MSc Economics

Peer effect estimation through a covariate-adjusted Regression Discontinuity Design

A Master's Thesis submitted for the degree of
"Master of Science"

supervised by

Justinas Pelenis

Simon Zuzek

1046358

Vienna, June 5, 2017

MSc Economics

Affidavit

I, Simon Zuzek

hereby declare

that I am the sole author of the present Master's Thesis,

Peer effect estimation through a covariate-adjusted Regression Discontinuity Design

28 pages, bound, and that I have not used any source or tool other than those referenced or any other illicit aid or tool, and that I have not prior to this date submitted this Master's Thesis as an examination paper in any form in Austria or abroad.

Vienna, June 5, 2017

Signature

Contents

List of Figures

List of Tables

Abstract

This thesis uses data on the Austrian labor market to obtain an estimate of a peer effect in parental leave decisions among young fathers. Identification arises through a reform of parental leave, when higher benefits led to an increase in the participation rate of 4.7 – 6.6 percentage points. I use the resulting discontinuity around the implementation date of the reform to implement a fuzzy Regression Discontinuity Design (RDD), which allows to obtain estimates of a peer effect of reform-window fathers on their coworkers. I include additional covariates on the observations in order to increase the precision of the estimator and discuss relevant theoretical results on asymptotic inference. The results indicate a quantitatively large increase of the coworker participation rate of 13 – 28 percentage points, however statistical significance of the peer effect is weak.

1 Introduction

Regression Discontinuity Designs (RDD) have seen a surge in popularity in applied economic research over the past decades. By what was later called the ”Credibility Revolution” of empirical economics by Angrist and Pischke (2010), RDDs are highly appreciated by applied researchers for their strong internal validity based on the exploitation of quasi-experimental variation induced by a discontinuous “jump” in an assignment variable.

In this thesis, I consider an application brought forth by Dahl et al. (2014), where a Regression Discontinuity can be used to obtain credible estimates of a peer effect in the workplace. Causal estimation of peer effects has proven to be difficult due to multiple endogeneity problems such as correlated unobservables and endogenous group membership among separate peer groups. I will make use of a reform in the Austrian parental leave law which caused a discontinuity of the benefits provided to young parents around the implementation date of the reform. Newly introduced income-dependent parental leave benefits resulted in an increased participation rate among fathers. By using extensive employment data from Austria, I construct a sample of fathers with births around the reform implementation date and corresponding coworkers whose children were born after the reform took place. The exogenous shock in the leave rate around the reform cutoff can then be used to credibly infer the peer effect which the participation decision of reform fathers had on their coworkers. A detailed discussion of the peer effect model and the dataset is given in Sections 4 & 5.

Before introducing the model, I first present a discussion of the theory behind Regression Discontinuity Designs. I devote some attention to novel results by Calonico et al. (2016) about the inclusion of covariates, since additional regressors are often added in applied work in an attempt to increase precision of the estimators. As an extension to their work, I argue that under their assumptions the inclusion of covariates reduces asymptotic variance of the estimator. In order to show the impact of violating a seemingly innocuous assumption, I further add a simulation example in which covariate-adjustment increases the mean squared error due to asymmetric sample sizes. Concerning covariates on fathers in the empirical application, I discuss whether including the additional regressors “age”, “income” and “Austrian nationality” is permissible in the example at hand.

The rest of the paper is structured as follows. Section 2 provides a literature review of theoretical results on Regression Discontinuity Designs and introduces the general model setup and notations. Section 3 provides results by the literature regarding asymptotic behavior of the covariate-adjusted treatment estimator and examines cases under which covariates improve or worsen desirable properties. Section 4 introduces the empirical model, where discontinuous father leave incentives due to a reform allow the identification of peer effects in the workplace. Section 5 introduces the dataset which I use to construct a sample of peers. The parental leave reform which caused the discontinuity is described in Section 5.1 and the inclusion of covariates into the model is discussed in Section 5.2. Section 6 presents empirical results. Section 7 concludes.

2 Theory of Regression Discontinuity Designs

Regression Discontinuity is a quasi-experimental research design viable when an assignment variable experiences an exogenous discontinuity. It often arises artificially as a by-product of some cutoff value mandated by regulation. An early example concerns the effect of class sizes on learning outcomes by Angrist and Lavy (1999), where school classes are to be separated after a maximum of 40 students is reached. Another popular example arises in the field of political economy, where a two-party system in combination with a majority voting mechanism results in election winners who obtain voting shares slightly above 50% and therefore close to their respective political rivals. Lee (2008) exploits the variation to investigate the effect that incumbency has on voting outcomes.

Parallel to its increasing popularity among applied researchers, theoretical results concerning RDDs have kept pace. In earlier applications such as Angrist and Lavy (1999), Instrumental Variables (IV) assumptions where used to identify causal effects, requiring exogeneity of the assignment variable. Hahn et al. (2001) put RDDs into the treatment effects framework and argue for weak functional form assumptions as identification strategies. In this framework, every unit has two possible states with respect to the outcome variable Y_i – treatment $Y_i(1)$ and control $Y_i(0)$. The assignment variable X_i determines whether a unit receives treatment, in which case T_i equals 1, or is part of the control group with T_i set equal to 0. Consider a "sharp" RD design, in which treatment is a deterministic function of the assignment variable, i.e. $T_i = 1$ if $X_i \geq \bar{x}$ and $T_i = 0$ if $X_i < \bar{x}$. In general, one would like to identify the treatment effect $\beta(x) \equiv E[Y_i(1) - Y_i(0)|X_i = x]$ as a function of x. In the case of school classes, for example, it is perceivable that splitting classes at a higher cutoff induces greater benefits. However, identifying the treatment effect as a function of x is not possible in general since observations of $Y_i(1)$ do not exist for $x < \bar{x}$ and vice versa. A Regression Discontinuity, on the other hand, will allow to identify a local treatment effect

$$\beta \equiv \beta(\bar{x}) = E[Y_i(1) - Y_i(0)|X_i = \bar{x}].$$

While in "sharp" designs treatment depends deterministically on the assignment variable, "fuzzy" designs describe an environment where treatment is a random variable whose conditional expectation $E[T_i(t)|X_i = x]$ depends on x. [1].

Let's introduce a setting where treatment is random. As an example, consider a legislation on firms which intends to reduce harmful activities Y_i, yet compliance is

[1]Note that sharp designs are a special case of fuzzy RDDs where $E[T_i|X_i = x] = T_i(1)$ for $x \geq \bar{x}$ and $E[T_i|X_i = x] = T_i(0)$ for $x < \bar{x}$.

imperfect. The treatment "compliant" ($T_i = 1$) is a random variable and value X_i serves as a predictor of compliance for each individual i. Think of X_i as firm size and $\bar{x}$ a threshold after which auditing becomes more likely. Consider the parameter of interest $\tau(X_i)$ as the effect which compliant behavior has on some outcome variable Y_i,

$$Y_i = \alpha + \tau(X_i)T_i + \epsilon_i,$$

where $Y_i(0) = \alpha + \epsilon_i$ and $Y_i(1) = \alpha + \tau(X_i) + \epsilon_i$ are the two possible states, of which only one is observed for each individual firm. The parameter of interest is the treatment effect $\tau(x) = E[\tau(X_i)|X_i = x] = E[Y_i(1) - Y_i(0)|X_i = x]$. Note that it is possible that the treatment effect depends on the firm size X_i.

Under standard regression assumptions, the error term ϵ_i is required to be uncorrelated with compliance T_i and then consistent estimates of the treatment effect $\tau(x)$ can be obtained. It is, however, not unreasonable to assume that the intensity of harmful behavior Y_i and the propensity to comply with regulations are related through unobserved firm characteristics. Some firms, for example, could be intrinsically more interested in reducing harmful activities and are therefore more willing to comply with related regulations. In such a case of omitted variable bias, the error term ϵ_i and "compliance" T_i are correlated and an identification of $\tau(x)$ is not possible in a standard OLS setting.

Hahn et al. (2001) consider a competing set of assumptions under which a local treatment effect at $\bar{x}$ can nonetheless be obtained under a fuzzy Regression Discontinuity Design.

Assumption 1: $lim_{x\downarrow\bar{x}}E[T_i|X_i = x] \neq lim_{x\uparrow\bar{x}}E[T_i|X_i = x]$

Assumption 2: $E[Y_i(0)|X_i = x]$ and $E[Y_i(1)|X_i = x]$ are continuous.

