Essays in Empirical Labor Economics

Omar Bamieh

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 24 October 2017

European University Institute
Department of Economics

Essays in Empirical Labor Economics

Omar Bamieh

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Examining Board

Prof. Andrea Ichino, EUI, Supervisor
Prof. Andrea Mattozzi, EUI
Prof. Samuel Bentolila, CEMFI
Prof. Francis Kramarz, CREST and ENSAE

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I, Omar Bamieh certify that I am the author of the work "Essays in Empirical Labor Economics" I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

Omar Bamieh

Florence, June the 28th 2017

...

Acknowledgements

Writing this thesis has kept me busy for the last five years of my life, over which I met many wonderful people whom I fear I may forget to mention in these acknowledgments. For fear of not missing anyone let me start the beginning.

First of all I should thank my colleagues at the European University Institute (EUI), from the first day of the PhD until the very last I always felt so lucky to be surrounded by such and amazing and supportive group of friends and colleagues. Thank you all!

The same applies to my supervisors who had to bear with me for most of my time the EUI. I thank my first supervisor, Andrea Ichino, from whom I learned a lot over the course of the years and thanks to Andrea the three chapters of my thesis went from mere ideas to concrete papers. Thank you also Andrea for taking us mountain biking and for having us over at your place for lunches and dinner, it was always great fun! I am also grateful to my second supervisor, Andrea Mattozzi, who has always been available to discuss my ideas and paper. Matthias Sutter has also helped me with the first chapter of this thesis and I am grateful to him, especially for helping me to understand football a bit better! Juan Dolado has been a constant support for me throughout the PhD, both personal and professional. I also need to thank Juan for introducing me to the world of Pop–Rock by lending me CDs (that is right CDs) on a regular basis for the last four years. Juan, what will I do now that I left the EUI?

During my PhD I started a new sport: climbing, which quickly turned into true passion. I thank Juan Masullo Jimenez and Francesco Arbi (also known as Checco or Gatto Guida) for introducing me to climbing. I thank all the people at the climbing gym Hubble in Florence for being such a special group, I am sure it is because of you that I came to love climbing so much. And thank you Rasmus for starting climbing together, I will never forget our motivation to go to Hubble together for endless bouldering session back in that Fall that we picked up this sport together. Thank you Noemie and Wojtek for taking me climbing outdoor for the first time, and for the numerous climbs together. Thank you Giovanni (Bomber) for the many intensive bouldering sessions at Hubble! Finally thank you Linda Gilbert for the financial support that the 4B offered to the EUI climbing club.

I shall not forget a special thank to my lifelong friend, Matteo (Gatto), whom joined me at the EUI one year after I started. We just could not live without each other, so we had to do the PhD in the same place! Thank you Matteo for your friendship, support and our many amazing and healthy dinners. Thank you Dominik and Mireia for all the nice dinners at your place. Thank you Andresa for our many nights out partying together and thank you Kristina for always making sure our glasses were never empty! Thank you John and Brais for all the great time together!

The EUI is a great place, not only for its amazing location but also for the people who work there. A special thank to Jessica, Lucia, Anne and Sarah for their administrative and personal support. You made everything easier! An even more special thank to Sonia and Loredana for keeping me well fed during my time in Florence and for our many chats. Thank you Giuseppe, who unfortunately was there only for less than half my time at the EUI, for always making sure everything would be delivered for me.

Thank you Fatma for being so supportive, you entered my life close to the end of the PhD but I hope you will never leave it.

Finally, I would like to thank my parents who always pushed me to pursue a PhD, even in the moments that I doubted myself.

Contents

Thesis summary

In the first chapter of my thesis I study the effect of firing costs and labor reallocation. Exogenous variation of expected firing costs is offered by the random allocation of judges to trials involving firms in a large Italian court. Judges may be slow or fast and therefore firms experience randomly assigned shorter or longer trial lengths in an institutional context in which longer trials imply higher employment protection. I find that a 1% increase in expected firing costs induced by the past experience of a longer trial reduces the hazard of hiring or firing by 0.4% after the end of the trial.

In the second chapter of my thesis I use administrative data from one large Italian court I quantify the extent to which lawyers can be held accountable for the slowness of Italian courts. Borrowing the methodology used in the analysis of employers-employees linked data, I estimate the contribution of unobservable time-invariant plaintiff and defendant lawyers' characteristics in explaining the variability of trials' length. I find that 27% of the variance in trials length is explained by unobservable time-invariant lawyers' characteristics.

In the third chapter I study how tournaments may motivate workers to provide effort, yet differences in relative abilities may undermine the incentives of workers to exert effort. I use a novel data set from professional football competitions and find that differences in relative abilities are associated with lower effort exerted by players. In this empirical setting, effort and relative abilities are measured as, respectively, the distance covered on the pitch by football players and relative winning probabilities, the latter derived from betting odds of professional bookmakers. I find that larger differences in betting odds of opposing teams lead to less distance covered on the pitch.

1 Firing Costs, Employment and Misallocation. Evidence from Randomly–Assigned Judges

Using a quasi–experimental setting I test the effect of firing costs on firms hiring and firing decisions and I provide a theoretical interpretation of these causal estimates. Exogenous variation of expected firing costs is offered by the random allocation of judges to trials involving firms in a large Italian court. Judges may be slow or fast and therefore firms experience randomly assigned shorter or longer trial lengths in an institutional context in which longer trials imply higher employment protection. I find that a 1% increase in expected firing costs induced by the past experience of a longer trial reduces the hazard of hiring or firing by 0.4% after the end of the trial. The same variation generates a 0.3% increase of average employment levels. These effects are not due to the sunk costs induced by past trials since they do not depend on how much the firm is liquidity constrained. They are, instead, smaller in size for older firms that, given more experience, are less likely to revise their expectations.

1.1 Introduction

Employment protection is widely believed to reduce firms firing and hiring while the effect on employment levels is ambiguous.[1] In this paper, I use a quasi–experimental setting to test the effect of firing costs on firms hiring and firing decisions and I provide a theoretical interpretation of these causal estimates.

Exogenous variation of expected firing costs is offered by the random allocation of judges to trials initiated by firms in a large Italian court. There are fast and slow judges, and firms have the same probability of being assigned to any judge. In the Italian context, longer trials imply higher firing costs for firms, independently of the trial outcome. Therefore, the exogenous variation in the length of trials experienced by different firms creates an exogenous variation of realized firing costs. Different realizations of firing costs may lead to different post–trial expectations of future firing costs potentially affecting hiring and firing decisions.

The empirical analysis uses administrative data from one large Italian labor court. This data set contains detailed information on the universe of cases filed between 2001 and 2012 including, specifically, the duration of the trial, the identity of the judge assigned to the case, and the identity of the parties. I match this information with data on firms monthly employment levels and balance sheet information taken, respectively, from the archives of the Italian Social Security Institute (INPS) and from CERVED.[2]

This information allows me to estimate how changes of expected firing costs, induced by the past experience of different trial lengths, affect the hazard of a variation of firm

[1] Seminal papers showing these results are Bentolila and Bertola 1990; Hopenhayn and Rogerson 1993.

[2] Centri Elettronici Reteconnessi Valutazione Elaborazione Dati, a private company collecting balance sheets of the universe of Italian limited companies.

employment. I find that a 1% increase in expected firing costs, measured as the increase in the experienced trial lengths, leads to a 0.4% decrease in the hazard rate out of the spell of employment inaction for firms. The same increase in expected firing costs generates a 0.3% increase of average employment levels.

These effects are not due to the cost paid for a long trial, because this cost is sunk and therefore it cannot affect the future optimal decisions of the firm. Reassuringly, the effects found do not depend on how much the firm is liquidity constrained, ruling out the possibility that, without perfect capital markets, the sunk cost induced by past trials could affect future decisions. Therefore, variations in experienced trial lengths affects firms' future decisions only as variations in expected firing costs.

Moreover, the effects found are smaller in size for older firms, supporting a possible interpretation that firms learn trial length with their experience in court. If firms do not know the exact trial length, then they may have different post–trial expectations on firing costs depending on their experienced trial lengths. The importance of a single experience in court to change firms' expectations is inversely related to the precision of their prior information. Presumably, younger firms have less precise knowledge of the exact trial length, making them react more to the newly acquired information through their direct experience in court.

These results confirm the theoretical predictions that firing costs introduce a corridor of inaction over which firms would prefer neither to hire nor to fire, thus, reducing employment adjustment over the business cycle. However, since both hiring and firing are reduced, the long run net effect on employment levels is ambiguous. Whether employment protection leads to higher or lower employment is an empirical question. According to my reduced form causal estimates firing costs increase employment levels.

Given the source of variation in expected firing costs considered, this paper further contributes to the literature by assessing the total effect of firing costs, which includes not only the transfer from the firm to the worker but also the tax component. In fact, firing costs have two separate dimensions: a transfer from the firm to the worker to be laid off, and a tax to be paid outside the firm–worker pair.

Variations in the tax component, associated with longer trials, are due to the following two facts. First, legal costs increase in trial length because Italian lawyers do not charge a flat fee but are paid according to the time spent on a case. Moreover firms need to cover at least their own legal expenses independently of the trial outcome. Second, since court cases represent a period of uncertainty, potentially affecting firms productivity negatively, firms may prefer short trials.[3] Besides, firms employing more than 15 employees are also sentenced to pay all forgone wages from the day of dismissal to the day of court ruling, if the judge rules in favor of the worker. This represents a variation in the transfer component associated with longer trials. Overall, longer trials imply higher firing costs, both in terms of the tax and of the transfer component. Considering separately the two subgroups of firms, employing 15 or less employees, and more than 15 employees, allows to identify exclusively the tax component of firing costs for the former subgroup of firms.

The remainder of this paper is structured as follows. Section 1.1.1 summarizes the literature related to my work. Section 1.2 introduces a simple model to derive testable hypothesis and interpret the causal estimates found. The same section also describes why, in the Italian context, longer trials imply higher firing costs, independently of the trial outcome. Section 1.3 describes the court and the firms level employment data. Section 1.4 describes the empirical strategy and how judges instruments are computed. Section 1.5

[3]The seminal paper showing the importance of uncertainty shocks is Bloom 2009.

reports the results. Section 1.6 concludes.

1.1.1 Related Literature

Since Lazear 1990 seminal paper there has been a large theoretical and empirical literature studying the effects of firing costs on firms' hiring and firing decisions. Some of these works relied on cross–countries comparisons using aggregate data (Lazear 1990) and firm level data (Haltiwanger, Scarpetta, and Schweiger 2008, 2014). Others used within–country variation of employment protection for different groups of firms (Autor, Kerr, and Kugler 2007; Kugler and Pica 2008). My paper contributes to this literature by using a true random allocation of expected firing costs to firms to identify the causal effect of employment protection on firms' hiring and firing decisions.

The random assignment of cases to judges has been exploited in other settings to identify the causal effects of, incarceration (Aizer and Doyle Jr 2013; Manudeep, Dahl, Løken, Mogstad, et al. 2016), disability insurance (Autor, Kostøl, and Mogstad 2015) and intergenerational transmission of welfare values (Dahl, Kostol, and Mogstad 2013). My paper fits into this literature as it exploits the natural experiment created by the random allocation of firms to judges within a large Italian labor court.

There are other works exploring within countries variations of legal practices. Gianfreda and Vallanti 2015 exploit the regional variation of courts' delay between Italian courts to compare firms' jobs flows and productivity. Fraisse, Kramarz, and Prost 2015 exploit the regional variation of the activity of French labor courts to study firms' jobs flows.

1.2 Background

This section is organized as follows. Section 3.2 introduces a standard model of employment adjustment with asymmetric costs. The model allows to clarify the hypothesis tested empirically in this paper. Section 1.2.2 explains why longer trials imply higher firing costs for Italian firms. Section 1.2.3 provides a possible explanation of how past experienced trial lengths could affect future expected firing costs. Section 1.2.4 explains the relation between employment flows and worker flows.

1.2.1 Theoretical framework

This section introduces a model of labor demand to clarify the theoretical prediction to be tested empirically, namely that a rise in firing costs reduces the firm's willingness to hire and fire. And how this affects employment levels. Following Bentolila and Bertola 1990; Bentolila and Saint-Paul 1994, firms' optimal behavior is described in terms of a band of revenue shocks within which inaction is optimal. A rise in firing costs increases this band and makes it more likely that firms do not change their employment level, I call this behavior employment inaction or simply inaction.

The profits of firm i, which employs homogeneous labor, n_{it}, as the sole input at time t, are given by

$$\pi_{it} = z_{it} f(n_{it}) - w_i n_{it} - F_i \max\{0, n_{it-1} - n_{it}\} \tag{1.1}$$

where z_{it} is a revenue shock identically distributed over time with cumulative density function G. w_i is the exogenous real wage, F_i is the firing cost.[4] The production function f is strictly increasing and strictly concave.

[4]Strictly speaking F_i represents an adjustment costs, however, abstracting from retirements and quits, the firm needs to fire workers in order to reduce its number of employees. Similarly, the firm need to hire worker to increase it number of employees. Section 1.2.4 elaborates more on this point.

The firm is risk neutral and chooses employment after the current shock realization is observed, to maximize the present discounted value of expected profits over an infinite horizon:

$$\max_{\{n_{it}\}_{t=0}^{\infty}} \sum_{t=1}^{\infty} \delta_i^t E\{z_{it} f(n_{it}) - w_i n_{it} - F_i \max\{0, n_{it-1} - n_{it}\}\} \quad s.t. \quad n_{it} \geq 0 \tag{1.2}$$

where E is the expectations operator and δ_i the discount factor of firm i ($\delta_i \in [0,1]$). In the absence of firing costs, $F_i = 0$, the problem of firm i would be a simple repeated static problem in which the firm, after observing the realization of the shock at time t, z_{it}, chooses the level of employment n_{it} that equates the marginal product of labor to the exogenous wage, w_i. In the presence of firing costs $F_i > 0$, however, the firm takes into account the previous level of employment, n_{it-1}, when choosing its employment level at time t. Technically speaking, n_{it-1} becomes a state variable in the optimization problem of the firm, which can be represent with a Bellman equation:

$$V(n_{it-1}, z_{it}) = \max_{n_{it} \geqslant 0} \; z_{it} f(n_{it}) - w_i n_{it} - F_i \max\{0, n_{it-1} - n_{it}\} + \delta_i E_t\{V(n_{it}, z_{it+1})\} \tag{1.3}$$

where V is the value function.

Due to firing costs the derivative of the objective function changes with the sign of the change in employment. The following first–order conditions are necessary and (by concavity) sufficient for optimality:

$$z_{it} f'(n_{it}) - w_i + F_i + \delta_i E_t \left(\frac{\partial V(n_{it}, z_{it+1})}{\partial n_{it}} \right) = 0 \qquad \text{(firing)} \tag{1.4}$$

$$z_{it} f'(n_{it}) - w_i + \delta_i E_t \left(\frac{\partial V(n_{it}, z_{it+1})}{\partial n_{it}} \right) = 0 \qquad \text{(hiring)} \tag{1.5}$$

The non–differentiability of adjustment costs at n_{it-1} creates a discontinuity in the firm's decision rule. Depending on the realization of the shock z_{it} it is optimal to satisfy neither equation (1.4) nor (1.5) but to maintain employment at the previous period's level. The optimal rule is: (refer to Appendix 1.A for the proofs of the results in this section)

Proposition 1. *(i) If*

$$z_{it} < \underline{z}_{it} \tag{1.6}$$

the firm fires and n_{it} is the solution to 1.4

(ii) If

$$z_{it} > \bar{z}_{it} \tag{1.7}$$

the firm hires and n_{it} is the solution to 1.5.

(iii) If

$$\underline{z}_{it} < z_{it} < \bar{z}_{it} \tag{1.8}$$

the firm is inactive: $n_{it} = n_{it-1}$, employment does not change in period t relatively to the previous period.

Therefore, it is not optimal for the firm to change employment at all, when the shock falls within an inaction range which is defined by two threshold values: $\underline{z}_{it}$ and $\bar{z}_{it}$ ($\bar{z}_{it} > \underline{z}_{it}$). Figure 1.1 shows a graphical representation of the optimal choice of labor as a function of the realization of the shock z_{it} and for a given value of the employment level in the previous period, n_{it-1}. Essentially, in the presence of firing costs the firm changes employment only if the shocks are either sufficiently high (which happens with probability $1 - G(\bar{z}_{it})$) or sufficiently low (which happens with probability $G(\underline{z}_{it})$), whenever shocks fall in between (which happens with probability $G(\bar{z}_{it}) - G(\underline{z}_{it})$) it is optimal for the firm not too change its employment level. It is easy to show that the size of the inaction range, $(\bar{z}_{it} - \underline{z}_{it})$, is increasing in the firing cost F_i, which leads to the following result:

Proposition 2. *An increase in the firing cost of firm i, F_i, increase the probability of inaction of the firm*

$$\frac{\partial[G(\bar{z}_{it}) - G(\underline{z}_{it})]}{\partial F_i} > 0 \tag{1.9}$$

Although optimal from the point of view of the firm, employment inaction is inefficient as it represents a deviation from a frictionless economy. Figure 1.2 compares the optimal employment of firms with positive firing costs, n_{it}, (red line) and firms with zero firing costs, n_{it}^{fl}, (blue line). The vertical differences between the blue and the red line represents the inefficiency introduced by firing costs.

The model presented in this section is a partial equilibrium model, for a more general statement about efficiency and welfare in a general equilibrium framework I refer to Hopenhayn and Rogerson 1993; Ljungqvist 2002. The essence of the result remains unchanged and firing costs reduce efficiency and welfare. Given that firing costs take the form of taxes, this result follows immediately from the First Welfare Theorem.

The effects of firing costs on employment levels is ambiguous. Firms are more inactive to fire but also to hire, leaving the net effect on employment levels undetermined. The first–order conditions (1.4) and (1.5) show that firing costs affect firing in the current period through F_i but, at the same time, firing costs have a discounted expected effect which is captured by the value function. When choosing employment in the current period the firm takes into account that the chosen level of employment will affect its payoff in the next period because firing is costly. In other words, a firing cost represents also an implicit hiring cost, because a firm hiring today must take into account the possibility of firing tomorrow, and firing is costly. How much this matters depends on the discount factor.

From equations (1.4) and (1.5) if the discount factor of the firm is zero, then only the current period matters for the firm. Firing costs do not affect hiring but reduce firing, thereby increasing the employment level at the firm. This point was point was first made by Bentolila and Bertola 1990, the net effect of firing costs on employment levels depends on the discount factor. The smaller is the firm's discount factor the more likely are firing costs to increase employment levels. Asymmetric adjustment costs and a positive discount factor produce a ratchet effect: the firm knows that workers may one day have a low marginal product of labor, (for example because of a recession), and firing costs will have to be paid to get rid of these workers, but this possibility is discounted since hiring occurs in good times, and bad times are far into the future. Expost (when bad times come), firing is less likely to occur due to firing costs, hence average employment increases.

Figure 1.3 illustrates this point, it reports the numerical solution of the model defined in equation (1.2) for different values of the discount factor. The figure shows how average employment levels change as firing costs increase, for firms with different discount factors. These results show that firing costs decrease the average employment level of firms with a high discount factor but increase the average employment level of firms with a low discount factor. This point is important to interpret the empirical results of section 1.5.4.

To empirically test the hypothesis that firing costs increase employment inaction (Proposition 2) and to test whether the effect of firing costs on employment levels is positive or negative, an exogenous variation in the firing cost F_i is needed. Section 1.2.2 explains that firing costs increase in the length of trials for Italian firms. Given that I have a randomized experiment with respect to trial lengths for firms going to court, I use this as the source of the exogenous variation in firing costs.

1.2.2 Longer trials imply higher firing costs

An exogenous variation of firing costs is needed in order to empirically test the effect of firing costs on employment inaction and on employment levels. The random allocation of firms to judges creates an exogenous variation in the length of trials experienced by firms. This sections explains different reasons why in Italy longer trials are more expensive for firms regardless of the outcome.

First, there is a consensus among legal scholars that Italian lawyers gain from longer trials, (see Marchesi 2003 for a review of the literature). In fact, Italian lawyers do not charge a flat fee but are paid accordingly to the number of hours worked on a case.[5] Moreover, Italian firms need to cover their own legal expenses even if they win the case and if they lose they also have to cover the legal expenses of workers. Therefore, longer trials are more costly for firms regardless of the outcome. Yet, one could argue that long trials involve many idle periods but the amount of work remains unchanged for lawyers. To rule out this possibility, Table 1.1 shows that there is a positive correlation between the length of the trial and the number of hearings to complete a case. Clearly, lawyers need to spend more time on cases taking more hearings to be completed.

Second, the trial represents a period of uncertainty and uncertainty shocks have been shown to affect the productivity of the firm, Bloom 2009. For this reason, in general firms should prefer shorter to longer trials.

Third, in the years considered in this study, 2001–2012, firms employing more than 15 employees had to pay all forgone wages from the day of dismissal to the day of court ruling if the judge ruled in favor of the worker. Additionally the firm had to pay a penalty to the

[5] More precisely, Italian lawyers are paid according to the number of tasks needed to assist their clients, see Marchesi 2003.

social security agency for delayed payment of social security contributions. This represents an expected cost because it depends on the probability of the firm losing the case against the worker.

Firing costs have two separate dimensions: a transfer from the firm to the worker to be laid off, and a tax to be paid outside the firm–worker pair. The first two points refer to the tax component, whereas the third refers to the transfer component associated with variations of trial length.

Finally, I rule out that longer trials lead to better outcomes for firms by showing that there is no correlation between the outcome and the length of the trial, (Table 1.2), and slower judges are not more likely to rule in favor of firms, (Figure 1.4).

It may still be that fast trials are bad for firms if, once firms realize that the assigned judge is slow, they accept costly, but fast, settlements to avoid being stuck in long trials. Therefore, even short trials could be costly for firms because they are the result of a fast, but costly, settlement between firms and workers. I rule out this possibility by showing that fast judges are not more likely than slow judges to induce settlements, (Figure A1).

Taken together, these facts suggest that longer trials imply higher firing costs. However, why should past experiences in court affect the future behavior of firms? Possibly, because firms learn the length of trials with their experience in court, hence they learn the degree of firing costs. Experiencing longer trials means experiencing higher firing costs. Consider a Bayesian updating rule where firm i has a prior on trial lengths, which implies a prior on firing costs.[6] Firm's i experienced trial length is a valuable signal for the firm to update its expectations on firing costs. Therefore, firms experiencing different trial lengths update

[6]This prior may come from different sources: aggregate statistics, lawyers or other firms. Regardless of the source of firm's i prior, each individual experience in court represents new information.

their expectations on firing costs differently and for this reason they behave differently after the end of the trial, that is after the new information has been acquired. Section 1.2.3 formalizes this argument.

1.2.3 Experienced trial lengths change expected firing costs

Suppose that firm's i prior on the true length of trials in the court considered in this study, ℓ, is Normally distributed with mean m_{0i} and variance $1/h_{0i}$ (i.e. precision h_{0i}). By going to court, firm i acquires a noisy signal, ℓ_i, about ℓ

$$\ell_i = \ell + \varepsilon_i \tag{1.10}$$

where ε_i are independent and identically Normally distributed, between different firms, with mean 0 and precision h_ε. ε_i are independent of ℓ. Given the discussion of section 1.2.2, the expected firing cost of firm i, F_i, depends on the expectation of firm i on the length of trials.

$$F_i \equiv E(\ell|\ell_i) \tag{1.11}$$

where $E(.|\ell_i)$ is the expectation operator. Given that both ℓ and ε_i are normally distributed, it is easy to show that, (DeGroot 2005)

$$F_i = \frac{h_{0i}m_{0i} + h_\varepsilon \ell_i}{h_{0i} + h_\varepsilon} \tag{1.12}$$

The firm weighs its prior, ℓ, and its signal, ℓ_i, according to their precisions. How does firm's i expected firing cost change when the signal changes?

$$\frac{\partial F_i}{\partial \ell_i} = \frac{h_\varepsilon}{h_{0i} + h_\varepsilon} \tag{1.13}$$

Remark 1. *Firm's i expected firing cost increases in the length of the experienced trial ℓ_i*

$$\frac{\partial F_i}{\partial \ell_i} > 0 \tag{1.14}$$

Therefore, the exogenous variation in the length of the trial experienced by firm i represents an exogenous variation in the expected firing cost of firm i, which can be used to empirically test the propositions presented in Section 1.2.1.

However, firm's i expected firing costs, F_i, may be not affected by its signal, ℓ_i, if the firm has a very precise prior relative to its signal. In other words, firms that already know the true length of trials, ℓ, do not change their expectations of firing costs because of one experience in court.

Remark 2. *As the precision of the prior, h_{0i}, increases relative to the precision of the signal, h_ε, the signal does not affect firm's i expected firing cost*

$$\lim_{\frac{h_{0i}}{h_\varepsilon} \to +\infty} \frac{\partial F_i}{\partial \ell_i} = 0 \tag{1.15}$$

Empirical results in Section 1.5.5 show that only young firms are affected by the length of the trial experienced in court. Presumably older firms have more precise priors on trial lengths compared to younger firms. Remark 2 rationalizes this finding.

1.2.4 Worker flows and employment flows

By construction, hires, terminations and net employment changes are related.[7] Let me abstract for simplicity from retirements and quits. For any given business and at any level of aggregation, the net change in employment between two points in time satisfies a

[7]I refer to Davis, Faberman, and Haltiwanger 2006 for a detailed discussion.

fundamental accounting identity:

$$\text{Net employment change} \equiv \underbrace{\text{Hires} - \text{Terminations}}_{\text{Worker Flows}} \equiv \underbrace{\text{Creation} - \text{Destruction}}_{\text{Employment/Jobs Flows}}{}^{8} \qquad (1.16)$$

Job creation is positive for an expanding or new business, and job destruction is positive for a shrinking or exiting business. While a single employer can either create or destroy jobs during a period, it can simultaneously have positive hires and terminations. Hence, the flow of hires exceeds job creation, and the flow of terminations exceeds job destruction. As an example, consider a firm with two terminations during the period and one hire. The worker flows at this business consist of two terminations and one hire, but there is a net employment change of one destroyed job.

This paper focuses on the effect of firing costs on employment flows. Section 1.2.1 shows that higher firing costs inefficiently reduces employment flows. Ideally, firms increase their labor force in upturns and decrease it in downturns. Firing costs hinder this adjustment process. In the absence of exogenous shocks, firms keep their labor force constant and the effect of firing costs would not be observed.

Several papers studying firing costs also use employment flows, (see for instance Autor, Kerr, and Kugler 2007; Kugler and Pica 2008). Unfortunately, due to the absence of worker flows data in my firms level administrative database, I cannot combine the joint analysis of employment and worker flows as is done in Kugler and Pica 2008.

1.3 Data Description

This section is organized as follows. Section 1.3.1 describes the different databases used. Section 1.3.2 describes how the final dataset is constructed from these databases.

[8] In the literature the terms "employment flows" and "jobs flows" are used interchangeably.

1.3.1 Data sources

The empirical analysis uses administrative data from one large Italian labor court. This data set includes detailed information for the universe of cases filed in the labor court between 2001 and 2012, including: the duration of the trial, the identity of the judge assigned to the case, and the identity of the plaintiff and of the defendant.[9] Information on firms going to court is recovered using data from the Italian Social Security Institute (INPS) and from the private company CERVED.[10] The former includes information on the monthly number of employees, and dates of incorporation and termination of the firm. The latter includes information on annual balance sheet data.

The court data is a unique database of 320,191 trials filed between 2001 and 2012 in a large Italian labor court.[11] For these cases I observe the complete history from the day of filing to the day of disposition, which takes place in one of the two main forms: a sentence by the judge or a settlement between parties. These cases are assigned to 82 judges of this court.[12] Judges are not involved in other tasks inside the tribunal and do not deal with trials of other kinds; their entire working time is dedicated to labor controversies. With this data I can construct a measure of how long each judge takes on average to complete his or her cases and I can assess the length of each trial involving a firm.

