

Tutorial Article

A Tutorial on the Use of Differences-in-Differences in Management, Finance, and Accounting

Um Tutorial Sobre o Uso de Diferenças em Diferenças em Administração, Finanças e Contabilidade



Rafael Felipe Schiozer¹
Frederico Abou Mourad²
Theo Cotrim Martins¹

ABSTRACT

Context: natural experiments or quasi-experiments have become quite popular in management research. The differences-in-differences (DiD) estimator is possibly the workhorse of these techniques. **Objective:** the goal of this paper is to provide a tutorial that serves as practical guide for researchers considering using natural experiments to make causal inferences. **Methods:** we discuss the DiD advantages, concerns, and tests of validity. We also provide an application of the technique, in which we discuss the effect of government guarantees on banks' degree of risk, using the 2008 financial crisis as a natural experiment. The database used, as well as the Stata and the R scripts containing the analyses, are available as online appendices. **Conclusion:** DiD may be used to tackle endogeneity concerns when treatment assignment is random.

Keywords: differences-in-differences; natural experiments; endogeneity; causal inference.

RESUMO

Contexto: Os métodos que usam experimentos naturais ou quase-experimentos têm se tornado populares na pesquisa em administração. O estimador de diferenças em diferenças (DiD) é possivelmente o mais usado desses métodos. **Objetivo:** o propósito deste artigo é fornecer um tutorial que sirva como guia prático para pesquisadores que estejam considerando usar experimentos naturais para fazer inferência causal. **Métodos:** nós discutimos as vantagens, preocupações e testes de validação do DiD. Também fazemos uma aplicação da técnica, na qual discutimos o efeito das garantias governamentais sobre o nível de risco dos bancos, usando a crise financeira de 2008 como experimento natural. Nossa base de dados e o arquivo com os comandos do Stata e do R são fornecidos como apêndices on-line. **Conclusão:** DiD pode ser usado para contornar problemas de endogeneidade quando o tratamento é aleatório.

Palavras-chave: diferenças em diferenças; experimentos naturais; endogeneidade; inferência causal.

1. Fundação Getúlio Vargas, Escola de Administração de Empresas de São Paulo, Departamento de Contabilidade, Finanças e Controle, São Paulo, SP, Brazil.
2. Banco Central do Brasil, São Paulo, SP, Brazil.

Cite as: Schiozer, R. F., Mourad, F. A., & Martins, T. C. (2021). A tutorial on the use of differences-in-differences in management, finance, and accounting. *Revista de Administração Contemporânea*, 25(1), e200067. <https://doi.org/10.1590/1982-7849rac2021200067>

of invited reviewers until the decision:

	1	2	3	4	5	6	7	8	9
1 st round									
2 nd round									
3 rd round									

JEL Code: B23, E5, A2.

Editor-in-chief: Wesley Mendes-Da-Silva (Fundação Getúlio Vargas, EAESP, Brazil)
Associate Editor: Henrique Castro Martins (PUC Rio, IAG, Brazil)
Reviewers: Cristiano Machado Costa (Unisinos, Campus de Porto Alegre, Brazil)
Eduardo Kayo (Universidade de São Paulo, FEA, Brazil)
One of the reviewers chose not to disclose his/her identity.

Received: March 15, 2020
Last version received: June 03, 2020
Accepted: June 05, 2020

INTRODUCTION

The use of the so-called ‘natural experiments’ (or quasi-experiments) using observational data has become quite popular in several areas of quantitative research in social sciences. Although a number of approaches can explore natural experiments, the differences-in-differences (DiD, or diff-in-diff) estimator is possibly the workhorse of these techniques (Atanasov & Black, 2016). DiD has been extensively used in a number of papers in several areas of management including finance (e.g., Jayaratne & Strahan, 1996), international business (e.g., Mithani, 2017), and accounting (e.g., Chen, Hung, & Wang, 2018).

The main goal of this paper is to discuss the proper use of DiD, allowing the researcher to make causal inferences from observational data. We choose not to focus on the derivation of the statistical properties of the estimators and the corresponding proofs. Instead, we emphasize the practical aspects and the intuition behind the use of DiD. We give particular emphasis to the discussion of the necessary assumptions for causal inference from the DiD model and describe variations of the model that allow the researcher to address both theoretical and empirical concerns. In doing so, we draw from a number of previous papers on the subject, from lecture notes of several researchers, and from our own experience as end-users of econometric tools. Therefore, the contribution of this paper is to serve as a practical guide for researchers considering the use of natural experiments (and the DiD technique in particular) to make causal inference from observational data.

We start by discussing the typical endogeneity problems that arise with traditional ordinary least squares (OLS) regressions and why these problems generally impede causal inference. We then discuss how — and under what assumptions — the DiD technique can be used to address these concerns and then describe typical falsification tests that may provide suggestive evidence about the validity of the assumptions of the model.

We continue with an application of the DiD technique using observational data. Namely, we exemplify the use of DiD by investigating whether implicit and explicit governmental guarantees provided to banks affect their degree of risk, using the 2008 financial crisis as a quasi-natural experiment. We briefly describe the theories supporting our hypothesis, the data, and the results of several variations of DiD specifications applied to our problem. We draw particular attention to the discussion of the necessary assumptions for using DiD in our case since, more often than not, researchers have to convince themselves, referees, and readers that their particular research setup satisfies the assumptions of the DiD model. Therefore, we present

a set of typical tests that allow the researcher to provide suggestive evidence in favor of (or against) the use of DiD.

We also discuss the limitations of the DiD technique and mention a series of possible extensions and alternative estimation methods that aim to address these limitations. We conclude with our view about the use of differences-in-differences in quantitative research in management, finance, and accounting.

PROBLEMS IN TRADITIONAL OLS REGRESSIONS

In this section, we discuss the three main sources of bias that may be present in traditional OLS estimations: omitted variable, simultaneity, and measurement error. Researchers generally refer to these issues as ‘endogeneity problems’ (Roberts & Whited, 2013). We give more emphasis on the first two (omitted variable and simultaneity) because DiD is more adequate to address these types of problems rather than measurement error in empirical applications.

The traditional OLS regression model is written as:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u \quad (1)$$

As researchers, we are generally interested in making causal inferences by providing answers to questions such as ‘what is the causal effect of a change in x_1 on y , holding all else constant?’ If we depart from a random sample of the population of interest, and assume no perfect collinearity among the independent variables, the best linear unbiased estimate of this effect will be the regression coefficient x_1 (i.e., β_1) under the following assumptions (Roberts & Whited, 2013):

$$(a) E[u] = 0;$$

Given that the regression has an intercept β_0 , the validity of this assumption is straightforward. If we suppose that $E[u] = k \neq 0$, then the intercept would become $\beta_0 + k$, and $E[u]$ would turn out to be equal to zero. Therefore, we generally assume that this assumption is met.

$$(b) E[u|x_1, x_2, \dots, x_k] = E[u];$$

This assumption is generally referred to as the conditional mean independence (CMI), and means that the average value of the error term u (the ‘unexplained’ part of y) does not depend on the values of the regressors x . CMI implies that the error term u is uncorrelated with each of the x 's.¹

The violation of the second condition is indeed one of the major issues that hinder causal inference. Broadly speaking, the term ‘endogeneity’ refers to the existence of

non-zero correlation between the error term and one or more regressors. We note that the error term u is not observable, and therefore this assumption cannot be tested². It is not very hard to think of reasons why the error term might be correlated with the regressors in typical OLS regressions in management research. In what follows, we present three main sources of problems that may cause the violation of the second condition (or, as generally said by researchers, create ‘endogeneity problems’).

Omitted variable

The omitted variable bias (OVB) is possibly the most common problem of using linear regressions. The basic concern is that the error term u contains a variable that is not in the regression model (call it z , for example) and is correlated with one or more of the regressors included in the model. To illustrate why OVB causes the violation of the second condition, suppose that we are interested in uncovering the causal effect of x on y . The true model should have two covariates, x and z , i.e.:

$$y = \beta_0 + \beta_1x + \beta_2z + v \tag{2}$$

but we omit z , and estimate a regression with only one covariate, x :

$$y = \beta_0 + \beta_1x + u \tag{3}$$

We skip the math, but it can be shown that the estimated coefficient $\widehat{\beta}_1$ will be:

$$\widehat{\beta}_1 = \beta_1 + \delta_{x,z}\beta_2 \tag{4}$$

where $\delta_{x,z} = Cov[x,z]/Var[x]$. The term $\delta_{x,z}\beta_2$ is the bias of this regression (i.e., the difference between the estimated coefficient $\widehat{\beta}_1$ and the true coefficient β_1). The magnitude of the bias depends on two main features of the omitted variable z : (a) how important it is in explaining y (β_2); and (b) how correlated it is with the variable of interest x . Therefore, even if we are not particularly interested in the effect of z on y , omitting z from the regression equation causes bias to the estimated β_1 . We emphasize that there is no bias if z is uncorrelated with x , because in this particular case δ_{xz} is equal to zero. This fact is important for the DiD estimation that we cover in section ‘The use of diff-in-diff to address endogeneity concerns’.

Gormley and Matsa (2014) show that, in the case of a simple regression with just one omitted variable, it is possible to determine at least the sign of the bias if one has theoretical arguments for the signs of β_2 and δ_{xz} . However, they also show that, if the regression involves more than one variable

of interest or more than one omitted variable, determining the sign of the bias becomes virtually impossible.

But how can a researcher address OVB? If the omitted variable z is observable, the solution is easy: one should simply add it to the estimation equation. However, in many practical problems in management, we are concerned with unobservable omitted variables, such as managerial talent, the risk aversion of shareholders and so on. As we show ahead, DiD involves finding a source of variation in the variable of interest x that is uncorrelated with potentially unobserved omitted variables.