Specifically, Assumption 1 imposes a discontinuity in expected treatment around the cutoff value $\bar{x}$ of the assignment variable, whereas Assumption 2 assures that the outcome variable behaves smoothly around $\bar{x}$. Intuitively, these assumptions guarantee that any observed discontinuity of the outcome at $\bar{x}$ can be attributed to the difference in expected treatment. In order to understand the relevance of these assumptions, it is instructive to restate the derivations of Hahn et al. (2001) under Theorem 1.

Theorem 1 by Hahn et al. (2001): Suppose Assumptions 1 & 2 hold. Then

$$\tau \equiv \tau(\bar{x}) = \frac{lim_{x\downarrow\bar{x}}E[Y_i|X_i = x] - lim_{x\uparrow\bar{x}}E[Y_i|X_i = x]}{lim_{x\downarrow\bar{x}}E[T_i|X_i = x] - lim_{x\uparrow\bar{x}}E[T_i|X_i = x]}$$

Proof by Hahn et al., 2001: Consider small $e > 0$. Then

$$E[Y_i|X_i = \bar{x} + e] - E[Y_i|X_i = \bar{x} - e] =$$
$$\tau(\bar{x} + e)E[T_i|X_i = \bar{x} + e] - \tau(\bar{x} - e)E[T_i|X_i = \bar{x} - e]+$$
$$E[\alpha + \epsilon_i|X_i = \bar{x} + e] - E[\alpha + \epsilon_i|X_i = \bar{x} - e]$$

Consider $e \to 0$. Since $Y_i(0) = \alpha + \epsilon_i$ and $Y_i(1) = \alpha + \tau(X_i) + \epsilon_i$, Assumption 2 allows

$$lim_{x \downarrow \bar{x}} E[Y_i|X_i = \bar{x}] - lim_{x \uparrow \bar{x}} E[Y_i|X_i = \bar{x}] =$$
$$\tau(\bar{x})\{lim_{x \downarrow \bar{x}} E[T_i|X_i = \bar{x}] - lim_{x \uparrow \bar{x}} E[T_i|X_i = \bar{x}]\}$$

Note that continuity of $\tau(x)$ is implied by the assumption about $Y_i(1)$. Finally, Assumption 1 guarantees that the denominator is defined.

□

Whereas these assumptions are notably weaker than IV conditions, estimation poses different questions. A practical strategy is to apply polynomial regressions of the outcome variable on the assignment on both sides of the cutoff to fit the unknown regression functions. Optimally, observations arbitrarily close on both sides of the cut-off value should be available to derive good estimates of the conditional expectations in the limit. As this will hardly be the case, an appropriate subset of the data around $\bar{x}$ has to be chosen by the researcher. Imbens and Kalyanaraman (2012), for example, develop a data-driven bandwidth algorithm which is optimal under squared error loss.

3 Covariates in RDDs

Applications of RDD often include covariates, although their inclusion is not necessary under the assumptions above. Indeed, even omitting covariates which are correlated with the assignment variable does not threaten identification of the treatment effect. Consider an example of endogeneity, where $E[T_i\epsilon_i] \neq 0$. It is easily seen that this moment condition, which prohibits a consistent estimator in an OLS setting, imposes no issues on the derivations of Theorem 1.

Nonetheless, covariates are often used in applied Regression Discontinuity settings driven by the outlook of potential efficiency gains. A theoretical foundation derives from considering Regression Discontinuities as local randomization devices which constitute an experiment near the cutoff. Since covariates are often used in experimental settings to increase precision of the parameters of interest, a similar reasoning in an RDD setting seems to be applicable. A discussion of RDDs under this framework can be found in Cattaneo et al. (2015) with an application to the data used by Lee (2008).

An early paper on the topic of covariate inclusion in Regression Discontinuity Designs is written by Frölich (2007), who derives conditions for the case of continuously valued covariates. I will focus below on the analysis by Calonico et al. (2016), who develop asymptotic properties, optimal bandwidth and robust standard errors for a covariate adjusted local polynomial estimator via the inclusion of further continuity assumptions and without invoking a local randomization framework.

Let's consider an extension of the example introduced in Section 2, where Z_i is a vector of covariates for the firms in question, for example data on the market structure. Moreover, consider that contrary to the linear model before, the exact functional forms of $E[Y_i(1)|X_i = x]$ and $E[Y_i(0)|X_i = x]$ are unknown, yet satisfy Assumptions 1 & 2. Covariates enter linearly into the model which is not a crucial assumption. Additionally, for simplification assume a "sharp" RD design where treatment is received for observations above $\bar{x}$ and not otherwise, i.e. all firms larger than a cutoff size will be audited with certainty and those which fall below the cutoff will not be audited.

$$Y_i = g_-(X_i - \bar{x}) + g_+(X_i - \bar{x}) + \mathbf{Z}_i'\gamma + \eta_i$$

Here, $g_-(.)$ and $g_+(.)$ represent the unknown relationship between firm size and compliant behavior for firms below and above the cutoff, respectively. By assumption, $g_-(x) = 0$ for $x \geq \bar{x}$ and $g_+(x) = 0$ for $x < \bar{x}$. The treatment effect is then defined as $\tau \equiv g_+(0) - g_-(0)$. In order to estimate the treatment effect, it is common to approximate the functions $g_-(.)$ and $g_+(.)$ by polynomial expansions. Let $r_-(.)$ and $r_+(.)$ be such polynomial expansions of order p with coefficients β_- and β_+, respectively, such that

$$r_+(x)'\beta_+ = \mathbb{1}(X_i \geq \bar{x})(\beta_{+,0} + \beta_{+,1}x + \cdots + \beta_{+,p}x^p)$$
$$r_-(x)'\beta_- = \mathbb{1}(X_i < \bar{x})(\beta_{-,0} + \beta_{-,1}x + \cdots + \beta_{-,p}x^p).$$

The proposed estimator for the treatment effect is $\tilde{\tau}(h) = \tilde{\beta}_{+,0}(h) - \tilde{\beta}_{-,0}(h)$, where $K_h(.)$ is a kernel function with a positive bandwidth h and the vectors $\tilde{\beta}_+(h)$ and $\tilde{\beta}_-(h)$ are derived by minimizing the sum of squared residuals as below. The estimator depends on a choice for the bandwidth h and a polynomial degree p as well as the selection of a particular kernel.

$$\begin{bmatrix} \tilde{\beta}_-(h) \\ \tilde{\beta}_+(h) \\ \tilde{\gamma}(h) \end{bmatrix} = \underset{\beta_-,\beta_+,\gamma}{argmin} \sum_{i=1}^{n} (Y_i - r_-(X_i - \bar{x})'\beta_- - r_+(X_i - \bar{x})'\beta_+ - \mathbf{Z}_i'\gamma)^2 K_h(X_i - \bar{x}).$$

Recall from Section 2 that continuity of the conditional expectations of the outcome variables with and without treatment, $Y_i(1)$ and $Y_i(0)$, was a necessary condition for correct identification of the treatment effect. Intuitively, the condition allowed to attribute any effect which arose as a result of the discontinuity in the assignment variable at $\bar{x}$ to the effect which treatment has on the outcome. Similarly, the inclusion of covariates requires a continuity assumption on the conditional expectations of the covariates next to the cutoff. Denote $Z_i(1)$ as the covariates of observations before the cutoff and $Z_i(0)$ as the covariates of observations after the cutoff, i.e. $Z_i = Z_i(1)\mathbb{1}(X_i \geq \bar{x}) + Z_i(0)\mathbb{1}(X_i < \bar{x})$. Calonico et al. (2016) employ the following assumptions for a vector of covariates Z_i, where p denotes the order of polynomials employed by the estimator.

For $\rho \geq p + 2$ and $x \in [x_l, x_u]$ with $x_l < \bar{x} < x_u$, consider the following assumptions (Assumptions 2a-2e in Calonico et al. (2016)):

Assumption 3a The Lebesgue density of X_i, denoted by $f(x)$, is continuous and bounded away from zero.

Assumption 3b $\mu_{Y-}(x) \equiv E[Y_i(0)|X_i = x]$ and $\mu_{Y+}(x) \equiv E[Y_i(1)|X_i = x]$ are ρ times continuously differentiable.

Assumption 3c $\mu_{Z-}(x) \equiv E[Z_i(0)|X_i = x]$ and $\mu_{Z+}(x) \equiv E[Z_i(1)|X_i = x]$ are ρ times continuously differentiable. Moreover, $\mu_{Z-}(x) \equiv E[Z_i(0)Y_i(0)|X_i = x]$ and $\mu_{Z+}(x) \equiv E[Z_i(1)Y_i(1)|X_i = x]$ are continuously differentiable.

Assumption 3d $V[(Y_i(t), Z_i(t)')']$ are continuously differentiable for $t \in \{0, 1\}$.