The private company CERVED collects balance sheets data for the universe of Italian

[9] For example, in a firing litigation the worker is the plaintiff and the firm is the defendant.

[10] Centri Elettronici Reteconnessi Valutazione Elaborazione Dati, a private company collecting balance sheets of the universe of Italian limited companies.

[11] The court data contains all cases filed in 2001–2014 which history is followed until the end of 2014, the time at which the data provider stopped collecting the data. I restrict my sample in order to limit censoring to 4%. That is, 4% of the cases filed in 2001–2012 are still pending by the end of 2014. For these cases I take the censoring date as end date, but dropping or including these censored cases does not change the results.

[12] These 82 judges represents the subset of 111 judges who worked on least 1,000 trials in the years considered. Even though this restriction is not required for identification, it allows to construct more precise instruments as described in Section 1.4.2.

limited companies, however, since the number of employees is not part of the information in the balance sheets I complement this with data from the Italian National Social Security (INPS) archives contains monthly employment for each establishment of the universe of Italian firms active between 1990 and 2013. Measuring employment at the establishment level is important for my purposes because the location of the establishment determines the court responsible for any litigation of the firm to which the establishment belongs. For example, if a firm is registered in city A, but it has another establishment in city B which is involved in a legal dispute, then the court of city B has jurisdiction over the case.

1.3.2 Sample construction

Table A1 describes the sample construction. There are 25,906 firms taking part in 82,518 trials filed between 2001 and 2012 in the labor court considered in this study. There are 220,341 firms operating in the geographical area where the labor court has jurisdiction. Firms are linked using their names as the only identifier. For this reason only 7617 firms are merged between the two data sets. Table A2 shows that the observable characteristics of trials do not differ between the group of firms in the labor court database linked to the CERVED–INPS database, and the group of firms for which this linkage is not possible.

1.4 Empirical framework

This section is organized as follows. Section 1.4.1 explains the timing of events relevant for the empirical analysis. Section 1.4.2 explains how the judges instruments are constructed. Section 1.4.3 describes the variation of the instrument and the first stage. Section 1.4.4 describes the final dataset used in the remainder of this paper.

1.4.1 Set up

Theory states without ambiguity that firing costs make firms less willing to change their labor force (employment inaction). Higher firing costs should increase the duration of the spell of employment inaction, which is the time that firms take to change their employment level. As explained in Section 1.2.2, if firing costs increase in trial lengths and if experienced trial lengths represent signals on the true length of trials, then firms' expectations of firing costs depend on these signals. Therefore, the empirical analysis is framed as a duration problem. Starting from the month in which trials end, the duration of the spell of firms' employment inaction, which is the time until firms change employment, should be longer the longer is the duration of the experienced trial.

Figures 1.8 shows the timing of the events. The analysis starts from the month in which the trials end. The duration analysis is framed using months from the end of the trial as the unit of elapsed time to test if firms experiencing longer trials wait longer to change their labor force shortly after the end of the trial, relative to firms experiencing short trials.

I focus on firms' first trials in my data. This choice is determined by the fact the firms' future decisions of going to court could be affected by firms' first experiences in court. As a robustness check, the analysis is restricted to the subgroup of firms born after 2001, because for these firms the first trial observed is certainly the first trial ever experienced.

The main outcome of interest is the employment inaction of firms, which is the number of months firms take to change employment after the end of their trials. Table A3 describes the censoring that mechanically arises in this setting. Trials end on a day in 2001–2014 and monthly employment is measured in 1990–2013. Therefore, all trials ending after December 2013 have 100% censoring with respect to my outcome variable of interest because I need to observe firms for at least two months. Older trials have less censoring (0% for trials ending in 2001) than more recent trials (42% for trials ending in 2013) because firms are observed for more months after the end of the trial.

Any assessment of the impact of trial lengths on employment inaction must address the problem posed by the correlation between trial lengths and factors such as the characteristics of the firm that are also likely to be correlated with the outcome. My empirical strategy uses the average time that randomly–assigned judges take to complete their cases as an instrument for the actual length of cases to which judges are assigned.

1.4.2 Instrumental variable calculation

For each firm I assign an instrument that corresponds to the average length of the judge assigned to the firm's first trial. The instrument, which is defined for each firm i assigned to judge $j(i)$ is simply a mean:

$$Z_{j(i)} = \left(\frac{1}{n_{j(i)}}\right)\left(\sum_{k=1}^{n_{j(i)}} \ell_k\right). \tag{1.17}$$

Here, $n_{j(i)}$ is the total number of cases seen by judge j excluding the cases of the firms for which I estimate the effect of trial lengths on employment inaction. Not excluding these trials could mislead us to believe there is a first stage even though the positive correlation between ℓ_i and $Z_{j(i)}$ is artificially created due to the fact that $Z_{j(i)}$ is a function of ℓ_i. ℓ_k is the length of the k–case seen by judge j.

In other words, I subtract the first cases of the 8,007 firms from the universe of 320,191 cases filed in order to compute the average length that each judge takes to complete his or her cases. This restriction is needed in order to assess the quality of the first stage, removing the positive correlation which is artificially created by the way in which the instrument is defined.

The instrument can therefore be interpreted as judges' average speeds to complete their cases. A slow judge has a high value of $Z_{j(i)}$ and a fast judge a low value of $Z_{j(i)}$. The validity of this instrument comes from the fact that each judge is assigned to a firm by a lottery, hence there is no correlation between the identity of the judge and the characteristics of the firms before going to court.

1.4.3 Judge variation

The analysis dataset includes 82 judges. The average number of cases per judge is 3,807 and the minimum number of cases seen by each judge is 1,000. Only one judge can hear each firm's case over time.[13] Each judge is monocratically responsible for the trials assigned to him or her. No jury or other judges are involved. The average number of months of each judge to complete his or her cases has mean 18 with a standard deviation of 5. Variation in the instrument can also be seen in Table 1.3 and Figure 1.5.

The fastest judge takes on average 9 months to complete a case, while the slowest judge takes on average 37 months. These differences can only be explained by the different ways in which judges work since cases are randomly allocated to judges within a court. In Italy, as in other countries, the law (Art. 25 of the Constitution) requires that judges receive a randomly assigned portfolio of new cases. My econometric strategy crucially relies on this random assignment, which is designed to ensure the absence of any relationship between the identity of judges and the characteristics of the cases assigned to them, including the characteristics of the firms involved in the cases. Section 1.5.2 provides evidence supporting

[13] If a judge retires or is transferred to a different court (for whatever reasons) his/her cases are either all randomly assigned to a new judge or they are distributed randomly to all the other judges in the court. For these cases the instrument is the average of the instruments of the different judges who worked on the cases.

the random allocation of firms to judges.

Figure 1.6 shows that there is a first stage, a positive correlation between the instrument defined in equation (1.17) and the length of the 8,007 firms' trials. This positive correlation is not artificially created because these 8,007 trials are not used to construct judges' instruments.

1.4.4 Sample description

The sample consists of firms that went to the court considered in this study in 2001–2012 which trials ended in 2001–2013. Table 1.3 reports descriptive statistics for the instrument and for the length of firms' trials. Table 1.4 reports descriptive statistics of firms' employment levels and durations of their spells of employment inaction. The latter is defined as the time, measured in months, until a firm changes its employment level after the end of the trial.

Table 1.5 describes the types of litigation of the 7617 used in the analysis. Since firms may learn trial lengths with their experience in court, the analysis is not restricted only to firms experiencing a trial following the termination of their employees but to any type of trial.[14] A firm may go to court because of a litigation related to the compensation of its employees, although the length of this particular trial does not imply a firing cost, firms can use this experience to infer the length of trials and form expectations about firing costs accordingly. Section 1.5.5 explores this point by comparing empirical results for the subgroups of firms experiencing firing trials and other types of trials.

[14]Firing cases refer only to individual dismissals and not to collective layoffs.

1.5 Results

This section is organized as follows. Section 1.5.1 provides some evidence that firing costs are binding for firms employment changes in the time period considered in this study. Section 1.5.2 tests for the random allocation of firms to judges. Section 1.5.3 presents the effect of expected firing costs on employment inaction. Section 1.5.4 empirically shows the consequences of firing costs, and the associated employment inaction, for employment levels. Section 1.5.5 explores heterogeneous effects on employment inaction with respect to several firms' characteristics. Section 1.5.6 rules out that employment inaction is caused by a first order effect of long trials on firms' balance sheet, instead of through firms' expectations on firing costs.

1.5.1 Firms employment variability in the post trial period

As shown in section 1.2.1, firing costs reduce employment adjustments when firms are hit by exogenous shocks. For example, consider two firms subject to the same exogenous revenue shocks but the first firm is subject to a low firing cost regime, whereas the second is subject to a high firing cost regime. Theory unambiguously predicts that the second firm will change its employment level less often than the first firm. Consider now the same two firms but in the absence of exogenous revenue shocks, in this case the two firms have no need to change their labor force, hence there will be no difference of employment changes between the low firing cost firm and the high firing cost firm.

This sections shows that firms considered in the analysis changed significantly their employment levels in the months following the end of their trial, suggesting that firing costs for these firms are likely to be binding. Table 1.6 reports summary statistics of firms monthly employment levels in the months after the end of their trials. The high within standard deviation of monthly employment levels suggests that firms had plenty of need to change their labor force. Yet, one may worry that all the variation comes only from a few firms, Figure 1.7 reports the distribution of the relative standard deviations (the ratio of the standard deviation to the mean) of monthly employment levels of each firm in the post trial period.

It is also important to quantify how often firms changed their employment levels after their experience in court. Table 1.7 reports the relative frequencies of positive and negative employment changes. In general firms changed their employment level and they experienced more negative than positive employment changes.

1.5.2 Instrument validity

Although I cannot directly test the validity of my instrument, I can provide evidence consistent with the condition being met. First, I have confirmed with court personnel that judges are assigned in a way that leads to a natural randomization of cases to judges: a computer randomly allocates cases to judges in such a way that at the end of a given period, all judges have been assigned an equal number of cases.[15] Second, I can partially test this empirically by examining whether the time–invariant and time–variant characteristics of firms, measured the year before the filing of their cases, differ by judge

Table 1.8 tests whether firms' characteristics are predictive of the average length that judges take to complete their cases. Essentially, I want to rule out that slow judges are assigned to particular groups of firms. Reassuringly, I find no relationship. Jointly, these variables explain less than 0.2 percent of the variation in the judge average length (joint

[15]This implies that slower judges accumulate more cases than faster judges, because all judges are continuously assigned new cases. Still, my identifying assumption does not hinge on judges being assigned the same the number of cases but solely on the random allocation of types of cases to judges, given the number of cases assigned.

p–value of 0.6355), and none is statistically significant at the 10 percent level.

1.5.3 The effect of firing costs on employment inaction

This section answers the following empirical question: do firms that expect higher firing costs take longer to change their employment level compared to firms that expect lower firing costs?

The empirical analysis is framed using "months from the end of the trial" as the unit of elapsed time. I follow the partial–likelihood approach proposed by Cox 1972 and specify the hazard that firm i changes employment t months after the end of the trial as:[16]

$$h_{it} = h_0(t)\exp\{\beta\ell_i\} \tag{1.18}$$

where $\exp\{\beta\ell_i\}$ captures the deviations from the baseline hazard, $h_0(t)$, in which I am interested. ℓ_i is the length of the trial experienced by firm i. Let T_i be the number of months firm i takes to change the employment level after its trial has ended. I call this the duration of the spell of employment inaction. The hazard, h_{it}, is the (limiting) probability that firm i leaves the state of inaction exactly at month t after the end of the trial.

According to the interpretation proposed in Section 1.2.3, β measures the causal effect of expected firing costs on the hazard of employment action. Theory states unambiguously that this coefficient should be negative, because higher firing costs make firms less willing to adjust their labor force, thereby increasing the duration of the spell of employment

[16]The hazard function represents the instantaneous probability of a failure event, for example the event that the firm changes its employment level, and is defined as,

$$h_{it} = \lim_{h \to 0^+} \left\{ \frac{\Pr(t \leq T_i < t + h | T_i \geq t, x_i)}{h} \right\}$$

where T_i is the time, from the end of the trial, until firm i changes employment. x_i can be any exogenous variable. In my application I am not interested in the shape of the baseline hazard but only in how the hazard is shifted with respect to the baseline by different expected firing costs.

inaction.

As in the case of omitted variable bias in linear regression, Maximum Likelihood estimation of the hazard does not guarantee that the causal effect in which I am interested is identified and estimated consistently because of the possibility that firm specific unobservables are not independent of the duration of the trial and at the same time affect the hazard. For example, if "bad" firms make trials last a long time and at the same time are very slow in adjusting their labor force, then the estimates from model 1.18 could represent only this selection bias.

Let U_i denote such a variable, for example, the unobservable quality of firm i, which could affect at the same time the length of the trial of firm i and its hazard, generating spurious correlations between these variables with no causal interpretation. Suppose that U_i is perfectly observable. Then the correctly specified hazard would be

$$\tilde{h}_{it} = h_0(t)\exp\{\beta\ell_i + \gamma U_i\} \tag{1.19}$$

and the following condition would hold

$$E[\tilde{h}_{it}|\ell_i, U_i] = 0. \tag{1.20}$$

This condition says that any shock affecting the actual completion hazard is random and independent of the determinants of the correctly specified parametric hazard $\tilde{h}_{it}$.

Now suppose that U_i is unobservable and that therefore I can only estimate the mis–specified hazard 1.18. Then, the expected difference between the true and the mis–specified hazard would be:

$$E[\tilde{h}_{it} - h_{it}|\ell_i] = h_0(t)(E[\exp\{\beta\ell_i\}\exp\{\gamma U_i\}|\ell_i] - \exp\{\beta\ell_i\}) \neq 0 \tag{1.21}$$

and the estimates of the causal parameter β would be inconsistent.

To address this problem I follow Palmer 2013; Coviello, Ichino, and Persico 2015 in using the control function approach proposed by Heckman and Robb 1985 adapted to the context of duration analysis. Consider the following first stage regression:

$$\ell_i = \delta_0 + \delta_1 Z_{j(i)} + v_i \tag{1.22}$$

where $Z_{j(i)}$ is an exogenous determinant the duration of the trial of firm i that is independent of U_i and does not affect the hiring hazard directly. The residual v_i capture the component of ℓ_i which depends on U_i.

Conditioning also on these residuals in the hazard 1.18 solves the endogeneity problem and delivers consistent estimates of the causal effects of interest. To see why, consider the following augmented specification of the hazard:

$$\bar{h}_{it} = h_0(t)\exp\{\beta\ell_i + g(v_i)\} \tag{1.23}$$

where $g(v_i)$ is a polynomial in the estimated residual from the first–stage regression 1.22. Using this specification, the expected difference between the true and the augmented hazard would be:

$$E[\tilde{h}_{it} - \bar{h}_{it}|\ell_i, v_i] = h_0(t)\exp\{\beta\ell_i\}(E[\exp\{\gamma U_i\}|\ell_i, v_i] - \exp\{g(v_i)\}) \tag{1.24}$$

which is equal to zero if the control function $\exp(g(v_i))$ is equal to the conditional expectation $E[\exp(\gamma U_i)|\ell_i, v_i]$. If the control function g(.) is linear and the conditional distribution of $\exp(\gamma U_i)$ is normal with appropriate mean and variance, then the equality holds exactly. Otherwise, identification relies on the quality of the polynomial control function in approximating the conditional expectation of $\exp(\gamma U_i)$. While this quality can be assessed by showing that results are robust to different specifications of the polynomial g(.), which is the case in my application[17], the crucial assumption on which this identification strategy stands is that the instrument $Z_{j(i)}$ is independent of the omitted determinant U_i of the hazard. To construct instruments that satisfy this condition I exploit the lottery that assigns cases to judges in the way explained in section 1.4.2.

The second column of Table 1.9 reports the estimates of the first stage regressions (1.22) which are strongly significant, the Cragg–Donald Wald F statistic is equal to 256. The control function estimates of the hazard (1.23), based on these first stage regressions, are in the second column of Table 1.9. The results show that firms experiencing longer trials are less likely to change employment in the months following the end of the trial. These results match the predictions of my theoretical framework: higher firing costs make firms less likely to change their employment levels.[18]

As shown in section 1.2.1, firing costs create an inaction range with respect to exogenous shocks. If shocks fall inside this range than the firm does not change its employment level. Given that the size of the inaction range increases in the firing costs, the higher the firing costs, the less likely is the firm to change employment in any period. As explained in section 1.2.2, longer trials imply higher costs for firms regardless of the outcome of the trial. The results in table 1.9 suggest that firms change their expectations on firing costs depending on the trial lengths they experience.

To understand the economic significance of the coefficient of column (2) in Table 1.9, note that it can be interpreted as the effect of a 1 unit change of the variable on the natural

[17]In Table 1.9, and in all other tables reporting estimates based on the control function, I present results based on a fifth degree polynomial, but I have experimented polynomials with different degrees obtaining similar results.

[18]Table A4 reports results of equations (1.22)–(1.23) augmented with a set of firms' control variables. Given my identification strategy, the inclusion of these controls should not change the estimates of β.

logarithm of the hazard ratio.[19] Based on these estimates suppose that the length of the trial experienced by the firm increases by 1%, at the median length of trials of 11 months this means making the trial approximately 3 days longer, then the hazard of employment inaction would decrease by 0.4%.[20]

To translate the effects of trial lengths on the hazard into effects on the durations of the spell of employment inaction some distributional assumptions are needed. As suggested by Arellano 2008, the Cox proportional hazard model can be written as a linear regression for the transformation $\Lambda(T_i) = \int_0^{T_i} u du$ of the underlying employment inaction duration T_i of firm i,

$$\ln(\Lambda(T_i)) = -\beta\ell_i + \eta_i \tag{1.25}$$

if the error term η_i has an extreme value distribution independent of the regressors. More specifically, if the baseline hazard $h_0(t) = 1$ then $\Lambda(T_i) = T_i$ and the regression simplifies to

$$\ln(T_i) = -\beta\ell_i + \eta_i. \tag{1.26}$$

This implies that the estimated coefficients of the Cox Proportional model can be interpreted as the effect of a one unit increase of the average length of trials on the logarithm of the duration of the spell of employment inaction.

Under these functional assumptions, using the estimates in the last column of Table 1.9, a 1% increase in firing costs, corresponding to a trial 3 days longer at the median length of trials of 11 months, would raise the average duration of the spell of employment inaction

[19] The hazard ratio is the ratio of the hazard rates corresponding to the conditions described by two levels of an explanatory variable. In my setting the hazard ratio for different levels $\Delta\ell$ of trial lengths is,

$$\frac{h(t|\ell + \Delta\ell)}{h(t|\ell)} = \exp(\beta\Delta\ell).$$

[20] $0.4\% = \exp(\beta_1 * \Delta\ell) - 1 = (\exp(-0.037 * 0.11) - 1) * 100.$

by 0.4%. This means increasing the duration of employment inaction by approximately 1 day, at the median duration of employment inaction of 4 months.

To probe the robustness of my findings with respect to alternative econometric specifications, I use a standard instrumental variable approach in a linear model. Consider the following two–stages equation:

$$\ell_i = \delta_0 + \delta_1 Z_{j(i)} + v_i \qquad \text{first–stage} \tag{1.27}$$

$$\log(T_i) = \beta_0 + \beta_1 \hat{\ell}_i + \varepsilon_i \qquad \text{second–stage} \tag{1.28}$$

where T_i is the number of months until firm i changes its employment level after the end of the trial. $\hat{\ell}_i$ are the fitted values from the first–stage regression. However, model (1.28) does not take into account the difference between the failure event and right censoring. For this reason, I do not consider firms which trial ended in 2013, since 40% of these firms are censored. For the remaining firms censoring ranges from 0 (firms which trials ended in 2001) to 8.8% (firms which trial ended in 2012) with an average censoring of 2.7%. Table A5 reports the estimates from the two stages (1.27) and (1.28). The coefficients in the second columns of Table 1.9 and Table A5 are very similar in absolute value. Reassuringly the results from the control function approach in the Cox Proportional Hazard Model are robust to alternative econometric specifications.

A possible interpretation of my findings is that firms update their believes on firing costs depending on the trial lengths experienced in court. (See Section 1.2.2). But what if two firms experience different trial lengths but observe each other? Then they should not behave differently after their experience in court because by observing each other they acquire exactly the same information. It is indeed a possibility and this kind of information spillover introduces measurement error in my treatment variable, which creates an attenuation bias of my results.

Essentially, it may be that I observe two firms receiving different treatments, (experiencing different trial lengths), not behaving differently after the end of the trial because these two firms know each other and shared their information on their experienced trial lengths.

1.5.4 The effect of firing costs on employment levels

Firing costs inhibit both firing and hiring. Therefore, the net effect of these offsetting factors on employment levels is ambiguous. My randomized experiment allows to understand which factor prevails by identifying the causal effect of firing costs on employment levels.

Normalizing the month at which the trial ends to $t = 0$, my two–stage equations are

$$\ell_i = \delta_0 + \delta_1 Z_{j(i)} + v_i \qquad \text{first–stage} \qquad (1.29)$$

$$\log(n_{it}) = \gamma_t + \alpha_t \hat{\ell}_i + \varepsilon_i \qquad t \in \{1, 2, ..., M\} \qquad \text{second–stage} \qquad (1.30)$$

where n_{it} is monthly employment at firm i at month t. Expected Firing cost are measured as the length of the trials experienced by each firm, ℓ_i. The instrument, $Z_{i(j)}$, is the average length judge j assigned to firm i takes to complete his or her cases, based on all the judge's other cases. $\hat{\ell}_i$ are the fitted values from the first–stage regression.

The estimated coefficients $\hat{\alpha}_1, ..., \hat{\alpha}_M$ flexibly capture the dynamic effect of an increase in firing costs after time $t = 0$, the month at which the trial ends. Because n_{it} are measured in 2001–2013 and the sample consists of firms which trial ended in 2001–2013, the composition of the pooled sample changes somewhat with t. For example and abstracting from plant closure, a firm which trial ended in January 2001 will exit the sample at $t = 156$, whereas a firm which trial ended in January 2013 will exit the sample at $t = 12$. A potential concern is that composition bias gives a distorted view of how employment levels change with time from the trial. To address this concern, I estimate the effect of trial lengths on employment levels when the sample is held fixed. I chose $M = 48$ and hold the sample fixed by considering firms which trials ended between January 2001 and January 2010. This means that I look at firms' monthly employment only in the 48 months following the end of their trial even for firms which trial ended in before January 2010.[21] I also consider an event–study framework and impose the restriction that $\alpha_t = \alpha$ for all $t = 1, ..., 48$. The event–study estimates increase statistical power and allows to present the findings in a more parsimonious way.

Figure 1.9 graphically depicts the coefficient estimates from equation (1.30), along with 95% confidence intervals, for the first 4 years after the end of the trial. After the end of the trial firms that have experienced longer trials have higher employment levels, suggesting that higher expected firing costs increase average employment levels. Employment levels are 3% higher for each extra month of duration of trial. Table 1.10 reports the event–study estimates for the $t = 1, ..., 48$ horizon

To interpret these empirical findings keep in mind the discussion of section 1.2.1. The lower is firms' discount factors the more likely are firing costs to increase average employment levels. Suppose there is an expansion, the firm knows that because of a future recession, workers may one day have a low marginal product of labor and firing costs will have to be paid to get rid of these workers, but this possibility is discounted since hiring occurs in good times, and bad times are far into the future. Expost, when the recession comes, firing is less likely to occur due to firing costs, hence average employment increases.

[21] Figure A2, reports estimates of equations (1.29) and (1.30) for different time horizons (different M) to assess the robustness of the results.

1.5.5 Heterogeneous effects

I conduct a number of analysis to explore whether the effect of firing costs on employment inaction differs across subgroups of firms. A possible interpretation of the effect presented in in section 1.5 is that firms learn trial lengths by going to court. The following heterogeneous effects support this interpretation of the results. Moreover, I rule out that employment inaction is caused by financial constraints rather than changes in firms' expectations on firing costs.

Firms' ages

The effect of experienced trial lengths on future firms' employment inaction can be interpreted as firms changing their expectations on the trial lengths, hence their expectations on firing costs. This interpretation suggests that firms with less informative priors on the distribution of trial lengths update more their believes on trial lengths, a point formalized in Section 1.2.3. Figure 1.12 reports the estimates of equations (1.23)–(1.22) for different subgroups of firms determined by the age of the firm at the end of its trial. According to these results, younger firms respond more to the same increase in expected firing costs compared to older firms. This finding supports the idea that firms learn the degree of firing costs by experiencing trial lengths. Presumably younger firms have more to learn about trial lengths than older firms, hence younger firms react more to the same new information about trial lengths.

Note, however, that the cleanest identification is achieved for firms born after 2001 because only for this subgroup of firms I can claim to be using the first trial ever experienced by firms as treatment. One may worry that in Figure 1.12 the effect is different between subgroups of firms because the effect for younger firms is better identified than for older firms. Figure 1.13 shows that even within the subgroup of firms born after 2001, young firms react more than old firms to the same experience in court.

Firing cases and other types of cases

Firms go to court for different reasons, not only when they fire a worker. Table 1.5 reports the distribution of the types of trials experienced by firms. If firms learn trial lengths with their experience in court, then their expectations of firings costs should be affected by any experience in court, regardless of the type trial. Table 1.11 reports estimates from equations (1.22)–(1.23) estimated for the two subgroups of firms experiencing non–firing and firing cases. The effect is similar in the two groups of firms and no statistically significant difference exists.

Stricter legislation for larger firms and tax component of firing costs

Firms employing more than 15 employees are subject to a stricter employment protection legislation because if the judge rules in favor of the worker, then the firm has to pay all forgone wages from the day of dismissal to the day of court ruling. Whereas firms employing 15 or less employees have to pay a severance payment decided by the judge of at least 2.5 and at most 6 months forgone wages. Smaller firms may still prefer shorter trials due to legal and organizational costs, however, for larger firms the costs of long trials are even higher.

Moreover, considering the subgroup of firms employing 15 or less employees allows to identify exclusively the tax component of firing costs. In fact, for this subgroup of firms the transfer component of firing costs does not depend on the trial length, whereas the tax component does in the form of a higher legal costs and longer periods of uncertainty.

Table 1.12 investigates if there are heterogenous effects with respect to subgroups of firms with different size. Firms' size is measured as the monthly average number of employees at the firm during the year before the firm goes to court. The results show that firms just above the 15 employees threshold, that is firms employing between 16 and 24 employees, have the largest effect among all subgroups of firms. These firms are sufficiency big to be subject to the stricter employment protection legislation and, at the same time, they are sufficiently small to be affected by one experience in court. Presumably, larger firms, although subject to the stricter legislation, are not affected by one single trial because, as Table 1.12 reports, they go to court more often and their marginal cost of a trial may be smaller, for example if they have a legal office within the firm.

1.5.6 Robustness check: financial constraints do not matter for inaction

In this section, I rule out that firms' employment inaction is due to the fact that firms have no resources to change their employment level after having paid for a costly long trial. In a world with perfect capital markets, this would not be a concern because realized trials' costs are sunk and hence they cannot affect future employment decisions of firms. Therefore, any effect of trial lengths must be due to the fact that firms' expectations of firing costs have changed because of the long experienced trial. However, without perfect capital markets firms may not have the resources to adjust their labor force if they just had to pay for a costly long trial. To assess this hypothesis, I investigate if there are heterogeneous effects with respect to various measures of firms' financial constraints.

The database CERVED[22] measures the available liquidity[23] at each firm in each fiscal year. This variable allows to construct a proxy of firms' financial constraints as the available

[22]CERVED collects annual balance sheet data for the universe of Italian firms, refer to Section 1.3.1 for a description of the data.

[23]Available liquidity is a balance sheet variable which measures the monetary resources of the firm, for example its lines of credit and its "cash".