Simultaneity

Simultaneity bias occurs whenever any of the regressors (x_1, x_2 , etc.) can be affected by changes in the dependent variable y . It is not hard to think of practical situations in management research in which the dependent and independent variables are simultaneously determined. A classic example in finance is the relationship between firm indebtedness and dividends. Arguably, firm leverage affects managers’ decisions about dividends, but dividends also affect leverage decisions. In fact, many of the firms’ outcomes such as sales, trade credit, level of inventories, governance choices, leverage, payout, corporate social relationship decisions and many others stem (at least in part) from decisions made by firm managers, and therefore these variables are arguably simultaneously determined.

To exemplify why simultaneity causes bias, take a model in which changes in x affect y , and changes in y also affect x , according to the equations:

$$y = \beta_0 + \beta_1x + u \tag{5}$$

$$x = \gamma_0 + \gamma_1y + v \tag{6}$$

If we estimate the first regression, the estimate of β_1 will be biased, because x is correlated with the error term. To see why this happens, just replace y from equation (5) in equation (6) to obtain:

$$x = \gamma_0 + \gamma_1(\beta_0 + \beta_1x + u) + v \tag{7}$$

resulting in

$$x = \left(\frac{\gamma_0 + \gamma_1\beta_0 + v}{1 - \gamma_1\beta_1}\right) + \left(\frac{\gamma_1}{1 - \gamma_1\beta_1}\right)u \tag{8}$$

Equation (8) shows that x ‘contains’ the error term u , meaning that x is correlated with the error term whenever γ_1 is not zero (i.e., whenever y affects x), thus violating the second condition for having an unbiased β_1 .

In a multivariate OLS regression setting, even if only one of the regressors suffers from simultaneity (i.e., is affected

by y), still all coefficients will be biased. For example, suppose that z affects x_2 (but not x_1) in the following regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \quad (9)$$

It can be shown that x_2 will be correlated with u , and all the coefficients (including $\widehat{\beta}_1$) will be biased (Wooldridge, 2010, for derivation).

Finally, simultaneity bias also arises when one regressor affects another. For example, suppose that x_1 affects x_2 in the regression of equation (9), i.e.:

$$x_2 = \delta_0 + \delta_1 x_1 + w \quad (10)$$

By replacing x_2 from equation (10) in equation (9), we obtain:

$$y = (\beta_0 + \beta_2 \delta_0) + (\beta_1 + \beta_2 \delta_1) x_1 + (u + \beta_2 w) \quad (11)$$

Equation (11) shows that the estimated coefficient of x_1 will be $\widehat{\beta}_1 = (\beta_1 + \beta_2 \delta_1)$. Therefore, instead of capturing the true effect β_1 , one would be wrongly attributing to x_1 part of the effect that x_2 has on y . This is what Angrist and Pischke (2008) call the ‘bad controls’ problem. We return to this point when we discuss the DiD method in section ‘The Use of Diff-in-Diff to Address Endogeneity Concerns’.

Measurement error

Another type of endogeneity problem arises when variables are measured with error, i.e., when the variable of interest is measured imprecisely. This problem may occur when the researcher cannot observe the perfect measure of some variable (e.g., the marginal cost of production) or when the variable of interest is not perfectly quantifiable (e.g., managerial talent or firm corporate governance), for which the researcher has to rely on proxies. It can be shown that, if the measurement error is correlated with the error term, then the estimated coefficients will be biased (Angrist & Pischke, 2008).

Because differences-in-differences cannot usually address measurement error problems, we do not develop the math behind it. Instead, we direct the reader to Angrist and Pischke (2008), Roberts and Whited (2013), and Gormley and Matsa (2014). For a more technical review on panel data models with measurement errors, we recommend Meijer, Spierdijk and Wansbeek (2017).

THE USE OF DIFF-IN-DIFF TO ADDRESS ENDOGENEITY CONCERNS

Following Roberts and Whited (2013), we start the discussion about how to solve endogeneity concerns with a practical advice to researchers. Before moving on to the use of any approach that aims to address endogeneity, a researcher should carefully think about the research question at hand and identify what are the main endogeneity concerns. In other words, one should ask: “what are the endogenous variables and why are they endogenous?” Generally, endogeneity concerns arise from the formulation of alternative hypotheses (also called ‘confounding effects’). In other words, researchers should always ask themselves questions like “is there anything else, other than my hypothesis, that would produce similar estimation results?”

Researchers often refer to the term ‘identification strategy’ as a methodological procedure that allows them to disentangle between two or more competing hypotheses. Any identification strategy departs from articulated assumptions about the method chosen to answer a particular research question. These assumptions generally include theoretical, economic, and managerial reasoning that mirrors statistical assumptions of the econometric/statistical models used.

Using natural experiments: why randomness is key

The use of natural experiments in management studies is inspired in controlled experiments (or controlled trials) that are often used in other areas of science. For example, to check if a certain drug reduces cholesterol, researchers can give a drug to a random sample of patients (treated patients), and placebo pills to others (untreated patients, or control group), and check how the cholesterol levels of patients evolve over time. The difference in the average change in cholesterol between the two groups of patients is referred to as the ‘average treatment effect’ (ATE). In regression format, the ATE is called the differences-in-differences estimator, as we explain in the next section. Provided that the assignment of the actual drug versus placebo among patients is random, the ATE is an estimate of the causal effect of the drug on cholesterol levels on average.

Unfortunately, it is very rare that researchers in management, finance, and accounting are able to apply controlled experiments, because that would involve imposing a ‘treatment’ (which could mean a change in management style, an increase in leverage, or whatever other source of change that the researcher is interested in) to a randomly assigned group of firms. Researchers normally cannot do such interventions in the real world. That is why researchers often resort to ‘natural experiments’ to uncover the ATE.

A natural experiment is a sharp change in one or more variables of interest that occurs for exogenous reasons, either by natural causes (e.g., natural disasters), or by some kind of human action, such as changes in regulation, economic policy, and political changes (generally referred to as ‘quasi-natural experiments’). We discuss under what circumstances the use of laws and regulation changes as quasi-natural experiments are appropriate for causal inference in section ‘Limitations and possible extensions of the diff-in-diff model’.

The key assumption for an experiment to allow causal inference is that the treatment assignment is random or at least ‘as good as random,’ in the sense that any other variable that is important to determine the outcome variable is uncorrelated with the treatment assignment. For example, in the cholesterol experiment, if the assignment method systematically assigns younger patients to the treated group compared to the control group, one potential concern is that the different resulting cholesterol levels might be due to different age profiles. Using statistical terms, the treatment is correlated with a variable (patient age) that supposedly affects the cholesterol level.

Single differences and the differences-in-differences estimator

To explain how the DiD estimator works in regression format, we start by describing the single difference regression at the cross-section level, then we move to time-difference regressions, and finally combine the two to build the DiD regression. To facilitate our argumentation, we define the observational unit to be a firm, but it can be generalized to any other type of observational unit (individuals, households, etc.). Hereafter, we call the natural experiment ‘the shock’ for simplicity.

The simplest approach to estimate the effect of treatment is the cross-sectional difference. It involves comparing the post-treatment values of the dependent variable y between treated and untreated firms. In regression format, this idea is written as:

$$y_i = \beta_0 + \beta_1 d_i + u_i \quad (12)$$

where d_i is a dummy variable that indicates whether firm i is treated by the experiment ($d_i = 1$) or not ($d_i = 0$). If the treatment is random, then d_i is uncorrelated with the error term and β_1 is an unbiased estimate of the average treatment effect (ATE).

This approach is useful when the researcher does not have data on the values of y previously to treatment. If the researcher is able to observe the values of the outcome variable y for several periods after treatment (i.e., there is a panel of firm-level observations), [Bertrand, Duflo and](#)

[Mullainathan \(2004\)](#) recommend using the average y_i over the several periods to account for the dependence across observations of the same firm³. In this case, the regression equation would be written as $\bar{y}_i = \beta_0 + \beta_1 d_i + u_i$, where \bar{y}_i is the average y of firm i along the periods after treatment.

For most natural experiments, it is virtually impossible to verify the assumption that assignment is truly random. Therefore, depending on the empirical problem at hand, one possible critique that can appear from the use of cross-sectional difference is that the average y of treated and untreated firms were different ex-ante (i.e., before the treatment). We show ahead how the DiD estimator can help us address this type of critique.

Now let us assume that we can observe y for the firms one period before the shock and one period after the shock, but we conjecture that all the firms are supposedly affected by the shock (for example, when a law or a change in regulation affects the outcome y of all the firms in the sample). A second approach to uncover the ATE involves comparing post-treatment values of y to pre-treatment values of y for all firms. This is called the time-difference approach, which in regression format looks as follows:

$$y_{i,t} = \beta_0 + \beta_1 p_t + u_{i,t} \quad (13)$$

where p_t takes value 1 for the observations in the post-shock period, and 0 for the observations in the pre-shock period. In this case, the ATE is given by β_1 , provided that no other event that affects y occurred between the pre-shock and the post-shock periods, i.e., there is no omitted variable, correlated to p_t , that affects y . The same regression can be used if one has more than one period of observation before and after the shock. β_1 will be the ATE if we add the assumption that the effect of treatment is constant along the post-shock periods⁴.

We discuss how to capture an effect that occurs gradually over time in our application in section ‘Application: bank risk and bailout probability’.

The differences-in-differences model combines the cross-sectional and the time-series differences models into a single model. The intuition behind DiD is to compute the difference of the change in y pre- versus post-treatment for the treated group and the change in y pre- versus post-treatment for the control group. To implement the DiD model, we need a panel of treated and untreated firms, with observations before the shock and after it. The DiD regression equation looks as follows:

$$y_{i,t} = \beta_0 + \beta_1 p_t + \beta_2 d_i + \beta_3 p_t \times d_i + u_{i,t} \quad (14)$$

where d_i and p_i are defined as in equations (12) and (13), respectively. The interpretation of the coefficients in equation (14) is as follows: β_1 captures the average change in y from the pre- to post-shock periods for the untreated group. Under the assumption that the treatment is random, β_1 is also the hypothetical change from the pre- to post-shock periods for the treated firms had they not been treated. Therefore, the key assumption here is that y of treated firms would ‘behave as’ y of untreated firms if treatment did not happen (i.e., the expected y of treated firms would change by as much as the observed average change in y for untreated firms).

