Title
Essays in Applied Microeconomics

Author
Hoffman, Mitchell H.

Publication Date
2012

Peer reviewed|Thesis/dissertation

ESSAYS IN APPLIED MICROECONOMICS

by

MITCHELL H. HOFFMAN

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor John Morgan, Chair
Professor Stephen Burks
Professor David Card
Professor Stefano DellaVigna
Professor Steven Tadelis

Spring 2012

ESSAYS IN APPLIED MICROECONOMICS

Abstract

ESSAYS IN APPLIED MICROECONOMICS

by

MITCHELL H. HOFFMAN
Doctor of Philosophy in Economics

University of California, Berkeley

Professor John Morgan, Chair

This dissertation consists of three essays. All are in personnel economics, using data from the trucking industry. Training by firms is a central means by which workers accumulate human capital, yet firms may be reluctant to train if workers can quit and use their gained skills elsewhere. "Training contracts" that impose a penalty for premature quitting can help alleviate this inefficiency. The first essay from this dissertation studies training contracts in the U.S. trucking industry where they are widely used, focusing on data from one leading firm. Exploiting two plausibly exogenous contract changes that introduced penalties for quitting, I confirm that training contracts significantly reduce quitting. To analyze the optimal design of training contracts and their welfare consequences, I develop and estimate a structural learning model with heterogeneous beliefs that accounts for many key features of the data. The estimation combines weekly productivity data with weekly subjective productivity forecasts for each worker and reveals a pattern of persistent overconfidence whereby many workers believe they will achieve higher productivity than they actually attain. If workers are overconfident about their productivity at the firm relative to their outside option, they will be less likely to quit and more likely to sign training contracts. Counterfactual analysis shows that workers' estimated overconfidence increases firm profits by over \$7,000 per truck, but reduces worker welfare by 1.5%. Banning training contracts decreases profits by \$4,600 per truck and decreases retention by 25%, but increases worker welfare by 4%. Despite the positive effect of training contracts on profits, training may not be profitable unless some workers are overconfident.

A robust finding in experimental psychology and economics is that people tend to be overconfident about their ability. However, much less is known about whether overconfidence can be reduced or eliminated, particularly in field settings. The second essay of this dissertation provides new evidence using data from the workplace. A field experiment with a large trucking firm shows that workers tend to systematically overpredict their productivity and that their overconfidence is unaffected by whether workers receive financial incentives of different sizes for accurate guessing. Randomly informing workers about other workers' overconfidence reduces overconfidence in the short-run, but the effect fades within two weeks. Neither the incentives or information treatments have any effect on worker satisfaction or search behavior. Using long-term survey data from a second firm, I show that experience re-

duces overconfidence, but only quite slowly. Although workers at both firms exhibit aspects of Bayesian updating, overconfidence appears to be sticky and difficult to change.

The third essay analyzes worker referrals. Many firms use referrals in their recruitment and hiring procedures. Are these practices profitable, and if so, why? A model is developed where referrals may improve selection and reduce moral hazard. The model is tested using extremely detailed personnel and survey data from a leading firm in the trucking industry. Referred workers are similar to non-referred workers across a large number of background characteristics and lab experimentally-measured dimensions of preferences. Referred workers are between 10-25% less likely to quit; the effects are strong across all groups of drivers, including new workers for whom the firm invests in expensive firm-sponsored general training. However, referred workers attain similar initial productivity and productivity growth as non-referred workers, and are no more likely to engage in various forms of moral hazard. The accumulation of friends *after* the starting work does not positively affect retention, productivity, or moral hazard. On net, the evidence is consistent with the idea that referrals benefit firms by selecting workers with a better fit for the job, as opposed to selecting workers with higher overall quality, by affecting worker behavior, or by changing job amenities.

To my parents.

Acknowledgments

I am extremely grateful to my dissertation chair, John Morgan. I start talking with John during my first weeks in graduate school. Our first conversation was about my undergraduate senior essay, as this was the only economics topic I could talk about. I have kept talking with him ever since. John has been a total believer in me every step of the way. His way of thinking about economics and attacking economic problems has profoundly influenced me. For his wisdom, time, and unconditional support, I am highly indebted.

I am deeply indebted to my committee members, Stephen Burks, David Card, Stefano DellaVigna, and Steven Tadelis.

From Stephen Burks, I learned the importance of a careful understanding of the institutions of organizations and industries. For our endless phone calls, emails, and visits, thank you Steve! His generosity and intellect have had a huge impact on me.

David Card taught me an enormous amount about doing applied economic research. From selecting research topics to analyzing data to structural estimation to organizing papers, I have learned a great deal from David over the last five years. Every time I go to David's office, I learn at least five new things.

Stefano DellaVigna taught me a great deal and introduced me to the world of empirical behavioral economics. Early on in graduate school, we would discuss my research ideas, nearly all of which were very poor, but I would learn why they were not interesting or not feasible. Stefano's dedication to doing top-quality research has been an inspiration for me.

Steven Tadelis instilled in me an excitement in studying contracts and organizations. He is a meticulous scholar and his class on organizational economics was my favorite class at Berkeley. My research has benefited enormously from his extremely detailed comments and suggestions, and I have benefited personally from his unyielding support.

Many other Berkeley faculty members generously shared their time and wisdom with me, particularly Fred Finan, Ben Handel, Ben Hermalin, Shachar Kariv, Pat Kline, Botond Koszegi, Jonathan Leonard, Ulrike Malmendier, Ted Miguel, Don Moore, Enrico Moretti, Denis Nekipelov, Matthew Rabin, Jesse Rothstein, Kenneth Train, and Felix Vardy.

Graduate school would have been difficult if not for my outstanding graduate student classmates. Mariana Carrera, Willa Friedman, Justin Gallagher, Hedvig Horvath, Jonas Hjort, Mauricio Larrain, Omar Nayeem, Valentina Paredes, Alex Poirer, Issi Romem, Alex Rothenberg, Michel Serafinelli, Edson Severini, Juan Carlos Suarez Serrato, James Zuberi and many others taught me economics and brightened my day.

Finally, I wish to thank my family. My parents have been with me every step of the way, always ready to offer their unconditional support. Thank you to my sisters for always listening to me and always offering such good advice. For all of their love, I am very grateful.

Contents

List of Figures

List of Tables

Chapter 1

Training Contracts, Worker Overconfidence, and the Provision of Firm-Sponsored General Training

1.1 Introduction

Training by firms is a central means by which workers accumulate human capital. However, since at least Pigou (1912), economists have recognized that the provision of general training is subject to a "hold-up" problem. If workers cannot credibly commit to stay with firms after receiving training, firms will under-invest in training. The canonical solution developed by Becker (1964) is for workers to pay for training themselves, but this may not be feasible, for example, if workers are credit-constrained. Indeed, growing evidence shows that a significant portion of training is paid for by firms.[1] Hold-up may have important implications for the overall level of training in the economy. High worker turnover in the United States may make firms reluctant to train, thereby contributing to lower levels of training than in countries with lower turnover (e.g. Blinder and Krueger, 1996). Understanding what makes training profitable for firms may thus be important for optimal human capital policy.

To discourage workers from quitting after receiving training, firms often use training contracts. In these contracts, the firm pays for training, and in exchange workers must agree to stay with the firm for some period of time. If workers leave early, they must pay penalties. Training contracts of this form are used for many workers, including truckers, policemen, firefighters, nurses, pilots, securities brokers, and federal employees, to name a few, metal-workers, mechanics, securities brokers, firm-sponsored MBAs, accountants, teachers, bank

[1]See e.g. Barron et al. (1999), Acemoglu and Pischke (1999a), and Cappelli (2004) for evidence that firms often pay for training, both nominally and in terms of incidence. U.S. firms appear to spend less on training than firms in other countries (Lynch, 1993; Brunello and Medio, 2001), but training expenditures are still substantial. For example, on tuition reimbursement alone, it is estimated that U.S. firms spent $10 billion in 2003 (Manchester, 2009). There are many reasons besides credit constraints that firms will pay for training, including labor market frictions, information asymmetries, and screening benefits (Acemoglu and Pischke, 1999a). Further discussion of the firm-sponsored training literature is given in Section 1.2.

workers, repairmen, social workers, and federal employees, but have received limited attention from economists.[2] [3] How do training contracts affect worker turnover, worker selection, and firm training? How do training contracts affect profits and welfare? How should firms design training contracts and how should they be regulated? While training contracts are legally permissible within some guidelines, some have argued that training contracts are exploitative and tantamount to a mild form of indentured servitude, and there have been recent legal challenges.[4] How should training contracts be regulated?

To address these questions, I use rich data from the U.S. trucking industry where I document that training contracts are frequently used. The main data in the paper are from a leading trucking firm, referred to as Firm A, at which there is plausibly exogenous contractual variation. [5] At Firm A, training was initially provided free of charge with no contractual obligations. In the early 2000s, a newly promoted manager suggested the idea of using a training contract, arguing it could improve retention, as well as help recover training costs. A training contract was created that required trained workers to stay 12 months or pay a penalty if they left early. The contract was phased into different training schools at different times, depending on how fast the contract was approved for use in different states. Around five years later, the company unrolled an 18-month contract with a higher initial quit penalty that decreased with tenure. Again, the contract was phased in gradually. I exploit the staggered timing of these two contract changes to estimate the impact of training contracts. The 12-month and 18-month contracts reduced quitting by 18 and 11 percent, respectively, relative to a situation with no training contract. The effects appear to be primarily driven by incentives instead of selection.

In Jovanovic's (1979) seminal theory of turnover, workers gradually learn about their productivity or job match, using their updated beliefs in deciding whether to quit. Thus, a big advantage of the Firm A data is that weekly panel data on worker subjective productivity

[2] See the law articles by Kraus (1993, 2008) for these examples, as well as a review of legal issues surrounding training contracts. Other workers with training contracts include metalworkers, mechanics, salesmen, paramedics, electricians, accountants, teachers, flight attendants, bank workers, repairmen, firm-sponsored MBAs, and social workers. In economics, there is a small related literature on firms providing tuition reimbursement, which I discuss in Section 1.2. In most of these studies, tuition reimbursement is provided as a benefit and not as part of a contract where the worker is obligated to stay for a length of time.

[3] For a review of legal issues surrounding training contracts, see Kraus (2008) and Kraus (1993). In empirical economics, there is a small related literature on firms providing tuition reimbursement, e.g. Manchester (2009) and Cappelli (2004). In most of these studies, tuition reimbursement is provided as a benefit, and not as part of a contract where the worker is obligated to stay for a length of time. For a recent theoretical paper which includes analysis of training contracts, see Peterson (2010).

[4] Arguments that training contracts are exploitative have been made, for example, in the context of police officers. As of 2006, the City of Los Angeles used a training contract requiring new officers to stay five years after receiving training. McGreevy (2006) quotes a non-L.A. police official arguing that the contracts constitute indentured servitude. The City of Oakland also requires police officers to stay five years after receiving training. In November 2010, an Oakland police officer sued to challenge her contract, with the case decided by the 9th Circuit of the U.S. Court of Appeals. See *Gordon v. Oakland*. In *Heder v. City of Two Rivers*, a firefighter argued his training contract constituted "involuntary servitude."

[5] Firm A's data is used for most of the reduced form and structural analysis. Firm B's data is used for the field experiment in Section 2.4.

forecasts are available for a large subset of drivers. Drivers are paid almost exclusively per mile driven (a piece rate), so beliefs about miles are highly consequential for how much drivers think they will earn. I analyze the belief data so as to better understand worker turnover in the presence of training contracts. Workers' beliefs about future productivity significantly predict quitting and future productivity. In addition, the data show that workers are substantially overconfident about their productivity, though there is significant heterogeneity. On average, workers' initial productivity beliefs exceed their productivity by roughly 25%. Overconfidence decreases over time, but persists throughout the two year panel.

Overconfidence raises important considerations for the efficacy and welfare consequences of training contracts. If workers are overconfident about their future earnings at the current job relative to the outside option, they will be more likely to sign up for training contracts and more likely to stay after training. This makes training more profitable for firms. Overconfidence may also be important for understanding whether training contracts are exploitative. In terms of welfare, while training contracts are legally permissible within some guidelines, some have argued that training contracts are exploitative and amount to a mild form of indentured servitude.[6] Arguments that workers should be legally restricted from entering into training contracts seem to fly in the face of standard economic reasoning; providing an additional option for financing training seems unlikely to hurt workers. However, if workers are systematically biased in their assessment of their ability at a job, the situation is less clear. Workers may overestimate how successful they will be at the job and end up owing penalties for training they would not have undertaken had they been rational. Specifically, workers may overestimate how successful they will be at the job and end up owing penalties for training they would not have undertaken had they not been overconfident.[7]

To better understand observed behavior and to quantify these considerations, I develop a dynamic model of turnover and belief formation. After training, workers solve an optimal stopping rule dynamic programming problem of when if ever to quit the firm. In many empirical models of turnover, workers are assumed to know their future productivity at the firm. However, in my model, productivity is initially unknown and is instead gradually learned about over time as in Jovanovic (1979). Using weekly productivity realizations, workers form expectations of their future productivity and earnings, and use this to decide whether to quit. Although workers update their beliefs in response to new information, the model does not impose that worker beliefs be fully rational, thereby nesting (a simplified version of) the Jovanovic model as a special case. The estimated structural model replicates several

[6] These arguments have been made, for example, in the context of police officers. As of 2006, the City of Los Angeles used a training contract requiring new officers to stay five years after receiving training. McGreevy (2006) quotes a non-LA police official arguing that the contracts constitute indentured servitude. The City of Oakland also requires police officers to stay five years after receiving training. In November 2010, an Oakland police officer sued to challenge her contract, with the case decided by the 9th Circuit of the U.S. Court of Appeals. See *Gordon v. Oakland*. In *Heder v. City of Two Rivers*, a firefighter argued his training contract constituted "involuntary servitude."

[7] The argument that contracts can be exploitative or unfair due to behavioral biases has been made in the legal literature, e.g. Kronman (1983), Eisenberg (1995), and Jolls and Sunstein (2006). Eisenberg (1995) argues that behavioral limitations constitute one of the major rationales for restricting the contracts people should be allowed to sign.

key features of the data including the quit-tenure curve, the productivity-tenure curve, and the belief-tenure curve. Both overconfidence and learning are key. Without overconfidence, the model predicts too much early quitting and fails to rationalize the subjective belief data. Additionally, without learning, the model does not generate the inverted U-shaped quit hazard observed in the data, nor does it predict that observed overconfidence will decrease over time. Estimating the model using workers with the 12-month contract, I show that the model can predict reasonably well out of sample, helping rationalize behavior under the no contract and 18-month contract regimes.

I use the estimates for counterfactual simulations. First, I show that the firm increased profits through its contractual changes, but decreased worker welfare. Second, I consider the counterfactual of reducing worker overconfidence. I show that reducing worker overconfidence would significantly reduce firm profits and worker retention, I show that reducing worker overconfidence would decrease firm profits by $6,000 to over $7,000, would reduce worker retention, and would increase worker welfare. Eliminating the observed amount of overconfidence would moderately increase worker welfare, but would substantially decrease firm profits and worker retention. Profits would decrease by over $7,000 per truck in the baseline case. Third, I analyze a government ban on training contracts. Firms are assumed to maximize profits subjects to workers' participation constraint. Because they believe it is unlikely they will want to quit, overconfident workers are willing to accept a large quit penalty in return for a small wage increase. Banning training contracts has the potential to improve welfare for overconfident workers and I find that a ban increases worker welfare by 4%. I also study optimal training contracts for firms, analyzing how optimal contracts depend on worker overconfidence and learning. I also study optimal training contracts for firms, showing that as worker overconfidence is reduced, the optimal training contract becomes smaller.

The results in my paper are specific to a particular industry, and as such, there are questions of external validity. However, there are several reasons why long-haul trucking is an interesting context for studying the effects of training contracts. While the results in my paper are specific to a particular industry, there are several reasons why long-haul trucking provides an interesting setting for studying the effects of training contracts. First, training contracts are a common mechanism by which general training is provided for truckdrivers, a large occupation employing 3.2 million Americans.[8] Second, and more importantly, trucking provides a natural setting for examining training contracts in the context of Jovanovic's (1979) model of turnover where workers gradually learn about their productivity. Long-haul truckers are typically paid a piece rate proportional to their productivity (the number of miles they drive per week). Unlike in many other industries, productivity in trucking (miles driven per week) is easily measurable, and is accurately recorded by firms given that it is used for determining worker payment. There is considerable variation in productivity across workers, but such differences are unlikely to be known *ex ante*. Third, trucking is an industry with high turnover, allowing for high-frequency retention analysis.

My study makes three contributions to the literature. First, I show that training contracts

[8]See Section 1.3 below for information on the extent of training contracts in trucking.

significantly reduce worker turnover, estimating the effects using plausibly exogenous intra-firm contractual variation. As discussed in the literature reviews by Prendergast (1999) and Chiappori and Salanie (2003), theory has often preceded measurement in economists' study of contracts. Firms' contractual choices are often difficult to observe, and contracts are unlikely to be randomly assigned across or within firms even when they are observable.[9] While Chiappori and Salanie (2003) argue that natural experiments may help researchers circumvent such endogeneity problems in studying contracts, relatively few such studies exist. Second, I provide long-term high-frequency field evidence on overconfidence, the longest I am aware of in the literature, and quantify its welfare impacts for workers.[10] To do so, I develop a structural learning model augmented with heterogeneous and potentially biased beliefs. I contribute to a small, but growing literature incorporating behavioral biases into structural models.[11] Third, I demonstrate that worker overconfidence benefits firms by increasing the profitability of training. Counterfactual simulations suggest that biased beliefs are quantitatively important in facilitating training; even when firms use training contracts, training would not be profitable for firms unless workers are also overconfident. Just as firms may benefit from consumers having time-inconsistent preferences or biased beliefs,[12] so too may firms benefit from their workers having biased beliefs.[13]

Whether it would be possible to reduce worker overconfidence is a separate question from what its impacts would be. I explore the feasibility of reducing overconfidence in a companion paper, Hoffman (2011b). In a field experiment with 254 workers at a different large trucking firm, I find that providing workers with information substantially reduces overconfidence, but that the effect decreases with time.

The paper proceeds as follows. Section 1.2 reviews the literature. Section 1.3 provides background on training contracts and trucking. Section 2.2 describes the data and analyzes the impact of the contractual changes. Section 1.5 analyzes the subjective belief data. Section 1.6 develops the dynamic model. Section 1.7 discusses estimation and identification. Section 1.8 provides structural results. Section 1.9 performs counterfactual simulations. Section 2.6 concludes.

---

[9]For example, it may be imagined that firms with more serious retention problems may be more likely to adopt training contracts. A regression predicting retention may incorrectly show that training contracts have zero or even a negative effect on retention. Alternatively, it may be that only the most successful firms or firms providing the best training think to adopt training contracts, in which case a regression predicting retention will overstate the effect of training contracts. By looking at multiple plausibly exogenous contract changes *within different segments of a single firm over time*, I provide credible estimates of causal effects.

[10]For other longer term evidence of overconfidence, see Massey et al. (2011) who study overconfidence in football fans over the four months of the NFL football season.

[11]See also Laibson et al. (2007), Conlin et al. (2007), Paserman (2008), Fang and Silverman (2009), Acland and Levy (2011), Bellemare and Shearer (2011), Crawford and Meng (2011), DellaVigna et al. (2012), and Grubb and Osborne (2011).

[12]See e.g. DellaVigna and Malmendier (2004, 2006), Eliaz and Spiegler (2006), Grubb (2009), and Ericson (2010).

[13]The point that firms may benefit from worker biases is also made in the recent experiment by Larkin and Leider (2011).

1.2 Related Literature

My paper adds to a growing literature on firm-sponsored training, which is critically surveyed in Acemoglu and Pischke (1999a).[14] The closest part of the literature on firm-sponsored training to my paper is that on tuition reimbursement. Employer-provided tuition reimbursement programs are quite common in the U.S. For firms with more than 20 employees, estimates of the share offering tuition reimbursement have ranged from 47% (Lynch and Black, 1998) to 85% (Cappelli, 2004). In a recent sample of MBA students, Manchester (2011) found that 87% received tuition assistance, with 42% of those obligated to come back to the firm for 12 or more months after completing the MBA. Manchester (2009) shows that workers receiving non-binding tuition reimbursement are more likely to stay with a firm. Other recent papers analyzing tuition assistance include Balmaceda (2005) and Gicheva (2009). For a recent theoretical paper on bonding and turnover, which includes analysis of training contracts, see Peterson (2010).

That people are overconfident has been referred to as "the most robust finding in the psychology of judgment" (De Bondt and Thaler, 1995), and is the subject of a vast literature in psychology and a growing literature in economics.[15] Moore and Healy (2008) provide an excellent review and discussion of the literature, and distinguish between three types of overconfidence: relative overconfidence or "overplacement" (thinking you are better than others), absolute overconfidence or "overestimation" (thinking you are better than you actually are), and excessive precision (thinking your beliefs are more precise than they actually are). This paper focuses on absolute overconfidence, that is, truckers thinking their productivity will be higher than it actually is, and I will refer to this hereafter simply as overconfidence. Overconfidence research has mostly focused on short-term behavior performing laboratory tasks, e.g. completing trivia games. This paper analyzes overconfidence using weekly data over two years on forecasts about individual productivity, an individually-important piece of information. I study how overconfidence affects turnover using reduced-form and structural approaches, and I also analyze structurally how contractual design is shaped by workers' overconfidence.[16]

My paper contributes to a large literature in labor economics on learning. In learning models, agents acquire information about an unknown economic parameter. Learning has been used to analyze wage growth (Harris and Holmstrom, 1982; Schoenberg, 2007; Kahn and Lange, 2011), wage discrimination (Altonji and Pierret, 2001), and occupational choice

[14]A small sample of the recent literature on firm-sponsored training includes Stevens (1994), Acemoglu (1997), Acemoglu and Pischke (1998), Acemoglu and Pischke (1999b), Parent (1999), Barron et al. (1999), Autor (2001), Buddin and Kapur (2005), Belzil et al. (2009), Lazear (2009), and Dustmann and Schoenberg (forthcoming).

[15]Recent papers on overconfidence in economics include Burks et al. (2010), Eil and Rao (2011), Benoit and Dubra (forthcoming), Mobius et al. (2010), Ericson (2011), Grossman and Owens (2010), Charness et al. (2010), Hoffman (2011a), and Hoffman (2011b).

[16]See Oyer and Schaefer (2005) and Bergman and Jenter (2007) for discussion of overconfidence and firms' decisions to issue stock options. See Sandroni and Squintani (2007) for how overconfidence affects insurance markets. See Spinnewjin (2010) for how overconfidence affects optimal unemployment insurance.

(e.g. Gibbons et al., 2005). Several recent papers in labor economics analyze learning using a structural approach, including Arcidiacono (2004), Papageorgiou (2010), Stange (forthcoming), Sanders (2011), James (2011), and Bojilov (2011). Out of these, my paper is most closely related to Bojilov (2011), who analyzes worker learning about match quality using data on call center workers. My largest point of departure from these papers is that I allow for both generalized and non-rational learning. Specifically, I present the first paper on labor markets (to my knowledge) to estimate a learning model with biased beliefs. Structural learning models have also been applied fruitfully in non-labor contexts, including macroeconomics, industrial organization, and political economy. Two papers in industrial organization, Goettler and Clay (2010) and Grubb and Osborne (2011), estimate biased learning models of cellular phone service demand. A main difference in my paper is that belief biases are identified using high-frequency subjective belief data.[17]

Finally, my paper relates to a literature in economics analyzing data on subjective beliefs. Pioneering work by Charles Manski and colleagues argues that economic agents can meaningfully report their subjective beliefs, and that these beliefs can be useful for understanding economic behavior.[18] A small number of papers have used subjective beliefs to estimate structural models. Delavande (2008) estimates a discrete choice model analyzing young women's contraceptive choices. van der Klaauw and Wolpin (2008) use subjective belief data to analyze retirement decisions. Erdem et al. (2005) estimate a structural model of active learning (where agents choose how much information to acquire) about computer purchases using data on price expectations. Wang (2010) estimates a model of smoking with belief bias and subjective expectations data. Pantano and Zheng (2010) show how to use subjective expectations data to more flexibly estimate unobserved heterogeneity in structural models.

1.3 Training Contracts and Training in Trucking

1.3.1 Theoretical Preliminaries

Training contracts attempt to solve the hold-up problem in the provision of general training. Consider a credit-constrained worker employed at a given firm. A socially optimal training investment is available which raises the worker's productivity by more than the cost of

[17]In Goettler and Clay (2010) and Grubb and Osborne (2011), biases are identified through contractual choices. There are advantages and disadvantages to using beliefs and contract choices to identify biases. An advantage of using contracts relative to using subjective beliefs is that economists are more trusting of "what people do" compared to "what people say." A disadvantage of using contracts is that repeated sub-optimal *ex post* choices may reflect factors other than biased beliefs, including inertia or switching costs. Because of the richness of the belief data, I am able to estimate heterogeneity in people's belief biases, e.g. some people are well-calibrated, some are moderately overconfident, some are very overconfident, etc. This heterogeneity is important both for rationalizing the data and also for considering optimal policies, as the welfare consequences of different policies (particularly banning training contracts or debiasing) will differ depending on a person's overconfidence.

[18]For an excellent discussion of the literature on subjective beliefs, see Manski (2004).

training. The training lasts a very short period of time so the cost of training cannot be deducted from worker wages, and training is general (not firm-specific). Becker's solution is for the worker to pay for training himself, but this may be infeasible due to the credit constraint. This situation can potentially be remedied by having the worker take on a training contract. The firm will pay for training, after which the worker must stay with the firm for some period of time. A training contract helps the firm recover training costs when a worker quits and may also reduce quitting.

It is not obvious, however, that a training contract will affect quitting. Suppose that workers and firms have no private information and that bargaining is costless. Then, by the Coase Theorem, turnover will be efficient, that is, it will occur if and only if the sum of the worker's and firm's outside options exceeds the value of the match. Moreover, turnover will be unaffected by a training contract. In the Coasean framework, a training contract merely represents a "property right" held by firms over the quit decision and thus will have no effect on turnover.[19,20]

In my context, however, it seems unlikely that the conditions of the Coase Theorem will hold. Workers likely have private information (about their taste for the job or their outside option) and renegotiating contracts with thousands of workers may be costly for a large firm like Firm A. In the Appendix, I present a model of training and turnover assuming workers have private information and assuming no renegotiation. I show that allowing for training contracts increases the profitability of training and reduces turnover.

One concern about training contracts is that they may harm workers if workers do not have accurate beliefs and/or if workers make bad decisions. If workers are overconfident about what their post-training ability will be relative to their outside option, they may be willing to take on a high-penalty training contract. In addition, after training, overconfidence may distort a worker's quitting decisions, making them stay longer with a firm than is rational. Though overconfidence may harm workers, it may benefit firms who use training contracts. As I show in the Appendix, overconfidence may make training more profitable for firms both by making it easier to get workers to sign training contracts and by making it easier to retain workers after training.[21]

[19]To see why, consider the case where it is socially optimal for the worker to quit, but disadvantageous for the firm. Without a training contract, the worker will quit; the firm will try to "bribe" the worker to stay, but the maximum bribe the firm is willing to offer will still not be high enough to retain the worker. With the same situation and a training contract in place, the worker and firm will bargain such that (after negotiation) the worker will still quit. Whether the worker must bribe the firm to let him quit may be affected by the training contract, but the quitting outcome will not be.

[20]In a related application of the Coase Theorem, Lazear (1990) analyzes job security provisions in Europe, where firms are "fined" (e.g. they must pay severance pay) for firing workers. He shows theoretically how the Coase Theorem may fail to hold and shows empirically that job security provisions do indeed affect firm firing.

[21]In fact, I show in the Appendix that if overconfidence is not permanent, then overconfidence and training contracts are complementary. Overconfidence increases the probability of training when training contracts are allowed, but not if they are not allowed.

1.3.2 Legal Issues Regarding Training Contracts

Courts in the U.S. and abroad have generally ruled that training contracts are legally permissible, arguing they serve the public good by promoting investment in worker training.[22] Most training contracts have the following form. Training is provided at no cost in exchange for the worker agreeing to stay with the company for some period of time. The worker faces penalties if they leave early. Contracts of many lengths have been employed. In trucking, the duration of training contracts is often 6-24 months, whereas for police officers, contracts for five years are sometimes used (Kraus, 2008). Although I use the word "penalty" to describe a training contract, courts have ruled that the amount owed under training contracts for early exit must be reasonable and no larger than the cost of training for firms. However, defining the actual "cost" of training is a difficult matter (e.g. there is the issue of average cost versus marginal cost, as well as the fact that one of the main costs of training is the time spent by employees working with trainees, which is hard to price). Courts have varied in how they have treated training contracts with large amounts owed.[23] Courts have generally held that enforceability does not depend on whether termination penalties decrease with tenure, holding that employees have the ability to bargain over this issue before signing a contract.[24]

1.3.3 Background on the U.S. Trucking Industry

Trucking is one of America's largest occupations, employing 3.2 million workers in 2008 (Bureau of Labor Statistics, 2010). Of these, around 25% work for for-hire motor carrier firms (i.e. "trucking firms"), with the remainder working for non-trucking firms that also employ truckdrivers (e.g. companies like Walmart and Safeway). This paper focuses on workers at trucking firms. The industry is usually divided into two segments, Less than Truckload and Truckload. Less than Truckload drivers deliver small to medium sized loads, usually make local deliveries, usually do not spend nights away from home, and have moderate rates of unionization. In contrast, truckload drivers deliver large loads across long distances, and have much lower rates of unionization.[25] Truckload drivers are usually paid by the mile (a piece rate) (Belzer, 2000).[26] Within the truckload segment, around 10% of miles in 1992 were driven by drivers who own their own truck (owner-operators), with the remainder driven by drivers driving company-owned trucks (company drivers) (Baker and Hubbard, 2004). This paper focuses on company drivers in the Truckload segment. Turnover in the industry is

[22] This section draws primarily on the excellent review articles by Kraus (1993, 2008).

[23] For example, *Heartland Securities Corp. v. Gerstenblatt* dealt with a case where new college graduates were provided computer training by an online brokerage company, in exchange for promising to stay with the company for two years, with a penalty of $200,000 for leaving. The court held this contract to be unenforceable. However, in *Tremco Incorporated v. Kent*, a case where a roofing products sales company sought the recovery of $42,000, the amount owed under a contract if a roofing salesmen trainee did not fulfill three years of service, the court deemed the contract to be enforceable.

[24] See e.g. Judge Richard Easterbrook's opinion in *Heder v. City of Two Rivers*.

[25] In 2008, 16% of all truckdrivers and driver/sales workers were union members or covered by union contracts (Bureau of Labor Statistics, 2010).

[26] For an analysis of productivity in trucking, see Hubbard (2003).

high at over 100% annually and is highest among new drivers.[27]

The main training for truckdrivers is training to obtain a commercial driver's license (CDL). It usually consists of a combination of classroom lecture, simulator driving, and actual behind-the-wheel truck driving, and provides the skills and knowledge necessary to safely operate a large truck. Most new drivers take an official CDL training course, and in some states it is required by law (Bureau of Labor Statistics, 2010). CDL training can be obtained at truckdriving schools, which are run by trucking firms or run privately, or can be obtained at some community colleges. Training courses usually last 2-4 weeks, and are certified by the Professional Truck Driver Institute (Bureau of Labor Statistics, 2010), which requires courses to contain at least 148 hours of instruction, including at least 44 hours of time spent driving (Professional Truck Driver Institute, 2010).[28] The market price for CDL training at private training schools varies, but is often around several thousand dollars.

Truckdriver hours are legally restricted per the federal Hours-of-Service Regulations. Truckdrivers can work up to 60 hours per week.[29] Despite the hours restrictions, however, there is large variation in average miles across drivers, as well as significant week-to-week variation within drivers. Driver miles are influenced by many factors, including time management, trip planning, and driver speed.[30] Factors such as weather, traffic, and variable loading/unloading time create substantial within-driver mileage variation. Thus, weekly miles, my measure of productivity, reflect both driver performance and factors that drivers do not control.

1.3.4 The Extent of Training Contracts in Trucking

To collect information on the extent of training contracts, I conducted phone interviews with the 20 largest dry-van and 10 largest refrigerated trucking companies in the U.S. I obtained the list of the largest trucking companies from Transport Topics (2009), a leading industry trade journal. I collected panel data on the type of training provided and on the training contracts used by each firm from 2001-2010. Interviews were conducted with someone familiar with the details of driver training, usually the director of human resources,

[27] In the first quarter of 2008, the turnover rate at large carriers (those with more than $300 million in revenue) was 103% and 80% for smaller Truckload carriers (Suzuki et al., 2009), where the turnover rate is defined as the number of workers who leave a company each year per 100 workers. The rate at large carriers has decreased in the 2008-2009 recession to 65%, but was at 130% in 2005 (Roth, 2009). For more general information on the trucking industry, see e.g. Burks et al. (2011).

[28] A secondary form of training for truckdrivers is on-the-job driving training, where a new driver drives with a veteran driver sitting in the passenger seat. This type of training varies in length and formality between different firms, ranging from a few hours to several weeks. Though this training is also costly for firms, it is much less so than CDL training. I do not focus on this training, since training contracts are written to cover CDL training.

[29] Drivers can alternatively work up to 70 hours over eight days. See http://www.fmcsa.dot.gov/rules-regulations/topics/hos/index.htm.

[30] One important factor is whether drivers arrive to a location on-time. Drivers who arrive late may have to wait for their truck to be unloaded, which can be highly detrimental to weekly miles, given that the 60 hours per week is hours of working time, including both driving and non-driving.

the director of training, or a driver recruiter. Further information on the interviews is given in the Appendix.

Firm-sponsored training and training contracts are widespread in trucking. 16 of the 30 largest trucking companies report operating their own training school at some point from 2001-2010. When firms provided training at a CDL school, it was almost always provided under a training contract: Only two companies (including the company studied in this paper) sometimes provided CDL training without using a training contract. In addition, many companies offer tuition reimbursement programs, where drivers can receive their training elsewhere, and have the amount paid back over time by the company according to a contract. Only six companies never offered either firm-sponsored training or tuition reimbursement at some point from 2001-2010. Many companies report that they engage in firm-sponsored training because it is often difficult to find enough qualified drivers. At least four companies either stopped training new drivers or cut back significantly on training during 2008-2010 in the wake of the Great Recession. Larger firms appear to be more likely to train.[31]

The form of training contracts varies across companies, though there are common elements. At one firm, drivers owe $2,995 if they quit in the first year. At another firm, the training contract lasts 26 months. Workers who quit during the first 13 months are required to pay back $3900 to the firm. After 13 months, the amount owed is reduced by $300 per month for 13 months; half of the monthly $300 deduction is deducted from the worker's paycheck. At another firm, the training contract lasts 12 months. Drivers who quit in the first 6 months are required to pay $3500 to the firm, and drivers who quit in months 7-12 are required to pay $1750.

1.4 Data and Reduced Form Analysis

1.4.1 Contract Changes

To examine the effects of training contracts, I make use of two large contract changes at Firm A, a leading U.S. trucking company. Firm A provides CDL training to thousands of new drivers at several geographically dispersed training schools. Prior to 2001, all training was provided at no cost to the worker with no contractual obligation.

In late 2000, a newly promoted manager proposed the idea of implementing a training contract. According to the Director of Driver Training, management had not been previously aware of the possibility of using a training contract. The primary motivation for implementing a contract was to help increase retention, with a secondary motivation being to help recover costs. In order to implement the contracts, it was necessary in each state to have the contract certified by the state.[32] This certification process took different amounts of time in different states. The contract was approved quickly in several training schools and was in use by April 2002. At another training school, the contract was not approved until

[31] In a regression of CDL training on log 2008 revenue, the coefficient on log 2008 revenue is 0.16 ($p = 0.06$).

[32] It was explained to me that the training schools are considered private colleges, and training contracts are counted as a form of loan contract.

the end of 2002, and in a couple states the contracts were never used as the certification process dragged on indefinitely.[33] The penalty for leaving varied slightly by training school and was between \$3,500 and \$4,000. The contract lasted 12 months and the quit penalty was constant throughout the 12 months. The contract applied for both quits and fires.[34]

After several years of the 12-month contract, management began to discuss increasing the duration of the contract as well as changing its form. According to a Senior Vice President, the interest in changing the contract stemmed from a desire to retain new drivers for longer. Management decided to switch to an 18-month contract. The initial penalty for leaving would be higher, but would decrease gradually with a worker's tenure. Again, the contract was phased in gradually. Under the new 18-month contract, the amount owed was initially around \$5,000, and was reduced by \$62.50 for each week of service. Of the \$62.50 per week, \$50 was paid by the firm, and \$12.50 was deducted from the worker's pay check. After two years with the company, the driver would be returned his \$12.50 payments over 78 weeks in the form of a single bonus payment. Both the adoption of the 12-month and 18-month contracts were made without additional changes in the drivers' wage.[35]

New drivers signed a written contract specifying the amount they would owe if they left within some period of time. No initial bond was posted. Upon early exit, drivers were contacted by the company to pay the amount due, either immediately or in monthly installments. If drivers did not pay promptly, they were referred to one of multiple collection agencies. Though comprehensive data on collection is not available, company records from late 2004 reveal that Firm A was collecting roughly 20 percent of the amount owed before referring accounts to collections. The collection agencies then collected an additional 5 to 10 percent.[36]

1.4.2 Data

The data from Firm A are highly advantageous for an analysis of the impact of training contracts due to the large sample size and high frequency of observation. The data contain weekly miles and earnings for thousands of drivers from mid 2001 to 2009. I focus exclusively on new inexperienced drivers who are trained by Firm A. Drivers are paid by the mile, with small payments for other tasks.[37] For each driver, there is also basic demographic

[33] The Director of Driver Training believed that the differences in time for state approval were idiosyncracies of the state bureaucracy, and not related to the type of impact the contracts might have.

[34] According to several managers at Firm A, the reason why the contract also covered fires was to prevent workers who wanted to quit from trying to get fired. According to these managers, the firm did not intentionally fire workers to collect training penalties.