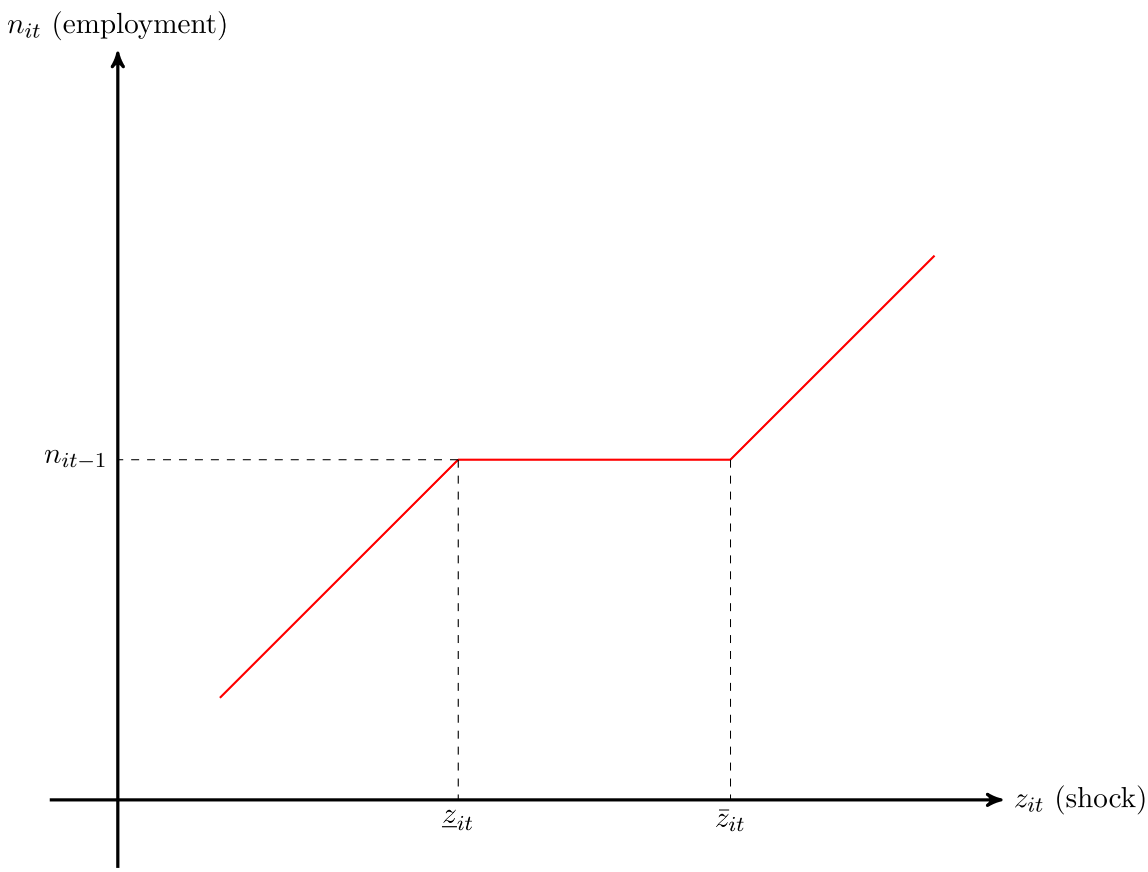
liquidity standardized by firm's size. I use two definitions of firm's size: the total assets of firms and the number of employees. Standardized available liquidity, measured the year before the firm goes to court, represents an appropriate moderator variable because it is determined before the treatment. Figure 1.11 reports estimates from equations (1.22)–(1.23) estimated in different subpopulations of firms, determined by the different quantiles of pre–treatment standardized available liquidity. Figure 1.10 and 1.10 show that the effect is similar across different quantiles of available liquidity, and since the confidence intervals of the estimated parameters overlap I cannot reject the null hypothesis of equal effects.

1.6 Conclusions

This paper provides evidence on the effect of firing costs on firms hiring and firing decisions. The key to my research design is that the Italian labor court system randomly assigns judges to firms in litigation. Some judges are systematically slower, which leads to random variation in the trial length a firm will experience. This may affect firms expected firing costs, given that in the Italian context longer trials imply higher firing costs. Then, I use this exogenous variation to examine the effects of firing costs on net employment changes, where there are unambiguous theoretical implications, and on employment levels where the predictions of theory are less clear cut. I find strong evidence that higher expected firing costs, measured as longer experienced trials, cause a reduction in the hazard of a variation of firm employment and an increase in employment levels after the end of the trial. These effects are larger in size for younger firms which, given less experience, may have less precise information on the exact length of trials and therefore react more to the newly acquired information with their experiences in court.

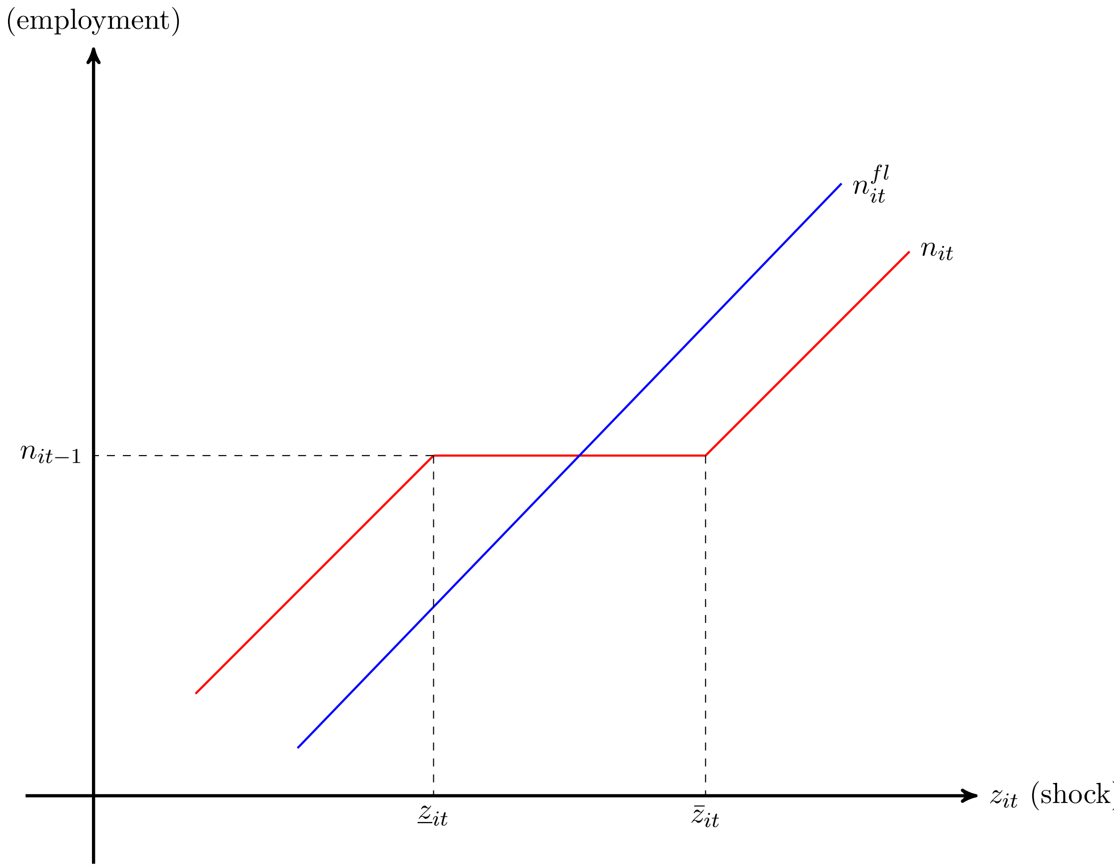
Trade unions often advocate the use of firing costs to protect jobs. Indeed, stricter employment protection decreases jobs flows and, as my results show, increases employment. However, unless a particular welfare function is considered or another friction in the economy is assumed, these lower employment volatility and higher employment levels are inefficient. Assessing the effects of employment protection legislation on welfare defines an agenda for future work.

Figure 1.1: Employment Inaction



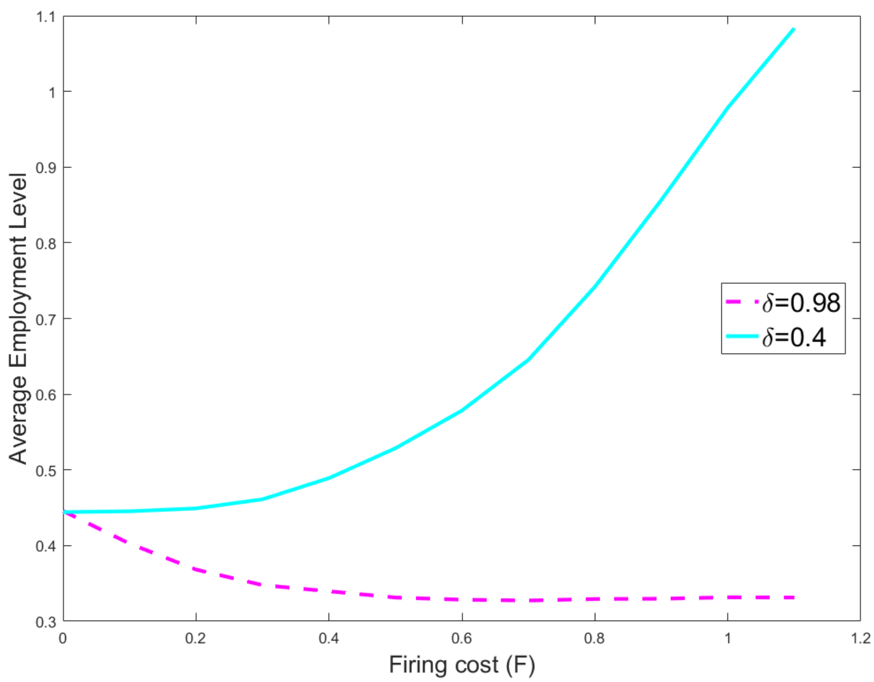
Note: The figure represents the optimal employment choice of firm i at time t for different values of the realization of the shock z_{it}, given the employment level of the firm at time $t-1$. Firing costs create an inaction range, $[\underline{z}_{it}, \bar{z}_{it}]$, which size is proportional to the degree of firing costs.

Figure 1.2: Employment Inaction and Efficiency



Note: The figure compares the optimal firm's employment choices with non zero firing costs, n_{it}, and with zero firing costs, n_{it}^{fl}. The absolute value of the vertical difference between the red and the blue line measures the inefficiency introduced by firing costs.

Figure 1.3: Firing costs and employment levels



Note: The figure reports the numerical solution of the model defined in equation (1.2). The figure shows how average employment levels change with firing costs, depending on the values of the firm's discount factor.

Table 1.1: There are more hearings for longer trials

Dependent variable	Log(trial lengths (months))
Number of hearings	0.1385***
	(0.0007)
Observations	320191

Note: The table shows that there is a positive correlation between the number of hearings needed to complete a trial and the length of the trial. There are on average 3 hearings for each trial with a standard deviation of 2. Standard errors in parenthesis are robust to heteroscedasticity. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.2: Outcome and length of trials are not positively correlated

Dependent variable	Trial's outcome	Trial's outcome
Sample	Only firms match emp. data	All firms
	(1)	(2)
trial lengths	-0.0085	-0.0050
	(0.0062)	(0.0060)
Observations	3,865	41,742

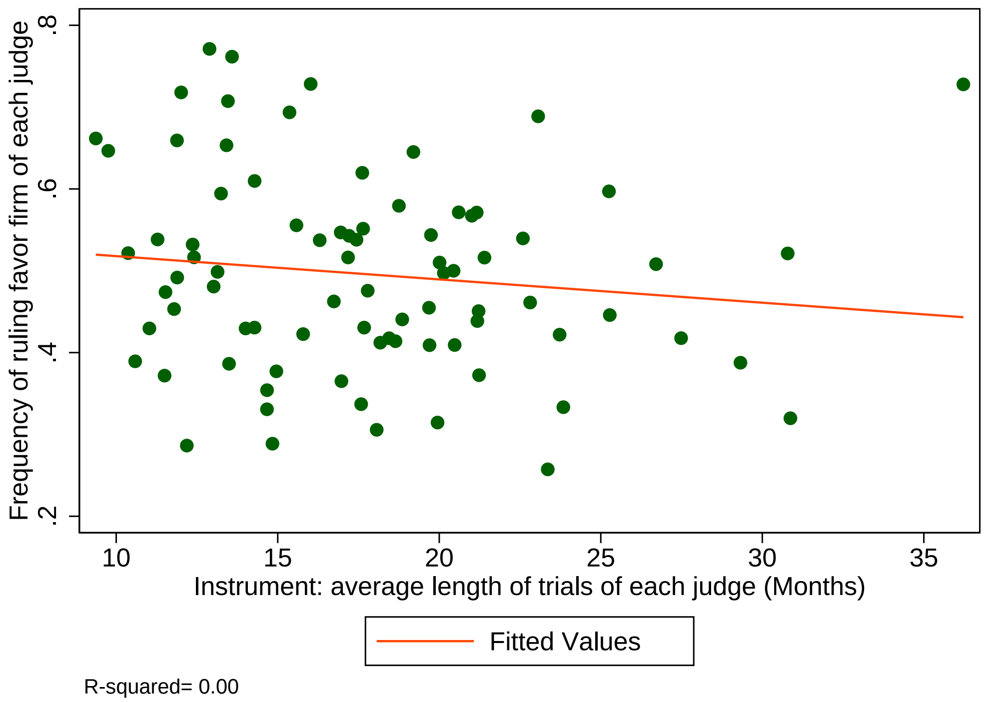
Note: The table shows that there is no positive correlation between the length and the outcome of trials. The sample is restricted to trials that ended with a sentence by the judge because only for this subgroup of trials it is possible to determine the trial outcome. For all the remaining trials, those ending with a settlements, it is not possible to determine the outcome of the settlement. The estimates in the table are from the following linear probability model:

$$y_{ij} = \alpha_0 + \alpha_1 \ell_i + u_i$$

$$y_{ij} = \begin{cases} 1 & \text{if judge } j \text{ in trial } i \text{ ruled in favor of the firm} \\ 0 & \text{otherwise} \end{cases}$$

and ℓ_i is trial length experienced by firm i. Column (1) refers to the subgroup of firms for which employment data is available, whereas column (2) refers to all firms in the court database. Standard errors in parentheses are clustered at the judge level.

Figure 1.4: Slow judges are not more likely to rule in favor of firms



Notes: The figure shows that judges taking on average long to complete their cases are not also more likely to rule in favor of firms. The fitted values are from the following regression:

$$R_j = \phi_0 + \phi_1 \tilde{Z}_j + \varepsilon_j$$

where R_j is the frequency judge j rules in favor of the firm and $\tilde{Z}_j$ is the average length judge j takes to complete his/her cases bases on all the cases assigned to the judge.

Table 1.3: Distribution of trial length and judges average trial length

Percentiles	Judges average length (months). All trials.	Trial length (months). Only firms trials.
1st	9	0.33
5th	11	2
10th	12	4
25th	13	7
50th	**18**	**11**
75th	21	19
90th	24	28
95th	28	35
99th	37	47
Mean	18	14
Standard deviation	5	10
Number of judges	82	82
Number of trials	320191	7617

Note: trial length is the duration of the trial, measured in month, from the filing to the disposition of the case. Judges average length is the mean length of all trials assigned to each judge.

Table 1.4: Distribution of firms average employment levels and inaction

Percentiles	Firms average employment (number of employees)	Firms duration employment inaction (months)
1st	1	2
5th	1	2
10th	1	2
25th	2	2
50th	**6**	**4**
75th	14	8
90th	55	14
95th	139	23
99th	830	52
Mean	74	7
Standard deviation	1041	10
Number of firms	7617	7617

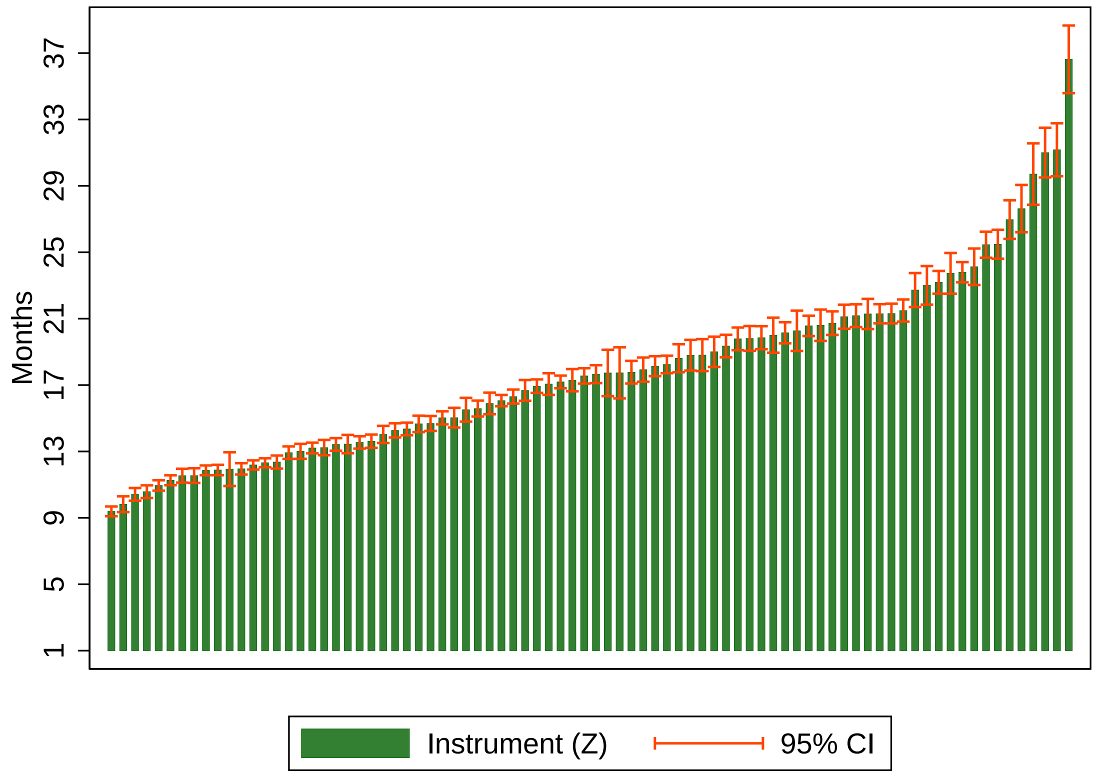
Note: Firms average employment is the average of monthly employment at each firm in 2001-2013. Firms duration of employment inaction is the time, measured in months, until a firm changes its employment level after the end of the trial.

Table 1.5: Distribution of firms' types of trials

Type of trial	Number of trials	Percentage (%)
Compensation	2325	30.52
Attendance allowance	1	0.01
Disability living allowance	3	0.04
Pension	1	0.01
Temporary work contract	243	3.19
Termination of employment	2433	31.94
Type of employment relationship	384	5.04
Other types of cases	2227	29.24
Total	7617	100

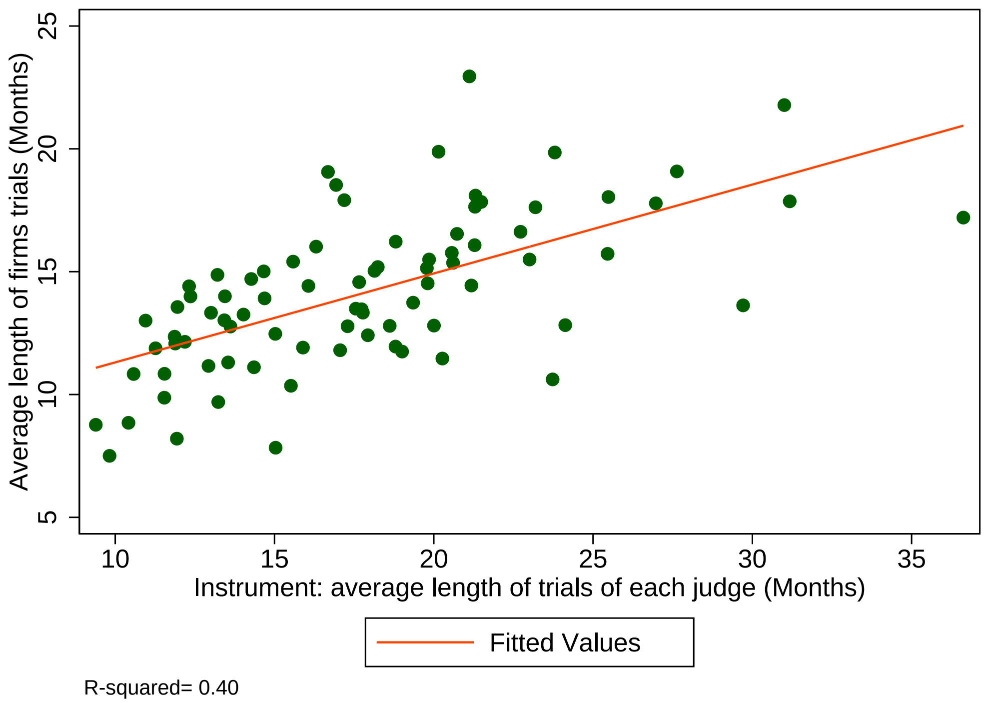
Note: The table reports the distribution of the types of trials of the 7617 firms used in the empirical analysis.

Figure 1.5: Average length of trials assigned to each judge



Notes: The figure shows a graphical representation of the instrument, the average number of months that each judge takes to complete his/her cases, ordered from left to right from the fastest to the slowest judge. Each vertical bar refers to one judge. The height of the bar measures the average length (in months) of the trials assigned to each judge. Red vertical lines show 95% confidence intervals. Only judges that were observed in at least 1,000 trials are considered.

Figure 1.6: First stage



Notes: The figure shows that there is a first stage. That is, a positive correlation between the average length of trials involving firms assigned to each judge and judges average trial lengths based on all other judges trials (trials not involving firms). The fitted values are from the following regression:

$$L_j = \psi_0 + \psi_1 Z_j + \varepsilon_j$$

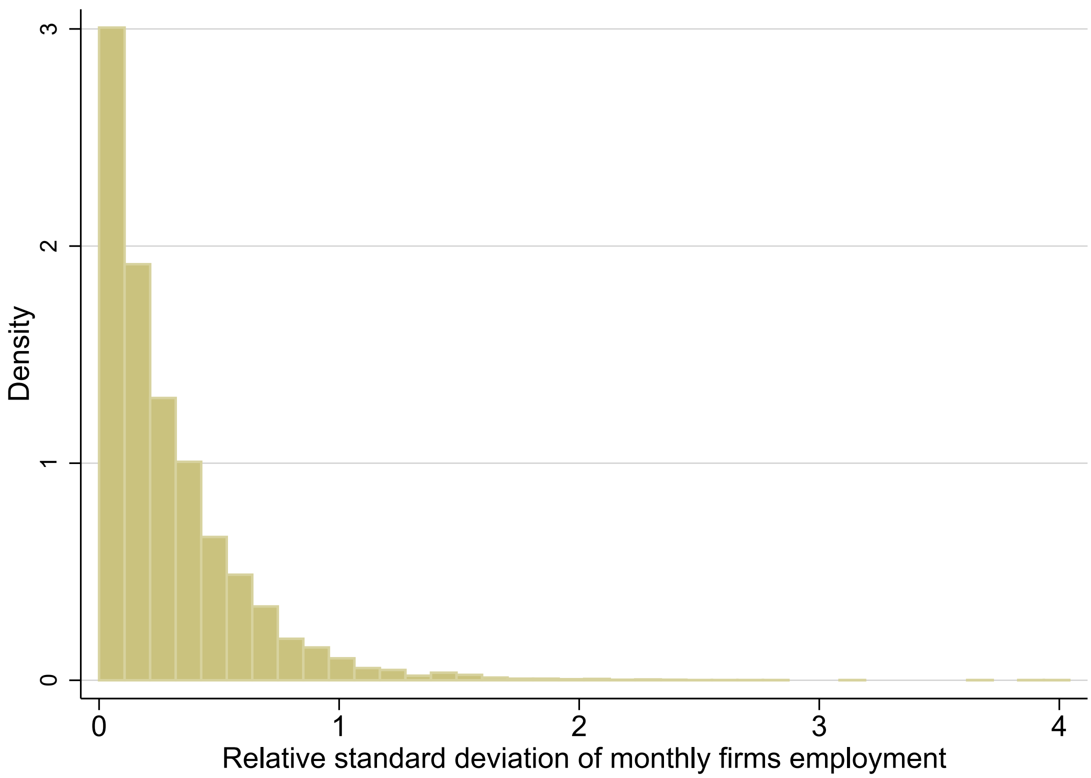
where L_j is the average length of trials involving firms assigned to judge j and Z_j is the average length that judge j takes to complete his/her cases bases on all other assigned cases to the judge.

Table 1.6: The variance of firms level monthly employment

	Mean	Standard deviation	Standard deviation between	Standard deviation within	Observations	Firms
Unbalanced full sample	116	1,581	1,044	282	363,141	7,617
12 months balanced sample	83	1,152	1,138	177	74,280	6,190
24 months balanced sample	95	1,314	1,298	202	56,736	4,728
36 months balanced sample	94	1,401	1,384	219	45,312	3,776
48 months balanced sample	98	1,457	1,437	241	37,128	3,094
60 months balanced sample	111	1,599	1,577	264	30,768	2,564
72 months balanced sample	133	1,781	1,757	294	24,768	2,064
84 months balanced sample	162	2,002	1,975	331	19,572	1,631

Note: The table reports summary statistics for monthly firm level employment for the post trial period of every firm considered in the analysis. The within firm standard deviation is especially important for the analysis since firing costs are binding only if firms need to change employment.

Figure 1.7: Firms monthly employment variation



Note: The figure reports the distribution of the within relative standard deviation (the ratio of the standard deviation to the mean) of firm level monthly employment after the end of the trial, computed for each firm.

Table 1.7: Positive and negative firm level shocks

	Relative frequency (Average of all firms)	Number of firms
Positive employment change (full sample)	0.1180	7,521
Negative employment change (full sample)	0.1573	7,521
Positive employment change (12 months balanced sample)	0.1280	6,190
Negative employment change (12 months balanced sample)	0.1650	6,190
Positive employment change (24 months balanced sample)	0.1359	4,728
Negative employment change (24 months balanced sample)	0.1675	4,728
Positive employment change (36 months balanced sample)	0.1407	3,776
Negative employment change (36 months balanced sample)	0.1694	3,776
Positive employment change (48 months balanced sample)	0.1432	3,094
Negative employment change (48 months balanced sample)	0.1676	3,094
Positive employment change (60 months balanced sample)	0.1460	2,564
Negative employment change (60 months balanced sample)	0.1699	2,564
Positive employment change (72 months balanced sample)	0.1530	2,064
Negative employment change (72 months balanced sample)	0.1742	2,064
Positive employment change (84 months balanced sample)	0.1581	1,631
Negative employment change (84 months balanced sample)	0.1800	1,631

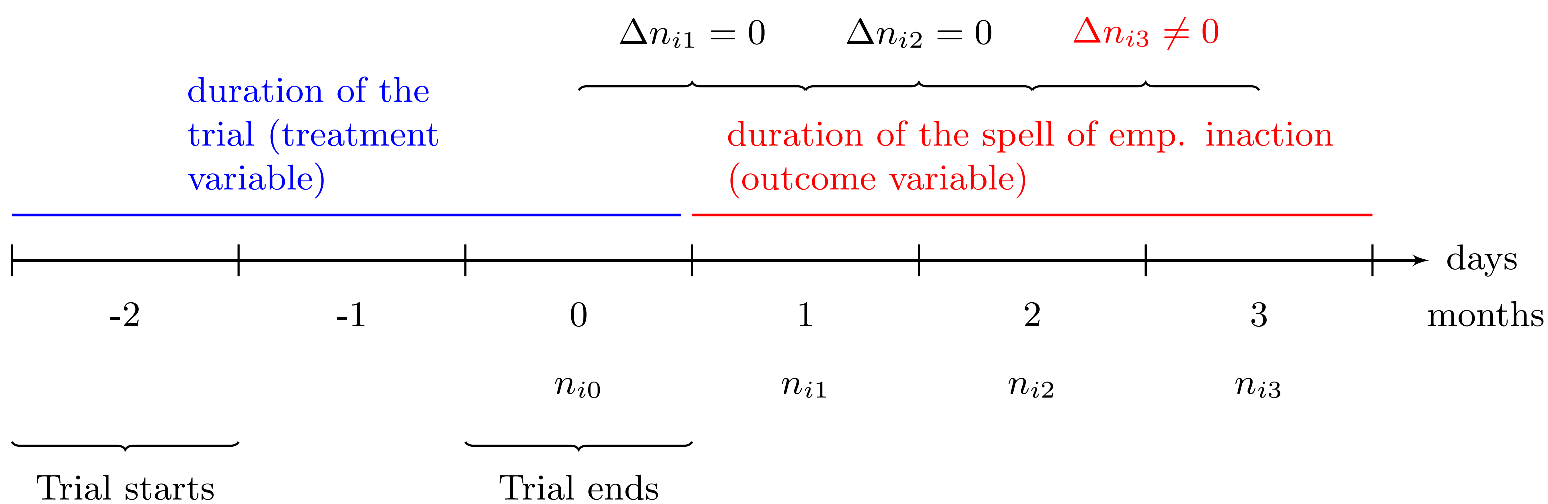
Note: The table reports the number of months firms increased or decreased their employment levels between two consecutive months following the end of the trial, normalized by the total number of months each firm is observed and averaged over all firms.