The coefficient β_2 captures the pre-shock difference in y between treated and untreated firms. Under the prior assumption, if treatment did not occur, this difference would have remained the same in the post-shock periods.

Finally, our main coefficient of interest is β_3 , which captures the effect of the shock, i.e., the ATE. β_3 is the average differential change in y from the pre- to post-treatment period for the treatment group relative to the change in y for the untreated group. β_3 is referred to as the ‘DiD coefficient.’ We reinforce that the main underlying assumption for interpreting β_3 as the causal effect of the treatment on y is that the expected outcome y of treated firms would change by as much as the observed change in y for untreated firms if the shock were absent. We can think of this hypothetical change absent treatment as an unobserved counterfactual. However, this assumption cannot be tested, because our counterfactual is not observable. Therefore,

robustness checks that provide suggestive evidence about the soundness of this assumption are helpful for the assessment of whether it is valid or not. We present a series of typical robustness checks ahead.

Practical aspects of DiD

Figure 1A illustrates the coefficients of the DiD regression for a stylized case where the pre- and post-shock average y of treated and untreated observations are constant (i.e., there is no trend in the data). This type of figure is helpful in presenting the data in a straightforward and visually intuitive manner. Although the visual inspection of a figure like this does not provide any statistically valid tool per se, understanding how y behaves over time for treated and control firms may provide several insights about the mechanisms causing the changes in the outcome variable. Figure 1B, instead, illustrates a case in which pre- and post-shock y increase at a constant trend. In this case, the coefficients of equation (14) capture the differences between averages, i.e., β_1 captures the difference between average pre- and average post-shock y for untreated observations, and β_2 picks up the average pre-shock difference between treated and untreated observations. Finally, β_3 captures the difference between observed average post-shock y and the average unobserved counterfactual y after the shock (i.e., the hypothetical value of y of treated observations absent the shock). Most papers that use DiD provide similar graphs, and they provide an intuitive depiction of the shock effect.

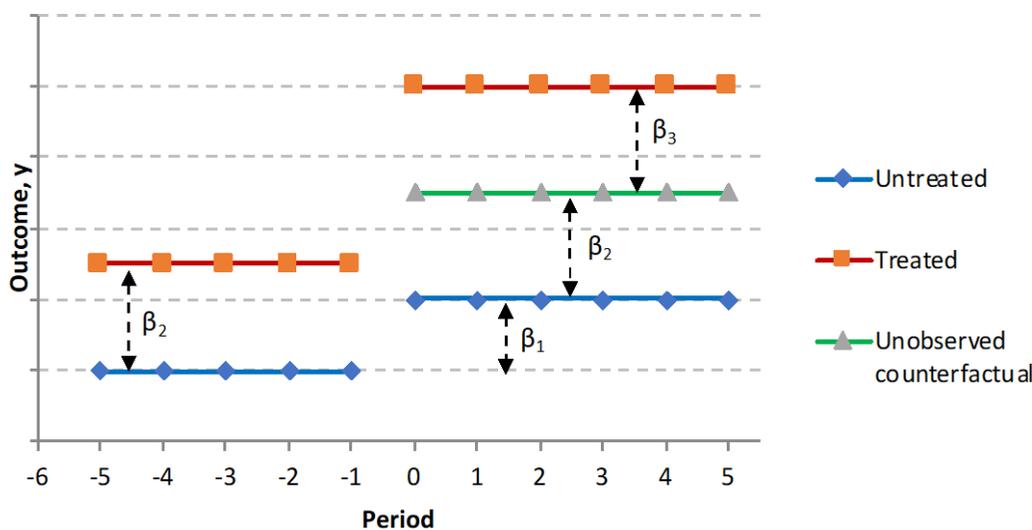


Figure 1A. Visual interpretation of the coefficients of the DiD model — without trends.

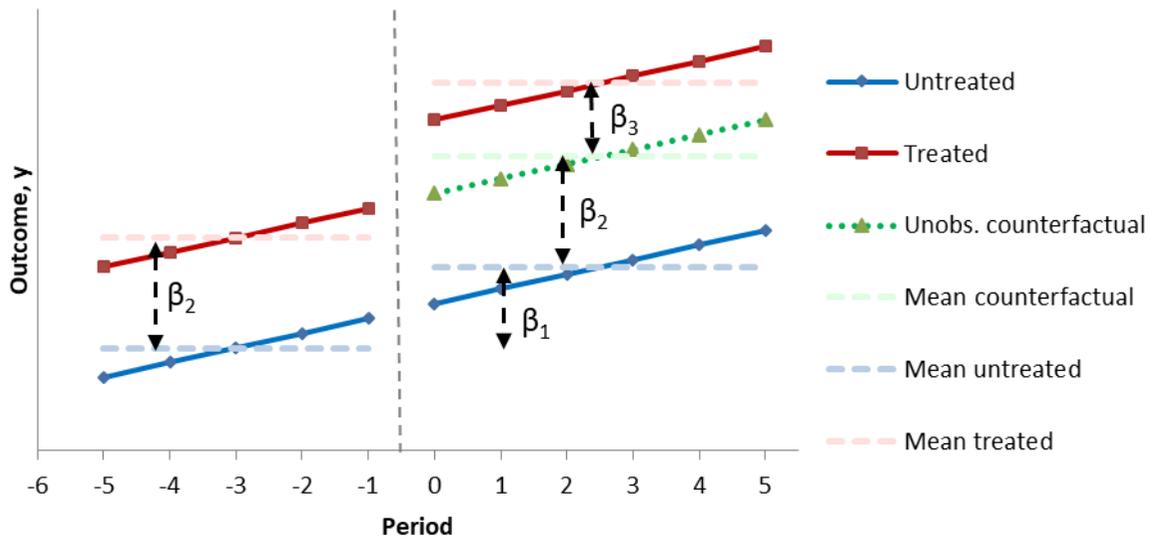


Figure 1B. Visual interpretation of the coefficients of the DiD model — with trends.

Inspecting the pre-shock averages for the treated and control group also helps us identify whether the outcome y of treated and control firms are similar prior to the shock. In the ideal framework, a truly random treatment would suggest that treated and control firms should be indistinguishable prior to the shock. However, even if treated and control firms present different averages for y prior to the shock, but their trends are parallel, then the required statistical assumption for causal inference with the DiD model (no correlation between the $p_i \times d_i$ variable and the error term) may still be valid. That is why the main assumption of the DiD model is sometimes referred to as the ‘parallel trends assumption.’ In visual terms, this assumption means that the pre-shock trends of y of treated and control firms are parallel (which can be visually inspected and statistically tested), and that they would remain parallel if treated firms had not been treated (which cannot be verified).

However, if treated and control firms present non-parallel trends prior to the shock, then the main assumption of the DiD model is most likely invalid, which undermines any attempt to draw causal inference from the model. This is why we advise building and presenting the graph of average y over time for treated and untreated firms. Since the graphs do not provide us with any formal test, it is generally enough to observe reasonably parallel trends, and address any specific concern with formal robustness tests.

Another practical matter in using DiD is defining how many pre- and post-shock periods to use. Although there is no theoretical background to give a definitive answer, we give some practical advice to support this decision. First, we discourage using too few periods (one or two) prior to the shock, as it does not allow checking for parallel trends.

On the other hand, if the pre- and/or post-shock period is too long, the analysis can be subject to the occurrence of other events (shocks) that also affect y , and these events may confound the analysis. In some cases, however, firms may be expected to respond to the shock gradually over time, and therefore one should use enough observations to pick up as much of the effect as possible.

Second, if the data used is not annual (say, monthly or quarterly) and the outcome variable presents a seasonal pattern, one should avoid potential seasonality problems by using the same calendar months or quarters in the pre- and post-shock periods. Finally, if the date of the natural experiment cannot be precisely defined (for example, a law may pass in a given date, but be effective only later on), it may be reasonable to exclude one or more periods from the analysis to avoid wrongly assigning a period to the pre- or post-shock.

DiD regressions easily allow for the inclusion of control variables. Adding control variables that explain y may be useful to increase the fit (i.e., the R-squared) of the regression and therefore may improve the precision of the estimates of β_1 , β_2 , and β_3 by reducing their standard errors. On the other hand, one should avoid including control variables that are themselves affected by the treatment, to avoid the ‘bad controls problem’ described in section ‘Problems in Traditional OLS Regressions’. Therefore, if the treatment is truly random, one should either not use controls or use pre-shock values as controls. It can be a good idea to do both.

One alternative to the use of control variables is to use firm fixed effects, aiming to capture all time-invariant firm

features (both observable and unobservable) that affect y . In this case, the coefficient β_2 is not identified because the treatment dummy is perfectly correlated with the firm fixed effects. In addition, one can also fully saturate the regression by using time fixed effects, aimed at capturing all the time variation in y that is common to all the firms (these may include economic cycles and other macroeconomic changes, as well as changes in law and regulations that affect all firms alike). Time fixed effects are particularly useful when we observe time trends in y , as depicted in figure 1B. In this case, β_1 is not identified because it is collinear with time fixed effects. This version of the DiD model is called generalized differences-in-differences:

$$y_{i,t} = \beta_0 + \beta_3 p_t \times d_i + \mu_i + \delta_t + u_{i,t} \quad (15)$$

where μ_i and δ_t are respectively firm and time fixed effects.

