EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Elisa Gamberoni, Katerina Gradeva,
Sebastian Weber

Firm responses to employment subsidies: a regression discontinuity approach to the 2012 Spanish labour market reform

No 1970 / October 2016

Abstract

This study focuses on the employment effect of a hiring subsidy available to firms with less than 50 employees, granted in the context of the 2012 Spanish labour market reform. Exploiting the arbitrary firm size threshold using regression discontinuity design, estimates show on average 2 percentage points higher employment growth for firms that became eligible for the scheme. However, tests and complementary regressions suggest that the higher employment growth for smaller firms in 2013 is driven by a 2010 reform, which imposes more stringent reporting requirements on larger firms. Accounting for this using difference-in-discontinuity regressions, we fail to find any significant effect of the subsidy on increasing employment of eligible firms. While our study suggests several pitfalls arising from size-contingent regulations, more data are needed to test for beneficial long-term effects from the hiring subsidy in addressing duality of the Spanish labour market.

JEL Classification: C21, D22, E24, H25

Keywords: Labour market reforms, employment subsidies, firm response, quasi-experiment, regression discontinuity design

Non-Technical Summary

Following the financial crisis, unemployment increased across large parts of Europe, reaching all-time highs in some countries. Constrained by rising debt levels and limited fiscal space, many European governments embarked on ambitious labour market reforms. Recent evidence points to the beneficial effects from a multitude of reform initiatives, including the reduction of entry barriers and similar deregulation in product markets as well as measures to increase wage flexibility, limit unemployment benefits and reduce labour tax wedges (IMF, 2016; Adhikari, Duval, Hu, and Loungani, 2016; Fiori, Nicoletti, Scarpetta, and Schiantarelli, 2012).

Employment subsidies are often part of broader reform packages. This has also been the case for Spain: As the unemployment rate increased above 25 percent, the government implemented a far-reaching labour market reform in 2012, among others, providing firms financial incentives to employ permanently additional personnel. While hiring subsidies are not a novelty, analysing their effects can be a daunting exercise as an evident control group or counter-factual is not always easily identifiable.

This study makes use of the differential treatment of firm groups to identify the impact of the employment subsidy scheme that was introduced in Spain in 2012. Specifically, the subsidy was only granted to firms that have less than 50 employees. The arbitrary size limit provides a natural candidate to analyse firm responses to the subsidy by comparing treated to control firms, with similar characteristics, in the neighbourhood of the 50 employees threshold. By considering only those firms in the close vicinity of the threshold, regression discontinuity methods are able to isolate the effect of being eligible for the new hiring subsidy on firm's employment growth from other factors such as a demand decrease, which affects firms just above and below the threshold likewise. Based on this approach and using the Amadeus firm-level dataset for Spain, we find that employment growth of eligible firms has been about 2 percent higher than employment growth of firms, which could not benefit from the subsidy. These estimation results rely, however, on two crucial assumptions. First, there should be no pre-existing (prior to 2012) firm-size contingent regulations that could affect firms around the 50 employees threshold differently. And second, firms should not have adjusted employment levels to sort below the eligibility threshold of 50 employees (possibly by not prolonging limited-term contracts) in order to benefit from the subsidy, when employing new staff. Based on placebo regressions and test statistics we cannot exclude, however, that these assumptions are violated. To control for the effect of pre-existing firm-size contingent regulations or sorting, several complementary estimation strategies are implemented: difference-in-discontinuities regressions and excluding firms that potentially sort purposefully below the threshold. Both methods make use of the time dimension of our data; in the first case by netting out possible confounding factors and sorting prevailing in 2011, whereas in the second case by excluding firms for which a pattern in employment levels is observed that would be consistent with sorting or employment adjustments in response to pre-existing rules. Irrespective of the applied robustness method, estimated effects decline to a level well below the initial estimate and imply that the difference between employment growth of eligible and control firms is not statistically significant anymore. Thus, we fail to find robust evidence of an effect on the overall employment growth of eligible firms as a result of the Spanish hiring subsidy scheme implemented in 2012.

The findings are consistent with those from other country studies which either find no impact on the employment level of the treatment group or only a short-lasting increase in the employability of an individual. Based on the lack of support for an employment enhancing effect of the subsidy and amidst the associated fiscal costs, caution may be warranted in the specific design of similar hiring subsidy schemes.

As more post-reform data become available, future research could assess possible long-term effects of the subsidy, including the impact on job security and productivity gains that could counterbalance the lack of positive employment dynamics amidst the negative fiscal short-run effects of the subsidy.

1 Introduction

Following the financial crisis, many European governments embarked on ambitious labour market reforms, often in the context of financial support programmes and as a response to fiscal and monetary policy constrains.

Expectations toward the returns from these structural reforms are high. While recent studies find support for the beneficial macroeconomic effects of several reform initiatives (IMF, 2016; Adhikari, Duval, Hu, and Loungani, 2016; Fiori, Nicoletti, Scarpetta, and Schiantarelli, 2012; ECB, 2015), empirical analyses remain relatively scarce.[1] There are multiple reasons for the limited availability of evidence on the effects of such reforms. First, available post-reform data are often insufficient for standard estimation strategies. Second, estimations are hampered by the challenge of identifying the reform effect, because appropriate control groups are not readily available. Furthermore, econometric difficulties arise when identifying the causal impact of structural reforms, because they are not randomly assigned and typically occur in conjunction with interventions in other policy areas. Thus, estimates of the impact of reforms on macroeconomic outcomes might suffer from endogeneity issues.[2]

The present study overcomes these limitations and focuses on a very specific aspect of the Spanish labour market reform introduced in 2012: a subsidy to firms that employ additional workers under a new permanent contract (Contrato de Apoyo a Emprendedores). The new contract entails several fiscal and hiring incentives and is exclusively available to firms with less than 50 employees. The design and implementation of the employment subsidy scheme provide a suitable context for the application of a local randomization inference approach (Calonico, Cattaneo, and Titiunik, 2014a, 2015; Cattaneo, Titiunik, and Vazquez-Bare, 2016b), exploiting the variation in the behavior of firms in the close vicinity of the arbitrary size limit (below and above 50 employees). Similar to other quasi-experimental evaluations, the application of this technique allows us to assess whether outcomes are realised due to the intervention rather than chance, economic environment, or participant selection. The latter aspects are addressed by moving beyond the basic regression discontinuity (RD) framework and employing a difference-in-discontinuities method, which combines the benefits from RD and difference-in-differences (DID) designs (Eggers, Freier, Grembi, and Nannicini, 2016; Grembi, Nannicini, and Troiano, 2016). A simple before-after comparison of beneficiary firms' outcome would be potentially biased due to the influence of the overall economic trend that might have changed while the reform was implemented. A standard DID approach, comparing the difference in outcomes between beneficiaries and non-beneficiaries before and after the reform, provides a possible way to address this concern. However, it includes observations that are far away from the threshold, which would have potentially evolved in a different way in the absence of the treatment. To control for this issue, we compare participants and non-participants in similar economic conditions. Comparing *changes* for the set of firms eligible for the subsidy to *changes* for a suitable control group that is not eligible, we are able to rule out the role of economic conditions and issues due to selection into eligibility.

In our empirical analysis, we fail to provide robust evidence that the subsidy had any significant effect in the short term on overall employment growth of small firms relative to firms that cannot benefit from the scheme. Our results suggest rather that a size-contingent regulation, introduced already in 2010, related to accounting procedures limited the growth for firms with more than 50 employees. First, using

[1] A larger part of the literature focuses on DSGE-based simulations from labour market reforms. These studies often rely on modeling the impact of the reforms via compression of mark-ups and total factor productivity growth effects.

[2] Some authors have attempted to overcome these shortcomings by using propensity score matching techniques (see for instance Bordon, Ebeke, and Shirono (2016)).

local randomization inference, we analyse to which extent employment growth at the firm level is affected by the new subsidy in 2013. Relying on this basic RD design and, thus, ignoring possible confounding factors, such as existing regulations, and sorting ahead of the reform into treatment, suggests that the hiring subsidy has increased employment growth of eligible firms relative to the relevant control group on average by two percentage points in 2013. However, placebo tests, applying the threshold criteria to four pre-reform years, indicate the presence of possible confounding factors or sorting in the pre-reform year 2011. Accounting for this aspect with the use of the difference-in-discontinuities approach and other robustness tests implies that the subsidy had no significant effect on employment growth. The differential growth of firms around the 50 employees threshold appears, instead, to derive from the lower growth of firms above 50 employees after 2011 that wish to circumvent tougher accounting requirements as a result of new regulations passed in the context of the 2010 reform.

In a narrow sense, our study adds to the ongoing assessment on the effect of the 2012 Spanish labour market reform by analysing its impact on employment growth at the firm level. In particular, our analysis complements an OECD study, which employs also a RD design at the aggregate country level based on time discontinuity (OECD, 2013).[3] The effect of the reform is identified through discontinuous patterns occurring at the time of its enforcement (February 2012). As a result, "it is impossible to distinguish its impact from that of other institutional changes occurring around the same date" (OECD, 2013). The study shows an increase in the number of permanent contracts in firms with up to 50 employees following the implementation of the reform. García-Pérez and Jansen (2015), instead, argue that the new contract has failed to achieve its purposes as 93 percent of firms report not to use it in 2013.[4] This percentage needs to be interpreted in the context of the fact that only about 30 percent of firms below 50 employees did at all increase the workforce in this year. Focusing also on the number of contracts, a report issued by the Spanish Ministry of Labour in 2013 claims that the reform limited the extent of job destruction as the number of permanent employment contracts fell only by 1.2 percent in the 12 months following the reform compared with 13.5 percent in the previous 12 months (Spanish Ministry of Labour, 2013). According to the report, the newly introduced permanent contract for small firms is estimated to account for 24 percent of all new permanent full-time contracts.[5] The European Commission in its 2016 country report of Spain concludes that the effects of "subsidies in promoting job creation on permanent contracts remain unclear" (EC, 2016) . Rather than focusing on the number of contracts, our analysis looks at the effect of the reform on employment growth at the firm level. Moreover, instead of considering the whole universe of firms below a certain size threshold, our estimation strategy identifies an optimal window around the threshold, allowing the construction of an appropriate counter-factual.

In a broader sense, our analysis is related to a growing literature that investigates the effects of labour market reforms using firm-level data. This includes Cappellari, Dell'Aringa, and Leonardi (2012) who provide one of the few impact evaluations of a labour market reform at the firm level. The study assesses the effect of new regulations for fixed-term and apprenticeship contracts which were introduced in 2001 in Italy. Employing a DID approach, the authors exploit the variation in the implementation of the new regulations across regions and industries to assess the impact of changes in the employment protection legislation on productivity. They find that easing the use of fixed-term contracts reduced firm-level productivity, whereas the reform on the apprenticeship contracts had a positive effect. Also using Italian

[3] The OECD (2013) study assesses additionally other measures of the reform using varying methods and data sources.

[4] However, the authors observe no increases in dismissals at the end of the probation period of the contract.

[5] The results are based on a comparison between employment demand forecasts and the observed employment for the time period between the second quarter of 2012 to the first quarter of 2013.

firm-level data, Boeri and Garibaldi (2007) find that a liberalisation of the use of temporary contracts has a short-term positive effect on employment growth followed, however, by a decrease in average firm productivity in the medium to long term. Leonardi and Pica (2013) combine the DID and RD designs in their assessment of yet another employment protection legislation reform in Italy on wages using a matched employee-employer sample for the time period 1989-1993. They find a wage reducing effect of the reform, which is however very heterogeneous and inversely related to the bargaining power of workers.