[35] Firm A pays slightly different mileage rates depending on driver regions. It increased its overall pay schedule twice during the period for which I have data, once, in early 2004, and a second time, in late 2007.

[36] A 25-30% collection rate may seem on face to be fairly low. However, collection efforts by Firm A appeared to be fairly strenuous. Drivers who did not pay received numerous strongly worded letters and phone calls from the company. For drivers who remained delinquent, credit agencies were notified.

[37] Drivers also receive small additional payments for non-miles related tasks such as going through customs, loading and unloading, scales weighing, working on trailers, and training other drivers. After one year, drivers are also eligible to receive a quarterly bonus. A small number of drivers are paid based on their activities or

information.[38]

Very detailed data are available for a subset of 895 new drivers. These drivers were trained at one of the firm's training schools in late 2005 and 2006. I refer to drivers in this group as the data subset, and I focus much of the structural analysis on this group. Data on these drivers were collected in Burks et al. (2009). Subjective belief data about next week's productivity are available only for drivers in the data subset.[39]

Table 2.1 provides summary statistics for the data. The top panel shows summary statistics under the different contractual regimes: No contract, 12-month contract, and the 18-month gradual contract. Many driver characteristics before and after the contract change look fairly similar. There are some statistically significant differences in characteristics across the regimes, but some of these reflect inter-school demographic differences (since contracts were in place in different schools for different lengths of time). The lower panel provides summary statistics for the data subset. Drivers' average schooling is slightly over 12 years. The median driver is male, white, and 35 years old.[40] The drivers have very low average credit scores. Of the 88% of drivers with credit scores (12% of drivers do not have a sufficient credit history to have a credit score), the average credit score is 586 and the median credit score is 564. According to CreditScoring.com, the median credit score for the U.S. general population is 723. While only 7% of the U.S. population have a credit score below 550, 43% of drivers have a credit score below 550.

1.4.3 Do Training Contracts Affect Quitting?

To analyze the effect of training contracts on quitting, I first plot survival curves under the different contractual regimes in Figure 1.1. Survival is significantly higher for both the 12-month and 18-month contracts relative to the no contract regime. Figure 1.2 compares the quitting hazards for the three contractual regimes. For drivers with the 12-month contract, the hazard decreases until the 52-week mark, where there is a large spike. There are also smaller bumps at the 52-week mark under the no contract and 18-month regimes.[41]

To more closely quantify these effects, I estimate Cox proportional hazard models of quitting. Every week, a driver faces some risk of quitting, which is influenced by his charac-

on salary instead of by the mile (e.g. drivers who work full time as driver trainers at the training schools).

[38] I restrict my sample to drivers from 5 training schools, excluding 3 schools where the training provided differed and/or the precise contract change dates were not available.

[39] Further information about the firm and the data is given in Burks et al. (2009), who show a strong relationship between cognitive skills and driver retention in the data subset. The number of 895 represents the drivers who successfully graduated from the training academy. From these 895 drivers, I drop any drivers who are ever seen working at non-piece rate trucking jobs where they are paid based on their activities or on salary (e.g. this drops drivers who ever go to work themselves as driver trainers at the training schools). This leaves a sample of 735 drivers. The structural analysis focuses on a sample of 699 drivers for whom one of several covariates is not missing.

[40] Truckdrivers must be at least 21 years old to cross state lines in a truck (Bureau of Labor Statistics, 2010) and Firm A requires new drivers to be over 21.

[41] Managers at Firm A suggested that the bumps under the no contract and 18-month regimes may result from workers being able to say that they worked for a full year at Firm A when applying for other jobs.

teristics, including whether or not he has a training contract:

$$\begin{aligned} log(h_{it\tau cs}) &= \alpha_t + \beta_1 * 12MCONTRACT_{sc} + \beta_2 * 18MCONTRACT_{sc} \\ &+ \beta_3 * UNEMP_{s\tau} + \beta_4 \bar{y}_{it} + \gamma_\tau + \delta_c + \theta_s + X_i\lambda + \epsilon_{it\tau cs} \end{aligned} \tag{1.1}$$

where $h_{it\tau cs}$ is the quit hazard of driver i with t weeks of tenure in year τ who is part of cohort (year of hire) c who attended training school s; $UNEMP_{s\tau}$ is the unemployment rate in state s at time τ; $\bar{y}_{it}$ is average productivity to date; α_t is a fixed effect for tenure t; γ_τ is a time fixed effect; δ_c is a fixed effect for year of hire c; θ_s is a school fixed effect; X_i are individual covariates; and $\epsilon_{it\tau cs}$ is an error. The coefficients of greatest interest are β_1 and β_2, representing the effects of the two contracts on the quit hazard. By including school fixed effects, I help rule out the possibility that the effects are simply due to the contracts being used for long periods of time at schools with high retention. By including time fixed effects, I help rule out the issue that factors other than the contracts changed over time. These results are shown in Table 1.2. As seen in column 1, the coefficients on the 12-month and 18-month contract variables are -0.18 and -0.11, respectively. That is, the 12-month and 18-month contract decrease the probability of quitting by roughly 18 and 11 percent. [42] A percentage point increase in state unemployment is estimated to reduce quitting by 5.6%. Thus, the estimated impact of the 12-month (18-month) training contracts is similar to a 3 (2) percentage point increase in the unemployment rate. When average productivity and demographics are controlled for, the estimates of β_1 and β_2 remain similar or become larger. The interaction effects in Table 1.2 reveal at what tenure levels the training contracts are reducing quitting. The effect of the 12-month contract is felt during the first 12 months (as expected), whereas the effect of the 18-month contract is experienced most in weeks 53-78.

To examine further how the effects of the contracts varied with tenure, I ran OLS regressions of quitting on interactions of the contracts with tenure blocks:

$$\begin{aligned} q_{it\tau cs} = &\sum_{r=1}^{8} \beta_{1r} * 12MCONTRACT_{sc} * TENUREQUARTER_r \\ &+ \sum_{r=1}^{8} \beta_{2r} * 18MCONTRACT_{sc} * TENUREQUARTER_r \\ &+ \beta_3 * UNEMP_{s\tau} + \beta_4 \bar{y}_{it} + \eta_t + \gamma_\tau + \delta_c + \theta_s + X_i\lambda + \epsilon_{it\tau cs} \end{aligned} \tag{1.2}$$

where $TENUREQUARTER_r$ is a dummy for a driver being in the rth quarter of tenure (e.g. $r = 1$ means the driver has $1 - 13$ weeks of tenure, $r = 2$ means the driver has $14 - 26$ weeks of tenure, etc.). The estimates are shown in Figure 1.3. Under the 12-month contract,

[42]Throughout the reduced form analysis on the full Firm A dataset, standard errors are clustered at the training school class (school-week of hire) level. Doing so allows for arbitrary correlation of the error within workers of the same training school class.

quitting is significantly lower (relative to no contract) in the 4th quarter (weeks 39-52), but is significantly higher (relative to no contract) in the 5th quarter (weeks 53-65). This postponement of quitting behavior likely reflects the sharp decline in the quitting penalty at one year. Under the 18-month contract, quitting is significantly lower (relative to no contract) in quarters 4-6, but then increases after that.

Event Study Analysis

Another approach to identification is to analyze new driver quit patterns before and after the contract changes using an event study methodology. For the transition from no contract to the 12-month contract, I analyze quitting in weeks 46-52. Under the 12-month contract, drivers may optimally wait until their year is up before quitting, whereas the same incentive is not present for drivers with no contract. For the transition from the 12-month to the 18-month contract, I analyze quitting in weeks 53-78. Those under the 12-month contract may have waited for the year to expire, whereas these weeks are still under contract for the 18-month contract. For the transition from the 12-month to the 18-month contract, the event study can be represented with the following regression equation:

$$Quit5378_{ics} = \alpha_s + \beta_c + \sum_{j=\underline{T}}^{\overline{T}} \theta_j D_{sc}^j + X_i\lambda + \epsilon_{ics} \tag{1.3}$$

where $Quit5378_{ics}$ is a dummy for whether worker i quits in weeks 53-78 (conditional on having stayed through week 52), α_s and β_c are school and cohort fixed effects, and ϵ_{ics} is an error. D_{sc}^j is a dummy for whether those in quarter of hire c at training school s are j periods from the introduction of the 18-month contract; formally, $D_{sc}^j = \mathbf{1}(c - e_s = j)$, where e_s is the quarter when school s adopted the 18-month contract. To avoid collinearity, I normalize $\theta_{-1} = 0$. For the switch from no contract to the 12-month (18-month) contracts, I assume that $\underline{T} = -3$ and $\overline{T} = 9$ ($\underline{T} = -7$ and $\overline{T} = 4$). Further, I "bin up" the end points by including dummies for the event time being less than $\underline{T}$ or greater than $\overline{T}$.[43] Results are shown in Figure 1.4. For the transition to the 12-month contract, the probability of quitting during weeks 46 to 52 drops by roughly 5 percentage points. Likewise, for the transition to the 18-month contract, there is a large decrease in the probability of quitting in months 53-78 occurring at the time of the contract change. Specifically, the probability of quitting decreases by roughly 20 percentage points. The average probability of quitting in the months 13-18, conditional on staying for one year, is about 53%. Thus, this effect represents a sizeable decrease in quitting in response to the contract.

Threats to Identification

Worker Sorting into Schools. One potential confound to identifying the impact of training contracts on quitting would be worker sorting into schools. For example, a worker who believed he had a high chance of quitting might prefer to attend a training school that did

[43] Note that I restrict the estimation sample in the event study to drivers who eventually exit the company.

not have a training contract. This is unlikely to be an issue at Firm A because drivers almost always attend the training facility closest to their home. Specifically, 83% of drivers live in a state where at least 90% of Firm A drivers attended the same training school.[44]

Endogenous Contract Changes. The estimation above assumes that the implementation of training contracts is orthogonal to unobserved factors affecting quitting. However, if training contracts were implemented in areas expected to have higher quitting, then I will underestimate the effect of training contracts on quitting. Alternatively, if it was easier to implement the training contracts where future quitting was expected to be lower, then I will overestimate the effect. Managers at Firm A thought this was unlikely to be a concern, believing the timing of training contract approval to be idiosyncratic, likely reflecting differences in the speed of state bureaucracy. In the data, implementation of the training contracts is not predicted by state unemployment rates (which may be correlated with unobserved factors affect quitting), as I show in the Web Appendix.

Tenure-Varying Contract Enforcement. Under the 12-month contract, a worker who quit after 51 weeks was technically responsible for the same penalty as a worker quitting just a few weeks after training. Could it have been that training contracts were enforced differently depending on worker tenure? Such a possibility does not threaten identification of the overall impact of the training contracts on quitting, but may affect the interpretation of impacts by tenure. Although disaggregated data on contract enforcement are not available, managers at Firm A said training contracts were enforced generally irrespective of worker tenure.[45]

1.4.4 Incentives or Selection?

A decrease in quitting from training contract penalties may result through incentive and/or selection effects. If a worker is penalized for quitting, he may become less likely to quit, that is, training contracts may have an incentive effect. However, adding a training contract for quitting may also affect the selection of workers who choose to work at the firm. If workers are fined for quitting early, low productivity workers or workers with a low taste for trucking may be less likely to sign up. Thus, training contracts may have a selection effect on worker quitting, drawing in higher ability workers or workers with a higher taste for the work.

One informal test for selection is to examine whether training contracts affect the rate of firing. If training contracts deterred low-quality drivers from working at Firm A, one would expect a decrease in the rate of firing. In the Web Appendix, I analyze the effect of the contracts on firing rates. None of the coefficients are statistically different from zero. There is thus no evidence that the contracts affected the rate of firing, though the estimates are less precise than those on quitting.

[44] Of the drivers that live in a state where less than 90% of the drivers attended the same training school, more than half live in one state approximately equidistant from two training schools. Excluding drivers from this state, 94% of drivers live in a state where at least 90% of the drivers attended the same training school.

[45] A training manager at Firm A raised one exception. Contracts were sometimes not enforced if workers were fired after very short tenures. However, other than that, I was told that contract enforcement did not depend on a driver's tenure.

A second test is to examine whether selection occurred on various observable characteristics. The most obvious characteristic to examine is productivity: Did adding a training contract lead more productive workers to begin working at the firm? As seen in Table 1.3, there is limited evidence that training contracts led to more productive workers selecting to work at the firm. One can also test for selection by looking at whether workers with other characteristics (potentially correlated with a worker's taste for the job or tendency to quit) are more likely to choose to work for the firm once training contracts are in place. There is evidence that the contracts may have affected selection on several characteristics (whether the driver is Hispanic, whether the driver smokes, and whether the driver applied online), but this evidence is not conclusive.

My third test of selection aims at testing whether there was selection on unobserved taste for trucking. Suppose that there are two types of drivers: "Good drivers," who are productive and who have a high taste for trucking, and "Bad drivers," who are less productive and have a low taste for trucking. The training contract would induce positive selection if it caused a greater share of new workers at the firm to be "Good drivers." If the contracts caused positive selection, controlling for productivity should reduce the estimated magnitude of the coefficients on the contract variables in quit hazard models. However, as can be seen in column 3 of Table 1.2, the contract dummy coefficients are roughly the same or become even larger in magnitude, once productivity is controlled for. Thus, the above test provides no evidence to support the idea that the contract induced positive selection.[46]

These tests provide support for some selection due to the training contracts, but the effects seem somewhat limited. Overall, the evidence suggests that the effect of the training contracts on quitting operated primarily through incentives. Given the strong evidence of selection effects of contracts in other personnel settings (e.g. Lazear, 2000), why does positive selection here appear to be small? One possibility is that workers lack private information about their productivity when signing up for the job. Long-haul trucking is very different from most other jobs and it may be very difficult to predict how good one will be at it.[47] Another possibility is that for some reason the contract may not have been salient to drivers when they signed up for the job.[48] A third possibility may be that selection is multidimensional.[49] Ultimately, however, I am unable to distinguish between these different

[46] This test is inspired by the test for selection in Lazear (2000), who tests for selection by analyzing whether the coefficient on the piece-rate dummy changes once individual fixed effects are added. He finds that the coefficient on the contract dummy falls by half, leading him to conclude that selection explains half the treatment effect of the contract. My test is significantly more indirect, given that I cannot observe the same individual under multiple contractual regimes.

[47] In my structural model, I will make the assumption that workers and firms do not have private information about the worker's productivity before he starts work.

[48] This seems unlikely to be the case, as a discussion of training contracts was a mandatory part of interviews at Firm A.

[49] Multidimensional selection has been observed, for example, in health insurance contracts (Finkelstein and McGarry, 2006). In my setting, workers could potentially be selecting both on productivity and their level of overconfidence. If less productive workers are also more likely to be overconfident about their productivity, then low and high productivity workers may have similar productivity beliefs and may not be differentially selected by different contracts.

possibilities.

1.4.5 Worker Learning

In the next section, I model the quitting decision as a product of worker learning. Workers are initially uncertain about how productive they will be as truckdrivers and gain information about their underlying productivity through weekly productivity signals. Workers who learn that they are less productive become more likely to quit. A testable implication of learning about productivity is that quitting should reflect selection on average productivity. Specifically, at every point in time, workers who are less productive should be more likely to quit. Figure 1.5 confirms this by comparing the average earnings per week of drivers who quit that week versus drivers who make it to that week and do not quit. Quitting drivers receive lower average earnings in prior weeks than non-quitting drivers. Selection on productivity is examined with controls in Table 1.2. As can be seen, an increase in past average productivity significantly reduces the hazard of quitting.

1.5 Worker Beliefs

Worker productivity beliefs are key in theories of turnover. I examine whether incorporating subjective beliefs can help better predict productivity and turnover under training contracts. Every week, drivers in the data subset were asked to predict their miles in the following week, being sent the below question over the Qualcomm message system in their truck:

About how many paid miles do you expect to run during your next pay week?

Drivers responded by typing in their answer. I interpret this question as asking drivers their beliefs about their average number of miles next week. Individual driver responses and participation were never shared with the company and this was emphasized to drivers. No incentives were used to incentivize accurate belief responses, though drivers were given $5 each week for completing the survey.[50] The average response rate across all drivers and weeks to the weekly beliefs question is 21%. Of the 699 drivers whom I focus on in the data subsample, 61% respond to at least one survey about mileage beliefs.[51]

In Table 1.4, I examine whether subjective productivity beliefs help predict productivity over and beyond other predictors. Specifically, I consider regressions of the form:

[50] At a different large trucking firm, I elicited miles expectations while randomizing whether or not drivers were given financial incentives for accurate beliefs (Hoffman, 2011b). Incentives ranged from up to $0, $10, or $50 per week for accurate guessing. I find no evidence that overconfidence is reduced by using incentives. Other field studies using incentives for belief elicitation include Grisley and Kellogg (1983) and Nelson and Bessler (1989).

[51] Women and minority drivers are less likely to respond to the survey, whereas workers with higher average productivity and older drivers are more likely to respond. Within a given driver, response is higher in weeks when the driver is more productive. Later on in the structural analysis, I perform a robustness check where I use Inverse Probability Weighting (Wooldridge, 2002) to address non-random response to the question.

$$y_{i,t} = \alpha + \beta b_{i,t-1} + \gamma \bar{y}_{i,t-1} + X_i \delta + \epsilon_{i,t} \tag{1.4}$$

where $y_{i,t}$ is driver i's productivity in his t^{th} week with the company, $b_{i,t-1}$ is his previous week's belief about his productivity in the next week, $\bar{y}_{i,t-1}$ is lagged average productivity to date, X_i includes controls, and $\epsilon_{i,t}$ is an error. Column 1 estimates with only lagged beliefs on the right-hand side; the estimated β is roughly 0.2. Once controls such as average productivity to date are included, the coefficient dips to roughly 0.06, as more productive people tend to have higher beliefs. The predictive power of productivity beliefs holds *within person* as well, that is, after individual fixed effects are included. Overall, worker subjective beliefs have informational content, but the effect is much less than one for one. The relatively low coefficients likely reflect attenuation bias due to measurement error in subjective beliefs.

Table 1.5 reports the impact of productivity beliefs in a proportional hazards model of quitting

$$log(h_{i,t}) = \alpha + \beta b_{i,t-1} + \gamma \bar{y}_{i,t} + X_i \delta + \epsilon_{i,t}, \tag{1.5}$$

where $h_{i,t}$ is the quit hazard. Average earnings to date, $\bar{y}_{i,t}$ is a sufficient statistic for beliefs about productivity in a basic normal learning model. Thus, the above regression asks whether driver quitting decisions reflect the additional information drivers have about their productivity. The results show that they do. A 100 mile increase in subjective miles predicts a 6 percent decrease in the probability a worker quits. The true effects are probably higher, with observed estimates biased downward due to measurement error. The effect is robust to the inclusion of different controls including average productivity to date. The coefficient on beliefs does not change very much across the different specifications. Compared to a standard setup where workers hold the same beliefs given their productivity signals and observed characteristics, it appears here that workers' heterogeneous subjective beliefs are in fact predictive of quitting.

Besides whether subjective beliefs are predictive, another basic question is whether or not they are well-calibrated. Figure 1.6 analyzes productivity and productivity beliefs by tenure. Productivity and productivity beliefs are collapsed by week of tenure and then smoothed using a local polynomial regression. As can be seen, workers on average consistently believe they will be more productive than they actually are. Beliefs exceed actual productivity both in terms of means and medians.[52] Workers are initially overconfident by roughly 500 miles per week, or approximately 25% of their productivity. This percentage declines over time, though it is quite persistent. Even after 100 weeks of signals, workers are still overconfident by around 150 miles per week.[53]

[52] I interpret the question "About how many paid miles do you expect to run during your next pay week?" as asking drivers for their mean expected miles next week. Another possible interpretation is that it is asking drivers for their median expected miles next week. In the data subset, mean and median miles are almost identical (the median of worker miles per week is 1% less than the mean miles per week). Thus, whether workers reported their mean or median expected miles seems unlikely to matter for the structural estimation.

[53] An important issue not addressed in Figure 1.6 is differential attrition, as low-productivity workers are more likely to quit. In the Web Appendix, I reproduce the figure restricting the sample to workers who

The results mask substantial heterogeneity and variability across and within drivers. Though on average workers are overconfident in almost every week, weeks where drive predictions are higher than next week's actual miles constitute only 65% of the data, whereas weeks where driver prediction are lower than next week's actual miles constitute 35% of the data. Thus, it is not the case that each individual driver is overpredicting his miles every week. Drivers also differ substantially in the share of the time they overpredict their miles, and in the level of their over- or under-prediction. Evidence on heterogeneity is plotted in Figure 1.7, which displays driver-level average mile beliefs, actual miles, and the difference between the two. Many drivers are moderately overconfident, but there are also many drivers who are well-calibrated, as well as some drivers who are very overconfident.

For overconfidence to reduce worker quitting in theory, the worker must be more overconfident about his current job earnings than his outside option. While this assumption is very difficult to test, I provide some suggestive evidence that the assumption is satisfied. Drivers in the data subset were asked what their earnings would have been had they not started work with Firm A.[54] I compare drivers' response to this question to what "similar-looking" people earned in the March 2006 Current Population Survey (CPS). Figure 1.8 shows the comparison. As can be seen, workers do not believe that they appear to earn significantly more than people like them in the CPS. That is, the perceived outside option workers would had earned had they not gone through training does not appear to be significantly higher than what people like them might be earning.

1.6 The Model

I develop a dynamic model of quitting and belief formation where workers learn about their productivity over time. The model is a discrete time extension of the model in Jovanovic (1979) allowing for biased beliefs. The goal of the model is to provide a framework for understanding the facts presented above and to serve as a basis for subsequent counterfactual analysis.

A worker decides each week whether to quit his job. The problem is an optimal stopping problem; once he quits, he cannot return. Quitting is the only decision to make – in particular, there is no effort decision. The worker is paid per mile driven. Workers have different underlying productivities, but productivity is initially unknown, both to the worker and the firm. Each week's miles provides a noisy signal about the worker's true productivity. The worker is forward-looking, basing his quit decision on his subjective expectation of his future productivity and the option-value of staying on with the firm to learn about his productivity. However, workers may be subject to belief biases. The worker's priors need not be accurate;

stay with the firm at least 75 weeks and the overall pattern is similar (though the sample is smaller). In estimating the level of overconfidence in the structural model, the attrition process is explicitly modeled.

[54] The exact wording was "Which range best describes the annual earnings you would normally have expected from your usual jobs (regular and part-time together), if you had not started driver training with [Firm A], and your usual jobs had continued without interruption?" The eight income ranges were $\$0 - \$10,000$, $\$10,000 - \$20,000$, and so on until $\$70,000+$.

he may believe that trucking will be very lucrative or he may hold more reasonable beliefs. Further, as new productivity information arrives, he may over- or underweight his prior relative to pure Bayesian updating. Biases in priors will diminish as new information comes in, but not necessarily quickly. In addition to heterogeneity in productivity and beliefs, workers also differ in their taste for their job or career. Quitting decisions will reflect a driver's underlying taste (e.g. how much one dislikes being away from home) as well as idiosyncratic shocks (e.g. having a fight with the boss).[55]

In the model below, both the piece rate and training contracts are taken as given; both are endogenized later in Section 1.9 when I consider optimal contract design for the firm.[56]

1.6.1 Model Setup

Workers have baseline productivity η, which is distributed $N\left(\eta_0, \sigma_0^2\right)$. Workers are paid by a piece rate, w_t, that depends on their tenure.[57] Workers know the piece rate-tenure profile, but believe that this profile will not be changed by the company at some future date. The time horizon is infinite and given in weeks 1, 2, ... A worker's weekly miles, y_t, are distributed $N\left(a\left(t\right) + \eta, \sigma_y^2\right)$, and weekly earnings are thus $Y_t = w_t y_t$. $a(t)$ represents a known learn-by-doing process, which I specify below. The worker also has an outside option r_t. Every period t, the worker makes a decision, d_t, whether to stay $(d_t = 1)$ or to quit $(d_t = 0)$. Workers make the decision to quit in t having observed their past miles $y_1, y_2, ..., y_{t-1}$, but not their current week miles, y_t. Workers are assumed to be risk-neutral and to have a discount factor given by δ.[58]

Stay-or-Quit Decisions. Workers make their stay-or-quit decisions every period to maximize expected utility:

$$V_t(\mathbf{x}_t) = \max_{d_t, d_{t+1}, ...} E_t\left(\sum_{s=t}^{\infty} \delta^{t-s} u_t\left(d_s, \mathbf{x}_s\right) | d_t, \mathbf{x}_t\right). \tag{1.6}$$

where $\mathbf{x}_t$ is the vector of state variables at time t. $\mathbf{x}_t$ includes past miles, $y_1, ..., y_{t-1}$, and is described further below. The maximization problem can also be written as the following Bellman Equation:

[55] I assume that workers' tastes for the job are realized immediately instead of learned about over time. That learning occurs only over productivity may be unrealistic; there is also likely learning about other aspects of job fit, as trucking is different from most other jobs (e.g. away from home for weeks at a time, working by oneself, etc.). Investigating turnover with multiple dimensions of learning is an important subject for future research.

[56] In addition, I abstract away from the firm's ability to fire workers.

[57] This is the standard pay system for the U.S. trucking industry. For example, a trucker who does not own his own truck may receive 30 cents per mile.

[58] Risk neutrality is assumed in many dynamic learning models, e.g. Crawford and Shum (2005), Nagypal (2007), Stange (forthcoming), and Goettler and Clay (2010), though not in all (for examples with risk aversion, see the recent survey paper by Ching et al. (2011)). Coscelli and Shum (2004) show that risk parameters are not identified in certain classes of learning models.

$$V_t(\mathbf{x}_t) = \max_{d_t} E_t\left(u_t\left(d_t, \mathbf{x}_t\right) + \delta V_{t+1}(\mathbf{x}_{t+1})|d_t, \mathbf{x}_t\right).$$

I assume that there is unobserved heterogeneity in how much drivers enjoy working in trucking. A worker's non-pecuniary taste for working in trucking is denoted by α. The per-period utility from staying is assumed to be given as the sum of utility from earnings, the taste for trucking, and an idiosyncratic shock:

$$u_t\left(1, \mathbf{x}_t\right) = \alpha + w_t y_t + \varepsilon_t^S,$$

where ε_t^S is an iid idiosyncratic error unobserved to the econometrician (but observed by the worker) with an Extreme Value-Type 1 distribution and scale parameter τ. I assume that the unobserved heterogeneity is drawn from a mass-point distribution (Heckman and Singer, 1984).[59]

If the worker quits, he may have to pay a fine associated with the training contract. Let the vector $\mathbf{k}$ denote the training contract, with k_t being the penalty for quitting at tenure t. The utility from quitting can be written as:

$$u_t\left(0, \mathbf{x}_t\right) = -k_t + \frac{r_t}{1-\delta} + \varepsilon_t^Q.$$

where ε_t^Q is an iid unobserved idiosyncratic error with the same distribution as ε_t^S. Note that the fraction $\frac{r_t}{1-\delta}$ includes a time subscript, expressing that a worker's outside option may depend on his tenure at quitting (though the reservation wage is assumed to be constant after some distant time T). It is useful to define the two choice-specific value functions. Let

$$\begin{aligned} V_t^S &\equiv E_t\left(u_t\left(1, \mathbf{x}_t\right) + \delta V_t(\mathbf{x}_{t+1})|1, \mathbf{x}_t\right) \\ V_t^Q &\equiv E_t\left(u_t\left(0, \mathbf{x}_t\right) + \delta V_t(\mathbf{x}_{t+1})|0, \mathbf{x}_t\right) \end{aligned}$$

be the value functions for staying and quitting, respectively. Using the continuation values, and the above expressions for $u_t\left(1, \mathbf{x}_t\right)$ and $u_t\left(0, \mathbf{x}_t\right)$, the choice-specific value functions can be written as:

$$\begin{aligned} V_t^Q &= -k_t + \frac{r_t}{1-\delta} + \epsilon_t^Q \equiv \overline{V}_t^Q + \epsilon_t^Q \\ V_t^S &= \alpha + E_t(w_t y_t|\mathbf{x}_t) + \delta E(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) + \epsilon_t^S \equiv \overline{V}_t^S + \epsilon_t^S, \end{aligned}$$

and the Bellman Equation can be re-written as

$$V_t(\mathbf{x}_t) = \max_{d_t \in \{0,1\}} \left(V_t^S\left(\mathbf{x}_t\right), V_t^Q\left(\mathbf{x}_t\right)\right).$$

Agents gradually learn their productivity as more and more productivity signals are

[59] I have also performed estimation allowing the unobserved heterogeneity to be normally distributed.

observed. Thus, after a sufficiently large number of periods, T, the value function can be approximated by the following asymptotic value functions:

$$\begin{aligned} V^Q &= \frac{r_T}{1-\delta} + \epsilon^Q \equiv \overline{V}^Q + \epsilon^Q \\ V^S &= \alpha + w_T \eta + \delta E(V(\mathbf{x}')|\mathbf{x}) + \epsilon^S \equiv \overline{V}^S + \epsilon^S \\ V(\mathbf{x}) &= \max_{d \in \{0,1\}} \left(V^S(\mathbf{x}), V^Q(\mathbf{x}) \right) \end{aligned}$$

Belief Formation. In a standard normal learning model, a worker's beliefs in period t about his productivity are given by:

$$E(y_t|y_1,...,y_{t-1}) = \frac{\sigma_y^2}{(t-1)\,\sigma_0^2 + \sigma_y^2}\eta_0 + \frac{(t-1)\,\sigma_0^2}{(t-1)\,\sigma_0^2 + \sigma_y^2} \frac{\sum_{s=1}^{t-1} y_s - a(s)}{s-1} + a(t) \tag{1.7}$$

This expression is a weighted sum of the agent's prior and his average productivity signals. As t increases, the agent will eventually shift all the weight to the $\overline{y}_{t-1}$ term over the η_0 term. I augment the standard learning model in two ways. First, I allow for agents to be overconfident. Instead of believing that their productivity η is drawn from a distribution $N(\eta_0, \sigma_0^2)$, agents believe that the productivity is drawn from a distribution $N(\eta_0 + \eta_b, \sigma_0^2)$. Second, I allow for agents to have a perception of signal noise that may be different from the true signal noise in the population. Specifically, workers may perceive the standard deviation of weekly productivity signals to be $\widetilde{\sigma_y}$ instead of σ_y. With these two assumption, an agent's subjective expectation of his productivity, which I denote by E^b (where b stands for belief) is:

$$E^b(y_t|y_1,...,y_{t-1}) = \frac{\widetilde{\sigma_y}^2}{(t-1)\,\sigma_0^2 + \widetilde{\sigma_y}^2}(\eta_0 + \eta_b) + \frac{(t-1)\,\sigma_0^2}{(t-1)\,\sigma_0^2 + \widetilde{\sigma_y}^2} \frac{\sum_{s=1}^{t-1} y_s - a(s)}{s-1} + a(t) \tag{1.8}$$

If the term η_b is greater (less) than zero, then agents exhibit positive (negative) mean bias or overconfidence (underconfidence). As more signals come in, agents will eventually learn not to be overconfident since they will put zero weight on the $(\eta_0 + \eta_b)$ term. The speed at which this occurs, however, will be determined by $\widetilde{\sigma_y}$.

I allow that workers' reported subjective beliefs include some measurement error. The rationale is that accurately reporting one's beliefs about productivity may be unusual or unfamiliar for a worker. I assume that reported beliefs equal subjective beliefs plus a normally distributed error. Specifically, the reported subjective belief of driver i in his t^{th} week of tenure, b_{it}, is distributed:

$$b_{it} \sim N\left(E^b\left(y_{it}|y_{i1},...,y_{it-1}\right), \sigma_b^2\right)$$

Summary of Within Period Timing. The within period timing in week t can be summarized as follows:

1. Workers form beliefs b_t given past earnings $y_1, y_2, ..., y_{t-1}$.
2. ϵ_t^S and ϵ_t^Q are realized and workers decide whether or not to quit.
3. y_t is realized, if they do not quit.

Learning by Doing and Skill Accumulation. Productivity increases over time with the function $a(t)$. I assume that $a(t) = a_1 * \Lambda(a_2 t)$ where $\Lambda(x) = \frac{e(x)}{1+e(x)}$ is the logit function and t is worker tenure in weeks.[60] Learning by doing depends only on tenure; thus, the speed of learning does not depend on the number of miles driven or on the ability of the driver. Workers fully anticipate the path of $a(t)$.

In addition, I account for skill accumulation immediately following CDL training. At Firm A and other trucking firms, truckdrivers who have just completed their CDL training often spend several weeks driving with an experienced driver riding along. At Firm A, this process lasts 4 to 6 weeks. During this time, drivers receive a flat rat of roughly \$375 per week and their productivity is not recorded in the payroll data. I thus assume that drivers do not begin learning from their earnings realizations until after 5 weeks. I also account for the fact that drivers who quit during this initial period of driving along a senior driver may lack some skills required to be a productive truckdriver. Specifically, I assume the reservation wage over time is given as follows:

$$
\begin{aligned}
r_t &= r - \frac{6-t}{5} s_0 \text{ for } t \leq 5 \\
r_t &= r \text{ for } t > 5
\end{aligned}
$$

With this specification, r is fixed using outside data and s_0, the value of skills gained while working with a senior driver, is a parameter to estimate.[61]

State Variables and Observed Heterogeneity. The state variables consist of past earnings, a vector of observed time-invariant individual characteristics, the piece rate, the training contract, taste heterogeneity, and a person's level of overconfidence: $\mathbf{x}_t = (y_1, ..., y_{t-1}, X, w_t, k_t, \alpha, \eta_b)$, where X is a vector of observable additional characteristics. Differences across people are accounted for by allowing for heterogeneity in taste for the job and in overconfidence. I have also allowed workers' taste and beliefs to depend on observable characteristics.

Solving the Model. To solve the model, I first solve for the asymptotic value functions (after all learning has taken placed) using value function iteration. With the asymptotic value

[60] The logit functional form is consistent with Jovanovic and Nyarko's (1996) micro-founded model of learning by doing in which the speed of learning decreases over time.

[61] I have also estimated allowing for skill accumulation to continue as driver tenure increases. Specifically, I assume that $r_t = r + \theta_1 \Lambda(\theta_2 t)$, where θ_1 and θ_2 are parameters to estimate.

functions in hand, backward recursion can then be applied to solve the dynamic programming problem. At the time T, the probability of staying, given the state variable, is:

$$\begin{aligned}
Pr(STAY_T|\mathbf{x}_T) &= Pr(V_T^S > V_T^Q|_1, ..., y_{t-1}, X, w_t, k_t, \alpha, \eta_b) \\
&= Pr(\alpha + X\overline{\alpha} + E^b(w_T y_T|y_1, ..., y_{T-1}) + \delta E(V(\mathbf{x})|\mathbf{x}_T) + \epsilon_T^S > -k_T + \frac{r}{1-\delta} + \epsilon_T^Q) \\
&= \Lambda\left(\frac{\alpha + X\overline{\alpha} + w_T E^b(y_T|y_1, ..., y_{T-1}) + \delta E(V(\mathbf{x})|\mathbf{x}_T) + k_T - \frac{r}{1-\delta}}{\tau}\right)
\end{aligned}$$

where $\Lambda(x) = \frac{e(x)}{1+e(x)}$ is the logit function. To evaluate this probability, it is necessary to calculate $E^b(y_T|y_1, ..., y_{T-1})$ and $E(V(\mathbf{x})|\mathbf{x}_T)$. This expectation depends on $y_1, ..., y_{T-1}$, which would imply that the state space has dimensionality of order K^{T-1} when y_t is discretized with K values. The key to avoiding a very high dimensional problem is that, in a normal learning model, and in the generalized learning model I consider, the worker's expectation of future productivity depends only on the average of all past productivity realizations. That is, since $E^b(y_t|y_1, ..., y_{t-1}) = \frac{\widetilde{\sigma_y}^2}{(t-1)\sigma_0^2+\widetilde{\sigma_y}^2}(\eta_0 + \eta_b) + \frac{(t-1)\sigma_0^2}{(t-1)\sigma_0^2+\widetilde{\sigma_y}^2}\overline{y}_{t-1}$, the *average of his past productivity is a sufficient statistic for the entire sequence* $y_1, ..., y_{t-1}$ (DeGroot, 1970). I show how to calculate $E(V(\mathbf{x})|\mathbf{x}_T)$ in the Appendix.

For a general period t, the probability of staying is given by:

$$\begin{aligned}
Pr(STAY_t|\mathbf{x}_t) &= Pr(V_t^S > V_t^Q|\mathbf{x}_t) \\
&= Pr(\alpha + X\overline{\alpha} + w_t E^b(y_t|y_1, ..., y_{t-1}) + \epsilon_t^S + \delta E(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) > -k_t + \frac{r}{1-\delta} + \epsilon_t^Q) \\
&= \Lambda\left(\frac{\alpha + X\overline{\alpha} + w_t E(y_t|y_1, ..., y_{t-1}) + \delta E(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) + k_t - \frac{r}{1-\delta}}{\tau}\right)
\end{aligned}$$

Calculating $E(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t)$ requires integrating expectations both of future earnings and future idiosyncratic shocks, and is done formally in the Appendix.

Exogenous Separations. In some specifications, I also allow for exogenous separations, that is, a probability that the worker-firm match will be destroyed.

1.6.2 Discussion of Model Assumptions

Outside of how workers form beliefs, most assumptions in the model are relatively standard. While worker learning is allowed to be non-standard, deviations such as overconfident priors will be determined by the data. The strongest assumption in the model is that workers are not overconfident about the outside option despite potentially being overconfident about their ability in trucking. Some evidence supporting the assumption was presented in Figure 1.8. However, the assumption is not required for overconfidence to have some effect on the profitability of training. As I show in the Appendix, firm profitability will be higher if there is some differential overconfidence, that is, the worker's perception of his current job income is more responsive to his productivity beliefs than his outside option. To the extent that workers are also overconfident about their outside option, this will attenuate the importance of overconfidence for firm profitability.