Table 1.8: Testing for Random Assignment of Cases to Judges

	ln(Judge avg. Length)	
ln(Firms Variables)	coeff.	s.e.
Revenue	0.0002	(0.0017)
Cost of labor	0.0008	(0.0017)
Cash flow	-0.0006	(0.0018)
Liquidity	-0.0009	(0.0015)
Assets	0.0006	(0.0016)
Capital	-0.0005	(0.0016)
Investments	0.0005	(0.0022)
Return on equity	0.0030	(0.0027)
Return on asset	0.0001	(0.0034)
Value added	0.0015	(0.0018)
employment	0.0020	(0.0017)
Firms' sector dummies		
Sector 1 dummy	-0.0280	(0.0468)
Sector 2 dummy	0.0548	(0.0515)
Sector 3 dummy	-0.0043	(0.0145)
Sector 4 dummy	0.0528	(0.0390)
Sector 5 dummy	-0.0005	(0.0141)
Sector 6 dummy	-0.0024	(0.0130)
Sector 7 dummy	0.00003	(0.0133)
Sector 8 dummy	-0.0108	(0.0169)
Sector 9 dummy	-0.0069	(0.0231)
Sector 10 dummy	0.0103	(0.0137)
Sector 11 dummy	0.0444	(0.0469)
Sector 12 dummy	0.0155	(0.0274)
Sector 13 dummy	0.0089	(0.0188)
R–sq from complete regression	0.16	
F–statistic for joint significance	0.80	
[p–value]	[0.6355]	
Observations	7617	

Note: This table displays the test of whether the court compiled with the random allocation of cases to judge. There are 82 judges. Each row of the first panel refers to OLS estimates from separate regressions of the instrument on firms' characteristics. As described in Section 1.4.2, the instrument is the average length that judges take to complete their assigned cases. Characteristics of firms are measured the year before firms go to court. The second panel reports OLS estimates of the instrument on 13 dummies for firms' sectors. The R–sq and F–statistic in the third panel are obtained from OLS estimation on the combined set of firms' characteristics.

Figure 1.8: Time line: the trial ends on a day in month 0



Note: Vertical tick marks indicate the first day of each month. The trial of each firm ends on a day in month 0. For example if a trial ends on July the 19th 2000, then month 0 is July 2000. Employment, n_{it}, is the monthly employment in month t at firm i. Employment change, Δn_{it}, is employment change in month t with respect to month $t-1$. The survival analysis starts from the month in which the trial ends. For example, in this figure the failure event happened at time 4. The firm changed employment 4 month after the end of the trial.

Table 1.9: The effect of firing costs on the hazard of employment action, Control Function

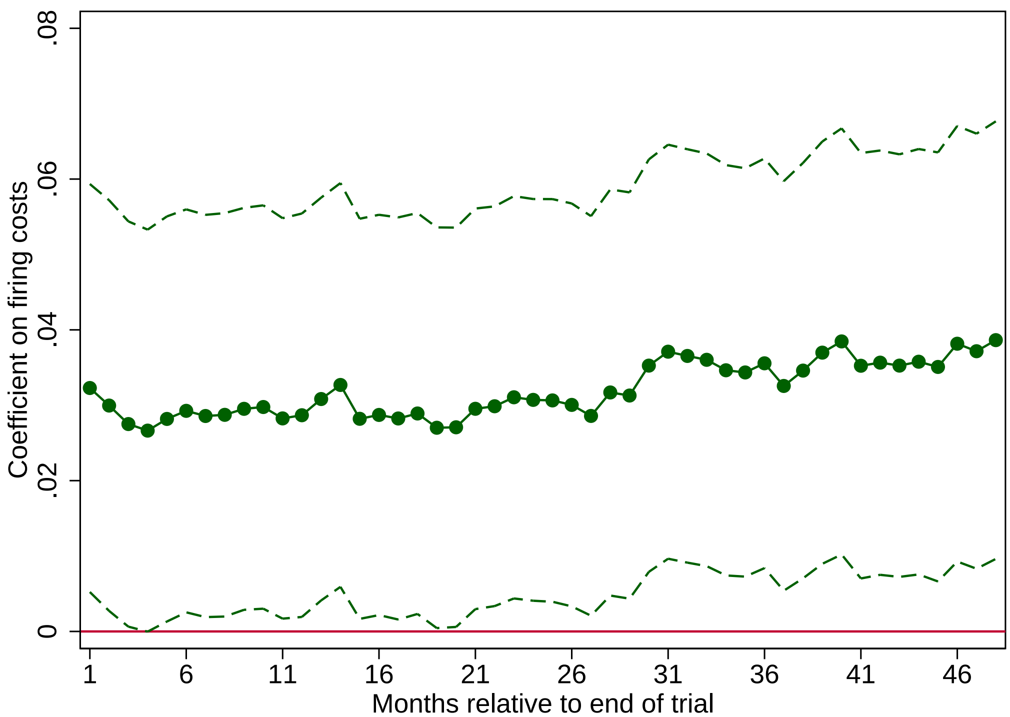
Dependent variable	Trial length	$h(t\|X)$
Estimation method	OLS	ML
Stage	First	Second
	(1)	(2)
Trial length		-0.0370***
		(0.0059)
		[0.0059]
Judge avg. length	0.4110***	
	(0.0257)	
Cragg–Donald Wald F statistic		256
Observations	7617	7617

Note: To implement instrumental variables in a Cox Proportional Hazard Model I use a control function approach. The table reports the estimates from the two stages:

$$\ell_i = \delta_0 + \delta_1 Z_{j(i)} + v_i \qquad \text{first–stage}$$
$$h_{it} = h_0(t)\text{exp}\{\beta\ell_i + g(v_i)\} \qquad \text{second–stage}$$

Expected Firing costs are measured as the length of the trials (measured in months) experienced by each firm, ℓ_i. The instrument, $Z_{i(j)}$, is the average length judge j assigned to firm i takes to complete his or her cases, based on all the judge's other cases. $g(v_i)$ is a fifth order polynomial in the estimated residuals v_i from the first stage regression. h_{it} is the hazard of a variation of firm employment t months after the end of the trial. $h_0(t)$ is the baseline hazard. Standard errors in round parentheses are clustered at the judge level in column (1). Standard errors in round parentheses are robust to heteroscedasticity and in squared parentheses are bootstrapped with 100 repetitions in column (2). * significant at 10%, ** significant at 5%, *** significant at 1%.

Figure 1.9: Effect of firing costs on employment levels (48–months fixed sample)



Notes: This figure reports estimates from the following two–stage least–squares:

$$\ell_i = \delta_0 + \delta_1 Z_{j(i)} + v_i \qquad \text{first–stage}$$

$$\log(n_{it}) = \alpha_{0t} + \alpha_{1t}\hat{\ell}_i + \varepsilon_i \qquad t \in \{1, 2, ..., 48\} \qquad \text{second–stage}$$

Expected Firing costs are measured as the length of the trials (measured in months) experienced by each firm, ℓ_i. The instrument, $Z_{i(j)}$, is the average length judge j assigned to firm i takes to complete his or her cases, based on all the judge's other cases. n_{it} is the monthly employment at firm i in month t following the end of the trial. $\hat{\ell}_i$ are the fitted values from the first–stage regression. Estimates of α_{1t} reported in the figure are from separate regressions for each month t and the dashed lines show 95% confidence intervals. The sample of firms is held fixed by considering only firms which trials ended between January 2001 and January 2010.

Table 1.10: Firing costs increase average employment levels (event–study)

Dependent variable	Trial length	ln(Employment)
Estimation method	OLS	IV
Stage	First	Second
	(1)	(2)
Trial length		0.0319**
		(0.0134)
Judge avg. length	0.4054***	
	(0.0427)	
Cragg–Donald Wald F statistic		96
Observations	3094	148512
Number of firms	3094	3094

Note: I use a two–stage least–squares procedure in a linear model. The table reports the estimates from the two stages:

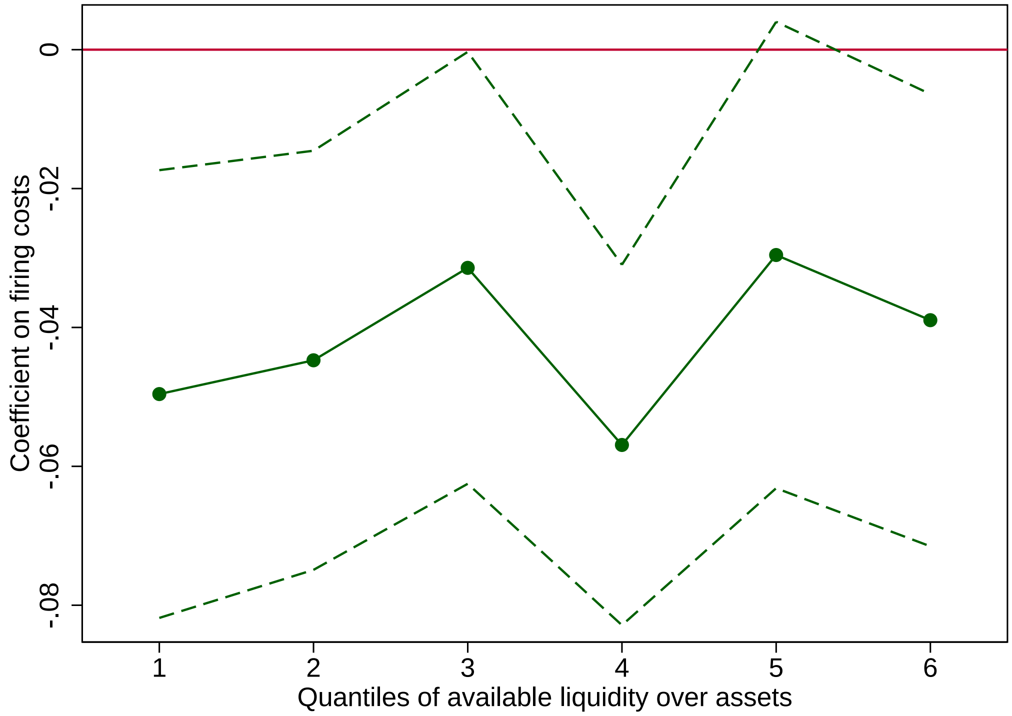
$$\ell_i = \delta_0 + \delta_1 Z_{j(i)} + v_i \qquad \text{first–stage}$$

$$\log(n_{it}) = \alpha_0 + \alpha_1 \hat{\ell}_i + \varepsilon_i \qquad t \in \{1, ..., 48\} \qquad \text{second–stage}$$

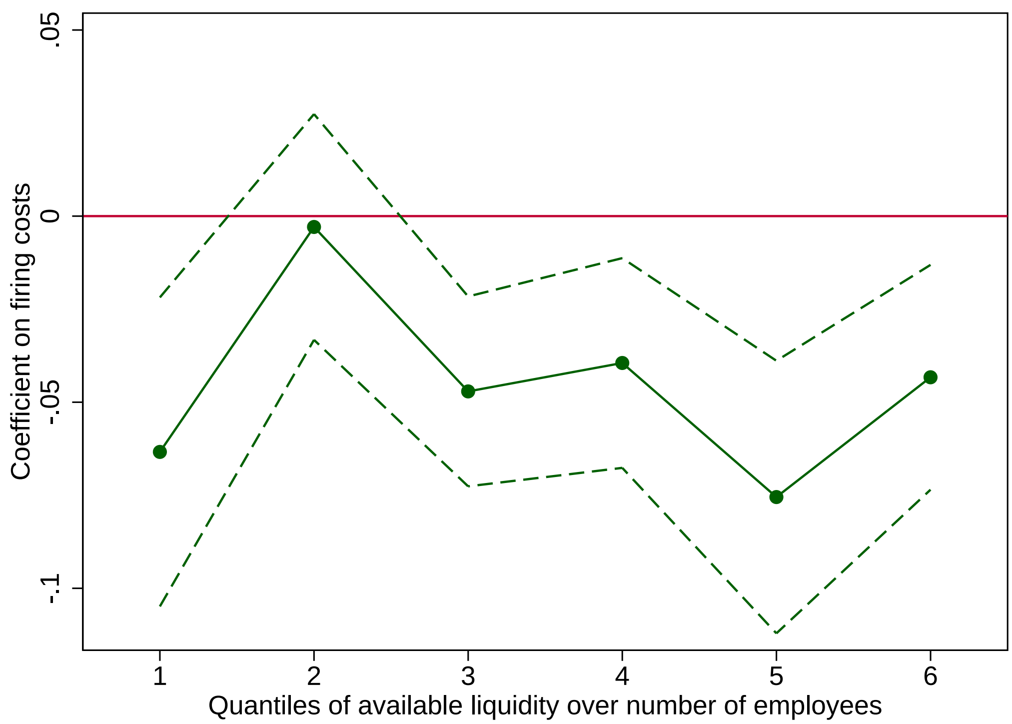
Expected Firing cost are measured as the length of the trials (measured in months) experienced by each firm, ℓ_i. The instrument, $Z_{i(j)}$, is the average length judge j assigned to firm i takes to complete his or her cases, based on all the judge's other cases. n_{it} is monthly employment at firm i in month t after the end of the trial. $\hat{\ell}_i$ are the fitted values from the first–stage regression. The sample of firms is held fixed by considering only firms which trials ended between January 2001 and January 2010. Standard errors in parentheses are clustered at the judge level in column (1) and at the firm level in column (2). * significant at 10%, ** significant at 5%, *** significant at 1%.

Figure 1.10: Heterogeneity by financial constraints

Financial constraints measured as available liquidity over assets



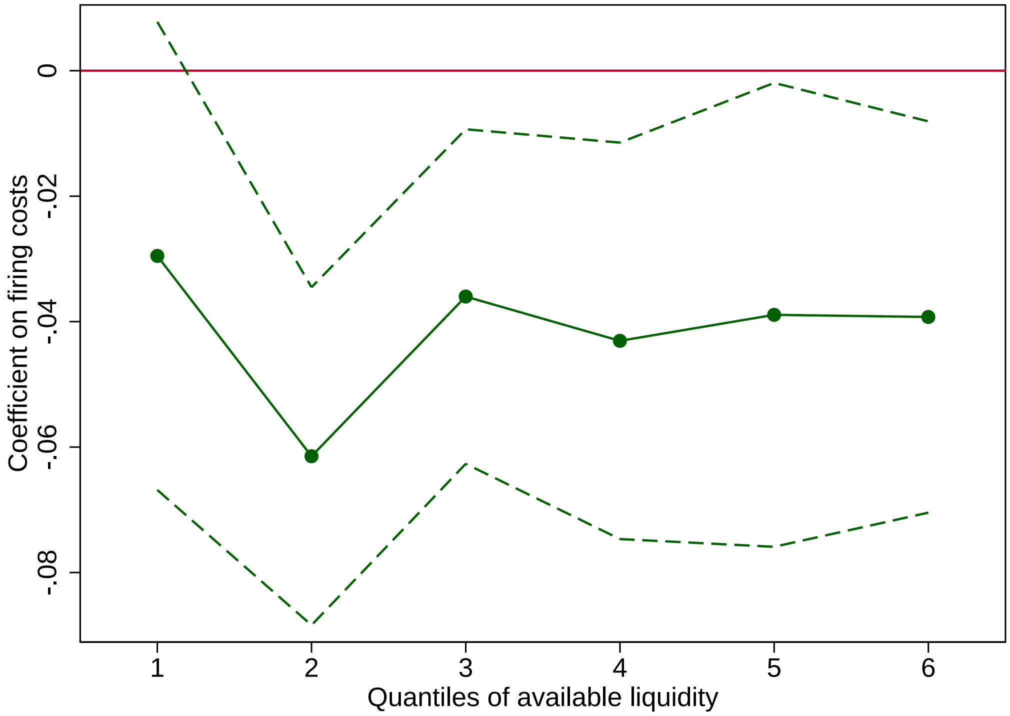
Financial constraints measured as available liquidity over employees



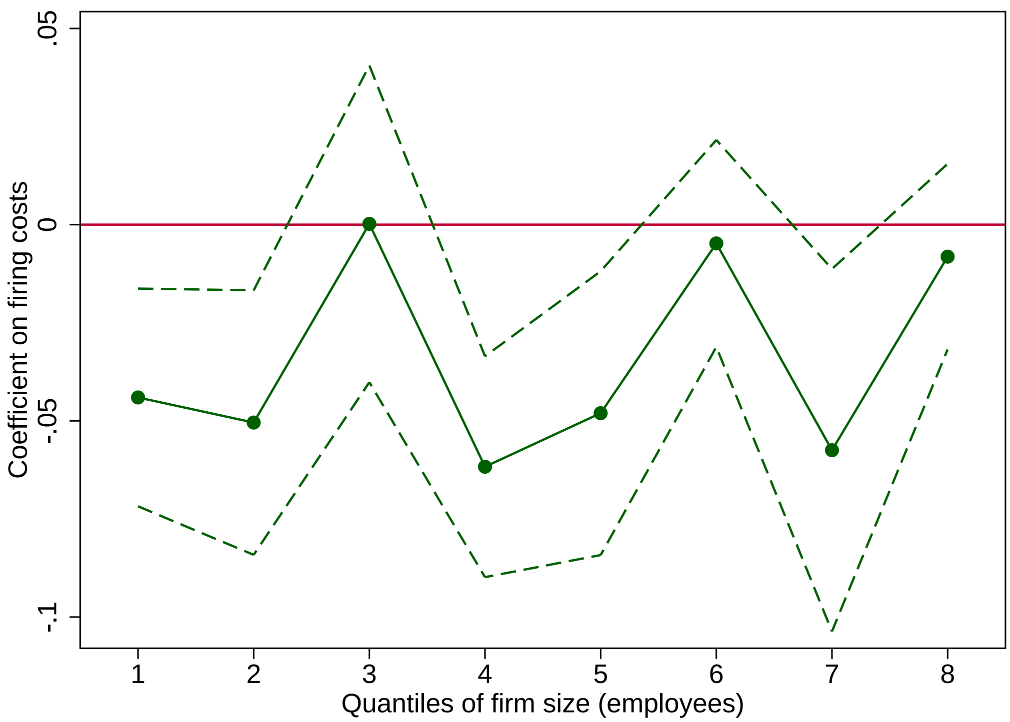
Note: The figure reports estimates from the two stages equations (1.22) and (1.23) estimated in different subpopulations of firms. Each quantile corresponds to a separate estimation and the dashed lines show 95% confidence intervals. Quantiles represent subpopulations with different levels of pre–treatment available liquidity over assets in figure 1.10 and over number of employees 1.10. Quantiles are reported in ascending order, hence the 1st quantile refers to firms more financially constraint whereas the 6th quantile refers to firms less financially constraint. Refer to Tables A6 and A7 for details of the results reported in the figures.

Figure 1.11: Heterogeneity by financial constraints and firm size

Financial constraints measured as available liquidity

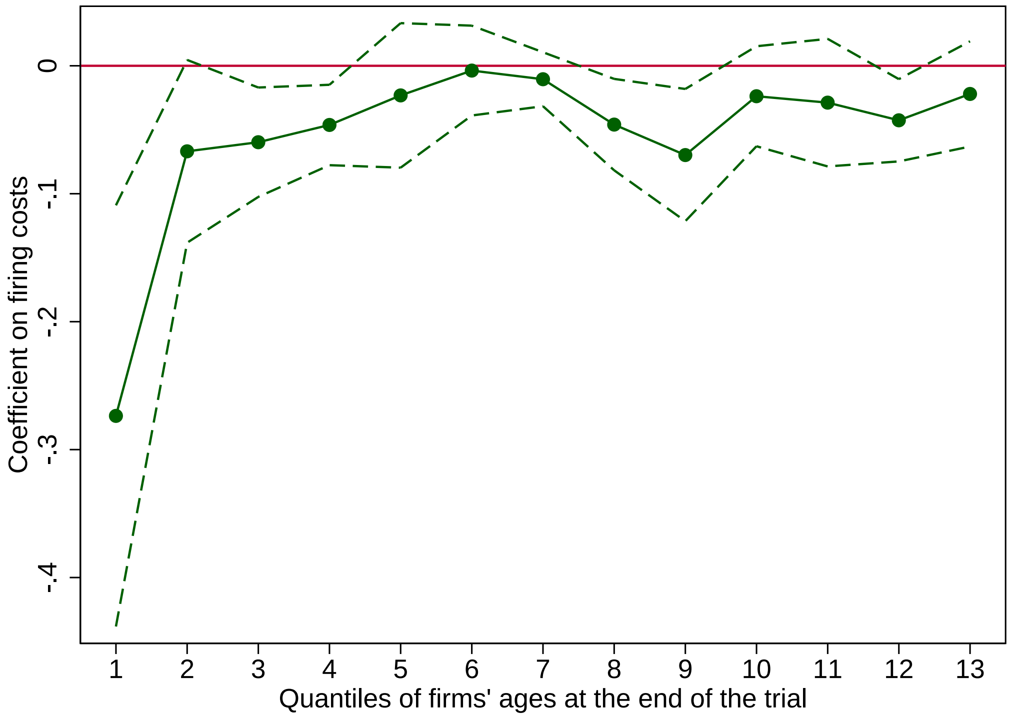


Size of the firm measured as number of employees



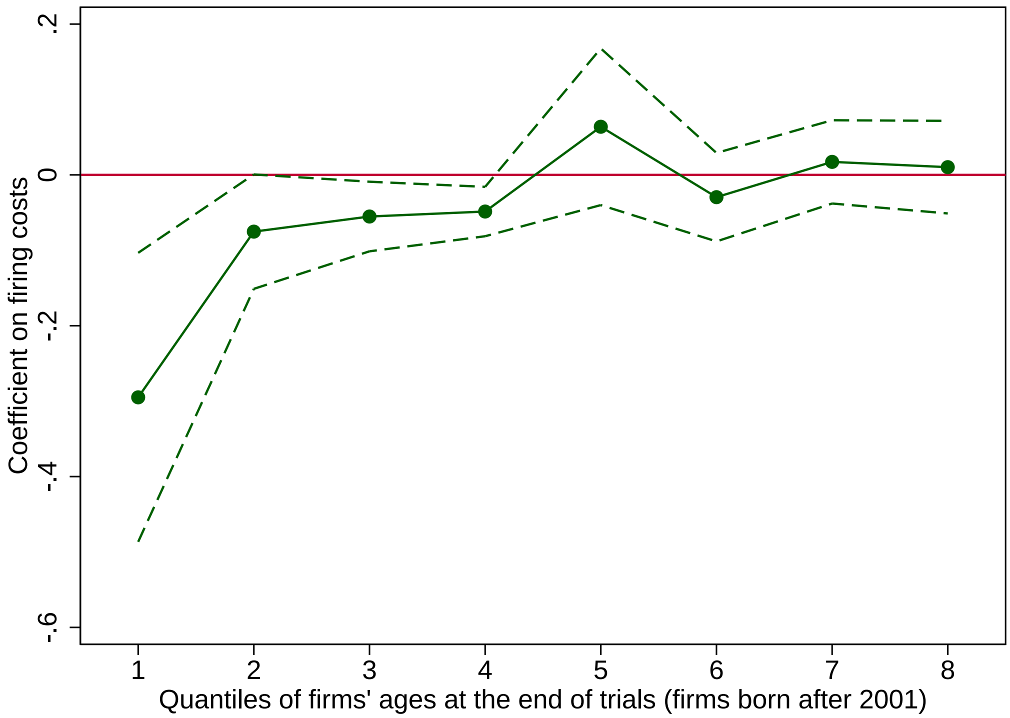
Note: The figure reports estimates from the two stages equations (1.22) and (1.23) estimated in different subpopulations of firms. Each quantile corresponds to a separate estimation and the dashed lines show 95% confidence intervals. Quantiles represent subpopulations with different levels of pre–treatment available liquidity in figure 1.11 and of number of employees 1.11. Quantiles are reported in ascending order.

Figure 1.12: Heterogeneity by firm age



Note: The figure reports estimates from the two stages equations (1.22) and (1.23) estimated in different subpopulations. Each quantile corresponds to a separate estimation and the dashed lines show 95% confidence intervals. Quantiles represent subpopulations with different firms' ages at the end of trials. Quantiles are reported in ascending order, hence the 1st quantile refers to younger firms whereas the 13th quantile refers to older firms. The 1st quantile contains firms that are at least 1 year old and at most 2 years old when trials end, the 2nd quantile firms 3 years old, the 3rd quantile firms 4 years old, the 4th quantile firms 5 years old, the 5th quantile firms 6 years old, the 6th quantile firms 7 years old, the 7th quantile firms that are at least 8 years old and at most 9 years old, the 8th quantile firms that are at least 10 years old and at most 11 years old, the 9th quantile firms that are at least 12 years old and at most 14 years old, the 10th quantile firms that are at least 15 years old and at most 18 years old, the 11th quantile firms that are at least 19 years old and at most 23 years old, the 12th quantile firms that are at least 24 years old and at most 31 years old, the 13th quantile firms that are at least 32 years old and at most 56 years old. Refer to Table A8 for details of the results reported in the figure.

Figure 1.13: Heterogeneity by firm age (firms born after 2001)



Note: The figure reports estimates from the two stages equations (1.22) and (1.23) estimated in different subpopulations. Each quantile corresponds to a separate estimation and the dashed lines show 95% confidence intervals. Quantiles represent subpopulations with different firms' ages at the end of trials, within the subgroup of firms born after 2001. Quantiles are reported in ascending order, hence the 1st quantile refers to younger firms whereas the 8th quantile refers to older firms. The 1st quantile contains firms that are at least 1 year old and at most 2 years old when trials end, the 2nd quantile firms 3 years old, the 3rd quantile firms 4 years old, the 4th quantile firms 5 years old, the 5th quantile firms 6 years old, the 6th quantile firms 7 years old, the 7th quantile firms that are at least 8 years old and at most 9 years old, the 8th quantile firms that are at least 10 years old and at most 12 years old. Refer to Table A9 for details of the results reported in the figure.