Assumption 3e $E[|(Y_i(t), Z_i(t)')'|], t \in \{0, 1\}$ are continuous, with $|.|$ denoting the Euclidean norm.

Assumptions 3b - 3e pertain to the continuity of relevant moments. The main contribution to the inclusion of covariates comes through Assumption 3c, which captures the intuition stated above. Assumption 3a puts a constraint on the distribution of the assignment variable X_i, whose relevance will be seen in the simulation study below.

It is worth noting that so far no reason was given as to why $\tilde{\tau}(h)$, the difference of intercept between the fitted polynomials before and after the cutoff as defined before, would be a good estimator for τ irrespective of the choice of bandwidth h. Indeed, $\tilde{\tau}(h)$ may well be different from the true parameter τ in any sample. However, Assumptions 1, 2 & 3 are sufficient for asymptotic arguments to apply on the estimator of interest. Specifically, Calonico et al. (2016) show that for $\tilde{\tau}(h)$ the following holds:

Lemma 1 by Calonico et al. (2016): Let the assumptions hold and $nh \to \infty, h \to 0$. Then

$$\tilde{\tau}(h) \underset{p}{\to} \tau - [\mu_{Z+} - \mu_{Z-}]'\gamma_Y,$$

with[2]

$$\gamma_Y = (\sigma^2_{Z+} + \sigma^2_{Z-})^{-1} E[(Z_i(0) - \mu_{Z-}(X_i))Y_i(0) + (Z_i(1) - \mu_{Z+}(X_i))Y_i(1)|X_i = \bar{x}]$$

Lemma 1 shows that $\tilde{\tau}(h)$ can indeed be a good estimator in an asymptotic sense. For a sufficiently small bandwidth and large number of observations, $\tilde{\tau}(h)$ converges in probability to an expression which could be τ if either $\mu_{Z+} = \mu_{Z-}$ or $\gamma_Y = 0$. Consistency of $\tilde{\tau}(h)$ depends therefore on imposing one additional moment condition. The first possible condition implies that treatment cannot induce jumps in the covariate expectations for observations at the cutoff value and again intuitively captures the assumption that any discontinuous behavior at $\bar{x}$ can be attributed to the effect of assignment on Y_i (and nothing else). The second condition is fulfilled if the partial effects of Z_i on Y_i before and after the cutoff cancel each other out and is rather unlikely to hold in a specific setting.

For the fuzzy estimator, additional continuity assumptions on treatment T_i have to hold. Consider that as in Theorem 1, the true parameter of interest in the fuzzy RD is $\tau_{fuzzy} \equiv \frac{\tau_Y}{\tau_T}$, which is the ratio of two sharp RD designs.[3]

[2]Notation $\sigma^2_{Z+} = V[Z_i(1)|X_i = \bar{x}]$, $\sigma^2_{Z-} = V[Z_i(0)|X_i = \bar{x}]$, $\mu_{Z+} \equiv \mu_{Z+}(\bar{x})$ and $\mu_{Z-} = \mu_{Z-}(\bar{x})$ and respectively for Y_i.

[3]Notation: $\tau_Y \equiv lim_{x \downarrow \bar{x}} E[Y_i|X_i = x] - lim_{x \uparrow \bar{x}} E[Y_i|X_i = x]$ and $\tau_T \equiv lim_{x \downarrow \bar{x}} E[T_i|X_i = x] - lim_{x \uparrow \bar{x}} E[T_i|X_i = x]$.

Assumption 4: $\mu_{T-}(x) \equiv E[T_i(0)|X_i = x]$, $\mu_{T+}(x) \equiv E[T_i(1)|X_i = x]$, $E[Z_i(0)T_i(0)]$ and $E[Z_i(1)T_i(1)]$ are continuous.[4]

Lemma 2 (Lemma 3 by Calonico et al. (2016)): Let assumptions 1, 2, 3 & 4 hold and $nh \to \infty, h \to 0$. Then

$$\tilde{\tau}(h)_{fuzzy} \underset{p}{\to} \frac{\tau_Y - [\mu_{Z+} - \mu_{Z-}]'\gamma_Y}{\tau_T - [\mu_{Z+} - \mu_{Z-}]'\gamma_T}$$

with γ_Y defined as above and

$$\gamma_T = (\sigma^2_{Z+} + \sigma^2_{Z-})^{-1} E[(Z_i(0) - \mu_{Z-}(X_i))T_i(0) + (Z_i(1) - \mu_{Z+}(X_i))T_i(1)|X_i = \bar{x}]$$

It is interesting to note that very similar conditions for consistency are sufficient as in the case of a sharp estimator, as can be seen by Lemma 2. In fact, the moment condition on the covariates, $\mu_{Z+} = \mu_{Z-}$, is sufficient for a consistent estimator in both sharp and fuzzy RD designs. On the other hand, the fuzzy case now requires that the partial effects of Z_i both on Y_i and T_i cancel out before and after the cutoff.

It is worth mentioning that the estimators do not allow for a different partial effect of the covariate before and after the cutoff. Moreover, such a specification poses a relevant threat to consistency as shown by Calonico et al. (2016).

The benefit of covariates hangs on the hope that additional observables might reduce noise and therefore lead to more precision of the estimator through a reduction in variance. The question boils down to comparing the variances of the estimators $\tilde{\tau}$ and $\hat{\tau}$, where the latter is the standard estimator derived without the inclusion of covariates, i.e. by setting γ equal to zero in the estimation function above. Consider their respective variances as provided by Calonico et al. (2016) [5]

$$V_{\tilde{\tau}} = (V[Y_i(0) - Z_i(0)'\gamma_Y|X_i = \bar{x}] + V[Y_i(1) - Z_i(1)'\gamma_Y|X_i = \bar{x}])e_0'\Lambda_{p+}e_0 f(\bar{x})^{-1}$$
$$V_{\hat{\tau}} = (V[Y_i(0)|X_i = \bar{x}] + V[Y_i(1)|X_i = \bar{x}])e_0'\Lambda_{p+}e_0 f(\bar{x})^{-1}.$$

Through the restriction that the partial effect of a covariate may not differ in the estimation specification, the question of ordering the two variances remains unclear. Calonico et al. (2016) note the interesting special case when $\gamma_Y = \gamma_{Y+} = \gamma_{Y-}$.[6] This is exactly true if the specification with different partial effects is equal to the

[4] Notation: $T_i = \mathbb{1}(X_i \geq \bar{x})T_i(1) + \mathbb{1}(X_i < \bar{x})T_i(0)$

[5] f(x) denotes the Lebesgue density of X_i at x. e_0 is a vector with 1 as its first component and 0 else. Λ_{p+} is a kernel integral. Most importantly, $e_0'\Lambda_{p+}e_0 f(\bar{x})^{-1}$ do not depend on the data process.

[6] Notation: $\gamma_{Y+} \equiv (\sigma^2_{Z+})^{-1}E[(Z_i(0) - \mu_{Z+}(X_i))Y_i(0)|X_i = \bar{x}]$ and $\gamma_{Y-} \equiv (\sigma^2_{Z-})^{-1}E[(Z_i(0) - \mu_{Z-}(X_i))Y_i(0)|X_i = \bar{x}]$

specification used in Lemma 1. Then, γ_Y is the best linear approximation on either side of the discontinuity and therefore

$$V[Y_i(t)|X_i = \bar{x}] \geq V[Y_i(t) - Z_i(t)'\gamma_Y|X_i = \bar{x}], t \in \{0, 1\}.$$

As an extension, I argue that the inclusion of covariates generally leads to a reduction in asymptotic variance.