Finally, there are several robustness checks or falsification tests that can be applied to provide suggestive evidence about the validity of the assumptions or deal with possible confounding effects. The choice of appropriate robustness checks depends mainly on the specific research question at hand and the particular concern that the researcher aims to address. We discuss some of these robustness checks in the context of the applied problem that we develop in the next section.

APPLICATION: BANK RISK AND BAILOUT PROBABILITY

In this section, we apply the DiD method to test the hypothesis that governmental guarantees to banks affect their risk. We use the global financial crisis of 2008 as a quasi-natural experiment.

A brief theoretical framework

The expected effect of governmental protection on bank risk is ambiguous. The charter value theory (Keeley, 1990) states that implicit guarantees cause protected banks to take less risk than unprotected banks, because this implicit protection allows them to fund at abnormally low costs, which is a source of value (charter value) that banks do not want to put at risk. On the other hand, the moral hazard hypothesis (Flannery, 1998) states that depositors of protected banks have less incentives to monitor the risk of the bank, thereby leading to an increase in risk-taking by these banks. The empirical evidence on the matter is also mixed. Some papers favor the charter value hypothesis (e.g., Forssbäck & Shehzad, 2015), whereas others, such as Dam and Koetter (2012), favor the moral hazard hypothesis.

During financial crises, governments typically adopt a series of measures to stabilize the financial system. Some of these measures (for example, ample liquidity provision and expansion of the safety net) occur at the macro level and benefit large and small banks alike, whereas other measures are aimed directly at avoiding the failure of systemically important financial institutions (also called too-big-to-fail). Recent evidence (Ueda & Di Mauro, 2013; Oliveira, Schiozer, & Barros, 2015) shows that systemically important banks benefit from a perception of protection by depositors and investors in general. This perception becomes more pronounced during financial crises, because financial authorities are more likely to bailout these large financial institutions if needed during times of turmoil than during normal times. Therefore, systemically important banks enjoy implicit guarantees that other smaller banks do not, and the perception of these guarantees increases during times of crisis, even in countries whose financial systems were not directly affected by the crisis.

The shock: the 2008 financial crisis

Although the liquidity crisis started in the US in the second half of 2007 (Acharya & Mora, 2015), Allen and Carletti (2010) argue that it was not until the failure of Lehman Brothers, in September of 2008, that the crisis actually spread internationally. In the US, government backing to systemically important financial institutions was key in restoring their deposits. Oliveira et al. (2015) show that the G-7 action plan launched in early October of 2008, which included a pledge to save systemically important banks, caused the perception of an increased bailout probability of large banks that went beyond the borders of the G-7 countries, including countries whose banking systems were not directly affected by the crisis.

We use the financial crisis as a quasi-natural experiment. As the banks in the US and other economies were at the origin of the crisis, we use a sample of banks from countries whose financial systems were not directly affected by the crisis. Based on the prior literature, we argue that the crisis changed the perception of governmental protection to large banks (the treatment group), whereas the perception of a bailout probability did not change much for less protected banks (the control group). Indeed, this perception derived from the observation that, in the US and other developed economies, many large banks were saved, whereas smaller, less protected banks were allowed to fail (Acharya & Mora, 2015).

Identification strategy: diff-in-diff model

To test whether governmental protection affects bank risk, we estimate the following DiD model:

$$\ln(Z_Score_{i,t}) = \beta_0 + \beta_1 Crisis_t + \beta_2 Protected_i + \beta_3 (Crisis \times Protected)_{i,t} + u_{i,t} \quad (16)$$

where the subscripts i and t refer to bank and year, respectively. The dependent variable is the Z_Score , a measure of bank risk traditionally used in the banking literature (Soedarmono, Machrouhb, & Tarazi, 2013; Schiozer, Mourad, & Vilarins, 2018). The Z_Score is defined as follows:

$$Z_Score_{i,t} = \frac{ROA_{i,t} + Capital\ ratio_{i,t}}{\sigma(ROA_{i,t})} \quad (17)$$

where ROA is the return on average assets, $Capital\ ratio$ is the bank's regulatory capital ratio, and $\sigma(ROA)$ is the standard deviation of ROA computed over the past three years of data. The intuition of the Z_Score as a proxy of risk is that it measures the bank's distance to default in terms of the standard deviation of ROA. Therefore, the smaller the Z_Score , the riskier the bank. We use the natural logarithm of the Z_Score instead of its raw value to obtain the ATE in relative (percentage) terms.

In our main regressions, we use data from 2005 to 2010. The first three years of data (2005, 2006, and 2007) are the pre-crisis period ($Crisis = 0$), and the last three years (2008, 2009, and 2010) are defined as the crisis period ($Crisis = 1$). To identify the banks that would be expected to receive governmental protection in case of need, we use data from one of the major rating agencies, which provides scores based on its assessment about the probability of the bank receiving external support. We assign $Protected = 1$ to the banks to which the agency assigns any probability of external support and $Protected = 0$ to banks without any expected external support⁵. Our main coefficient of interest is β_3 : a positive (negative) β_3 favors the charter value (moral hazard) hypothesis.

One possible concern of our definition of the treatment and control groups in investigating bank risk is that protected and unprotected banks might be differently exposed to 'toxic assets' (i.e., subprime assets and other types of assets that ultimately gave rise to the financial crisis). Therefore, the crisis could be considered endogenous, in the sense that the pre-existing exposure to toxic assets could be considered an omitted variable that explains bank risk (our outcome variable) and is correlated to our main regressor (protected bank). To avoid this endogeneity concern, we use a sample of banks located in countries of the OECD (Organization for Economic Cooperation and Development) whose banks were not directly affected by the crisis, according to the definition of Laeven and Valencia (2012). As such, we mitigate the concern that changes in the Z_Score might be stemming from pre-existing exposure to crisis-related assets.

We also run a series of alternative specifications including control variables at the bank-level and macroeconomic controls at the country-level that have been previously shown to affect bank risk (Gropp, Hakenes, & Schnabel, 2011; Schiozer et al., 2018), as well as country, bank, and time fixed effects. The bank-level controls are bank size (measured by the natural logarithm of assets) and liquidity (measured by the ratio between liquid assets and short-term liabilities). The values are set at pre-crisis dates to avoid the 'bad controls' problem described in section 'Problems in Traditional OLS Regressions'. Macroeconomic controls follow Schiozer et al. (2018) and include the concentration of the banking market (measured by the Herfindahl-Hirschman index), the countries' GDP per capita, the first two lags of GDP growth, and the ratio of credit to the private sector to GDP.

We use financial statements and regulatory data at the bank level from one of the world's major provider of information on banks. Data from this bureau has been previously used in several different studies in related subjects (Gropp, Hakenes, & Schnabel, 2011; Drechsler, Drechsel, Marques-Ibanez & Schnabl, 2016; Schiozer et al., 2018). We collect data on the variables of interest for banks⁶ from OECD member countries that were not directly affected by the global financial crisis according to Laeven and Valencia (2012)⁷. After excluding missing data, we end up with an unbalanced panel of 3,324 observations from 900 banks. This dataset, and the Stata and R commands used in the analysis, are provided as online appendices to this paper. Our dataset also includes additional data (from 2011 onwards) that we use in the robustness checks described further in the paper.

Descriptive statistics and parallel trends

Table 1 presents the descriptive statistics for the variables at the bank level, splitting between protected (treated) and unprotected (control) banks, both before and after the shock. Prior to the crisis, the average $\ln(Z_Score)$ of unprotected banks is slightly smaller than that of protected banks. However, the difference between the average $\ln(Z_Score)$ of the two groups is not statistically significant, meaning that protected and unprotected banks had approximately the same level of risk prior to the crisis on average. Both protected and unprotected banks increase their risk (i.e., decrease their Z_Score) from the pre-crisis to the crisis period, but the average $\ln(Z_Score)$ of protected banks decreases more than that of unprotected banks, consistent with the moral hazard hypothesis.

Table 1. Descriptive statistics.

Variables	Unprotected				Protected				Wilcoxon rank-sum test
	Mean	Median	Std. dev.	Obs.	Mean	Median	Std. dev.	Obs.	
Panel A - Before the crisis									
ln (Z-score)	4.54	4.67	1.22	490	4.62	4.74	1.2	318	-1.23
Liquidity	17.81	7.84	37.19	484	15.66	8.3	15.58	316	-2.09***
assets	13.4	3.23	60.44	490	71.79	28.09	174.42	318	-17.24***
Panel B - During the crisis									
ln (Z-score)	4.17	4.15	1.44	2,102	3.94	3.89	1.04	414	2.83***
Liquidity	26.69	22.36	37.94	2,085	17.02	10.84	22.61	410	-12.17***
Assets	9.14	1.94	78.78	2,102	96.14	40.52	215.66	414	-27.30***

Note. Crisis is considered between 2008 and 2010; Protected (unprotected) banks are observations considered as treated (non-treated) due to the existence (nonexistence) of external support as assessed by a major credit rating agency. ln (Z-score) refers to the natural logarithm of the Z-score winsorized at 1% (see section ‘Application: Bank Risk and Bailout Probability’ for further details); Liquidity is the ratio between liquid assets and short-term liabilities (in percentage points); Assets refers to the total assets (in billions of USD). Data is restricted to pre-crisis (2005-2007) and during the crisis (2008-2010) periods. The commands for the comparison of means between protected and unprotected banks are provided in the online appendix.

The statistics described in Table 1 also show that, prior to the crisis, unprotected banks had a slightly larger ratio of liquid assets to short-term liabilities compared to protected banks on average, but the difference is not statistically significant. Both groups increase their liquidity ratios during the crisis. Finally, protected banks are significantly larger than unprotected banks, consistent with the idea that the protected banks are typically large, systemically important financial institutions.

Figure 2 depicts the average *Z_Scores* of protected (treated) and unprotected (control) banks from 2005 to 2010. In the pre-crisis years, the *Z_Scores* of both groups are roughly constant (or, if anything, the average *Z_Score* is slightly increasing for treated banks and slightly decreasing for banks of the control group from 2005 to 2007). For the time being, we consider that the pre-crisis trends of the groups are roughly parallel, but we return to this issue in our robustness checks in the next section.