How reasonable is it to think that overconfidence affects the current job earnings more than the outside option? Suppose that the outside option is a non-trucking job. While current earnings are proportional to ability in trucking, most jobs in the U.S. do not pay piece rates. Thus, even if workers are overconfident about their productivity at their outside option, it is unclear how much this affects workers' perceptions of their earnings at their outside option.[62] Suppose now that the outside option is another trucking job. Insofar as drivers select the job at which they believe their ability will be the highest, they will be differentially overconfident about their ability at the current job relative to the outside option.[63]

1.7 Structural Estimation

1.7.1 Estimation

The model is estimated using maximum likelihood. Let $L_i = L(d_{i1}, ..., d_{it}, y_{i1}, ..., y_{it}, b_{i1}, ..., b_{it})$ be the likelihood of driver i for an observed sequence of quitting decisions, earnings realizations, and subjective beliefs. The likelihood can be expanded as follows:

$$\begin{aligned} L_i &= \int L(d_{i1}, ..., d_{it}, y_{i1}, ..., y_{it}, b_{i1}, ..., b_{it} | \alpha, \eta_b) f(\alpha, \eta_b)\, d\alpha d\eta_b \\ &= \left(\prod_{s=1}^{t} L(y_{is} | y_{i1} ..., y_{is-1}) \right) \left[\int \prod_{s=1}^{t} L(d_{is} | y_{i1}, ..., y_{is-1}, \alpha, \eta_b) \prod_{s=1}^{t} L(b_{is} | y_{i1} ..., y_{is-1}, \eta_b) f(\alpha, \eta_b)\, d\alpha d\eta_b \right] \\ &\equiv L_i^2 \int L_i^1(\alpha, \eta_b) L_i^3(\eta_b) f(\alpha, \eta_b)\, d\alpha d\eta_b \end{aligned}$$

$L_i^1(\alpha, \eta_b)$ is defined as the contribution to the likelihood from quitting decisions; L_i^2 is the contribution to the likelihood from earnings realizations; and $L_i^3(\eta_b)$ is the contribution to the likelihood from subjective beliefs. That the likelihood can be decomposed in this way

[62] Performance pay is used in only 37% of U.S. jobs, comprises a median of 4% of total pay across jobs, and is less common in blue-collar jobs than white-collar jobs (Lemieux et al., 2009). Of course, other pecuniary aspects of a job, e.g. the perceived probability of being promoted to a higher wage, may be affected by overconfidence.

[63] The argument that self-selection can promote belief biases is made in Van den Steen (2004). In the case where the outside option is another trucking job, the assumption that the worker is more overconfident at the current wage than at his outside option may be testable. Assume that the worker's overconfidence is correlated across firms with different multiplicative values, $\eta_{b,f} = \theta_f * \theta_b$, where $\eta_{b,f}$ is a worker's belief bias at firm f. In the regressions of quitting on beliefs in Table 1.5, I showed that workers who have higher beliefs are less likely to quit.

is shown formally in the Appendix. For a driver who quits in period t, $L_i^1(\alpha,\eta_b)$, L_i^2, and $L_i^3(\eta_b)$ can be written as

$$\begin{aligned} L_i^1(\alpha,\eta_b) &= \left(\prod_{s=t}^{t-1} Pr\left(STAY_{is}|\mathbf{x}_{is}\right)\right)\left(1-\Pr(STAY_{it}|\mathbf{x}_{is})\right) \\ L_i^2 &= f(y_{i1}) * \prod_{s=2}^{t} f(y_{is}|y_{i1},...,y_{is-1}) \\ L_i^3(\eta_b) &= f(b_{i1}) * \prod_{s=2}^{t} f(b_{is}|y_{i1},...,y_{is-1}) \end{aligned}$$

with

$$\begin{aligned} f(y_{i1}) &\sim N(\eta_0,\sigma_0^2+\sigma_y^2) \\ f(y_{is}|y_{i1},...,y_{is-1}) &\sim N((1-\gamma_{s-1})\,\eta_0+\gamma_{s-1}\bar{y}_{is-1},\Omega_{s-1}) \text{ for } s>1 \\ f(b_{i1}) &\sim N(\eta_0+\eta_b,\sigma_b^2) \\ f(b_{is}|y_{i1},...,y_{is-1}) &\sim N(\left(1-\gamma_{s-1}^b\right)(\eta_0+\eta_b)+\gamma_{s-1}^b\bar{y}_{is-1},\sigma_b^2) \text{ for } s>1 \end{aligned}$$

and where $\gamma_s = \frac{s\sigma_0^2}{s\sigma_0^2+\sigma_y^2}$, $\Omega_s = \frac{\sigma_0^2\sigma_y^2}{s\sigma_0^2+\sigma_y^2}+\sigma_y^2$, and $\gamma_s^b = \frac{s\sigma_0^2}{s\sigma_0^2+\widetilde{\sigma_y}^2}$. The overall likelihood is computed, first, by integrating over the unobserved heterogeneity for each individual's likelihood, and then by taking the product over all people. Since the unobserved heterogeneity is mass-point distributed, the integral becomes a sum.

$$\begin{aligned} L &= \prod_i \int L_i^1(\alpha,\eta_b)L_i^2L_i^3(\eta_b)f(\alpha,\eta_b)d\alpha d\eta_b \\ &= \prod_i \left(\int L_i^1(\alpha,\eta_b)L_i^3(\eta_b)f(\alpha,\eta_b)d\alpha d\eta_b\right)L_i^2. \\ \log(L) &= \sum_i \log\left(\sum_{\alpha,\eta_b} L_i^1(\alpha,\eta_b)L_i^3(\eta_b)f(\alpha,\eta_b)\right)+\sum_i \log\left(L_i^2\right) \end{aligned}$$

1.7.2 Identification

I discuss which data features allow me to identify the model parameters.

Productivity and skill parameters. The productivity parameters σ_0, σ_y, and η_0 are identified primarily by the productivity data. σ_0 reflects the degree of permanent productivity differences *across* individuals whereas σ_y reflects differences *within* individuals in productivity. The parameter η_0 reflects mean average ability of workers in the population. The skill gain parameter, s_0, is identified based on turnover levels during the first five weeks when workers are driving with an experienced driver and receiving a flat wage instead of a piece rate.

Taste heterogeneity. The taste heterogeneity parameters are identified from persistent differences between individual quitting behavior and the predictions of the model. Suppose that the data contained many low-productivity workers who nevertheless kept choosing not to quit. This would cause the model to estimate that there is a large amount of unobserved taste heterogeneity.

Belief parameters. The subjective expectations data are critical for identification of all the belief parameters. The standard deviation of beliefs, σ_b, is identified by how much subjective expectations differ from the mathematical expectation of future productivity. The greater the variation in subjective expectations from mathematical expectations (that is, the greater the measurement error in state beliefs), the greater is σ_b. The believed standard deviation of productivity shocks, $\widetilde{\sigma_y}$, determines the subjective speed of learning in the model. Because workers use this subjective speed of learning both to make quit decisions and to form subjective beliefs, the parameter is over-identified. The faster that agents begin to weight their productivity realizations to date in making their quit decisions, the smaller that $\widetilde{\sigma_y}$ will be. Likewise, the faster that agents' initial overconfidence in subjective beliefs begins to disappear, the smaller that $\widetilde{\sigma_y}$ will be. The belief heterogeneity is identified from persistent differences across individuals regarding how their subjective beliefs compare to the model's mathematical expectation of future productivity. Suppose that some workers persistently report productivity beliefs that are much larger than their actual productivity, whereas other workers are well-calibrated. This would lead to estimating a large amount of belief heterogeneity.

Other parameters. The scale parameter of the idiosyncratic shock, τ, is identified based off of how much quitting behavior in the data differs from that predicted by a model with individual unobserved heterogeneity, but not time-varying uncertainty.

1.7.3 Implementation

To solve the dynamic programming problem numerically, I discretize productivity into K values. In my baseline estimation, I let productivity range in increments of 100 from 100 to 4,000 miles per week, *i.e.* $K = 40$.[64] Transitions between earnings states are given by:[65]

$$\Pr(y_s^k|y_1, ..., y_{s-1}) = \Phi\left(\frac{y_s^k + .5 * kstep - E(y_s^k|y_1, ..., y_{s-1})}{\sqrt{\Omega_{s-1}}}\right) - \Phi\left(\frac{y_s^k - .5 * kstep - E(y_s^k|y_1, ..., y_{s-1})}{\sqrt{\Omega_{s-1}}}\right)$$

where $kstep$ is the distance between earnings realizations.

The outside option, r, is taken to be the median full-time earnings from the 2006 March Current Population Survey of workers like the "median" driver (35-year old males with a

[64] As a sensitivity check, I have also estimated with finer and coarser discretizations, as explored for example in Rust (1987). The results are qualitatively similar except when the grid is very coarse. I have also estimated using linear interpolation instead of discretization and the results are similar.

[65] See Rust (1996) and Stange (forthcoming) for similar formulas.

high school degree), which is \$32,000 per year. I convert this to a weekly wage of \$640. The weekly discount factor, δ, is assumed rather than estimated. In my preferred specification, I take $\delta = 0.9957$, corresponding to an annual discount factor of 0.8.[66]

A few other points are worth noting. First, I assume that drivers act as if the contract is perfectly enforceable.[67] Second, I report results here without including covariates like education and age.[68] Third, the data contain a number of zero mile, zero earnings weeks for drivers. During these weeks, the driver is not working. These weeks do not count toward the earnings component of the likelihood, and average earnings to date (in terms of the quit decision) are given by the prior week's average earnings to date.

1.8 Structural Results

Table 1.6 displays the main structural estimates. As a benchmark, column 1 provides estimates assuming no belief bias. Column 2 allows workers to have a mean bias in their productivity expectations, that is, to be over- or underconfident about their ability. Workers are estimated to have an initial belief bias of 589 miles. Given the estimated mean of the prior productivity distribution of 2,025 miles, workers are estimated to be overconfident by around 29%. One should note that the productivity parameters are also different from column 1 and more reasonable in size.[69] The believed standard deviation of productivity shocks is roughly 2.5 times higher than the actual standard deviation of productivity shocks. This implies that the rate at which agents update their quit-related productivity expectations and reported subjective productivity expectations is a fair amount slower then the rate predicted by the productivity data alone. Recall that the weight agents place on their signals relative to their prior is $\frac{t\sigma_0^2}{t\sigma_0^2+\widetilde{\sigma_y}^2}$. After 20 weeks, the worker will place weight 0.31 on his signals, whereas if $\widetilde{\sigma_y} = \sigma_y$, he would place weight 0.77 on his signals.

Column 3 allows for heterogeneity in people's belief bias. There is again a considerable improvement in the log-likelihood. The data reveal the majority of people to be moderately

[66] I have experimented with a broad range of discount factors in sensitivity analysis. Model fit appears best for discount factors in this range, though it still is quite reasonable for those corresponding to annual discount factor of 0.90. An annual discount factor of 0.80 is "low," but is comparable or higher than discount factors used or estimated in other analyzing dynamic choices of blue-collar or low-income workers (e.g. Paserman, 2008; Fang and Silverman, 2009; Warner and Pleeter, 2001).

[67] As described above, even though only a portion of the amount owed was collected, I assume that drivers act as if the utility cost of quitting is equivalent to the utility loss from paying the contract penalty. The firm was very firm with new drivers about its intention to collect money owed upon a quit. Combining this with the actual aggressiveness of the collection efforts, as well as the reporting of delinquency to credit agencies, this assumption may not be tremendously unrealistic. I have also considered robustness checks where drivers act as if the utility loss from quitting is only a portion of the contract penalty.

[68] I have estimated models including covariates as described above. However, including covariates has little effect on model fit and on the estimates of the other parameters.

[69] For example, the mean of the prior productivity distribution is 2,025 miles per week, down roughly 20% from 2,468 miles per week in column 1. In column 1, the prior productivity distribution needs to explain the earnings data. But it also needs to explain the quitting and subjective beliefs data, which "pulls" the estimate substantially upward.

overconfident and a small group of people to be severely overconfident. There is also significant heterogeneity in people's taste / non-monetary preference for working in trucking. All three models suggest three types of workers: Workers with a strong distaste for working in trucking, workers with moderate distaste, and workers with a moderate positive taste. Given that workers are forced to be away from home for several weeks at a time, that many workers dislike this is unsurprising.

The fit of the model is assessed graphically in Figure 1.11. The quit hazard-tenure profile, the productivity-tenure profile, and the beliefs-tenure profile observed in the data are plotted using an Epanechnikov kernel. Predicted quitting behavior is plotted by simulating the careers of 40,000 drivers. The model appears to match several patterns in the data. The hazard of quitting is initially increasing in both the model and the data. In the model, this is due to learning about productivity. When workers are uncertain about their productivity, they face an incentive to wait and see how productive they will be before deciding to quit. There is a large spike in quitting after 52 weeks as predicted by the model, the time at which drivers come off the 12-month contract. Learning also helps rationalize the gradually decreasing beliefs-tenure profile, though the fit is imperfect.

To further assess model fit, I can compare the observed number of drivers quitting (O_t) with the number predicted from the model (E_t) using a chi-squared test.[70] The chi-squared statistic for the model-predicted quit behavior is given by $\sum_t \frac{(E_t - O_t)^2}{E_t}$. Statistical inference is conducted using T degrees of freedom, where T is the maximum number of weeks a driver is observed. For column 1, the chi-squared statistic is 504.7684, whereas for columns 2 and 3, the chi-squared statistics are 216.1135 and 202.9855. Thus, although the data reject model 3 ($p = 1.6856e-007$), the fit in terms of turnover is a considerable improvement over models without belief bias or with no belief heterogeneity. This occurs even though the belief bias parameters are simultaneously fitting both the belief data and turnover data.

Table 1.7 takes the baseline model and adds learning by doing and logit skill accumulation. The fit is better in all three specifications than in Table 1.6. Many of the parameters are qualitatively similar to before, but there are some differences. In column two, the mean prior bias is larger than before, estimated at 722 miles. The estimated taste heterogeneity is also somewhat different. Further results with learning by doing are given in the Web Appendix where instead of allowing for gradual skill acquisition, the models estimated allow for an exogenous probability of match separation after one year. The fit is inferior to that in Table 1.7, with lower log-likelihoods.

Robustness. I conduct a number of robustness checks of my main findings. I vary the discount factor, the number of mass points, the maximum allowed subjective belief, and the maximum time during which learning occurs. Results are in the Web Appendix. Comparing log-likelihoods under the assumed annualized discount factor of 0.80 with a fully myopic alternative ($\delta = 0$), the data soundly reject the myopic model ($p = 0.00$ in a likelihood ratio

[70] Chi-squared tests are often used to assess the fit of dynamic models, e.g. Keane and Wolpin (1997) and Card and Hyslop (2005). As noted in Card and Hyslop (2005), it is more correct to think of the calculated chi-squared statistic as an informal measure of fit, since the predicted numbers are created from the same data being used for the observed cell entries.

test). The model fit is slightly better for an annualized discount factor of 0.80 than 0.90 or 0.95, but the difference is not statistically significant. To account for the potential impact of differential response to the subjective belief questions, I perform a robustness check using Inverse Probability Weighting (Wooldridge, 2002). In the first stage, I fit a probit model of whether a driver responds to the survey as a function of time-invariant observables. I then use the inverse of the predicted values to weight each driver's contribution to the likelihood. Results are shown in the Web Appendix and are quite similar to the unweighted estimates.

Out-of-Sample Fit. A more demanding task of the model is to see whether it can predict well out-of-sample. The model is estimated using workers facing the 12-month contract, as this is the only data for which subjective beliefs at Firm A were collected. It is natural to ask whether the model can be simulated to predict behavior under the no contract and 18-month contract regimes. As I show in the Web Appendix, the model can predict the basic shape of the no contract and 18-month contract conditions, but the fit is imperfect.

1.9 Counterfactual Simulations

The section uses the structural estimates to assess the quantitative importance of training contracts and biased beliefs, as well as to analyze potential firm or government policies.

1.9.1 Profits and Welfare Under the Three Contracts

I first asses how different training contracts affect profits and welfare by simulating worker behavior under the three observed training contracts. I use two measures of firm profits: profits per worker and profits per truck.[71] Average profits per worker is defined as average profits brought in from a worker during his time with the company. Implicit in such a formulation is that when a worker quits, he is not replaced.

$$\begin{aligned} \pi(k) &= \text{Trucking Profits + Training Contract Penalties - Training Costs} \qquad (1.9) \\ &= \sum_{t=1}^{\infty} \delta^{t-1}(1-Q_t)((P-w_t-mc)y_t - FC) + \sum_{t=1}^{\infty} \delta^{t-1}\theta k_t q_t - TC \end{aligned}$$

where q_t is a dummy for quitting in week t, $Q_t = \sum_{s=1}^{t} q_s$ is a whether a driver has quit in the first t weeks, y_t is a driver's productivity, P_t is the price the firm charges for one mile of shipment, mc is the non-wage marginal cost per mile (i.e. truck wear and gas costs), θ is the share of the training contract penalty collected by the firm, FC is fixed costs per week (i.e. support for the drivers and the opportunity cost of the truck), and TC is training cost per worker. I assume values for P, mc, θ, FC, and TC based on consultation with Firm A.[72]

[71] See Bojilov (2011) for similar profit formulas in a different context.

[72] Specifically, I assume that $P = \$1.80$ per mile, $mc = \$1.20$ per mile, $\theta = 0.3$, $FC = \$475$ per week, and

Profits can also be defined in terms of profits per truck. In this formulation, when a worker quits, he is replaced by the company the next period at some cost, which I set equal to the cost of the training. The formula for profits per truck is:

$$\pi = \sum_{t=1}^{\infty} \delta^{t-1} \left[(1 - Q_t)((P - w_t - mc)y_t - FC) + Q_t \left(-TC + \theta k_t + \pi \right) \right] \tag{1.10}$$

which can be re-arranged to yield

$$\pi = \frac{\sum\limits_{t=1}^{\infty} \delta^{t-1} \left[(1 - Q_t)((P - w_t - mc)y_t - FC) + Q_t \left(-TC + \theta k_t \right) \right]}{1 - \sum\limits_{t=1}^{\infty} \delta^{t-1} Q_t} \tag{1.11}$$

Since workers have biased beliefs, average actual (experienced) utility will differ from a worker's expected utility. Worker welfare is measured using experienced utility by summing earnings, taste for trucking, and idiosyncratic shocks. I simulate 3,000 workers for up to 1,300 weeks each. As seen in Table 1.8, both measures of profits are higher under the 12-month contract than no contract, and higher under the 18-month contract than the 12-month contract. These results suggest that Firm A increased profits through its contractual changes, though also lowered worker welfare.

1.9.2 Debiasing: Reducing Worker Overconfidence

Workers exhibit overconfidence and overconfidence improves the fit of the structural model. To examine quantitatively how overconfidence affects quitting behavior, worker welfare, firm profits, and firms' contractual choices, a natural counterfactual is to eliminate overconfidence, which I refer to as debiasing.

Before describing the counterfactual, it is worth considering whether reducing people's overconfidence is feasible at all. Psychologists have long been interested in whether overconfidence and other behavioral biases can be eliminated, focusing primarily on laboratory settings.[73] In Hoffman (2011b), I show using a field experiment at another trucking firm that informing workers about other workers' overconfidence can reduce confidence. Alternatively, overconfidence might be reduced if workers are informed about the distribution of productivity or if they receive individual-level feedback about their productivity and productivity

---

$TC = \$2,500$. It should be noted that even within Firm A there is variation in P and MC depending on the shipment and the driving conditions. The price per mile charged by Firm A to a shipper may depend on the type of commodity being hauled and on whether it is driven by team drivers (two drivers sharing a truck) who will deliver the good more quickly than solo drivers. In addition, workers may have different marginal costs depending on how they use the equipment. Incorporating such heterogeneity, however, is beyond the scope of this paper.

[73] Fischoff (1982) provides an excellent early summary of the literature. Many papers provide support for the feasibility of laboratory debiasing (e.g. Arkes et al., 1987; Lau and Coiera, 2009), but many also do not (e.g. Sanna et al., 2002; Fleisig, 2011).

forecasts.[74] To the extent that debiasing is feasible, the counterfactual can be viewed not only as quantifying overconfidnece, but also as a potential public policy experiment, e.g. a government information campaign that reduced overconfidence.

To examine the effects of debiasing, I simulate the structural model with the overconfidence mass points reduced. In addition to a full elimination of overconfidence, I also simulate a reduction of overconfidence by one half, recognizing that debiasing may be incomplete in practice. Results can be seen in Table 1.9. I find that the intervention substantially reduces worker retention; when workers are overconfident, they become much more likely to quit because they no longer unrealistically foresee themselves as being highly productive in the future. This also raises worker welfare since worker quitting decisions become less distorted by overconfidence. In addition, debiasing significantly reduces profits per worker and profits per truck. Because workers are quitting earlier, the firm has less time to make profits on a given worker, lowering profits per worker. Higher turnover from debiasing also means the firm must spend more on training costs to train new drivers.

Comparison Counterfactual: Reveal Ability to Worker After Training. To illustrate the importance of learning, I consider a counterfactual where learning is "turned off." That is, I examine what happens if worker ability is revealed to workers immediately after training. In addition to eliminating uncertainty about ability, such a counterfactual eliminates all overconfidence since workers do not rely on their potentially biased priors to infer their ability. As seen in Table 1.9, revealing ability reduces retention at 20 weeks below that of 100% debiasing since many workers would quit immediately after discovering that they had low ability. The counterfactual also raises worker welfare and firm profits compared to 100% debiasing due to better allocation; productive workers become more likely to stay with Firm A and unproductive workers become more likely to leave. However, compared to the baseline, firm profits are substantially lower. Thus, it appears that firms benefit and workers lose because ability is gradually revealed in this market; if ability were revealed immediately, overconfidence would disappear, which as argued above, is bad for firms.

1.9.3 Debiasing with Optimal Contractual Responses

In the above counterfactuals, I take firms' training contracts as fixed. However, in reality, firms might adjust their training contracts in response to debiasing. When workers' overconfidence is reduced, workers will become more likely to quit after receiving training, and firms may wish to change the quitting penalties in response. In addition, if workers are debiased before they sign up for training, it may be more difficult to get them to sign a training contract.

[74] Outside the psychology literature, legal scholars have also taken an interest in debiasing, e.g. Jolls and Sunstein (2006). In addition to discussing information as a form of debiasing, Jolls and Sunstein (2006) also discuss how legal institutions may reduce behavioral biases. For example, one rationale for having an outside member of a corporate board is that they may be less prone to overoptimism about company performance than inside members.

In the optimal contracting problem I consider, the firm sets the piece rate and quit penalty for every week $1, 2, ..., \infty$. Rather than solving analytically, I attack the problem using computational methods.[75] Even with modern computational tools, I need to simplify the problem. The training contract penalty is assumed to be a flat level $\bar{k}$ for weeks 1-52 and zero thereafter. The wage is restricted to be a scale-factor $\bar{w}$ times the actual wage profile. The firm solves the problem:

$$max_{\bar{k},\bar{w}} E\pi(\bar{k}, \bar{w}) \text{ s.t. } E^b V(\bar{k}, \bar{w}) \geq \bar{u} \tag{1.12}$$

where $E\pi(\bar{k}, \bar{w})$ is the firm's expected profit per worker and $E^b V(\bar{k}, \bar{w})$ is the worker's perceived expected utility under $\bar{k}$ and $\bar{w}$. $\bar{u}$ is the minimum utility level at which the worker will sign the training contract. Since worker perceived welfare is strictly decreasing in quit penalties in increasing in the piece rate, the problem is condensed to a one-dimensional optimization problem. The firm chooses training contract penalties and piece rate schedules lying along an indifference curve.

I perform the analysis with $\bar{u}$ equal to the perceived utility a worker would have gotten in each of the counterfactuals in Table 1.9. Estimates are shown in Table 1.10. Column 1 of Table 1.10 reports the optimal contract in the baseline situation, before any debiasing has occurred. The optimal contract involves a higher quit penalty and higher wage than that observed in practice. When overconfidence or learning are removed, the optimal quit penalties decrease. This occurs because overconfident workers believe it is unlikely they will quit and they are thus willing to accept a high quit penalty for a small wage increase.[76] Other aspects of the counterfactuals are similar to before. Debiasing decreases worker retention and firm profits, but increases worker welfare. Thus, the qualitative conclusions of the counterfactuals are unchanged when firms optimally set contracts compared to counterfactuals when contracts are fixed. In fact, the effects of debiasing appear to be amplified when contracts are optimally set, as it is "cheaper" for firms to satisfy participation constraints when workers are overconfident.[77]

1.9.4 Banning Training Contracts

If workers have accurate beliefs about their expected productivity, then restricting the type of contracts workers and firms are allowed to sign seems unlikely to be welfare-enhancing.

[75] Several other papers use computational methods to analyze optimal contracts in rich contracting environments. For example, Armstrong et al. (2010) computationally analyze optimal compensation contracts with both moral hazard and adverse selection.

[76] In addition, higher piece rates are more appealing to workers who are more overconfident.

[77] In these counterfactuals, I have let $\bar{u}$ vary by counterfactual to equal what workers would get in perceived expected utility as in the case with no contractual response. This implies that when overconfidence is reduced, workers' perceived outside option goes down as well. An alternative approach is to use the same $\bar{u}$ in each counterfactual. Results are in the Web Appendix. Optimal profits are often lower than in Table 1.10 because as workers become debiased, their perceived expected utility goes down, and they require either a lower quit penalty or higher wage to have as high perceived utility as before. However, the other qualitative conclusions remain the same.

However, if workers are overconfident, it might be possible that they are taking on contracts that are lowering their welfare. In this case, worker welfare may be higher when firms are restricted from using training contracts. Courts could, for example, decide that training contracts should be unenforceable.

I model a ban in training contracts as a restriction on the firm's optimization problem. Instead of (1.12), the firm solves the restricted problem:

$$max_{\bar{w}} E\pi(\bar{k} = 0, \bar{w}) \text{ s.t. } E^b V(\bar{k} = 0, \bar{w}) \geq \bar{u} \tag{1.13}$$

Results are shown in the last two columns of Table 1.10. Worker welfare is 3.7% higher when training contracts are banned compared to when the firm optimally sets training contracts unrestricted. I also examine a combined policy of banning training contracts and 50% debiasing. Worker welfare is 2.7% higher under this counterfactual compared to 50% debiasing when firms optimally set training contracts. Note that if workers were fully well-calibrated, banning training contracts would have no effect since firms are optimally choosing to set $k = 0$. The value of using training contracts for a profit-maximizing firm appears to be linked to workers having biased beliefs.[78]

Robustness: Competition. I have studied optimal training contracts and banning training contracts for the case of a single monopsonistic firm. While the trucking industry as a whole is closer to perfect competition than monopsony, analyzing one firm may be a reasonable assumption for the market for firm-sponsored training. However, it is worth examining how competition would affect my conclusions. Under perfect competition, the impact of banning training contracts on worker welfare is more ambiguous than before. Although overconfident workers benefit by not taking on high quit penalties, they may be harmed since firms may need to substantially decrease wages to reach zero profits. The positive relationship between optimal quit penalties and overconfidence will remain under perfect competition since more overconfident workers dislike high quit penalties and like higher piece rates more than non-overconfident workers.[79]

[78] The analysis here assumes that firm-sponsored training would continue if training contracts were banned. It is possible, however, that banning contracts would make some firms no longer willing to train. In the Web Appendix, I modify the counterfactual so that banning training contracts means that firms stop providing training. New truckers who received training from private schools would continue to get trained that way, but truckers receiving firm-sponsored training (the ones I study in this paper) could no longer enter the industry. This counterfactual reduces worker welfare, as might be expected. However, the impacts vary substantially by worker overconfidence. In my baseline model with two types of overconfidence, welfare for very overconfident workers actually increases whereas welfare for moderately overconfident workers decreases. Since the quitting decisions of very overconfident workers are strongly distorted, they are actually better off just receiving their outside option and not getting trained.

[79] In current work, I am repeating the simulations for the case of perfect competition.

1.10 Conclusion

The question of how firms profitably train despite the potential for hold-up is a critical one both for economic theory and for policy. This paper explores the joint role of training contracts and biased beliefs in alleviating hold-up.

Using plausibly exogenous contractual variation, I show that implementing a training contract reduced quitting by 10 to 20 percent, an effect equivalent to a 2-4 percentage point increase in the driver's home state unemployment rate. I find only limited evidence of positive worker selection induced by training contracts, suggesting the effects operated through incentives instead of selection. Data on workers' subjective productivity beliefs predict actual quitting and productivity, but also indicate a significant and persistent pattern of overconfidence whereby workers believe they will achieve higher miles than they actually attain. To understand key patterns in the quits and beliefs data, I structurally estimate a quitting model with learning and overconfidence. I show that both training contracts and overconfidence are critical for the profitability of training. If training contracts were banned or overconfidence eliminated, firm profits from training would fall substantially, sometimes from positive to negative.

My results are specific to one industry, so it is worth considering to what extent the conclusions may generalize. Training contracts are used in many jobs, both other blue-collar jobs (e.g. mechanics, electricians, and metalworkers) and high-skill jobs (e.g. pilots, accountants, and securities brokers). Future work should examine whether the impacts of training contracts in these jobs appear to be similar. Is worker overconfidence relevant for firm training in other industries? The piece rate compensation system in trucking makes overconfidence easy to measure and consequential for expected earnings. Even in jobs without piece rates, though, if the jobs people choose are the ones where they are most overconfident, differential overconfidence will also be present, thereby making workers less likely to quit. While I focus on firm training, overconfidence may also be relevant for other types of human capital investment, e.g. going to college, or for occupational choice in general.[80]

To reach the above conclusions, a number of simplifying assumptions were made. I focused only on the quitting decision of workers without modeling the firing decision of firms.[81] In addition, I study optimal contract choice by one firm instead of considering firms offering competing training contracts. Finally, I analyze optimal contract design when the firm chooses a single contract instead of allowing the firm to offer a menu of training contracts.

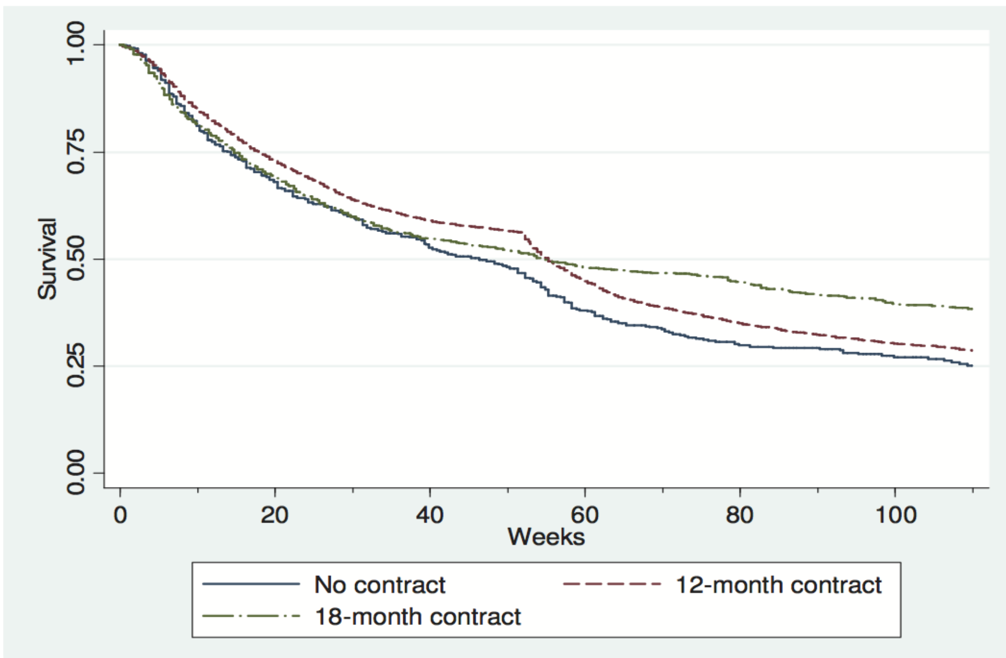
Worker overconfidence and learning may be important for many aspects of optimal job design and compensation. For example, when firms can choose to pay flat wages or piece

[80] See Stinebrickner and Stinebrickner (2011) for evidence that college students are initially overconfident about their likely performance in college. There has been recent discussion, particularly related to law schools, that students may be overly optimistic about their future job prospects when taking on student loans, e.g. Segal (2011) and Goodwin (2011).

[81] Although firing is about five times less common than quitting in the data, future work may wish to consider the game played between workers and firms where workers may quit and firms may fire.

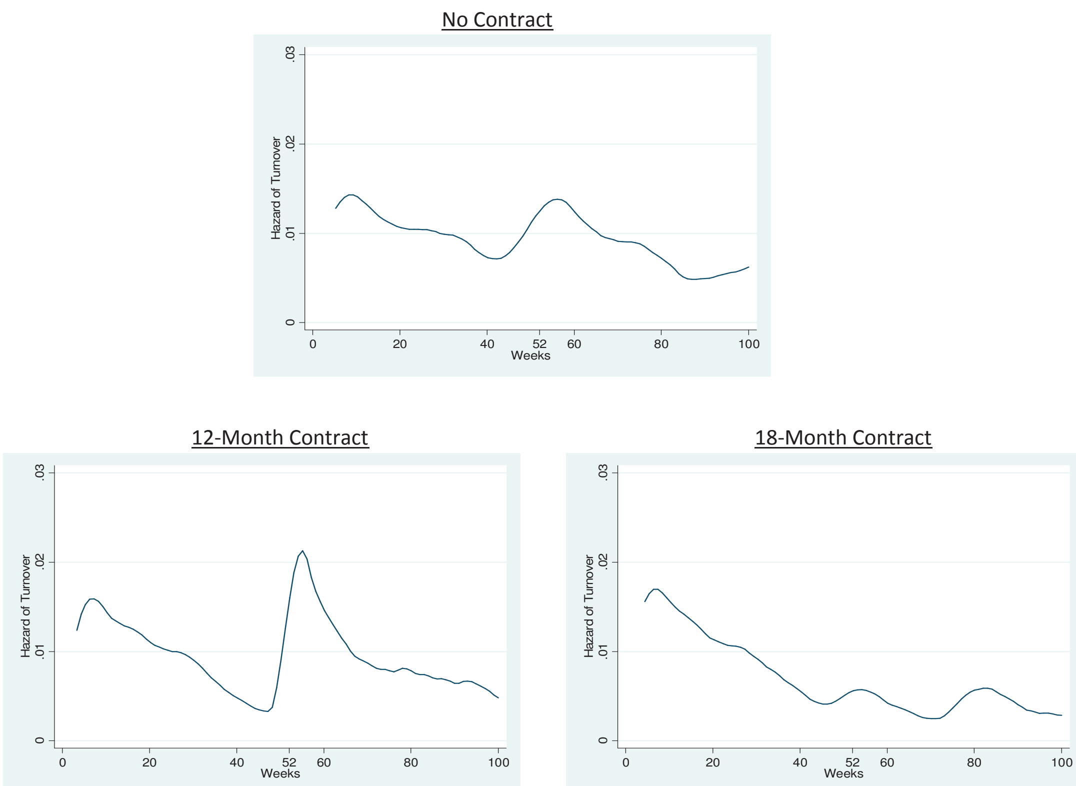
rates, paying a piece rate may be appealing if workers are overconfident since overconfident workers perceive they may earn more than they actually will (Larkin and Leider, 2011). Future work should analyze the equilibrium interplay between worker behavior and firms' contractual choices, both for training contracts and other types of contracts.

Figure 1.1: Effect of Training Contracts on Driver Retention



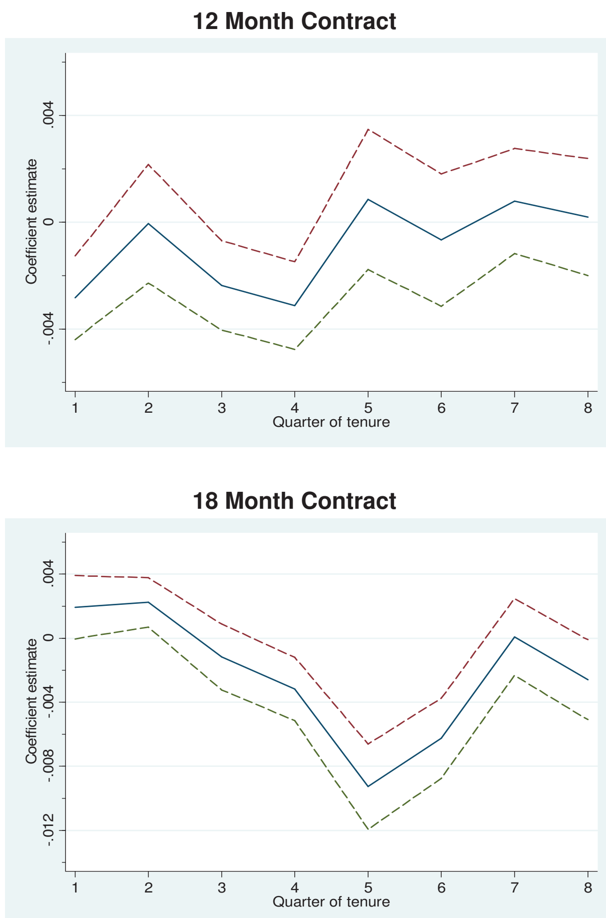
Notes: This graph plots the share of drivers surviving at any week under the 3 contractual regimes, using new system drivers starting at Firm A between 2001 and 2009. The figure is a Kaplan-Meier survival curve, and focuses only on exits by quitting (fires are treated as censored). Wilcoxon test for equality of three curves: $p < 0.01$. Wilcoxon test for equality of No contract and 12m contract regimes: $p < 0.01$. Wilcoxon test for equality of 12m contract and 18m contract regimes: $p < 0.01$.