Finally, our analysis complements the impact evaluation literature on the effects of various forms of employment subsidy schemes, which are mostly based on individual-level data as opposed to firm-level data. A notable exception is Hujer, Caliendo, and Radic (2001) who estimate a conditional DID regression using West-German firm-level data. They find no effect of existing wage subsidies on the employment level, citing as a main reason possible substitution from non-subsidised to subsidised employment. Huttunen, Pirttila, and Uusitalo (2013)' assessment of the low-wage subsidy in Finland comes also to the conclusion that there has been no effect on the employment rate of the eligible groups, which is identified using a DID approach based on the eligibility criteria for the relevant workers. Groh, Krishnan, McKenzie, and Vishwanath (2012) analyse the impact of training and wage subsidies programmes on female young employees in Jordan based on a randomized experiment, in which a group of participants was randomly assigned a job voucher to reduce employer costs. The authors find that the job voucher led to a large increase in employment in the short run, however, the impact is no longer statistically significant four months after the voucher period has ended. To a similar conclusion come Galasso, Ravallion, and Salvia (2001) in their study on the effectiveness of a job voucher and training programme in Argentina (Proempleo). Based on a randomized assignment some participants received vouchers entitling their new employer to a sizable wage subsidy. In the short term treated individuals experienced a higher probability to be employed, however, the effect was not sustained. Differently from most of these studies, our focus is on the firm dynamics rather than on the individual effect and the analysis is based on a subsidy scheme that includes a mandatory tenure. However, the effects for aggregate employment are similar although the sustained effect of the subsidy is still to be tested as more post-reform data become available.

In the remainder of the paper, Section 2 provides an overview of the main elements of the 2012 Spanish labour market reform and the relevance of size-contingent regulations in Spain. Section 3 discusses the data. Section 4 describes the empirical framework and relevant test results for the validity of the estimation methods. The results of the estimations are presented in Section 5, followed by a robustness analysis in Section 6. Finally, Section 7 concludes.

2 The 2012 Spanish labour market reform

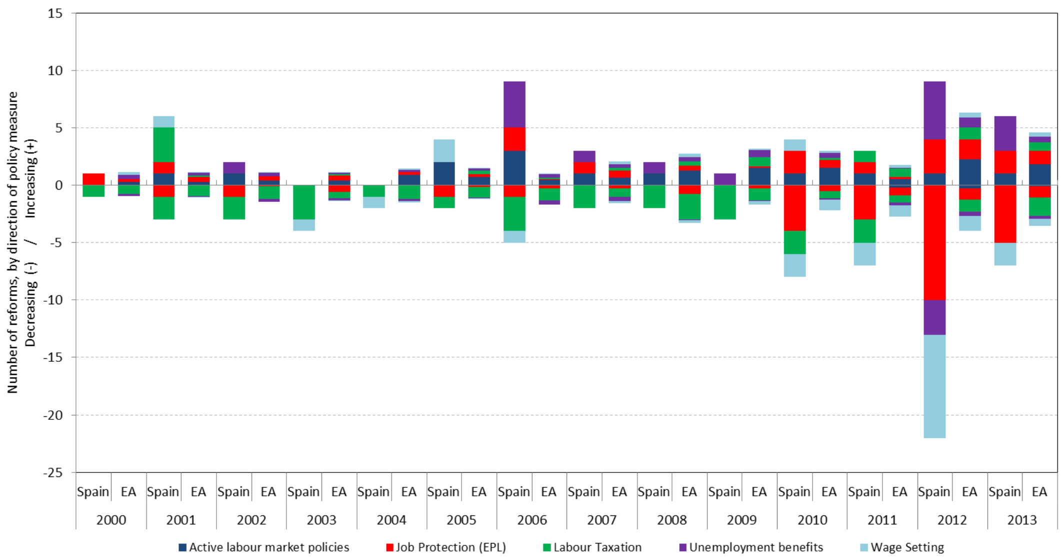
Following the parliamentary elections on November 20, the conservative Popular Party took power in December 2011 ending about seven years of center-left (PSOE) rule. The new government quickly announced that the previous reforms were deemed insufficient and gave two months time to the employers' confederation and the main trade unions to agree on further reform proposals to be introduced by February 2012 (Bentolila, Cahuc, Dolado, and Le Barbanchon, 2012). In the absence of an agreement, the government would legislate unilaterally. The employer and employee organizations met for the first time end-December 2012. While the broad direction of the labour reform was already known earlier on the basis of the election programme of the Popular Party, policies in the programme were relatively vague.[6]

[6]There was no explicit mention of the new permanent contract nor of a subsidy for such a contract.

The 2012 reform was officially unveiled on February 10, 2012 and approved by Royal Decree-Law 3/2012, and subsequently ratified in July.

The 2012 Spanish labour market reform implied a multitude of changes to the Spanish labour legislation. The number of measures taken by the government exceeded not only the reform momentum in recent Spanish history, but also the average reform efforts undertaken by other Euro Area countries (Figure 1).[7]

Figure 1: Spanish and Euro Area reform measures over time



Number of labour market reform measures for Spain and average number for Euro Area countries. Direction of the measures indicates whether the intervention increased (positive) or decreased (negative) the underlying policy settings. Source: LABREF database.

The 2012 reform modified three main aspects of labour regulations aiming at increasing flexibility and decreasing duality of the Spanish labour market. First, it shifted wage bargaining more to the firm level and made it easier for firms to opt-out from a collective agreement. Second, it tightened unemployment benefits, amongst others by reducing the replacement rate after six months of unemployment. Third, it altered employment protection legislation toward a convergence in dismissal costs of permanent and temporary staff.

In addition, a new permanent contract (Contrato de Apoyo a Emprendedores) was introduced.[8] This new indefinite contract for young and previously unemployed workers can be used by companies that have less than 50 employees (prior to making use of the contract). This contract, which can be used for both full-time and part-time employment, entails an extended probation period of one year (with the possibility to end the contract at will during that time) and several financial incentives (applicable until the overall unemployment rate is below 15 percent). Financial incentives include tax breaks and reductions in social security contributions. Specifically, for hiring workers below 30 years of age tax deductions of up to EUR 3,000 are provided after the completion of the probation period. Firms are also granted tax deduction of 50 percent of the unemployment benefits that an unemployed would receive

[7] A caveat applies to this simple measure of reform effort, as it only traces the number of measures as recorded by the European Commission's LABREF database, but cannot capture the intensity and depth of the relevant reform. For a description of previous significant labour market reforms in Spain in 1984, 1994, and 1997 see Ferreir and Serrano (2001) and Dolado, Garcia-Serrano, and Jimeno (2002).

[8] Royal Decree Law 3/2012 of 10 February 2012. See also the LABREF database for a detailed description of the contract.

at the moment she is hired with the new permanent contract.[9] Recruiting firms will also be entitled to additional fiscal incentives in the form of further social contribution reductions: EUR 1,000, 1,100 and 1,200 per year in the first, second, and third year, respectively, for each young unemployed recruited (between 16 and 30 years old), and EUR 1,300 per year for each long-term unemployed over 45 years old over three years.[10] Incentives are conditional on keeping the worker at least three years in the firm (some exceptions are foreseen). Firms cannot have engaged in collective or unfair dismissals in the six months prior to the starting date defined in the new contract in order to be eligible for participation. It should be noted however, that there is no provision in the contract limiting "fair" firings or termination of temporary contracts of other workers already working for the firm. This could potentially imply a switching from temporary to permanent contracts with no net effect on overall employment. Together with the other elements of the 2012 Spanish labour market reform, the establishment of the new permanent contract was approved by the government in February 2012 and confirmed with no substantial modifications in July 2012.

There are three other provisions in the 2012 reform and one related provision implemented in 2013, which have a firm size specific reference.[11] First, firms with fewer than 10 employees can apply for reduced working time collective procedures (García-Pérez and Jansen, 2015). Second, the 2012 reform extended the existing severance-pay subsidy to all cases of fair dismissal, but limited this to firms with less than 25 workers (OECD, 2013). This reduced the ordinary severance cost for employers by 40 percent.[12] However, it had essentially no real effect for firms with less than 25 employees, because a transitory norm contained in the 2010 reform had already extended this subsidy to all firms, including even the case of unfair dismissal (for contracts stipulated after June 2010). Thus, effectively the new rule only put firms above 25 at a disadvantage relative to those below 25 and relative to their own position in 2010/2011. In 2013, the support for firms with fewer than 25 employees was tightened by eliminating the compensation in case of dismissal motivated by economic, technical, organisational or production grounds (Act 22/2013). Third, profitable companies with more than 100 employees incur additional costs if they engage in collective redundancy procedures with workers aged 50 and older. These costs are reflected by contributions to the Treasury to offset the cost caused to the state by the dismissal (unemployment benefits).[13]

The application of different provisions for firms below 10, 25, 50, and 100 employees (see Table 1) imply different incentives to hire (and fire) conditional on firm size. In summary, as a result of the reform, firms with more than 100 employees, under certain conditions, face more costly firing procedures than prior to the the reform. Firms between 50 and 100 employees face comparable labour regulations, with no specific incentive that is peculiar to them. Firms between 25 and 50 employees face comparable labour regulations, but more favourable incentives for hiring workers compared to larger firms, due to the new permanent contract. Firms with less than 25 employees maintain additionally the benefit of lowered severance pay, which was abolished for larger firms as part of the 2012 reform. Finally, firms with less

[9] The unemployed hired with the new contract can choose to receive for one year 25 percent of the unemployment benefits on the top of her salary or keep unused unemployment benefit rights.

[10] Amounts are increased by EUR 100 per year in the case of young female employees and by EUR 200 for female employees over 45 years.

[11] In 2013 there has also been a provision to promote self-employment, which however, is not relevant in the context of this study, because we exclude single-employee firms given their special status.

[12] According to the OECD (2013) this puts the cost of ordinary severance payment in Spain to the level of the OECD average for firms of that size, for which the severance pay applies.

[13] Contributions are reduced when the dismissed workers are relocated within six months after the dismissal into alternative employment, within or outside the company (Royal Decree 1484/2012 of 29 October 2012, based on Law 27/2011 "Actualizacion, adecuacion y modernizacion del sistema de Seguridad Social").

than 10 employees receive the most favourable treatment, because they benefit from additional flexibility in the use of working time arrangements and reduced severance pay compared to all other firms.

Table 1: Summary of 2012 reform measures, by firm size

Measure	Firm size (E_i)				
	$E_i > 100$	$100 \geq E_i \geq 50$	$50 > E_i \geq 25$	$25 > E_i \geq 10$	$10 > E_i$
Higher firing cost in selected cases	✓	✗	✗	✗	✗
Incentives for hiring workers	✗	✗	✓	✓	✓
Reduced severance pay	✗	✗	✗	✓	✓
Use of reduced working time	✗	✗	✗	✗	✓

Summary of the measures introduced by the 2012 labour market reform by eligible firm size.

Based on these reform elements, any control group should be confined to firms with employees between 50-100. Concerning the treatment group the various reform elements require to exclude firms with less than 25 employees.[14] Therefore, we consider only firms with 25 to 75 employees (a bandwidth of 25 around the threshold) in our empirical analysis. This way, control and treatment groups are affected differently by the 2012 labour market reform only due to the differential treatment in terms of the employment subsidy and the window for control and treatment group is symmetric around the threshold.

While to the best of our knowledge there are no other reforms in the years 2012/2013 which affect firms within the neighborhood of 25-75 employees differently, size-contingent legislation is not without precedent in Spain. Two long-standing regulations dating back to 1995, stipulate that firms with more than 50 employees are obliged to establish a 5-member workers' representation committee and appoint two risk prevention delegates for each workers and employers' side (OECD, 2014; EC, 2016). Additionally from 2011 onwards, firms with more than 50 employees are required to design a social plan to facilitate the transition of dismissed employees (Royal decree 801/2011). Next to labour related regulations, accounting regulations can be different for firms that have less than 50 employees. However, employment size is but one possible determinant in this regulation, which was implemented as part of the 2010 reform package.[15] These regulations potentially limit firm growth for firms below 50 employees and at the same time could trigger firms initially larger than 50 employees to sort below the 50 employees threshold, which in principle could play a role for employment growth starting in 2011.[16]

[14]This is also supported by OECD (2013) findings that firms with less than 25 employees experienced a larger increase in new permanent contracts compared to firms with 26-50 employees, suggesting that the severance pay subsidy, which is available only for firms with less than 25 employees, might have favoured the higher increase for this group.