Table 1.11: Heterogeneity by type of trial

Type of trial	Non–firing	Firing
	(1)	(2)
Dependent variable	$h(t\|X)$	$h(t\|X)$
Estimation method	ML	ML
Stage	Second	Second
Trial length	-0.0373***	-0.0409***
	(0.0067)	(0.0116)
	[0.0069]	[0.0120]
Dependent variable	Trial length	Trial length
Estimation method	OLS	OLS
Stage	First	First
Judges avg. length	0.4317***	0.3729***
	(0.0319)	(0.0429)
Cragg–Donald Wald F stat.	183	76
Observations	5184	2433

The table reports estimates from the two stages equations (1.22)–(1.23) estimated in different subpopulations. Each column corresponds to a separate estimation. Column (1) refers to the subpopulation of firms experiencing trials not related to the termination of an employee, whereas column (2) refers to firms experiencing trials related to the termination of an employee. Standard errors in round parentheses are clustered at the judge level in the second panel. Standard errors in round parentheses are robust to heteroscedasticity and in squared parentheses are bootstrapped with 100 repetitions in the first panel. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 1.12: Heterogeneity just above the 15 employees threshold

Number of employees intervals	(0,15]	(15,24]	(24,43]	(43,107]	(107,74744]
Average number of trials experienced	2	3	3	5	23
Dependent variable	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$
Estimation method	ML	ML	ML	ML	ML
Stage	Second	Second	Second	Second	Second
Trial length	-0.0394***	-0.0903**	-0.0385	-0.0081	-0.0127
	(0.0066)	(0.0450)	(0.0250)	(0.0186)	(0.0138)
	[0.0068]	[0.0470]	[0.0290]	[0.0196]	[0.0168]
Dependent variable	Trial length	Trial length	Trial length	Trial length	Trial length
Estimation method	OLS	OLS	OLS	OLS	OLS
Stage	First	First	First	First	First
Judges avg. length	0.4396***	0.1969**	0.3204***	0.3659***	0.3631***
	(0.0300)	(0.0991)	(0.0920)	(0.1048)	(0.1018)
CraggâDonald Wald F stat.	214	4	12	12	13
Observations	5721	475	473	474	474

Note: The table reports estimates from the two stages (1.22)–(1.23) estimated in different subpopulations of firms' size. Each column corresponds to a separate estimation. The subpopulations of firms are determined according to the number of employees in the year before firms go to court. Firms employing more than 15 employees face higher expected costs of long trials because if the judge rules in favor of the worker the firm has to pay all forgone wages from the day of dismissal to the day of court ruling. Firms employing 15 or less employees pay a severance payment of at least 2.5 and at most 6 months wages which is determined by the judge if this rules in favor of the worker. Therefore for these firms the variation in trial length represents only a variation of the tax component of firing costs. The table also reports the average number of trials experienced by firms within each size interval. Unsurprisingly, larger firms experience more trials. Standard errors in round parentheses are clustered at the judge level in the second panel. Standard errors in round parentheses are robust to heteroscedasticity and in squared parentheses are bootstrapped with 100 repetitions in the first panel. * significant at 10%, ** significant at 5%, *** significant at 1%.

Appendix 1.A Proof of propositions in sections 3.2

Proof of Proposition 1.

Case (i): If,

$$\overbrace{z_t f'(n_{t-1}) + \delta E_{t-1}\left(\frac{\partial V(n_{t-1}, z_{t+1})}{\partial n_{t-1}}\right)}^{\text{MB of increasing labor at } t} > \overbrace{w}^{\text{MC of increasing labor at } t} \tag{1.31}$$

then it is optimal to increase labor in period t relatively to period $t-1$,

$$n_t > n_{t-1}$$

This is true since the LHS of 1.31 is decreasing in the labor force. In fact, by assumption of decreasing marginal returns $f'' < 0$ and

$$\frac{\partial^2 V(n_{t-1}, z_{t+1})}{\partial^2 n_{t-1}} = 0$$

since the costs of labor are linear. Rearranging 1.31 shows that the firm increases its labor force only if the realization of the shock is sufficiently high,

$$z_t > \frac{w - \beta E_{t-1}\left(\frac{\partial V(n_{t-1}, z_{t+1})}{\partial n_{t-1}}\right)}{f'(n_{t-1})} \equiv \bar{z}_t \tag{1.32}$$

Therefore, if condition 1.31 holds then the firm increases labor and the optimal level of labor in period t is given by the following first order condition:

$$z_t f'(n_t) - w + \delta E_t\left(\frac{\partial V(n_t, z_{t+1})}{\partial n_t}\right) = 0. \tag{1.33}$$

Case (ii): If,

$$\overbrace{w}^{\text{MB of decreasing labor at } t} > \overbrace{z_t f'(n_{t-1}) + \delta E_{t-1}\left(\frac{\partial V(n_{t-1}, z_{t+1})}{\partial n_{t-1}}\right) + F}^{\text{MC of decreasing labor at } t} \tag{1.34}$$

then it is optimal to decrease labor in period t relatively to period $t-1$,

$$n_t < n_{t-1}$$

This is true since the RHS of 1.34 is decreasing in the labor force. Rearranging 1.34 shows that the firm decreases its labor force only if the realization of the shock is sufficiently low,

$$z_t < \frac{w - F - \beta E_{t-1}\left(\frac{\partial V(n_{t-1}, z_{t+1})}{\partial n_{t-1}}\right)}{f'(n_{t-1})} \equiv \underline{z}_t \tag{1.35}$$

Therefore, if condition 1.34 holds then the firm decreases labor and the optimal level of labor in period t is given by the following first order condition:

$$z_t f'(n_t) - w + F + \delta E_t\left(\frac{\partial V(n_t, z_{t+1})}{\partial n_t}\right) = 0. \tag{1.36}$$

Case (iii): If,

$$w - F < z_t f'(n_{t-1}) + \delta E_{t-1}\left(\frac{\partial V(n_{t-1}, z_{t+1})}{\partial n_{t-1}}\right) < w \tag{1.37}$$

then it is optimal for the firm not to change employment in this period relatively to the previous period.

$$n_t = n_{t-1}$$

Rearranging 1.37 shows that the firm does not change its labor force if the realization of the shock is neither too high nor too law,

$$\underline{z}_t < z_t < \bar{z}_t$$

Proof of Proposition 2.

From Proposition 1 it follows that firms chose neither to hire nor to fire with probability $G(\bar{z}_t) - G(\underline{z}_t)$. Since $\bar{z}_t > \underline{z}_t$, it follows that $G(\bar{z}_t) - G(\underline{z}_t) > 0$. Moreover,

$$\frac{\partial \bar{z}_t}{\partial F} = \frac{\delta G(\underline{z}_t)}{f'(n_t)[1 - \delta(G(\bar{z}_t) - G(\underline{z}_t))]} > 0 \tag{1.38}$$

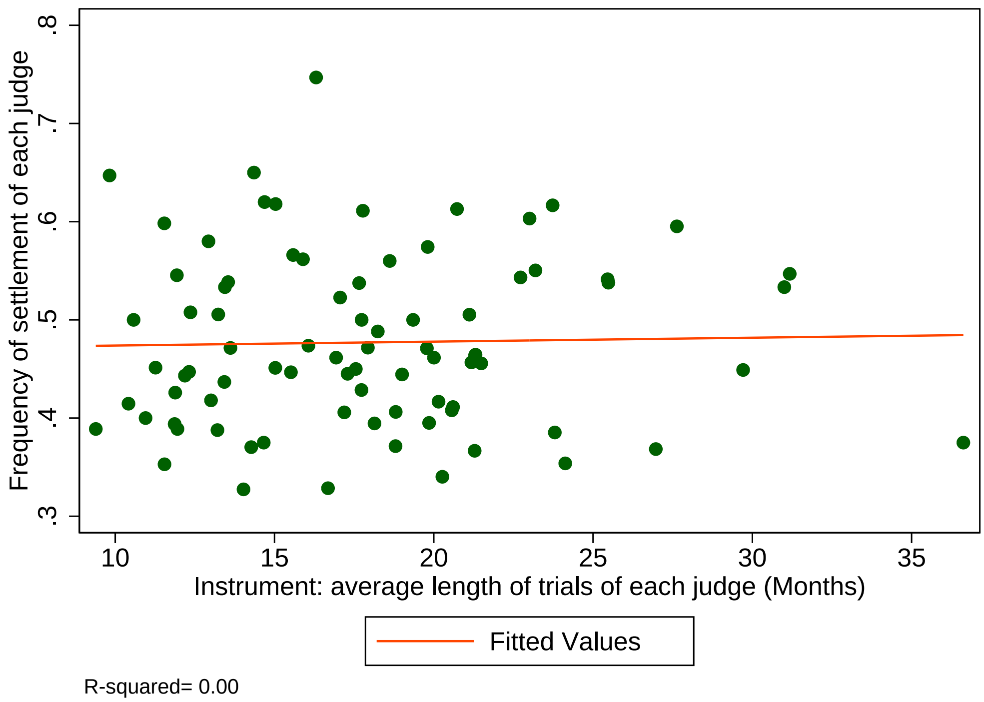
and

$$\frac{\partial \underline{z}_t}{\partial F} = -\frac{(1 + \delta G(\underline{z}_t))(1 - \delta G(\bar{z}_t))}{1 - \delta(G(\bar{z}_t) - G(\underline{z}_t))} < 0 \tag{1.39}$$

Therefore, an increase in the firing cost increases the probability of employment inaction.

$$\frac{\partial[G(\bar{z}_t) - G(\underline{z}_t)]}{\partial F} > 0 \tag{1.40}$$

Figure A1: Fast judges are not more likely to induce a settlement



Notes: The figure shows that judges that take on average short to complete their cases are not also more likely to induce a settlement. The fitted values are from the following regression:

$$S_j = \omega_0 + \omega_1 \tilde{Z}_j + \varepsilon_j$$

where S_j is the frequency judge j induces settlements and $\tilde{Z}_j$ is the average length judge j takes to complete his/her cases bases on all the cases assigned to the judge.

Table A1: Data sources and sample construction

Data sources:	
Firms balance sheet database (CERVED)	1993–2014
Monthly firms employment database, Italian National Social Security (INPS) archives	1990-2013
Labor court database:	
Cases filed	2001-2012
Trials ended	2001-2014
Data are linked using firms names as identifier:	
Firms in CERVED–INPS database operating in the geographical area where the labor court has jurisdiction	220,341
Firms in labor court database	25,906
Firms linked between labor court and CERVED–INPS databases	7,617

Note: this table summarizes the steps to obtain the final data set of firms which went to court between 2001 and 2012 and for which it was possible to recover information on their monthly employment using the CERVED–INPS database. Table A2 shows that the observable characteristics of trials do not differ between the group of firms linked between the labor court database and the CERVED–INPS database (7,617 firms), and the group of firms for which this linkage is not possible (18,289 firms).

Table A2: Comparison of trials of firms linked and not linked between databases

Variables	Averages: Firms not linked	Firms linked	p–value for H_0: equal means
Object of controversy: *Overall % of trials with given object*			
Compensantion	0.2842	0.2965	.000
29%	(0.4510)	(0.4567)	
Attendance allowance	0.0004	0.0004	.942
0.04%	(0.0189)	(0.0192)	
Other hypothesis	0.1976	0.2078	.000
20%	(0.3982)	(0.4057)	
Other controversies	0.0338	0.0329	.469
3%	(0.1807)	(0.1783)	
Disability living allowance	0.0002	0.0001	.236
0.02%	(0.0157)	(0.0115)	
Pension	0.0002	0.0002	.813
0.02%	(0.0134)	(0.0126)	
Temporary work contract	0.0506	0.0464	.005
5%	(0.2192)	(0.2103)	
Termination of employment	0.1809	0.2039	.000
19%	(0.3849)	(0.4029)	
Type of employment relationship	0.0575	0.0454	.000
5%	(0.2328)	(0.2082)	
Other types of cases	0.1947	0.1665	.000
18%	(0.3960)	(0.3726)	
Red code case	0.2175	0.2316	.000
22%	(0.4125)	(0.4219)	
Number of parties involved in trials	2.41	2.41	.893
Overall average: 2.41	(2.50)	(2.36)	
Number of trials	44,552	37,966	
Number of firms	18,289	7,617	

Notes: The table shows that the observable characteristics of trials do not differ between the group of firms linked between the labor court database and the CERVED–INPS database, and for firms for which this linkage is not possible. The p–value refers to the t–tests of the equality of means between these two groups of firms. Red code cases is a dichotomous aggregation of the objects of controversy in red code versus green code cases, by analogy with what happens in a hospital emergency room, where red code cases are those that, according to judges, are urgent and/or complicated, thus requiring immediate action and/or greater effort. Values in parentheses are standard deviations.

Table A3: Percentage of firms for which no monthly employment change is observed (censored)

Year end of trial	Number of firms	Number of firms censored	Percentage of firms censored (%)
2001	29	0	0
2002	394	0	0
2003	512	2	0.39
2004	589	3	0.51
2005	689	5	0.73
2006	649	6	0.92
2007	607	5	0.82
2008	551	7	1.27
2009	508	10	1.97
2010	600	16	2.67
2011	712	43	6.04
2012	981	86	8.77
2013	796	325	40.83
Overall	7617	508	6.67

Note: The table reports the number of firms for which no monthly employment change is observed after the end of their trials. Since the court data contains firms going to court in 2001–2012, there are only 29 firms going to court in 2001 and experiencing a trial that lasted less than 13 months.

Table A4: Including Controls

	(1)	(2)
Dependent variable	$h(t\|X)$	$h(t\|X)$
Estimation method	ML	ML
Stage	Second	Second
trial lengths	-0.0370***	-0.0381***
	(0.0059)	(0.0058)
Dependent variable	trial lengths	trial lengths
Estimation method	OLS	OLS
Stage	First	First
Judges avg. length	0.4110***	0.4137***
	(0.0257)	(0.0257)
CraggâDonald Wald F stat.	256	260
Observations	7617	7617
Controls	No	Yes

Notes: The table shows the irrelevance of controls for the estimates of the causal parameter of interest. Each column corresponds to a separate estimation. Column (1) reports estimates from the two stages equations (1.22)–(1.23). Column (2) adds to these equations a set of firm level control variables,

$$\ell_i = \delta_0 + \delta_1 Z_{j(i)} + \delta_2 X_{i,-1} + v_i \qquad \text{first–stage}$$
$$h_{it} = h_0(t)\text{exp}\{\beta\ell_i + \psi X_{i,-1} + g(v_i)\} \qquad \text{second–stage}$$

where $X_{i,-1}$ is a vector of controls that includes, calendar monthly and yearly dummies to control for time effects, including seasonality, in the most flexible way, 13 sectors dummies and a set of time–varying baseline covariates measured in the year prior to the filing of firm's i case. The covariates are: revenue, cost of labor, cash flow liquidity, assets, capital, investments, return on equity, return on assets, value added and employment. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A5: The effect of firing costs on the duration of employment inaction, IV

Dependent variable	Trial length	$\log(T)$
Estimation method	OLS	IV
Stage	First	Second
	(1)	(2)
Trial length		0.0388***
		(0.0058)
Judge avg. length	0.4141***	
	(0.0289)	
Cragg–Donald Wald F statistic		229
Observations	6821	6821

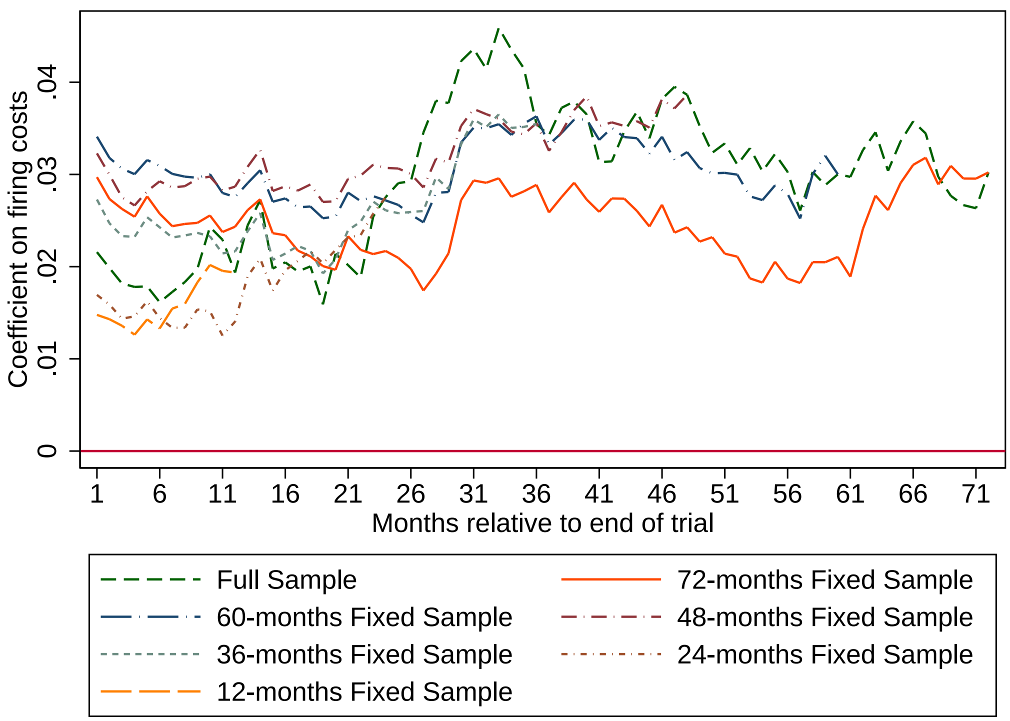
Note: The table reports the results on the effect of trial length on the duration of the spell of employment inaction, that is the number of months until a firm changes employment after the end of its trial. Censoring occurs if variations in the employment levels of a firm are not observed. Considering only firms which trial ended before 2013, censoring ranges from 0 (trials ending in 2001) to 8.8% (trials ending in 2012) with an average censoring of 2.7%. For firms which trial ended between 2001 and 2012, I use a two–stage least–squares procedure in a linear model. The table reports the estimates from the two stages:

$$\ell_i = \delta_0 + \delta_1 Z_{j(i)} + v_i \qquad \text{first–stage}$$

$$\log(T_i) = \beta_0 + \beta_1 \hat{\ell}_i + \varepsilon_i \qquad \text{second–stage}$$

Expected Firing cost are measured as the length of the trials experienced by each firm, ℓ_i. The instrument, $Z_{i(j)}$, is the average length judge j assigned to firm i takes to complete his or her cases, based on all the judge's other cases. T_i is the number of months until firm i changes employment from the end of the trial. $\hat{\ell}_i$ are the fitted values from the first–stage regression. Standard errors in parentheses are clustered at the judge level in column (1). * significant at 10%, ** significant at 5%, *** significant at 1%.

Figure A2: Effect of firing costs on employment levels with fixed samples



Notes: The figure shows the effect of firing costs for different time horizons when the sample is held fixed. The 12–months estimates include firms which trials ended between January 2001 and January 2013, the 24–months estimates firms which trials ended between January 2001 and January 2012, the 36–months estimates firms which trials ended between January 2001 and January 2011, the 48–months estimates firms which trials ended between January 2001 and January 2010, the 60–months estimates firms which trials endend between January 2001 and January 2009 and the 72–months estimates firms which trials ended between January 2001 and January 2008.

Table A6: Heterogeneity by firms' financial constraints (available liquidity/assets)

Dependent variable	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$
Estimation method	ML	ML	ML	ML	ML	ML
Stage	Second	Second	Second	Second	Second	Second
	(1)	(2)	(3)	(4)	(5)	(6)
Trial length	-0.0496***	-0.0447***	-0.0314**	-0.0569***	-0.0296*	-0.0390**
	(0.0164)	(0.0154)	(0.0159)	(0.0132)	(0.0171)	(0.0166)
Cragg–Donald Wald F stat.	26	38	32	46	33	26
Dependent variable	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS
Stage	First	First	First	First	First	First
	(1)	(2)	(3)	(4)	(5)	(6)
Judgeâs avg. length	0.3715***	0.4432***	0.4106***	0.4786***	0.3912***	0.3584***
	(0.0722)	(0.0721)	(0.0728)	(0.0705)	(0.0683)	(0.0703)
Observations	1048	1048	1049	1047	1048	1047

Note: The table reports estimates from the two stages equations (1.22) (second panel of the table) and (1.23) (first panel of the table) estimated in different subpopulations of firms depending on their available liquidity over assets. Each column refers to a different quantile of available liquidity over assets which are reported in ascending order, hence the 1st quantile (column (1)) refers to firms more financially constraint whereas the 6th quantile (column (6)) refers to firms less financially constraint.
* significant at 10%, ** significant at 5%, *** significant at 1%.

Table A7: Heterogeneity by firms' financial constraints (available liquidity/employees)

Dependent variable	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$
Estimation method	ML	ML	ML	ML	ML	ML
Stage	Second	Second	Second	Second	Second	Second
	(1)	(2)	(3)	(4)	(5)	(6)
Trial length	-0.0634***	-0.0029	-0.0471***	-0.0395***	-0.0755***	-0.0433***
	(0.0212)	(0.0155)	(0.0130)	(0.0144)	(0.0187)	(0.0154)
Cragg–Donald Wald F stat.	17	38	53	39	23	29
Dependent variable	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS
Stage	First	First	First	First	First	First
	(1)	(2)	(3)	(4)	(5)	(6)
Judgeâs avg. length	0.2917***	0.4195***	0.5125***	0.4751***	0.3469***	0.3695***
	(0.0717)	(0.0680)	(0.0707)	(0.0760)	(0.0716)	(0.0682)
Observations	1048	1048	1048	1048	1048	1047

Note: The table reports estimates from the two stages equations (1.22) (second panel of the table) and (1.23) (first panel of the table) estimated in different subpopulations of firms depending on their available liquidity over number of employees. Each column refers to a different quantile of available liquidity over number of employees which are reported in ascending order, hence the 1st quantile (column (1)) refers to firms more financially constraint whereas the 6th quantile (column (6)) refers to firms less financially constraint. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A8: Heterogeneity by firms' ages

Dependent variable	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$
Estimation method	ML	ML	ML	ML	ML	ML	ML	ML	ML	ML	ML	ML	ML
Stage	Second	Second	Second	Second	Second	Second	Second	Second	Second	Second	Second	Second	Second
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Trial length	-0.2736***	-0.0669*	-0.0597***	-0.0462***	-0.0231	-0.0037	-0.0105	-0.0460**	-0.0698***	-0.0238	-0.0288	-0.0426***	-0.0220
	(0.0840)	(0.0365)	(0.0218)	(0.0160)	(0.0288)	(0.0179)	(0.0108)	(0.0182)	(0.0264)	(0.0199)	(0.0254)	(0.0164)	(0.0210)
Cragg–Donald Wald F stat.	6	12	23	39	9	23	69	16	11	21	9	26	13
Dependent variable	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Stage	First	First	First	First	First	First	First	First	First	First	First	First	First
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Judgeâs avg. length	0.1000**	0.1883***	0.3557***	0.5737***	0.3019***	0.5152***	0.7727***	0.4549***	0.3480***	0.4178***	0.3416***	0.5580***	0.3781***
	(0.0419)	(0.0542)	(0.0734)	(0.0914)	(0.0998)	(0.1078)	(0.0932)	(0.1139)	(0.1041)	(0.0923)	(0.1169)	(0.1104)	(0.1049)
Observations	592	735	691	582	538	465	674	592	550	567	531	538	562
Quantiles [min, max] Age	[1,2]	[3,3]	[4,4]	[5,5]	[6,6]	[7,7]	[8,9]	[10,11]	[12,14]	[15,18]	[19,23]	[24,31]	[32,56]

Note: The table reports estimates from the two stages equations (1.22) (second panel of the table) and (1.23) (first panel of the table) estimated in different subpopulations of firms, within the subgroup of firms born after 2001, depending on their age at the end of trials. For example, in column (1) results refer to firms that are at least 1 year old and at most 2 years old when the trials ends. Each column corresponds to a separate estimation. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A9: Heterogeneity by firms' ages (firms born after 2001)

Dependent variable	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$	$h(t\|X)$
Estimation method	ML	ML	ML	ML	ML	ML	ML	ML
Stage	Second	Second	Second	Second	Second	Second	Second	Second
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Trial length	-0.2949***	-0.0752*	-0.0552**	-0.0486***	0.0638	-0.0295	0.0173	0.0103
	(0.0977)	(0.0387)	(0.0235)	(0.0167)	(0.0530)	(0.0299)	(0.0282)	(0.0313)
Cragg–Donald Wald F stat.	5	12	21	36	2.6	11	11	12
Dependent variable	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length	Trial length
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Stage	First	First	First	First	First	First	First	First
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Judgeâs avg. length	0.0929**	0.2021***	0.3649***	0.6527***	0.2152	0.4557***	0.5261***	0.5623***
	(0.0423)	(0.0587)	(0.0805)	(0.1088)	(0.1337)	(0.1389)	(0.1592)	(0.1647)
Observations	551	658	579	443	356	266	310	228
Quantiles [min, max] Age	[1,2]	[3,3]	[4,4]	[5,5]	[6,6]	[7,7]	[8,9]	[10,12]

Note: The table reports estimates from the two stages equations (1.22) (second panel of the table) and (1.23) (first panel of the table) estimated in different subpopulations of firms, within the subgroup of firms born after 2001, depending on their age at the end of trials. For example, in column (1) results refer to firms that are at least 1 year old and at most 2 years old when the trials ends. Each column corresponds to a separate estimation. * significant at 10%, ** significant at 5%, *** significant at 1%.

2 Are Lawyers responsible for Trial Delays?

Using administrative data from one large Italian court I quantify the extent to which lawyers can be held accountable for the slowness of Italian courts. Longitudinal court data allows to observe the same lawyer over time in different trials, acting in different roles, of plaintiff and defendant. Borrowing the methodology used in the analysis of employers–employees linked data, I estimate the contribution of unobservable time–invariant plaintiff and defendant lawyers' characteristics in explaining the variability of trials' length. I find that 27% of the variance in trials length is explained by unobservable time–invariant lawyers' characteristics. There is, instead, no evidence that lawyers tend to specialize in certain types of controversies or of assortative matching between plaintiff and defendant lawyers. Long delays make it harder to do business and whether lawyers are partially responsible for the excessive length of Italian trials is a hypothesis that has been often suggested, but this is the first paper that tries to find convincing evidence for it.

2.1 Introduction

My goal is to assess the importance of lawyers in determining the duration of trials. In every trial there is a plaintiff and a defendant lawyer, this paper investigates the role of unobservable time–invariant lawyers' heterogeneity (lawyers' fixed effects) in explaining the length of trials. Using all cases filed between 2001 and 2012 in a large Italian court, it is possible to observe the same lawyers working in different cases and hence to identify her time–invariant contribution to trials length. I find that 27% of the variance in trials length is explained by unobservable time–invariant lawyers' characteristics. The slowness of the Italian judicial system has been identified as major drag on the Italian economy.[1] Whether lawyers are partially responsible for the excessive length of trials in Italy is a hypothesis that has been often suggested, to the best of my knowledge, this is the first paper that tries to find convincing evidence for it.

A striking fact about Italy is that there are many more lawyers than in other European countries. According to the Council of Bars and Law Societies of Europe (CCBE), there are 233,852 Italian lawyers compared to 53,744 French lawyers even though the number of inhabitants is similar in Italy (61 millions) and France (66 millions). Besides, according to the European Commission for the Efficiency of Justice (CEPEJ), there are 11 and 10.7 judges for 100,000 inhabitants in Italy and French respectively. Finally, it takes 493 days in Italy compared to 279 days in France to terminate a pending case. Therefore, despite having the same resources and similar populations, Italy has more lawyers and more delayed trials than France.

Anecdotally, this army of Italian lawyers is held responsible for the number and the delay

[1] Giacomelli and Menon (2016) provide empirical evidence on the negative effect of slow courts on Italian firms.

of trials. Since there are many lawyers looking for jobs, lawyers create their own demand by inducing their clients to start lawsuits or by delaying their cases. The adverse incentives are clear: Italian lawyers never charge a flat fee but are paid according to the number of hours worked on a client's case until the case is resolved.