Figure 1.2: Training Contracts and the Hazard of Quitting



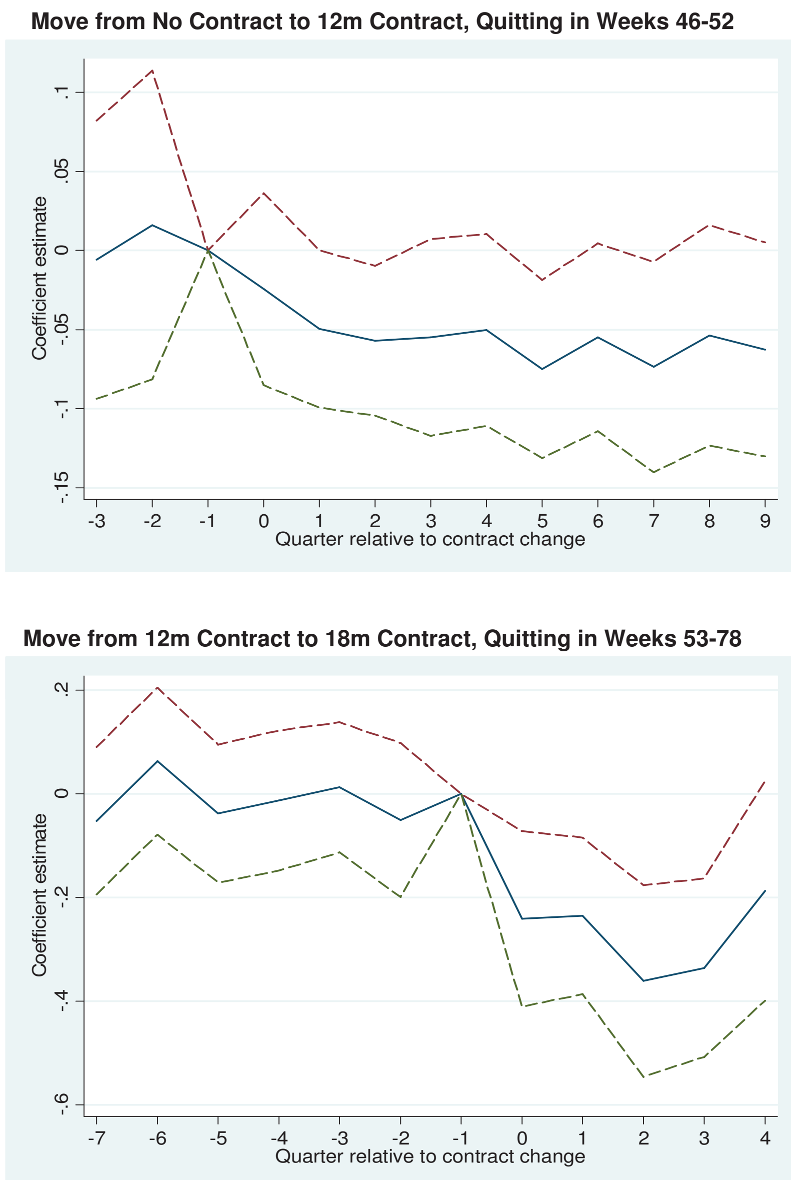
Notes: The sample Size is $0.09N$ with no contract, $0.73N$ with 12m contract, $0.18N$ with 18m contract, $N >> 5,000$. These figures plot the quitting hazard at any week under the 3 contractual regimes, using all drivers starting at Firm A between 2001 and 2009. The top figure is for the no contract regime, the bottom left is for the 12-month contract regime, and the bottom right is for the 18-month contract regime. An Epanechnikov kernel is used. The bandwidth is 4 weeks for the no contract regime, 2 weeks for the 12-month contract regime, and 3 weeks for the 18 month contract regime. The number of drivers under each of the 3 regimes is listed as the share of all new drivers (N). The exact N is withheld to protect the confidentiality of the firm.

Figure 1.3: The Impact of Training Contracts on Quitting by Quarter of Tenure (with 95% Confidence Intervals)



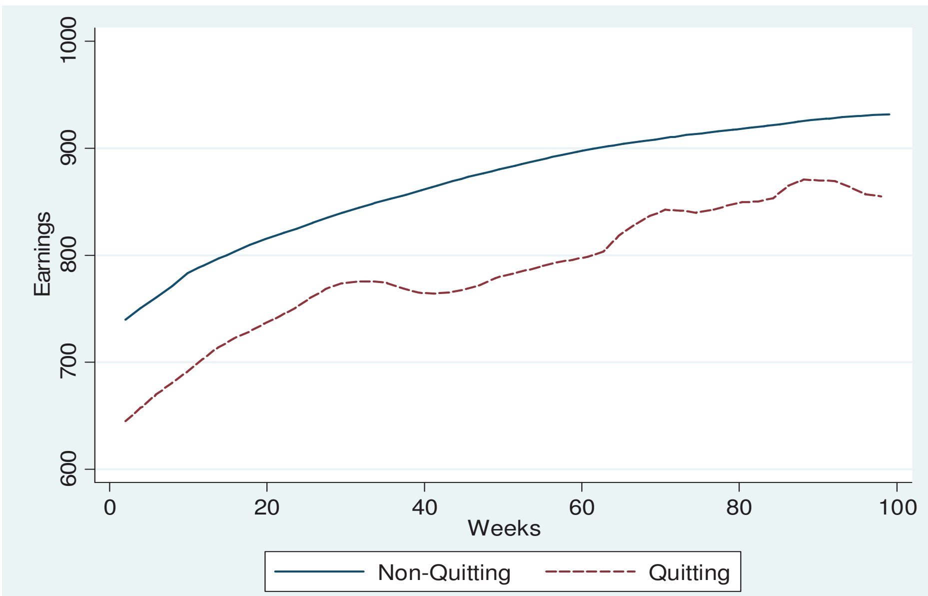
Notes: This figure plots the estimated effect of the two training contracts on quitting at different tenure levels. The solid line denotes the coefficient estimate, with the dotted lines denoting the 95% confidence interval. The coefficients are from OLS regressions of quitting (0 or 1) for a driver in a given week on contract-quarter of tenure interactions and controls, as in Equation (1.2). Controls include tenure week dummies, year dummies, year of hire dummies, school dummies, annual state unemployment rate, and demographic controls (gender, race , marital status, and driver age). Standard errors are clustered at the school-week of hire level. Under the 12-month contract, quitting is significantly lower (relative to no contract) in the 4th quarter (weeks 39-52), but is somewhat higher (relative to no contract) in the 5th quarter (weeks 53-65). Under the 18-month contract, quitting is significantly lower (relative to no contract) in quarters 4-6, but then increases after that.

Figure 1.4: Event Studies: The Impact of Training Contracts on Quitting, Comparing Before and After the Contract Changes (with 95% Confidence Intervals)



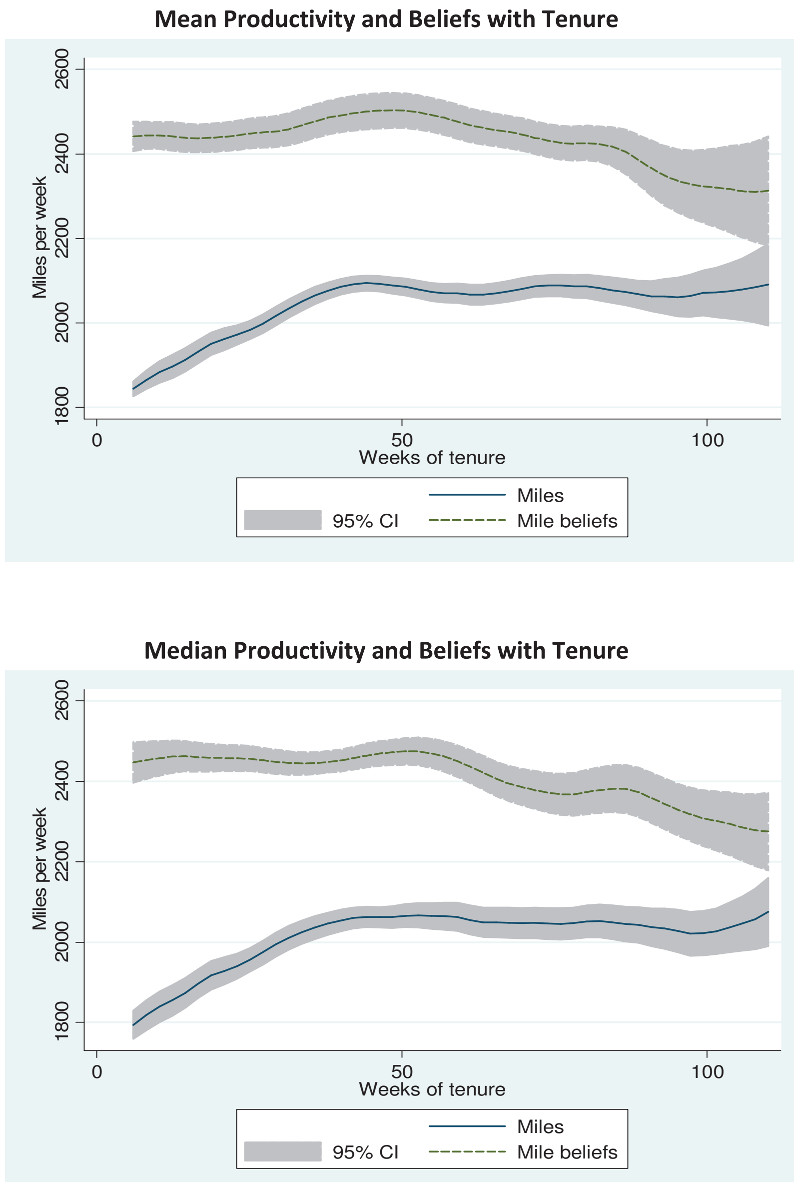
Notes: This figure provides event study estimates of the impacts of the 12-month and 18-month contracts on quitting at the different training schools. The solid line denotes the coefficient estimate, with the dotted lines denoting the 95% confidence interval. The top figure analyzes quitting in weeks 46-52 before and after the change to the 12-month contract whereas the bottom figure analyzes quitting in weeks 53-78 before and after the change to the 18-month contract. The x-axis denotes "event time," reflecting the contracts being changed at different schools at different times. Each "quarter" refers to the workers hired in a 3-month block. Quarter 0 is the first quarter after the introduction of each training contract. Controls include school fixed effects, year of hire fixed effects, and demographic controls (gender, race , marital status, and driver age), as in Equation (1.3). Standard errors are clustered at the school-week of hire level.

Figure 1.5: Quitting and Selection on Productivity: Average Earnings per Week of Quitting and Non-Quitting Drivers



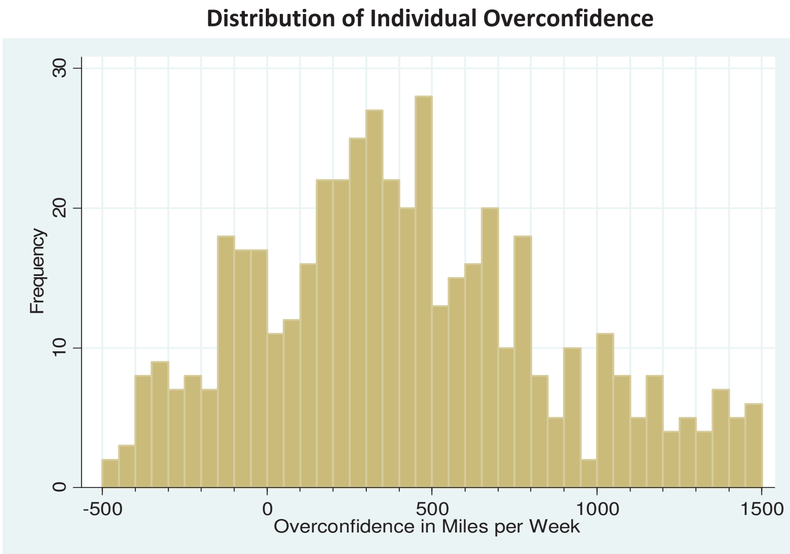
Notes: This figure analyzes the average earnings to date in each week of all new drivers from 2001 to 2009 at Firm A. In each week, it compares the earnings of drivers who quit with those of drivers who do not quit that week. For example, it can be used to compare the average earnings per week (from weeks 1-19) of drivers who quit in the 20th week with the average earnings per week (from weeks 1-19) of drivers who survive to the 20th week, but do not quit. The lines are local polynomials plotted with Stata's lpoly command. Zero mile weeks are excluded.

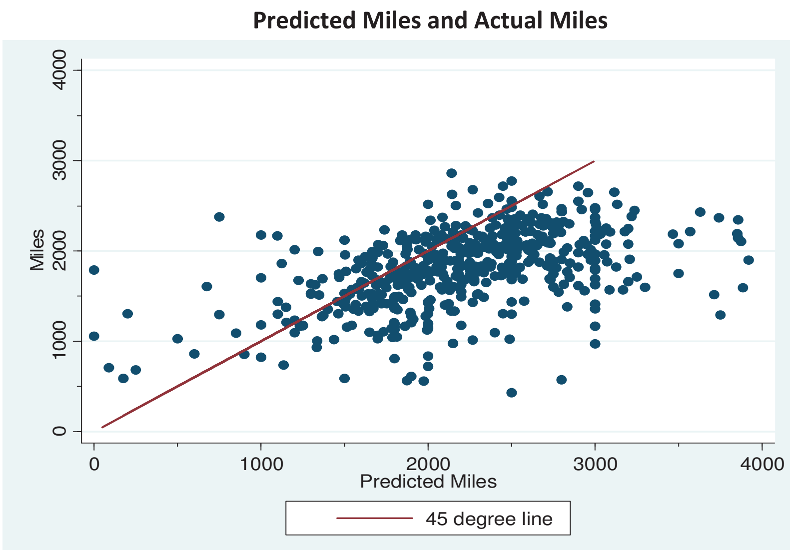
Figure 1.6: Overconfidence: Comparing Subjective Productivity Forecasts with Actual Worker Productivity (as a Function of Worker Tenure)



Notes: This figure analyzes actual and believed productivity for 699 drivers in the data subset. The figures are plotted using a local polynomial regression with an Epanechnikov kernel. A bandwidth of 5 weeks is used for the productivity data and a bandwidth of 7 weeks is used for the belief data. In the top figure, the productivity and belief data are collapsed into weekly means before local polynomial smoothing. In the bottom figure, the productivity and belief data are collapsed into weekly medians before local polynomial smoothing.

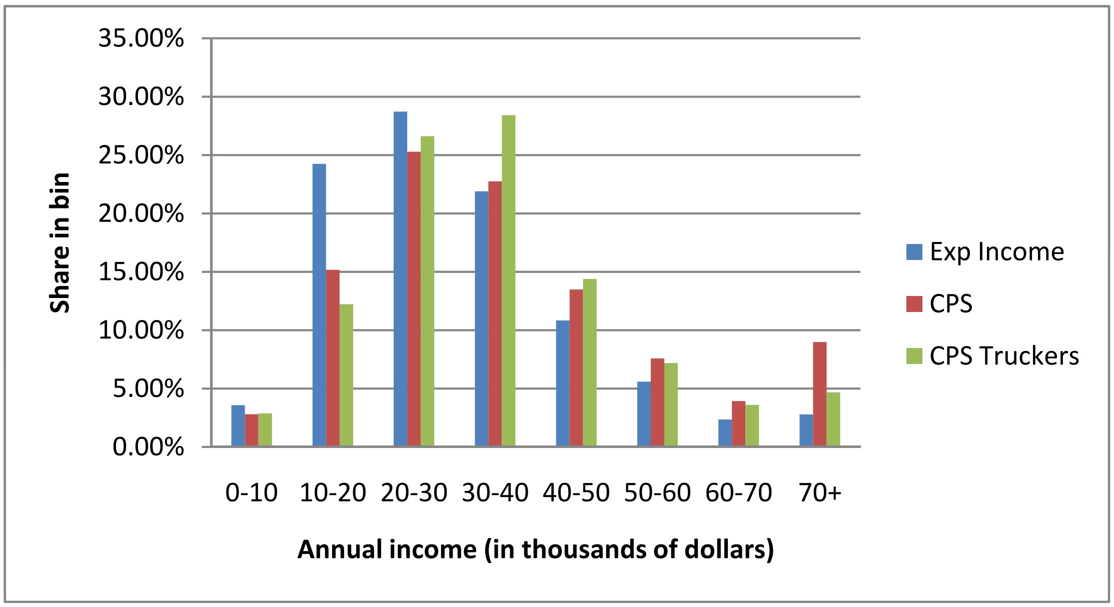
Figure 1.7: Distribution of Overconfidence Across Drivers





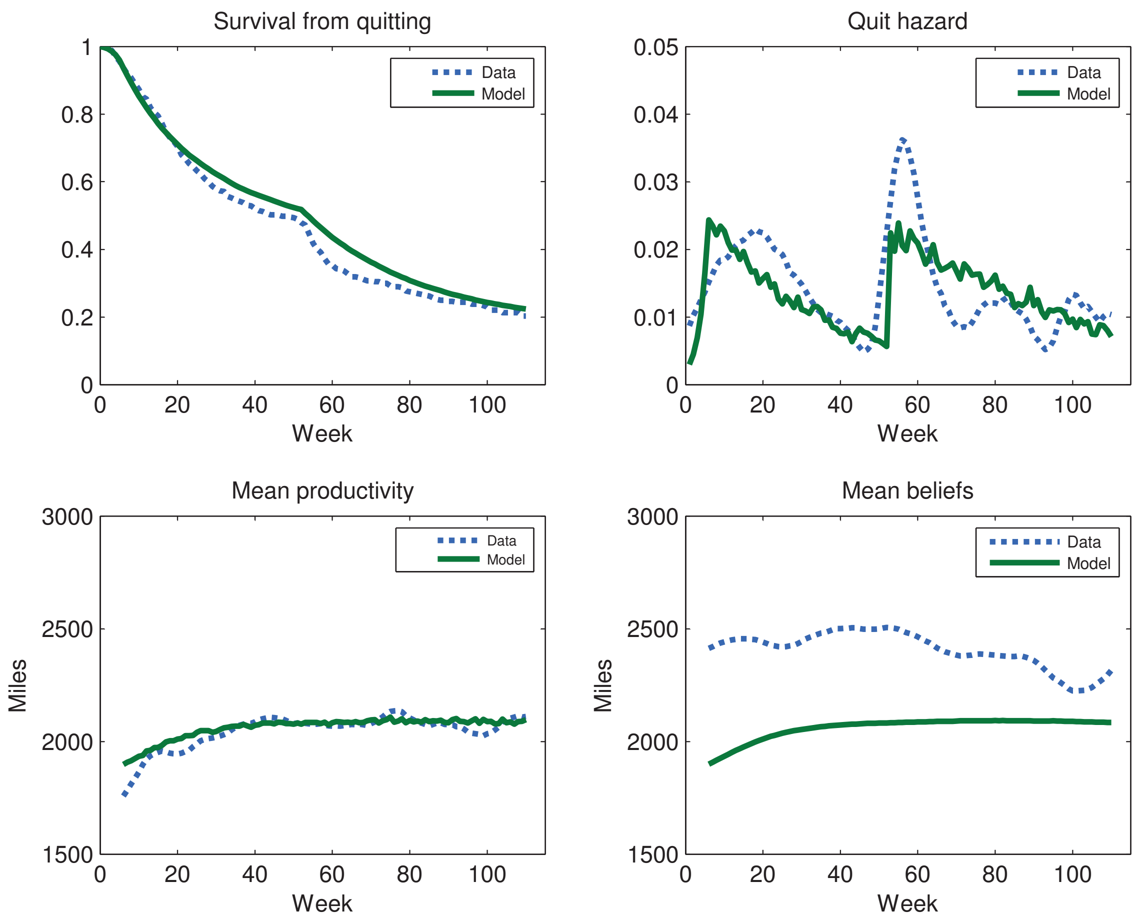
Notes: This figure analyzes overconfidence among 699 drivers in the data subset. It presents reduced-form evidence on the distribution of overconfidence across drivers. The top figure plots a histogram of driver-level overconfidence, where overconfidence is defined as the difference between beliefs and productivity. A driver's overconfidence level is calculated by averaging over all the weeks with productivity beliefs and actual productivity. In the bottom figure, each driver is represented by a dot located at their average productivity and beliefs (averaged over all weeks the driver is observed).

Figure 1.8: Are Workers Overconfident About their Outside Option? A Comparison of Firm A Workers' Believed Outside Option with Earnings of Similar Workers in the CPS



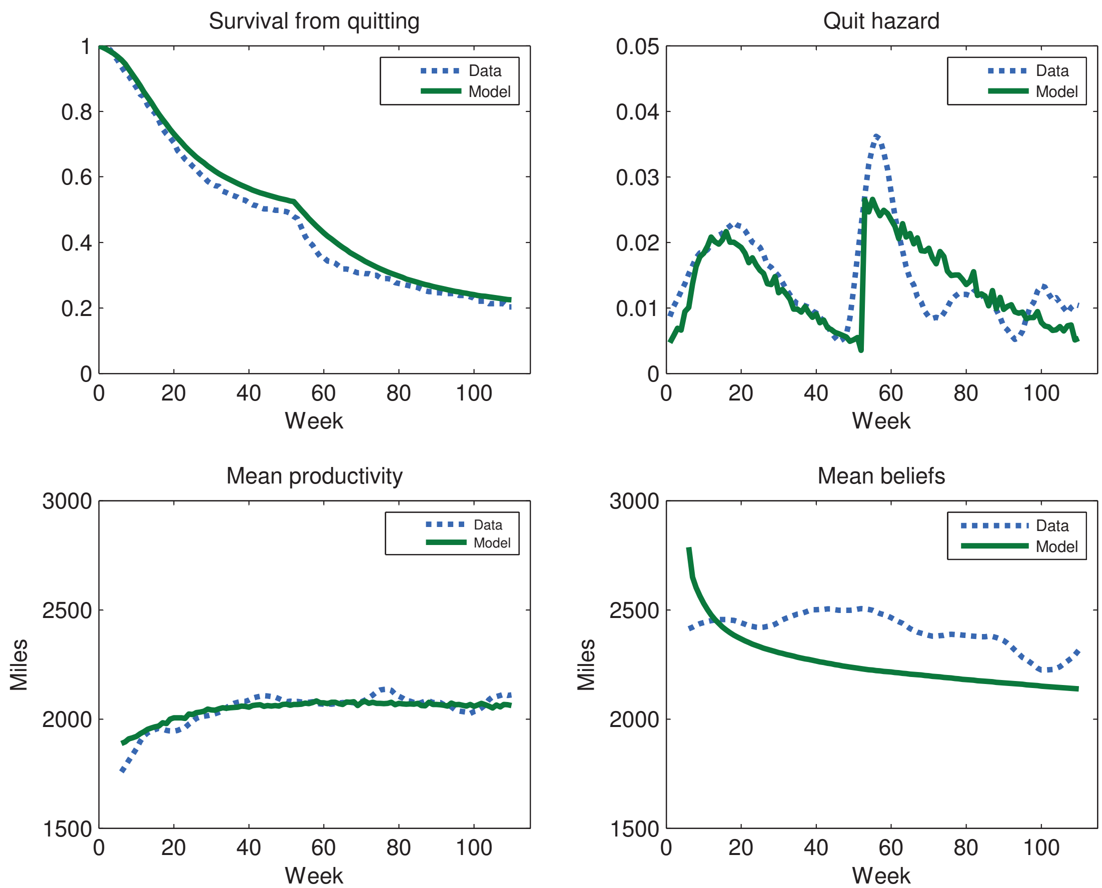
Notes: This figure analyzes worker beliefs about their outside option. During driver training, workers at Firm A were asked "Which range best describes the annual earnings you would normally have expected from your usual jobs (regular and part-time together), if you had not started driver training with [Firm A], and your usual jobs had continued without interruption?" Answers were given in eight intervals: $\$0 - \$10,000$, $\$10,000 - \$20,000$, $\$20,000 - \$30,000$, $\$30,000 - \$40,000$, $\$40,000 - \$50,000$, $\$50,000 - \$60,000$, $\$60,000 - \$70,000$, $\$70,000+$. The CPS comparison data is from the 2006 March CPS. "CPS" is the average income and earnings for 35-year old male workers with a high school degree who worked full-time last year. "CPS Truckers" is the average income and earnings for 30-40 year old male workers with a high school degree who work as truckdrivers (Occ=913).

Figure 1.9: Model Fit, Model Estimated Without Belief Data



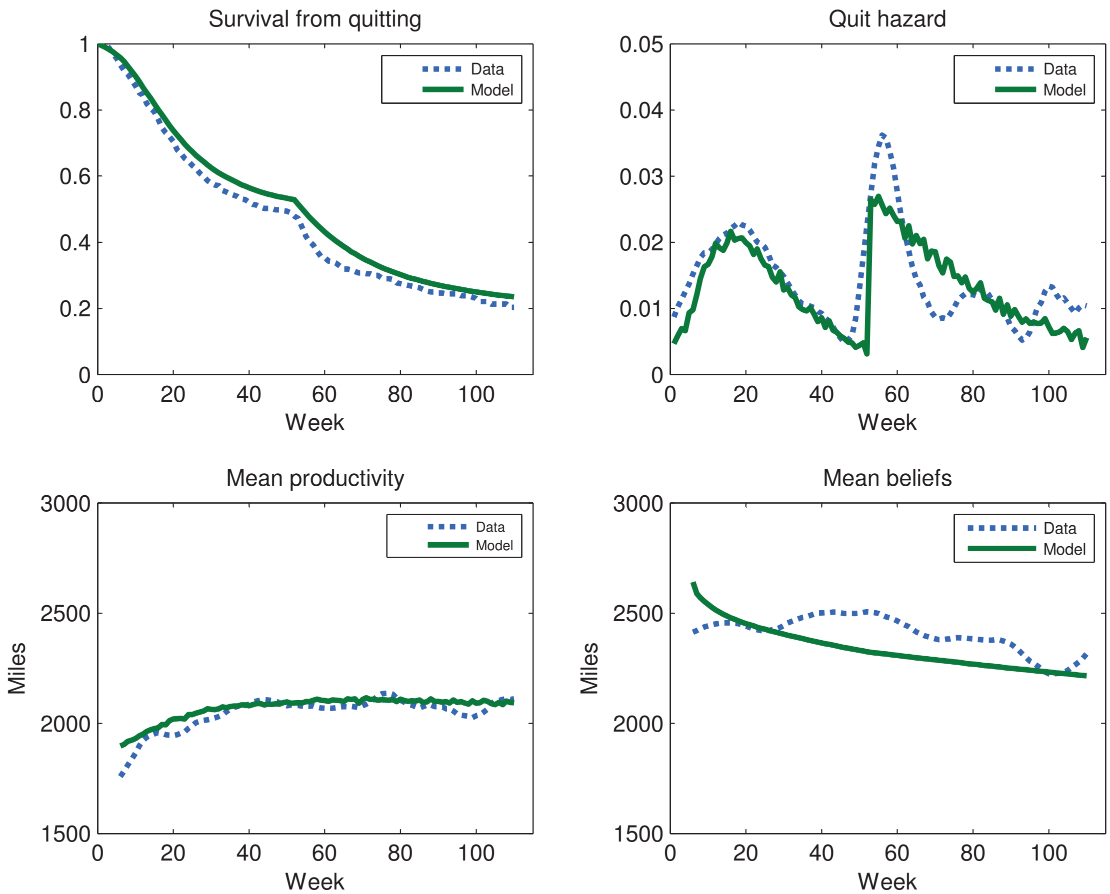
Note: These figures plot the actual hazard rate of quitting against a hazard created from 40,000 simulated drivers. The paths of quits, earnings, and beliefs over tenure are plotted using an Epanechnikov kernel. The bandwidths are 6 weeks, 5 weeks, and 10 weeks for quits, earnings and beliefs, respectively. The data are from 699 drivers under the 12-month contract. A weekly discount factor of 0.9957 is assumed for workers and firms, corresponding to an annual discount factor of 0.8. The model is estimated with learning by doing and allowing for an increasing outside option.

Figure 1.10: Model Fit, Model Estimated With Overconfidence and Standard Learning



Note: These figures plot the actual hazard rate of quitting against a hazard created from 40,000 simulated drivers. The paths of quits, earnings, and beliefs over tenure are plotted using an Epanechnikov kernel. The bandwidths are 6 weeks, 5 weeks, and 10 weeks for quits, earnings and beliefs, respectively. The data are from 699 drivers under the 12-month contract. A weekly discount factor of 0.9957 is assumed for workers and firms, corresponding to an annual discount factor of 0.8. The model is estimated with learning by doing and allowing for an increasing outside option.

Figure 1.11: Model Fit for Full Model: Overconfidence and Generalized Learning



Note: These figures plot the actual hazard rate of quitting against a hazard created from 40,000 simulated drivers. The paths of quits, earnings, and beliefs over tenure are plotted using an Epanechnikov kernel. The bandwidths are 6 weeks, 5 weeks, and 10 weeks for quits, earnings and beliefs, respectively. The data are from 699 drivers under the 12-month contract. A weekly discount factor of 0.9957 is assumed for workers and firms, corresponding to an annual discount factor of 0.8. The model is estimated with learning by doing and allowing for an increasing outside option.

Table 1.1: Summary Statistics

Panel A: All New Drivers at Firm A

Variable	**No Contract**	**12-Month Contract**	**18-month Contract**
African-American	0.19	0.19	0.19
Hispanic	0.07	0.04	0.03
Female	0.08	0.08	0.08
Married	0.35	0.35	0.38
Age	37.27	37.00	37.25
Online Application	0.47	0.55	0.70
Smoker	0.30	0.42	0.40
Drivers	$0.09N$	$0.73N$	$0.18N$

Panel B: Drivers in Data Subset

Variable	**Obs**	**Mean**	**Min**	**Max**
African-American	895	0.11	0	1
Hispanic	895	0.02	0	1
Female	895	0.10	0	1
Married	895	0.41	0	1
Age	894	36.46	21.06	69.21
Number of Kids	895	0.96	0	7
Online Application	889	0.67	0	1
Smoker	787	0.46	0	1
Years of Schooling	895	12.85	9	18
High School Dropout	895	0.04	0	1
High School Graduate	895	0.40	0	1
Some College	895	0.34	0	1
Technical School	895	0.14	0	1
College Degree or More	895	0.08	0	1
Credit Score	784	585.96	407	813
No Credit Score	895	0.12	0	1

Notes: Panel A provides summary statistics for all trained drivers at Firm A from 2001 to 2009. Panel B restricts to drivers in the data subset. Data subset drivers and were hired in late 2005 or 2006. The number of drivers under each of the three contractual regimes is listed as the share of all new drivers (N). The exact N is withheld to protect the confidentiality of the firm.

Table 1.2: Impact of the Training Contracts on Quitting – Cox Model, Diff-in-Diff

	(1)	(2)	(3)	(4)	(5)	(6)
12m contract	-0.179***	-0.173***	-0.208***	-0.202***	-0.182***	
	(0.046)	(0.046)	(0.052)	(0.055)	(0.058)	
18m contract	-0.108*	-0.118*	-0.116*	-0.127*	-0.107	
	(0.062)	(0.062)	(0.069)	(0.071)	(0.074)	
State unemployment rate		-0.055***	-0.048***	-0.062***	-0.049***	-0.054***
		(0.011)	(0.013)	(0.013)	(0.014)	(0.014)
Avg miles to date				-0.060***	-0.048***	-0.049***
				(0.003)	(0.004)	(0.004)
12m contract * (wks$\leq$52)						-0.371***
						(0.074)
12m contract * (52$<$wks$\leq$78)						0.066
						(0.081)
12m contract * (wks$>$78)						0.043
						(0.090)
18m contract * (wks$\leq$52)						-0.021
						(0.087)
18m contract * (52$<$wks$\leq$78)						-0.891***
						(0.112)
18m contract * (wks$>$78)						-0.100
						(0.132)
Time FE (yr)	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE (yr of hire)	Yes	Yes	Yes	Yes	Yes	Yes
Training School FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	No	No	Yes	Yes
Observations	M	M	$0.89M$	$0.89M$	$0.79M$	$0.79M$

Notes: An observation is a driver-week. The regressions are Cox proportional hazard models, with standard errors clustered by school-week of hire in parentheses. Driver tenure is controlled for non-parametrically. State unemployment is a state's annual unemployment rate. Demographic controls include gender, race dummies, marital status, and driver age. Average miles to date is a driver's average weekly productivity to date and is given in terms of hundreds of miles driven per week. Column (3) differs from column (2) differs in that it restricts the sample to be the same as in column (4). The exact M is withheld to protect firm confidentiality, $M >> 100,000$. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.3: Selection Effects of Training Contracts

Panel A: Selection on Productivity

	All Weeks		Exclude 0 Mile Weeks		Trim 5/95%	
	(1)	(2)	(3)	(4)	(5)	(6)
12m contract	0.17	-7.01	24.69	21.1	23.74	22.01
	(28.59)	(28.55)	(23.95)	(23.63)	(18.68)	(18.58)
18m contract	-21.56	-25.44	11.78	8.31	13.62	10.37
	(31.01)	(31.29)	(28.72)	(28.60)	(22.06)	(22.12)
Demog controls	No	Yes	No	Yes	No	Yes
Mean Dep. Var:	1661	1690	1963	1956	1951	1947
Observations	0.99M	0.86M	0.84M	0.74M	0.75M	0.67M
R-squared	0.24	0.22	0.05	0.05	0.04	0.04

Panel B: Selection on Other Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var:	Black	Hispanic	Female	Married	Age	Smoker	Online Ap
12m contract	-0.009	0.021**	0.012	-0.016	0.111	0.050*	0.055**
	(0.014)	(0.010)	(0.010)	(0.017)	(0.399)	(0.027)	(0.022)
18m contract	0.006	0.014	0.023	-0.005	0.574	0.062*	0.047*
	(0.020)	(0.013)	(0.015)	(0.024)	(0.541)	(0.033)	(0.028)
Observations	N	N	N	N	0.72N	0.70N	0.82N
R-squared	0.03	0.03	0.01	0.01	0.01	0.02	0.07

Notes: Panel A reports OLS regressions of productivity (miles per week) on training contract dummies and controls. In Panel A, an observation is a driver-week. "Trim 5/95%" refers to trimming the lowest 5% and highest 5% of the miles observations (ignoring all 0 mile weeks). Panel B reports OLS regressions of driver characteristics on training contract dummies and controls. In Panel B, an observation is a driver. Standard errors clustered by school-week of hire in parentheses. The exact M (driver weeks) and N (drivers) are withheld to protect firm confidentiality, $M >> 100,000$. All regressions include time fixed effects (for each month), cohort fixed effects (by year of hire), tenure fixed effects (by week), work type controls, and the annual state unemployment rate. Demographic controls include gender, race dummies, marital status, and driver age.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.4: Do Productivity Beliefs Predict Productivity? OLS Regressions

	(1)	(2)	(3)	(4)	(5)
L. Pred miles	0.195***	0.066***	0.064***	0.079***	0.086***
	(0.023)	(0.016)	(0.016)	(0.022)	(0.023)
L. Avg miles to date		0.789***	0.689***		-0.188*
		(0.037)	(0.038)		(0.108)
Tenure FE	No	Yes	Yes	Yes	Yes
Demographic Controls	No	No	Yes	No	No
Education Controls	No	No	Yes	No	No
Work Type Controls	No	No	Yes	No	No
Subject FE	No	No	No	Yes	Yes
Observations	8,449	8,435	8,435	8,449	8,435
R-squared	0.05	0.17	0.18	0.29	0.29

Notes: The dependent variable is miles driven per week. An observation is a driver-week. Standard errors clustered by driver in parentheses. Demographic controls include gender, race dummies, marital status, and age bin dummies for the different age groups: 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, and 60-80. Education controls are dummies for high school graduate, some college, and college. Work type controls are dummies for different work configurations and for receiving any salary or activity-based pay. Productivity is given in terms of hundreds of miles driven per week. All drivers are from the same training school and were hired in late 2005 or 2006. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.5: Do Productivity Beliefs Predict Quitting?

	(1)	(2)	(3)	(4)	(5)
Predicted miles	-0.059***			-0.059***	-0.067***
	(0.018)			(0.020)	(0.021)
Avg miles to date		-0.079***	-0.112***	-0.002	-0.062
		(0.010)	(0.039)	(0.036)	(0.042)
Demographic Controls	No	Yes	Yes	No	Yes
Education Controls	No	Yes	Yes	No	Yes
Work Type Controls	No	Yes	Yes	No	Yes
Observations	8,500	38,381	8,500	8,500	8,500

Notes: An observation is a driver-week. The regressions are Cox proportional hazard models, where the dependent variable is quitting. Events where the driver is fired are treated as censored. Standard errors clustered by worker are in parentheses. Demographic controls include gender, race dummies, marital status, and age bin dummies for the different age groups: 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, and 60-80. Education controls are dummies for high school graduate, some college, and college. Work type controls are dummies for different work configurations and for receiving any salary or activity-based pay. Productivity is given in terms of hundreds of miles driven per week. All drivers are from the same training school and were hired in late 2005 or 2006. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.6: Baseline Structural Estimates

		No Bias	**Belief Bias**	**Belief Bias 2 Types**
τ	Scale param of idiosyncratic shock	1618	2206	3726
		(136)	(291)	(449)
	Productivity Parameters			
η_0	Mean of prior productivity dist	2464	2025	2024
		(9)	(17)	(18)
σ_0	Std dev of prior productivity dist	475	286	284
		(16.5)	(10.1)	(9.9)
σ_y	Std dev of productivity shocks	707	706	706
		(1.5)	(1.5)	(1.6)
	Skill Gain Parameter			
s_0	Value of skills gained in wks 1-5	14.9	8.6	31.7
		(3.5)	(5.9)	(9.0)
	Taste UH Parameters			
μ_1	Mass point 1 of taste UH	-248	-259	-736
		(9.0)	(12.6)	(41.9)
μ_2	Mass point 2 of taste UH	-106	-135	-150
		(14.6)	(12.1)	(10.0)
μ_3	Mass point 3 of taste UH	139	135	191
		(39.3)	(33.3)	(63.1)
p_1	Probability of type 1, taste	0.55	0.34	0.12
		(0.04)	(0.06)	(0.03)
p_2	Probability of type 2, taste	0.24	0.43	0.67
		(0.03)	(0.06)	(0.04)
	Belief Parameters			
σ_b	Std dev in belefs	299	298	271
		(0.3)	(0.3)	(0.3)
$\widetilde{\sigma_y}$	Believed std dev of productivity shocks	3650	1888	2068
		(134)	(82)	(81)
η_b	Belief bias		589	
			(22)	
$\eta_{1,b}$	Mass point 1 of belief UH			426
				(20)
$\eta_{2,b}$	Mass point 2 of belief UH			3649
				(41)
$p_{1,b}$	Probability of type 1, beliefs			0.94
				(0.16)
	Log-likelihood	-91064	-90865	-89882
	Number of workers	699	699	699

Notes: This table presents estimates of the structural parameters. The idiosyncratic shock, skill gain, and taste parameters are given in terms of dollars whereas the productivity and belief parameters are given in terms of miles. Standard errors are in parentheses and are calculated by inverting the Hessian. All specifications assume a normal learning model. A weekly discount factor of 0.9957 is assumed for workers and firms, corresponding to an annual discount factor of 0.8. The data are from 699 drivers in the data subset, all of whom face the 12-month training contract.