[15]More specifically, the Royal Legislative Decree 1/2010, creates the obligation for companies to submit full balance sheets (as opposed to simplified ones) if at the end of the fiscal year and over two consecutive years, the company fulfills at least two of the following three conditions: Total assets more than EUR 4 million; turnover more than EUR 8 million; average number of workers greater than 50. The same Decree also specifies that an auditor must review the financial statement if at the end of the fiscal year and over two consecutive years, the company fulfills at least two of the following three conditions: Total assets more than EUR 2.85 million; turnover more than EUR 5.7 million; average number of workers greater than 50. In our empirical analysis we will consider the former (higher) thresholds for total assets and turnover in order to control for the overall effect of the Decree on employment growth of firms around the 50 employees threshold (the case when both regulations apply).

[16]The conditions stipulate that two of the three conditions must hold such that firms can also trade-off the options for adjustment over two years, implying possible effects on employment levels in any of the years after 2010.

3 Data

The main data source for the empirical analysis is the Amadeus database provided by Bureau van Dijk.[17] Amadeus is a commercial cross-country firm-level dataset based on the information from firms' balance sheets. Because access to the firm-level data from national statistical agencies and tax offices is often limited due to confidentiality concerns, Amadeus provides an alternative for researchers to explore dynamics at the firm-level.[18] The database covers 934,033 Spanish firms in total for the time period 2005-2014[19] and provides information about general characteristics of the firms such as industry (at the NACE 4-digit level), legal status, year of incorporation, location and ownership as well as about firm size, financial and stock data. The database includes roughly 65 percent of the actual existing number of firms in Spain with one or more employees in 2012.[20]

According to Kalemli-Ozcan, Sorensen, Villegas-Sanchez, Volosovych, and Yesiltas (2015) the Amadeus database covers about 80 percent of output of the economy compared to Eurostat aggregate data and the relative weight of the manufacturing sector in the Amadeus database is broadly comparable to Eurostat in the case of Spain (21 versus 19 percent). Concerning the distribution across firm size, for the total sample there is a bias for an overweight of larger firms in the Amadeus database, compared to Eurostat data. However, this type of comparison is less informative, because many firms in Eurostat have zero employment (self-employed), while this is not the case in the Amadeus database.

The outcome variable in our empirical analysis is employment growth at the firm level, which is defined as the difference in the log-values of average firm size between two consecutive years.[21] Unfortunately, Amadeus does not provide individual firm-level data on the take up of the hiring scheme. Hence, we can only draw conclusions about the impact of the reform based on the eligibility of firms to take up the subsidy. Because of possible imperfect compliance (not all eligible firms take up the hiring incentives), the estimated impact of the reform can be interpreted as an "intent-to-treat" effect (Lee and Lemieux, 2010).

Based on complementary information by the Employment Agency of Spain, Figure 2 indicates that in the time period 2012-2013 around 4 percent of all newly signed permanent contracts by firms with less than 50 employees have been subsidised under the new scheme.[22] Usage of the new contract was particularly strong in the first two quarters of 2012. After 2013, the share of the subsidised contracts decreases to around 2 percent with the last quarters of 2015 showing a slight increase again. Figure 2 demonstrates that at least in the first two years of implementation there has been a pronounced interest by the eligible employers to use the new contract. However, it is uncertain whether the subsidised contracts reflect an effective increase in permanent employment by providing incentives for employers to convert temporary positions into permanent ones.

[17] http://www.bvdinfo.com/en-gb/our-products/company-information/international-products/amadeus.

[18] For a detailed description see Gal (2013) and Force (2014).

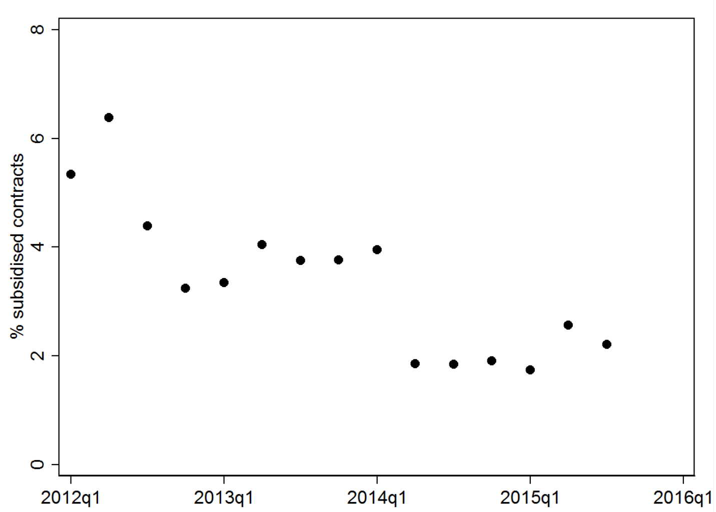
[19] The data have been retrieved on September, 4, 2015.

[20] According to the central business registry of the National Statistical Institute (INE) of Spain, the number of firms in Spain in 2012 was about 3.2 million of which 1.76 million firms without employees (See INE statistics: http://www.ine.es/dynt3/inebase/en/index.htm?padre=51&dh=1).

[21] Ideally, we would also differentiate between the number of permanent and temporary employees per year and firm, because one of the main goals of the reform was to decrease the duality of the Spanish labour market. However, Amadeus does not provide information on the type of employment (permanent vs. temporary) or the hours worked (full-time vs. part-time). Therefore, we use the growth rate of total firm employment in order to proxy for changes in the firm size induced by the reform.

[22] https://www.sepe.es/contenidos/que_es_el_sepe/estadisticas/datos_estadisticos/contratos/datos/estadisticas_nuevas.html.

Figure 2: Percentage of subsidised permanent contracts of all newly signed permanent contracts by small firms



Share of subsidised contracts (Contrato de Apoyo a Emprendedores) of all newly signed permanent contracts by firms with up to 50 employees, by quarter. Source: Employment Agency of Spain (SEPE).

Despite the wide coverage of the Amadeus database, we face several constraints.[23] First, Table 2 shows that for roughly half of the sample, information on employment growth (data on number of employees in two consecutive years) is missing. Second, post-reform data are not fully available. Data for the year 2014 are still incomplete at the date of retrieval. Therefore, we exclude 2014 from our empirical analysis and focus only on 2013 as full post-reform year. We look in particular at the results for 2013, because the reform was not in force throughout the entire year of 2012 and employment levels are based on the year-average. Thus, the outcomes of our empirical analysis provide evidence on the short-term impact of the reform on employment growth of eligible firms.

Table 2: Number of available observations, by year

Year	Total number of firms	Firms with available employment growth data
2009	825992	466600
2010	859482	474445
2011	890917	472695
2012	918476	468872
2013	933667	433721
2014	934032	44687

Total number of firms per year available in the Amadeus dataset and number of firms with data on employment growth (number of employees available for two consecutive years).

[23] Amadeus only contains information on surviving firms which in principle could lead to a selection bias as we cannot take into account the firms that went bankrupt. However, the outcome variable for the purpose of our study is employment growth at the firm level. Thus, by definition we can only consider surviving firms to estimate the effects of the reform.

4 Estimation strategy

The core issue for empirical research on policy reforms is to assess whether or not outcomes are realised due to the intervention or chance, economic conditions, or participant selection. If we could compare the outcome for a unit with and without treatment at a given point in time, we could rule out potential endogeneity issues resulting from a selection bias or changing economic environment. Because this is impossible, we have to rely on a counter-factual using a suitable comparison group that has *not* benefited from the intervention. If we simply compare the treated firms before and after the reform, we cannot rule out the influence of economic conditions that may have changed while the reform was implemented. If we compare participants and non-participants in the post-treatment period given the same business environment, we face a selection bias: firms who benefit from the reform may be in general different from those which are excluded. Thus, the main challenge of conducting an impact evaluation study is to accurately identify the most relevant comparison for units eligible for the programme. Among the various quasi-experimental approaches, the design of the reform allows us to apply a RD estimation. Compared to other evaluation methods, RD designs resemble more closely randomized experiments (within a small window around the threshold) and provide a straightforward strategy to identify the casual effect of a reform (Lee and Lemieux, 2010).[24]

Two key assumptions are necessary in order to identify the average treatment effect of an intervention: overlap and unconfoundedness (also referred to as the conditional independence assumption, selection on observables or exogeneity). The overlap assumption requires that both treated and control units are observed for all values of the covariates (Imbens, 2004; Imbens and Lemieux, 2008). This assumption is violated in a RD framework, because it is not possible to observe treated and non-treated individuals for a given value of the assignment variable. In order to compensate for this assumption, continuity of the outcome variable's distribution is required (Lee and Lemieux, 2010). Continuity implies that the conditional distribution function of the outcome is smooth in the treatment-determining covariate so that any discontinuity at the threshold can be attributed to the intervention. Therefore, the average treatment effect in a RD framework is estimated by the difference in the conditional expectation of the outcome variable arbitrarily close below and above the threshold for units with covariate values, which are comparable. Intuitively, by choosing units close to the threshold, we maximize the probability that all the other covariates are "evolving smoothly with respect to the assignment variable" (Lee and Lemieux, 2010). The unconfoundedness assumption asserts that conditional on observed characteristics, the treatment indicator is independent of the error term in the regression, meaning that the treatment status is exogenous (all relevant characteristics are controlled for excluding the possibility of omitted variable bias). Hence, any endogeneity can be ruled out and the estimated treatment effect has a causal interpretation (Imbens, 2004; Imbens and Lemieux, 2008).

The following sub-sections introduce the two RD estimation approaches that we apply in our empirical analysis:[25] the local randomization inference method, which is based on comparing firms closely below and

[24]For example, compared to propensity score matching, RD estimations do not require the inclusion of all observable covariates in order to ensure an overlap between the treated and the control group (Lee and Lemieux, 2010). In contrast to a DID approach, the RD method excludes observations that are far away from the threshold and which would have potentially evolved in a different way in the absence of the treatment (non-testable parallel trend assumption in a DID framework).

[25]The OECD (2013) study largely relies on time-based RD approaches to identify the effect of the reform focusing on the discontinuity in time in various outcome variables, which was induced by the introduction of the reform in February 2012. In the authors' view "the comprehensive nature of the 2012 reform makes its evaluation a difficult task [and] the inclusion of a large number of provisions, sometimes explicitly targeted at different groups, does not allow the identification of a suitable control group." We argue that it is exactly the differential treatment of firm groups (based on their size) which

above the threshold of 50 employees in a post-reform environment[26] and the difference-in-discontinuities method, which exploits additionally the time dimension of our data. The latter allows controlling for possible confounding factors (e.g. due to the presence of other regulations) or sorting of firms. Finally, the validity of the assumptions underlying the methods is tested.

4.1 Local randomization

The recently introduced local randomization inference method, following Cattaneo, Frandsen, and Titiunik (2015) and Cattaneo, Titiunik, and Vazquez-Bare (2016a), provides a suitable approach for a RD estimation in the case of a discrete treatment-assigning variable.[27] Two main features need to be fulfilled for the validity of the method. First, a window around the threshold should exist where treatment assignment is known as in a randomized experiment. Second, the outcome variable can be transformed in a way which eliminates its direct dependence on the assignment variable (exclusion restriction). The new method builds upon the idea of Lee (2008) by considering the treatment status around the threshold in a RD framework "as if random". Thus, in some small neighborhood around the threshold, treated and control units possess the same distribution of baseline covariates as in a randomized experiment.

The crucial step in the estimation procedure is to select an optimal window of observations around the threshold depending on the assumed functional form of the relationship between the outcome and treatment-assigning variable.[28] The aim is to choose a window such that covariates between treated and control units should not be significantly different from each other. The idea is to start from a small window around the threshold and to increase it to the largest possible window such that the minimum p-value from all covariate tests is equal or higher to a predetermined significance level.[29] For a conservative choice of the selected window, Cattaneo, Frandsen, and Titiunik (2015) suggest choosing a significance level of 0.15, higher than conventional levels. The choice of the functional form is the other essential step in the estimation of the treatment effect. In the case of a discrete assignment variable, as in our study, the causal effect of the intervention cannot be identified without assuming a parametric form of the function, because the gap between the control ("just above") the threshold and the treatment ("just below") observations is irreducible, meaning that we cannot choose units which are closer to the threshold than a window of [-1;1] (Lee and Card, 2008).