Ideally, one would like to randomize the number of lawyers in different areas to assess if lawyers do cause litigation and delay. In this spirit, Carmignani and Giacomelli (2010) show a positive correlation between the number of lawyers and litigation across Italian provinces. To provide causal evidence, they also use the number of law schools in each province as an exogenous variation for the number of lawyers.

Instead of looking for an exogenous source of variation for the number of lawyers, I consider one large Italian court and study if different lawyers contribute differently to the duration of trials. I find that lawyers explain a significant proportion of the variation of trials length. Moreover, by controlling for the complexity of each controversy I can rule out that part of this effect is due to the fact that lawyers tend to specialize in certain types controversies.

Other works study the effect of delayed trials on economic outcomes. For example, previous empirical works show that the uncertainty and costs associated with longer trials reduce firms' size Kumar, Rajan, and Zingales (1999); Giacomelli and Menon (2016), job turnover rates and firms' productivity Gianfreda and Vallanti (2015), the efficiency of credit markets Bianco, Jappelli, and Pagano (2001); Fabbri (2010), trade flows Nunn (2007) and economic development Chemin (2009, 2012). My work contributes to the literature by explaining why trials are delayed in the first place and I show that part of this delay is due to unobservable time–invariant lawyers' characteristics.[2]

[2]My work is related to Coviello, Ichino, and Persico (2015), they find as a determinant of trials length the

2.2 Measuring the determinants of trials' length

This section introduces the econometric framework for disentangling the components of the variation of the duration of trials attributable to lawyer–specific heterogeneity and other variables.

The data set contains observations on T trials. There are L lawyers taking part in these trials. Let L_p and L_d denote the set of plaintiff and defendant lawyers respectively. Since the same lawyer can work as plaintiff and as defendant, on different cases, the intersection of L_p and L_d is not empty. Functions $p(t)$, $d(t)$, $j(t)$ give respectively the identity of the unique plaintiff lawyer, defendant lawyer and judge working on trial t. I assume that the duration y_t of trial t is the sum of a plaintiff lawyer component $\pi_{p(t)}$, a defendant lawyer component $\delta_{p(t)}$, a judge component $\psi_{j(t)}$, a set of observable characteristics of the case x_t and an error component ϵ_t:

$$y_t = \pi_{p(t)} + \delta_{d(t)} + \psi_{j(t)} + x_t'\beta + \epsilon_t \tag{2.1}$$

I interpret the plaintiff $\pi_{p(t)}$ and defendant $\delta_{p(t)}$ lawyers effects as all unobservable time–invariant characteristics of lawyers that affect the duration of trials. Likewise, I interpret $\psi_{j(t)}$ as all unobservable time–invariant judges' characteristics affecting trials length. Instead, I interpret $x_t'\beta$ as the effect of a combination of trial–specific characteristics. I include in x_t a set of dummy variables which identify the type of case, for example a firing litigation, the number of parties involved in the trial and a set of year dummies to control for the year in which the case is filed.

There are two reasons to include $x_t'\beta$ in equation (2.1). First, because I want to measure

the way in which judges organize their work.

the contribution of trial–specific characteristics in explaining the variance of trials length. Second, because omitting $x_t'\beta$ in equation (2.1) could bias the estimates of lawyers' fixed effects. To see this consider the following example: suppose that a lawyer is always working on complicated cases which last many days. If I do not measure the complexity of the cases, then I would conclude that the lawyer is slow even though the long duration of her trials is not due to her individual fixed effect but to the complexity of the cases on which the lawyer works. In other words, it is possible that lawyers specialize in long and complicated cases, hence not controlling for the observable characteristics of the cases in equation (2.1) could bias the estimates of lawyers fixed effects.

Let y denote the stacked $T \times 1$ vector of duration of trials, P a $T \times |L_p|$ design matrix of plaintiff lawyer indicators, D a $T \times |L_d|$ design matrix of defendant lawyer indicators, J a $T \times N_j$ design matrix of judge indicators, X a $T \times K$ matrix of trial–specific covariates, and ϵ a $T \times 1$ vector of errors. Then the model can be written in matrix notion as:

$$\begin{aligned} y &= P\pi + D\delta + J\psi + X\beta + \epsilon \\ &= Z'\xi + \epsilon \end{aligned} \tag{2.2}$$

where $Z \equiv [P, D, J, X]$ and $\xi \equiv [\pi', \delta', \psi', \beta']$.

Equation (2.2) is estimated by ordinary least squares (OLS). These estimates solve the standard normal equations:

$$Z'Z\xi = Z'y. \tag{2.3}$$

A unique solution requires that the matrix $Z'Z$ has full rank. Applying the results of Abowd, Kramarz, and Margolis (1999) and Abowd, Creecy, and Kramarz (2002) to my context, the plaintiff and defendant lawyers effects are only separately identified within a "connected set" of lawyers that are linked by the fact that lawyers face different lawyers in distinct trials. In my sample almost all lawyers belong to a single connected set of lawyers.[3] I normalize the lawyers effects within this connect set of lawyers by omitting the last lawyer dummy. Appendix 2.A provides details of the procedure used to obtain a solution to the normal equations (2.3). In brief, I use an iterative conjugate gradient algorithm proposed by Abowd, Creecy, and Kramarz (2002) to solves for the vector of coefficients ξ without actually inverting the matrix $Z'Z$.

I would like to point out two differences between our setting and the setting of matched employer–employee data sets, from which I borrow the techniques to identify and estimate lawyers' fixed effects. First, as shown by Abowd, Kramarz, and Margolis (1999) it is possible to identify the fixed effect of an employee who never changes employer as long as there is at least another employee working for the same employer who has also moved to another employer at some point. However, as Abowd, Kramarz, Lengermann, and Pérez-Duarte (2004) and Andrews, Gill, Schank, and Upward (2008) show, when employee mobility is limited, there may be an estimation bias that causes the employee and employer effects to be estimated imprecisely. The intuition is to use the mover information to figure out the employer fixed effects. When the number of movers goes up, there is more information and thus the employer fixed effects (and then the employee fixed effects) can be more precisely estimated.

This is not a concern in my context, because each lawyer is matched with a different lawyer in different trials. In other words, there is a lot of mobility among lawyers which guarantees connectedness among them and a more connected sample increases the precision of the model estimates.

[3]Section 2.3 reports exact figures of the number of lawyers not belonging to the largest "connected set" of lawyers.

Second, for OLS to identify the underlying parameters of interest, I need the following orthogonality condition to hold:

$$E[\epsilon_t | t, p(t), d(t), j(t), x_t] = 0. \tag{2.4}$$

This condition is likely to be satisfied in my case because plaintiff and defendant lawyers are not endogenously matched, instead they are independently chosen by their clients. In the setting of matched employer–employee, employees can to some extent chose their own employers, hence violating this orthogonality condition.

2.2.1 Variance decomposition

Similarly to Gruetter and Lalive (2009) in the context of matched employers–employees databases, I measure the importance of lawyers for the duration of trials by the contribution of lawyers' fixed effects to the overall variance of trials' duration. Using Equation (2.1), the total variance in trials length $\mathrm{var}(y_t)$ can be decomposed into two terms reflecting the covariance of trials length with time invariant characteristics of plaintiff and defendant lawyers, a term reflecting the covariance of trials length with time invariant characteristics of judges, a term reflecting the covariance of trials length with trials characteristics and a term reflecting the covariance of trials length with the error term, i.e.

$$\begin{aligned} \mathrm{var}(y_t) = \mathrm{cov}(y_t, y_t) = \mathrm{cov}(y_t, \pi_{p(t)}) + \mathrm{cov}(y_t, \delta_{d(t)}) + \mathrm{cov}(y_t, \psi_{j(t)}) \\ + \mathrm{cov}(y_t, x_t'\beta) + \mathrm{cov}(y_t, \epsilon_t). \end{aligned} \tag{2.5}$$

Moreover, defining the importance of each component in determining trials length relative to the overall variance

$$\nu_p = \frac{\operatorname{cov}(y_t, \pi_{p(t)})}{\operatorname{var}(y_t)} = \frac{\operatorname{var}(\pi_{p(t)})}{\operatorname{var}(y_t)} + \frac{\operatorname{cov}(\delta_{d(t)} + \psi_{j(t)} + x_t'\beta, \pi_{p(t)})}{\operatorname{var}(y_t)} \tag{2.6}$$

$$\nu_d = \frac{\operatorname{cov}(y_t, \delta_{d(t)})}{\operatorname{var}(y_t)} = \frac{\operatorname{var}(\delta_{d(t)})}{\operatorname{var}(y_t)} + \frac{\operatorname{cov}(\pi_{p(t)} + \psi_{j(t)} + x_t'\beta, \delta_{d(t)})}{\operatorname{var}(y_t)} \tag{2.7}$$

$$\nu_j = \frac{\operatorname{cov}(y_t, \psi_{j(t)})}{\operatorname{var}(y_t)} = \frac{\operatorname{var}(\psi_{j(t)})}{\operatorname{var}(y_t)} + \frac{\operatorname{cov}(\delta_{d(t)} + \pi_{p(t)} + x_t'\beta, \psi_{j(t)})}{\operatorname{var}(y_t)} \tag{2.8}$$

$$\nu_x = \frac{\operatorname{cov}(y_t, x_t'\beta)}{\operatorname{var}(y_t)} = \frac{\operatorname{var}(x_t'\beta)}{\operatorname{var}(y_t)} + \frac{\operatorname{cov}(\delta_{d(t)} + \psi_{j(t)} + \pi_{p(t)}, x_t'\beta)}{\operatorname{var}(y_t)} \tag{2.9}$$

$$\nu_\epsilon = \frac{\operatorname{cov}(y_t, \epsilon_t)}{\operatorname{var}(y_t)} = \frac{\operatorname{var}(\epsilon_t)}{\operatorname{var}(y_t)} \tag{2.10}$$

ensures a full decomposition of the total variance in trials length into a plaintiff lawyer component ν_p, a defendant lawyer component ν_d, a judge component ν_j, a component due to trials' characteristics ν_x and a transitory component ν_ϵ.[4]

This decomposition is useful because it allows to quantify, by estimating ν_p and ν_d, the percentage of the variance in trials length explained by time invariant lawyers' characteristics. Furthermore, equations (2.6) and (2.7) show that lawyers can be important in determining trials length for two reasons. First, lawyers generate strong trials length differentials if there are reasons for lawyers to make otherwise observationally equivalent trials longer. This shows up as a high relative variance of lawyers' contribution, $(\operatorname{var}(\pi_{p(t)})/\operatorname{var}(y_t)$ and $\operatorname{var}(\delta_{d(t)})/\operatorname{var}(y_t))$. Second, lawyers can also amplify the total variance in trials length by attracting more complex cases or by being matched to slower lawyers or judges. Random assignment of cases to judges excludes the possibility that slow judges are assigned to certain lawyers, whereas the possibility that lawyers attract more complex cases or assortatively match with their colleagues is assessed empirically using equations (2.6) and

[4]Note that the covariance between each variable, including lawyers' contributions, with the error term is zero by the exogeneity assumption in Equation (2.4).

(2.7).

In the analysis that follows I use a feasible version of this decomposition that replaces each term with its corresponding sample analogue.

2.3 The data

I use data on 207,001 trials filed between 2002 and 2011 in one of the largest Italian labor court[5]. 15,257 lawyers participated in these trials and for each of them it is recorded whether the role was that of plaintiff or defendant. The testing strategy described in Section 2.2 requires me to focus on the subset of lawyers belonging to the largest connected set of lawyers. Out of 15,257 lawyers, only 230 do not belong to the largest connected set. This leaves me with 15,027 lawyers taking part in 206,735 trials.

The court used in this study is a subset of the civil court which specializes in labor controversies. Table 2.1 reports the distribution of the types of controversies which are filed in this court. Table 2.2 shows the large variance of trials length. Table 2.3 reports the distribution of the number of trials for each lawyer in order to give an idea of the number of observations available for each lawyer. Lawyers fixed effects are estimated using the whole sample, however, for the fixed effects estimates to be meaningful, only the fixed effects of lawyers who took part in at least 2 trials are used in Section 2.4.

[5]This is a subset of the 412,395 cases filed during the same period in the same court. Appendix 2.B explains why I am not using the universe and provides additional information about the data, beyond what is summarized in this section

2.4 Evidence

In this section I present and discuss the results obtained from estimating equation (2.1).[6] I decompose trials length into its components as described in section 2.2.1 and compute the extent to which each contributes to the variance in trials length.

Table 2.4 reports the estimated decomposition of variance presented in equation (2.5), based on the estimates of equation (2.1). Results indicate that 17% of the variance in trials length is due to differences in unobserved time invariant characteristics of plaintiffs lawyers. In comparison, the contribution of differences in unobserved time invariant characteristics of defendant lawyers to the variance in trials length is 10%. Judges' unobserved time invariant characteristics contribute 2%. Trials' characteristics such as, the type of controversy, the year in which the trial is filed and the number of parties involved in the controversy, contribute 9%. The error term still accounts to for the majority of variance in trials length 62%.

Columns (2) to (5) in Table 2.4 decompose, respectively, the total contribution of plaintiff lawyers, defendant lawyers, judges, and observable trials characteristics. The variance in plaintiff lawyer effects is 15% of the overall variance in trials length and the variance in defendant lawyer effect is 9% of the overall variance in trials length. The covariance between plaintiff and defendant lawyers effects is essentially zero (0.64%), as it is the covariance between lawyers effects and observable trials characteristics, (1% for the plaintiff lawyer and 0.61% for the defendant lawyer). Reassuringly, the covariance between judges effects

[6] The whole sample consisting of 206735 trials and 15027 lawyers is used to estimate equation (2.1). Note, however, that some lawyers are observed only in a few trials, hence the fixed effects of these lawyers are expected to be estimated quite noisily. To alleviate this problem, I use only the lawyers fixed effects of lawyers observed in at least 5 trials to estimate the covariance decomposition defined in equation (2.5). Equation (2.1) is estimated using the user–written Stata command a2reg (Ouazad (2008)), which uses a Conjugate gradient method to solve the minimum squares problem.

and other variables is essentially zero, as random assignment of cases to judges would imply. Finally, the covariance between all variables and the estimated error term is zero as required for the least squares solution.

My findings suggest that plaintiff lawyers are more important than defendant lawyers in contributing to trials length. There is no assortative matching of lawyers, that is slow plaintiff lawyers do not tend to work with slow defendant lawyers in the same trials. Finally, lawyers do not tend to sort into certain types of cases.

2.5 Conclusions

A growing body of research argues that delayed trials lead to worse economic outcomes, yet little is known why courts take so long to resolve legal disputes. Knowing the reasons behind slow trials could direct policy interventions aimed at improving the quality of judicial systems. My analysis of Italian data finds that lawyers are accounted for 27% of the variance in trials length. Moreover, plaintiff lawyers are more important than defendant lawyers in contributing to trials length. The former explain 17%, whereas the latter explain only 10% of the variance in trials length. I do not find, instead, any evidence of lawyers sorting into complex cases or lawyers' assortative matching, based on lawyers contribution to trials' delay.

Future research should explain what drives the large heterogeneity of lawyers' contribution to trials length. For example, are older and more experienced lawyers faster than younger ones? Or, are faster lawyers more in demand than slower ones? These and other questions can be answered by following my approach of adapting the machinery of the analysis of employers–employees linked data to a setting in which the link is between lawyers in trials.

Table 2.1: Types of Trials

	Number of Trials	Percent
Compensation	34,950	17
Attendance allowance	41,304	20
Disability living allowance	4,713	2
Pension	9,850	5
Temporary work contract	12,959	6
Termination of employment	17,377	8
Type of employment relationship	5,231	3
Other types of cases	80,351	39
Total	206,735	100

Note: The table reports the distribution of the types of trials aggregated in 8 different categories.

Table 2.2: Summary statistics of trials length

	Trials Length (Days)
1st	25
5th	88
10th	139
25th	227
Median	351
75th	551
90th	882
95th	1,259
99th	4,102
Mean	525
Standard Deviation	677
Number of Trials	206,735

Note: The table reports the percentiles, mean and standard deviation of trials length measured as the number of days from the filing to the disposition of the case.

Table 2.3: Number of trials in which lawyers take part

	Number of Trials Lawyer (any role)	Number of Trials Defendant Lawyer	Number of Trials Plaintiff Lawyer
1st	1	0	0
5th	1	0	0
10th	1	0	0
25th	1	0	1
Median	3	1	1
75th	10	4	4
90th	30	13	14
95th	71	28	32
99th	454	147	240
Mean	28	14	14
Standard Deviation	208	164	99
Number of Lawyers	15,027	15,027	15,027

Note: The table reports the percentiles, mean and standard deviation of the number of trials in which lawyers take part. Half of the lawyers in the sample took part in at least 3 trials.

Table 2.4: The Components of Trials Length

		Covariance with				
	Total Contribution	Plaintiff Effect $\hat{\pi}$	Defendant Effect $\hat{\delta}$	Judge Effect $\hat{\psi}$	Obs. Trial Char. $X\hat{\beta}$	Error Term $\hat{\epsilon}$
	(1)	(2)	(3)	(4)	(5)	(6)
Plaintiff Effect $\hat{\pi}$	16.87	15.09	0.64	0.00	1.03	0.11
Defendant Effect $\hat{\delta}$	10.40	0.64	9.13	-0.05	0.61	0.08
Judge Effect $\hat{\psi}$	1.88	0.00	-0.05	2.06	-0.14	0.01
Obs. Trial Char. $X\hat{\beta}$	8.86	1.03	0.61	-0.14	7.25	0.11
Error Term $\hat{\epsilon}$	61.99	0.11	0.08	0.01	0.11	61.68
Total	100.00	16.87	10.40	1.88	8.86	61.99

Note: The table reports the estimated decomposition of variance presented in equation (2.5), based on the estimates of equation (2.1). Column (1) reports the estimated decomposition of the variance of trials length into the covariance between trials length and the following five terms: the time invariant characteristics of plaintiff lawyers, defendant lawyers and judges, trials characteristics and the error term. Columns (2) to (6) decomposes further each covariance reported in each row of column (1) as defined in equations (2.6) to (2.10). For example, column (2) decomposes further the covariance between trials length and the time invariant characteristics of plaintiff lawyers(ν_p) into the variance of the time invariant characteristics of plaintiff lawyers and its covariance with the remaining terms: the time invariant characteristics of defendant lawyers and of judges, and trials characteristics. The observable trials characteristics include, a set of dummy variables which identify the type of case, for example a firing litigation, the number of parties involved in the trial and a set of year dummies to control for the year in which the case is filed. Moreover, to avoid using lawyers fixed effects which are noisily estimated, the covariances in the table are computed using only the subset of 3753 plaintiff lawyers and 3686 defendant lawyers (the same lawyer can be both plaintiff and defendant in distinct trials) observed in at least 5 different trials. This reduces the number of observations used to compute the covariances in the table to 181045 trials.

Appendix 2.A Identification of the lawyers specific structural parameters π and δ

Consider the following two–way fixed effects model, where the dependent variable y_t is the duration of trial t:

$$y_t = \pi_{p(t)} + \delta_{d(t)} + \epsilon_t \tag{A.1}$$

trials are indexed by $t = 1, ..., T$, plaintiff and defendant lawyers are indexed by $p = 1, ..., |L_p|$ and $d = 1, ..., |L_d|$ respectively. L_p and L_d are the sets of plaintiff and defendant lawyers respectively. $|A|$ is the cardinality of an arbitrary set A. The functions $p(t)$ and $d(t)$ map trial t to plaintiff lawyer p and defendant lawyer d. Since the same lawyer cannot be both plaintiff and defendant in the same trial, $p(t) \neq d(t)\ \forall t$. The first component of equation (A.1) is the contribution of the plaintiff lawyer to the duration of the trail, $\pi_{p(t)}$. The second component of equation (A.1) is the contribution of the defendant lawyer to the duration of the trail, $\delta_{d(t)}$. The third component of equation (A.1) is noise, ϵ_t. (For the moment I abstract from trial–specif covariates affecting the duration of trials, but their inclusion in equation (A.1) is straightforward).

The model in equation (A.1) in matrix form is,

$$y = P\pi + D\delta + \epsilon \tag{A.2}$$

where y is a T vector of the duration of trials, P is a $T \times |L_p|$ design matrix of plaintiff lawyer indicators, π is a $|L_p|$ vector of plaintiff effects, D is a $T \times |L_d|$ design matrix of defendant lawyer indicators, δ is a $|L_d|$ vector of defendant effects and ϵ is a T vector; (i.e. P and D have only one entry equal to 1 for each row, all other elements of each row are 0). The least squares solution to the estimation problem for equation (A.2) solves the following normal equations:

$$\begin{bmatrix} P'P & P'D \\ D'P & D'D \end{bmatrix} \begin{bmatrix} \hat{\pi} \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} P'y \\ D'y \end{bmatrix} \tag{A.3}$$

However, the sum of the columns of P and the sum of the columns of D are both equal to the column vector of ones, $\mathbf{1}$. Therefore, the rank of matrix $[P, D]$ is less than full column rank, hence the system defined in equation (A.3) is indeterminate and parameters π and δ cannot be identified.

In what follows I show that even though the absolute contribution of the plaintiff (π)

and of the defendant (δ) lawyers to the duration of trials cannot be identified, it is still possible to identify the contribution of each lawyer to the duration of trials with respect to the contribution of a reference lawyer. This is possible as long as every lawyer is connected to every other lawyer either directly or indirectly through other lawyers.

Before stating the general results which allows to identify the parameters of interest consider an example to clarify the intuition. Consider the following trials:

$$y_1 = \pi_3 + \delta_1 \tag{A.4}$$

$$y_2 = \pi_3 + \delta_2 \tag{A.5}$$

$$y_3 = \pi_2 + \delta_3 \tag{A.6}$$

$$y_4 = \pi_2 + \delta_4 \tag{A.7}$$

$$y_5 = \pi_2 + \delta_3 \tag{A.8}$$

$$y_6 = \pi_2 + \delta_4 \tag{A.9}$$

the system of equations (A.4)–(A.9) is indeterminate. To solves this system I would need to impose two restrictions on the parameters. For example, if I impose $\delta_1 = 0$ then from the first two equations

$$\delta_1 = 0 \Rightarrow \pi_3 = y_1 \Rightarrow \delta_2 = y_2 - y_1$$

and if I impose $\delta_3 = 0$ then from the last four equations

$$\delta_3 = 0 \Rightarrow \pi_2 = y_3 \Rightarrow \delta_4 = y_4 - y_3$$

hence only four out of six parameters are identified. Consider now the following variation of equations (A.4)–(A.9) where δ_3 is replaced by δ_1 in equation (A.8):

$$y_1 = \pi_3 + \delta_1 \tag{A.10}$$

$$y_2 = \pi_3 + \delta_2 \tag{A.11}$$

$$y_3 = \pi_2 + \delta_3 \tag{A.12}$$

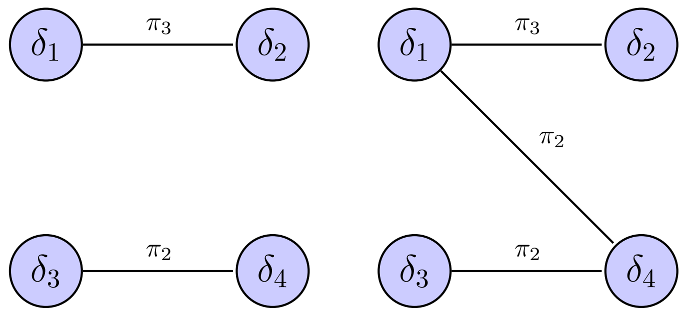
$$y_4 = \pi_2 + \delta_4 \tag{A.13}$$

$$y_5 = \pi_2 + \delta_1 \tag{A.14}$$

$$y_6 = \pi_2 + \delta_4 \tag{A.15}$$

both systems of equations defined in (A.4)–(A.9) and (A.10)–(A.15) are indeterminate,

Figure A.1: Notation: edges are plaintiffs and nodes are defendants.



(a) two connected components

(b) one connected component

however, the difference between the two is that in the former the first two equations do not have any parameter in common with the remaining four whereas in the latter δ_1 is common between the first two equations and the remaining four. For example, if I impose $\delta_1 = 0$ then all the remaining parameters are identified

$$\delta_1 = 0 \Rightarrow \pi_3 = y_1 \text{ and } \pi_2 = y_5 \Rightarrow \delta_2 = y_2 - y_1 \ ; \ \delta_3 = y_3 - y_5 \ ; \ \delta_4 = y_4 - y_5$$

hence five out of six parameters are identified.

The two systems of equations defined in (A.4)–(A.9) and (A.10)–(A.15) are represented as graphs in Figure A.1. Nodes represent defendant lawyers and edges represent plaintiff lawyers.[7] For example an edge 3 between nodes 1 and 2 means that lawyer 3 took part as plaintiff in a trial with lawyer 1 defendant and another trial with lawyer 2 defendant. In the language of graph theory, graph (a) in figure A.1 has two connected components because there is no path between nodes 1,2 and nodes 3,4. Whereas graph (b) in figure A.1 has only one connected component because the trials between lawyer 2 plaintiff and lawyers 1 and 4 defendants creates one path between every node.

Following Kramarz, Machin, and Ouazad (2010) consider now a general framework. $n_{p,d}$ is the number of trials between plaintiff p and defendant d. The **connection graph** for defendants G_d is an undirected graph $(\mathcal{D}, \mathcal{C})$ where $\mathcal{D}$ is the set of defendants and $\mathcal{C} \subset \mathcal{D} \times \mathcal{D}$ is the set of connections. Two defendants (d, d') are connected if and only if there is a plaintiff p that took part in trials both with defendant d and defendant d', i.e.

[7] this choice is arbitrary and nothing changes if one represents plaintiff lawyers as nodes and defendant lawyers as edges as shown in the appendix Figure A.2.

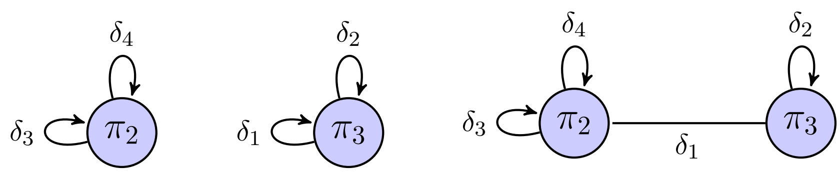
$\exists p$ s.t. $n_{p,d} > 0$ and $n_{p,d'} > 0$. The **connected components** of G_d are the sets of largest subgraphs of G_d that are each connected. **Sufficient connection**: the connection graph for defendants is *sufficiently connected* if it has only one connected component.

Proposition 3. *If the connection graph for defendants (or equivalently for plaintiffs) is sufficiently connected and* $E(\epsilon|P, D) = 0$ *then* $\hat{\pi}$ *and* $\hat{\delta}$ *are unbiased estimators of* π *and* δ*, up to a constant, i.e.*

$$E(\hat{\pi}|P, D) = \pi + K_\pi$$
$$E(\hat{\delta}|P, D) = \delta + K_\delta$$

For example, if $K_\pi = -\pi_1$ and $K_\delta = -\delta_1$ then the effect of each lawyers is identified relative to the effect of the lawyer with ID equal to 1.

Figure A.2: Notation: edges are defendants and nodes are plaintiffs.



(a) two connected components (b) one connected component

Proof. of proposition 3: See Kramarz, Machin, and Ouazad (2010). □

Abowd, Creecy, and Kramarz (2002) show the same result that within each connected components c, $N_c^1 - 1 + N_c^2 - 1$ effects are identified. Where N_c^1 and N_c^2 are the number of first and second level fixed effects in component c. The authors show that the effects π_p and δ_d in component c are identified up to constraints of the form:

$$\sum_{p \in \{comp.\ c\}} w_p \pi_p = 0 \quad \text{and} \quad \sum_{d \in \{comp.\ c\}} v_d \delta_d = 0$$

where w_p and w_d are arbitrary weights at least one of which is nonzero. For example a possible choice of weights is $v_1 = 1$ and $v_d = 0\ \forall d \neq 1$ as in the example above. Note that a different normalization is required for each component and, as argued by Cornelissen (2008), the estimated effects should be compared only within the same component. Comparing the estimated effects between different components is not meaningful because the effects are identified with respect to different normalizations.