Table 1.7: Structural Estimates with Learning by Doing and Skill Accumulation

		No Bias	**Belief Bias**	**Belief Bias 2 Types**
τ	Scale param of idiosyncratic shock	1891	1605	4207
		(260)	(180)	(925)
	Productivity and Skill Parameters			
η_0	Mean of prior productivity dist	2307	1595	1742
		(28)	(41)	(33)
σ_0	Std dev of prior productivity dist	521	275	276
		(20.8)	(10.5)	(10.5)
σ_y	Std dev of productivity shocks	706	706	705
		(3.6)	(3.6)	(3.6)
b_1	Learning by doing level	251	485	343
		(33)	(38)	(34)
b_2	Learning by doing speed	0.08	0.11	0.07
		(0.01)	(0.02)	(0.01)
θ_1	Skill gain level	92	228	534
		(53)	(54)	(140)
θ_2	Skill gain speed	0.68	0.03	0.05
		(0.30)	(0.01)	(0.01)
	Taste UH Parameters			
μ_1	Mass point 1 of taste UH	-195	-345	-988
		(56.7)	(19.7)	(211.6)
μ_2	Mass point 2 of taste UH	-61	-81	27
		(58.1)	(30.0)	(58.5)
μ_3	Mass point 3 of taste UH	180	231	634
		(73)	(55)	(170)
p_1	Probability type 1	0.55	0.60	0.42
		(0.07)	(0.04)	(0.06)
p_2	Probability of type 2	0.25	0.19	0.35
		(0.06)	(0.04)	(0.06)
	Belief Parameters			
σ_b	Std dev in belefs	298	297	271
		(1.4)	(1.4)	(1.3)
$\widetilde{\sigma_y}$	Believed std dev of productivity shocks	3481	1295	1759
		(167)	(82)	(108)
η_b	Belief bias		722	
			(31)	
$\eta_{1,b}$	Mass Point 1 of belief UH			466
				(25)
$\eta_{2,b}$	Mass Point 2 of belief UH			3899
				(139)
p_1	Prob of Type 1			0.93
				(0.01)
	Log-likelihood	-91024	-90731	-89787
	Number of workers	699	699	699

Notes: This table presents estimates of the structural parameters. The idiosyncratic shock, skill gain, and taste parameters are given in terms of dollars whereas the productivity and belief parameters are given in terms of miles. Standard errors are in parentheses and are calculated by inverting the Hessian. All specifications assume a normal learning model. A weekly discount factor of 0.9957 is assumed for workers and firms, corresponding to an annual discount factor of 0.8. The data are from 699 drivers in the data subset, all of whom face the 12-month training contract.

Table 1.8: Profits and Welfare Under Different Contracts

	No contract	12 month	18 month
Profits per worker	\$363	\$1,625	\$1,875
Profits per truck	-\$2,544	\$1,856	\$2,641
Welfare per worker	\$159,580	\$156,590	\$156,375

Notes: This table presents profits per worker and profits per worker under the three different training contracts used by Firm A. Profits per worker and profits per firm are defined in Section 1.9 of the text. Profits are calculated assuming a Fixed Cost of \$600 per week, a price of \$1.80 per mile, a non-wage marginal cost of \$1.16 per mile, a sunk cost of \$2.50 per worker per week, and a marginal cost of training of \$2,500, and a collection rate of 30%. Profits and welfare are calculated by simulating 40,000 workers under each of the three regimes. A weekly discount factor of 0.9957 is assumed for workers and firms, corresponding to an annual discount factor of 0.8. The model is estimated with 3 taste mass points and 2 overconfidence mass points. The model simulated has no learning by doing and assumes a flat outside option after the first 5 weeks.

Table 1.9: Counterfactual Simulations, No Contractual Response

	Baseline	50% debias	100% debias	Reveal ability
Retention at 20 wks	0.74	0.62	0.49	0.45
Retention at 40 wks	0.55	0.44	0.32	0.32
Retention at 60 wks	0.43	0.35	0.26	0.26
Welfare per worker	$156,590	$157,771	$158,708	$158,959
Profits per worker	$1,625	$793	-$83	$562
Profits per truck	$1,856	-$1,159	-$5,225	-$2,650
Ability at 20 wks	2,037	2,034	2,036	2,085
Ability at 40 wks	2,058	2,055	2,053	2,115
Ability at 60 wks	2,072	2,062	2,057	2,114

Notes: This table reports the reports the results of the counterfactual simulations described in the text, while assuming that training contracts are not adjusted in response. Under the 50% debias and 100% debias counterfactuals, worker overconfidence is reduced by 50% or 100% (by reducing the overconfidence mass points η_{b1} and η_{b2} by 50% or 100%). Under the reveal ability counterfactual, the worker's ability is revealed to the worker after training. Profits are calculated assuming a Fixed Cost of $600 per week, a price of $1.80 per mile, a non-wage marginal cost of $1.16 per mile, a sunk cost of $2.50 per worker per week, and a marginal cost of training of $2,500, and a collection rate of 30%. A weekly discount factor of 0.9957 is assumed for workers and firms, corresponding to an annual discount factor of 0.8. The model is estimated with 3 taste mass points and 2 overconfidence mass points. The model simulated has no learning by doing and assumes a flat outside option after the first 5 weeks.

Table 1.10: Counterfactual Simulations Allowing for Optimal Contractual Responses

	Baseline	50% debias	100% debias	Reveal ability	Ban training contracts	Ban training contracts + 50% debias
Optimal 12m penalty	\$11,400	\$8,600	\$0	\$0	NA	NA
Wage factor	1.022	1.026	0.918	0.909	0.983	0.965
Retention at 20 wks	0.86	0.76	0.38	0.36	0.64	0.51
Retention at 40 wks	0.80	0.65	0.23	0.22	0.41	0.33
Retention at 60 wks	0.66	0.52	0.20	0.18	0.31	0.26
Welfare per worker	\$153,622	\$155,653	\$159,177	\$159,265	\$159,247	\$159,852
Profits per worker	\$2,070	\$1,311	\$705	\$1,278	\$708	\$334
Profits per truck	\$3,129	\$969	-\$2,904	-\$317	-\$1,508	-\$3,566
Ability at 20 wks	2,025	2,032	2,033	2,084	2,035	2,031
Ability at 40 wks	2,032	2,049	2,041	2,115	2,056	2,047
Ability at 60 wks	2,052	2,066	2,043	2,109	2,057	2,049

Notes: This table reports the reports the results of the counterfactual simulations described in the text, while assuming that training contracts are optimally adjusted in response. Under the 50% debias and 100% debias counterfactuals, worker overconfidence is reduced by 50% or 100% (by reducing the overconfidence mass points η_{b1} and η_{b2} by 50% or 100%). Under the reveal ability counterfactual, the worker's ability is revealed to the worker after training. Under the ban training contracts counterfactual, firms are forbidden from charging a quitting penalty. Profits are calculated assuming a Fixed Cost of \$600 per week, a price of \$1.80 per mile, a non-wage marginal cost of \$1.16 per mile, a sunk cost of \$2.50 per worker per week, and a marginal cost of training of \$2,500, and a collection rate of 30%. A weekly discount factor of 0.9957 is assumed for workers and firms, corresponding to an annual discount factor of 0.8. The model is estimated with 3 taste mass points and 2 overconfidence mass points. The model simulated has no learning by doing and assumes a flat outside option after the first 5 weeks.

1.A Model for Section 1.3.1

In this section, I present a model of training contracts and turnover to accompany the discussion in Section 1.3.1. I show how training contracts can increase training and reduce turnover. In addition, I show how worker overconfidence about ability can also increase training, and how overconfidence and training contracts can be complimentary. The model has one period and abstracts from the dynamics considered in the structural model. The model is fairly similar to that in Peterson (2010), with one difference being that I assume monopsony in the market for training whereas Peterson (2010) assumes perfect competition. Both of us allow for competition in the post-training labor market.

Consider a firm which trains its workers. Workers have an initial productivity of zero at the firm and an outside option of r. A training investment is available at cost c that raises productivity from 0 to η. Workers have some non-pecuniary taste for the job ε, which they learn after training. I assume that ε has a distribution function F and has support over the entire real line. Let $\Psi(x) = 1 - F(x)$, that is, $\Psi(\cdot)$ is the survival function. I assume that the function $x \mapsto x\Psi^{-1}(x)$ is concave. If the firm chooses to train, it also chooses a piece rate w to pay. That is, the worker's total wage will be given by $W = w\eta$. The firm may also employ a training contract k, which is a penalty the worker pays if they quit. If the worker quits, they receive an outside option of $\overline{W}$ minus any training contract k. I assume only that the outside option after training is greater than or equal to the worker's outside option before training $\left(\overline{W} \geq r\right)$. The case of $\overline{W} = r$ corresponds to the worker choosing whether to go to another occupation whereas $\overline{W} = \eta$ may correspond to the worker choosing to leave for another firm within the same occupation (if training is portable across firms, but occupation-specific). I assume that the training contract is experienced fully by the worker, but that only a share $\theta \in [0, 1]$ of the contract is collected.

The timing of the model is as follows:

1. The firm chooses whether to train, and if so, sets w and the level of the training contract.
2. The worker decides whether or not to accept the contract.
3. The taste shock ε is realized, the worker decides whether to quit, and payoffs are realized.

The firm's problem can be written as:

$$\begin{aligned} & \max_{w,k}(1 - F(-w\eta - k + \overline{W})) * (\eta - w\eta) + F(-w\eta - k + \overline{W})\theta k - c \\ r \quad \leq \quad & E \max\left(w\eta + \varepsilon, \overline{W} - k\right) \end{aligned}$$

I first prove the following proposition.

Proposition 1. *Allowing firms to use training contracts increases the probability of training and increases retention compared to when training contracts are not available.*

Proof. Note that the IR constraint must bind at an interior solution. To see why, differentiate the Lagrangean to get

$$f\left(-W - k + \overline{W}\right)(\eta - W - \theta k) - \left(1 - F\left(-W - k + \overline{W}\right)\right) + \lambda \frac{\partial Emax}{\partial W} = 0$$
$$f\left(-W - k + \overline{W}\right)(\eta - W - \theta k) + \theta F\left(-W - k + \overline{W}\right) + \lambda \frac{\partial Emax}{\partial k} = 0$$

If $\lambda = 0$, an interior solution exists only if $\left(1 - F\left(-W - k + \overline{W}\right)\right) = -\theta F\left(-W - k + \overline{W}\right)$, which is not possible for $\theta > 0$.[82] Because it is not possible to satisfy the IR constraint while setting $k = 0$ (since $\overline{W} \geq r$

[82] The boundary solutions for the unconstrained problem are to have W to go minus infinity and k go to

and ε has full support over the real line), $k = 0$ cannot be optimal.[83]

To analyze retention, let P denote the retention probability.[84] Note then that $W = \left(\bar{W} - k - \Psi^{-1}(P)\right)$. Profits are then given by:

$$P \times (\eta - W) + (1-P)\theta k - c = P \times \left(\eta - \left(\bar{W} - k - \Psi^{-1}(P)\right)\right) + (1-P)\theta k - c \qquad (1.14)$$
$$= P(\eta - \bar{W}) + k(P + (1-P)\theta) + P\Psi^{-1}(P).$$

Note that the IR constraint can be written as $r \leq \mathbb{E}max\left\{\bar{W} - \Psi^{-1}(P) - k + \varepsilon, \bar{W} - k\right\}$ or as $r \leq \mathbb{E}max\left\{\varepsilon - \Psi^{-1}(P), 0\right\} + \bar{W} - k$. By inspection, the right hand is strictly decreasing in k, but also strictly increasing in P. Thus, the IR constraint defines a strictly increasing function $k = k\left(P\right)$.

In the case where $k = 0$, the first order condition is

$$\eta - \bar{W} + \Psi^{-1}(P) + P\Psi^{-1\prime}(P) = 0\,. \qquad (1.15)$$

When the firm can optimally set k, the first order condition is

$$\eta - \bar{W} + \Psi^{-1}(P) + P\Psi^{-1\prime}(P) + k'(P) \times (P + (1-P)\theta) + k(1-\theta) = 0\,. \qquad (1.16)$$

Given that $k'(P) \times (P + (1-P)\theta) + k(1-\theta)$ is positive for all P and that $\Psi^{-1}(P) + P\Psi^{-1\prime}(P)$ is decreasing in P (by the concavity of $P\Psi^{-1}(P)$), it follows that the P that solves Equation (1.16) (where the firm optimally sets k) is greater than the P that solves Equation (1.15) (where $k = 0$).[85] ∎

Next, I examine the impact of productivity overconfidence of workers becoming more overconfident on the probability of training. The idea is that as workers become more overconfident about their productivity, it is easier to get them to agree to work for a given wage at a given training contract and they become less likely to quit afterward for a given wage. I first consider overconfidence that persists after the worker goes through training. After this, I will consider overconfidence that goes away after training via learning. When workers have biased beliefs, the firm's problem can be written as

$$\arg\max_{w,k}(1 - F(-w\eta' - k + \overline{W})) * (1-w)\eta + F(-w\eta' - k + \overline{W})\theta k - c$$
$$Emax(w\eta' + \epsilon, \overline{W} - k) \geq r$$

where η' is the worker's belief about his productivity.

Proposition 2. *An increase in permanent overconfidence leads to a higher likelihood of training, both when training contracts are available and when they are unavailable, provided that the worker perceives his current earnings to be more responsive to his current productivity than his outside option. Formally, suppose that* $w > \frac{d\overline{W}}{d\eta'}$*, where* w *solves* $\eta - w\eta = G(-w\eta + \overline{W})$*; then an increase in* η' *increase the probability of training.*

Proof. Note that the outside option $\overline{W}$ may depend on η'. Start with when k is restricted to be zero. Let w^* denote the optimal salary if workers had rational beliefs. If workers believe their productivity will be

positive infinity faster (all worker stay and get paid minus infinity) or to have W go to minus infinity and have k go to minus infinity slower (all workers quit and firm collects infinity from them). Both of these solutions violate the IR constraint.

[83]Setting W to infinity is clearly suboptimal. When W is negative and large in magnitude, the derivative wrt W may be positive. When k is a large number, the derivative may be negative. Setting k to minus infinity is clearly suboptimal.

[84]I am grateful to Ben Hermalin for greatly simplifying the proof that retention increases.

[85]Indeed, one can show that retention is highest in the problem where the firm optimally sets k compared to any other situation where k is exogenous.

$\eta' > \eta$, setting a wage of w^* is still feasible (as $Emax(w\eta' + \epsilon, \overline{W}) > Emax(w\eta + \epsilon, \overline{W})$). Further, given the rational wage w^*, an increase in productivity beliefs increases profits so long as $sgn\left(\frac{\partial\left(1-F(-w\eta'-k+\overline{W})\right)}{\partial\eta'}\right) = sgn\left(-f*\left(-w+\frac{d\overline{W}}{d\eta'}\right)\right) = sgn\left(w-\frac{d\overline{W}}{d\eta'}\right) = +$. Since using the rational belief wage results in higher profits when workers have biased beliefs than when they are rational, optimal wage profits must increase when workers have biased beliefs. When the firm sets optimal w and k, the argument is similar. Let (w^*, k^*) denote the optimal wage and training contract when workers are rational. Setting (w^*, k^*) is still feasible if workers come to believe their productivity will be $\eta' > \eta$ and increasing beliefs from η to η' increases profits under (w^*, k^*) so long as $w > \frac{d\overline{W}}{d\eta'}$ and $(1-w)\eta - \theta k > 0$. ■

From this claim, we learned that overconfidence can make training profitable for firms when they are using training contracts. In the next claim, I consider the case where workers are overconfident, but this overconfidence disappears after training. In this case, overconfidence increases training when training contracts are available, but not if training contracts are unavailable.

Proposition 3. *An increase in non-permanent overconfidence leads to a higher likelihood of training when training contracts are available, but has no effect when training contracts are not available.*

Proof. Recall that the IR constraint is slack when a training contract is not available. Since non-permanent overconfidence will only affect the IR constraint, the overconfidence has no effect on profitability. However, when a training contract is available, the IR constraint must bind. An increase in overconfidence can be combined with either a decrease in wage or an increase in k, and the IR constraint will still hold, thereby increasing profits. ■

1.B Structural Model Derivations

One of the key equation in the model is the expected maximum or Emax term. I have that:[86]

$$\begin{aligned}
E(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) &= E_{y_t}E_{\epsilon|y_t}(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) \\
&= E_{y_t}E_{\epsilon}(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) \\
&= E_{y_t}E_{\epsilon}(max\{\overline{V}^S_{t+1}(\mathbf{x}_{t+1}) + \epsilon^S_{t+1}, \overline{V}^Q_{t+1} + \epsilon^Q_{t+1}\}|\mathbf{x}_t) \\
&= \int E_{\epsilon}(max\{\overline{V}^S_{t+1}(\mathbf{x}_{t+1}) + \epsilon^S_{t+1}, \overline{V}^Q_{t+1} + \epsilon^Q_{t+1}\}|\mathbf{x}_{t+1})f^b(y_t|y_1, ..., y_{t-1})dy_t \\
&= \int \tau \log\left(\exp(\frac{\overline{V}^S_{t+1}(\mathbf{x}_{t+1})}{\tau}) + \exp(\frac{\overline{V}^Q_{t+1}}{\tau})\right) f^b(y_t|y_1, ..., y_{t-1})dy_t \\
&= \sum_k \tau \log\left(\exp(\frac{\overline{V}^S_{t+1}(\mathbf{x}_{t+1})}{\tau}) + \exp(\frac{\overline{V}^Q_{t+1}}{\tau})\right) \Pr(y^k_t|y_1, ..., y_{t-1}).
\end{aligned}$$

The first equality expresses that the value function depends on future earnings and future idiosyncratic shocks. The second equality follows because future earnings are independent of the idiosyncratic shocks. The third equality uses the definition of V and that V^Q_s is independent of the state variable. The fourth equality integrates out y_t, which are not observed when the driver makes his period t decision, but are observed in the future. The fifth equality integrates out the idiosyncratic shocks. The last equality follows because, in implementation, earnings will be discretized into K possible values. The probability $\Pr(y^k_t|y_1, ..., y_{t-1})$ can be easily shown to depend only on $\overline{y}_{t-1}$, and is expressed below. As long as average

[86]Similar derivations can be found in Rust (1987) and Stange (forthcoming).

earnings to date by time $t+1$, $\overline{y}_t$, is a sufficient statistic for x_{t+1}, then it follows that average earnings by time t, $\overline{y}_{t-1}$, is a sufficient statistic for x_t.

Collecting terms from the model, I show how to derive the likelihood function:

$$
\begin{aligned}
L_i &= \int L(d_{i1},...,d_{it},y_{i1},...,y_{it},b_{i1},...,b_{it}|\alpha,\eta_b)f(\alpha,\eta_b)\,d\alpha d\eta_b \\
&= \int\{L(d_{i1},...,d_{it}|y_{i1},...,y_{it},b_{i1},...,b_{it},\alpha,\eta_b)*L(b_{i1},...,b_{it}|y_{i1},...,y_{it},\alpha,\eta_b) \\
&\quad *L(y_{i1},...,y_{it}|\alpha,\eta_b)f(\alpha,\eta_b)\,d\alpha d\eta_b\} \\
&= \int\{L(d_{i1},...,d_{it}|y_{i1},...,y_{it},b_{i1},...,b_{it},\alpha,\eta_b)*L(b_{i1},...,b_{it}|y_{i1},...,y_{it},\eta_b) \\
&\quad *L(y_{i1},...,y_{it})f(\alpha,\eta_b)\,d\alpha d\eta_b\} \\
&= \{\int L(d_{i1},...,d_{it}|y_{i1},...,y_{it},\alpha,\eta_b)*L(b_{i1},...,b_{it}|y_{i1},...,y_{it},\eta_b)f(\alpha,\eta_b)\,d\alpha d\eta_b\}L(y_{i1},...,y_{it}) \\
&= \left[\int\prod_{s=1}^{t}L(d_{is}|d_{i1},...,d_{is-1},y_{i1},...,y_{it},\alpha,\eta_b)*\prod_{s=1}^{t}L(b_{is}|b_{i1},...,b_{is-1},y_{i1},...,y_{it},\eta_b)f(\alpha,\eta_b)\,d\alpha d\eta_b\right] \\
&\quad *L(y_{i1},...,y_{it}) \\
&= \left[\int\prod_{s=1}^{t}L(d_{is}|y_{i1},...,y_{is-1},\alpha,\eta_b)\prod_{s=1}^{t}L(b_{is}|y_{i1}...,y_{is-1},\eta_b)f(\alpha,\eta_b)\,d\alpha d\eta_b\right]\left(\prod_{s=1}^{t}L(y_{is}|y_{i1}...,y_{is-1})\right) \\
&\equiv \{\int L_i^1(\alpha,\eta_b)L_i^3(\eta_b)f(\alpha,\eta_b)\,d\alpha d\eta_b\}L_i^2
\end{aligned}
$$

The first equality and second equalities follow by the law of total probability. The third equality follows because productivity is unaffected by the taste and overconfidence heterogeneity and because beliefs are unaffected by the taste heterogeneity. The fourth equality holds because, since earnings do not depend on the unobserved heterogeneity, they can be taken outside the integral. The fifth equality follows because (a) future earnings are not observed when a worker decides to quit and (b) quit decisions are independent of reported subjective beliefs conditional on the overconfidence unobserved heterogeneity. The sixth equality follows because (a) since the ϵ shocks are iid, the decision to quit is conditionally independent of all prior decisions to quit (given the earnings realizations and the unobserved heterogeneity) and (b) reported subjective beliefs are conditionally independent of past reported subjective beliefs conditional on productivity and the belief heterogeneity. In the seventh equality, I define the part of the likelihood due to the quitting decisions as $L_i^1(\alpha,\eta_b)$, the part due to the earnings realizations as L_i^2, and the part due to subjective beliefs as $L_i^3(\eta_b)$.

Once the likelihood function has been derived, it is important to given an expression for $L(y_{is}|y_{i1}...,y_{is-1})$. This is done as follows:

$$
\begin{aligned}
f(y_{i1}) &\sim N(\eta_0,\sigma_0^2+\sigma_y^2) \\
f(y_{is}|y_{i1},...,y_{is-1}) &\sim N((1-\gamma_{s-1})\,\eta_0+\gamma_{s-1}\overline{y}_{is-1},\Omega_{s-1}) \text{ for } s>1
\end{aligned}
$$

and where $\gamma_s=\frac{s\sigma_0^2}{s\sigma_0^2+\sigma_y^2}$ and $\Omega_s=\frac{\sigma_0^2\sigma_y^2}{s\sigma_0^2+\sigma_y^2}+\sigma_y^2$.[87]

[87] This follows by applying the standard formula for the conditional density for a multivariate normal distribution: $X_1|(X_2=x_2)\sim N(\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2),\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

Chapter 2

Overconfidence at Work and the Evolution of Beliefs: Evidence from a Field Experiment

2.1 Introduction

How do people form beliefs? Classical economic theory assumes that people are Bayesian, with accurate priors and a rational updating process. Growing work, however, has shown some significant departures from this assumption. Notably, a voluminous literature in psychology and a growing literature in economics documents that agents tend to be overconfident.[1] Both the economics and psychology literatures usually examine overconfidence using laboratory experiments where subjects engage in simple short-term tasks (e.g. trivia tests). The laboratory provides researchers with exceptional experimental control and researchers on overconfidence have recently made substantial strides, particularly in testing competing theories of overconfidence. However, economists may be concerned to what extent these findings generalize outside the laboratory. Perhaps overconfidence exists on trivia tests, but is absent in day-to-day economic activities? Within individuals, is overconfidence unchanging and fixed, or would certain interventions or policies eliminate it? Might the accumulation of experience and information cause overconfidence to disappear?

This paper analyzes overconfidence in examining how workers form beliefs about their productivity. I define the term "overconfidence" as people believing their performance will be better in absolute terms than it actually is.[2] A field experiment was conducted with a large

[1]In economics, recent experimental papers include Burks et al. (2010), Eil and Rao (2011), Benoit and Dubra (forthcoming), Mobius et al. (2010), Ericson (2010), Mayraz (2011b), Grossman and Owens (2010), and Hoffman (2011a). Non-experimental empirical papers include Barber and Odean (2001), Malmendier and Tate (2005, 2008), DellaVigna and Malmendier (2006), and Bergman and Jenter (2007).

[2]This definition of overconfidence is what Moore and Healy (2008) refer to as "absolute overconfidence." Absolute overconfidence can be distinguished from "relative overconfidence," in which people believe they are better than others, and "over-precision," in which people believe their beliefs are more accurate than

U.S. trucking firm, referred to hereafter as Firm B, in which new drivers were asked every week to predict their weekly productivity.[3] At this firm (as well at most long-haul trucking firms), workers are paid almost exclusively by piece rate, with output defined as the number of miles driven each week. Thus, the prediction task is likely to be of significant personal interest to workers since it is directly linked to their pay. The goal was to examine whether workers were overconfident and, if so, to see whether overconfidence might be eliminated using interventions. The experiment randomly assigned 254 new drivers to receive incentives for accurate forecasting or to guess without incentives. After several weeks, drivers were assigned to one of three additional treatments: information, increased incentives, or control. The information treatment consisted of informing workers about the existence of overconfidence from workers at another firm. Increased incentives consisted of using increased stakes (up to $50 per week) for accurate guessing.

I find that workers exhibit substantial overconfidence about their productivity. Overconfidence exists with and without incentives and there is no evidence that using incentives (either small or larger sized) reduces overconfidence. However, due to a limited sample size, I cannot rule out that incentives may have economically meaningful impacts on the level of overconfidence. In contrast, I find that the information intervention substantially reduces worker overconfidence. Workers' subjective beliefs decline by around 200 miles. However, the effect of the information diminishes within several weeks. The short-run reduction in overconfidence has no effect on workers' self-reported job satisfaction or job search behavior.

The field experiment shows that overconfidence appears to be difficult to alter, at least in the short-run. Even if interventions have little effect, it could still be the case that overconfidence diminishes on its own over time. To shed light on this, I use data from second large trucking firm, referred to hereafter as Firm A. Around 1,000 new drivers were asked every week to predict their weekly productivity for over two years. In another paper, Hoffman (2011), I use this data to estimate a structural quitting model with biased beliefs. Workers at Firm A are substantially overconfident about their productivity, but bias only slowly diminishes over time. After more than one year with the company, workers are overconfident by roughly 300 miles per week (or 15% of their average productivity); only toward the end of the Firm A data, after around 100 weeks, is overconfidence no longer statistically significant, though the point estimate is still positive and moderately-sized.

Despite the overconfidence observed at both firms, I find that workers update beliefs in a way consistent with aspects of Bayesian learning. In response to new productivity realizations, workers revise their subjective productivity beliefs in the corresponding direction. Moreover, consistent with Bayesian updating, the greater the number of signals accumulated, the more weight agents placed on the average of past signals.

I also explore heterogeneity in overconfidence. Despite average predictions exceeding average productivity, there are many workers at both firms who appear well-calibrated or

they actually are.

[3]Firm B is a different firm from the firm considered in Hoffman (2011), which I label Firm A. Data from Firm A are also considered later on in this paper.

even slightly underconfident about their productivity. In addition, some workers continue to severely overpredict their miles. This overconfidence appears to be unrelated to most observable characteristics. Using the structural model developed in Hoffman (2011), I show that this heterogeneity remains even after non-random attrition and measurement error in belief elicitation are accounted for.

The large body of evidence in psychology on overconfidence prompted De Bondt and Thaler (1995) to refer to overconfidence as "the most robust finding in the psychology of judgment." However, a large portion of research on overconfidence relies on unincentivized belief elicitation, which may lead to biases. Lack of incentives for belief elicitation might be thought an even more serious concern in field settings where the researcher exercises less control over the subject's environment. A main contribution of this paper is to demonstrate that incentives do not appear to affect the level of overconfidence. In addition, it suggests that subjects' stated subjective forecasts are not merely aspirational targets and that overconfidence is unlikely to be completely accounted for in this setting by utility from beliefs. Whether overconfidence can be reduced or eliminated has also been an object of study in psychology, though it has not in economics.[4] I contribute to this literature by showing in a field setting that information can reduce overconfidence. However, the effect does not appear to be long-lasting.

My results on the stickiness of overconfidence are also highly relevant for theories in labor economics which utilize information on worker beliefs.[5] For example, in the canonical theory of turnover due to Jovanovic (1979), workers decide whether to quit a job based on their beliefs about their future productivity, taking into account the option value from future learning. Worker beliefs are also an integral part of theories of occupational choice (e.g. Gibbons et al. (2005); James (2011); Sanders (2011)) and compensating wage differentials (e.g. Viscusi and O'Connor (1984); Viscusi and Moore (1991)). Accounting for biased beliefs within such theories may help enrich these models and improve the fit of the models to the data. In addition, worker overconfidence may be important when considering policy related to training, long-term contracts, and human capital investment. Hoffman (2011) studies contracts where firms provide general training, but then require workers to stay for some period of time after receiving training; he argues that worker overconfidence is important for allowing these contracts to operate. More generally, this paper contributes to a growing literature analyzing subjective beliefs.[6]

[4]Psychologists refer to efforts to eliminate overconfidence and other biases as debiasing. A comprehensive early survey on debiasing is given in Fischoff (1982). Many papers argue that debiasing in laboratory environments is possible (Arkes et al. (1987); Lau and Coiera (2009); Haran et al. (2010)), but some do not (Sanna et al. (2002); Fleisig (2011)).

[5]Other research has examined overconfidence with truckdrivers. Walton (1999) shows that drivers believe they are safer than other drivers. The excellent study by Burks et al. (2010) is particularly relevant. Using the same sample of drivers from Firm A as in my study, Burks et al. (2010) ask drivers to predict their quintile for performance on an IQ test. My study differs in that it focuses on absolute rather than relative overconfidence, it focuses on productivity overconfidence instead of IQ overconfidence, and it analyzes overconfidence over a long period of time. The data on workers at Firm B were newly collected for this paper.

[6]See e.g. Dominitz and Manski (1997), Dominitz (1998), Hurd and McGarry (2002), Khwaja et al. (2007),

The rest of the paper is structured as follows. Section 2.2 briefly describes the institutional setting and the data. Section 2.3 provides a basic theory of worker beliefs. Section 2.4 analyzes the data from the field experiment at Firm B. Section 2.5 analyzes belief data at Firm A. Section 2.6 concludes.

2.2 Data and Institutional Setting

2.2.1 Background

Firms A and B are two large long-haul U.S. trucking firms. Trucking is one of America's largest occupations, with roughly 3 million workers (Bureau of Labor Statistics, 2010). Truckers work for for-hire trucking firms (like Firm A or Firm B) or private companies that employ truckers (like Walmart or Safeway). Long-haul drivers, also called Over the Road or Truckload drivers, refer to drivers who transport goods over long distances. Most long-haul driver are paid by piece rate (Belman et al., 2005); new drivers at a large trucking firm might be paid $0.30 per mile.

Productivity in trucking is measured in terms of the number of driven per week.[7] According to the Department of Transportation Hours of Service Regulations, drivers may only work up to 60 hours per week.[8] Though data on hours is not available for Firm A or Firm B, both companies report that most workers attempt to work the maximum number of hours per week. Once a worker delivers a load, they can proceed to delivering another one. At Firm A, driver loads are assigned by a central company dispatching system, making favoritism in route assignment unlikely.

Despite limits on the number of hours, there are large differences across drivers in average productivity. For example, Hoffman (2011) finds using structural estimation that the standard deviation of permanent productivity across drivers is almost 300 miles per week. What is the cause of these differences? In interviews, managers at Firm A and Firm B reported time management, trip planning, speed, and getting lost as four main sources of potential differences. In addition, there is also widespread week-to-week variation in worker productivity. Productivity can be affected by weather, the length of the trip assigned, and whether the shipper unloads trucks in a timely fashion. Drivers generally report that getting longer loads helps increase weekly productivity. Hoffman (2011) documents that the weekly variation in productivity within drivers is almost 700 miles per week. If this variation is largely idiosyncratic, we should not expect that drivers will perfectly predict their productivity every week. In addition, this week-to-week noise means that a worker's inference problem about underlying productivity is not totally finished by the end of the first week.

Delavande (2008), and Zafar (2010). See Manski (2004) for a thorough discussion of the literature.

[7]For another analysis of productivity in trucking, see Hubbard (2003).

[8]Alternatively, drivers may work up to 70 hours over an 8 day period.

2.2.2 Data Description

Personnel data is available on all workers at Firm A from 2001 through 2009 and all workers at Firm B from 2005 through 2009. At Firm A, there is weekly information on miles, earnings, and quitting behavior, as well as various demographic information including race, gender, age, and other information obtained in the driver's job application. For a subset of around 1,000 workers starting training with Firm A in late 2005, much richer data are available, including a wide variety of survey and preference characteristics. This data was collected by Burks et al. (2010) and includes the weekly subjective belief question: "About how many paid miles do you expect to run during your next pay week?" I interpret this question as asking drivers their beliefs about their average number of miles during the next week. No incentives were used to motivate accurate belief responses, though drivers were given $5 each week for completing the survey. Further information on the data at Firm A can be found in Hoffman (2011).

Data collected from Firm B are very similar and consist of matching weekly miles to weekly subjective beliefs. The description of the experimental design is given in Section 2.4.

Table 2.1 provides summary statistics for the data. The top panel gives data for Firm A while the bottom panel gives data for Firm B. The median worker at Firm A is white, male, and unmarried, with an average age of 36 years. All the drivers at Firm A are brand-new to trucking, with no prior long haul trucking experience. Though drivers at Firm B have low tenure levels, most have prior trucking experience. The average past trucking experience is 8.01 years, though the distribution is right-skewed, with 28% of drivers having zero to one years of experience and 33% of drivers having two to three years of experience.

2.3 Theory

I present a simple theory of how workers form subjective beliefs about their ability. The model is from Hoffman (2011). Workers differ from one another in their underlying productivity. At the end of the week, they observe their productivity for that week, and use this information in making their next week's forecast.

Workers have baseline productivity η, which is distributed $N\left(\eta_0, \sigma_0^2\right)$. This baseline productivity is unknown to both workers and firms. A worker's weekly miles, y_t, are distributed $N\left(a\left(t\right) + \eta, \sigma_y^2\right)$, where $a(t)$ is a term associated with learning by doing that can be estimated from the data. In creating his subjective belief forecast, the worker thus faces a classic signal extraction problem. He must use his observed productivity draws to form a prediction of his future productivity.

If a worker exhibited standard Bayesian learning, the expectation could be written as:

$$E(y_t|y_1, ..., y_{t-1}) = \frac{\sigma_y^2}{\left(t-1\right)\sigma_0^2 + \sigma_y^2}\eta_0 + \frac{\left(t-1\right)\sigma_0^2}{\left(t-1\right)\sigma_0^2 + \sigma_y^2}\frac{\sum_{s=1}^{t-1} y_s - a(s)}{s-1} + a(t) \qquad (2.1)$$

The term $\frac{\sigma_y^2}{(t-1)\sigma_0^2+\sigma_y^2}$ is the weight the agent places on the mean of his prior, whereas the term $\frac{(t-1)\sigma_0^2}{(t-1)\sigma_0^2+\sigma_y^2}$ is the weight the agent places on his (de-meaned) average productivity draws. As tenure t increases, the agent puts more and more weight on his productivity as he learns his true productivity. This leads to a testable implication that workers should increase their subjective beliefs in response to increases in average past productivity, and that the responsiveness of subjective beliefs to average productivity should increase over time.

However, as I will show shortly in the data, workers' subjective beliefs are systematically higher than their actual observed productivity. This difference is large at first and decreases slowly over time. One way of modeling this is to relax the assumption that workers' priors exactly match the actual distribution of productivity in the population. In particular, workers believe the average productivity in the population is given by $\eta_0 + \eta_b$ instead of η_0. I allow the term η_b to be heterogeneous. If $\eta_b > 0$, then the worker is overconfident. If $\eta_b < 0$, then the worker is underconfident. The rate at which overconfidence disappears depends on how the weight on signals versus priors changes over time. Overconfidence decreasing slowly over time can be modeled by having the perceived noise in the productivity signals be greater than the actual noise in the productivity signals. Let $\widetilde{\sigma_y}$ denote the perceived noise in the productivity signals. Then the model can be written out as:

$$E^b(y_t|y_1,...,y_{t-1}) = \frac{\widetilde{\sigma_y}^2}{(t-1)\,\sigma_0^2 + \widetilde{\sigma_y}^2}(\eta_0+\eta_b) + \frac{(t-1)\,\sigma_0^2}{(t-1)\,\sigma_0^2 + \widetilde{\sigma_y}^2}\frac{\sum_{s=1}^{t-1} y_s - a(s)}{s-1} + a(t) \quad (2.2)$$

In particular, the larger $\widetilde{\sigma_y}$ is, the slower the agents' overconfidence will disappear.