Having determined the optimal window around the threshold for the chosen functional form, the difference in means estimator provides an unbiased estimate of the average treatment effect for the relevant interval (Cattaneo, Titiunik, and Vazquez-Bare, 2016b). The main specification, estimated for the post-reform year can be described as follows:

$$Y_i = \mu + \alpha \cdot \delta_i + \beta \cdot f(E_i) + \gamma \cdot \delta_i \cdot f(E_i) + \epsilon_i \qquad (1)$$

allows for the identification of the impact of different components of the reform.

[26] In a robustness section we also employ flexible parametric estimations.

[27] Conventional methods for the empirical application of a RD estimation include local non-parametric polynomial techniques, which rely on the assumption that the assignment variable is continuously distributed and compare outcomes "just below" and "just above" the threshold (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). This assumption is, however, violated in our case because the assignment variable is discrete (number of employees), meaning that we cannot employ local non-parametric techniques.

[28] In general, higher-order polynomials tend to "overfit" the data if the chosen bandwidths are small (Lee and Lemieux, 2010).

[29] When choosing the optimal bandwidth, researchers face the trade-off between bias and precision: Smaller bandwidths reduce potential bias at the expense of increasing variance as the number of observations decline (Calonico, Cattaneo, and Titiunik, 2014b; Imbens and Kalyanaraman, 2012; Ludwig and Miller, 2007).

The treatment indicator δ_i is defined as:

$$\delta_i = \begin{cases} 1 \text{ if } E_i < E^T \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

where the threshold level is given by $E^T = 50$ and the assignment variable, the employment level in the previous year of firm i (E_i), is centered at the threshold of 50 employees. This simplifies the interpretation of the results, because the intercept (α) captures the shift in the outcome variable at the threshold. The estimation sample is restricted to observations with an employment level in the previous year within the interval $E_i \in (E^T - b, E^T + b)$, where b is the optimal bandwidth according to the covariate tests. As discussed in the previous section we focus on average employment growth in 2013 as the first full post-reform year and differentiate between firms that had average employment levels of below 50 and above 50 employees in 2012. While in later sections we also provide some results when additionally classifying 2012 as a post-reform year, this may create some downward bias as the contract was only available after March 2012 and firms report the average employment level in a given year. The dependent variable, Y_i, is given by the average employment growth of firm i. The function of the assignment variable ($f(E_i)$) in Equation 1 controls for the relationship between the assignment and the dependent variable allowing for different slopes on either side of the threshold reflected by the inclusion of the interaction term $\delta_i \cdot f(E_i)$. We estimate Equation 1 including different order polynomials (from zero- to second-order) for each respective optimal window around the threshold computed by the covariate tests. The confidence interval for the estimated treatment effect is corrected for finite-sample inference.

4.2 Difference-in-discontinuities

The local randomization inference relies on the assumption of unconfoundedness. However, this condition might be violated due to sorting around the threshold or confounding factors such as other legislation from previous years that affect the estimates. If this is the case, making use of the time dimension of our dataset would allow nevertheless for an assessment of the 2012 subsidy. The difference-in-discontinuities design, recently proposed by Grembi, Nannicini, and Troiano (2016) and Eggers, Freier, Grembi, and Nannicini (2016), provides a tool for netting out a potential bias in a RD estimation caused by such factors. In order to be applicable, we need to have available two sets of observations: one observation in which the policy of interest is in place together with the confounding factor and another observation when only the confounding factor plays a role and the policy of interest is not in place (Eggers, Freier, Grembi, and Nannicini, 2016; Grembi, Nannicini, and Troiano, 2016). Furthermore, the estimate relies on two assumptions: the existence of local parallel trends and separability. The first assumption requires that the effect of the confounding policy (e.g. a previously existing reform) is constant over time. This assumption is analogous to the parallel trend assumption in the general DID framework, but limited to a more narrow range around the threshold. The plausibility of the assumption may be tested by extending the cross-sectional test of the continuity of the density at the threshold (used in the local randomization inference context) to test for the continuity of the *difference* in the densities before and after the policy of interest is in place. If the first assumption holds the estimates are unbiased (Eggers, Freier, Grembi, and Nannicini, 2016). The assumption of separability requires that there is no interaction between the treatment and the confounding policy discontinuity and it is similar in spirit to the additivity condition in the DID set-up.[30] It ensures that the estimate is equivalent to the average treatment effect. Intuitively, the difference-in-discontinuities estimator can be understood as a simple two-period extension of the local RD estimator and is described by:[31]

$$\begin{aligned} Y_{it} = \quad & \mu + \alpha_1 \cdot \delta_i + \beta_1 \cdot f(E_{it-1}) + \gamma_1 \cdot \delta_i \cdot f(E_{it-1}) + \\ & T_t \cdot [\alpha_2 \cdot \delta_i + \beta_2 \cdot f(E_{it-1}) + \gamma_2 \cdot \delta_i \cdot f(E_{it-1})] + \epsilon_{it} \end{aligned} \tag{3}$$

for $t \in (t_0, t_k)$ and where T_t is an indicator variable for the post-2012 reform period. The effect of the reform in this case is measured by the the coefficient α_2 as the treatment is captured by $T_t \cdot \delta_i$. Standard errors are clustered at both the firm and the number of employees level.

A priori, it is not possible to exclude the issue of confoundedness when applying a conventional RD estimation. Apart from legislation already in place for several years (see Section 2), two additional reforms started to play a role in 2011. From 2011 onwards, firms with more than 50 employees are required to design a social plan to facilitate the transition of dismissed employees. Potentially more relevant for our estimation, exceeding the threshold of a firm size of 50 employees became one of three possible factors to determine after two fiscal years whether a firm will be subject to more stringent accounting and reporting rules. To our knowledge, there is no other regulation after 2011 aside from the introduction of the subsidy in 2012, which is also governed by the threshold of 50 employees (or a threshold between 25 and 75 employees). Thus, we cannot exclude the possibility of a bias in the estimation of Equation 1 due to confoundedness from these previous reforms. Similarly, sorting cannot be excluded ex-ante. While the 2012 reform was implemented relatively quickly and the law foresees that firms, which fired staff unfairly in the previous months cannot benefit from the subsidy scheme, it is still possible for firms to adjust their level of employment in anticipation of the law.[32] In the absence of confounding factors or sorting, the difference-in-discontinuities estimator should provide comparable results to the local randomization inference approach (Grembi, Nannicini, and Troiano, 2016).

4.3 Validity of the RD design

We start the analysis by testing the validity of the general assumptions underlying the RD framework. First, we assess whether there exists a small enough window around the threshold such that the covariates of the treated and control firms are balanced, a necessary condition to fulfill the overlap assumption underlying the local randomization inference approach. Second, we analyse the distribution of firm size measured by the number of employees in the time period before 2012, to test for potential confounded treatments and evidence of sorting around the threshold before the introduction of the scheme. Third, we test for constancy in this density over time, which would support the validity of the parallel trends assumption of the difference-in-discontinuity estimator. Finally, we provide graphical evidence of the main outcome variable, employment growth, as a first assessment on whether it is smoothly distributed in the treatment-assigning variable (number of employees) before 2012, and whether it shows a substantial

[30] We provide no formal test for this assumption, given the specific reform context. However, the data allow us to disentangle the effect of the 2012 reform from confounding factors or sorting by focusing on firms which are unaffected by the 2010 reform or by excluding firms that potentially sorted below the 50 employees threshold to benefit from the subsidy. The results are discussed in the robustness section.

[31] Leonardi and Pica (2013) apply a similar approach, combining the features of the DID and RD methods. However, we follow Grembi, Nannicini, and Troiano (2016) because their method allows a more flexible estimation.

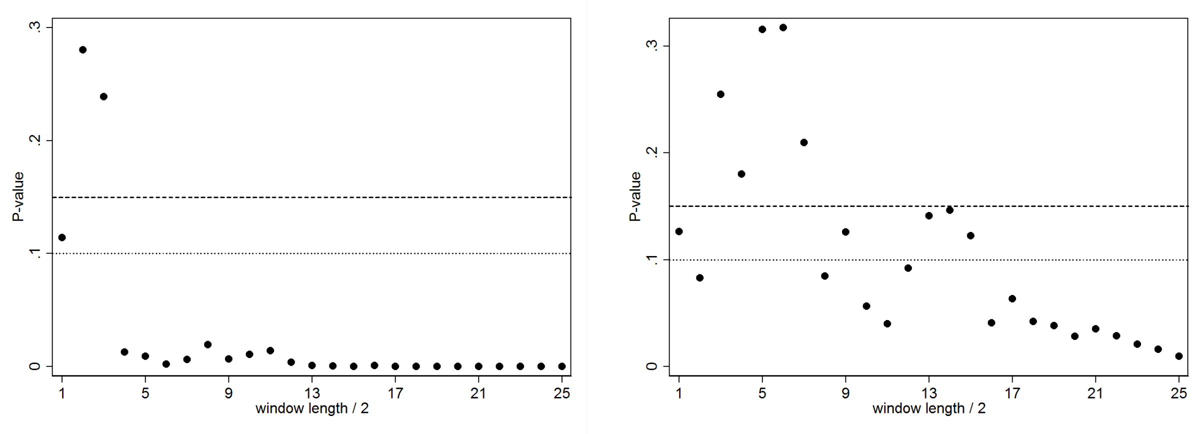
[32] Based on data provided by the Employment Agency of Spain (SEPE), firms in Spain employ about 30 percent of their workforce under temporary contracts on average. Because these contracts have on average a duration of about 60 days, a firm with employees in the range of 50 to 75 could in principle let all its temporary contracts run out within a few months, reducing the number of employees to below 50.

discontinuity at the threshold after the introduction of the reform.

Figure 3: Minimum p-values from covariate tests, 2013

(a) no model transformation

(b) first-order polynomial



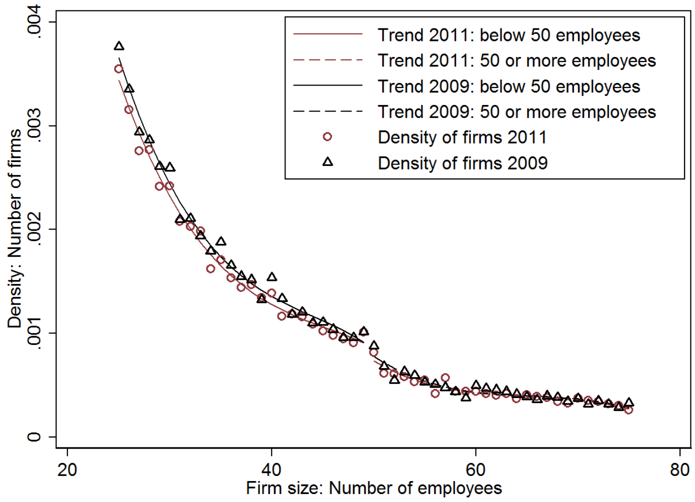
Minimum p-values from covariate tests. Covariate tests based on five firm characteristics: industry, region, age, legal form and capital per employee ratio. The first-order polynomial is fitted separately for treatment (below 50 employees) and control (above 50 employees) firms (maximum window length of 25).

The window choice around the threshold is a key step of the RD estimation. We implement a data-driven selection procedure by running covariate tests in order to identify a window around the threshold of 50 employees such that the treatment assignment can be regarded "as if random" (Cattaneo, Frandsen, and Titiunik, 2015; Cattaneo, Titiunik, and Vazquez-Bare, 2016a,b). The covariate tests are based on five variables and namely the age of the firm, the geographical location, the industry, the legal form and the capital per employee. The optimal windows are robust to changes in the selected covariates.

Because the idea of the local randomization inference in a RD design is to choose a small window around the threshold such that characteristics of treated and control firms are balanced, specifications including higher-order polynomials might "overfit" the data (Cattaneo, Titiunik, and Vazquez-Bare, 2016b). Therefore, we run the covariate tests either without any transformation in the outcome variable or including a first-order polynomial, which is allowed to vary on both sides of the threshold. We start from the smallest possible window [-1;1] comparing firms with 49 and 51 employees, to the largest possible one in our sample [-25;25]. As expected, not taking into account any polynomial leads to choosing a smaller window compared to including a first-order polynomial. We consider two significance levels to identify the optimal window length; the conventional level of 0.10 and the more conservative one of 0.15. Figure 3a shows that the optimal window for 2013 without any transformation in the outcome variable is the same, [-3;3], for the two significance levels (0.15 and 0.10). Controlling for a first-order polynomial, the optimal window is either [-7;7] at the more conservative significance level of 0.15 or [-15;15] for the conventional significance level of 0.10 (Figure 3b). In the empirical analysis we provide estimates for all optimal windows around the threshold as a way to validate the robustness of our findings.