Appendix 2.B Data Appendix

In the period 2002-2011, 412,395 trials were filed in a large Italian Labor Court. Of these trials, 97,459 (24% of the total) were special cases involving only one lawyer and thus irrelevant for the question studied in this paper. Of the remaining 314,936 trials, I had to further drop those in which information was missing on the key variables that I need for the econometric analysis:

- the identify of the judge assigned to the trial;
- the object of the controversy;
- the dates of filing and disposition of the trial, that are needed to compute its duration;
- the identify of the lawyers acting as plaintiff and defendant in the trial.

The identity of the judge, the object of the controversy and the date of filing are never missing. The situation is more complex for the date of disposition. First, 13,564 cases were not completed yet by July 1, 2014, which is the date in which the information at my disposal was extracted from the court's archive. Two years and a half after filing, these cases should have had plenty of time to complete. In fact the average duration for the non-censored cases is 385 days, while all the censored cases lasted at least 914 days and have an average duration of 3255 days. At this preliminary stage of the analysis I use the date of censoring as date of disposition for these cases, controlling econometrically for the occurrence of censoring and checking the robustness of my results with respect to alternative ways to address the censoring problem. Second, and less problematic given the small number of affected observations, 266 of the non–censored cases had a truly missing disposition date and had to be dropped from the analysis.

The existence of missing information on the identity of lawyers is more problematic for this study. Lawyers and their roles cannot be identified in 107,669 trials that have, therefore to be dropped as well from the analysis. I am trying to get in touch with the data provider to understand why the problem occurs. For the moment, I can only show that the missing information is missing almost at random from a substantive point of view. This is shown in Table B.1 which reports the mean of observable characteristics of the trials that I have to drop because of the missing identity of lawyers and of those that instead I can keep. Due to the large number of observations, the differences between the means of two groups is always statistically significant but quantitatively small. In the end, the trials that I can use for the econometric analysis are 207,001.

Table B.1: Comparison of trials with missing or non-missing lawyers' ID

	Averages		
Variables	ID lawyers present	ID lawyers missing	p–value for H_0: equal means
Object of controversy:			
% of trials with given object			
Compensation	0.17	0.21	.000
18.45%	(0.37)	(0.41)	
Attendance allowance	0.20	0.25	.000
21.69%	(0.40)	(0.43)	
Other hypothesis	0.13	0.09	.000
11.99%	(0.34)	(0.29)	
Other controversies	0.07	0.10	.000
8.18%	(0.26)	(0.30)	
Disability living allowance	0.02	0.02	.625
2.27%	(0.15)	(0.15)	
Pension	0.05	0.06	.000
5.09%	(0.21)	(0.23)	
Temporary work contract	0.06	0.03	.000
5.11%	(0.24)	(0.17)	
Termination of employment	0.08	0.07	.000
7.78%	(0.28)	(0.25)	
Type of employment relationship	0.03	0.02	.000
2.31%	(0.16)	(0.14)	
Other types	0.18	0.15	.000
17.14%	(0.39)	(0.36)	
Timing information:			
Filing date	30 April 2007	1 November 2006	.000
	(1068.08)	(1060.92)	
Trial's length (days)	525.16	478.68	.000
	(677.13)	(649.68)	
Observations	207,001	107,669	

Notes: values in parentheses are standard deviations. The p–value refers to the t–tests of the equality of means between trials in which the ID of lawyers is missing and trials in which the ID of lawyers is present. By ID of lawyers missing I mean: the ID of either the plaintiff or of the defendant lawyer is missing, or the ID of the plaintiff and defendant lawyers coincide. Only trials with disposition date not missing are considered in this table.

3 The Adverse Incentive Effects of Heterogeneity in Tournaments–Empirical Evidence from The German Bundesliga

Tournaments may motivate workers to provide effort, yet differences in relative abilities may undermine the incentives of workers to exert effort. I use a novel data set from professional football competitions and find that differences in relative abilities are associated with lower effort exerted by players. In this empirical setting, effort and relative abilities are measured as, respectively, the distance covered on the pitch by football players and relative winning probabilities, the latter derived from betting odds of professional bookmakers. Assuming that betting odds control for all unobservable variables affecting at the same time relative abilities and effort, I find that larger differences in betting odds of opposing teams lead to less distance covered on the pitch.

3.1 Introduction

Competition encourages workers to provide effort. In fact, tournament–style competitions in which firms reward workers according to their relative performance are common in many contexts, pitting workers against each other for promotion and awards. In their seminal work, Lazear and Rosen 1981 show that tournaments alleviate the principal–agent problem by tying workers' rewards to their relative output. The effectiveness of tournaments is, however, inversely related to the degree of heterogeneity between tournaments' participants. For example, if workers of unequal abilities compete, then the weaker (optimally) gives up whereas the stronger wins by providing little effort. In the end, both weak and strong players slack off when they compete against each other.

This theoretical prediction is tested using data from the German Bundesliga (Germany's primary football competition). This data allows to measure both effort, as the distance covered by football players on the pitch, and heterogeneity, as teams' relative winning probabilities, the latter derived from betting odds of professional bookmakers. I find that when heterogeneity increases by 1%, players exert 0.7% less effort. For example, if the ratio between the winning probabilities of two opposing teams increases from 1(completely homogenous teams) to 1.01, then, at the mean distance covered of 120 kilometers, both teams cover approximately 1 kilometer less on the pitch.

A key contribution of this paper is that, differently than in the literature, football is a context in which effort can be explicitly measured as numbers of kilometers run by each player. Also key and novel in my analysis is the use of betting odds to proxy football teams' relative abilities.

My identifying assumption relies on the betting market being efficient, hence prices (betting odds) should incorporate all information simultaneously affecting effort and relative difference, including differences in abilities, between opposing teams. Given this assumption, the playing strengths, player injuries, the importance of a single match for a team and all other variables affecting the provision of effort should be incorporated in the betting odds. Therefore, measuring heterogeneity between teams using winning probabilities allows to control for potential confounding factors.

I complement my analysis with further evidence supporting distance covered as measure of effort. First, I show that a team is more likely to win when its players cover more distance on the pitch, for a given distance covered of the opposing team and controlling for teams fixed effects. Second, I rule out that distance covered measures output instead of effort by showing that, over the course of a season, strong and weak teams cover on average the same distance on the pitch. That is, strong teams are better at converting distance covered (input) into goals scored (output) than weak teams even though both strong and weak teams exert on average the same effort.

The difficulty of testing predictions of Tournament Theory comes from the fact that measures of effort and ability are often not available in Personnel Economics databases. This study contributes to the literature by proposing a novel measure of effort as the distance covered by football players on the pitch. Previous works have relied on measures of output as proxies for effort.

The remainder of this paper is structured as follows. Section 3.1.1 summarizes the literature related to my work. Section 3.2 presents a tournament model and describes the identification strategy. Section 3.3 describes the data. Section 3.4 presents the empirical results. Section 3.5 discusses the attenuation bias of my estimates. Section 3.6 presents further evidence to justify distance covered as a measure of effort. Section 3.7 rules out alternative explanations for the results of section 3.4. Section 3.8 concludes.

3.1.1 Related literature

Since Lazear and Rosen 1981 seminal paper there has been a large theoretical and empirical literature studying tournaments. Given the limitations of Personnel Economics databases, several papers have relied on data from professional sport competitions.

Brown 2011 exploits the presence of a superstar golfer (Tiger Woods) as an exogenous variation in the relative abilities between golfers to show that when Tiger Woods participates in a game all other golfers score worse. This identifies the effect of the presence of a superstar player on the performances of players who are weaker than the superstar. Instead, I estimate the adverse effect of variations in relative abilities on the provision of effort of all players, strong and weak players alike. Sunde 2009; Lallemand, Plasman, and Rycx 2008 examine professional tennis data to study heterogeneity in elimination tournaments. They find that lower–ranked players tend to underperform in more heterogeneous matches. Nieken and Stegh 2010 use data from the German Hockey League, while Berger and Nieken 2010 use data from the Handball–Bundesliga to study heterogeneity in tournaments. While the two papers use different datasets, they both use the number of two minutes penalties as a measure of effort. Bach, Gürtler, and Prinz 2009 use the end time to finish the Olympic Rowing Regatta as a measure of effort. All these works use measures of output as implicit measures of effort (input), whereas I propose a novel measure of effort.

Deutscher, Frick, Gürtler, and Prinz 2013 find that low ability teams exert more sabotage than high ability teams using data from the German Bundesliga. They also use betting odds to determine relative differences of opposing teams. Balafoutas, Lindner, and Sutter 2012 use data from Judo competitions and they find that higher costs of sabotage lead to less sabotage. They also find that contestants of relatively higher ability are more likely to be the targets of sabotage and less likely to engage in it.

3.2 Theoretical framework and identification

3.2.1 Tournament Model

This section introduces the theoretical framework to study the effect of heterogeneity between teams on their effort provision. I derive a simple game theoretical model to define the hypothesis to be tested empirically. Note that throughout this paper teams are treated as unitary players.

The team (i) and its opponent ($-i$) play a pairwise match. Both teams chose efforts simultaneously. The marginal costs of effort are c_{im} and c_{-im} for the team and its opponent in match m. There is only one winner for each match. The benefits of winning the match m for the team and for its opponent are $B_{im} > 0$ and $B_{-im} > 0$. The benefits of not winning the match are normalized to 0 for both team and opponent. There is no private information, all teams are risk neutral and they have linear costs of effort. The (endogenous) probability that team i wins against its opponent $-i$ in a match m is given by the following Tullock contest success function, (see Tullock 2001):[1]

$$P_{im} = \frac{a_{im}e_{im}}{a_{im}e_{im} + a_{-im}e_{-im}} \tag{3.1}$$

where a_{im}, a_{-im} are the abilities and e_{im}, e_{-im} are the efforts of team i and its opponent $-i$ in match m. Team i solves the following problem:

$$\underset{e_{im}\geq 0}{\text{Max}} \quad \left\{ \frac{a_{im}e_{im}}{a_{im}e_{im} + a_{-im}e_{-im}} B_{im} - c_{im}e_{im} \right\}$$

[1] Appendix 3.A provides a micro foundation of the Tullock contest success function in terms of output produced by the team.

Before describing the equilibrium of the game I should briefly explain how the Bundesliga tournament works. The Bundesliga is contested by 18 teams in each season running from August to May. Each team plays every other team once at home and once away. A victory is worth three points, a draw is worth a single point and zero points are given for a loss. The club with the most points at the end of the season becomes German champion. The top three clubs in the table qualify automatically for the group phase of the UEFA Champions League, while the fourth-place team enters the Champions League at the third qualifying round. The two teams at the bottom of the table are relegated into the 2nd Bundesliga, which is the Second Division of professional football in Germany.

In principle each team solves a dynamic model for the optimal choice of effort in each match. Instead of modeling such a complicated scenario, I take a shortcut and allow the following three exogenous variables: ability (a_{im}), benefit of winning the match (B_{im}) and marginal cost of effort (c_{im}) to be match specific and hence capture the dynamic features of the Bundesliga tournament. For example, if team i is in the third or second position from the bottom of the table in its last match m of the season, then team i has an almost null marginal cost of effort (c_{im}) because it is the last match of the season. In the meanwhile team i has a very high benefit of winning the match (B_{im}) because by gaining a few extra points it can avoid relegation.

Therefore, even the static tournament model captures the dynamic features of the Bundesliga tournament once the exogenous variables determining the provision of effort are allowed to change in each match of the Bundesliga tournament.

The first order condition for the problem of team i is given by

$$\frac{a_{im}a_{-im}e_{-im}}{(a_{im}e_{im}+a_{-im}e_{-im})^2}=\frac{c_{im}}{B_{im}} \tag{3.2}$$

symmetrically for opponent $-i$

$$\frac{a_{im}a_{-im}e_{im}}{(a_{im}e_{im}+a_{-im}e_{-im})^2}=\frac{c_{-im}}{B_{-im}} \tag{3.3}$$

Defining $V_{im}\equiv B_{im}/c_{im}$ and $V_{-im}\equiv B_{-im}/c_{-im}$ as the values of match m for the team and its opponent reduces the number of exogenous variables. Intuitively, increasing the benefit of winning the match or decreasing the cost of effort for the same match have the same effect on the provision of effort. Combining (3.2) and (3.3) yields the following expression for the provision of effort of team i:

$$e_{im}(V_{im},H_{im})=V_{im}\frac{H_{im}}{(1+H_{im})^2} \tag{3.4}$$

where the heterogeneity between teams in match m, H_{im}, is defined as following:

$$H_{im}\equiv\frac{a_{im}V_{im}}{a_{-im}V_{-im}}$$

all the differences between teams in a given match m are summarized by H_{im}. If $H_{im}=1$, there is no heterogeneity in match m. The following remark summaries the key empirical hypothesis of this paper:

Remark 3. *Effort decreases as the heterogeneity between teams increases and it is maximized when there is no heterogeneity:*

$$\begin{aligned}
\frac{\partial e_i(\bar{V}_{im},H_{im})}{\partial H_{im}}&>0 \quad \textit{if} \;\; 0<H_{im}<1\\
\frac{\partial e_i(\bar{V}_{im},H_{im})}{\partial H_{im}}&=0 \quad \textit{if} \;\; H_{im}=1\\
\frac{\partial e_i(\bar{V}_{im},H_{im})}{\partial H_{im}}&<0 \quad \textit{if} \;\; H_{im}>1
\end{aligned}$$

Figure 3.1 plots effort e_{im} as a function of heterogeneity H_{im} (for a given V_{im}). Effort is maximized when there is no heterogeneity, i.e. $H_{im}=1$.

3.2.2 Identification

The empirical challenge in testing remark 3 comes from the fact that the following exogenous variables: V_{im}, V_{-im}, a_{im} and a_{-im} are not observed, however, using betting odds from professional bookmakers it is possible to construct the winning probabilities of teams. This section shows that observing the winning probabilities of teams is equivalent to observing all exogenous variables determining the degree of heterogeneity between teams.

Taking the ratio between (3.2) and (3.3) yields the following relation:

$$\frac{e_{im}}{e_{-im}} = \frac{V_{im}}{V_{-im}} \tag{3.5}$$

Moreover, taking the ratio between the winning probabilities of team i and its opponent $-i$, (given by the Tullock contest success function (3.1))

$$\frac{P_{im}}{P_{-im}} = \frac{a_{im}e_{im}}{a_{-im}e_{-im}} \tag{3.6}$$

combining (3.6) and (3.5) gives the following relation between the ratio of the endogenous winning probabilities and the exogenous variables determining the heterogeneity between team and opponent:

$$\frac{P_{im}}{P_{-im}} = \frac{a_{im}V_{im}}{a_{-im}V_{-im}} = H_{im} \tag{3.7}$$

Remark 4. *the ratio between the winning probabilities captures all unobservable variables determining the degree of heterogeneity between teams in each match.*

Figures 3.2 and 3.3 give a graphical representation of remark 4. Figure 3.2 shows that heterogeneity is equal to the ratio of winning probabilities. Hence, the equilibrium relation between effort and the ratio of winning probabilities shown in figure 3.3 is equal to the causal effect of heterogeneity on effort shown in figure 3.1. For example, if a key player of team i is injured and cannot play in match m, then the ability of team i decreases in match m and the winning probabilities adjust accordingly. Also, if winning match m is important for team i or the cost of effort of team i is low in match m, then the value of the match is high and the winning probabilities incorporate this.

3.3 Data

The effort of a team in a given match is measured as the sum of all the distance covered by its players on the pitch. I use a novel data set which I created using the statistics freely available on the website of the German Bundesliga.[2] This data set includes information on the distance covered (measured in km) by each team in each match of the following three seasons of the Bundesliga: 2014–2013, 2013–2012 and 2012–2011. Over the three seasons considered there are 918 matches but due to some missing observations concerning statistics on distance covered there are 905 matches which can be used for the analysis. I consider this 13 missing observations to be random since they are presumably due to a temporary malfunctioning of the tracking system. This is shown in tables A.1 and A.2 which report the mean of observable characteristics of the matches that I have to drop because of the missing distance covered of teams on the pitch and of those that instead I can keep.

Since the outcome variable of interest is distance covered by the team, each match gives two distinct observations, one for the team and another for its opponent, for a total of 1810 observations.

The equilibrium winning probabilities are computed from the betting odds of professional bookmakers, (historical data on betting odds are freely available online[3]). This data set

[2] http://www.bundesliga.com/en/stats/matchday
[3] http://football-data.co.uk/germanym.php

contains historical data from seven different bookmakers and in order to have more precise estimates of the winning probabilities, for each match I consider the average between the betting odds of different bookmakers.

In each season of the Bundesliga every team has to play against every other team twice. This is good for two reasons. First, teams cannot select their opponent, hence any bias due to teams sorting into particular matches can be ruled out. Second, all combinations of heterogeneity between opposing teams are observed.

3.4 Empirical Results

In order to fit model (3.4) to the data I consider the following functional form:

$$e_{im} = \exp\{\beta_0\} \left(\frac{H_{im}}{(1 + H_{im})^2} \right)^{\beta_1} \exp\{u_{im}\} \tag{3.8}$$

taking the natural logarithm of equation (3.8) yields the following log-linear model:

$$\log(e_{im}) = \beta_0 + \beta_1 \log \left(\frac{H_{im}}{(1 + H_{im})^2} \right) + u_{im} \tag{3.9}$$

where e_{im} is the total distance covered by team i in match m and H_{im} is the ratio between the winning probabilities of team i and its opponent in match m. u_{im} is given by the following equation:

$$u_{im} = \delta \log(V_{im}) + \xi_{im} \tag{3.10}$$

which is the unexpected innovation in effort of team i in match m, (ξ_{im}), plus the effect of the value of match m for team i, (V_{im}). ξ_{im} captures all variations in effort which are not anticipated by the betting market; for instance the injury of a player during the match. V_{im} is directly proportional to the benefit that team i receives from winning match m and inversely proportional to the cost of effort of team i in match m.

The parameter β_1 is not of interest by itself, but only to the extent to which it is informative on the effect of heterogeneity on effort. It is more interesting to consider the elasticity of effort with respect to heterogeneity given by the following equation:[4]

$$\varepsilon(H_{im}) \equiv \frac{\partial \mathbb{E}\{\log(e_{im})|H_{im}\}}{\partial \log(H_{im})} = \beta_1 \left[1 - 2\frac{\partial \log(1+H_{im})}{\partial \log(H_{im})}\right] \tag{3.11}$$

$$= \beta_1 \frac{1-H_{im}}{1+H_{im}} \tag{3.12}$$

given that V_{im} is not observed, the estimates of this elasticity are biased. However, as section 3.5 shows, the estimated elasticity is downward biased. Therefore, the results of this section are conservatives estimates of the elasticity of effort with respect to heterogeneity.

Figure 3.4 plots the relation between effort and heterogeneity, both variables are transformed in their natural logarithm. Effort is maximized when two teams have similar odds of winning the match and it decreases as teams have different probabilities of winning the match. Figure 3.5 plots the estimates of the elasticity of effort with respect to heterogeneity given by equation (3.12) for different values of H_{im}.

Computing $\varepsilon(H_{im})$ at the mean value of H_{im} gives the average effect of heterogeneity on the provision of effort. The first column of table 3.1 shows that the average estimated elasticity of effort with respect to heterogeneity is -.677%. This results confirms the prediction of tournament theory, higher heterogeneity causes players to exert less effort.

To determine whether this effect is big or small it should be compared to the effect of other instruments inducing players to exert more effort. For example, players exert more effort if the monetary prize of the tournament increases. I cannot easily compute

[4] The last equality follows from the fact that $y = \log(x) \Leftrightarrow x = e^y$, hence

$$\frac{\partial \log(1+x)}{\partial \log(x)} = \frac{\partial \log(1+e^y)}{\partial y} = \frac{e^y}{1+e^y} = \frac{x}{1+x}$$

the monetary prize of each match of the Bundesliga, however, in their two papers using data from golf competitions, Ehrenberg and Bognanno 1990a,b, report that the elasticity of effort with respect to the monetary prize is .035%. If the monetary prize of the tournament increases by 1%, then golfers increase on average effort by .035%. Therefore, the degree of heterogeneity between tournament participants plays an important role in the provision of effort, not less important than the monetary prize of the tournament.

As a robustness check, the estimates of the effect of heterogeneity on the provision of effort should not change once control variables are included in the empirical model. For example, the results should not change if the ability of teams is included in the model, since ability is already measured by the betting odds. The control variables considered here for the abilities of teams are the teams fixed effects. (The same team is treated as a different team for each season to take into account the fact that teams change characteristics from one season to the other). The last column of table 3.1 shows that the estimated mean elasticities of effort with respect to heterogeneity do not vary considerably between different models. As theory predicted, the ratio between the winning probabilities of teams already measures the relative abilities of teams in each match.

3.5 Characterization of the Bias

This section shows that the estimates presented in section 3.4 are conservative due to the fact that the value of each match for each team is not observed.

Consider the error term given by equation 3.10 and the following function for convenience:

$$g(H_{im}) = \frac{H_{im}}{(1 + H_{im})^2}$$

theory says that $\delta > 0$, while the sign of $Cov(V_{im}, g(H_{im}))$ can be inferred from $\partial g / \partial V_{im}$

since H_{im} is itself a function of V_{im}. By the chain rule,

$$\frac{\partial g}{\partial V_{im}} = \frac{\partial g}{\partial H_{im}} \cdot \frac{\partial H_{im}}{\partial V_{im}}$$

where

$$\frac{\partial H_{im}}{\partial V_{im}} = \frac{a_{im}}{a_{-im}V_{-im}} > 0$$

and

$$\begin{cases} \frac{\partial g}{\partial H_{im}} > 0 \quad \text{if} \quad H_{im} < 1 \\ \frac{\partial g}{\partial H_{im}} < 0 \quad \text{if} \quad H_{im} > 1 \end{cases}$$

since

$$\frac{\partial g}{\partial H_{im}} = \frac{1 - H_{im}}{(1 + H_{im})^3}$$

hence the following remark:

Remark 5.

$$\frac{\partial g}{\partial V_{im}} < 0 \quad for \quad H_{im} > 1$$

This implies $Cov[\log(V_{im}), \log(g(H_{im}))] < 0$ for values of H_{im} larger than 1. Therefore, the estimates of β_1 are downward biased, (see appendix 3.B for derivation of this result).

Remark 6.

$$\mathbb{E}\{\hat{\beta}_1 | H_{im}\} < \beta_1$$

Table 3.1 confirms that the estimate of β_1 is downward biased because it is smaller than 1, which would be the value of β_1 predicted by the tournament model presented in section 3.2.1.

Recall that the elasticity of effort with respect to heterogeneity is

$$\varepsilon(H_{im}) = \beta_1 \frac{1 - H_{im}}{1 + H_{im}} \tag{3.13}$$

and the estimated elasticity is

$$\hat{\varepsilon}(H_{im}) = \hat{\beta}_1 \frac{1 - H_{im}}{1 + H_{im}} \tag{3.14}$$

given that the estimate of β_1 is downward biased, the absolute value of the estimated elasticity is smaller than the absolute value of the true elasticity.

Remark 7.

$$|\hat{\varepsilon}(H_{im})| < |\varepsilon(H_{im})| \quad \forall \quad H_{im} \in (1, +\infty).$$

Therefore, the estimates of the elasticity of effort with respect to heterogeneity are conservative for values of H_{im} larger than 1.

3.6 Further Evidence: Effort and the Probability of Winning

In this section I provide further evidence to justify the use of the distance covered by players on the pitch as a measure of effort. As common in tournament literature equation (3.1) defines the winning probability of a team as an increasing function of the effort and the ability of the team, and as a decreasing function of the effort and the ability of the opponent. I test this hypothesis with the following econometric specification:

$$y_{im} = \alpha_0 + \alpha_1 e_{im} + \alpha_2 e_{-im} + f_i + f_{-i} + v_{im} \tag{3.15}$$

where y_{im} is a binary variable taking value 1 if team i won match m and 0 otherwise, e_{im} and e_{-im} are respectively the distances covered by team i and its opponent $-i$ in match m. f_i and f_{-i} are team and its opponent fixed effects. v_{im} is the error term. If distance covered is an appropriate measure of effort, then $\alpha_1 > 0$ and $\alpha_2 < 0$, as tournament theory predicts. Table 3.2 confirms this by reporting the estimates of these parameters from the linear probability model of equation (3.15).

In order to claim that α_1 and α_2 are identified as the effect of effort on the winning probability of the team I need to make the following assumption: teams fixed effects measure the ability of teams. In fact, the ability of a team affects both its probability of winning and its effort. Failing to control for ability can result in biased estimates of α_1 and α_2 for the effect of interest.

I cannot measure the ability of teams directly, but I can control for teams fixed effects. To asses how much the ability of a team changes from one match to the other I use different definition of teams fixed effect in each column of table 3.2. In the first column each team has a fixed effect for all the matches it played over the three seasons considered. In the second column each team has a season specific fixed effect for all the matches it played in a given season. For example, Bayern Munich playing the 2012-2011 season is considered as a different team from Bayern Munich playing the 2014-13 or the 2013-2012 seasons. In other words, Bayern Munich is considered as if it was three different teams. In the remaining columns each team is treated as a different team a progressively higher number of times. For example in the last column Bayern Munich is treated as 18 different teams. The fact that the results of different columns are similar suggests that the ability of teams does not change significantly over time. Therefore, including teams fixed effects controls for the unobserved ability of teams.

3.7 Alternative Explanations

3.7.1 Effort, Output or Ability?

The results of sections 3.4 and 3.6 show that distance covered measures the effort of teams. This subsection rules out the alternative explanations that distance covered is instead a measure of the ability of teams or a measure of the output of teams.

If distance covered was a measure of ability or output instead of effort, then one should observe higher ability teams to cover on average more distance than lower ability teams. The final ranking of teams offers a measure of the average ability of teams in each season. The best team of each season is placed first in the ranking while the worst team is placed 18th. Table 3.3 shows that the nine best teams and the nine worst teams cover on average the same distance in each season. This rules out the possibility that distance covered is measuring either the ability or the output of teams.

The fact that high and low ability teams cover on average the same distance does not contradict the fact that high ability teams are more likely to win individual matches than low ability teams, (winning is the output of the team). As shown in subsection 3.2.1 and section 3.6, the probability of winning of the team is an increasing function of the ability of the team as well as of the effort of the team, equation (3.1). Hence, if two teams cover the same distance on the pitch the high ability team is still more likely to win.

3.7.2 Mechanical Effect

If high ability teams keep low ability teams in their side of the pitch, then teams cover less distance on the pitch in more heterogeneous matches because all the action is taking place in a restricted area of the pitch. If this was the case, then the result of section 3.4 could be due to this mechanical effect instead of the incentive effect of heterogeneity on the provision of effort. This alternative explanation is ruled out for two reasons.

First, the fact that high ability teams keep low ability teams in their side of the pitch depends on the style of play of both teams, which is accounted for by including teams fixed effects.

Second, as table 3.4 shows, high ability teams do not keep low ability teams in their side of the pitch. Table 3.4 reports the number of times all players of a team touched the ball in a given area of the pitch over the total number of times all players of that team touched the ball in any area of the pitch.[5] The pitch is divided in three areas: the own half of the team, the middle and the opponent half. The data shows that the percentage of touches in the opponent half of the pitch by high ability teams is not significantly different from the percentage of touches in the opponent half of the pitch by low ability teams. Therefore, high ability teams do not keep low ability teams in any particular side of the pitch.

3.7.3 Calendar of the Bundesliga

Players may get tired during the season, they may have more energies to run at the beginning of the season than at the end. One might worry that if more homogeneous matches take place at the beginning of the season, then players run more on the pitch not because they react to incentives but simply because they have more energies to run.