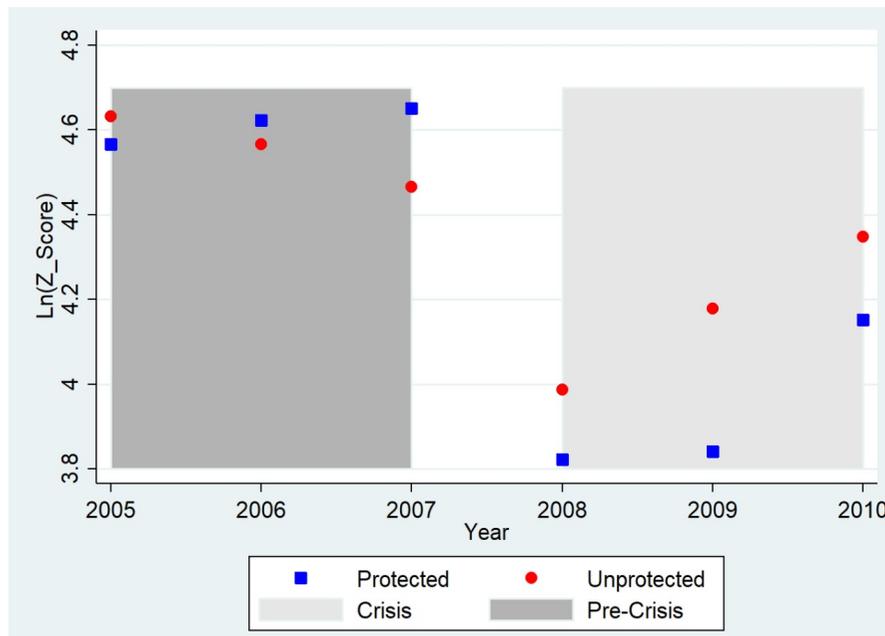


Figure 2. Parallel trends.

The natural logarithm of the Z-score is our measure of bank risk (see section ‘Application: Bank Risk and Bailout Probability’ for further details); Crisis is considered to be in 2008; Protected (unprotected) banks are observations considered as treated (non-treated) due to the existence (nonexistence) of external support as assessed by a major credit rating agency. Data here is restricted to pre-crisis (2005-2007) and during the crisis (2008-2010) periods.

From 2007 to 2008 (the first year of crisis), we observe a dramatic decrease in $\ln(Z_Score)$ (i.e., an increase in bank risk), which is more pronounced for protected banks than for banks in the control group. In 2009 and 2010, both groups gradually increase the $\ln(Z_Score)$, but the measure remains lower than pre-crisis figures for both groups, and the average value for the treated banks remains lower than for the control group throughout the three years of crisis.

DIFF-IN-DIFF REGRESSION RESULTS

Table 2 presents the estimation of equation (16) with several variations. We start with the traditional DiD equation — i.e., exactly the specification described in equation (16) — in column 1. The estimated coefficient β_1 (for *Crisis*) indicates that the Z_Score is reduced by approximately 36.8% on average for the control group from the pre-crisis to the crisis period (statistically significant at the 1% level). The coefficient of *Protected* (β_2) is not statistically significant at the usual levels, indicating that the average pre-crisis Z_Scores of the treated and control groups are not statistically significantly different. Finally, the main coefficient of interest, β_3 , indicates that the average Z_Score of the protected banks group decreases by approximately 31% more than that of their unprotected counterparts does. This finding is the central result of our analysis. This result supports the moral hazard hypothesis (Flannery, 1998), meaning that the increased perception of governmental guarantees has a positive and economically significant effect on bank risk.

In the specifications that follow, we add other features to the basic DiD regression. In column 2 of Table 2, we add the bank-level controls *Size and Liquidity*. To avoid the bad controls problem, we use pre-shock (2007) values. The purpose of including controls is twofold. First, it can help falsify the main assumption. One alternative story to the results in column 1 could be that investors tend to shift resources to larger, more diversified banks or to banks with greater liquidity in troubled times, and this excess inflow of funds could lead banks to invest in riskier assets. Since bank size and liquidity are correlated to our *Protected* dummy, we would have an omitted variable problem, and would be wrongly attributing the increase in risk of protected banks to the implicit guarantees, and not to their size or liquidity. The second reason to include controls is simply to improve precision and regression fit. The results in column 2 show that bank *size* and *liquidity* do not materially affect the Z_Score , as their coefficients are statistically insignificant. More importantly, the coefficients β_1 and β_3 remain rather stable in comparison to the estimates reported in column 1, and β_2 remains statistically insignificant, implying that our previous inferences stand up to the inclusion of these control variables.

In column 3 of Table 2, we add control variables at the country-level to address the possible concern that country features might be important in determining bank risk. However, we lose more than 700 observations in this regression due to missing data. Nevertheless, we find that none of the country-level controls is statistically significant. Importantly, our main inferences about β_1 and β_3 remain qualitatively unchanged, although the magnitude of the coefficients vary in relation to the results in columns 1 and 2 (which may be due to the missing observations), whereas β_2 remains statistically insignificant. In column 4 of Table 2, we replace country-level controls with country fixed effects. Country fixed effects may capture other time-invariant country features that are not captured by macroeconomic controls, such as market microstructure, quality of regulation, enforcement of the law, and other institutional aspects. Yet, the signs and statistical significance of β_1 , β_2 , and β_3 remain practically unchanged relative to the previous specification, although the magnitude of the coefficients varies.

In the specification reported in column 5 of Table 2, we add bank fixed effects. Because bank fixed effects are perfectly collinear with the *Protected* dummy, country fixed effects and the control variables, the coefficients of these variables are no longer identified. The coefficient β_1 remains negative and statistically significant, and its magnitude is larger than in the previous regressions. The DiD coefficient β_3 remains negative, but it is smaller than in the previous regressions and statistically insignificant at the usual levels. The remarkable increase in R-squared compared to the other specifications shows that bank fixed effects explain great part of the variation of the dependent variable (in other words, the risk of a given bank is relatively stable over time). Finally, we report in column 6 the estimation results of a generalized DiD, i.e., a model that includes both time and bank fixed effects. In this model, only the main coefficient of interest, β_3 , is identified. The DiD coefficient β_3 is negative, but statistically insignificant, and its magnitude is similar to the one obtained in column 5.

Indeed, it is rather common that coefficients of interest lose statistical significance as one saturates the regression specification with more fixed effects. As the bank fixed effects are correlated by construction with the *Protected* dummy and time fixed effects are correlated with the *Crisis* dummy, they will capture part of the variation that is captured by the dummies in the specifications without fixed effects. Many times, examining the change in the magnitude and stability of the coefficient of interest is as important as its statistical significance.

Table 2. Differences-in-differences regressions (2005–2010).

Variables	Ln (Z-score)					
	(1)	(2)	(3)	(4)	(5)	(6)
Crisis	-0.368*** (0.0852)	-0.345*** (0.0852)	-0.293** (0.115)	-0.427*** (0.0841)	-0.586*** (0.0997)	
Protected	0.0781 (0.123)	-0.0242 (0.141)	0.196 (0.157)	0.0564 (0.144)		
Crisis × Protected	-0.310** (0.129)	-0.358*** (0.126)	-0.466*** (0.155)	-0.280** (0.126)	-0.0840 (0.152)	-0.0775 (0.153)
Size2007		0.0415 (0.0261)	0.0717** (0.0354)	0.0358 (0.0311)		
Liquidity2007		-0.148 (0.140)	0.360*** (0.119)	0.104 (0.0973)		
HHI2007			-1.038 (1.274)			
GDP per capita2007			0.00698 (0.0106)			
GDP growth2006			-4.898 (10.35)			
GDP growth2005			1.244 (3.811)			
Credit / GDP2007			-0.0555 (0.531)			
Observations	3,324	3,296	2,593	3,296	3,293	3,293
R-squared	0.022	0.028	0.039	0.065	0.764	0.773
Bank FE	NO	NO	NO	NO	YES	YES
Country FE	NO	NO	NO	YES	---	---
Year FE	NO	NO	NO	NO	NO	YES
# of clusters (banks)	900	891	705	891	869	869

Note. The dependent variable in all models is our measure of bank risk, the natural logarithm of the Z-score as defined in section ‘Application: Bank Risk and Bailout Probability’; Crisis is equal to 1 in 2008, 2009, and 2010. Protected (unprotected) banks are considered treated (non-treated) due to the existence (nonexistence) of external support as assessed by a major credit rating agency; Size is the natural logarithm of Assets (in billions of USD); Liquidity is the ratio between liquid assets and short-term liabilities; HHI is the country’s banking system Herfindahl index; GDP per capita is the yearly country GDP over its population (in current USD); GDP growth is the countries’ GDP annual growth; Credit (private credit) is the domestic credit to private sector by banks as a percentage of GDP. Robust standard errors clustered at the bank level are shown in parentheses. The symbols *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

The fact that the magnitude estimated coefficient β_3 is smaller in the regressions with bank fixed effects (columns 5 and 6) than in the previous regressions weakens the interpretation in favor of the moral hazard hypothesis. One possibility is that, as the sample is not balanced along the pre- and post-shock periods (i.e., there are banks entering and leaving the sample along the years), the results obtained in specifications (1) to (4) are possibly capturing changes in the composition of the sample. If unprotected banks that enter the sample in the post-shock period are less risky

than protected banks on average, this leads to a decrease in the magnitude of β_3 with the introduction of fixed effects. We address this possible sample phenomenon in the next section.