As outlined in Section 4.1, another requirement for the application of a RD estimation in our setup is that firms have not self-selected into being in the treatment group. Figure 4 shows a distribution of the number of firms smoothly decreasing in the number of employees for companies with a number of employees between 25 and 75 in the years 2009 and 2011. The plot presents two third-order polynomials which are fitted on both sides of the threshold for 2009 and 2011, respectively. As revealed by Figure

Figure 4: Distribution of firms over the number of employees in 2009 and 2011



Distribution of number of firms over the number of employees for 2009 and 2011. Two third-order polynomials fitted for treatment (below 50 employees) and control (above 50 employees) firms and for each year respectively.

4, there is no substantial difference in the distribution of firms shortly before the introduction of the reform compared to previous years. We formalise the analysis by fitting a regression of polynomial trends of the number of employees on the number of firms per employee-level, allowing the trend to differ on the left and right side of the threshold.[33] Results are reported in Table A.1. When including a third-order polynomial trend, we find no significant break in the distribution in the pre-reform years, with the exception of 2011 for a bandwidth of 15 employees around the threshold. When limiting the polynomial trend to the second-order, which fits the data less well, we find support for a significant jump at the threshold for all pre-reform years. This might be explained by the fact that firms with less than 50 employees are granted also other exemptions or benefit from special regulations, as discussed in Section 2.[34]

Importantly, we find no significant *change* in the estimated "jump" at the 50 employees threshold when comparing 2011 to any of the years before. This holds irrespective of whether a second or third order polynomial trend is employed. Furthermore, when comparing distribution densities in 2013 and 2011 (see Figure A.1 in the Appendix), there is also no evidence of a change in a possible discontinuity at the threshold following the 2012 reform. Based on this evidence, it appears that the local parallel trend assumption underlying the difference-in-discontinuities estimator is supported.

Finally, we provide a graphical visualisation of our main variable of interest, employment growth, as a function of the firm size. Using local sample means over non-overlapping bins for all firms with number of employees between 25 and 75, Figures 5a - 5e implement the mimicking variance evenly spaced method proposed by Calonico, Cattaneo, and Titiunik (2015). The method captures the variability of the raw data and estimates two smooth second-order polynomial regressions, for treatment and control firms separately.[35] All graphs reveal a negative employment growth for both treatment and control

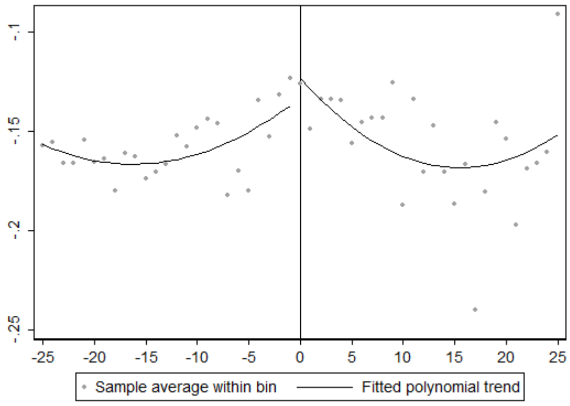
[33]The analysis follows Lee and Card (2008), because the procedure proposed by McCrary (2008) is not applicable for discrete variables.

[34]The annual report by Banco de España (2015) shows that the discontinuity at the 50 employees threshold has existed already in the time period 1995-2007.

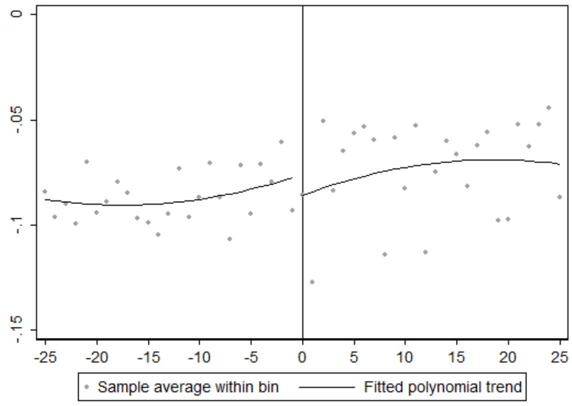
[35]Cattaneo, Titiunik, and Vazquez-Bare (2016a) argue that higher-order approximations are better suited to provide an

Figure 5: Employment growth in various years, by treatment status

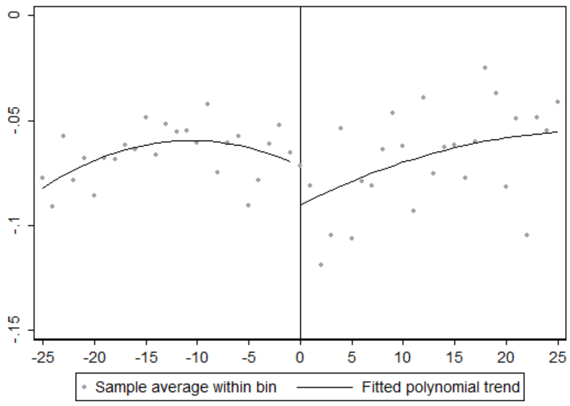
(a) 2009



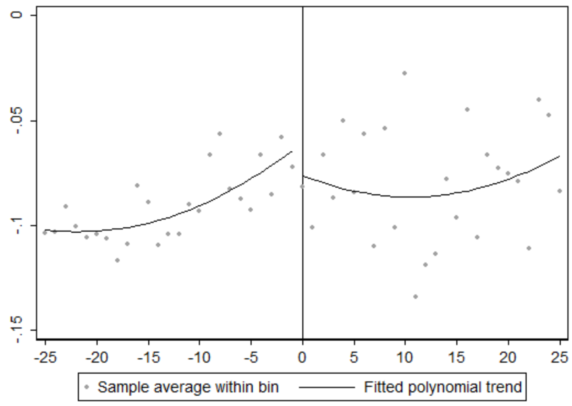
(b) 2010



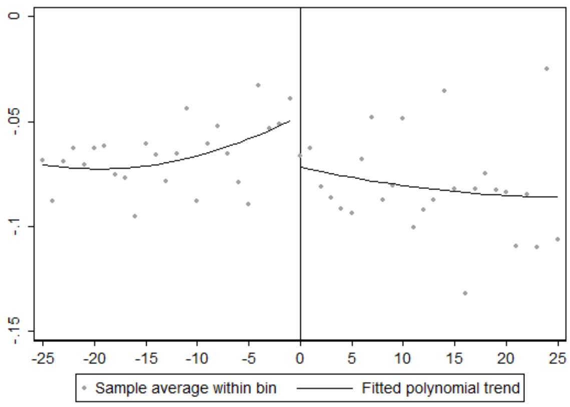
(c) 2011



(d) 2012



(e) 2013



Local sample means over non-overlapping bins for firms with a number of employees between 25 and 75. Second-order polynomials fitted separately for treatment (less than 50 employees) and control (more than 50 employees) firms. x-axis centered at the 50 employees threshold (50=0).

groups. In 2009 and 2010, firms with a few more employees than 50 registered no particular different employment growth than the firms just below the threshold. Starting in 2011, a more noticeable gap starts to open up, with employment growth falling less for smaller firms compared to larger firms above 50 employees. This is, in principle, consistent with sorting of firms below the 50 employees threshold due to both, anticipation effects related to the 2012 subsidy for smaller firms and reaction to the 2010 reform, according to which a firm's size of more than 50 employees is one of the possible determinants for tighter auditing and reporting requirements. By 2013 (Figure 5e) a substantial discontinuity emerges with control firms (above 50 employees) growing at a lower pace than treated ones (below 50 employees). The plots illustrate the dispersion of the data and offer evidence for a discontinuity in the distribution of employment growth at the threshold of 50 employees after the introduction of the reform, but also in the pre-reform year 2011. Taken together, these preliminary tests suggest that despite the reform's speedy implementation, possible sorting and in particular confoundedness cannot be excluded.

5 Results

The following section formally assesses the descriptive evidence suggested by Figure 5e. We estimate the difference in employment growth for firms eligible for the subsidy scheme relative to employment growth for firms in the relevant control group using local randomization, followed by placebo estimation analysis, and difference-in-discontinuities methods to adjust for possible confoundedness.

5.1 Local randomization

Table 3 shows the results from the local randomization inference approach as described in Equation 1 focusing on 2013 for different window lengths and order of polynomials.[36] The bold cells capture the outcome for the optimal window around the threshold based on the minimum p-values of the covariate tests given the selected order of the polynomial, zero- or first-order, and the significance level, 0.15 or 0.10 (see Figures 3a, 3b). As discussed in Section 4.3, assuming a zero-order polynomial leads always to the same optimal window length, [-3;3], independently of the selected significance level for the covariate tests. However, the choice of the significance level matters when controlling for a first-order polynomial in terms of the optimal bandwidth around the threshold. Therefore, we present results for the two significance levels.

As Table 3 reveals, the local randomization approach suggests that being eligible for the new permanent contract increases significantly employment growth in 2013. Focusing on the estimates based on the optimal window lengths shows that the overall impact of eligibility for the hiring subsidies is associated with a roughly 2.3-4 percentage points higher employment growth rate of treated firms compared to the control group in 2013. Based on the smallest window and, thus the most conservative result, we estimate that firms just below the threshold of 50 employees have increased on average their employment growth by 2.5 percentage points compared to the firms just above the threshold. Controlling for a first-order polynomial, results for the larger bandwidth [-7;7] suggest a higher growth in employment of 4 and for the largest bandwidth [-15;15] higher growth of 2.3 percentage points. Controlling for a second-order instead of first-order polynomial leads to very similar findings.[37] Comparing our results to the ones of the OECD

overall representation of the RD design and falsification testing for other discontinuities.

[36] Results for 2012 are comparable, although imply a somewhat lower impact, and are available upon request.

[37] Including covariates directly in the regression has negligible effects on the coefficient estimates. This is to be expected

Table 3: Local randomization: Employment growth

	(1)	(2)	(3)
	Employment growth 2013		
	[-3;3]	[-7;7]	[-15;15]
Polynomial (0)	**0.025***	0.016	0.013**
	(0.086)	(0.131)	(0.056)
Polynomial (1)		**0.04*****	**0.023*****
		(0.000)	(0.000)
Polynomial (2)		0.03***	0.032***
		(0.001)	(0.000)
Sample size treated	1,313	3,167	8,360
Sample size control	1,169	2,063	3,700
Observations	2,482	5,230	12,060
Significance level	0.15/0.10	0.15	0.10

Each coefficient is the result of a separate regression and captures the total treatment effect (including the higher-order polynomials). Bold coefficients present the results for the optimal window based on the choice of the polynomial and the significance level of the covariate tests.
Note: p-values, corrected for a finite sample, in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

(2013), shows a similar pattern in terms of a positive employment growth effect of the new permanent contract. The OECD (2013) concludes that at least 25,000 new permanent contracts per month in firms with 50 or less employees can be attributed to the 2012 reform, while no significant effect is observable for larger firms.[38]

5.2 Placebo estimations

Motivated by the preliminary tests on the validity of the basic RD design and to scrutinise the results from the local randomization inference estimation, we run several placebo estimations. The first set of placebo estimations pretend that the intervention has taken place before 2012, in order to exclude the possibility that the estimated treatment effect reflects a pre-existing difference in employment growth between firms just below and above the threshold and the potential role of treatment confoundedness. We estimate the placebo effect for each year of the time period 2008-2011. Table 4 depicts the results for the same window lengths around the threshold as in the estimations for 2013. Table 4 shows that there is no robust and significant difference in employment growth between firms closely below and above the threshold of 50 employees in the years 2008-2010. In 2008 the sign of the coefficient depends on the functional form chosen, in 2009 and 2010 almost none of the coefficients are significant. However, in 2011 there is strong evidence that employment growth of smaller firms is higher than employment

if there is no discontinuity in the covariates around the threshold, which otherwise would indicate support for sorting into treatment. Results are not reported to save space and available on request.