Proposition 1: Let the assumptions hold and the variances of $\tilde{\tau}$ and $\hat{\tau}$ therefore be as derived by Calonico et al. (2016). Then

$$\begin{gathered} V[Y_i(0) - Z_i(0)'\gamma_Y|X_i = \bar{x}] + V[Y_i(1) - Z_i(1)'\gamma_Y|X_i = \bar{x}] \\ \leq \\ V[Y_i(0)|X_i = \bar{x}] + V[Y_i(1)|X_i = \bar{x}] \end{gathered}$$

Proof: To see this, consider first that for $t \in \{0, 1\}$ and $-, +$ respectively, γ_{Y-} and γ_{Y+} can be written as

$$\begin{aligned} \gamma_{Y\mp} &= (\sigma^2_{Z\mp})^{-1} E[Y_i(t)(Z_i(t) - \mu_{Z\mp})|X_i = \bar{x}] \\ &= (\sigma^2_{Z\mp})^{-1} E[Y_i(t)Z_i(t) - Y_i(t)\mu_{Z\mp}|X_i = \bar{x}] \\ &= (\sigma^2_{Z\mp})^{-1} Cov(Y_i(t), Z_i(t)|X_i = \bar{x}). \end{aligned}$$

Then

$$\begin{aligned} &V[Y_i(t) - Z_i(t)\gamma_Y|X_i = \bar{x}] - V[Y_i(t)|X_i = \bar{x}] = \\ &V[Z_i(t)|X_i = \bar{x}]\gamma_Y^2 - 2Cov(Y_i(t), Z_i(t)|X_i = \bar{x})\gamma_Y \\ &= \sigma^2_{Z\mp}\gamma_Y^2 - 2\gamma_Y\sigma^2_{Z\mp}\gamma_{Y\mp} \end{aligned}$$

Moreover, γ_Y can be written as a linear combination of γ_{Y+} and γ_{Y-}

$$\begin{aligned} \gamma_Y &= (\sigma^2_{Z+} + \sigma^2_{Z-})^{-1} E[Y_i(0)(Z_i(0) - \mu_{Z-}) + Y_i(1)(Z_i(1) - \mu_{Z+})|X_i = \bar{x}] \\ &= (\sigma^2_{Z+} + \sigma^2_{Z-})^{-1}(\sigma^2_{Z-}\gamma_{Y-} + \sigma^2_{Z+}\gamma_{Y+}). \end{aligned}$$

So Proposition 1 can be written as

$$\begin{aligned} \sigma^2_{Z+}\gamma_Y^2 + \sigma^2_{Z-}\gamma_Y^2 &\leq 2\gamma_Y(\sigma^2_{Z-}\gamma_{Y-} + \sigma^2_{Z+}\gamma_{Y+}) \\ \sigma^2_{Z+}\gamma_Y^2 + \sigma^2_{Z-}\gamma_Y^2 &\leq 2\gamma_Y^2(\sigma^2_{Z+} + \sigma^2_{Z-}) \end{aligned}$$

where the inequality always holds.

$\square$

Lemma 1 and Proposition 1 show that under the general assumptions put forth by Calonico et al. (2016), the inclusion of covariates is asymptotically not harmful.

In order to show that the inclusion of covariates does not generally result in more precise estimates, I present a numerical example where the inclusion of covariates is asymptotically harmful, even though all necessary continuity and moment conditions hold. As put forth by Freedman (2008), a potential pitfall arises when the ratio of observations before and after the cutoff differs. Such a systematic difference in the number of observations may arise in a setting where the cutoff results in different sampling probabilities. Consider the above example, where $\bar{x}$ was a cutoff value after which auditing becomes more likely. If the data was collected during audits, the sampling process would also systematically include more observations above the cutoff.

Note that this example hinges on a violation of Assumption 3a, stating that the Lebesgue density of the assignment variable is continuous in an interval around the cutoff value. Consider now a data generating process X_i and errors $\epsilon_{Z,i}$ and $\epsilon_{Y,i}$,

$$X_i \sim \begin{cases} U(-1,0), & \text{with probability } \theta \\ U(0,1), & \text{with probability } 1-\theta \end{cases}$$
$$\begin{pmatrix} \epsilon_{Z,i} \\ \epsilon_{Y,i} \end{pmatrix} \sim N(0,\Sigma), \Sigma = \begin{pmatrix} \sigma_z^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$$

where $\theta \in (0,1)$ is the fraction of observations before the cutoff. Note that the data generating process violates Assumption 3a, as the density of X_i changes discontinuously at $\bar{x}$. The outcome variable Y_i and covariate Z_i are generated around the cutoff value $\bar{x} = 0$ in the following way:

$$Z_i = \beta_{Z-}X_i + \mathbb{1}(X_i \geq \bar{x})(\beta_{Z+} - \beta_{Z-})X_i + \epsilon_{Z,i}$$
$$Y_i = \beta_{Y-}X_i + \gamma_{Y-}Z_i + \mathbb{1}(\tau + X_i \geq \bar{x})(\beta_{Z+}X_i - \beta_{z,-}X_i + \gamma_{Y+}Z_i - \gamma_{Y-}Z_i) + \epsilon_{Y,i}$$

I consider 5000 simulations of a sample size of 1000 each for the following parameters: $\gamma_{Y+} = -1, \gamma_{Y-} = 2, \beta_{Z+} = 5, \beta_{Z+} = 2, \beta_{Y+} = -3, \beta_{Y-} = 4, \tau = 0.5, \sigma_Y = 1$ and $\sigma_Z = 1$. Most importantly, $\theta = 0.2$ allocates a large portion of the observations to the right of the cutoff value. Table 1 reports mean squared error, empirical coverage rate of the 95% confidence interval, the mean length of this interval and mean bias for a simulation with the above parameters. On the left, I report results with a uniform kernel on the full bandwidth, which includes all observations and weighs equally. The right side reports results for a RD estimate which employs a coverage rate optimal bandwidth and triangular kernel. Both estimates are obtained through the "rdrobust" package of Calonico et al. (2015) with linear trends on each side of the cutoff. In both cases, the inclusion of covariates increases the mean squared error and the average length of the confidence interval, whereas mean bias and coverage rate stay largely unaffected. Note that the perceived superiority of a uniform kernel with full bandwidth is owed to the fact that the linear specification which was employed in the estimation corresponds to the true model and should not be taken as an argument against more conservative bandwidth and kernel choices.

Table 1: Simulation results (fraction of observations before the cutoff $\theta = 0.2$)

	Uniform & full bandwidth		Triangular & CER optimal	
	With cov.	Without cov.	With cov.	Without cov.
MSE	0.1411	0.1092	0.9906	0.8571
Coverage Rate	0.9458	0.946	0.9058	0.9124
Bias	-0.0032	-0.0016	0.0002	-0.0046
Interval Length	1.474	1.298	3.512	3.061
Bandwidth			0.2459	0.2054

4 Identifying peer effects with a Regression Discontinuity Design

As mentioned in the introduction, I consider a model and identification strategy introduced by Dahl et al. (2014), who examine how participation in parental leave by new fathers trickles through a personal network. Such a peer effect has relevant implications from two perspectives. First, encouraging fathers to take parental leave has been an important policy goal in many countries, yet participation in newly introduced policies can be sluggish. Early participants may have an effect on their peers via the transmission of information or via signaling a change of social culture in their peer group. Therefore, targeting programs to make use of relevant personal networks has potential social welfare gains and is an important component of policy analysis.

Second, a general interest on the impact of social groups on economic decisions has resurfaced in economics, where the impact of various social networks on participation in a general government programs poses an interesting side question. Brown and Laschever (2012), for example, study retirement decisions by public school teachers and find positive effects of retirement decisions on coworkers of peers. A different line of research on the peer effect in economic decisions concerns workplace productivity through unobserved effort, where Mas and Moretti (2009) study the effect which highly productive employees have on the productivity of their coworkers.

However, peer effects are not easily identified from observational data due to endogenous group membership and correlated unobservables, as discussed by Manski (1993). A good identification strategy is therefore imperative to any empirical analysis. Similar to Dahl et al. (2014), I utilize a reform of the parental leave regulation and extensive data from the Austrian security system to implement a Regression Discontinuity Design, where the assignment variable is the birth date of a father's child with a cutoff at the implementation date of the reform. Parents whose children were born shortly after the cutoff were eligible for higher benefits than parents with children born shortly before the cutoff. I argue that the reform incentivized men to take father leave as the option of income-dependent benefits greatly reduced the opportunity costs of staying at home. To the extent that parents can not influence the exact date of birth, any variation in assignment to the new vs. old parental leave system can be considered random.

