools` (Bravo 2018), `kmatch` (Jann 2017), and `plotplain` (Bischof 2017). Some analysis used `R` version 4.0.2. Packages used include `tidyverse` (Wickham et al. 2019), `HonestDiD` (Rambachan and Roth 2020b), and `did` (Callaway and Sant'Anna 2020).

Variables

Variables used from the raw IRP5 data include: `taxyear taxrefno payereferenceno dateofbirth gender idno passportno province_geo busprov_geo periodemployedfrom periodemployedto totalperiodsinyearofassessment totalperiodsworked employertaxincentiveamnt employmenttaxincentiveind`.

Employment income was created from the following IRP5 amount codes: `amt3601 amt3605 amt3606 amt3607 amt3615 amt3616`. A record of employment-related allowances was created from the following IRP5 amount codes: `amt3701 amt3704 amt3710 amt3711 amt3712 amt3713 amt3715`. A fringe benefits variable was created from the following IRP5 amount codes: `amt3801 amt3802 amt3803 amt3804 amt3805 amt3806 amt3807 amt3808 amt3809 amt3810 amt3813 amt3815 amt3816 amt3817 amt3820 amt3821 amt3825 amt3828 amt3829 amt3830 amt3831 amt3832 amt3833`.

An ETI-applicable employment remuneration variable was created from the sum of the three quantities above. IRP5 employment records were identified by records that had non-zero income or allowances; those with zero or missing income and allowances data are dropped from the analysis.

Variables used from the CIT-IRP5 data include: `taxyear taxrefno g_sales k_ppe k_faother imp_mic_sic7_1d comp_prof_sic5_1d cust_impexpindicator`.

Cleaning and sample notes

PAYE entities without CIT tax reference numbers are excluded from the sample, on the basis of public sector firms not being eligible for the ETI subsidy. CIT entities not matched to PAYE entities are also excluded, as this is primarily an employment analysis and the ETI is only applicable to firms that have employees. Our ETI amount variable was cleaned at the individual level before aggregation, by setting it to zero when it was claimed for individuals who were too old, or who were hired at the PAYE entity before the 2013 eligibility date. In cases where the annual claim value was larger than ZAR2,000 in the 2014 tax year or larger than ZAR12,000 in subsequent tax years, it was reduced to these levels as these are the maximum subsidies claimable over these periods. We did not adjust or drop ETI claims on monthly remuneration which looked too large to be eligible, given the difficulties associated with creating monthly earnings data out of the annual IRP5 data. While the analysis in this paper uses counts of employment at the PAYE entity, results are robust to using full-time equivalents based on employing starting date and ending date variables. We use the unbalanced firm-level panel, which is only balanced after creating stacked events. These notes represent some particularly noteworthy data cleaning and sample construction decisions, but we cannot outline all such decisions here without reproducing our many thousands of lines of code; users are referred to our do-files and R scripts, which are available at the NT-SDF.