Further refinements, falsification, and robustness tests

In this section, we exemplify a few typical falsification tests and robustness checks that can be used in papers using

DiD. As we mentioned before, the choice of robustness checks fundamentally depends on the practical concern facing the researcher. As such, we give examples of robustness checks applied to the concerns that may stem from the inferences derived from the results of Table 2, as well as challenge some of our identifying assumptions.

One typical type of robustness check that is used in DiD papers is checking for treatment reversals (see, for example, Oliveira et al., 2015). The basic idea is that, if the treatment is subsequently reversed, we should expect the opposite effect of what was observed with treatment. If one observes the opposite effect when the treatment is reversed, the treatment effect becomes more credible, in the sense that it is harder to attribute the original results to an alternative story.

We conjecture that the acute phase of the crisis lasted up to 2010. By 2011, the turmoil had passed, and therefore, the perception of an increased guarantee to protected banks by investors diminished⁸. Therefore, we consider that a ‘reversal’ of treatment occurred in 2011. To check for the effect of the reversal of treatment, we use data from 2008 to 2013 and run the following regression:

$$\ln(Z_Score_{i,t}) = \beta_0 + \beta_1 PostCrisis_t + \beta_2 Protected_i + \beta_3 (PostCrisis \times Protected)_{i,t} + u_{i,t} \quad (18)$$

where *PostCrisis* is equal to 0 in 2008, 2009, and 2010, and equal to 1 in 2011, 2012, and 2013. The specifications are analogous to the ones reported in Table 2. Under this definition, the coefficient β_1 compares the post-crisis to the crisis $\ln(Z_Score)$, β_2 reflects the average difference between the two groups during the crisis period, and β_3 measures the treatment reversal ATE. These estimations are reported in Table 3.

In the estimation of column 1, the coefficient β_1 shows that the average *Z_Score* of unprotected banks increases by approximately 135% from the crisis to the post-crisis period⁹, statistically significant at the 1% level. β_2 shows that, during the crisis, the average *Z_Score* of protected banks was approximately 23% smaller than that of unprotected banks (statistically significant at the 5% level). Our main coefficient of interest, β_3 , shows that the absolute difference in the average *Z_Scores* of protected and unprotected banks is increased by approximately 18.9% from the crisis to the post-crisis period (statistically significant at the 10% level). This is the estimated ATE of the treatment reversal, and is consistent with the moral hazard hypothesis.

As we add controls and country fixed effects in the specifications reported in columns 2 to 4, the signs and significance of β_1 and β_2 remain, although their magnitudes

change somewhat across different specifications, the same applying to β_2 on specification 5, which includes bank fixed effects. The magnitude of β_3 is remarkably stable, ranging from 0.184 to 0.211 across all specifications reported in Table 3, although it loses statistical significance in specifications 3 and 6 (their p-values are approximately 12% and 10% in these regressions, respectively). The fact that protected banks reduce their risks in comparison to unprotected banks when the bailout perception of protected banks decreases reinforces the previous evidence in favor of the moral hazard hypothesis.

Another common type of robustness check is the non-parametric version of DiD (Gonçalves, Schiozer, & Sheng, 2018), described in equation (19):

$$y_{i,t} = \beta_0 + \beta_1 \varphi_t + \beta_2 d_i + \sum_t \omega_t (d_i \times \varphi_t) + u_{i,t} \quad (19)$$

where φ_t is equal to 1 if *year* = *t*, and zero otherwise. In other words, we have a series of interactions of the treatment dummy with each year of data, except one of them, generally the first year of data. Therefore, the coefficients ω_t capture the difference in *y* between treated and untreated firms in each year, in relation to the difference that existed in the first year. The non-parametric regression is particularly useful when the timing of treatment is not a ‘sharp’ event, such as our case, because it does not rely on a subjective judgment about the timing of treatment. It also helps in verifying parallel pre-trends: if any of the ω prior to the shock are statistically different from 0, then the pre-shock trends of treated and untreated firms are not parallel. Finally, this regression helps in identifying effects that occur gradually over time, or fade away over time, in which case we would observe a gradual variation in ω over time. One can also easily add controls and fixed effects to the non-parametric DiD regression.

The main disadvantage of the non-parametric version of DiD is that coefficients may contain a great amount of noise, because of the excess of parameters to be estimated. Therefore, one should not care too much about statistically insignificant coefficients that may appear. Many times, the results of the non-parametric DiD are presented in a graph rather than a table (e.g., Ponticelli & Alencar, 2016; Oliveira et al., 2015), as it allows an easy visualization of the coefficients.

Table 3. Differences-in-differences estimator (reversal: 2008–2013).

Variables	Ln (Z-score)					
	(1)	(2)	(3)	(4)	(5)	(6)
PostCrisis	0.855*** (0.0477)	0.861*** (0.0478)	0.956*** (0.0557)	0.872*** (0.0477)	0.858*** (0.0522)	
Protected	-0.232** (0.0914)	-0.691*** (0.112)	-0.413*** (0.153)	-0.405*** (0.129)		
PostCrisis × Protected	0.189* (0.103)	0.211** (0.103)	0.194 (0.126)	0.211** (0.103)	0.184* (0.113)	0.186 (0.113)
Size ₂₀₁₀		0.134*** (0.0248)	0.146*** (0.0342)	0.108*** (0.0288)		
Liquidity ₂₀₁₀		-0.314** (0.143)	0.0925 (0.105)	-0.0355 (0.0981)		
HHI ₂₀₁₀			-0.814 (0.524)			
GDP per capita ₂₀₁₀			0.00967 (0.0170)			
GDP growth ₂₀₀₉			1.647 (2.978)			
GDP growth ₂₀₀₈			10.18** (4.271)			
Credit / GDP ₂₀₁₀			1.020 (0.833)			
Observations	5,000	4,943	3,913	4,943	4,977	4,977
R-squared	0.098	0.127	0.164	0.157	0.687	0.693
Bank FE	NO	NO	NO	NO	YES	YES
Country FE	NO	NO	NO	YES	---	---
Year FE	NO	NO	NO	NO	NO	YES
# of clusters (banks)	900	891	705	891	869	869

Note. The dependent variable in all models is our measure of bank risk, the natural logarithm of the Z-score as defined in section ‘Application: Bank Risk and Bailout Probability’. PostCrisis is equal to 1 in 2011, 2012, and 2013. Protected (unprotected) banks are considered treated (non-treated) due to the existence (nonexistence) of external support as assessed by a major credit rating agency; Size is the natural logarithm of Assets (in billions of USD); Liquidity is the ratio between liquid assets and short-term liabilities; HHI is the country’ banking system Herfindahl index; GDP per capita is the yearly country GDP over its population (in current USD); GDP growth is the countries’ GDP annual growth; Credit (private credit) is the domestic credit to private sector by banks as a percentage of GDP. Robust standard errors clustered at the bank level are shown in parentheses. The symbols *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

We estimate the non-parametric DiD using our main dataset (from 2005 to 2010) without any controls. We use the first year of data (2005) as our reference. The vector of coefficients ω_p , along with their 95% confidence intervals, is depicted in Figure 3. The figure shows that the hypothesis of parallel trends in the pre-shock period cannot be rejected, because the confidence intervals for ω_{2006} and ω_{2007} cross the horizontal axis, meaning that these coefficients are

not statistically different from zero. The coefficients ω_{2008} and ω_{2009} are negative and statistically different from zero, confirming the hypothesis that protected banks increase their risk more than unprotected banks, and the effect seems to fade away in 2010 as ω_{2010} is also negative, but its magnitude is smaller than in the previous two years, and it is not statistically significant.

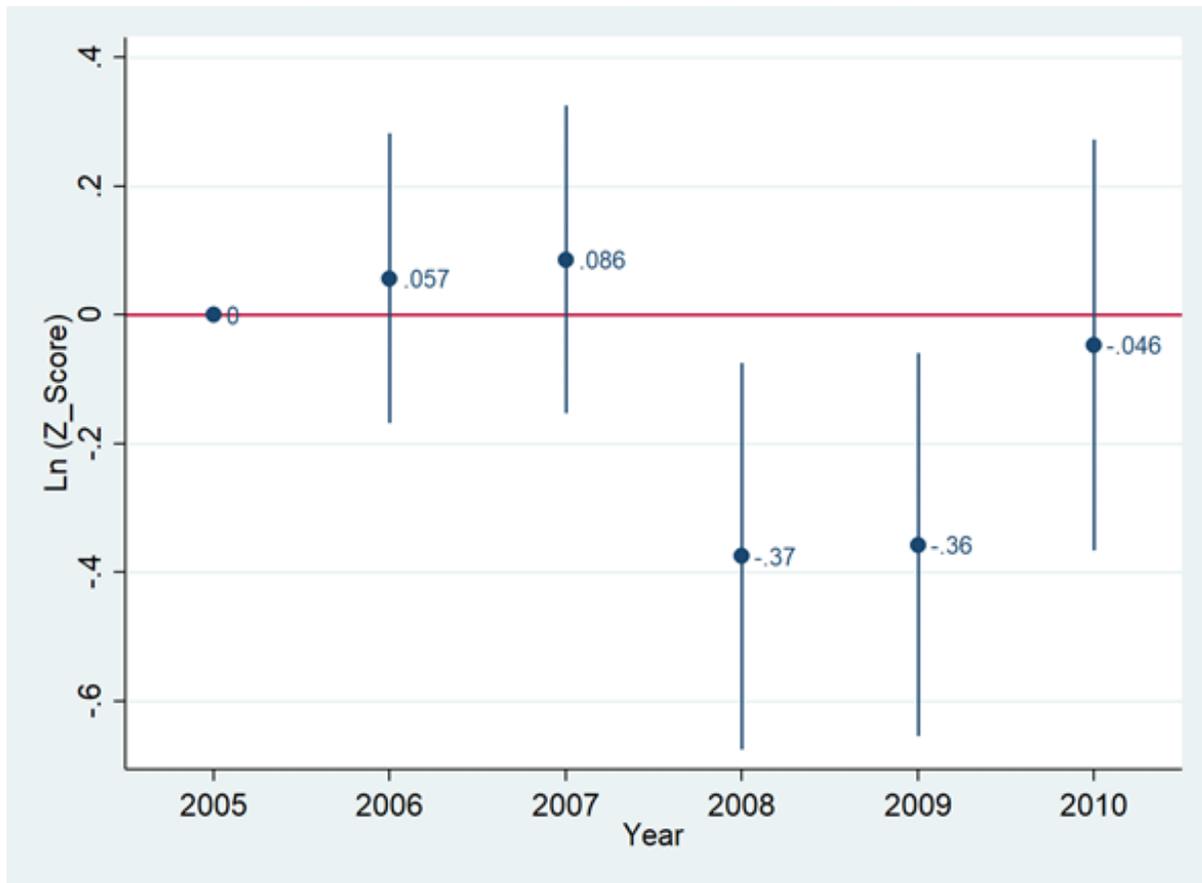


Figure 3. Treatment effect over time (without controls).