[38]The results are not strictly comparable to the estimates presented here, because they reflect the absolute number of contracts, which will naturally be determined by the firm size (the larger the firm the higher the number of contracts for a given percentage increase in employment) and the number of firms for a given size.

growth of larger firms. The results are to some extent foreshadowed by Figure 5c, which shows a strong decline in employment growth in 2011 for firms that in 2010 were just above the 50 employees threshold. This pattern would be consistent with larger firms sorting just below the threshold in 2011 in order to become eligible for the new contract to be introduced in 2012 or to avoid the more stringent reporting requirements as a result of the reform implemented in September 2010. This could have been made possible by not prolonging temporary contracts.

The second set of placebo estimations considers a different firm size threshold, which defines the eligible firms for treatment. Columns 5-7 in Table 4 present the results pretending that the threshold for benefiting from the new contract has been set at the 45, 55 or 60 number of employees-level (instead of 50). The few coefficients that are significant point rather in the opposite direction of the results in Table 3, however, their sign depends on the chosen functional form and bandwidth.

Table 4: Placebo tests: different years and firm size thresholds

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Employment growth						
	Placebo years				*Placebo firm sizes*		
	2008	2009	2010	2011	45	55	60
Poly.(0), [-3;3]	-0.002	-0.001	0.011	0.031**	-0.008	-0.010	0.008
	(0.898)	(0.944)	(0.473)	(0.019)	(0.497)	(0.575)	(0.676)
Poly.(0), [-7;7]	-0.006	-0.015	- 0.007	0.020**	-0.007	0.013	0.001
	(0.538)	(0.112)	(0.458)	(0.044)	(0.418)	(0.264)	(0.933)
Poly.(0), [-15;15]	- 0.001	-0.010	-0.010	0.016**	-0.004	0.014*	0.018*
	(0.932)	(0.173)	(0.124)	(0.014)	(0.473)	(0.088)	(0.061)
Poly.(1), [-7;7]	-0.020**	0.019*	0.026***	0.022**	-0.016**	-0.028**	-0.001
	(0.036)	(0.057)	(0.008)	(0.024)	(0.050)	(0.025)	(0.925)
Poly.(1), [-15;15]	-0.010	0.009	0.009	0.020***	-0.010*	0.010	-0.007
	(0.156)	(0.205)	(0.166)	(0.003)	(0.088)	(0.237)	(0.414)
Poly.(2), [-7;7]	0.022**	0.020**	0.000	0.033***	-0.004	-0.020*	-0.019
	(0.023)	(0.041)	(0.981)	(0.001)	(0.633)	(0.096)	(0.178)
Poly.(2), [-15;15]	-0.023***	0.002	0.011	0.021***	-0.008	-0.024**	-0.021**
	(0.000)	(0.784)	(0.114)	(0.001)	(0.214)	(0.004)	(0.019)
Observations [-3;3]	3,091	3,220	2,949	2,850	3,292	1,654	1,444
Observations [-7;7]	6,670	6,930	6,336	6,072	7,097	3,992	3,126
Observations [-15;15]	14,680	15,599	14,368	13,899	15,645	9,429	7,539

Results are obtained using a local randomization inference approach. Each coefficient is the result of a separate regression and captures the total treatment effect (including the higher-order polynomials). Chosen bandwidths based on the covariate tests for the 2013.
Note: p-values, corrected for a finite sample, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.3 Difference-in-discontinuities

The results from the placebo estimates for 2011 in Section 5.2 underpin the suspicion from the preliminary tests in Section 4.3 that sorting and confoundedness cannot be excluded. Thus, the local randomization inference results are unlikely to provide an unbiased estimate of the true effect of the reform. We, therefore, proceed with the estimation of the difference-in-discontinuity design as described in Section

4.2. In line with the timing of the reform, the placebo estimates and test results, confounding factors and pre-reform sorting could start playing a role in 2011. The only difference between the years 2013 and 2011 in terms of differential treatment of firms just above and just below the 50 employees threshold is the introduction of the subsidised permanent contract in 2012. To net out the possible confounding effect of the 2010 reform and possible sorting, we estimate the effect of the subsidy by Equation 3 using $t_k = 2013$, $t_0 = 2011$ and the indicator variable $T_t = 1$ for 2013. Additionally, we provide estimates exploring the full time dimension including the two years prior to the implementation of the reform (2010 and 2011) as well as the reform year 2012 and 2013 as post-reform years. In this case, the indicator variable $T_t = 1$ for 2012 and 2013 and Equation 3 includes time dummies.

Table 5: Difference-in-discontinuities: Employment growth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Employment growth 2011 vs. 2013						
	[-3;3]	[-7;7]	[-15;15]	[-7;7]	[-15;15]	[-7;7]	[-15;15]
δ	0.031***	0.020**	0.016***	0.022*	0.020	0.033*	0.021
	(0.011)	(0.009)	(0.006)	(0.013)	(0.012)	(0.017)	(0.014)
$\delta \cdot 2013$	-0.006	-0.004	-0.003	0.019	0.004	-0.003	0.011
	(0.009)	(0.011)	(0.008)	(0.013)	(0.013)	(0.024)	(0.015)
Polynomial	0	0	0	1	1	2	2
Observations	5,332	11,302	25,959	11,302	25,959	11,302	25,959
R-squared	0.002	0.001	0.000	0.001	0.001	0.002	0.001
	Employment growth before (2010-2011) vs. after (2012-2013)						
	[-3;3]	[-7;7]	[-15;15]	[-7;7]	[-15;15]	[-7;7]	[-15;15]
δ	0.021**	0.006	0.003	0.023*	0.014*	0.016	0.016
	(0.008)	(0.007)	(0.004)	(0.013)	(0.008)	(0.018)	(0.012)
$\delta \cdot post$	-0.003	0.003	0.003	0.004	0.002	0.015	0.007
	(0.008)	(0.007)	(0.007)	(0.011)	(0.008)	(0.014)	(0.012)
Polynomial	0	0	0	1	1	2	2
Observations	11,051	23,592	53,608	23,592	53,608	23,592	53,608
R-squared	0.001	0.000	0.001	0.001	0.001	0.001	0.001

Results are obtained using a local parametric method. Each of the regressions includes a constant, time fixed effects and the reported order of polynomial, which is allowed to differ on both sides of the threshold as well as between the two time periods (2011-2013 or before-after). δ captures the treatment effect at the threshold, while $\delta \cdot 2013$ and $\delta \cdot post$ reflect the treatment effect of the subsidy. Chosen bandwidths based on the covariate tests for 2013. Robust standard errors, clustered both at the firm level and at the number of employees, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Coefficient estimates for the difference-in-discontinuities estimates using only 2011 and 2013 confirm (though not for all bandwidth-polynomial constellations) that starting in 2011 employment growth of smaller firms has been higher than of firms with more than 50 employees (δ estimate reported in the first panel of Table 5). The coefficient on the indicator variable reflecting the effect of the subsidy ($\delta \cdot 2013$) on employment growth is now closer to zero and never significant. Using the time period from 2010 to 2013 and a post-reform indicator for 2012 and 2013 confirms this pattern, yielding no single significant estimate for the effect of the reform on employment growth ($\delta \cdot post$ estimate reported in the second panel of Table 5). Using this sample, coefficient estimates for the difference in the pre-reform period (2010-2011) for employment growth of small firms relative to larger firms (δ) are in general lower and less significant, pointing to the relevance of 2011 as the year which marks the start of the differential growth trend, as indicated also by the placebo regressions.

Thus, the difference-in-discontinuities regression results suggest that the initial findings from the local randomization inference estimation showing a higher employment growth for firms that are eligible for the subsidy are likely to be driven by confounding factors and possible sorting that pre-date the reform.

6 Robustness

The following section provides additional robustness estimates that aim at mitigating the possible effect from sorting and confoundedness using alternative methods. Rather than netting out the confounding effects by using the difference-in-discontinuities approach, the alternative methods try to tackle the effect from confounding regulations and sorting by excluding specific firms from the estimation. We consider three variations. First, a symmetric exclusion of firms in the close proximity of the threshold following the "donut" approach. Second, a targeted exclusion of firms that (passively or actively) could potentially have sorted around the threshold to either benefit from the subsidy or (to a lesser extent) avoid stricter reporting requirements. And third, a targeted exclusion of firms that may have had an incentive to adjust the employment level to avoid tighter reporting regulations as a consequence of the 2010 reform.

6.1 "Donut" RD design

One alternative to the difference-in-discontinuities method that allows adjusting for possible confounding factors or sorting and has been proposed in the literature are "donut" regressions (Barreca, Guldi, Lindo, and Waddell, 2011). This approach is most suited when sorting appears in the close proximity of the threshold. The principle underlying this approach is that it should be possible to estimate the effect of the reform also on a sample of firms within the optimal bandwidth, however, excluding a more narrow set of firms in the close proximity of the threshold. Naturally, there is a trade-off in choosing the size of the window that is excluded: The smaller the window the closer in spirit is the estimation to the local randomization approach and fulfills the requirements on similarity in covariates (overlap assumption). At the same time, the smaller the window, the less likely it is that all potential sorting cases are effectively excluded from the estimation.

Given the optimal bandwidth selections in our case (3,7, and 15) and the need to exclude a relevant set of potential sorters, the donut approach appears only meaningful at the higher bandwidths. In an attempt to balance the need to exclude a meaningful set of possible sorters, and keep a large enough set of employment levels to the left and right of the threshold, we consider excluding firms in the windows [-3;3], and [-5;5] and focus on the bandwidth of 15 employees around the threshold. For completeness, we also provide results when excluding only firms within the window [-1;1] and report also results for a bandwidth of 7. Regression results are reported in Table 6.

The donut regressions show that by excluding only one firm size-level to the left and right of the threshold, [-1;1], coefficient estimates remain broadly in line with the baseline regressions in Table 3. However, once the excluded window is increased to consider a larger set of possible sorting firms estimates indicate no significant difference in employment growth in 2013 for eligible firms below the 50 employees threshold and firms above the threshold.

Table 6: Donut regressions: Employment growth

	(1)	(2)	(3)	(4)	(5)	(6)
	Employment growth					
	Excluding [-1;1]					
	[-7;7]	[-15;15]	[-7;7]	[-15;15]	[-7;7]	[-15;15]
δ	0.017	0.014**	0.068***	0.030**	0.005	0.048**
	(0.010)	(0.006)	(0.017)	(0.012)	(0.045)	(0.018)
Polynomial	0	0	1	1	2	2
Observations	4,167	10,997	4,167	10,997	4,167	10,997
Adj. R-squared	0.000	0.000	0.001	0.000	0.001	0.000
	Excluding [-3;3]			*Excluding [-5;5]*		
	[-15;15]	[-15;15]	[-15;15]	[-15;15]	[-15;15]	[-15;15]
δ	0.012	0.024	0.059	0.007	-0.004	-0.099
	(0.007)	(0.022)	(0.058)	(0.008)	(0.024)	(0.093)
Polynomial	0	1	2	0	1	2
Observations	9,578	9,578	9,578	8,178	8,178	8,178
Adj. R-squared	0.000	-0.000	-0.000	-0.000	-0.000	-0.000

Results are obtained using a local parametric method. Each of the regressions includes a constant and the reported order of polynomial, which is allowed to differ on both sides of the threshold. δ captures the treatment effect at the threshold. Chosen bandwidths based on the covariate tests for 2013.
Robust standard errors, clustered both at the firm level and at the number of employees, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The results should, however, be considered with caution, because the method implies a significant reduction in the sample. The number of observations excluded are 1,063, 2,482 and 3,882 in the case of the excluded windows of [-1;1], [-3;3] and [-5;5], respectively. This corresponds to a reduction in the sample by 20 and 9 percent in the case of the [-1;1] window for the bandwidths of 7 and 15, respectively. The windows [-3;3] and [-5;5] reduce the sample by 20 and 32 percent.[39] The fall in the significance level may, thus, also be a result from reducing the number of observations and excluding possibly valid information.