This explanation is ruled out because the calendar of each season of the Bundesliga is randomly determined. For example, the fact Augsburg played against Bayern Munich on November the 6th in 2011 instead of any other day was randomly decided.

Table 3.5 shows that heterogeneity between matches is not correlated with the day on which each match is scheduled. Therefore, more homogenous matches are uniformly distributed throughout different matchdays of each season.

3.8 Conclusion

This paper shows that differences in relative players' characteristics, including abilities, reduce players' effort in tournaments. The key to my research design is that, using data from the German Bundesliga, effort can be measured explicitly as the number of kilometers

[5]this data is freely available here: http://www.whoscored.com

run by football players on the pitch. This represents a contribution to the exist literature which relies on measures of output to indirectly infer effort. Moreover, in an efficient betting market, prices represented by betting odds should capture all information determining teams' provision of effort. Therefore, differences in relative players' characteristics measured with betting odds of professional bookmakers account for all factors determining effort provision.

The following two facts provide additional evidence supporting distance covered on the pitch as a measure of effort. First, the probability of winning of a team increases as its distance covered on the pitch increases, holding constant, the distance covered by its opponent, the ability of the team, and the ability of the opponent. Second, on average higher ability teams do not cover more distance than lower ability teams, ruling out that distance covered is measuring output instead of effort.

Tournaments are widely used to motivate workers. My empirical results show, however, that the effectiveness of these promotion schemes is severely reduced when workers differ in their relative abilities.

Figure 3.1: The measure of heterogeneity is on the horizontal axis, while effort is on the vertical axis. Effort is maximized when there is no heterogeneity between teams, ($H_{im} = 1$).

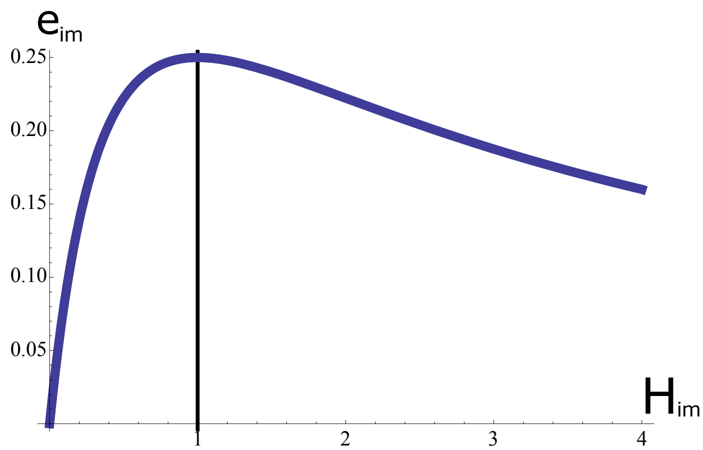


Figure 3.2: Proportional relation between the exogenous variable H_{im} and the endogenous variable P_{im}/P_{-im} in equilibrium.

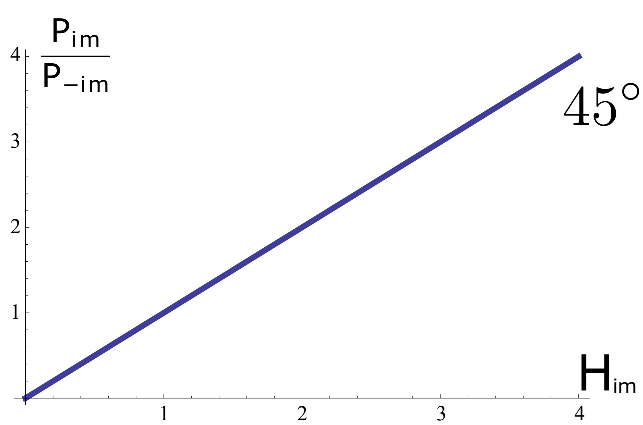


Figure 3.3: The equilibrium relation between effort and the ratio of winning probabilities is the causal effect of heterogeneity on effort.

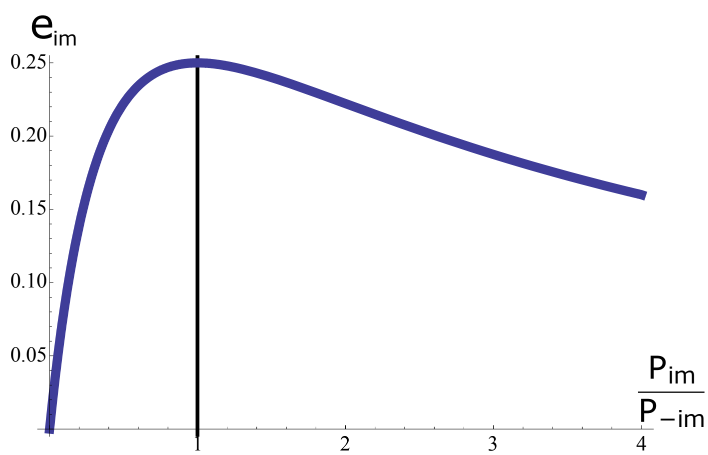
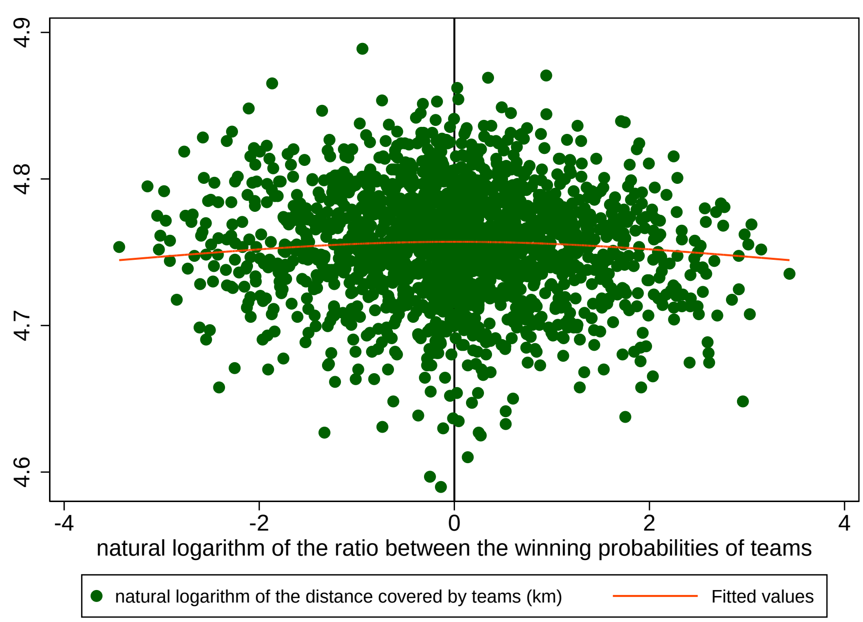
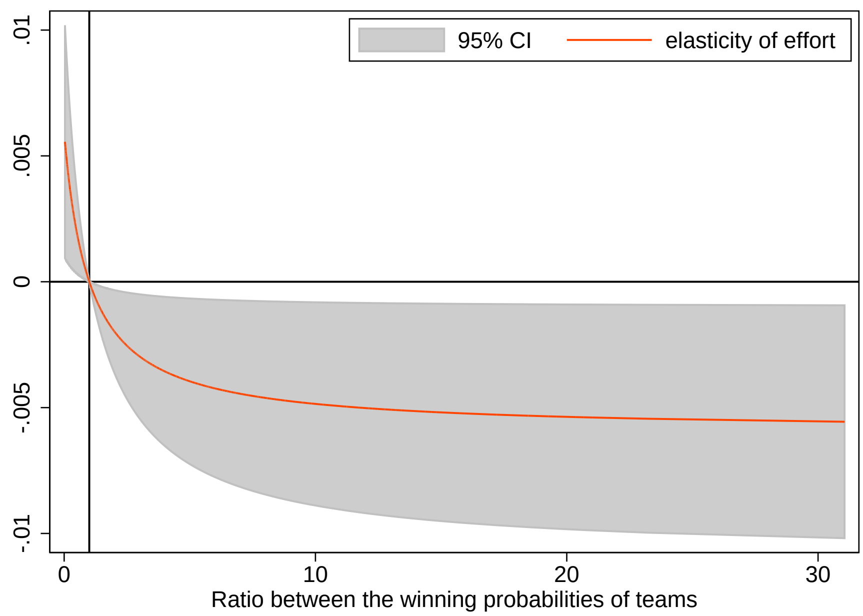


Figure 3.4: The adverse effect of heterogeneity on the provision of effort.



Notes: the fitted values are computed from model (3.9). The highest level of effort is reached when there is no heterogeneity, that is when the two teams have the same probability of winning the match.

Figure 3.5: The elasticity of effort with respect to heterogeneity estimated from function $\varepsilon(H_{im})$ in equation (3.12) for different values of heterogeneity (H_{im}).



Notes: each point estimate is associated to its 95% confidence intervals given by the shaded area. Standard errors are computed by clustering at each match.

Table 3.1: Elasticity of effort with respect to heterogeneity

	(1)	(2)
$\hat{\varepsilon}(\bar{H})\%$	-0.677%	-0.641%
	(0.0017)	(0.0027)
$\hat{\beta}_1$	0.013	0.012
	(0.0033)	(0.0053)
Teams Fixed Effects	✗	✓
Range of values of H_{im}	[1, 31]	[1, 31]
Observations	905	905

Notes: this table reports estimates of the effect of a 1% increase in heterogeneity, (measured by the betting odds), between opposing teams on the distance covered on the pitch by the same two teams. If heterogeneity increase by 1%, then teams cover on average 0.7% less distance on the pitch. Teams fixed effects are constructed considering the same team playing in different seasons as a different team. Standard errors in parenthesis are computed clustering by each match.

Table 3.2: Estimates of the effect of the efforts of the team and of its opponent on the probability of winning of the team.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	y_{im}	y_{im}	y_{im}	y_{im}	y_{im}	y_{im}	y_{im}
e_{im}	0.0419	0.0489	0.0425	0.0375	0.0411	0.0421	0.0546
	(0.00465)	(0.00535)	(0.00530)	(0.00655)	(0.00692)	(0.00849)	(0.0113)
e_{-im}	-0.0546	-0.0636	-0.0548	-0.0533	-0.0546	-0.0521	-0.0598
	(0.00422)	(0.00471)	(0.00495)	(0.00566)	(0.00610)	(0.00788)	(0.00948)
Teams FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of FE for each team	1	3	6	9	12	15	18
Number of matches	905	905	905	905	905	905	905

Notes: The dependent variable is the binary outcome (winning or not) of the match for the team. (Draws are treated as loses in the definition of the outcome variable). Each column of the table differs by the number of times that a team is treated as a different team in the definition of its fixed effects. Column (1) is the simplest case in which each team is treated as the same team in all matches. In column (2) each team is treated as a different team in each season of the Bundesliga, hence for each team there are 3 fixed effects because the sample covers three seasons of the Bundesliga. In the remaining columns each team is treated as a different team a progressively higher number of times. For example in column (7) each team is treated as 18 different teams. Hence, in the first column the ability of each team is assumed to remain constant for three years, whereas in the last column the ability of each team is assumed to remain constant only in a few matches. The fact that the results of different columns are similar suggests that the ability of teams does not change significantly over time. Standard errors in parenthesis are robust to heteroscedasticity.

Table 3.3: Average Distance Covered in each match by Favorite and Underdogs Teams

	Average Distance Covered (km) per Season by Teams		
Season	Favorite Teams	Underdog Teams	H_0 : Favorite and Underdog Teams Cover the Same Distance on the pitch (t–Test)
2014–2013	117.1 (1.32)	116.8 (2.15)	p–value=0.766
2013–2012	115.4 (2.83)	116.9 (1.35)	p–value=0.189
2012–2011	115.8 (2.78)	115.6 (3.32)	p–value=0.886

Notes: the table reports the average distance covered (measured in kilometers) by the top nine teams (favorite teams) and the bottom nine teams (underdogs teams) in each season. The second column refers to teams whose ranking at the end of a given season was above the median (favorite teams) while the third column refers to teams whose ranking at the end of a given season was below the median (underdogs teams). For example, in the 2014–2013 season the strongest nine teams covered on average 117 kilometers while the weakest nine teams covered on average 116.8 kilometers in each match. This suggests that distance covered cannot be a measure of ability because stronger teams do not covered more distance on the pitch than weaker teams. Values in parenthesis are standard deviations.

Table 3.4: Average Percentage of Ball Touches by Favorite and Underdog Teams on Different Sides of the Pitch

	Favorite Teams			Underdog Teams			H_0 : Favorite and Underdog have the same percentage of touches in
Season	Own Half	Middle	Opponent Half	Own Half	Middle	Opponent Half	the opponent side (t–Test)
2014–2013	27.4	43.8	28.6	29	43.3	27.6	p–value=0.368
	(0.024)	(0.012)	(0.031)	(0.015)	(0.015)	(0.007)	
2013–2012	26.1	46.4	27.4	27.1	46.2	26.6	p–value=0.460
	(0.025)	(0.013)	(0.026)	(0.010)	(0.009)	(0.015)	
2012–2011	26.1	46.8	27	26.8	46.3	26.7	p–value=0.853
	(0.023)	(0.017)	(0.030)	(0.010)	(0.011)	(0.018)	

Notes: the table reports the percentage of touches in different sides of the pitch by the top nine teams (favorite teams) and the bottom nine teams (underdogs teams) in each season. Favorite teams are teams whose ranking at the end of a given season was above the median while underdog teams are teams whose ranking at the end of a given season was below the median. The percentage of touches of a team in a given area of the pitch is defined as the number of times every player of the team touched the ball in that area over the total number of times every player of the team touched the ball in any area. For example 28.6 and 27.6 in the first row mean that on average over the 2014-2013 season 28.6 and 27.6 percent of the touches of the favorite and underdog teams were in the opponent side of the pitch. This suggests that stronger teams do not keep weaker teams on one side of the pitch, otherwise the percentage of touches of the stronger teams in the opponent side should have been significantly higher than the percentage of touches of the weaker team in the opponent side. Values in parenthesis are standard deviations.

Table 3.5: Random Calendar of the Bundesliga

	Dependent Variable: Ratio Winning Probabilities home/away team
Matchday Season 2014–2013	-0.00946
	(0.0116)
Matchday Season 2013–2012	-0.0156
	(0.0119)
Matchday Season 2012–2011	-0.0147
	(0.0121)
Number of Matches	918

Notes: this table shows that heterogeneity between matches is not correlated with the day on which each match is scheduled (i.e. the matchday of a given season). Standard errors in parenthesis are robust to heteroscedasticity.

Appendix 3.A Micro Foundations of the Tullock Contest Success Function (CSF)

In section 3.2 I have taken the Tullock CSF as a primitive. There are various justifications of the Tullock CSF in the literature (see Konrad et al. 2009 for a literature review on the Tullock CSF). The one I consider here is the simplest and it assumes that players' output is a multiplicative function of ability, effort and a random luck component distributed as an exponential with mean 1. The output of player i takes the following form: $y_i = a_i \cdot e_i \cdot \epsilon_i$, where a_i, e_i and ϵ_i are respectively the ability, the effort and the random luck component of player i.

Consider an arbitrary fixed value of $\epsilon_j = \bar{\epsilon}_j$. The probability that the output of player i is lager than the output of his opponent player j is,

$$\begin{aligned} P(y_i \geq \bar{y}_j) &= P(a_i e_i \epsilon_i \geq a_j e_j \bar{\epsilon}_j) \\ &= 1 - P\left(\epsilon_i < \frac{a_j e_j \bar{\epsilon}_j}{a_i e_i}\right) \\ &= \exp\left\{-\frac{a_j e_j \bar{\epsilon}_j}{a_i e_i}\right\} \end{aligned}$$

To compute the probability that the output of player i is larger than the output of player j for all realization of ϵ_j one needs to integrate the expression above over ϵ_j. Hence,

$$\begin{aligned} P(y_i \geq y_j) &= \int_0^{+\infty} \exp\left\{-\frac{a_j e_j \bar{\epsilon}_j}{a_i e_i}\right\} \exp\{-\epsilon_j\} d\epsilon_j \\ &= \int_0^{+\infty} \exp\left\{-\frac{a_i e_i + a_j e_j}{a_i e_i}\ \epsilon_j\right\} d\epsilon_j \\ &= \frac{a_i e_i}{a_i e_i + a_j e_j} \end{aligned}$$

which is the Tullock CSF introduced in section 3.2.

Appendix 3.B Downward Bias of Estimates

Consider the following model:

$$\log(e_{im}) = \beta_0 + \delta \log(V_{im}) + \beta_1 \log(g(H_{im})) + \xi_{im}$$

since I cannot measure V_{im} the observed model is the following:

$$\log(e_{im}) = \beta_0 + \beta_1 \log(g(H_{im})) + \tilde{\xi}_{im}$$

Therefore, the expected value of the OLS estimator of β_1 is given by the following formula:

$$\mathbb{E}\{\hat{\beta}_1 | H_{im}\} = \beta_1 + \frac{Cov[\log(g(H_{im})), \log(V_{im})]}{Var[\log(g(H_{im}))]}\delta$$

since theory says that $\delta > 0$ and H_{im} can be defined such that $Cov[\log(g(H_{im})), \log(V_{im})] < 0$, the bias in the estimation of β_1 is negative.

Table A.1: Comparison of matches with missing or non-missing distance covered by teams

	Averages		
Variables	distance covered not missing	distance covered missing	p–value for H_0: equal means
matchday	17.60	11.85	0.036
	(9.78)	(10.08)	
home team Team score	1.67	1.38	0.453
	(1.37)	(1.12)	
away team Team score	1.32	0.92	0.243
	(1.23)	(0.76)	
home team Yellow+Red cards	1.70	2.15	0.176
	(1.20)	(1.46)	
away team Yellow+Red cards	2.14	2.15	0.980
	(1.74)	(1.28)	
home team Fouls (overall: fouls+cards)	16.37	14.92	0.291
	(4.90)	(3.84)	
away team Fouls (overall: fouls+cards)	17.57	16.46	0.430
	(4.96)	(5.71)	
home team offsides	2.92	2.08	0.147
	(2.08)	(1.85)	
away team offsides	2.66	2.69	0.952
	(2.08)	(2.43)	
home team corners	5.47	5.54	0.944
	(2.93)	(2.82)	
away team corners	4.22	4.77	0.419
	(2.44)	(2.74)	
home team attempts on goal	14.66	14.54	0.933
	(5.21)	(5.21)	
away team attempts on goal	11.78	13.15	0.283
	(4.59)	(4.74)	
Matches	905	13	

Notes: the p–value refers to the t–tests of the equality of means between matches in which the distance covered by teams on the pitch is missing and matches in which the distance covered by teams on the pitch is present. Values in parenthesis are standard deviations.

Table A.2: Comparison of matches with missing or non-missing distance covered by teams

Variables	Averages		
	distance covered not missing	distance covered missing	p–value for H_0: equal means
home team tackles won	50.52	53.48	0.006
	(3.84)	(3.95)	
away team tackles won	49.49	46.52	0.006
	(3.84)	(3.95)	
home team passes completion rate	79.45	78.89	0.928
	(22.55)	(5.93)	
away team passes completion rate	78.16	77.21	0.854
	(18.44)	(8.96)	
home team number of touches	625.05	610.54	0.624
	(107.66)	(95.87)	
away team number of touches	599.29	611.31	0.687
	(106.52)	(129.15)	
home team possession	51.04	50.17	0.683
	(7.66)	(8.08)	
away team possession	48.96	49.83	0.683
	(7.67)	(8.08)	
home team number of crosses	11.93	11.38	0.729
	(5.74)	(5.30)	
away team number of crosses	9.54	10.69	0.402
	(4.96)	(2.78)	
home win probability	0.45	0.46	0.856
	(0.17)	(0.13)	
away win probability	0.30	0.28	0.704
	(0.15)	(0.12)	
draw probability	0.25	0.26	0.513
	(0.04)	(0.03)	
Matches	905	13	

Notes: the p–value refers to the t–tests of the equality of means between matches in which the distance covered by teams on the pitch is missing and matches in which the distance covered by teams on the pitch is present. Values in parenthesis are standard deviations.

Bibliography

Abowd, J., F. Kramarz, P. Lengermann, and S. Pérez-Duarte (2004): "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States," *Unpublished Manuscript*.

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Longitudinal employer-household dynamics technical papers, Center for Economic Studies, U.S. Census Bureau.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67(2), 251–333.

Aizer, A., and J. J. Doyle Jr (2013): "Juvenile incarceration, human capital and future crime: Evidence from randomly-assigned judges," Discussion paper, National Bureau of Economic Research.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008): "High wage workers and low wage firms: negative assortative matching or limited mobility bias?," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3), 673–697.

Arellano, M. (2008): "Duration Models," *Class notes. www. cemfi. es/arellano/duration-models. pdf*.

Autor, W. R. Kerr, and A. D. Kugler (2007): "Does employment protection reduce productivity? Evidence from US states," *The Economic Journal*, 117(521), F189–F217.

Autor, D., A. R. Kostøl, and M. Mogstad (2015): "Disability Benefits, Consumption Insurance, and Household Labor Supply," .

Bach, N., O. Gürtler, and J. Prinz (2009): "Incentive Effects in Tournaments with Heterogeneous Competitorsâan Analysis of the Olympic Rowing Regatta in Sydney 2000," *management revue*, pp. 239–253.

Balafoutas, L., F. Lindner, and M. Sutter (2012): "Sabotage in tournaments: evidence from a natural experiment," *Kyklos*, 65(4), 425–441.

Bentolila, S., and G. Bertola (1990): "Firing costs and labour demand: how bad is eurosclerosis?," *The Review of Economic Studies*, 57(3), 381–402.

Bentolila, S., and G. Saint-Paul (1994): "A model of labor demand with linear adjustment costs," *Labour Economics*, 1(3-4), 303–326.

Berger, J., and P. Nieken (2010): "Heterogeneous Contestants and Effort Provision in Tournaments-an Empirical Investigation with Professional Sports Data," .

Bianco, M., T. Jappelli, and M. Pagano (2001): "Courts and Banks: Effects of Judicial Enforcement on Credit Markets," *Journal of Money, Credit and Banking*, 37(2), 223–44.

Bloom, N. (2009): "The impact of uncertainty shocks," *econometrica*, 77(3), 623–685.

Brown, J. (2011): "Quitters never win: The (adverse) incentive effects of competing with superstars," *Journal of Political Economy*, 119(5), 982–1013.

Carmignani, A., and S. Giacomelli (2010): “Too many lawyers? Litigation in Italian civil courts,” *Bank of Italy working paper*, (745).

Chemin, M. (2009): “Do judiciaries matter for development? Evidence from India,” *Journal of Comparative Economics*, 37(2), 230–250.

——— (2012): “Does Court Speed Shape Economic Activity? Evidence from a Court Reform in India,” *Journal of Law, Economics and Organization*, 28(3), 460–485.

Cornelissen, T. (2008): “The Stata command felsdvreg to fit a linear model with two high-dimensional fixed effects,” *Stata Journal*, 8(2), 170.

Coviello, D., A. Ichino, and N. Persico (2015): “THE INEFFICIENCY OF WORKER TIME USE,” *Journal of the European Economic Association*, 13(5), 906–947.

Cox, D. (1972): “Regression Models and Life-Tables,” *Journal of the Royal Statistical Society*, 34(2), 187–220.

Dahl, G. B., A. R. Kostol, and M. Mogstad (2013): “Family welfare cultures,” Discussion paper, National Bureau of Economic Research.

Davis, S. J., R. J. Faberman, and J. Haltiwanger (2006): “The flow approach to labor markets: new data sources and micro–macro links,” *The Journal of Economic Perspectives*, 20(3), 3–26.

DeGroot, M. H. (2005): *Optimal statistical decisions*, vol. 82. John Wiley & Sons.

Deutscher, C., B. Frick, O. Gürtler, and J. Prinz (2013): “Sabotage in tournaments with heterogeneous contestants: empirical evidence from the soccer pitch," *The Scandinavian Journal of Economics*, 115(4), 1138–1157.

Ehrenberg, R. G., and M. L. Bognanno (1990a): "Do Tournaments Have Incentive Effects?," *The Journal of Political Economy*, 98(6), 1307–1324.

——— (1990b): "The incentive effects of tournaments revisited: Evidence from the European PGA tour," *Industrial & Labor Relations Review*, 43(3), 74S–88S.

Fabbri, D. (2010): "Law Enforcement and Firm Financing: Theory and Evidence," *Journal of the European Economic Association*, 8(4), 776–816.

Fraisse, H., F. Kramarz, and C. Prost (2015): "Labor Disputes and Job Flows," *ILR Review*, p. 0019793915591989.

Giacomelli, S., and C. Menon (2016): "Does weak contract enforcement affect firm size? Evidence from the neighbourâs court," *Journal of Economic Geography*, p. lbw030.

Gianfreda, G., and G. Vallanti (2015): "Institutions and firms' adjustments: measuring the impact of courts' delays on job flows and productivity," *Mimeo*.

Gruetter, M., and R. Lalive (2009): "The importance of firms in wage determination," *Labour Economics*, 16(2), 149–160.

Haltiwanger, J., S. Scarpetta, and H. Schweiger (2008): "Assessing job flows across countries: the role of industry, firm size and regulations," Discussion paper, National Bureau of Economic Research.

——— (2014): "Cross country differences in job reallocation: the role of industry, firm size and regulations," *Labour Economics*, 26, 11–25.

HECKMAN, J. J., AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions: an Overview," *The Journal of Econometrics*, 30, 239–267.

HOPENHAYN, H., AND R. ROGERSON (1993): "Job turnover and policy evaluation: A general equilibrium analysis," *Journal of political Economy*, pp. 915–938.

KONRAD, K. A., ET AL. (2009): "Strategy and dynamics in contests," *OUP Catalogue*.

KRAMARZ, F., S. MACHIN, AND A. OUAZAD (2010): "Using compulsory mobility to identify heterogeneous school quality and peer effects," *Unpublished working paper*.

KUGLER, A., AND G. PICA (2008): "Effects of employment protection on worker and job flows: Evidence from the 1990 Italian reform," *Labour Economics*, 15(1), 78–95.

KUMAR, K. B., R. G. RAJAN, AND L. ZINGALES (1999): "What Determines Firm Size?," (7208).

LALLEMAND, T., R. PLASMAN, AND F. RYCX (2008): "Women and competition in elimination tournaments evidence from professional tennis data," *Journal of sports economics*, 9(1), 3–19.

LAZEAR, E. P. (1990): "Job security provisions and employment," *The Quarterly Journal of Economics*, pp. 699–726.

LAZEAR, E. P., AND S. ROSEN (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy*, 89(5), 841–864.

LJUNGQVIST, L. (2002): "How Do Lay-off Costs Affect Employment?," *The Economic Journal*, 112(482), 829–853.

Manudeep, B., G. B. Dahl, K. V. Løken, M. Mogstad, et al. (2016): "Incarceration, recidivism and employment," Discussion paper.

Marchesi, D. (2003): *Litiganti, avvocati e magistrati: diritto ed economia del processo civile*. Il mulino.

Nieken, P., and M. Stegh (2010): "Incentive Effects in Asymmetric Tournaments Empirical Evidence from the German Hockey League," .

Nunn (2007): "Relationship-Speci?city, Incomplete Contracts, and the Pattern of Trade," *The Quarterly Journal of Economics*, 122(2), 569–600.

Ouazad, A. (2008): "A2REG: Stata module to estimate models with two fixed effects," *Statistical Software Components*.

Palmer, C. (2013): "Why did So Many Subprime Borrowers Default During the Crisis," *Manuscript*.

Sunde, U. (2009): "Heterogeneity and performance in tournaments: a test for incentive effects using professional tennis data," *Applied Economics*, 41(25), 3199–3208.

Tullock, G. (2001): "Efficient rent seeking," in *Efficient Rent-Seeking*, pp. 3–16. Springer.