The natural logarithm of the Z-score is our measure of bank risk (see section ‘Application: Bank Risk and Bailout Probability’ for details); Crisis is between 2008 and 2010; Protected (unprotected) banks are observations considered as treated (non-treated) due to the existence (nonexistence) of external support as assessed by a major credit rating agency.

In many papers (Khwaja & Mian, 2008), the DiD model can also be expressed in differences, as in equation (20):

$$\Delta \bar{y}_i = \delta_0 + \delta_1 Treatment_i + v \quad (20)$$

In this case, $\Delta \bar{y}_i$ is the difference between the average post-shock y_i and the average pre-shock y_i . The treatment value may be either a dummy or a continuous variable. If it is a dummy variable, the coefficient δ_1 has the same interpretation of β_3 in equation (14) (i.e., the ATE). If one uses a continuous treatment variable, δ_1 captures the expected change in y caused by a one-unit change in the *Treatment* variable. A continuous treatment variable is adequate when the intensity of the shock varies across firms. See Schiozer and Oliveira (2016) for an example of continuous treatment with Brazilian data.

We do not report the results of the estimation of equation (20) to our case for the sake of brevity, but the

estimation commands are found in the Stata and R scripts of the online appendices. In these estimations, the *Treatment* variable is the *Protected* dummy.

Finally, there is a number of other robustness checks that can be performed to mitigate specific concerns about our previous inferences. We do not report all these tests to save space, but we invite the reader to go through the online appendix that contains the Stata and R scripts to perform these tests. Namely, we: (a) restrict the sample to banks that have observations both before and after the crisis to check whether the reduction in the magnitude in specifications 5 and 6 of Table 2 is due to the entry of low-risk banks along the sample period. We warn that this procedure may lead to survivorship bias, and therefore one must be very careful in interpreting the results; (b) re-run the non-parametric DiD using 2007 (the pre-shock year) as the reference, instead of 2005 (the first year of data).

Other typical robustness checks for DiD include: (a) using a placebo timing of treatment; (b) if there is any

theory supporting that some treated individuals should be more sensitive to treatment than others, one can add triple differences. See [Campello, Ladika, & Matta \(2019\)](#) for an excellent example of triple differences.

LIMITATIONS AND POSSIBLE EXTENSIONS OF THE DIFF-IN-DIFF MODEL

In this section, we discuss alternative manners of running a DiD regression, the main limitations of the DiD model, and possible ways to address these limitations through extensions of the model or alternative techniques.

Possibly the greatest source of limitation to the DiD technique is related to the non-verifiability of its assumptions. In most applications in management, treatment assignment is not completely random or exogenous. Particularly when changes in law, regulations, or policy action are used as experiments, the researcher must keep in mind that these human-driven changes happen for a reason, with an objective in perspective. For example, a government may pass a law that affects small firms (treated group) exactly to improve the business environment for these firms. As such, one should be careful in claiming that an event is exogenous and treatment is 'as good as random.'

We repeat that researchers must think of possible confounding effects (i.e., omitted variables that are correlated to treatment) that could yield the same results and address these confounding effects. In other words, researchers should always ask themselves if there is any *ex-ante* difference (observable or unobservable) between the treated and untreated groups, and if this difference could make each group react differently to the shock at hand. We direct the reader to [Atanasov and Black \(2016\)](#) for an excellent discussion about the validity of the shocks used in papers published in major accounting, economics, finance, law, and management journals.

Returning to our application on governmental guarantees and bank risk, one possible concern is that protected banks are larger than unprotected banks on average. This could raise a suspicion that the risk of protected banks might have increased more than that of unprotected banks not because of increased moral hazard, but because treated banks are inherently different from unprotected banks.

For example, they may have more sophisticated risk management techniques that allow them to take more risks than smaller unprotected banks in turbulent times. As a first attempt to address this concern, one could include in the regression the interaction of bank size with the crisis dummy, and check if this addition changes the main inferences drawn in the traditional DiD. In addition, one could look at more qualitative data on risk management

(e.g., experience of the Chief Risk Officer) or compensation incentives of managers, if such type of data is available. For an excellent paper using such type of information, we direct the reader to [Fahlenbrach and Stulz \(2011\)](#).

Another possible concern is that smaller banks do not have as much access to sources of liquidity as larger banks do and, as a result, smaller banks decide to reduce their risk by holding on to liquid assets. To disentangle between the moral hazard and the liquidity stories, it could be helpful to inspect banks' liquidity sources and types of investors as in [Oliveira et al. \(2015\)](#).

A second manner of addressing such type of concern is by using matching techniques. The basic idea of matching is to identify one or more firms from the untreated group that are similar to each treated firm prior to the shock, and then apply the DiD method between treated and their counterpart 'matched' firms. There are several possible matching techniques (i.e., manners of identifying 'similar' firms among the treated and untreated group), and we refer the reader to [Imbens \(2015\)](#) for a detailed review of matching techniques. In our example of section 'Diff-in-diff Regression Results', one could match treated and untreated banks based on features such as country of origin, size, and liquidity to identify treated and untreated banks that are similar along these dimensions. Still, if treated and untreated banks are different in unobservable dimensions, and these dimensions are important in determining bank risk, then the omitted variable problem persists. We refer the reader to [Almeida, Campello, Laranjeira and Weisbenner \(2011\)](#) and [Sampaio, Gallucci, Silva and Schiozer \(2020\)](#) for applications using matching with US and Brazilian data, respectively.

Other extensions of the DiD model can be applied to very specific situations, such as regression discontinuity design (RDD) and selection models. These techniques fall outside the scope of this paper. For examples of applications of RDD with Brazilian data in finance and management, we direct the reader to [Martins and Novaes \(2012\)](#) and [Arvate, Galilea and Todescat \(2018\)](#), respectively.

Finally, while we have presented how to deal with a single treatment event, the DiD estimator can be used in the context of multiple events with minor adjustments. For examples of using DiD with multiple events, we recommend the papers by [Jayaratne and Strahan \(1996\)](#), [Bertrand et al \(2004\)](#), and [Gormley and Matsa \(2011\)](#).

CONCLUSION

The goal of this paper is to provide researchers with a guide for using the differences-in-differences (DiD) estimator to make causal inference in management, finance,

and accounting. First, we discuss the typical endogeneity problems that generally impede standard ordinary least squares (OLS) from making causal inference.

We then present the underpinnings of the DiD estimator, and why it is considered the workhorse technique for causal inference from observational data. Nevertheless, we also show DiD's potential flaws.

The paper provides a practical example of DiD to help the reader apply the technique using real data. The example

investigates whether government guarantees can affect banks' risk taking, using the 2008 financial crisis as a quasi-natural experiment. We present variations of the traditional DiD specification, followed by a set of robustness and falsification tests. Finally, we present the reader to possible extensions of the model.

The readers can use the Stata and R scripts provided in the online appendices to replicate our model and tests, and adapt them in their own research.