Let $Y_{re,i}$ indicate whether a father from the reform group has gone on parental leave, $X_{re,i}$ be his child's date of birth and $\bar{x}$ the cutoff date for the reform. Moreover, let $Y_{co,i}$ be the leave behavior of a coworker of the reform father. Consider polynomial expansions $r_{re,+}$, $r_{re,-}$ and corresponding coefficients δ_+ and δ_- of the assignment variable at either side of the cutoff as defined in Section 3. RDD can be implemented by a fuzzy design, in which the assignment to pre- and post-reform groups among fathers induces a change of behavior in expectation, denoted by $\delta_{+,0} - \delta_{-,0}$. The model equation for individuals in the reform group is

$$Y_{re,i} = r_{re,-}(X_{re,i} - \bar{x})'\delta_{-} + r_{re,+}(X_{re,i} - \bar{x})'\delta_{+} + \epsilon_{re,i}. \tag{1}$$

In turn, the participation of fathers induces a peer effect τ on their colleagues which gives rise to the model equation for coworkers,

$$\begin{aligned} Y_{co,i} &= \tau Y_{re,i} + \epsilon_{co,i} \\ &= \tau[r_{re,-}(X_{re,i} - \bar{x})'\delta_{-} + r_{re,+}(X_{re,i} - \bar{x})'\delta_{+} + \epsilon_{re,i}] + \epsilon_{co,i} \\ &= r_{re,-}(X_{re,i} - \bar{x})'\beta_{-} + r_{re,+}(X_{re,i} - \bar{x})'\beta_{+} + \tilde{\epsilon}_{co,i}, \end{aligned} \tag{2}$$

with $\beta_{+} = \tau\delta_{+}$ and $\beta_{-} = \tau\delta_{-}$. Any omitted effects on the participation rates are summarized in the error terms $\epsilon_{re,i}$ and $\epsilon_{co,i}$. The problem with estimating τ directly from the first line of equation 2 are several possible sources of endogeneity. For one, firm characteristics such as a particular lenient or sympathetic attitude can lead to fathers with a desire to take a leave to strategically seek employment at such firms. Leave behavior would then be correlated due to endogenous group membership. Likewise, people who work at the same company tend to be similar and therefore exhibit correlated unobservable characteristics. Finally, it is possible that coworkers have an influence on reform fathers. A coworker who is already determined to take father leave in the future might convince fathers who face the decision today.

A Regression Discontinuity Design provides a remedy to these issues of endogeneity. Note that it is not necessary to assume exogeneity of the assignment variable and the errors. Assumption 1 and 2 are sufficient to derive consistent estimates of the peer effect even if $E[X_{re,i}\epsilon_{\cdot,i}] \neq 0$. For the functional forms to be correct, however, it is necessary to assume that the only way in which the birth date of the treatment-control fathers affects the leave decision of the coworkers is through the respective leave decision of the treatment-control fathers. Analogously to Hahn et al. (2001) in Section 2, the peer effect τ can then be derived by the fraction

$$\tau = \frac{\beta_{+,0} - \beta_{-,0}}{\delta_{+,0} - \delta_{-,0}}.$$

The condition $\delta_{+}^{0} \neq \delta_{-}^{0}$ is implied by Assumption 1 and practically imposes that the reform must have at least a first-stage effect on the fathers in the treatment-control group. It is neat to notice that the first-stage effect is in essence a sharp RDD with the treatment variable as the outcome of the assignment variable. The fuzzy RDD is therefore simply the ratio of two sharp RDDs. It is also worth noting that no control variables are needed for a consistent estimate of τ. As is common practice, I will include covariates in my analysis and discuss the relevant assumptions as described above and compare the estimates derived with the inclusion of covariates to the vanilla estimator. Additionally, I will discuss the relevant covariate assumptions which have to be made in this particular application.

5 Data

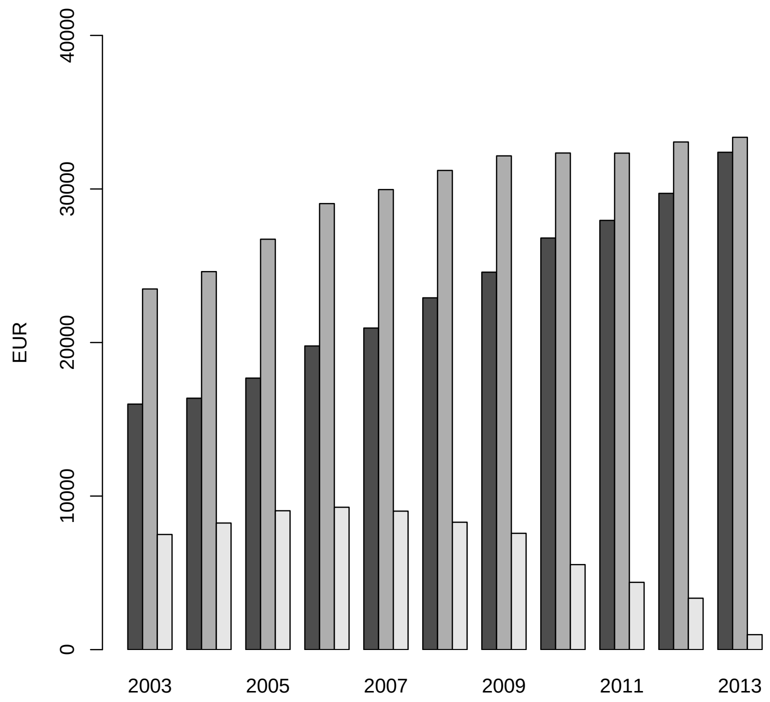
The Austrian *Arbeitsmarktdatenbank* (AMDB) is a dataset on the Austrian labor market based on administrative entries in the social security system and contains information on the universe of Austrian employees who have taken up registered employment in Austria. As long as a person has been employed once, the database can be used to keep track of subsequent entries and exits into the labor market. Aside from unemployment, the dataset records certain other labor market spells such as retirement, subsidized education programs, military service and parental leave. Additionally, the employment information of workers is linked to a personal record, which indicates demographic variables such as age, gender and nationality. Austrian firms, which have to transfer social security contributions on behalf of their employees, are linked throughout the sample via an identifying number. Therefore, information on firm size and dynamics is available too and the firm identifiers further allow to obtain coworker relationships. Due to the universal structure of the dataset (any regular employment is covered), it is well suited for investigating network effects among employees. This component of the AMDB data has been applied before by Saygin et al. (2014), who use plant closures to identify labor market shocks and examine to what extent displaced workers use their existing coworker network in their job search.

Apart from networks among employees and firms, the dataset allows for the identification of core families via the use of coinsurance data. A drawback of the coinsurance information consists in the fact that there is no mandatory requirement for spouses to be co-insured with their significant other. Moreover, the data does not keep track of the status of relationships. It is not possible to tell whether a coinsurance between spouses ends due to a divorce or due to other reasons and therefore the dataset is unsuitable to address interesting questions in the intersection of labor markets and families.

The dataset itself does not contain information on new fathers. On the other hand, births are recorded for mothers via the mandatory maternity leave policy, under which soon-to-be mothers may not be employed 8 weeks before the (expected) date of birth and 8 weeks thereafter. The data also contains rudimentary information on children born in Austria, who have to be co-insured with at least one of their parents (or other legal guardians). The dataset is imperfect for children born before 1992 and at the end of the sample - other then that, the numbers in the social security data align well with official Austrian birth statistics. Due to these informations, it is possible to link fathers to mothers and therefore obtain information on when a male worker becomes a father. To my knowledge, this thesis is the first to capitalize on the coinsurance information of children and to construct family data with this dataset, and therefore I will provide detailed information on how the father data is obtained.

Figure 1: Gross income during the year before having a child

Fathers who took parental leave (dark grey) vs. fathers who did not participate (light grey), difference (white)



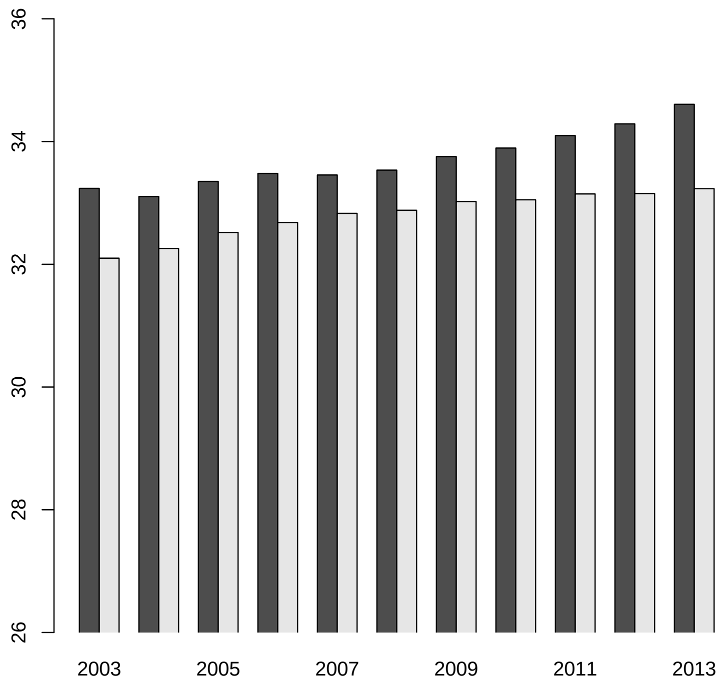
Out of approximately 2.5 mio. children in the AMDB dataset, 1.65 mio. have coinsurance spells with two parents. I subsequently extract a list of co-insurers of children in the years 2002 until 2015 and use the mother's labor market data to obtain a birth date. The event "birth" is not an official social security status in the data. However, the mandatory maternity leave of ca. 16 weeks[7] is accounted for by the data and additionally, almost all mothers have a labor market entry "Sonstige Versicherungszeiten" ("other insured time") of exactly one day after the first 8 weeks of maternity leave. I assume that this day marks the birth date. Using the maternity leave data would be sufficient to obtain an estimate of the birth date, yet using the one-day social security entry increases precision. I then look for spells of the type "parental leave" in the labor market history of the corresponding father up to three years after the birth date to determine whether the individual took a parental leave.