6.2 Adjusting for "switchers"

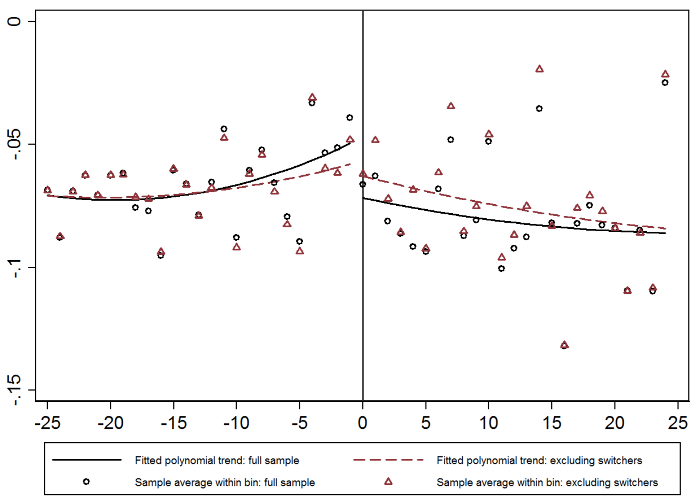
As a second robustness, we propose an approach that is similar in spirit to the donut RD design, but it is more selective in excluding observations. Making use of the time dimension of the dataset, we eliminate possible candidate observations, which could reflect firms that may have sorted around the threshold and perform the local randomization inference on the remaining set of firms. If there is sorting, the relevant firm would need to feature an employment level that is above the threshold in t-2, move then to

[39] Barreca, Guldi, Lindo, and Waddell (2011) consider sample reductions up to 11 percent.

a level below the threshold in t-1 to make use of the subsidy when employing new workers in t and as a consequence move broadly back to the initial level of employment. Because firms are also affected by other factors over time, the employment level is unlikely to end up at exactly the same level in t as the starting level in t-2. We therefore opt for a relatively broader concept. Specifically, we exclude all those firms from the estimation which in t-2 were within the range of 50 to 75 employees, in t-1 between 25 to 49 employees and in t again within the range of 50 to 75 employees.[40] Because sorting can occur in the pre-reform year and in any of the following years, the exclusion applies to $t \in (2012, 2013)$. Finally, as in the case of the donut regressions, firms with the exact opposite pattern are excluded as well from the estimation to ensure symmetric treatment. This is necessary, because if there was no sorting, excluding firms only in a one-sided manner, would lead to biased results. By construction, the excluded firms may also include those firms which may have reduced (and then increased) the employment level in order to circumvent the tighter reporting requirements following the 2010 reform.[41]

The advantage of this approach is that it allows for a wider window around the thresholds to be excluded, and targets directly those firms that show a pattern over time that would be consistent with sorting. The donut approach, instead, excludes all firms that fall within the chosen window, even if their historical employment level pattern would suggest that they have not engaged in sorting. Table 7 provides an overview of the results using this procedure for 2013 and 2011, reporting also the number of observations that are excluded.

Figure 6: Employment growth, 2013: Full sample vs. excluding “switching” firms



Local sample means over non-overlapping bins for firms with a number of employees between 25 and 75. Second-order polynomials fitted separately for treatment (less than 50 employees) and control (more than 50 employees) firms.

Focusing on the results for 2013, the local randomization inference approach applying this procedure

[40]This is broadly consistent with the average share of temporary contracts of 30 percent. Firms with more than 70 employees would need to let go workers on permanent contracts to fall below the 50 employees threshold, if their share of temporary workers would be 30 percent or lower. Similarly, firms with exactly 50 employees, could reduce their workforce (temporarily) to 35 employees without the need to let go workers under permanent contract, if their share of temporary workers would be 30 percent. This would imply a range of 35 to 70 employees. However, an exact matching is not needed as some firms do have a higher share of temporary contracts.

[41]A full separation is difficult given the overlap in the eligibility criteria and implied incentives.

Table 7: Excluding switchers: Employment growth

	(1)	(2)	(3)	(4)	(5)	(6)
	Employment growth					
	2011			2013		
	[-3;3]	[-7;7]	[-15;15]	[-3;3]	[-7;7]	[-15;15]
Polynomial (0)	**0.019**	0.006	0.006	**0.009**	0.002	0.003
	(0.211)	(0.517)	(0.412)	(0.565)	(0.848)	(0.660)
Polynomial (1)		**0.009**	**0.003**		**0.023****	**0.007**
		(0.387)	(0.623)		(0.027)	(0.286)
Polynomial (2)		0.023**	0.009		0.007	0.015**
		(0.024)	(0.198)		(0.505)	(0.032)
Observations	2,615	5,670	13,332	2,173	4,758	11,456
Excluded obs.	235	402	567	309	472	604

Results are obtained using a local randomization inference approach. Each coefficient is the result of a separate regression and captures the total treatment effect (including the higher-order polynomials). Bold coefficients present the results for the optimal window based on the choice of the polynomial and the significance level of the covariate tests.
Note: p-values, corrected for a finite sample, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

implies point estimates, which are broadly half the values obtained from the baseline regression results reported in Table 3. Moreover, while nearly all estimates were significantly different from zero in the baseline regressions, excluding potential switchers renders the vast majority of coefficient estimates non-significant, with only two cases showing a significant effect at the 5 percent level. Figure 6 provides a graphical representation of how results change at the threshold when excluding potential switchers (red dotted line). Compared to the full sample (black solid line replicating Figure 5e) the fitted polynomials to the left and right of the threshold are now very close. This is largely due to an upward adjustment of the employment growth of firms with more than 50 employees, consistent with sorting below the threshold. This casts additional doubt on the existence of an effect from the subsidy scheme on overall employment growth and underpins the possible relevance of firm sorting as a driver of the observed difference in employment growth between smaller and larger firms in the raw data. The number of observations excluded are 309, 472 and 604, which corresponds to a reduction in the sample by 12, 9, and 5 percent in the case of a bandwidth of 3, 7, and 15, respectively. This is well below the fall in the sample required for the donut regression and consistent with the range of the sample reduction considered in Barreca, Guldi, Lindo, and Waddell (2011), which ranges from 2 to 11 percent.

It is worth noting that also in 2011, we cannot find anymore a significant difference between employment growth of firms, which are just below the threshold compared to those just above. The stark contrast to the local randomization inference results in Section 5.1, where for all cases a positive significant difference between small and large firms was found, is remarkable given that in one case we only exclude 4 percent of the observations.[42] Based on these findings it appears, thus, plausible that sorting

[42]The numbers of observations excluded for the 2011 estimates are 235, 402 and 567, which corresponds to a reduction in the sample by 8, 7 and 4 percent in the case of a bandwidth of 3, 7 and 15, respectively.

in response to the 2010 and 2012 reforms may have been the main reason for differential employment growths at the threshold in the years 2011-2013.

6.3 Controlling for firms affected by the 2010 reform

In a final robustness exercise, we re-estimate Equation 1 for 2011 and 2013, but focus on firms which with certainty are unaffected by the 2010 reform concerning their employment decision. The 2010 reform creates the obligation for companies to externally audit accounts and to submit full balance sheets (as opposed to simplified ones) if at the end of the fiscal year and over two consecutive years, the company fulfills at least two of the following three conditions: Total assets higher than EUR 4 million; turnover in excess of EUR 8 million; or average number of workers greater than 50. As a consequence, firms which in two consecutive years have assets and turnover above the threshold would be subject to the stricter reporting regulation irrespective of their level of employment. Conversely, firms which in two consecutive years have assets and turnover below the threshold would *not* be subject to the stricter regulation irrespective of their level of employment. We use these two groups for the final robustness test as representing a very clear cut sample of firms which are unaffected by the 2010 reform, when it comes to their decision on the level of employment.[43] Again, it cannot be excluded that this definition also effectively excludes firms which sort below the threshold in order to benefit from the subsidy.

Table 8: Excluding firms affected by the 2010 reform: Employment growth

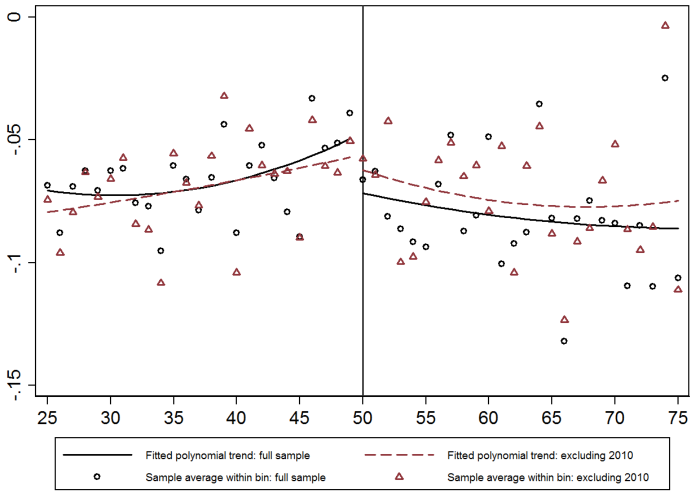
	(1)	(2)	(3)	(4)	(5)	(6)
	Employment growth					
	2011			2013		
	[-3;3]	[-7;7]	[-15;15]	[-3;3]	[-7;7]	[-15;15]
Polynomial (0)	**0.025**	0.004	0.009	**0.006**	0.006	0.006
	(0.125)	(0.730)	(0.265)	(0.740)	(0.632)	(0.483)
Polynomial (1)		**0.011**	**0.004**		**0.012**	**0.005**
		(0.362)	(0.555)		(0.315)	(0.538)
Polynomial (2)		0.034***	0.005		0.005	0.010
		(0.002)	(0.483)		(0.687)	(0.214)
Sample size treated	949	2342	6389	860	2105	5572
Sample size control	945	1687	2965	819	1439	2599

Results are obtained using a local randomization inference approach. Each coefficient is the result of a separate regression and captures the total treatment effect (including the higher-order polynomials). Bold coefficients present the results for the optimal window based on the choice of the polynomial and the significance level of the covariate tests. Note: p-values, corrected for a finite sample, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

While the advantage of this approach is that it allows focusing narrowly on those firms, which are clearly unaffected in their employment decision as a result of the 2010 reform, it leads to a significant

[43] There are other possible constellations where in a single year the level of employment would be irrelevant for reporting in time t. However, it would have possible repercussions for reporting requirements in $t + 1$. Thus, these cases are not considered here.

Figure 7: Employment growth, 2013: Full sample vs. excluding firms affected by the 2010 reform



Local sample means over non-overlapping bins for firms with a number of employees between 25 and 75. Second-order polynomials fitted separately for treatment (less than 50 employees) and control (more than 50 employees) firms.

reduction in the sample size of around 30 percent depending on the respective bandwidth. Table 8 provides an overview of the results using this procedure for 2013 and 2011.

Similar to the results when excluding potential 'switchers", results for 2013 imply that coefficient estimates are much smaller and insignificant. This re-enforces the doubts that the subsidy has been the reason for the initial estimate of a positive difference between employment growth of smaller firms compared to firms above the 50 employees threshold. As before, it is also here the case that coefficients for 2011 are not significant anymore as it would be expected, when no firm's employment decision is altered by the 2010 reform.

Figure 7 makes evident how the exclusion of the firms that may switch or may be affected by the 2010 reform alters the average employment growth above and below the 50 employees threshold. While there is essentially no change in the pattern of firms below 50 employees, firms with more than 50 employees displayed across the board higher employment growth when excluding firms that could have been affected in their employment decision by the 2010 reform.

7 Conclusion

Analysing the effect of recent labour market reforms in Europe is a daunting task given the lack of an evident control group or a suitable counter-factual. This study exploits an arbitrary size limit on firm eligibility for a hiring subsidy scheme under a new permanent contract, which was part of the Spanish labour market reform in 2012. We employ recent advances in regression discontinuity design to gauge the impact of the scheme on firm behavior and assess its impact on aggregate employment using local randomization inference and difference-in-discontinuities approaches.