REFERENCES

- Acharya, V. & Mora, N., (2015). A crisis of banks as liquidity providers. *Journal of Finance*, 1, 1-43. <https://doi.org/10.1111/jofi.12182>
- Allen, F. & Carletti, E. (2010). An overview of the crisis: causes, consequences, and solutions. *International Review of Finance*, 10(1), 1-26. <https://doi.org/10.1111/j.1468-2443.2009.01103.x>
- Almeida, H., Campello, M., Laranjeira, B., & Weisbenner, S. (2011). Corporate debt maturity and the real effects of the 2007 credit crisis. *Critical Financial Review*, 1, 3-58. <http://doi.org/10.3386/w14990>
- Angrist, J. & Pischke, S. (2008). Mostly harmless econometrics: An empiricists' companion. Princeton: Princeton University Press.
- Arvate, P. R., Galilea, G. W. & Todescat, I. (2018). The queen bee: a myth? The effect of top-level female leadership on subordinate females. *Leadership Quarterly*, 29(5), 533-548. <https://doi.org/10.1016/j.leaqua.2018.03.002>
- Atanasov, V. & Black, B. (2016). Shock-based causal inference in corporate finance and accounting research. *Critical Finance Review*, 5(2), 207-304. <https://doi.org/10.1561/104.00000036>
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences - in-differences estimates? *Quarterly Journal of Economics*, 119(1), 249-275. <https://doi.org/10.1162/003355304772839588>
- Campello, M. Ladika, T., & Matta, R. (2019). Renegotiation frictions and financial distress resolution: Evidence from CDS spreads. *Review of Finance*, 29(3), 513-556. <https://doi.org/10.1093/rof/rfy021>
- Chen, Y., Hung, M., & Wang, Y. (2018). The effect of mandatory CSR disclosure on firm profitability and social externalities: Evidence from China. *Journal of Accounting and Economics*, 65(1), 169-190. <https://doi.org/10.1016/j.jacceco.2017.11.009>
- Dam, L., & Koetter, M. (2012). Bank bailouts and moral hazard: empirical evidence from Germany. *Review of Financial Studies*, 25(8), 2343-2380. <https://doi.org/10.1093/rfs/hhs056>
- Drechsler, I., Drechsel, T., Marques-Ibanez, D. Schnabl, P. (2016) Who Borrows from the Lender of Last Resort?, *Journal of Finance*, 71(5), 1933-1974. <https://doi.org/10.1111/jofi.12421>
- Fahlenbrach, R., & Stulz, R. M. (2011). Bank CEO incentives and the credit crisis. *Journal of Financial Economics*, 99(1), 11-26. <http://doi.org/10.1016/j.jfineco.2010.08.010>
- Flannery, M. J. (1998). Using market information in prudential bank supervision: A review of the US empirical evidence. *Journal of Money, Credit and Banking*, 30(3), 273-305. <https://doi.org/10.2307/2601102>
- Forssbäck, J., & Shehzad, C. T. (2015). The conditional effects of market power on bank risk – cross-country evidence. *Review of Finance*, 19(5), 1997-2038. <https://doi.org/10.1093/rof/rfu044>
- Gonçalves, A. B., Schiozer, R. F. & Sheng, H. H. (2018). Trade credit and product market power during a financial crisis. *Journal of Corporate Finance*, 49, 308-323. <https://doi.org/10.1016/j.jcorpfin.2018.01.009>
- Gormley, T., & Matsa, D. (2014) Common errors: How to (and not to) control for unobserved heterogeneity. *Review of Financial Studies*, 27(2), 617-61. <https://doi.org/10.1093/rfs/hht047>
- Gormley, T., & Matsa, D. (2011) Growing out of trouble: Corporate responses to liability risk. *Review of Financial Studies*, 24(8), 2781-2821. <https://doi.org/10.1093/rfs/hhr011>
- Gropp, R., Hakenes, H., & Schnabel, I. (2011). Competition, risk-shifting, and public bail-out policies. *Review of Financial Studies*, 24(6), 2084-2120. <https://doi.org/10.1093/rfs/hhq114>

- Imbens, G. W. (2015). Matching methods in practice: Three examples. *Journal of Human Resources*, 50(2), 373-419. <https://doi.org/10.3368/jhr.50.2.373>
- Jayarathne, J., & Strahan, P. E. (1996). The finance-growth nexus: Evidence from Bank Branch Deregulation. *Quarterly Journal of Economics*, 111(3), 639-670. <https://doi.org/10.2307/2946668>
- Keeley, M. (1990). Deposit Insurance, Risk, and Market Power in Banking. *American Economic Review*, 80(5), 1183-1200. <https://www.jstor.org/stable/2006769>
- Khwaja, A., & Mian, A. (2008). Tracing the impact of bank liquidity shocks: Evidence from an emerging market. *American Economic Review*, 98(4), 1413-1442. <https://doi.org/10.1257/aer.98.4.1413>
- Laeven, L., & Valencia, F. (2012). Systemic banking crises database: an update. *Working Paper International Monetary Fund*, 12/163, 1-32. <https://doi.org/10.2139/ssrn.2096234>
- Martins, T.C., & Novaes, W. (2012). Mandatory dividend rules: Do they make it harder for firms to invest? *Journal of Corporate Finance*, 18(4), 953-967. <https://doi.org/10.1016/j.jcorpfin.2012.05.002>
- Meijer, E., Spierdijk, L., & Wansbeek, T. (2017). Consistent estimation of linear panel data models with measurement error. *Journal of Econometrics*, 200(2), 169-180. <https://doi.org/10.1016/j.jeconom.2017.06.003>
- Mithani, M. A. (2017). Liability of foreignness, natural disasters, and corporate philanthropy. *Journal of International Business Studies*, 48, 941-963. <https://doi.org/10.1057/s41267-017-0104-x>
- Oliveira, R. F., Schiozer, R., & Barros, L. A. B. C. (2015). Depositors' perception of "too-big-to-fail". *Review of Finance*, 19(1), 191-227. <https://doi.org/10.1093/rof/rft057>
- Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1), 435-480. <https://doi.org/10.1093/rfs/hhn053>
- Ponticelli, J., & Alencar, L. S. (2016). Court enforcement, bank loans, and firm investment: Evidence from a bankruptcy reform in Brazil. *Quarterly Journal of Economics*, 131(3), 1365-1413. <https://doi.org/10.1093/qje/qjw015>
- Roberts, M. R., & Whited, T.M. (2013). Endogeneity in empirical corporate finance. *Handbook of the Economics of Finance*, 2, 493-572.
- Sampaio, J. O., Gallucci Neto, H., Silva, V. A. B., & Schiozer R. (2020). Mandatory ifrs adoption, corporate governance, and firm value. *Revista de Administração de Empresas*. Retrieved from <http://bibliotecadigital.fgv.br/ojs/index.php/rae/article/view/81318/77663>
- Schiozer, R., Mourad, F. A., & Vilarins, R. S. (2018). Bank risk, bank bailouts and sovereign capacity during a financial crisis: a cross-country analysis. *Journal of Credit Risk*, 14(4), 1-28. <https://doi.org/10.21314/JCR.2018.246>
- Schiozer, R. F., & Oliveira, R. F. (2016). Asymmetric transmission of a bank liquidity shock. *Journal of Financial Stability*, 25, 234-246. <https://doi.org/10.1016/j.jfs.2015.11.005>
- Soedarmono, W., Machrouhb, F., & Tarazi, A. (2013). Bank competition, crisis and risk taking: Evidence from emerging markets in Asia. *Journal of International Financial Markets, Institutions and Money*, 23, 196-221. <https://doi.org/10.1016/j.intfin.2012.09.009>
- Ueda, K., & di Mauro, B. W. (2013). Quantifying structural subsidy values for systemically important financial institutions. *Journal of Banking and Finance*, 37(10), 3830-3842. <https://doi.org/10.1016/j.jbankfin.2013.05.019>
- Wooldridge, J. M. (2010). *Econometric analysis of cross-section and panel data* (2nd ed). Massachusetts: MIT Press.

ENDNOTES

1. While independence implies no correlation, the reverse is not always true. Nevertheless, the absence of correlation between the regressors and the error term will yield consistent estimators, and is generally a sufficient condition for causal inference. For more details, see Wooldridge (2010).
2. The error term in the model, u , should not be confounded with the estimated error in the regression, which is uncorrelated with the regressors by construction.
3. Dependence among observations of the same firm through time may cause the incorrect estimation of standard errors in the regression. For more details on the correct estimation of standard errors through clustering, we direct the reader to Bertrand, Duflo, and Mullainathan (2004) and Petersen (2009).
4. To account for autocorrelation of the error term of observations of the same firm through different periods, one needs to cluster the standard errors at the firm level to avoid underestimating them. See Petersen (2009) for details.
5. We use the latest available information prior to the crisis.
6. The financial institutions in our sample are commercial banks, savings banks, cooperative banks, mortgage banks, and government credit institutions. We call all of them "banks" for simplicity.
7. The countries are Australia, Canada, Czech Republic, Estonia, Finland, Israel, Japan, Mexico, New Zealand, Norway, Poland, Republic of Korea, Slovakia, and Turkey.
8. Another important fact that may have contributed to the decrease in bailout expectations of protected banks is that, by 2011, regulators had recognized that heterogeneous bailout expectations caused competitive distortions, and started discussing alternative measures to the bailout of large institutions, such as the creation of contingent capital, bail-in provisions and others.
9. The computation of the effect is as follows: $e^{0.855} - 1 \approx 1,35 = 135\%$.

Authorship

Rafael Felipe Schiozer*

Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo, Departamento de Contabilidade, Finanças e Controle. Av. Nove de Julho, nº 2029, Bela Vista, 01313-902, São Paulo, SP, Brazil.

E-mail address: rafael.schiozer@fgv.br

 <https://orcid.org/0000-0003-3890-6345>

Frederico Abou Mourad

Banco Central do Brasil. Av. Paulista, nº 1894, Bela Vista, 01310-922, São Paulo, SP, Brazil.

E-mail address: frederico.mourad@bcb.gov.br

 <https://orcid.org/0000-0001-9087-9634>

Theo Cotrim Martins

Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo, Departamento de Contabilidade, Finanças e Controle. Av. Nove de Julho, nº 2029, Bela Vista, 01313-902, São Paulo, SP, Brazil.

E-mail address: theoem@gmail.com

 <https://orcid.org/0000-0001-6768-3799>

* Corresponding Author

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Banco Central do Brasil.

Funding

The authors thank the Ministério da Ciência, Tecnologia e Inovação, Conselho Nacional de Desenvolvimento Científico e Tecnológico, grant number 305423/2018-5, for the financial support.

Conflict of Interests

The authors have stated that there is no conflict of interest.

Authors' Contributions

1st author: conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); project administration (lead); software (supporting); writing-original draft (lead); writing-review & editing (lead).

2nd author: conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (lead); writing-original draft (supporting); writing-review & editing (supporting).

3rd author: conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (supporting); writing-original draft (supporting); writing-review & editing (supporting).

Copyrights

RAC owns the copyright to this content.

Plagiarism Check

The RAC maintains the practice of submitting all documents approved for publication to the plagiarism check, using specific tools, e.g.: iThenticate.

Peer Review Method

This content was evaluated using the double-blind peer review process. The disclosure of the reviewers' information on the first page is made only after concluding the evaluation process, and with the voluntary consent of the respective reviewers.

Data Availability

All data and materials were made publicly available through the Harvard Dataverse platform and can be accessed at:



Schiozer, Rafael F.; Mourad, Frederico A.; Martins, Theo, 2020, "Replication Data for: A tutorial on the use of differences-in-differences in management, finance and accounting research", <https://doi.org/10.7910/DVN/AEENUD>, Harvard Dataverse, V1.