One drawback of using coinsurance data to obtain information on fathers is that families with only a single co-insurer are not included by construction. Unfortunately, there is little to overcome such shortcomings and results have to be taken with a grain

[7] 8 weeks before the expected date of birth and usually 8 weeks after the actual date of birth, except for multiple births and unexpected difficulties.

Figure 2: Age of fathers at the birth of their child

Fathers who took parental leave (dark grey) vs. fathers who did not participate (light grey)

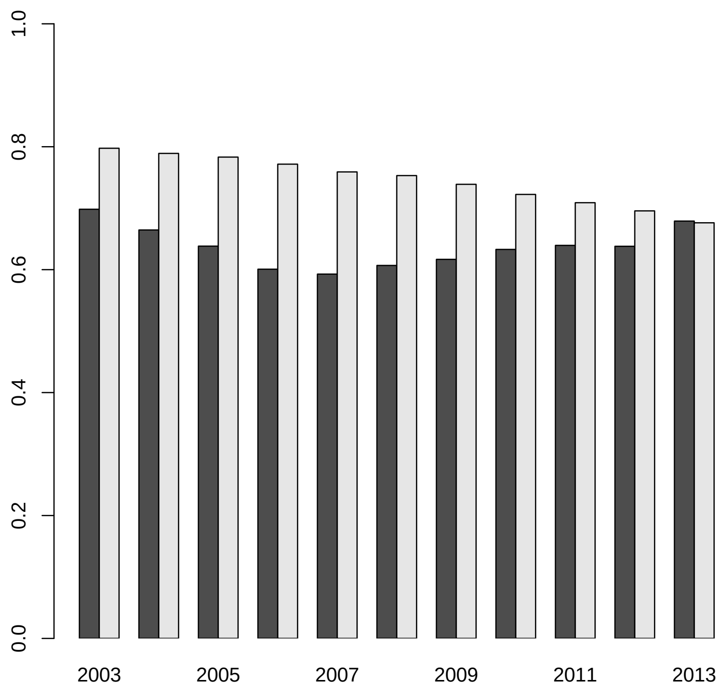


of salt as to the applicability for various types of families. An interpretation is most likely to hold for working couples where both partners are part of the labor force, whereas single-breadwinner households are omitted.

Additionally, I obtain covariates for all fathers. These include age, an indicator for Austrian citizenship and gross income in the year before the birth of a child, where the income information is based on social security contributions. As there is no information on working hours available, the data cannot be used to study labor market decisions on the intensive margin and also information on wages cannot be obtained. Nonetheless, gross income is certainly a relevant factor for household decisions - and particularly so for parental leave. Figure 1 aggregate data on the mean income of fathers who took a parental leave vs. fathers who did not. It appears that father leave used to be consumed by lower income fathers, yet the gap has closed significantly in recent years. Fathers who take leave are also on average around 1.5 years older and are less likely to have an Austrian citizenship, although the latter difference has diminished over time. Recall from Proposition 1 that covariates help to reduce asymptotic variance of the estimator if they have explanatory power for the outcome variable. From a glimpse at the aggregate data presented in Figures 1 - 3, it appears that leave

Figure 3: Fraction of fathers with Austrian nationality

Fathers who took parental leave (dark grey) vs. fathers who did not participate (light grey)



takers are on average different to non-leave takers with respect to these factors. In the case of income, a potential reason is due to the increasing opportunity costs of staying at home. Younger fathers, too, could be less involved in their respective careers and therefore more likely to take a father leave. Overall, the inclusion of these covariates could explain some variation in father leave and increase asymptotic efficiency.

5.1 Identification through a parental leave reform

As in Dahl et al. (2014), identification comes through a reform of parental leave which changed the incentive structure for novice fathers. Before 2008, parental leave benefits in Austria where granted for a maximum duration of 3 years and payed a flat benefit of around €435 per month. Note that parental leave benefits are independent of the legal requirement for employers to allow a parental leave of up to 2 years (with job security), which is the reason why most parents choose to return to work after two years. Most importantly, the benefit scheme before 2008 payed a relatively low amount for a long period of time and did not differentiate between different lengths of leave. As displayed in Figure 1, the mean income of fathers who took leave was substantially lower for fathers who took leave, which can be seen as an indication for the fact that higher earners were less incentivized to participate in father leave.

In order to increase flexibility for young families, a reform in 2008 introduced the option to take shorter leaves with the motivation of a larger monthly benefit of around €800 per month for a leave of up to 15 months (18 if both partners take a share).

A further reform in 2010 introduced the option to take an even shorter parental leave of 12 months (15 if both partners take a share) and a flat benefit of around €1000 per month. Crucially, the reform also introduced an income dependent benefit scheme which pays out 80% of previous income (up to a maximum of €2000) for up to 12 months (15 if both partners take a share). The last options have greatly changed the incentives provided to new parents, as the costs of staying at home for a short time are now heavily subsidized for middle to high earners. It is worth noting, however, that the application for parental leave has to be filed jointly, and therefore both partners have to take the same benefit option - e.g. it is not possible for a mother to take a short leave and receive income-dependent benefits whereas the father stays at home for up to three years. Nonetheless, especially families with one high-earning parent now face lower opportunity costs of splitting parental leave more evenly. Indeed, the difference in mean income between leave takers and non-participants has decreased strongly from 2009 to 2010 and thereafter, indicating a stronger participation rate among higher income fathers.

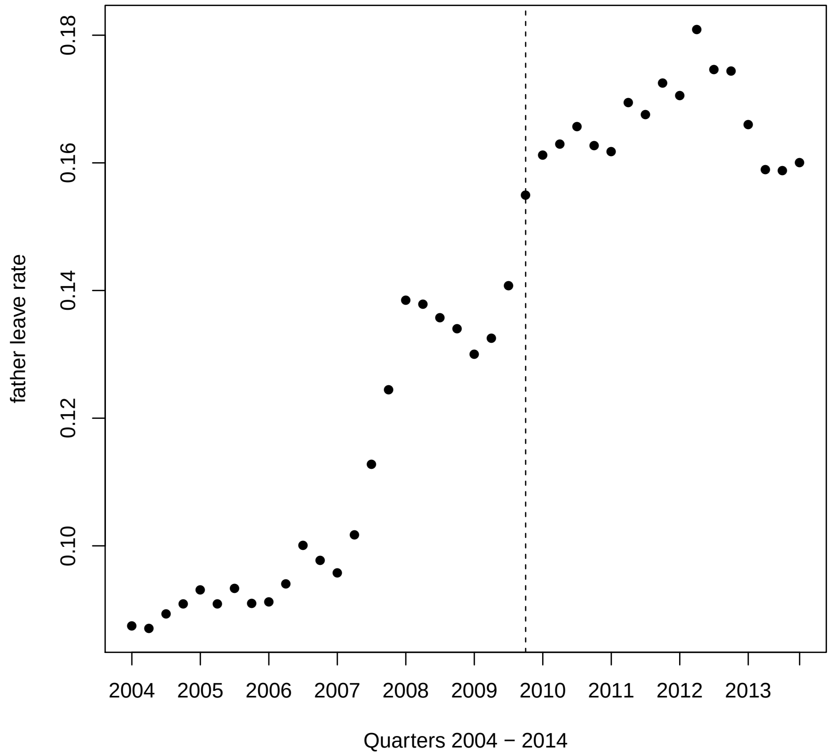
As can be seen in Figure 4, the incentives provided by the reform of 2010 translate into a strong effect on the participation rate of new fathers. In response to the reform indicated by the dashed line, the rate has gone up from around 13.5% before the reform to around 16% in the months afterwards [8]. The reform of 2008 also seems to have had an effect, although the jump is not as pronounced. The leave rate prior to 2008 is already increasing even before the reform was introduced by the end of the year.