The baseline local randomization inference results suggest that firms, which could make use of the reform, had about a 2 percentage points higher growth in employment in 2013, following the implementation of the reform. However, test results and placebo regressions, imply that the unconfoundedness assumption underlying the local randomization inference approach may not be satisfied. Adjusting for possible sorting by firms below the eligibility threshold and the existence of confounding factors using difference-in-discontinuities regressions and robustness tests, we fail to find robust evidence of an effect on the overall employment growth of eligible firms as a result of the subsidy. This finding is consistent with the fall in employment for firms just above the 50 employees threshold in 2011, possibly reflecting the non-renewal of temporary contracts. Based on the lack of support for an employment enhancing effect of the subsidy and amidst the fiscal costs associated with such a subsidy, caution may be warranted in considering the application and design of size-contingent hiring subsidy schemes.

Two caveats remain given lack of available data: First, based on the dataset, we could not investigate whether the subsidy has helped effectively increase the prevalence of permanent contracts to address duality of the Spanish labour market. Second, the analysis remains silent on the broader employment effect of the 2012 Spanish labour market reform, which is beyond the scope of this paper. As more post-reform data become available, future research could assess possible long-term effects of the subsidy, including the impact on job security and productivity gains that could counterbalance the lack of positive employment dynamics amidst the negative fiscal short-run effects of the subsidy.

References

Adhikari, B., R. Duval, B. Hu, and P. Loungani (2016): "Can Reform Waves Turn the Tide? Some Case Studies Using the Synthetic Control Method," IMF Working Paper 171.

Banco de España (2015): "Annual Report 2014," Discussion paper.

Barreca, A. I., M. Guldi, J. M. Lindo, and G. R. Waddell (2011): "Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification," *The Quarterly Journal of Economics*, 126(4), 2117–2123.

Bentolila, S., P. Cahuc, J. J. Dolado, and T. Le Barbanchon (2012): "Two-tier Labour Markets in the Great Recession: France versus Spain," *The Economic Journal*, 122(562), F155–F187.

Boeri, T., and P. Garibaldi (2007): "Two Tier Reforms of Employment Protection: A Honeymoon Effect?," *The Economic Journal*, 117(521), F357–F385.

Bordon, A., C. Ebeke, and K. Shirono (2016): "When Do Structural Reforms Work? On the Role of the Business Cycle and Macroeconomic Policies," IMF Working Paper 62.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014a): "Robust Data-Driven Inference in the Regression-Discontinuity Design," *The Stata Journal*, 14(4), 909–946.

——— (2014b): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82(6), 2295–2326.

——— (2015): "Optimal Data-Driven Regression Discontinuity Plots," *Journal of the American Statistical Association.*

Cappellari, L., C. Dell'Aringa, and M. Leonardi (2012): "Temporary Employment, Job Flows and Productiviy: A Tale of Two Reforms," *The Economic Journal*, 122(562), F188–F215.

Cattaneo, M. D., B. R. Frandsen, and R. Titiunik (2015): "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate," *Journal of Causal Inference*, 3(1), 1–24.

Cattaneo, M. D., R. Titiunik, and G. Vazquez-Bare (2016a): "Comparing Inference Approaches in RD designs: A Reexamination of the Effect of Head Start on Child Mortality," mimeo.

——— (2016b): "Inference in Regression Discontinuity Designs under Local Randomization," *Stata Journal*, 16(2), 331–367.

Dolado, J. J., C. Garcia-Serrano, and J. F. Jimeno (2002): "Drawing Lessons from the Boom of Temporary Jobs In Spain," *The Economic Journal*, 112(721), F270–F295.

EC (2016): "Country Report Spain 2016," European Commission Country Report.

ECB (2015): "Progress with Structural Reforms Across the Euro Area and Their Possible Impacts," ECB Economic Bulletin Issue 2.

Eggers, A. C., R. Freier, V. Grembi, and T. Nannicini (2016): "Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions," *The American Journal of Political Science*, forthcoming.

Ferreir, J., and F. Serrano (2001): "The Spanish Labour Market: Reforms and Consequences," *International Review of Applied Economics*, 15(1), 31–53.

Fiori, G., G. Nicoletti, S. Scarpetta, and F. Schiantarelli (2012): "Employment Effects of Product and Labour Market Reforms: Are There Synergies?*," *The Economic Journal*, 122(558), F79–F104.

Force, C. T. (2014): "Micro-based Evidence of EU Competitiveness, The CompNet Database," ECB Working Paper Series 1634.

Gal, P. N. (2013): "Measuring Total Factor Productivity at the Firm Level using OECD-ORBIS," OECD Economics Department Working Papers 1049.

Galasso, E., M. Ravallion, and A. Salvia (2001): "Assisting the Transition from Workfare to Work: Argentina's Proempleo Experiment," Development Research Group, World Bank.

García-Pérez, J. I., and M. Jansen (2015): "Assessing the Impact of Spain's Latest Labour Market Reform," *Spanish Economic and Financial Outlook*, 4(3), 5–15.

Grembi, V., T. Nannicini, and U. Troiano (2016): "Do Fiscal Rules Matter?," *American Economic Journal: Applied Econometrics*, 8(3), 1–30.

Groh, M., N. Krishnan, D. McKenzie, and T. Vishwanath (2012): "Soft Skills or Hard Cash? The Impact of Training and Wage Subsidy Programs on Female Youth Employment in Jordan," World Bank Impact Evaluation Series 62.

Hujer, R., M. Caliendo, and D. Radic (2001): "Estimating the Effects of Wage Subsidies on the Labour Demand in West-Germany Using the IAB Establishment Panel," *Ifo-Studien*, 47(2), 163–197.

Huttunen, K., J. Pirttila, and R. Uusitalo (2013): “The Employment Effects of Low-wage Subsidies,” *Journal of Public Economics*, 97(C), 49–60.

Imbens, G., and K. Kalyanaraman (2012): “Optimal Bandwidth Choice for the Regression Discontinuity Estimator,” *Review of Economic Studies*, 79(3), 933–959.

Imbens, G. W. (2004): “Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review,” *The Review of Economics and Statistics*, 86(1), 4–29.

Imbens, G. W., and T. Lemieux (2008): “Regression Discontinuity Designs: A Guide to Practice,” *Journal of Econometrics*, 142(2), 615–635.

IMF (2016): “Time for a Supply-Side Boost? Macroeconomic Effects of Labor and Product Market Reforms in Advanced Economies,” IMF WEO Chapter 3.

Kalemli-Ozcan, S., B. Sorensen, C. Villegas-Sanchez, V. Volosovych, and S. Yesiltas (2015): “How to Construct Nationally Representative Firm Level Data from the ORBIS Global Database,” NBER Working Paper Series 21558.

Lee, D. S. (2008): “Randomized Experiments from Non-Random Selection in U.S. House Elections,” *Journal of Econometrics*, 142(2), 675–697.

Lee, D. S., and D. Card (2008): “Regression Discontinuity Inference with Specification Error,” *Journal of Econometrics*, 142(2), 655–674.

Lee, D. S., and T. Lemieux (2010): “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature*, 48(2), 281–355.

Leonardi, M., and G. Pica (2013): “Who Pays for it? The Heterogeneous Wage Effects of Employment Protection Legislation,” *The Economic Journal*, 123(573), 1236–1278.

Ludwig, J., and D. L. Miller (2007): “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design,” *The Quarterly Journal of Economics*, 122(1), 159–208.

McCrary, J. (2008): “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test,” *Journal of Econometrics*, 142(2), 698–714.

OECD (2013): “The 2012 Labour Market Reform in Spain: A Preliminary Assessment,” OECD publishing 2013.

——— (2014): “OECD Economic Surveys Spain,” OECD publishing 2014.

Spanish Ministry of Labour (2013): “Report Evaluating the Impact of the Labour Reform,” Discussion paper.

Appendix

Figure A.1: Difference in distributions, 2013 vs. 2011

Difference in the density distributions of firms (2013-2011) with a number of employees between 25 and 75 .Two third-order polynomials fitted separately for treatment (less than 50 employees) and control (more than 50 employees) firms. Black lines show the lower and upper bound of the 95%-confidence interval.

Table A.1: Distribution of firm size: Discontinuity test at the threshold

	Polynomial: second-order					Polynomial: third-order				
	(A1)	(A2)	(A3)	(A4)	(A5)	(B1)	(B2)	(B3)	(B4)	(B5)
Bandwidth	15.00	20.00	25.00	30.00	35.00	15.00	20.00	25.00	30.00	35.00
2011	108.91**	143.53***	198.63***	298.48***	522.14***	118.67**	81.02	67.59	56.15	-45.60
	(42.70)	(35.19)	(37.19)	(57.58)	(118.49)	(52.01)	(58.00)	(53.10)	(56.72)	(85.42)
2010	90.46	105.49*	189.47***	302.29***	526.53***	64.91	46.16	10.56	7.65	-72.08
	(58.37)	(53.93)	(53.25)	(69.68)	(127.83)	(65.98)	(71.03)	(73.85)	(71.58)	(93.81)
2009	100.00**	126.79***	200.35***	335.08***	571.65***	69.50	51.12	36.51	11.01	-65.75
	(45.11)	(43.28)	(44.01)	(72.56)	(131.98)	(49.41)	(58.22)	(58.19)	(62.87)	(82.05)
2008	159.86***	216.52***	274.24***	403.46***	634.98***	118.26***	87.73**	110.73***	68.75	-2.10
	(32.01)	(42.65)	(47.07)	(82.58)	(134.94)	(42.74)	(43.37)	(39.51)	(56.89)	(71.92)
Observations	124	164	204	244	284	124	164	204	244	284
R-squared	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.99	1.00	1.00
Test (2010)	0.80	0.56	0.89	0.97	0.98	0.52	0.70	0.53	0.60	0.83
Test (2009)	0.89	0.76	0.98	0.69	0.78	0.49	0.72	0.69	0.59	0.87
Test (2008)	0.34	0.19	0.21	0.30	0.53	1.00	0.93	0.52	0.88	0.70

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.2: Local parametric results: Employment growth 2013

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					Employment growth				
	[-3 ; 3]	[-7,7]	[-15 ; 15]	[-3 ; 3]	[-7,7]	[-15 ; 15]	[-3 ; 3]	[-7,7]	[-15 ; 15]
δ	0.025***	0.016	0.013**	0.029***	0.040***	0.023**	0.048***	0.030	0.032***
	(0.006)	(0.009)	(0.006)	(0.005)	(0.009)	(0.010)	(0.002)	(0.019)	(0.010)
E_i				-0.008***	-0.000	-0.000	-0.001	-0.018***	-0.003
				(0.001)	(0.002)	(0.001)	(0.006)	(0.004)	(0.003)
E_i^2							-0.002	0.003***	0.000
							(0.002)	(0.000)	(0.000)
$1[E_i < E^T] \cdot E_i$				0.015***	0.006*	0.001	0.029***	0.028*	0.009**
				(0.002)	(0.003)	(0.001)	(0.006)	(0.015)	(0.004)
$1[E_i < E^T] \cdot E_i^2$							0.008**	-0.002	0.000
							(0.002)	(0.002)	(0.000)
Constant	-0.073***	-0.075***	-0.076***	-0.063***	-0.074***	-0.076***	-0.065***	-0.059***	-0.070***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.007)	(0.006)	(0.002)	(0.007)	(0.005)
Observations	2,482	5,230	12,060	2,482	5,230	12,060	2,482	5,230	12,060
Adj. R-squared	0.001	0.000	0.000	0.000	0.001	0.000	-0.000	0.001	0.000

Note: Standard errors, clustered at the number of employees values, in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Acknowledgements
The views expressed here are those of the authors and not necessarily those of the ECB. We are greateful to Isabel Vansteenkiste, Matthias Schüdeln, Beatrice Pierluigi, Ettore Dorrucci, Robert Anderton, and various conference participants for their helpful comments.

Elisa Gamberoni
European Central Bank, Frankfurt am Main, Germany; email: elisa.gamberoni@ecb.europa.eu

Katerina Gradeva
European Central Bank, Frankfurt am Main, Germany; email: katerina.gradeva@ecb.europa.eu

Sebastian Weber
European Central Bank, Frankfurt am Main, Germany; email: sebastian.weber@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

ISSN 1725-2806 (pdf)
ISBN 978-92-899-2218-0 (pdf)
DOI 10.2866/559447 (pdf)
EU catalogue No QB-AR-16-087-EN-N (pdf)