The goal of the paper is to document overconfidence and explore its relation to economic behavior, as opposed to explaining why agents are overconfident. Thus, I will only briefly speculate on why agents may have biased priors. One possibility is that people are endowed with high levels of overconfidence.[9] A second possibility is that people may have varying levels of overconfidence about their performance at different jobs, but they may choose the job at which they are most overconfident. Consider the following simple Roy Model. Suppose that mean prior bias η_b is distributed across the population of potential workers according to a distribution $F(\eta_b)$. Workers decide to go into trucking if the expected value of the job is higher than their outside option. Because workers with higher prior bias perceive the value of the job to be higher, they are more likely to go into the job than other workers. Hence, the beliefs of workers in a firm may reflect a Winner's Curse situation. High levels of overconfidence may exist in observed jobs even if the overall population is less overconfident (or even perfectly rational). A third possibility is that agents' beliefs may adjust to rationalize their choice of employer or occupation (Mayraz, 2011a).

[9]For example, Cesarini et al. (2009) document that relative overconfidence is correlated within pairs of identical twins.

2.4 Field Experiment at Firm B

2.4.1 Experimental Procedure

All drivers in their first four months of tenure with the company were selected for the field experiment. Firm B does not train its own drivers (unlike Firm A), so all drivers in the study had already received a commercial driver's license. The experiment was conducted via weekly phone surveys. A total of 254 drivers participated.

The experiment contained two randomizations. First, workers were randomized to receive incentives or not for correctly guessing about their mileage. Drivers assigned to the No Incentives condition were asked every week to predict their miles and earnings for the next week without incentives for accuracy. Drivers assigned to the Incentives condition were asked to predict their miles and earnings for next week, with accurate guesses rewarded under a quadratic scoring rule. Quadratic scoring rules are a common means by which experimental economists elicit agents' expectations in an incentive-compatible manner.[10] Under a quadratic scoring rule, the agent's payoff is given by $A - b(B - x)^2$, where b is the agent's stated belief, x is the realization of the random variable of interest, and A and B are constants chosen by the researcher. This mechanism is incentive-compatible if the subject is risk-neutral.[11] There were two levels of incentives randomly assigned. In the most common level, drivers could make up to \$10 per week for accurate guessing.[12] Drivers in the Incentive condition were paid based on their predictions of earnings or miles, with which one determined randomly. The payment for guessing miles was $10-10*((x_m-b_m)/1000)^2$ and the payment for guessing earnings was $10-40*(x_e-b_e)/1000)^2$, where x_m and x_e are actual miles and earnings and b_m and b_e are predicted miles and earnings. In order to see whether larger incentives mattered, some drivers were paid up to \$50 per guess and faced sharper penalties

[10]Some recent papers assessing the ability of quadratic scoring rules and alternative scoring rules to accurately elicit beliefs include Offerman et al. (2009), Hossain and Okui (2010), and Schlag and van der Weeley (2009). Recent papers using quadratic scoring rules include Mobius et al. (2010), Holt and Smith (2009), and Hoffman (2011a).

[11]To see why, consider the problem of choosing b in order to maximize one's payoff. Let $f(x)$ denote the agent's subjective assessment of the distribution of x:

$$argmax_b \int A - B(x-b)^2 f(x)dx$$
$$= argmin_b \int (x-b)^2 f(x)dx$$
$$FOC: \int \frac{d}{dx}(x-b)^2 f(x)dx = 0$$
$$b = \int xf(x)dx$$

[12]The amount was chosen so as to be large enough to be salient for drivers, but small enough to be unlikely to influence their driving behavior. For example, if drivers could receive hundreds or thousands of dollars for guessing accurately, they might choose to stop driving exactly when they reached their guess. In addition, paying large incentives to some workers but not others could lead to equity concerns.

for mistakes. These drivers were paid according to the rules $50 - 200 * ((x_m - b_m)/1000)^2$ and $50 - 800 * (x_e - b_e)/1000)^2$. All subjects were paid a \$5 participation fee for each survey taken.

After two to four weeks of making miles predictions with or without incentives, drivers were randomized to receive debiasing or not. Drivers selected for debiasing were read the following script:

Before we get started, we'd like to share with you some of our findings so far.

At another trucking company we studied, workers over-estimated their next week's miles by around 500 miles per week during their first few months with the company. That is, they thought they were going to drive 500 more miles per week than their actual average miles. Even after more than one year with the company, people were still over-predicting their miles by 300 miles per week; for example, many people thought they would average 2,400 miles per week, but they ended up only driving 2,100 miles per week.

Please think for a moment about the last few weeks. Were your predictions of your mileage high or low? Also think about the week ahead. Are there any factors that might decrease you mileage, e.g. bad weather, bad traffic, or a late unloading?

In our survey, your average prediction per week has been [INSERT MILES NUMBER] miles.

After the first paragraph, drivers were asked if they had any questions or comments.

Drivers selected not to receive debiasing were simply told:

In our survey, your average prediction per week has been [INSERT MILES NUMBER] miles.

2.4.2 Evidence of Overconfidence

Before I describe the experimental results, it is useful to describe how worker beliefs compare with actual productivity. I focus on the period before the information intervention. Figure 2.2 shows how driver productivity and beliefs evolve as a function of weeks in the experiment. I collapse the data by week and plot results over time. As can be seen, beliefs exceed actual productivity on average. This holds both in terms of the mean and median of the distribution, evidence that the results are not driven by outliers in productivity beliefs. When they began the experiment, drivers at Firm B had worked there for different lengths of time. Figure 2.3 takes advantage of this variation to show how beliefs and productivity vary with tenure. The gap between productivity and beliefs is present here as well.

2.4.3 Experimental Results

Table 2.2 shows the effect of using incentives on people's beliefs. It reports regressions of the form:

$$b_{it} = \alpha + \beta_t + \gamma_1 INCENT_i + X_{it}\delta + \epsilon_{it},$$

where b_{it} is agent i's subjective belief at tenure t, $INCENT_i$ is the main regressor of interest, X_{it} is other control variables, and ϵ_{it} is an error. Given multiple observations per person, I cluster standard errors at the person level. In addition, I focus on the effect of incentives in the period before debiasing occurs. As can be seen in Table 2.2, there is little evidence that using incentives significantly reduces agents' productivity beliefs.

Table 2.3 shows the effect of the debiasing intervention on workers' beliefs. Panel A reports regressions of the form:

$$b_{it} - b_i^{\bar{pre}} = \alpha + \beta_t + \gamma_1 INCENT_i + \gamma_2 DEBIAS_i + X_{it}\delta + \epsilon_{it},$$

where $b_i^{\bar{pre}}$ is person i's average beliefs before debiasing and $DEBIAS_i$ is a dummy for being assigned the debiasing treatment. In the first week of treatment, debiasing reduces an individual's productivity beliefs by roughly 230 miles. In the week after treatment, the estimated treatment effect is roughly -350 miles. Two weeks after treatment, the estimated treatment effect is -152 miles, though this is not statistically significantly different from zero (though the standard error is quite large). The effects in three and four weeks after also move in the expected direction, though they are small and very large standard errors prevent any meaningful inference. Panel B observes similar effects running the regression:

$$b_{it} = \alpha + \beta_t + \theta b_i^{\bar{pre}} + \gamma_1 INCENT_i + \gamma_2 DEBIAS_i + X_{it}\delta + \epsilon_{it},$$

that is, with the coefficient on $b_i^{\bar{pre}}$ not restricted to be one. Finally, Table 2.4 reports effects on search intention (whether drivers intend to look for a new job in the next six months, measured on a 3 point scale) and job satisfaction (measured on a 4 point scale) using ordered probits. The experiment has no apparent effects on either measure, though the standard errors are limited by small sample size.

The experiment suggests that agents adapt their beliefs in response to a treatment informing them about the existence of overconfidence. However, the impact significantly persists for only a week or two after treatment, although I cannot rule out that it persists longer.

2.5 Analysis of Firm A Belief Data

2.5.1 Survey Response

The average response rate across all drivers and weeks to the weekly beliefs question was 21%. Of the 699 drivers whom I focus on in the data subsample, 61% responded to at least one survey about mileage beliefs. Of the drivers who respond at least once to the survey, the average response rate was 26%. Columns 1-2 of Table 2.5 report regressions predicting response to the survey questions. Women and minority drivers were less likely to respond to the survey, whereas workers with higher average productivity and older drivers are more likely to respond. Column 2 presents some *within driver* results (where driver fixed-effects are added). Within a given driver, response is higher in weeks when the driver is more productive. These results suggest that more productive drivers were more likely to respond to the survey and that drivers were more likely to respond when they are more productive; these factors should work against detecting overconfidence in the data. Nevertheless, these issues of selection must be taken into account when interpreting the following results.

2.5.2 Belief Updating

What determines workers' productivity beliefs? This question is investigated in Columns 3-8 of Table 2.5. The normal learning model presented in Section 2.3 predicts that workers should increase their productivity forecasts in response to higher than expected productivity realizations. In particular, an increase in lagged average productivity should increase productivity forecasts. The data confirms this prediction. The coefficient on lagged average productivity is highly statistically significant and in some specifications it is close to one. For example, the coefficient of 0.72 in column 5 implies that a 10 mile increase in lagged average productivity increases the productivity forecast by 7.2 miles. Columns 7 and 8 reveal that this effect holds *within driver*. An additional prediction of the model is that average productivity should become more predictive of productivity forecasts at greater levels of tenure. This is because the weight on productivity realizations versus the prior increases with tenure. In column 6, the interaction term of lagged average productivity and tenure is positive and statistically significant. However, in column 8, once driver fixed effects are controlled for, the estimated coefficient is essentially zero.

To summarize, consistent with theory, drivers revise upward their productivity forecasts in response to increases in past average productivity. However, there is not strong evidence that this effect increases with tenure. It may be difficult to interpret this last result due to unobserved heterogeneity across drivers and the fact that the selection of drivers across time out of the sample was highly non-random. The structural estimates presented in Section 2.5.5 may be useful in this regard.

2.5.3 Overconfidence

To examine whether workers appear overconfident, Figures 2.4 and 2.5 compare productivity and productivity beliefs over time. These figures collapse the productivity and beliefs data by week of tenure to either the weekly mean or median (averaged across all drivers). As can be seen, drivers tend to systematically believe they will achieve higher productivity than they actually attain. Two potential confounds are that more productive drivers are more likely to answer the beliefs survey and that there is differential selection out of the data. To address these issues, the data is plotted restricted to drivers who respond to the survey[13] and also restricted to drivers who stay with the company at least 75 weeks.

Despite the overall level of overconfidence averaged across drivers, there is considerable *within driver* variation both in productivity and predicted productivity. Drivers do not

[13] Specifically, I drop all productivity observations where the driver did not respond to the survey in that week.

overpredict their miles in every week. Weeks where drive predictions are higher than next week's actual miles constitute 65% of the data, whereas weeks where driver prediction are lower than next week's actual miles constitute 35% of the data.[14]

2.5.4 Persistence and Adapting from Past Prediction Errors

Workers' predictions are higher than their actual productivity, even though they make their predictions for many weeks. Do workers adapt their forecasts in response to forecast errors? Although workers may not have fully rational expectations, I analyze here whether they have adaptive expectations, that is, do their beliefs adapt in response to new information?[15] To consider this question, I regress current forecast errors on past forecast errors to see whether workers adjust in response to past forecast errors. Specifically, I consider the specification:

$$\text{overconf}_{it} = \alpha + \theta_i + \beta_t + \gamma * \text{overconf}_{it-1} + X_i\delta + \epsilon_{it} \tag{2.3}$$

where $\text{overconf}_{it} = b_{it} - y_{it}$. A coefficient of $\gamma < 0$ is evidence for adaptive expectations.

Columns 1-3 of Table 2.7 analyze the above regression looking at the last time the agent answered the survey, whereas columns 4-6 restrict attention to cases where the last survey occurred in the previous week. Columns 4-6 will analyze fewer data points than columns 1-3 given that workers do not answer surveys in many weeks. Columns 1-2 and 4-5 consider the above regressions, but without individual fixed effects. I focus first on Panel A, where all drivers are used. The estimated coefficient on past overconfidence is positive and statistically significant. This means that agents who are overconfident one week tend also to be overconfident the previous week. However, once individual fixed effects are controlled for, the estimated coefficients become negative. The results in columns 3 and 6 suggest that agents do indeed exhibit adaptive forecasting. Panel B repeats the specifications from Panel A while eliminating drivers who share the same truck. [16] Under this restriction, the evidence on adaptive forecasting becomes stronger. Overall, the results seem to indicate heterogenous amounts of overconfidence across people, coupled with a tendency of workers to adapt after being too high or too low in their productivity forecasts.

2.5.5 Structural Examination of Belief Heterogeneity

To accurately account for non-response and selective attrition, I utilize a structural model of optimal quitting decisions developed in Hoffman (2011). The model is presented formally in the Appendix. Informally, every week the worker makes a decision whether or not to quit the firm. Workers are endowed with different abilities, but abilities are *ex ante* unknown. Rather,

[14] There are also nine instances where the forecast exactly equaled the number of miles.

[15] Whether agents have rational expectations or adaptive expectations is an important question in macroeconomics.

[16] This includes drivers who are paired together as a fixed team, as well as drivers who alternate teams. In these settings, both drivers are in the same truck, with one driver working and the other driver riding in the passenger seat or sleeping.

workers learn about their ability over time through their weekly productivity realizations. In addition, workers hold different priors about the mean of the distribution from which their ability is drawn, which can be interpreted as different beliefs about the productivity of an average truckdriver. Thus, having observed the same signal draws, different workers will have different posteriors about their true productivity because they hold different priors. Over time, however, as more and more productivity signals accumulate, the effect of workers' prior biases will fade away.

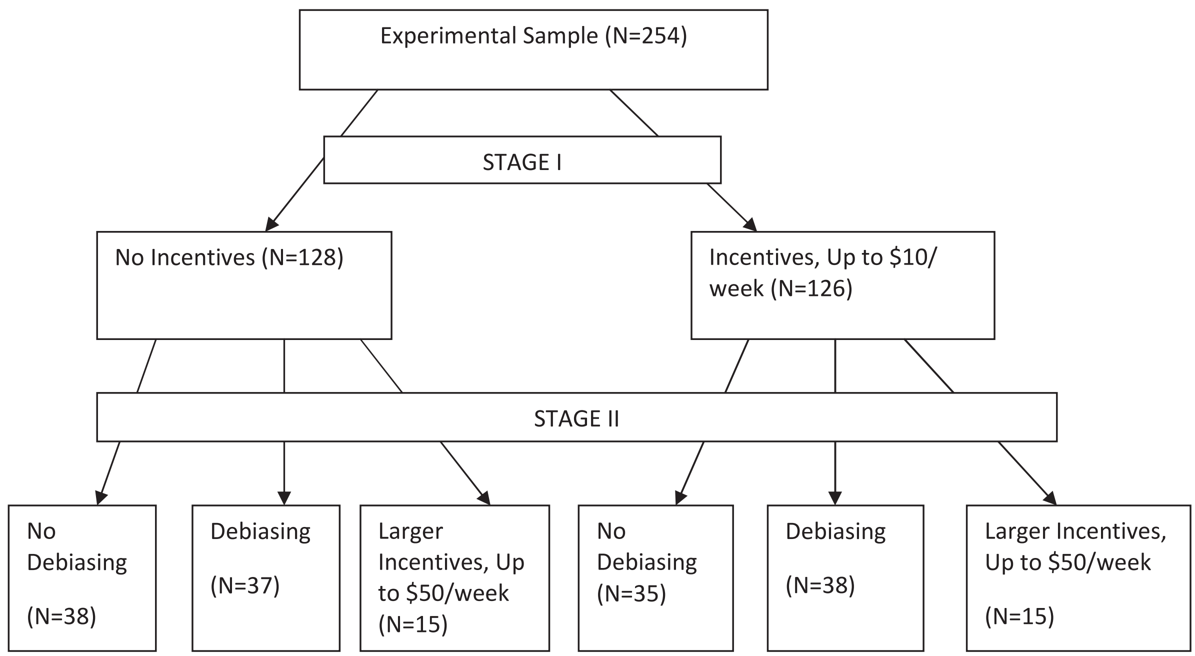
The structural model is useful because it allows estimation of the underlying structure while modeling the (non-random) selection out of the company in an internally consistent manner. Figure 2.7 plots the distribution of overconfidence estimates from the structural model. I assume that overconfidence is drawn from a mass-point distribution (Heckman and Singer, 1984), specifically, one with six mass points. The median (and modal) driver has a prior bias of slightly more than 500 miles per week, which is about 25% of productivity. The second most common mass point is just slightly above zero, corresponding to a situation of perfect calibration. In addition, there is a smaller number of drivers who are estimated to be somewhat under-confident, as well as a sizeable right tail exhibiting substantial overconfidence. This figure suggests substantial heterogeneity in overconfidence across drivers even when selection issues are explicitly controlled for.

2.6 Conclusion

This paper analyzes belief formation in the workplace. Using a field experiment, I show that workers are overconfident about their productivity, consistently over-predicting the number of miles they will drive. The robustness of overconfidence was examined through incentive and information interventions and by analyzing how overconfidence changes over two years. Incentives appear to have very limited impact on beliefs. Information can reduce overconfidence, though it appears that the effects do not last more than a few weeks. Over time, overconfidence does decrease, but only quite slowly. Workers do, however, exhibit aspects of Bayesian updating. They increase their productivity forecasts in response to positive news and adjust somewhat in response to prediction errors.

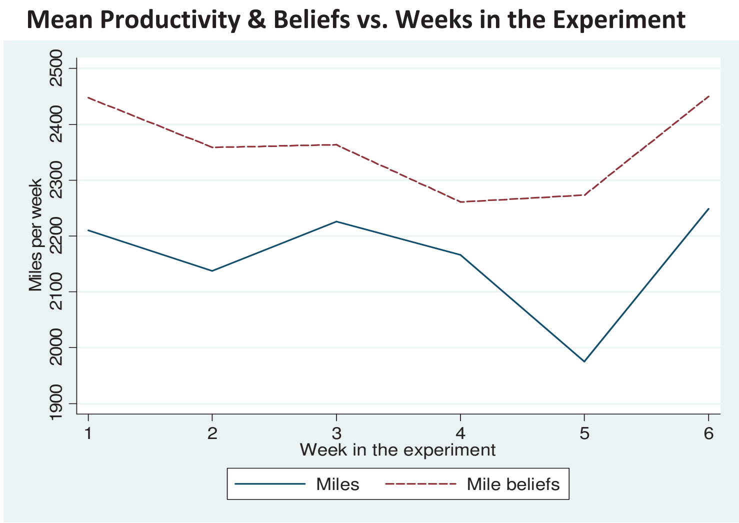
The implications of overconfidence in the workplace seem very rich. For example, overconfidence may be important for better understanding heterogeneity in managerial investment (Malmendier and Tate, 2005), promotion patterns (Kaniel et al., 2011), and the development of organizational culture (Van den Steen, 2010). Further research is clearly needed, particularly to further bridge lab findings and actual workplace behavior.

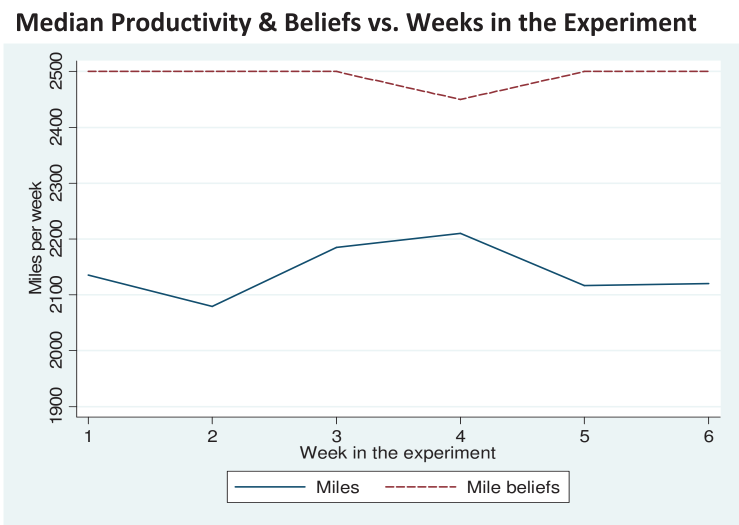
Figure 2.1: Experimental Design



Notes: This figure presents the experimental design. Stage I and Stage II each listed 2-6 weeks. When assigned to debiasing or no debiasing in Stage II, workers kept their assignment to non-incentivized or incentivized guessing. Beliefs were incentivized using the quadratic scoring rules described in the text. All workers are from Firm B.

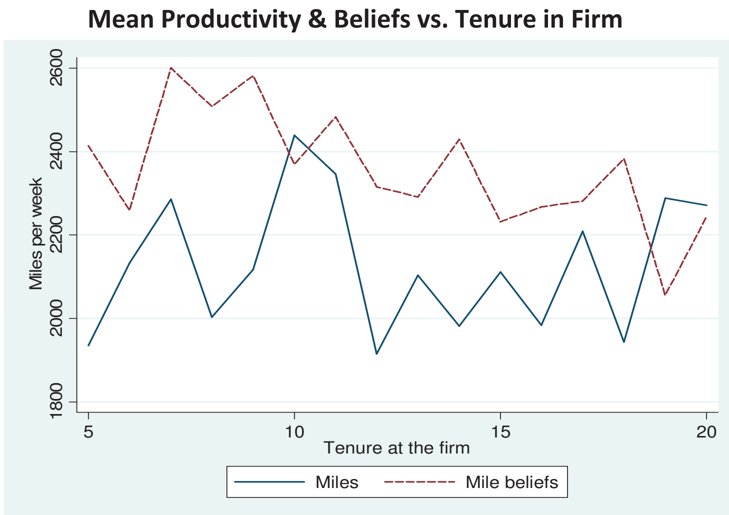
Figure 2.2: Overconfidence: Productivity and Believed Productivity by Weeks in the Experiment (Firm B)

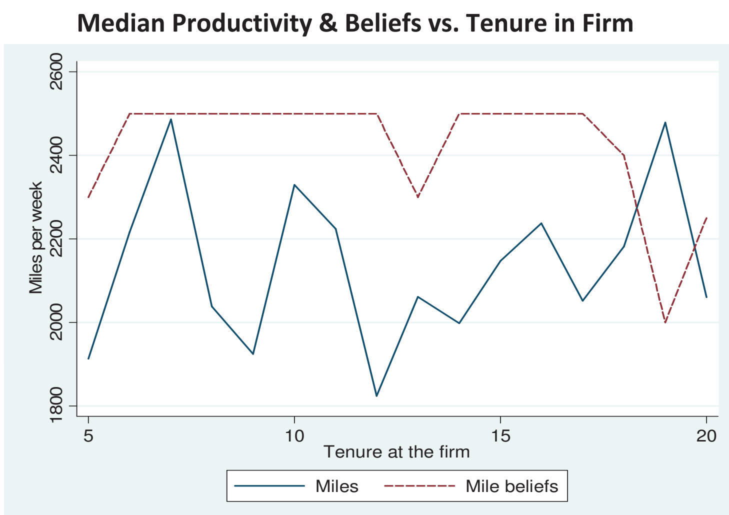




Notes: These figures analyzes actual and believed productivity for drivers in at Firm B. The data is collapsed by week. The figures are based on all the data prior to the provision of the information treatment. The top figure collapses the data by mean miles and beliefs whereas the bottom figure collapses the data by median miles and beliefs.

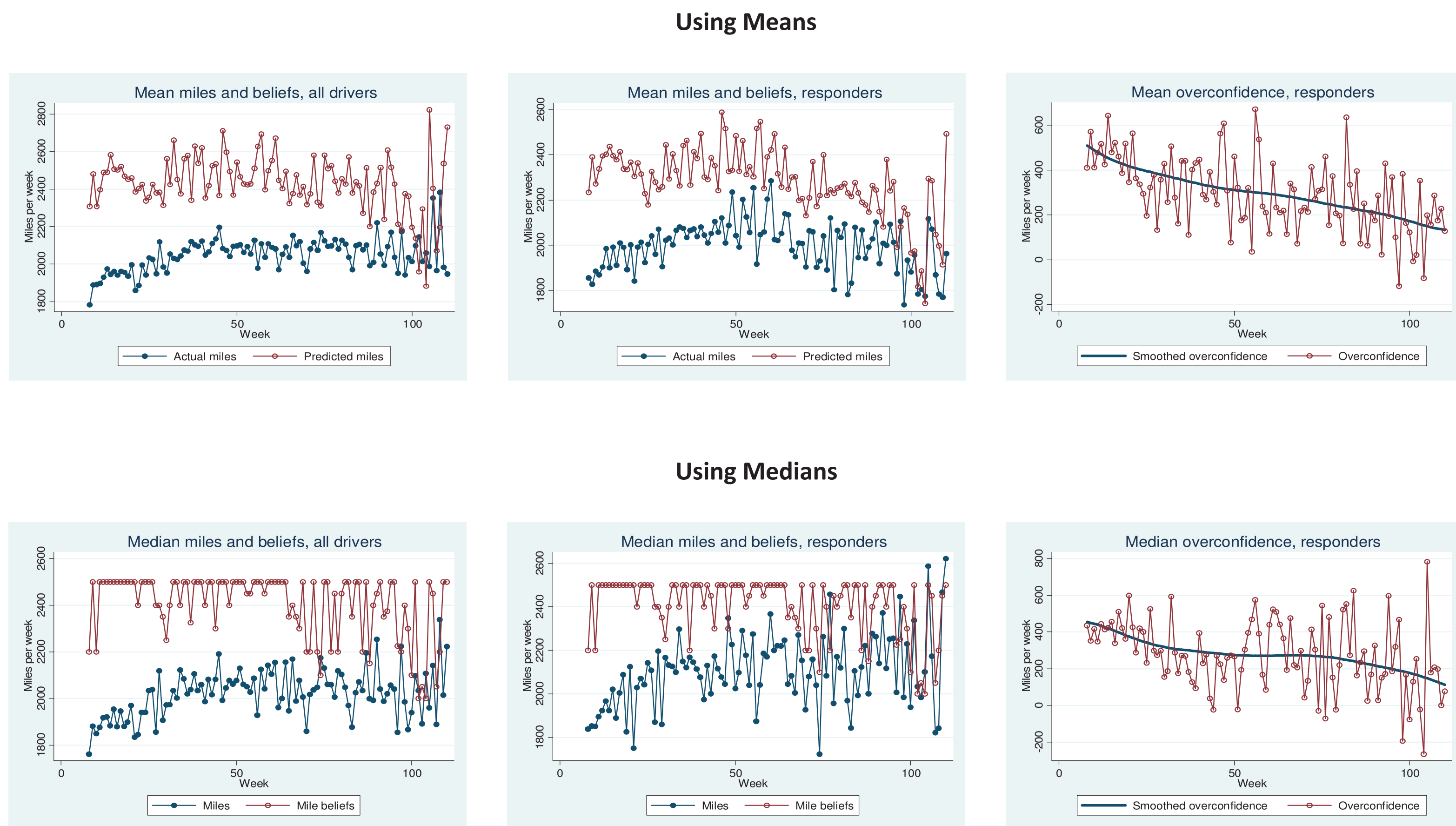
Figure 2.3: Overconfidence: Productivity and Believed Productivity by Tenure at the Firm (Firm B)





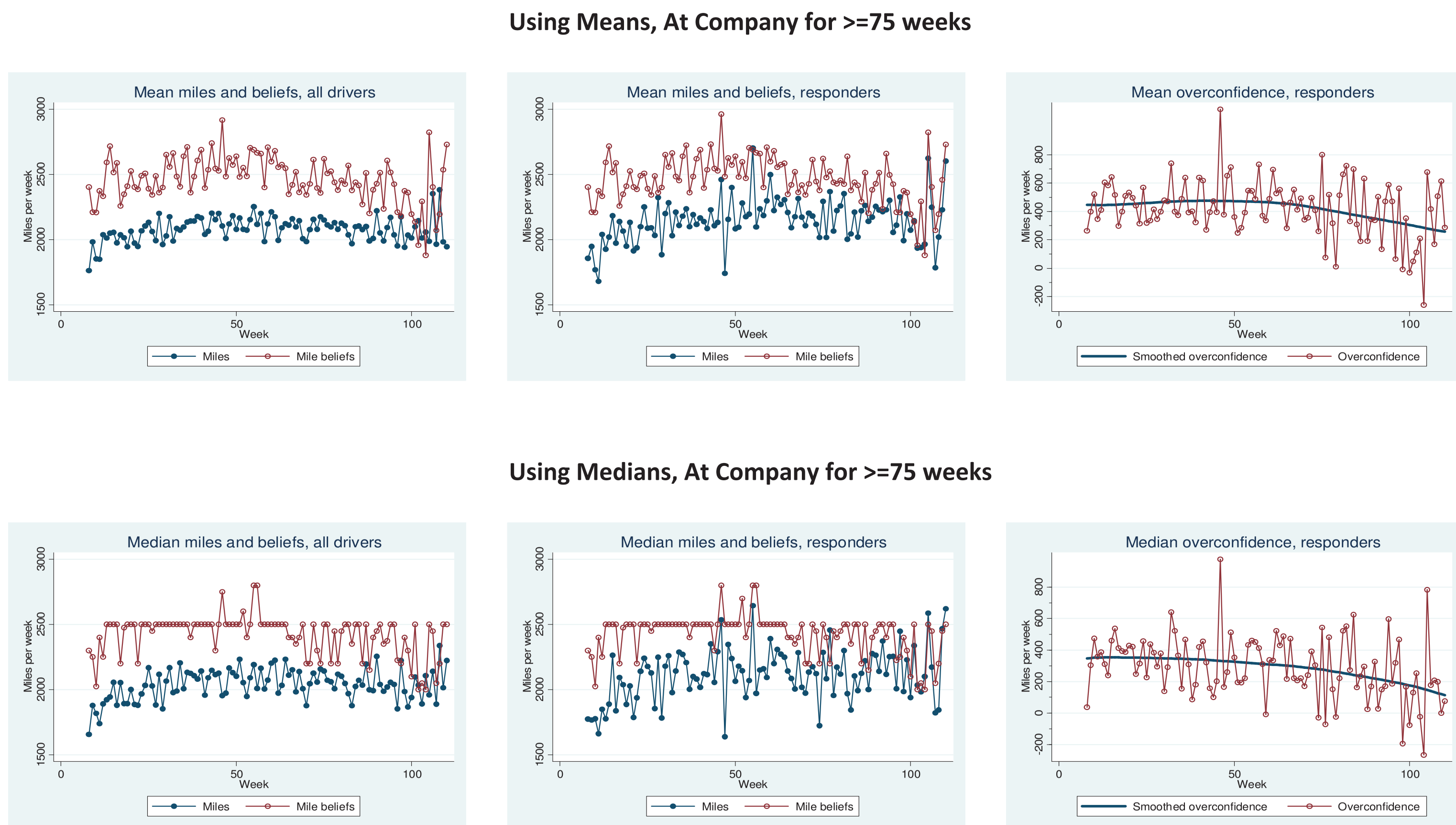
Notes: These figures analyzes actual and believed productivity for drivers in at Firm B. The data is collapsed by week. The figures are based on all the data prior to the provision of the information treatment. The top figure collapses the data by mean miles and beliefs whereas the bottom figure collapses the data by median miles and beliefs.

Figure 2.4: Productivity and Believed Productivity by Tenure



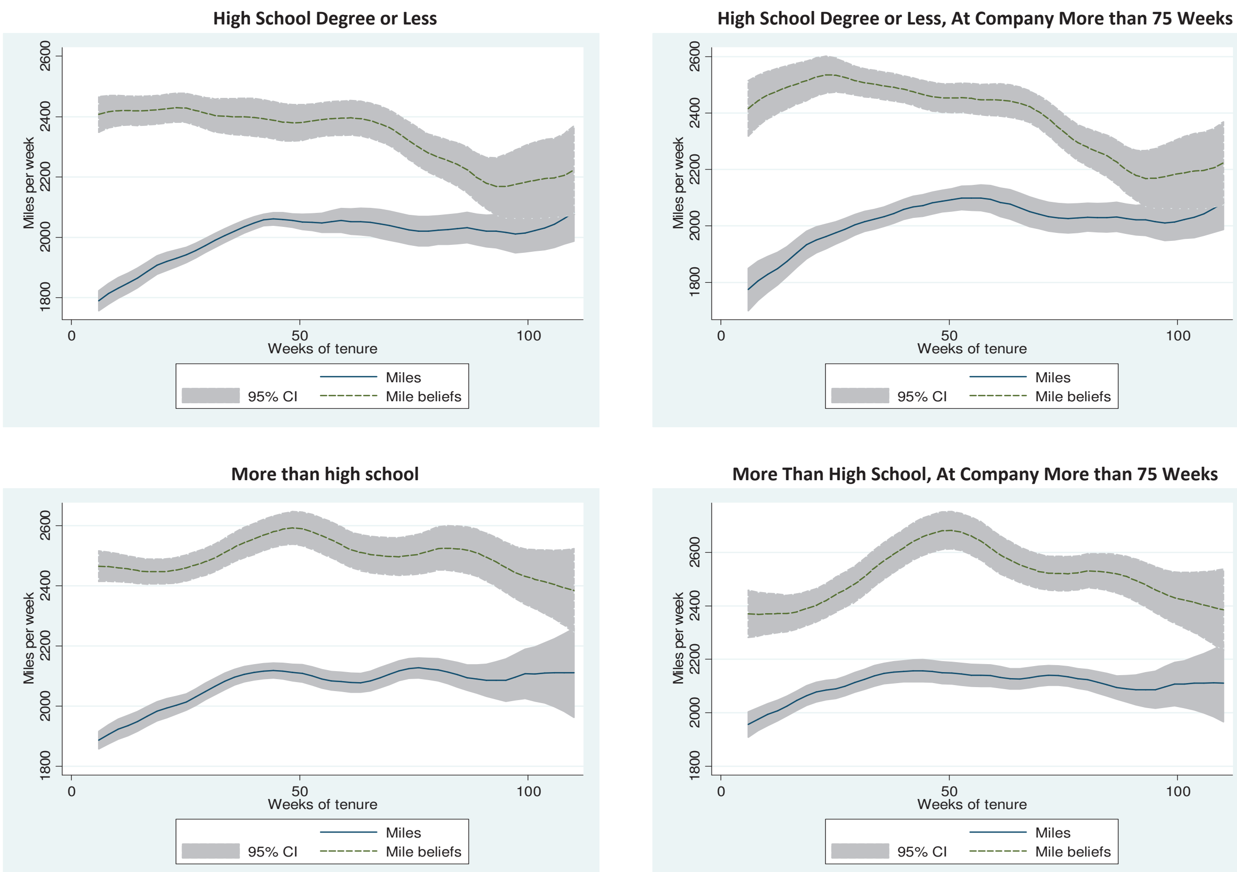
Notes: This figure analyzes actual and believed productivity for the 849 drivers at Firm A. The actual and predicted productivity data is collapsed by week, either in means or medians.

Figure 2.5: Productivity and Believed Productivity by Tenure, Restricted to Drivers Who Stay with Company At Least 75 Weeks



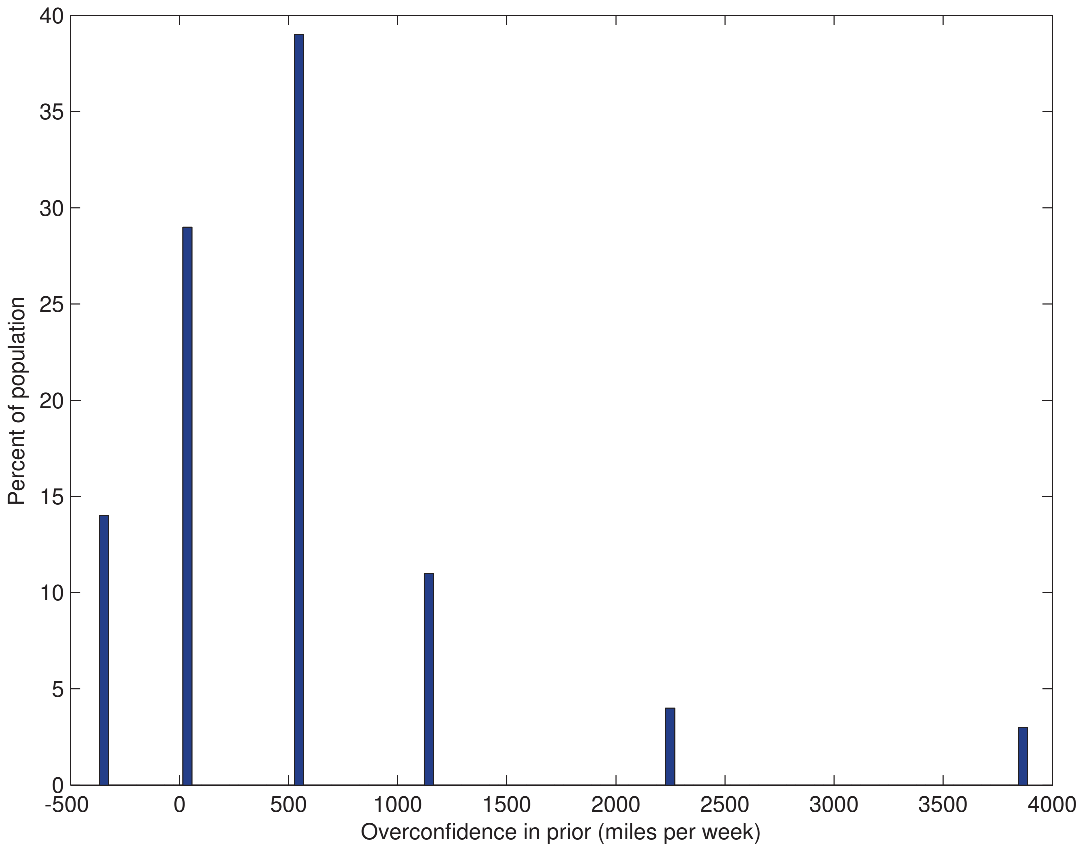
Notes: This figure analyzes actual and believed productivity for drivers at Firm A. In order to account for selection out of the data (quitting) based on productivity, it restricts to drivers who stay with the company for most of the sample, that is, for at least 75 weeks.