One important component of the reforms which helps to explain the perceived non-linearity is that they were introduced with a phasing-in period. Whereas the provisions of the reforms had to be applied on births after September 30th 2007 (2009 respectively), parents of children born before January 1st 2008 (2010) had the option to file for parental leave benefits under the old provisions. This phasing-in effect seems more pronounced for the reform of 2008 and threatens identification in a RDD as the imposed non-linearity invalidates RDD designs. For this reason, I will employ the reform of 2010 as my identification strategy.

I proceed by constructing a sample of employees who became fathers in an interval

[8] The data shows higher leave rates than normally reported as I also include parental leave takers who are administratively registered as “parental leave without active employment”. I include these observations for the possibility that people register under this category for administrative reasons only. I control for people who use parental leave as a substitute to unemployment by excluding fathers in my final dataset who are not employed with the same employer before and after birth of their child.

Figure 4: Fraction of fathers who take leave - dashed line indicates cutoff date for the reform of 2010



around the reform, where fathers with a child born after the reform cutoff date have been incentivized under the new regulations whereas fathers with earlier births have not. I will refer to the group of fathers in this interval as “reform group” or “reform fathers” and their respective peers as “coworkers”. My interval of observation for the reform group contains births from June 1st, 2009 to March 31st, 2010. In order to control for the phasing-in period which lasted until December 31st, 2009, my interval includes more observations after the cutoff than before.

The group of coworkers are fathers with children born after April 1st, 2010 who are employed at the same company as a reform father. In order to guarantee that coworkers actually observe the leave behavior, I filter for fathers who do not return to the same company after taking their leave or who are employed for the same company for less than 6 months after birth of their child. In doing so, I also control for fathers who might use father leave as a substitute to unemployment while transitioning between jobs. To cleanly define a peer effect of one person on another, I further restrict the sample to companies where only one father was employed in the reform interval.

An important assumption on the behavior of reform families is that individuals with due dates close to the reform cutoff cannot influence the date of birth in order to be eligible. Such sorting behavior would likely be linked to unobservable characteristics and therefore threatens identification of the first-stage and subsequently the peer effect. Even though the possibility of timing births is limited, planned cesareans could still be scheduled accordingly. Similar to Dahl et al. (2014), I will exclude observations in a one-week window around the reform. This also mitigates potential measurement errors of the exact birth date, since this date is only derived from social security entries and not separately collected by the labor market agency. The final sample consists of 14583 observations, where 6010 observations have birth dates of the reform father before and 8573 after the reform cutoff date. Table 2 provides summary statistics for the relevant variables in the sample.

Table 2: Summary statistics of the coworker sample

	Mean	Standard deviation
Reform father leave rate	0.111	0.314
Coworker leave rate	0.137	0.344
Coworker income	31110	15821
Coworker age	33.02	6.065
Coworker Austrian	0.700	0.458

5.2 Inclusion of covariates

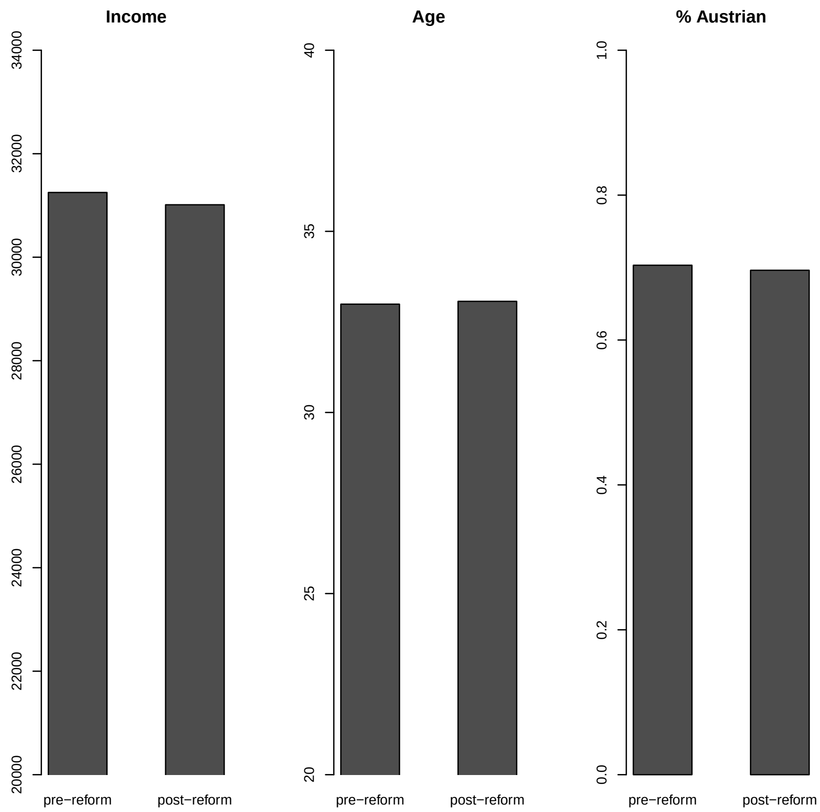
I use information on income, age and nationality of the coworkers as covariates in the Regression Discontinuity Design. As discussed before, the inclusion of covariates requires corresponding continuity assumptions to hold for the covariates, i.e. $\mu_{Z-}(x) \equiv E[Z_i(0)|X_i = x]$ and $\mu_{Z+}(x) \equiv E[Z_i(1)|X_i = x]$ are continuous (Assumption 3c). Since the covariates are obtained for the coworkers and the assignment variable x only concerns their respective reform fathers, coworkers are unaffected by the reform itself and these continuity assumptions are likely to hold. Moreover, a sufficient condition for consistency of the estimates in a RDD with covariates is that the conditional expectation of the covariates at the cutoff point is equal, i.e. $\mu_{Z+} = \mu_{Z-}$ (see Lemma 1).

Here, μ_{Z+} and μ_{Z-} correspond to the limit of the conditional expectations in the covariates for coworkers whose reform fathers had a child just before and just after the cutoff, respectively. In other words, $\mu_{Z+} - \mu_{Z-}$ is the treatment effect of the assignment variable on the covariates at the cutoff and can therefore be estimated by a sharp RD. Table 3 presents results for these estimations with respect to the three covariates, where it can be seen that neither variable had received a significant treatment effect at the cutoff date. An additional graphical representation in Figure 5 shows the sample average of each covariates for the coworkers before and after the reform date. Absence of distinct jumps between the means indicate a well behaved distribution of the covariates throughout the sample and provides further evidence that the assumption $\mu_{Z+} = \mu_{Z-}$ holds.

Table 3: Sharp RD estimates of the effect of the reform on covariates

	Estimate	Std. Err.	p	bw
Income	-2247.29	1526.49	0.141	20.659
Age	-0.267	0.503	0.596	29.286
Austrian	-0.046	0.058	0.425	13.261

Figure 5: Covariate distribution for the coworkers around the reform window



6 Regression Discontinuity results

For estimation purposes, I rely on the package "rdrobust" by Calonico et al. (2015) for the statistical software R. Throughout the analysis, I employ a data-driven bandwidth based on optimality with respect to the coverage error rate of the confidence intervals, as well as triangular kernel weights to discount observations further away from the cutoff.

A neat feature of RDDs is the possibility to portray the effect in a graphical way. Figure 6 shows the leave rate for fathers whose children were born around the cutoff date and fits linear approximations on both sides of the discontinuity. The validity of a fuzzy RD design hangs on a sufficiently strong response of the treatment to the assignment variable. Graphically, the jump between both lines at the point 0 is an estimate for this first-stage effect of the assignment variable "date of birth" on the leave behavior of the reform fathers. In the notation introduced before, the gap corresponds to estimates $\hat{\delta}_{+,0} - \hat{\delta}_{-,0}$. Table 4 reports estimates for the first-stage under a sharp regression discontinuity design. Absent of covariates, I estimate that the reform induced a 6.6 percentage point jump in the participation rate of fathers. The second line also includes the covariates income, age and nationality of the reform fathers as additional covariates for the first stage, where the estimate reduces to 4.7 percentage points.

Table 4: RDD estimates of the first-stage effect of the reform on father leave

	Estimate	Std. Err.	p	bw
First stage without covariates	0.066	0.027	0.014	23.268
First stage with covariates	0.047	0.025	0.064	25.440

Figure 7 displays the effect which the reform had on coworkers whose children were born after the reform became effective. On the abscissa are the birth dates for their respective peer father. The only way through which these dates can affect the leave decision of coworkers is through whether the peer father was assigned before or after the cutoff. The jump at 0 shows the effect which reform had on coworkers and corresponds to $\hat{\beta}^0_+ - \hat{\beta}^0_-$ of my model. I use linear estimations on each side of the cutoff to approximate the discontinuity, as Figures 6 and 7 do not show signs of non-linear behavior.