Figure 2.6: Belief Updating by Education



Notes: These figures analyzes actual and believed productivity for drivers in at Firm A. The data is collapsed by week and then smoothed using a local polynomial regression with an Epanechnikov kernel. For the productivity lines, the bandwidth is 5 weeks. For the productivity forecast line, the bandwidth is 7 weeks three graphs.

Figure 2.7: Estimated Distribution of Overconfidence in the Population



Notes: This figure presents estimates of the distribution in prior bias (overconfidence) in the population. The model is estimated using 6 mass points, with each bar in the picture corresponding to one of the estimated points. The median (and modal) prior bias is slightly above 500 miles per week. The data is estimated using drivers from Firm A.

Table 2.1: Summary Statistics

Firm A Drivers

Variable	**Obs**	**Mean**	**Min**	**Max**
African-American	895	0.11	0	1
Hispanic	895	0.02	0	1
Female	895	0.10	0	1
Married	895	0.41	0	1
Age	894	36.46	21.06	69.21
Exemptions	846	0.40	0	97
Number of Kids	895	0.96	0	7
Online Application	889	0.67	0	1
Smoker	787	0.46	0	1
Years of Schooling	895	12.85	9	18
High School Dropout	895	0.04	0	1
High School Graduate	895	0.40	0	1
Some College	895	0.34	0	1
Technical School	895	0.14	0	1
College Degree or More	895	0.08	0	1
Credit Score	784	585.96	407	813
Insufficient Credit for Score	895	0.12	0	1
BMI	698	28.19	16.39	52.54
Overweight (BMI>25)	698	0.64	0	1
Obese (BMI>30)	698	0.34	0	1

Firm B Drivers

Variable	**Obs**	**Mean**	**Min**	**Max**
Female	249	0.08	0	1
Age	254	41.31	21.7	66.26
Years of experience	244	8.01	0	41
0 or 1 years of experience	244	0.28	0	1
2 or 3 years of experience	244	0.33	0	1
Weeks of tenure	254	12.68	1	28

Notes: The data at Firm A is restricted to new drivers with over 1,000 weekly miles on average. Married at Firm A is compared to single and unspecified. Female at Firm A is compared to male and unspecified.

Table 2.2: Field Experiment: Effect of Incentives on Productivity Forecasts

Panel A: $10 Incentives vs. Control

Dep Var:	Miles Prediction		Overconfidence		Absolute Error	
	(1)	(2)	(3)	(4)	(5)	(6)
$10 Incentive	-40.01	-13.52	41.83	79.20	-0.55	9.67
	(58.93)	(59.64)	(73.81)	(76.58)	(44.77)	(46.05)
Trucking experience (yrs)		-0.59		3.26		4.34
		(4.00)		(4.98)		(3.25)
Age		4.68		0.17		-1.61
		(3.28)		(4.31)		(2.74)
Tenure Dummies	No	Yes	No	Yes	No	Yes
Observations	575	568	398	394	398	394
R-squared	0.08	0.13	0.03	0.09	0.00	0.08

Panel B: Effect of $50 Incentives

Dep Var:	Miles Prediction		Overconfidence		Absolute Error	
	(1)	(2)	(3)	(4)	(5)	(6)
$50 Incentive	-6.48	1.79	-102.00	-44.68	21.37	75.38
	(65.18)	(86.68)	(100.00)	(124.68)	(76.94)	(114.15)
Trucking experience (yrs)	-4.43		10.95		8.03**	
	(6.13)		(6.63)		(3.26)	
Age	0.32		-12.82		3.75	
	(5.77)		(7.74)		(4.14)	
Tenure Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	No	Yes	No	Yes	No	Yes
Observations	152	152	124	124	124	124
R-squared	0.30	0.73	0.22	0.54	0.07	0.34

Notes: This table reports the effect of providing incentives for accuracy on agents' productivity forecasts using OLS regressions. The field experiment was conducted with Firm B. In "Miles Prediction," the dependent variable is the productivity expectation. In "Overconfidence," the dependent variable is the productivity expectation minus the realized miles. In "Absolute Error," the dependent variable is the absolute value of the productivity expectation minus the realized miles. The sample in Panel A is restricted to the weeks before debiasing. Standard errors are clustered by driver in parentheses. The table shows that there does not appear to be much of a difference between incentivized and non-incentivized beliefs. This suggests that the overconfidence results with Firm A are not explained away by lack of incentives.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.3: Field Experiment: Effect of Debiasing on Productivity Forecasts

	Panel A: Dep Var = Prediction - Avg Pre-Debias Prediction				
Week relative to debiasing	w_0	w_1	w_2	w_3	w_{4-6}
	(1)	(2)	(3)	(4)	(5)
Debiasing	-231.58***	-346.55***	-140.46	-49.36	-145.59
	(84.64)	(115.23)	(102.10)	(97.13)	(95.18)
Incentive	104.53	264.05**	-56.68	-42.80	-78.28
	(79.17)	(101.91)	(94.28)	(84.77)	(83.70)
Exper (yrs)	4.93	2.16	-1.84	-2.01	5.81
	(4.29)	(5.87)	(4.74)	(4.86)	(6.26)
Observations	107	82	68	66	132
R-squared	0.33	0.43	0.47	0.64	0.30

	Panel B: Dep Var = Prediction				
Week relative to debiasing	w_0	w_1	w_2	w_3	w_{4-6}
	(1)	(2)	(3)	(4)	(5)
Debiasing	-206.92**	-320.22***	-141.29	-70.83	-143.40
	(82.46)	(117.37)	(98.14)	(96.87)	(89.49)
Incentive	88.97	233.13**	-67.46	-60.54	-82.39
	(76.85)	(105.38)	(90.76)	(84.45)	(80.11)
Exper (yrs)	5.1	1.31	-2.85	-2.22	5.09
	(4.15)	(5.91)	(4.58)	(4.80)	(6.00)
Pre-debias avg prediction	0.77***	0.86***	0.77***	0.84***	0.83***
	(0.09)	(0.12)	(0.11)	(0.11)	(0.10)
Observations	107	82	68	66	132
R-squared	0.62	0.64	0.72	0.82	0.66

Notes: This table reports the impacts of the debiasing experiment with Firm B using OLS regressions. Standard errors are clustered by driver in parentheses. The week relative to debiasing refers to the number of weeks since the debiasing information intervention was provided. Thus, w_0 refers to the week of debiasing, w_1 refers to one week after debiasing, ..., with w_{4-6} referring to 4-6 weeks after debiasing. Weeks 4-6 are combined together due to smaller sample size for these weeks. All specifications include dummies for days not worked and weeks of tenure. Incentive refers to whether a driver had incentivized guessing (up to $10 per week). The table shows that the debiasing treatment substantially reduced agents' productivity forecasts for a few weeks, but that the effect appears to fade afterward. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.4: Field Experiment: Effect of Debiasing on Search Intention and Job Satisfaction

	(1) Search Intention	(2) Job Satisfaction
Debiasing	-0.05	0.01
	(0.26)	(0.22)
Incentives	-0.19	0.01
	(0.26)	(0.22)
Exper (in years)	0.02	-0.03***
	(0.01)	(0.01)
Observations	99	319

Notes: This table reports the reports of the debiasing experiment with Firm B. Standard errors are clustered by driver in parentheses. The models are ordered probits. The question about search intention was asked only once, coming 1-3 weeks after debiasing. The question about job satisfaction was asked multiple weeks. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.5: The Determinants of Responding to the Belief Questions and of Productivity Beliefs

	Respond		Beliefs					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
L. Avg miles			0.933***		0.724***	0.518***	0.455***	0.467***
			(0.085)		(0.081)	(0.086)	(0.064)	(0.080)
Avg miles	0.011			1.023***				
	(0.001)***			(0.088)				
L. Avg miles * Tenure						0.005***		-0.000
						(0.002)		(0.001)
Miles		0.007			0.247***	0.241***	0.175***	0.175***
		(0.000)***			(0.019)	(0.019)	(0.016)	(0.016)
Black	-0.035		-0.225	-0.201	-0.366	-0.481		
	(0.028)		(0.996)	(0.989)	(0.973)	(0.994)		
Hispanic	-0.159***		-3.364*	-3.559**	-3.371**	-3.303*		
	(0.033)		(1.863)	(1.767)	(1.641)	(1.806)		
Female	-0.062**		1.038	1.150	1.063	1.070		
	(0.028)		(1.400)	(1.415)	(1.405)	(1.399)		
Married	0.029*		-0.285	-0.254	-0.277	-0.305		
	(0.017)		(0.643)	(0.640)	(0.631)	(0.630)		
Age when start	0.004***		0.038	0.036	0.031	0.030		
	(0.001)		(0.029)	(0.028)	(0.027)	(0.027)		
Schooling	0.008		-0.086	-0.096	-0.139	-0.126		
	(0.006)		(0.223)	(0.222)	(0.214)	(0.214)		
Driver FE	No	Yes	No	No	No	No	Yes	Yes
Observations	40,258	39,023	8,713	8,713	8,614	8,614	8,614	8,614
R-squared	0.07	0.30	0.17	0.19	0.20	0.21	0.63	0.63

Notes: This table presents OLS regressions, where an observation is a driver-week. In columns 1-2, the dependent variable is responding to the belief survey. In columns 3-8, the dependent variable is a driver's expectation of his productivity in the next week in hundreds of miles. Standard errors clustered by driver in parentheses. All regressions include weekly fixed effects. Drivers who are white, older, and more productive are more likely to respond to the survey. Beliefs about future productivity respond strongly to increases in average productivity. All drivers are from the same training school and were hired in late 2005 or 2006. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.6: Who is Overconfident About their Productivity?

	(1)	(2)
Average miles	-0.333***	-0.313***
	(0.068)	(0.099)
Black	-1.837	-2.039
	(1.229)	(1.392)
Hispanic	-1.306	-0.417
	(3.206)	(3.518)
Female	-0.532	-0.250
	(1.317)	(1.544)
Married	-0.431	-0.490
	(0.763)	(0.865)
Age	-0.054	-0.057
	(0.036)	(0.041)
Schooling	-0.243	-0.078
	(0.244)	(0.276)
Constant	16.145***	8.109
	(3.368)	(9.298)
Tenure FE	No	Yes
Observations	556	556
R-squared	0.06	0.22

Notes: This table reports OLS regression using values averaged over drivers. The dependent variable is overconfidence about productivity, defined as a worker's actual miles in week $t+1$ (next week) minus his forecast of next week's miles in week t (the current week). The dependent variable is given in terms of hundreds of miles. Standard errors clustered by driver in parentheses. More productive drivers are less overconfident. In addition, female, older, and minority drivers are less overconfident, but the effect is not statistically significant. All drivers are from the same training school and were hired in late 2005 or 2006.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2.7: The Persistence of Overconfidence and Evidence for Adaptive Expectations

Panel A: All Drivers

	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Overconfidence (from last survey)	0.329***	0.327***	-0.024			
	(0.039)	(0.040)	(0.023)			
Lagged Overconfidence (from last week)				0.272***	0.268***	-0.078**
				(0.043)	(0.043)	(0.033)
Constant	3.019***	6.691	9.118***	2.868***	7.315	4.290***
	(0.258)	(4.808)	(0.470)	(0.218)	(4.870)	(0.094)
Tenure FE	No	Yes	No	No	Yes	No
Demographic Controls	No	Yes	No	No	Yes	No
Education Controls	No	Yes	No	No	Yes	No
Work Type Controls	No	Yes	No	No	Yes	No
Subject FE	No	No	Yes	No	No	Yes
Observations	7,740	7,740	7,740	4,445	4,445	4,445
R-squared	0.11	0.13	0.36	0.07	0.10	0.37

Panel B: Eliminate "Team" Drivers

	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Overconfidence (from last survey)	0.171***	0.147***	-0.070***			
	(0.025)	(0.026)	(0.023)			
Lagged Overconfidence (from last week)				0.117***	0.089***	-0.140***
				(0.031)	(0.032)	(0.029)
Constant	2.620***	3.534	3.323***	2.690***	-0.491	-4.419***
	(0.234)	(4.110)	(0.066)	(0.275)	(3.891)	(0.475)
Tenure FE	No	Yes	No	No	Yes	No
Demographic Controls	No	Yes	No	No	Yes	No
Education Controls	No	Yes	No	No	Yes	No
Work Type Controls	No	Yes	No	No	Yes	No
Subject FE	No	No	Yes	No	No	Yes
Observations	7,019	7,019	7,019	4,044	4,044	4,044
R-squared	0.03	0.06	0.25	0.01	0.06	0.28

Notes: The dependent variable is overconfidence about productivity, defined as a worker's actual miles in week $t+1$ (next week) minus his forecast of next week's miles in week t (the current week). An observation is a driver-week. Standard errors clustered by driver in parentheses. Demographic controls include gender, race dummies, marital status, and age bin dummies for the different age groups: 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, and 60-80. Education controls are dummies for high school graduate, some college, and college. Work type controls are dummies for different work configurations and for receiving any salary or activity-based pay. Productivity is given in terms of hundreds of miles driven per week. All drivers are from the same training school and were hired in late 2005 or 2006. * significant at 10%; ** significant at 5%; *** significant at 1%

Chapter 3

The Value of Referrals

3.1 Introduction

An important question for firms is how to hire. Human capital is one of the most valuable assets of modern organizations and is a critical determinant of organizational performance. For firms to achieve an excellent workforce, it is important that they hire well.

This paper seeks to examine to what extent firms should rely on employee referrals in their hiring decisions and how to design hiring and referral systems. Social networks appear to play an important role in job-finding, with roughly 50 percent of jobs found through a friend or family member. A growing literature has begun to explore the implications of social networks for job-finding, wage patterns, and employment outcomes, both theoretically and empirically. For example, Calvo-Armengol and Jackson (2004) argue that social networks can have important effects on wage differentials and lifetime inequality. Given the prevalence of social networks as a means of job-finding, how should firms respond?

Economic intuitions on the value of referrals seem mixed. On the one hand, making referrals a prominent part of hiring may be deleterious and subject to nepotism. Workers may refer close friends or family members even if they are not well-suited for the job. On the other hand, social networks may serve positive roles. They may provide a mean of increased monitoring in order to combat moral hazard. Or if current workers at the firm have superior information about potential hires, they may help locate higher-ability workers or workers who are better-suited to the job. Finally, hiring via referrals may be beneficial if people enjoy or are more productive working in an organization with their friends.

I develop a simple model of referrals that formalizes these intuitions. Existing workers may refer new workers, but suffer a loss for referring workers who are bad matches. If existing workers have information about match, they will be more likely to refer workers who are a better match, and after starting, these workers will be more likely to stay. In addition, after starting, the presence of an existing worker at the firm can help incentivize effort if existing workers help monitor new workers.

To test the model, I use very detailed data on tens of thousands of workers from a leading firm in the trucking industry. At the firm, roughly 20% of workers report being hired via referral. On many dimensions, workers who are referred are similar to workers who are not referred. In addition to standard demographic information for all drivers, extremely detailed data on driver background and preferences are available for a subset of the workers. On most of these measures, referred and non-referred workers are similar. On the basic of overall background characteristics, referred drivers do not seem to be advantageously or adversely selected.

I show, though, that there are large differences in the retention of referred and non-referred workers. Both in basic regressions and in specifications with many controls, referred workers are 10-25% less likely to quit. Effects are present across both brand-new truckdrivers who receive training from the firm and experienced truckdrivers, and are strong across the business cycle. The effect is significant both statistically and economically. The effect of being a referred driver on quitting is similar to that from a two-percentage point increase in the driver's home unemployment rate and almost as large as the drop in quitting from assessing a $3,500 penalty for quitting in the first 12 months.

An advantage of using data from the trucking industry is that one can directly assess how productivity differs between referred and non-referred drivers. In long-haul trucking, drivers are paid almost exclusively by the number of miles driven. I show that referred workers achieve very similar miles to non-referred drivers. In addition, the data contain other measures of performance including excessive speeding, the share of work time spent working, and measures of the quality of communication between workers and their bosses. Across all of these measures, referred and non-referred workers are very similar.

These results seem to suggest that referrals shape behavior by selecting workers rather than affecting moral hazard. To address this issue more directly, I use survey data collected from workers after they started work about friend formation. Unlike one's social connection upon hiring, the number of friends formed after the start of work has no effect on quitting (or on productivity or moral hazard). I rule out several other explanations exploiting the richness of the survey data. For example, it is not the case that referred drivers are more overconfident or achieve greater happiness from work than non-referred workers, or that differences in quitting are being driven by differences in the outside option. Rather, referred drivers report better dimensions of match quality. They report greater satisfaction with the amount of time spent at home and are less likely to report that the job is interfering with their family life, factors which are important given the nature of life as a long-haul truckdriver.

It appears that the referral process appears to generate selection based on match or taste for the job. To quantify the difference in taste for the job induced by referrals, and to measure profits and welfare, I estimate a structural model of quitting.

My paper complements a few other recent papers analyzing the performance of referred vs. non-referred workers. Heath (2011) analyzes referred workers using data on garment workers in Bangladesh. Brown et al. (2012) analyze performance among salaried workers in the US – their data are advantageous in that it allows them to analyze workers across lower-tier, mid-level and high-level position. In contrast, the data I use focus on a very large number of homogenous workers. As such, I can precisely estimate the impact of referrals on turnover, productivity, and moral hazard, but can much less about the generalizability of my results. In the sociology literature, Fernandez et al. (2000) analyze differences between referred and non-referred workers at a call center. More generally, my paper adds to a growing literature analyzing the importance of social networks for worker and firm performance (Dustmann et al., 2012; Bayer et al., 2008; Wahba and Zenou, 2005).

3.2 Conceptual Model

Consider a worker employed at a firm who randomly meets prospective new workers. During these meetings, the worker learns about the employee's match. For the worker, referring a worker provides some combination of monetary and/or social benefit. However, there is also a potential cost to making a referral. Specifically, if the new worker ends up being a low match (or quitting), the old worker suffers a prestige penalty. If the worker applies to the job and is accepted.

Formally, the timing of events is as follows:

- with some probability the worker W has contact with the prospective new worker N. If the two have contact, N's match quality $m \in \{0,1\}$ is observed by W
- W makes a decision $r \in \{0,1\}$ whether to refer N to apply to the firm
- a first period of work occurs during which m is realized to the worker. After this period of work, N decides whether to quit
- if the worker stays on to a second period, he decides on his effort $e \in \{0,1\}$

The new worker's utility if he accepts the job is $u_1^N = w - (1-m)k$ in the first period and $u_1^N = w - (1-m)k - k_r r(1-e) - c_e e$ in the second period. That is, in the first period, the worker is paid a wage w, but suffers a penalty k if the match withe job is poor. In addition, in the second period, the worker pays a cost c_e for exerting effort; if he doesn't exert effort, he also suffers a disutility k_r if he obtained the job via a referral. The idea is that a worker suffer disutility from exerting low effort if the worker he referred him is still at the firm. If the new worker does not accept the job, he earns an outside payoff of zero. In deciding whether to refer the new worker, the old worker receives a benefit b for making the referral, but suffers a cost c_{br} if the worker ends up being a bad match. Formally, the old worker has utility $b - (1-m)c_{br}$ if he decides to make a referral, so chooses to make a referral as long as $b > (1-m)c_{br}$. I assume that $b < c_{br}$ such that the firm will not wish to refer bad workers.

Solving the Model. The game is solved backwards. First, conditional on deciding not to quit, the new worker needs to decide whether to exert effort. If the worker was not referred his utility from exerting effort is $-c_e$ compared to zero for not exerting effort, so a non-referred worker never exerts effort. In contrast, a referred worker will exert effort if $k_r > c_e$. Thus, if $k_r > c_e$, referred workers will exert more effort. Before choosing effort, the new worker chooses whether to quit. If he was referred, he chooses to stay if $w-(1-m)k-min\{k_r, c_e\} > 0$ whereas if he was not referred, he chooses to stay if $w - (1-m)k > 0$. At the stage before, conditional on the old worker and the new worker having made contact, the old worker will choose to refer iff $m = 1$. Among workers who are referred, thus, all will have $m = 1$, so referred workers will stay if $w - min\{k_r, c_e\} > 0$. In contrast, half of non-referred workers will choose to stay if $w > 0$ and half will stay if $w - k > 0$.

Proposition 1. *Suppose that the cost of a bad match for a worker is large and that the cost of effort or social pressure is relatively low. (Formally, assume that $w - min\{k_r, c_e\} > 0$ and that $w - k < 0$.) Then, referred workers will be more likely to stay than non-referred workers. Second, assume that the cost of social pressure is larger than the cost of effort. (Formally, assume that $k_r > c_e$.) Then, referred workers will exert more effort than non-referred workers.*

3.3 Data

The data are from a large long-haul US trucking firm. Trucking is a large occupation in the U.S., employing over 3 million workers. Truckers work in for-hire firms (i.e. trucking firms) or for private firms. In addition, for-hire firms are further divided into long-haul or short-haul firms. The firm studied here is a long-haul firm. Workers are non-union and paid almost exclusively by piece rate. Further information on the firm can be found Burks et al. (2009) and Hoffman (2011).

I use data on all drivers at the firm who start between 2001 and 2009. For these workers, data are available on weekly productivity, quits, and a number of background characteristics. Very detailed personnel and survey data are available for roughly 1,000 new drivers who start at the firm in late 2005 and 2006, with the data collected by Burks et al. (2009). These workers were interviewed extensively during their training with the firm. In addition, they were surveyed every week about their productivity expectations and happiness. Drivers and the family members of the driver were also interviewed at several times during the driver's first two years with the company, and drivers and family members were also surveyed upon driver exit. This information is highly useful in analyzing the degree of match between referred and non-referred drivers.

3.4 Descriptive Results

Characteristics. Table 3.1 summarizes the characteristics of referred and non-referred drivers. Roughly 20% of the workers at the firm are referred out of those for whom information is available as to how they found out about the job. Referred drivers look similar to non-referred drivers on most characteristics. One difference is that referred drivers are more likely to be female. Part of this reflects that a decent percentage of female drivers drive as "team drivers" with their husband. In addition, non-referred drivers are more likely to apply online. Indeed, many of the non-referred drivers report that they found out about the job through the internet.

Table 3.1 also provides information about drivers in the data subset. Non-referred drivers have 13 years of schooling, whereas referred drivers have 12.8 years of schooling, a difference which is not statistically significant. Both referred and non-referred drivers have very poor credit scores on average. Via survey questions asked during the training school, it is seen that referred and non-referred workers have similar work backgrounds. Non-referred and referred workers report having had 1.6 and 1.7 jobs, respectively, in the last two years. One difference is that referred workers report having slightly lower income had they continued at their past jobs.[1] This is consistent with the finding of Loury (2006), who argues using NLSY data that referred workers may have slightly worse outside options. As I show later, this difference in outside appears to explain a very small portion of the large differences in quit rates.

One can imagine many ways in which workers who are referred may differ from workers who are not referred. For example, referred workers might be more likely to stay because they are more patient, because they have a greater risk tolerance for the weekly swings in truckdriver income, or because they are more altruistic. As can be seen in the data, referred and non-referred workers appear similar across most of these measures. One difference is that referred workers are more likely to trust in a sequential prisoner's dilemma that non-referred workers. While the data provide (possibly imperfect) measurements of many previously unobservable characteristics, it is possible, of course, that referred and non-referred workers differ from one another in other unobservable dimensions.

Quitting. Table 3.3 analyzes whether referred workers are less likely to quit. Specifically, it analyzes Cox Proportional Hazard models of the form:

$$\begin{aligned} log(h_{it\tau cs}) &= \alpha_t + \beta_0 * REFERRAL_i + \beta_1 * 12MCONTRACT_{sc} + \beta_2 * 18MCONTRACT_{sc} \\ &+\beta_3 * UNEMP_{s\tau} + \beta_4 \bar{y}_{it} + \gamma_\tau + \delta_c + \theta_s + X_i\lambda + \epsilon_{it\tau cs} \end{aligned} \quad (3.1)$$

where $h_{it\tau cs}$ is the quit hazard of driver i with t weeks of tenure in year τ who is part of cohort (year of hire) c who attended training school s; $REFERRAL_i$ is a dummy for whether driver i is referred, $UNEMP_{s\tau}$ is the unemployment rate in state s at time τ; $\bar{y}_{it}$ is average productivity to date; α_t is a fixed effect for tenure t; γ_τ is a time fixed effect; δ_c is a fixed effect for year of hire c; θ_s is a school fixed effect; X_i are individual covariates; and $\epsilon_{it\tau cs}$ is an error. The main coefficient of interest is β_0 expressing the difference in quit rate between referred and non-referred drivers. The estimate range from $-.10$ to $-.16$, implying that referred drivers between 10% and 16% less likely to quit at any given time than non-referred drivers. This effect is similar in magnitude to a 2-3 percentage point increase in the driver's home state unemployment rate.

Table 3.4 repeats the same exercise using all drivers instead of only inexperienced drivers. The impact of referrals is similar in the broad sample. Table 3.5 analyzes whether the

[1] Specifically, drivers were asked: "Which range best describes the annual earnings you would normally have expected from your usual jobs (regular and part-time together), if you had not started driver training with [Firm A], and your usual jobs had continued without interruption?" The answers were $10,000 increments from $0-$10,000, $10,000 to $20,000, and so on, up to more than $70,000.

difference between referred and non-referred drivers in quitting varies with different factors. For example, does the difference between referred and non-referred drivers differ with the unemployment rate, worker tenure, or whether the worker is bound with a quit penalty contract? There is little evidence this is the case. For example, in column 1, the coefficient on referred is -0.19, and the coefficient on the interaction of being a referred driver with the state unemployment rate is 0.01 (standard error of 0.015), meaning that the difference between referred and non-referred drivers quitting is 1 percent less in magnitude for each additional point increase in a state's unemployment rate. That is, referred drivers would be 15 percent less likely to quit when the unemployment rate is 4 percent, but 13 percent less likely to quit when the unemployment rate is 6 percent.

Productivity. Table 3.6 analyzes differences in productivity between referred and non-referred drivers. I show results exclusively for inexperienced workers and for all workers. There is no evidence that referred workers are more productive. Using all the data, it appears actually to be the case that referred workers are slightly less productive, by around 25-30 miles per week, which is about 1 to 1.5% of productivity. However, once the zero mile weeks are eliminated, the difference between referred and non-referred worker productivity tends to very close to zero.

Moral Hazard. Table 3.7 analyzes differences in moral hazard behavior between referred and non-referred workers. In addition to driver miles, there are several other dimensions of driver performance that are important for firms in the trucking industry. The data include measurements on driver speeding, the percent of work time spent actually working, and the degree to which drivers send messages back to their supervisors. If referred workers exert greater effort due to increased monitoring or social pressure from the referring drive, one would expect that referred workers would exhibit lower levels of moral hazard. However, there is no evidence in Table 3.7 to support this view.

More generally than these particular measures of moral hazard, it appears that referred drivers are differentially selected instead of affected by the presence of the referring driver at the firm. Data on friendships formed after the driver starts provide a more direct way of assessing this claim. After one week in training, drivers were asked about the number of friends formed since the start of training. There is substantial variation in the number of friends formed, with some drivers reporting zero or one friends, and other drivers twenty friends or more. I examine how the number of friends formed after training affects quitting, productivity, and moral hazard in Table 3.8. As can be see, there is no evidence suggesting that new friends improve worker performance on these dimensions; if anything, drivers who have more friends are slightly more likely to quit and have slightly lower productivity.

3.4.1 Alternative Explanations

Worse Outside Option. I argue that referrals improve quitting by selecting workers who are better matches for the job. Another possibility is that drivers who are referred have lower outside options and are thus more likely to stay (Loury, 2006). As I showed in Table 3.1, referred drivers do indeed report slightly lower outside options. To examine the importance of difference in outside options for observed quitting differences, I consider quitting models both with and without controls for the outside option, as seen in Table 3.10. Controlling for the outside option has a very small impact on the size of referral dummy, though there estimates are somewhat imprecise.

Differences in Happiness or beliefs. Another possibility is that referred and non-referred drivers could differ in their beliefs. Using the same set of workers in the data subset, I show in Hoffman (2011) that workers who are more overconfident about their productivity are less likely to quit. However, as seen in Table 3.9, it is seen that referred and non-referred workers do not exhibit different average beliefs about their productivity or a different average level of overprediction (beliefs minus actual productivity). Further, referred and non-referred workers also exhibit similar weekly happiness.

3.4.2 Survey-based Evidence on Match

Survey evidence is also useful for illuminating whether referred workers are being selected based on their fit for the job. In their continuing driver survey, workers were asked a number of questions about various aspects of the job and work-life seemed to them compared to what they expected. Since truckers work away from home for long periods of time, there were particularly a good number of questions about driver and familial satisfaction with the driver being away from home. If referrals help select people who are better suited to being away from home, then referred drivers may score hire on these satisfaction variables, and these variables may help explain the difference in quitting between referred and non-referred workers.

In Table 3.11, I first consider whether referral status helps predict various measures of job match. Referred workers are less likely to feel that the demands of the job interfered with their family life, less likely to feel bothered by an unexpectedly low paycheck, and more likely to be believe that they were home an acceptable number of times per month. It should be noted, however, that the response to the continuing driver survey is somewhat, and there are only around 220 respondents for each question. I then show that including these variables in a Cox quitting model reduces the magnitude of the coefficient on the referred dummy. Specifically, I consider Cox models of the form:

$$log(h_{i,t}) = \alpha + \beta REFERRAL_i + SURVEYMATCH_i\gamma + X_i\delta + \epsilon_{i,t}, \qquad (3.2)$$

The coefficient β on the referral variable becomes smaller once the vector of survey questions measuring worker match is introduced.[2]

[2]One could imagine that having a friend at work could also affect a worker's perception of whether work is interfering with their family life, as opposed to merely selecting workers whose family lives are amenable to trucking.

3.5 Conclusion

This paper aims to examining whether hiring through referrals is profitable for firms, and if so, how firm should structure their referral program. I find that referred workers yield significantly more profits on average than non-referred workers. Referrals appear to play a substantial role in generating better matches between workers and firms.

Table 3.1: Comparing Referred and Non-Referred Workers

	All Workers			Inexperienced		Experienced	
Panel A: Full Data	No Referral	Referral		No Referral	Referral	No Referral	Referral
African-American	0.15	0.14		0.15	0.13	0.15	0.15
Hispanic	0.05	0.05		0.05	0.05	0.06	0.06
Female	0.07	0.13		0.07	0.16	0.05	0.08
Married	0.37	0.38		0.39	0.38	0.35	0.39
Age when start	38.8	38.7		37.1	36.7	41.0	41.2
Online application	0.72	0.55		0.78	0.63	0.68	0.48
Smoker	0.34	0.37		0.38	0.42	0.3	0.31
Miles	1653	1680		1647	1656	1663	1713
Miles, miles greater than zero	1915	1940		1947	1951	1868	1926
Earnings	754	783		716	729	815	857
System driver	0.44	0.37		0.46	0.37	0.43	0.36
Team driver	0.09	0.19		0.10	0.20	0.07	0.17
Dedicated driver	0.25	0.26		0.25	0.25	0.25	0.27
Number in each	24380	6102		12911	3134	10464	2684
Share in each	0.80	0.20		0.80	0.20	0.80	0.20
Panel B: Data Subset	No Referral	Referral					
Demographics:							
African-American	0.09	0.08					
Hispanic	0.02	0.01					
Female	0.08	0.19	*				
Married	0.42	0.40					
Age when start	36.01	36.69					
Number of kids	1.05	0.90					
Online application	0.84	0.64	*				
Smoker	0.42	0.47					
Years of schooling	12.97	12.78					
High school dropout	0.04	0.03					
High school graduate	0.36	0.46	*				
Some college	0.36	0.31					
Technical school	0.14	0.13					
College degree or more	0.11	0.07					
Credit score	588	584					
No credit score	0.12	0.10					
Cognitive Ability and Preferences:							
IQ (std deviations)	-0.01	-0.11					
Numeracy (std deviations)	-0.01	-0.19					
CRRA risk aversion V1	0.48	0.46					
CRRA risk aversion V2	0.11	0.07					
Patient option chosen (share)	0.60	0.58					
Beta from HB Model, TUnit=Day	0.84	0.83					
Delta from HB Model, TUnit=Day	0.98	0.97					
Trust (P1 sending, SPD Game)	3.41	2.90	*				
Altruism_V1 (P2 return, SPD Game)	1.52	1.55					
Altruism_V2 (P2 return, SPD Game)	3.74	3.46					
Work background:							
Regular jobs in last 2 years	1.56	1.69					
Annual income if continued at past jobs	31.05	27.90	*				
Maximum years at a previous job	7.91	7.97					
# months holding reg jobs in last 2 years	17.73	18.37					
Parent worked in trucking	0.12	0.15					
Trainees known before training	0.15	0.19					
Friends made since training	5.45	6.33					
Exper (yrs) w/large onroad vehicle	1.12	1.09					

* $p<0.05$

Table 3.2: Correlates of Having Been Referred

VARIABLES	(1) referral	(2) referral1
Rookie driver, getting basic training	-0.043***	-0.023***
	(0.007)	(0.004)
Unemployment rate, county	0.008***	0.005***
	(0.002)	(0.002)
Unemployment rate, state	-0.001	0.000
	(0.004)	(0.003)
African-American	-0.009	-0.014***
	(0.007)	(0.005)
Hispanic	-0.016	-0.019**
	(0.012)	(0.007)
Female	0.113***	0.078***
	(0.012)	(0.008)
Married	0.004	0.007**
	(0.006)	(0.004)
Age when start	-0.000	-0.001***
	(0.000)	(0.000)
Observations	21,710	34,715
R-squared	0.04	0.04

VARIABLES	(1) referral	(2) referral1
Rookie driver, getting basic training	-0.048***	-0.022***
	(0.009)	(0.006)
miles_10k	-0.022	-0.018
	(0.021)	(0.013)
Unemployment rate, county	0.009***	0.005**
	(0.003)	(0.002)
Unemployment rate, state	-0.008	-0.003
	(0.005)	(0.003)
African-American	-0.004	-0.009
	(0.012)	(0.007)
Hispanic	-0.017	-0.016
	(0.018)	(0.011)
Female	0.140***	0.104***
	(0.017)	(0.013)
Married	0.014	0.010*
	(0.008)	(0.005)
Age when start	-0.000	-0.001***
	(0.000)	(0.000)
R-squared	0.04	0.04

Notes: This table examines correlates of which new workers are referred by workers at the company versus those who are not. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3.3: Referred Drivers are Less Likely to Quit: Inexperienced Drivers

	(1)	(2)	(3)	(4)	(5)
Referral	-0.163***	-0.161***	-0.105***	-0.102***	-0.127***
	(0.029)	(0.029)	(0.032)	(0.034)	(0.039)
12m contract	-0.410***	-0.380***	-0.355***	-0.344***	-0.402***
	(0.091)	(0.091)	(0.100)	(0.103)	(0.111)
18m contract	-0.283***	-0.268***	-0.207*	-0.216*	-0.257**
	(0.102)	(0.102)	(0.111)	(0.115)	(0.121)
State unemployment rate		-0.060***	-0.051***	-0.064***	-0.042**
		(0.015)	(0.017)	(0.017)	(0.019)
Avg miles to date				-0.065***	-0.054***
				(0.004)	(0.005)
Time FE (yr)	Yes	Yes	Yes	Yes	Yes
Cohort FE (yr of hire)	Yes	Yes	Yes	Yes	Yes
Training School FE	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	No	No	Yes
Observations	840,596	839,840	748,601	748,601	670,984
	(6)	(7)	(8)	(9)	(10)
Referral1	-0.156***	-0.153***	-0.088***	-0.079**	-0.113***
	(0.027)	(0.027)	(0.030)	(0.031)	(0.036)
12m contract	-0.178***	-0.172***	-0.208***	-0.202***	-0.181***
	(0.046)	(0.046)	(0.052)	(0.055)	(0.058)
18m contract	-0.109*	-0.118*	-0.116*	-0.127*	-0.107
	(0.062)	(0.062)	(0.069)	(0.071)	(0.074)
State unemployment rate		-0.054***	-0.048***	-0.062***	-0.048***
		(0.011)	(0.013)	(0.013)	(0.014)
Avg miles to date				-0.060***	-0.048***
				(0.003)	(0.004)
Time FE (yr)	Yes	Yes	Yes	Yes	Yes
Cohort FE (yr of hire)	Yes	Yes	Yes	Yes	Yes
Training School FE	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	No	No	Yes

Notes: This table shows that referred drivers are less likely to quit in a Cox Proportional Hazards Model. The sample is restricted to inexperienced drivers. Standard errors clustered by driver in parentheses. 'Referral' means driver who found job via a referral versus those who found the job in other ways, whereas 'Referral1' means drivers who found the job via a referral versus everyone else (including if the how found a job variable is missing).