Table 5 shows estimates for the peer effect in a setting without covariates. Utilizing the optimal bandwidth, the estimate of the peer effect is an increase in the leave rate of coworkers of 28 percentage points, yet statistically insignificant. The estimate is also somewhat sensitive to the chosen bandwidth, as the second and third row indicate. I further include all covariates in a seperate estimation, to test their impact on the estimates. Recall that the estimator converges in probability to $\frac{\tau_Y-(\mu_{Z+}-\mu_{Z-})\gamma_Y}{\tau_T-(\mu_{Z+}-\mu_{Z-})\gamma_T}$ by Lemma 2. Consistency of the peer effect estimate depends therefore on the assumption $\mu_{Z+} = \mu_{Z-}$, which, as I argued above, is supported by the data. Under this condition, I do not expect the estimate to differ upon the inclusion of covariates, yet precision might increase due to a possible reduction of the asymptotic variance as shown in Section 3. Table 6 shows that the estimator for the CER optimal bandwidth shrinks to 13.4 percentage points, while standard errors increase irrespective of the chosen bandwidth. Overall, the data does not provide sufficient evidence for the existence of a peer effect and also the inclusion of covariates did not help to improve precision in this specific example.

Table 5: RDD estimate of the peer effect - without covariates

	Estimate	Std. Err.	p	bw
CER optimal bandwidth	0.284	0.466	0.542	24.718
Half bandwidth	0.222	0.458	0.627	12.359
Double bandwidth	0.316	0.405	0.435	49.436

Table 6: RDD estimate of the peer effect - with covariates

	Estimate	Std. Err.	p	bw
CER optimal bandwidth	0.134	0.772	0.862	23.558
Half bandwidth	0.257	0.733	0.726	11.779
Double bandwidth	0.232	0.466	0.619	47.116

Figure 6: Leave behavior of fathers around the cutoff date October 1st 2009

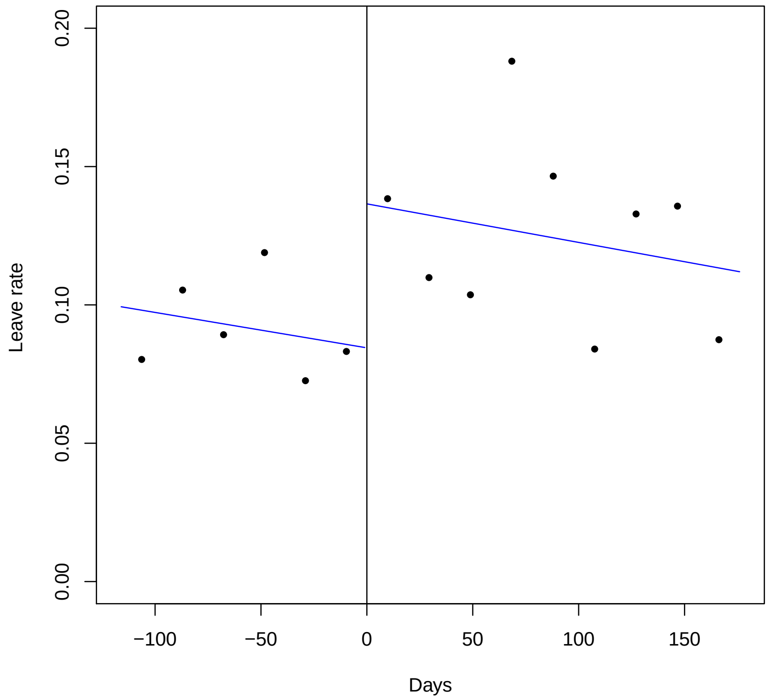
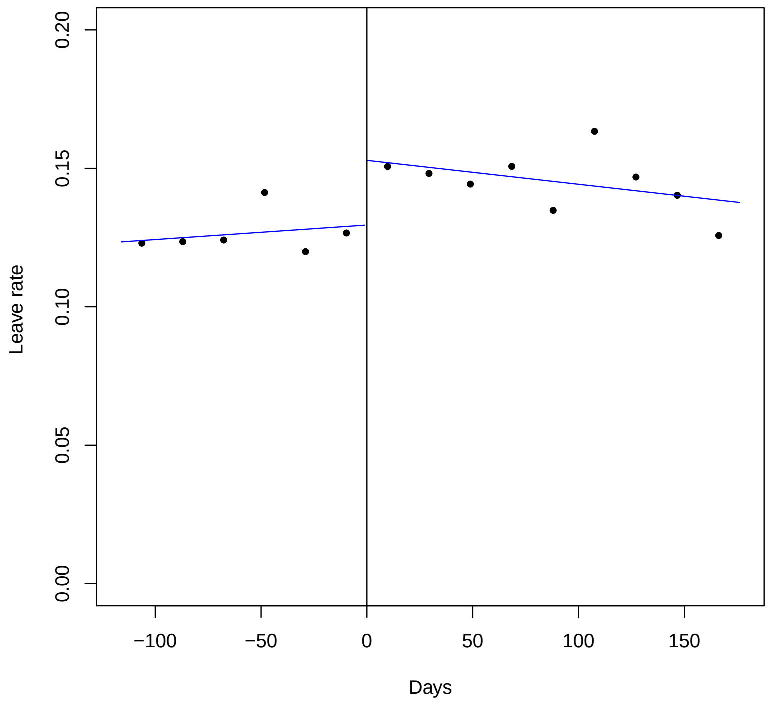


Figure 7: Leave behavior of coworkers with peer fathers around the cutoff date October 1st 2009



7 Conclusion

I implemented a fuzzy Regression Discontinuity Design on Austrian labor market data to obtain estimates of a peer effect among male coworkers, where the discontinuity arises around the implementation date of a reform of the parental leave benefit system. Motivated by potential efficiency gains, I included additional regressors in the estimation and discussed the relevant theoretical background. Among further continuity assumptions, a moment condition about covariates around the cutoff date has to hold in order to guarantee consistency of the covariate-adjusted estimator. To receive an improvement in asymptotic variance, the included covariates should have explanatory power over the outcome variable. I discussed how these assumptions hold up in my sample and find that covariates are both permissible and useful in my application. Both estimators, however, turn out to be insignificant and the research design does not support the specific peer effect in the workplace.

I further discussed the general conditions under which covariates improve estimators and show an example of harmful asymptotic behavior of the covariate-adjusted estimator in the case of a discontinuous density of the assignment variable around the cutoff. In a simulation, the covariate-adjusted estimator fares generally worse for large sampling asymmetries. A derivation of the exact variance term of the estimator under asymmetric sampling sizes would be an interesting future extension of this work.

In order to obtain a sample of fathers, I made use of coinsurance information in the AMDB dataset. In particular, I linked fathers and mothers to their respective children. An interesting future research concerning peer effects could make use of this family component of the AMDB dataset and investigate intergenerational relationships in the labor market outcomes for parents and their offspring. For example, one could ask whether children make use of their parents coworker network to find employment. As of now, the first generation whose data are contained in the dataset have yet to fully arrive on the labor market and addressing this question has to be postponed, although the depth and universal structure of the dataset makes future peer effects research promising.

References

Angrist, J. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics 114*(2), 533–575.

Angrist, J. D. and J.-S. Pischke (2010, June). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Brown, K. M. and R. A. Laschever (2012). When they're sixty-four: Peer effects and the timing of retirement. *American Economic Journal: Applied Economics 4*(3), 90–115.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2016). Regression discontinuity designs using covariates. *Working paper, University of Michigan*.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2015). rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs. *The R Journal 7*(1), 38–51.

Cattaneo, M. D., B. Frandsen, and R. Titiunik (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the us senate. *Journal of Causal Inference 3*(1), 1–24.

Dahl, G. B., K. V. Løken, and M. Mogstad (2014, July). Peer effects in program participation. *American Economic Review 104*(7), 2049–74.

Freedman, D. A. (2008). On regression adjustments to experimental data. *Advances in Applied Mathematics 40*(2), 180 – 193.

Frölich, M. (2007, September). Regression Discontinuity Design with Covariates. *Working paper, Institute for the Study of Labor (IZA)* (3024).

Hahn, J., P. Todd, and W. V. der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*(1), 201–209.

Imbens, G. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933.

Lee, D. S. (2008, February). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics 142*(2), 675–697.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Mas, A. and E. Moretti (2009, March). Peers at work. *American Economic Review 99*(1), 112–45.

Saygin, P. O., A. Weber, and M. Weynandt (2014, May). Coworkers, Networks, and Job Search Outcomes. IZA Discussion Papers 8174.