Table 3.4: Referred Drivers are Less Likely to Quit: All Drivers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Referral	-0.132***	-0.134***	-0.088***	-0.094***	-0.092***		
	(0.022)	(0.022)	(0.024)	(0.025)	(0.028)		
Inexperienced	0.088	0.064	0.071	0.073	0.134*	0.064	0.134*
	(0.058)	(0.058)	(0.066)	(0.069)	(0.076)	(0.058)	(0.076)
12m contract	-0.226***	-0.203***	-0.245***	-0.270***	-0.379***	-0.203***	-0.380***
	(0.060)	(0.059)	(0.066)	(0.069)	(0.077)	(0.059)	(0.077)
18m contract	-0.219***	-0.199***	-0.228***	-0.258***	-0.398***	-0.199***	-0.399***
	(0.067)	(0.067)	(0.075)	(0.077)	(0.083)	(0.067)	(0.083)
State unemployment rate		-0.045***	-0.043***	-0.052***	-0.037***	-0.045***	-0.037***
		(0.011)	(0.013)	(0.013)	(0.014)	(0.011)	(0.014)
Avg miles to date				-0.052***	-0.045***		-0.045***
				(0.003)	(0.003)		(0.003)
Referral * (wks≤52)						-0.138***	-0.096***
						(0.026)	(0.032)
Referral * (52<wks≤78)						-0.081	-0.098
						(0.058)	(0.070)
Referral * (wks>78)						-0.166***	-0.071
						(0.059)	(0.069)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Referral1	-0.091***	-0.093***	-0.040*	-0.044*	-0.051*		
	(0.021)	(0.021)	(0.023)	(0.023)	(0.026)		
Inexperienced	0.056	0.047	0.058	0.043	-0.001	0.049	-0.000
	(0.035)	(0.035)	(0.041)	(0.042)	(0.047)	(0.035)	(0.047)
12m contract	-0.137***	-0.131***	-0.187***	-0.200***	-0.211***	-0.132***	-0.212***
	(0.035)	(0.035)	(0.040)	(0.042)	(0.046)	(0.035)	(0.046)
18m contract	-0.171***	-0.167***	-0.200***	-0.211***	-0.254***	-0.167***	-0.254***
	(0.045)	(0.045)	(0.051)	(0.052)	(0.055)	(0.045)	(0.055)
State unemployment rate		-0.037***	-0.041***	-0.053***	-0.045***	-0.037***	-0.044***
		(0.009)	(0.010)	(0.010)	(0.011)	(0.009)	(0.011)
Avg miles to date				-0.049***	-0.040***		-0.040***
				(0.002)	(0.003)		(0.003)
Referral * (wks≤52)						-0.120***	-0.080***
						(0.024)	(0.031)
Referral * (52<wks≤78)						-0.050	-0.091
						(0.056)	(0.067)
Referral * (wks>78)						-0.001	0.095

Notes: This table shows that referred drivers are less likely to quit in a Cox Proportional Hazards Model. Standard errors clustered by training facility-week of hire in parentheses. ‘Referral’ means driver who found job via a referral versus those who found the job in other ways, whereas ‘Referral1’ means drivers who found the job via a referral versus everyone else (including if the how found a job variable is missing).

Table 3.5: Interaction Effects for Quitting Models: All Drivers

	(1)	(2)	(3)	(4)	(5)
Referral	-0.193**	-0.230***	-0.132***	-0.132***	-0.092***
	(0.091)	(0.077)	(0.041)	(0.041)	(0.028)
Referral * State unemployment rate	0.010				
	(0.015)				
Referral * County unemployment rate		0.018			
		(0.012)			
Referral * Inexperienced			0.014	0.119	
			(0.051)	(0.116)	
Referral * 12m contract				-0.099	
				(0.114)	
Referral * 18m contract				-0.162	
				(0.126)	
Referral * Tenure					-0.001*
					(0.000)
Inexperienced	0.064	0.087	0.086	0.064	0.086
	(0.058)	(0.063)	(0.064)	(0.067)	(0.063)
12m contract	-0.203***	-0.217***	-0.218***	-0.197***	-0.218***
	(0.059)	(0.064)	(0.064)	(0.068)	(0.064)
18m contract	-0.199***	-0.228***	-0.228***	-0.197***	-0.227***
	(0.067)	(0.071)	(0.071)	(0.075)	(0.071)
State unemployment rate	-0.047***				
	(0.012)				
County unemployment rate		-0.012*	-0.009	-0.008	-0.009
		(0.007)	(0.007)	(0.007)	(0.007)

	(1)	(2)	(3)	(4)	(5)
Referral1	-0.115	-0.159**	-0.064*	-0.064*	-0.096***
	(0.086)	(0.074)	(0.038)	(0.038)	(0.027)
Referral * State unemployment rate	0.004				
	(0.014)				
Referral * County unemployment rate		0.013			
		(0.012)			
Referral * Inexperienced			-0.029	0.131	
			(0.047)	(0.109)	
Referral * 12m contract				-0.156	
				(0.107)	
Referral * 18m contract				-0.232*	
				(0.121)	
Referral * Tenure					0.000
					(0.000)
Inexperienced	0.047	0.042	0.045	0.034	0.042
	(0.035)	(0.037)	(0.037)	(0.038)	(0.037)
12m contract	-0.131***	-0.117***	-0.115***	-0.106***	-0.116***
	(0.035)	(0.037)	(0.037)	(0.037)	(0.037)
18m contract	-0.167***	-0.162***	-0.161***	-0.141***	-0.162***
	(0.045)	(0.047)	(0.047)	(0.049)	(0.047)
State unemployment rate	-0.038***				
	(0.009)				
County unemployment rate		-0.007	-0.006	-0.006	-0.006
		(0.005)	(0.005)	(0.005)	(0.005)

Notes: This table analyzes Cox Proportional Hazard models of quitting, looking at interaction effects of Referral status with other variables. Standard errors clustered by training facility-week of hire in parentheses. 'Referral' means driver who found job via a referral versus those who found the job in other ways, whereas 'Referral1' means drivers who found the job via a referral versus everyone else (including if the how found a job variable is missing).

Table 3.6: Referrals and Worker Productivity

Panel A: Inexperienced Drivers	All Weeks		Exclude 0 Mile Weeks		Trim 5/95 %	
	(1)	(2)	(3)	(4)	(5)	(6)
Referral	-33.871***	-22.294	-11.957	-5.030	-9.155	-2.913
	(12.664)	(13.770)	(10.630)	(11.393)	(10.091)	(10.777)
12m contract	14.235	25.605	69.603**	81.130***	64.770**	76.314***
	(34.949)	(34.936)	(31.284)	(31.113)	(29.369)	(29.320)
18m contract	-4.711	4.266	50.451	56.140	47.406	53.078
	(38.894)	(39.418)	(35.124)	(35.085)	(33.232)	(33.318)
Demog Controls	No	Yes	No	Yes	No	Yes
Mean Dep Var						
R-squared	0.26	0.24	0.07	0.07	0.07	0.07
Panel B: All Drivers, Referral	All Weeks		Exclude 0 Mile Weeks		Trim 5/95 %	
	(1)	(2)	(3)	(4)	(5)	(6)
Referral	-30.775***	-25.278**	-3.979	-1.082	-3.225	-0.674
	(10.747)	(11.444)	(8.169)	(8.608)	(7.801)	(8.208)
Rookie driver, getting basic training	-4.820	21.563	-10.771	4.389	-11.397	3.421
	(29.328)	(30.419)	(28.831)	(29.168)	(28.116)	(28.465)
12m contract	-53.376*	-50.666	-7.364	-4.805	-5.516	-3.673
	(30.146)	(30.890)	(29.521)	(29.785)	(28.791)	(29.087)
18m contract	-123.808***	-135.930***	-11.977	-18.619	-7.819	-14.935
	(30.302)	(31.364)	(29.133)	(29.554)	(28.127)	(28.558)
Demog Controls	No	Yes	No	Yes	No	Yes
Mean Dep Var						
R-squared	0.19	0.17	0.08	0.09	0.08	0.08
Panel C: All Drivers, Referral1	All Weeks		Exclude 0 Mile Weeks		Trim 5/95 %	
	(1)	(2)	(3)	(4)	(5)	(6)
Referral1	-32.558***	-25.154**	-2.542	2.217	-2.093	2.201
	(10.389)	(11.164)	(7.873)	(8.340)	(7.515)	(7.944)
Rookie driver, getting basic training	-44.152**	-23.823	-42.392**	-29.958	-40.913**	-29.249
	(21.220)	(21.477)	(18.857)	(18.655)	(18.467)	(18.316)
12m contract	-26.159	-20.098	0.621	2.115	3.208	4.553
	(21.897)	(22.297)	(19.633)	(19.661)	(19.162)	(19.220)
18m contract	-81.543***	-88.835***	12.265	6.201	14.961	8.943
	(21.936)	(22.299)	(19.451)	(19.325)	(18.765)	(18.692)
Demog Controls	No	Yes	No	Yes	No	Yes
Mean Dep Var						
R-squared	0.19	0.17	0.08	0.09	0.08	0.09

Notes: This table examines whether a worker's referral status predicts their productivity in miles. All specifications are OLS regressions with time fixed effects (for each month), cohort fixed effects (by year of hire), tenure fixed effects (by week), work type controls, and the annual state unemployment rate. An observation is a driver-week. "Trim 5/95%" refers to trimming the lowest 5% and highest 5% of the miles observations (ignoring all 0 mile weeks). Standard errors clustered at the driver level in parentheses. Demographic controls include gender, race dummies, marital status, and age bin dummies for the different age groups: 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, and 60-80. Work type controls are dummies for different work configurations and for receiving any salary or activity-based pay. 'Referral' means driver who found job via a referral versus those who found the job in other ways, whereas 'Referral1' means drivers who found the job via a referral versus everyone else (including if the how found a job variable is missing).

Table 3.7: Referrals and Moral Hazard

VARIABLES	(1) over_speed_percentage	(2) working_percentage	(3) Macro4Macro34
Referral	0.162	-0.062	-0.007
	(0.251)	(0.948)	(0.033)
Observations	20,362	20,362	17,042
R-squared	0.00	0.05	0.03

VARIABLES	(1) over_speed_percentage	(2) working_percentage	(3) Macro4Macro34
Referral1	-0.001	0.006	-0.015
	(0.253)	(0.926)	(0.032)
Observations	27,370	27,370	22,929
R-squared	0.00	0.05	0.02

Notes: This table examines whether a worker's referral status predicts their moral hazard behavior at work. All specifications are OLS regressions with demographic controls, education controls, and work type controls. Standard errors clustered at the driver level in parentheses. Demographic controls include gender, race dummies, marital status, and age bin dummies for the different age groups: 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, and 60-80. Education controls are dummies for high school graduate, some college, and college. Work type controls are dummies for different work configurations and for receiving any salary or activity-based pay. All drivers are from the same training school and were hired in late 2005 or 2006.

Table 3.8: Isolating the Treatment Effects of Social Networks While Shutting Down the Selection Effects: The Effect of New Friends at Work on Quitting and Productivity

VARIABLES	(1) _t	(2) _t	(3) _t	(4) _t	(5) _t	(6) _t
Friends made since training	0.002	0.006				
	(0.010)	(0.010)				
Referral1		-0.234		-0.249		-0.228
		(0.160)		(0.167)		(0.160)
Annual income if continued at past jobs		0.011***		0.011***		0.011***
		(0.004)		(0.004)		(0.004)
Log(Friends made since training)			0.044	0.075		
			(0.076)	(0.076)		
Trainees known before training					0.000	0.016
					(0.113)	(0.121)
Observations	29,743	29,670	27,587	27,514	29,743	29,670

VARIABLES	(1) miles	(2) miles	(3) miles	(4) miles	(5) miles	(6) miles
Friends made since training	-7.23	-6.53				
	(6.77)	(4.27)				
Referral1		45.35		45.81		38.12
		(54.33)		(55.84)		(53.80)
Log(Friends made since training)			-117.49***	-73.25***		
			(35.24)	(27.99)		
Trainees known before training					-56.46	-96.06*
					(55.24)	(51.29)
Observations	30,194	30,194	28,042	28,042	30,194	30,194
R-squared	0.20	0.28	0.20	0.28	0.20	0.28

Notes: Standard errors clustered at the driver level in parentheses. This table examines whether a worker's referral status predicts their moral hazard behavior at work. The quitting specifications are are Cox Proportional Hazard models with demographic controls, education controls, and work type controls. Demographic controls include gender, race dummies, marital status, and age bin dummies for the Education controls are dummies for high school graduate, some college, and college. Work type controls are dummies for different work configurations and for receiving any salary or activity-based pay. All drivers are from the same training school and were hired in late 2005 or 2006.

Table 3.9: Referrals, Happiness, and Beliefs

VARIABLES	(1) MeanHappinessThisWk	(2) pmiles	(3) overconf
Referral	-0.024	73.205	30.165
	(0.145)	(61.399)	(55.134)
Avg miles to date		0.539***	
		(0.067)	
Observations	6,386	6,361	6,293
R-squared	0.04	0.34	0.15

VARIABLES	(1) MeanHappinessThisWk	(2) pmiles	(3) overconf
Referral1	-0.025	34.298	0.399
	(0.141)	(58.629)	(52.683)
Avg miles to date		0.548***	
		(0.066)	
Observations	8,741	8,713	8,628
R-squared	0.04	0.31	0.12

Notes: This table examines whether a worker's referral status predicts their moral hazard behavior at work. All specifications are OLS regressions with demographic controls, education controls, and work type controls. Standard errors clustered at the driver level in parentheses. Demographic controls include gender, race dummies, marital status, and age bin dummies for the different age groups: 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, and 60-80. Education controls are dummies for high school graduate, some college, and college. Work type controls are dummies for different work configurations and for receiving any salary or activity-based pay. All drivers are from the same training school and were hired in late 2005 or 2006.

Table 3.10: Is the Quitting Difference Between Referred and Non-Referred Workers Due to Differences in the Outside Option?

VARIABLES	(1) _t	(2) _t
Referral	-0.186	-0.162
	(0.143)	(0.144)
Annual income if continued at past jobs		0.010**
		(0.004)
Observations	28,706	28,706

VARIABLES	(1) _t	(2) _t
Referral1	-0.242*	-0.224
	(0.137)	(0.137)
Annual income if continued at past jobs		0.008**
		(0.003)
Observations	38,381	38,381

Notes: This table examines whether a worker's referral status predicts their moral hazard behavior at work. All specifications are Cox Proportional Hazard models with demographic controls, education controls, and work type controls. Standard errors clustered at the driver level in parentheses. Demographic controls include gender, race dummies, marital status, and age bin dummies for the Education controls are dummies for high school graduate, some college, and college. Work type controls are dummies for different work configurations and for receiving any salary or activity-based pay. All drivers are from the same training school and were hired in late 2005 or 2006.

Table 3.11: Are Referred Workers Better Matched and How Much of the Referral Quit Difference Can They Explain? Evidence Using Survey Measures of Match Quality

	(1)	(2)	(3)
VARIABLES	DrvContS_Q23	DrvContS_Q39	DrvContS_Q4
Referral1	-0.531***	-0.516**	0.473**
	(0.181)	(0.244)	(0.223)
Observations	223	215	226
R-squared	0.04	0.02	0.02

	(1)	(2)
VARIABLES	_t	_t
Referral1	-0.860***	-0.544*
	(0.306)	(0.328)
Feel bothered when received an unexpectedly low paycheck		0.037
		(0.101)
Demands of the job interfered with family life		0.175**
		(0.088)
Acceptable number of times at home per month		-0.178**
		(0.086)
Observations	16,634	14,674

Notes: This table examines whether a worker's referral status predicts survey measures of job match, and in turn, how much of the increased retention from being referred can be explained by these variables. The models are OLS regressions in the top panel and Cox Proportional Hazard models in the bottom panel. Standard errors clustered at the driver level in parentheses. All drivers are from the same training school and were hired in late 2005 or 2006. The survey questions are asked on a 5-point scale from Strongly Disagree (-2) to Strongly Agree (+2).

Bibliography

Acemoglu, Daron, "Training and Innovation in an Imperfect Labour Market," *Review of Economic Studies*, 1997, *64* (3), 445–464.

_ **and Jorn-Steffen Pischke**, "Why Do Firms Train? Theory And Evidence," *Quarterly Journal of Economics*, 1998, *113* (1), 78–118.

_ **and** _ , "Beyond Becker: Training in Imperfect Labour Markets," *Economic Journal*, 1999, *109* (453), F112–42.

_ **and** _ , "The Structure of Wages and Investment in General Training," *Journal of Political Economy*, 1999, *107* (3), 539–572.

Acland, Dan and Matthew Levy, "Habit Formation, Naivete, and Projection Bias in Gym Attendance," 2011. Mimeo, UC Berkeley.

Altonji, Joseph G. and Charles R. Pierret, "Employer Learning And Statistical Discrimination," *Quarterly Journal of Economics*, 2001, *116* (1), 313–350.

Arcidiacono, Peter, "Ability Sorting and the Returns to College Major," *Journal of Econometrics*, 2004, *121* (1-2), 343–375.

Arkes, Hal R., Caryn Christenson, Cheryl Lai, and Catherine Blumer, "Two Methods of Reducing Overconfidence," *Organizational Behavior and Human Decision Processes*, 1987, *39*, 133–144.

Armstrong, Christopher S., David F. Larcker, and Che-Lin Su, "Endogenous Selection and Moral Hazard in Compensation Contracts," *Operations Research*, 2010, *58* (4-Part-2), 1090–1106.

Autor, David H., "Why Do Temporary Help Firms Provide Free General Skills Training?," *Quarterly Journal of Economics*, 2001, *116* (4), 1409–1448.

Baker, George P. and Thomas N. Hubbard, "Contractibility And Asset Ownership: On-Board Computers and Governance In U. S. Trucking," *Quarterly Journal of Economics*, 2004, *119* (4), 1443–1479.

Balmaceda, Felipe, "Firm-Sponsored General Training," *Journal of Labor Economics*, 2005, *23* (1), 115–134.

Barber, Brad M. and Terrance Odean, "Boys Will Be Boys: Gender, Overconfidence, And Common Stock Investment," *Quarterly Journal of Economics*, 2001, *116* (1), 261–292.

Barron, John M., Mark C. Berger, and Dan A. Black, "Do Workers Pay for On-The-Job Training?," *Journal of Human Resources*, 1999, *34* (2), 235–252.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa, "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, December 2008, *116* (6), 1150–1196.

Becker, Gary, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, Chicago, IL: University of Chicago Press, 1964.

Bellemare, Charles and Bruce Shearer, "On The Relevance And Composition Of Gifts Within The Firm: Evidence From Field Experiments," *International Economic Review*, 08 2011, *52* (3), 855–882.

Belman, Dale L., Kristen A. Monaco, and Taggert Brooks, *Sailors of the Conrete Sea: A Portrait of Truck Drivers' Work and Lives*, Lansing, MI: Michigan State University Press, 2005.

Belzer, Michael, *Sweatshops on Wheels: Winners and Losers in Trucking Deregulation*, New York, NY: Oxford University Press, 2000.

Belzil, Christian, J. Hansen, and Nicolai Kristensen, "Estimating Complementarity Between Education and Training," Mimeos, Ecole Polytechnique Mimeo 2009.

Benoit, Jean-Pierre and Juan Dubra, "Apparent Overconfidence," *Econometrica*, forthcoming.

Bergman, Nittai K. and Dirk Jenter, "Employee Sentiment and Stock Option Compensation," *Journal of Financial Economics*, 2007, *84* (3), 667–712.

Blinder, Alan S. and Alan B. Krueger, "Labor Turnover in the USA and Japan: A Tale of Two Countries," *Pacific Economic Review*, 1996, *1* (1), 27–57.

Bojilov, Raicho, "Incentives to Work or Incentives to Quit?," Mimeo, Columbia University 2011.

Brown, Meta, Elizabeth Setren, and Giorgio Topa, "Do Informal Referrals Lead to Better Matches? Evidence from a Firm's Employee Referral System," Mimeo, Federal Reserve Bank of New York 2012.

Brunello, Giorgio and Alfredo Medio, "An explanation of international differences in education and workplace training," *European Economic Review*, February 2001, *45* (2), 307–322.

Buddin, Richard J. and Kanika Kapur, "The Effect of Employer-Sponsored Education on Job Mobility: Evidence from the U.S. Navy," *Industrial Relations*, 2005, *44*, 341–363.

Bureau of Labor Statistics, "Occupational Outlook Handbook, 2010-11 Edition: Truck Drivers and Driver/Sales Workers, http://www.bls.gov/oco/ocos246.htm," 2010.

Burks, Stephen, Jeffrey Carpenter, Lorenz Goette, and Aldo Rustichini, "Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment," *Proceedings of the National Academy of Science*, 2009, *106* (19), 7745–7750.

_ , _ , _ , and _ , "Overconfidence is a Social Signaling Bias," 2010. IZA Working paper.

_ , Michael Belzer, Q. Kwan, S. Pratt, and S. Shackelford, "Trucking 101: An Industry Primer," *Transportation Research Board*, 2011.

Calvo-Armengol, Antoni and Matthew O. Jackson, "The Effects of Social Networks on Employment and Inequality," *American Economic Review*, June 2004, *94* (3), 426–454.

Cappelli, Peter, "Why Do Employers Pay for College?," *Journal of Econometrics*, 2004, *121* (1-2), 213–241.

Card, David and Dean R. Hyslop, "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers," *Econometrica*, 2005, *73* (6), 1723–1770.

Cesarini, David, Magnus Johannesson, Paul Lichtenstein, and Bjorn Wallace, "Heritability of Overconfidence," *Journal of the European Economic Association*, 2009, *7* (2-3), 617–627.

Charness, Gary, Aldo Rustichini, and Jeroen van de Ven, "Overconfidence, Subjectivity, and Self-Deception," 2010. Mimeo, UCSB.

Chiappori, Pierre-Andre and Bernard Salanie, "Testing Contract Theory: A Survey of Some Recent Work," in L. Hansen M. Dewatripont and S. Turnovsky, eds., *Advances in Economics and Econometrics*, 2003.

Ching, Andrew, Tulin Erdem, and Michael Keane, "Learning Models: An Assessment of Progress, Challenges and New Developments," 2011. Mimeo, University of Toronto.

Conlin, Michael, Ted O'Donoghue, and Timothy J. Vogelsang, "Projection Bias in Catalog Orders," *American Economic Review*, 2007, *97* (4), 1217–1249.

Coscelli, Andrea and Matthew Shum, "An Empirical Model of Learning and Patient Spillovers in New Drug Entry," *Journal of Econometrics*, 2004, *122* (2), 213 – 246.

Crawford, Gregory S. and Matthew Shum, "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 2005, *73* (4), 1137–1173.

Crawford, Vincent P and Juanjuan Meng, "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, 2011, *101* (5), 1912–32.

De Bondt, Werner and Richard Thaler, "Financial Decision-making in Markets and Firms: a Behavioral Perspective," in R. A. Jarrow, V. Maksimovic, and W. T. Ziemba, eds., *Handbooks in Operations Research and Management Science*, North Holland, 1995, pp. 385–410.

DeGroot, Morris, *Optimal Statistical Decisions*, New York, NY: Mcgraw-Hill College, 1970.

Delavande, Adeline, "Pill, Patch, Or Shot? Subjective Expectations And Birth Control Choice," *International Economic Review*, 2008, *49* (3), 999–1042.

DellaVigna, Stefano and Ulrike Malmendier, "Contract Design and Self-control: Theory and Evidence," *The Quarterly Journal of Economics*, 2004, *119* (2), 353–402.

_ **and** _ , "Paying Not to Go to the Gym," *American Economic Review*, June 2006, *96* (3), 694–719.

_ **, John List, and Ulrike Malmendier**, "Testing for Altruism and Social Pressure in Charitable Giving," *Quarterly Journal of Eocnomics*, 2012, *Forthcoming.*

Dominitz, Jeff, "Using Expectations Data To Study Subjective Income Expectations," *Review of Economics and Statistics*, 1998, *80* (3), 374–388.

_ **and Charles F. Manski**, "Using Expectations Data To Study Subjective Income Expectations," *Journal of the American Statistical Association*, 1997, *92* (439), 855–867.

Dustmann, Christian, Albrecht Glitz, and Uta Schoenberg, "Referral-based Job Search Networks," 2012. Mimeo, UCL.

_ **and Uta Schoenberg**, "What Makes Firm-based Vocational Training Schemes Successful? The Role of Commitment," *American Economic Journal: Applied Economics*, forthcoming.

Eil, David and Justin Rao, "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 2011, *Forthcoming.*

Eisenberg, Melvin Aron, "The Limits of Cognition and the Limits of Contract," *Stanford Law Review*, 1995, *47* (2), pp. 211–259.

Eliaz, Kfir and Ran Spiegler, "Contracting with Diversely Naive Agents," *The Review of Economic Studies*, 2006, *73* (3), 689–714.

Erdem, Tulin, Michael Keane, Sabri Oncu, and Judi Strebel, "Learning About Computers: An Analysis of Information Search and Technology Choice," *Quantitative Marketing and Economics*, 2005, *3* (3), 207–247.

Ericson, Keith, "Market Design when Firms Interact with Inertial Consumers: Evidence from Medicare Part D," 2010. Mimeo, Harvard University.

_ , "Forgetting We Forget: Overconfidence and Prospective Memory," *Journal of the European Economic Association*, 2011, *Forthcoming.*

Fang, Hanming and Dan Silverman, "Time-Inconsistency And Welfare Program Participation: Evidence From The NLSY," *International Economic Review*, 2009, *50* (4), 1043–1077.

Fernandez, Roberto M., Emilio J. Castilla, and Paul Moore, "Social Capital at Work: Networks and Employment at a Phone Center," *American Journal of Sociology*, 2000, *105* (5), pp. 1288–1356.

Finkelstein, Amy and Kathleen McGarry, "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market," *American Economic Review*, 2006, *96* (4), 938–958.

Fischoff, Baruch, "Debiasing," in Paul Slovic Daniel Kahneman and Amos Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases*, University of Chicago Press, 1982, pp. 422–444.

Fleisig, Dida, "Adding Information May Increase Overconfidence In Accuracy Of Knowledge Retrieval," *Psychological Reports*, 2011, *108* (2), 379–392.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent, "Comparative Advantage, Learning, and Sectoral Wage Determination," *Journal of Labor Economics*, 2005, *23* (4), 681–724.

Gicheva, Dora, "Worker Mobility, Employer-Provided Tuition Assistance, and the Choice of Graduate Management Program Type," 2009. Mimeo, Yale University.

Goettler, Ronald and Karen Clay, "Tariff Choice with Consumer Learning and Switching Costs," Mimeo, University of Chicago 2010.

Goodwin, Liz, "Law grads sue school, say degree is 'indentured servitude'," *Yahoo News*, August 2011.

Grisley, William and Earl D. Kellogg, "Farmers' Subjective Probabilities in Northern Thailand: An Elicitation Analysis," *American Journal of Agricultural Economics*, 1983, *65* (1), pp. 74–82.

Grossman, Zachary and David Owens, "An Unlucky Feeling: Overconfidence and Noisy Feedback," 2010. Mimeo, UCSB.

Grubb, Michael, "Selling to Overconfident Consumers," *American Economic Review*, 2009, *99* (5), 1770–1807.

_ **and Matthew Osborne**, "Cellular Service Demand: Tariff Choice, Usage Uncertainty, Biased Beliefs, and Learning," Mimeo, MIT 2011.

Haran, Uriel, Don A. Moore, and Carey K. Morewedge, "A simple remedy for overprecision in judgment," *Judgment and Decision Making*, 2010, *5* (7), 467–476.

Harris, Milton and Bengt Holmstrom, "A Theory of Wage Dynamics," *Review of Economic Studies*, 1982, *49* (3), 315–333.

Heath, Rachel, "Why do Firms Hire using Referrals? Evidence from Bangladeshi Garment Factories," 2011. Mimeo, University of Washington.

Heckman, James J. and Burton Singer, "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 1984, *52* (2), 271–320.

Hoffman, Mitchell, "Training Contracts, Worker Overconfidence, and the Provision of Firm-Sponsored General Training," 2011. Mimeo, UC Berkeley.

_ , "How is Information (Under-) Valued? Evidence from Framed Field Experiments," 2011a. Mimeo, UC Berkeley.

_ , "Overconfidence at Work and the Evolution of Beliefs: Evidence from a Field Experiment," 2011b. Mimeo, UC Berkeley.

Holt, Charles A. and Angela M. Smith, "An Update on Bayesian Updating," *Journal of Economic Behavior and Organization*, 2009, *69* (2), 125 – 134.

Hossain, Tanjim and Ryo Okui, "The Binarized Scoring Rule of Belief Elicitation," 2010. Mimeo, University of Toronto.

Hubbard, Thomas N., "Information, Decisions, and Productivity: On-Board Computers and Capacity Utilization in Trucking," *American Economic Review*, 2003, *93* (4), 1328–1353.

Hurd, Michael D. and Kathleen McGarry, "The Predictive Validity of Subjective Probabilities of Survival," *Economic Journal*, 2002, *112* (482), 966–985.

James, Jon, "Ability Matching and Occupational Choice," Mimeo, Duke University 2011.

Jolls, Christine and Cass R. Sunstein, "Debiasing through Law," *Journal of Legal Studies*, 2006, *35*, 199–242.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 1979, *87* (5), 972–90.

_ **and Yaw Nyarko**, "Learning by Doing and the Choice of Technology," *Econometrica*, 1996, *64* (6), 1299–1310.

Kahn, Lisa and Fabian Lange, "Employer Learning, Productivity and the Earnings Distribution: Evidence from Performance Measures," 2011. Mimeo, Yale University.

Kaniel, Ron, Cade Massey, and David Robinson, "The Importance of Being an Optimist: Evidence from Labor Markets," Mimeo, Duke 2011.

Keane, Michael P. and Kenneth I. Wolpin, "The Career Decisions of Young Men," *Journal of Political Economy*, 1997, *105* (3), 473–522.

Khwaja, Ahmed, Frank Sloan, and Sukyung Chung, "The Relationship Between Individual Expectations and Behaviors: Mortality Expectations and Smoking Decisions," *Journal of Risk and Uncertainty*, 2007, *35*, 179–201.

Kraus, Anthony W., "Repayment Agreements for Employee Training Costs," *Labor Law Journal*, 1993, *44* (1), 49–55.

_ , "Employee Agreements for Repayment of Training Costs: The Emerging Case Law," *Labor Law Journal*, 2008, pp. 213–226.

Kronman, Anthony T., "Paternalism and the Law of Contracts," *The Yale Law Journal*, 1983, *92* (5), pp. 763–798.

Laibson, David, Andrea Repetto, and Jeremy Tobacman, "Estimating Discount Functions with Consumption Choices over the Lifecycle," 2007. Mimeo, Harvard University.

Larkin, Ian and Stephen Leider, "Incentive Schemes, Sorting and Behavorial Biases of Employees: Experimental Evidence," *American Economic Journal: Microeconomics*, 2011, *forthcoming.*

Lau, Annie Y.S. and Enrico W. Coiera, "Can Cognitive Biases during Consumer Health Information Searches Be Reduced to Improve Decision Making?," *Journal of the American Medical Informatics Association*, 2009, *16* (1), 54 – 65.

Lazear, Edward, "Job Security Provisions and Employment," *Quarterly Journal of Economics*, 1990, *105* (3), 699–726.

_ , "Performance Pay and Productivity," *American Economic Review*, 2000, *90* (5), 1346–1361.

_ , "Firm-Specific Human Capital: A Skill-Weights Approach," *Journal of Political Economy*, 2009, *117* (5), 914–940.

Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent, "Performance Pay and Wage Inequality," *Quarterly Journal of Economics*, 2009, *124* (1), 1–49.

Loury, Linda Datcher, "Some Contacts Are More Equal than Others: Informal Networks, Job Tenure, and Wages," *Journal of Labor Economics*, 2006, *24* (2), 299–318.

Lynch, Lisa M, "The Economics of Youth Training in the United States," *Economic Journal*, 1993, *103* (420), 1292–302.

Lynch, Lisa M. and Sandra E. Black, "Beyond the Incidence of Employer-Provided Training," *Industrial and Labor Relations Review*, 1998, *52* (1), pp. 64–81.

Malmendier, Ulrike and Geoffrey Tate, "CEO Overconfidence and Corporate Investment," *Journal of Finance*, December 2005, *60* (6), 2661–2700.

_ **and** _ , "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction," *Journal of Financial Economics*, 2008, *89* (1), 20 – 43.

Manchester, Colleen, "How Does General Training Increase Retention? Examination Using Tuition Reimbursement Programs," 2009. Mimeo, University of Minnesota.

_ , "Learning Through the Lens of Your Job: The Effect of Employer-Financed MBA Education on Turnover," 2011. Mimeo, University of Minnesota.

Manski, Charles F., "Measuring Expectations," *Econometrica*, 2004, *72* (5), 1329–1376.

Massey, Cade, Joseph P. Simmons, and David A. Armor, "Hope Over Experience," *Psychological Science*, 2011, *22* (2), 274–281.

Mayraz, Guy, "Priors and Desires: a Model of Payoff Dependent Beliefs," 2011. Mimeo, Oxford University.

— , "Wishful Thinking," 2011. Mimeo, Oxford University.

McGreevy, Patrick, "LAPD Suing Former Officers," *Los Angeles Times*, March 2006.

Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya Rosenblat, "Managing Self-Confidence: Theory and Experimental Evidence," 2010. Mimeo, Harvard University.

Moore, Don A. and Paul J. Healy, "The Trouble With Overconfidence," *Psychological Review*, 2008, *115* (2), 502 – 517.

Nagypal, Eva, "Learning by Doing vs. Learning About Match Quality: Can We Tell Them Apart," *Review of Economic Studies*, 2007, *74* (1), pp. 537566.

Nelson, Robert G. and David A. Bessler, "Subjective Probabilities and Scoring Rules: Experimental Evidence," *American Journal of Agricultural Economics*, 1989, *71* (2), pp. 363–369.

Offerman, Theo, Joep Sonnemans, Gijs Van De Kuilen, and Peter P. Wakker, "A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes," *The Review of Economic Studies*, 2009, *76* (4), 1461–1489.

Oyer, Paul and Scott Schaefer, "Why Do Some Firms Give Stock Options to All Employees?: An Empirical Examination of Alternative Theories," *Journal of Financial Economics*, 2005, *76* (1), 99–133.

Pantano, Juan and Yu Zheng, "Using Subjective Expectations Data to Allow for Unobserved Heterogeneity in Hotz-Miller Estimation Strategies," 2010. Mimeo, Washington Univeristy in St. Louis.

Papageorgiou, Theodore, "Learning Your Comparative Advantages," 2010. Mimeo, Pennsylvania State University.

Parent, Daniel, "Wages and Mobility: The Impact of Employer-Provided Training," *Journal of Labor Economics*, 1999, *17* (2), pp. 298–317.

Paserman, M.Daniele, "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation," *Economic Journal*, 2008, *118* (531), 1418–1452.

Peterson, Jonathan, "Employee Bonding and Turnover Efficiency," 2010. Mimeo, Cornell University.

Pigou, Arthur C., *Wealth and Welfare*, London: Macmillan, 1912.

Prendergast, Canice, "The Provision of Incentives in Firms," *Journal of Economic Literature*, 1999, *37* (1), pp. 7–63.

Professional Truck Driver Institute, "Certification Standards & Requirements for Entry-Level Tractor-Trailer Driver Courses, http://www.ptdi.org/errata/CERTIFICATION_STANDARDS1.pdf," 2010.

Roth, Alex, "Layoffs Deepen Pool of Applicants for Jobs Driving Big Rigs," *Wall Street Journal*, 2009. 2/27/2009.

Rust, John, "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 1987, *55* (5), 999–1033.

_ , "Numerical Dynamic Programming in Economics," in H. M. Amman, D. A. Kendrick, and J. Rust, eds., *Handbook of Computational Economics*, Vol. 1 of *Handbook of Computational Economics*, Elsevier, 1996, chapter 14, pp. 619–729.

Sanders, Carl, "Skill Uncertainty, Skill Accumulation, and Occupational Choice," Mimeo, University of Wisconsin 2011.

Sandroni, Alvaro and Francesco Squintani, "Overconfidence, Insurance, and Paternalism," *American Economic Review*, 2007, *97* (5), 1994–2004.

Sanna, Lawrence J., Norbert Schwarz, and Shevaun L. Stocker, "When Debiasing Backfires: Accessible Content and Accessibility Experiences in Debiasing Hindsight," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2002, *28* (3), 497 – 502.

Schlag, Karl H. and Joel van der Weeley, "Eliciting Probabilities, Means, Medians, Variances and Covariances without assuming Risk Neutrality," 2009. Mimeo, Universitat Pompeu Fabra.

Schoenberg, Uta, "Testing for Asymmetric Employer Learning," *Journal of Labor Economics*, 2007, *25*, 651–691.

Segal, David, "Law Students Lose the Grant Game as Schools Win," *New York Times*, April 2011.

Spinnewjin, Johannes, "Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs," 2010. Mimeo, LSE.

Stange, Kevin, "An Empirical Examination of the Option Value of College Enrollment," *American Economic Journal: Applied Economics*, forthcoming.

Stevens, Margaret, "A Theoretical Model of On-the-Job Training with Imperfect Competition," *Oxford Economic Papers*, 1994, *46* (4), 537–62.

Stinebrickner, Ralph and Todd Stinebrickner, "Learning about academic ability and the college drop-out decision," Mimeo, University of Western Ontario 2011.

Suzuki, Yoshinori, Michael R. Crum, and Gregory R. Pautsch, "Predicting Truck Driver Turnover," *Transportation Research Part E: Logistics and Transportation Review*, 2009, *45* (4), 538 – 550.

Transport Topics, "Transport Topics: Top 100 For-Hire Carriers," 2009.

Van den Steen, Eric, "Rational Overoptimism (and Other Biases)," *American Economic Review*, 2004, *94* (4), 1141–1151.

— , "On the origin of shared beliefs (and corporate culture)," *RAND Journal of Economics*, 2010, *41* (4), 617–648.

van der Klaauw, Wilbert and Kenneth I. Wolpin, "Social security and the retirement and savings behavior of low-income households," *Journal of Econometrics*, 2008, *145* (1-2), 21–42.

Viscusi, W. Kip and Charles J. O'Connor, "Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?," *The American Economic Review*, 1984, *74* (5), pp. 942–956.

— **and Michael J. Moore**, "Worker Learning and Compensating Differentials," *Industrial and Labor Relations Review*, 1991, *45* (1), pp. 80–96.

Wahba, Jackline and Yves Zenou, "Density, social networks and job search methods: Theory and application to Egypt," *Journal of Development Economics*, December 2005, *78* (2), 443–473.

Walton, D., "Examining the self-enhancement bias: professional truck drivers' perceptions of speed, safety, skill and consideration," *Transportation Research Part F: Traffic Psychology and Behaviour*, 1999, *2* (2), 91 – 113.

Wang, Yang, "Subjective Expectations: Test for Bias and Implications for Choices," Mimeo, Lafayette College 2010.

Warner, John T. and Saul Pleeter, "The Personal Discount Rate: Evidence from Military Downsizing Programs," *American Economic Review*, 2001, *91* (1), 33–53.

Wooldridge, Jeffrey, *Econometric Analysis of Cross Section and Panel Data*, New York, NY: MIT Press, 2002.

Zafar, Basit, "How Do College Students Form Expectations?," Mimeo, New York Federal Reserve Bank 2010.