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

Mitchell Hoffman[†]

Stephen Burks[‡]

February 2, 2014

Abstract

Training by firms is a central means by which workers accumulate human capital, yet firms may be reluctant to provide general training if workers can quit and use their gained skills elsewhere. “Training contracts” that impose a penalty for premature quitting can help alleviate this inefficiency. Using two exogenously staggered contract changes at a leading trucking firm, we show that training contracts significantly reduce post-training quitting and increase the profitability of training. We demonstrate further that training contracts have especially large effects on training profits in our setting because of their interaction with worker overconfidence. If workers are overconfident about their own productivity at the current firm relative to their outside option, they will be more likely to sign training contracts and less likely to quit after training. Combining weekly productivity data with weekly productivity beliefs, we show that workers systematically overpredict their productivity, both with and without randomized financial incentives for accurate prediction. To quantify the impact of overconfidence for optimal contracts and welfare, we develop and estimate a structural learning model with biased beliefs that accounts for many key features of the data. Eliminating worker overconfidence would moderately increase worker welfare by 1.7%, but would decrease training profits by over \$8,000 per truck and substantially alter the optimal training contract. We confirm the feasibility of reducing overconfidence using an information experiment at a second large trucking firm. Despite the positive effect of training contracts on profits, training may not be profitable unless workers are overconfident.

JEL Classifications: M53, J24, D03, J41

Keywords: Firm-sponsored general training; training contracts; overconfidence; biased learning

*Hoffman is deeply indebted to David Card, Stefano DellaVigna, John Morgan, and Steve Tadelis for their advice and encouragement. We especially thank the many trucking industry managers and drivers who shared their insights with us. We thank managers at Firms A and B for sharing their data, for facilitating on-site data collection (Firm A), and for helping design the field experiment (Firm B). We are grateful for comments and suggestions from many people, but for especially detailed comments, we thank Jason Abaluck, Ben Handel, Ben Hermalin, Ken Judd, Lisa Kahn, Pat Kline, Ulrike Malmendier, Matthew Rabin, Jesse Rothstein, and Felix Vardy. We also thank seminar participants at Arizona, Berkeley, Carnegie Mellon, Chicago, the CIRPEE/Laval Conference on Subjective Expectations (2012), Columbia Business School, the Econometric Society (Summer 2011), Firm A, Harvard Business School, LSE, MIT, NBER Summer Institute (Labor/Personnel Economics, 2012), New York Fed, Northwestern Kellogg, Ohio State, RAND, Rochester Simon, SITE (Psychology and Economics, 2011), SOLE (2012), Stanford GSB, Toronto, Toronto Rotman, UBC Sauder, WashU Olin, and Yale SOM. We thank Mark Gergen, Anthony Kraus, and Gregory Lemmer for guidance in understanding relevant legal issues. Christina Chew, Sandrena Frischer, Will Kuffel, Amol Lingnurkar, and Irina Titova provided outstanding research assistance. Hoffman acknowledges financial support from the National Science Foundation IGERT Fellowship and the Kauffman Foundation. Burks and the Truckers & Turnover Project acknowledge financial support from the MacArthur Foundation, the Sloan Foundation, Firm A, the Trucking Industry Program at Georgia Tech, and University of Minnesota, Morris.

[†]University of Toronto Rotman School of Management; mitchell.hoffman@rotman.utoronto.ca

[‡]University of Minnesota, Morris; svburks@morris.umn.edu

1 Introduction

Training by firms is a central means by which workers accumulate human capital (e.g. [Acemoglu and Pischke, 1998, 1999a; Autor, 2001](#)). However, it has long been recognized that training may be subject to a “hold-up” problem: firms may be unwilling to train if workers are likely to quit after training. [Becker’s \(1964\)](#) canonical solution is for workers to pay for training themselves, but this may fail for a number of reasons, e.g., credit constraints. Understanding what makes training profitable for firms is important given policy concerns that US workers receive less training than workers in countries with lower turnover.

Training is often provided by firms using a “training contract:” the firm pays for training and the worker is fined for premature exit. These contracts are used in many jobs (e.g., truckers, policemen, nurses, pilots, to name a few), but have received limited attention from economists. We argue that training contracts can play an important role in reducing post-training quitting and increasing the profitability of investment in training. Further, at least in the setting we study, it is not training contracts alone, but rather their interaction with a common behavioral bias, worker overconfidence about ability, that significantly increases the profitability of training. Specifically, if workers are overconfident about their post-training ability relative to their outside option, they will be more likely to sign training contracts and less likely after training to quit.

We use quasi-experimental contractual variation from the trucking industry to show that training contracts significantly reduce quitting. Initially, workers at a very large long-haul trucking firm, which we call Firm A, received training at no cost with no contractual obligation. At the suggestion of a newly promoted manager, the firm introduced a training contract fining workers for quitting in their first 12 months. The contract was phased into different training schools at different times, depending on how fast the contract was approved for use in different states. The contract was later changed a second time, in a similarly staggered fashion. Exploiting these exogenously staggered contract changes, we estimate that the two contracts reduced quitting by 18 and 11 percent, respectively.

To better understand quitting under training contracts, we turn to [Jovanovic’s \(1979\)](#) canonical model of turnover, where quitting decisions reflect the evolution of worker beliefs about job match or productivity. We analyze beliefs by having 900 workers with training contracts predict their weekly productivity for two years. We show that workers who expect higher productivity end up achieving higher productivity and are less likely to quit, consistent with theory. However, the data also reveal a pattern where workers tend to systematically overpredict their productivity. Overprediction fades only very slowly. The overprediction we observe without financial incentives remains even when belief-elicitation is made incentive-compatible using randomized financial incentives for accurate prediction. We refer to this overprediction as “overconfidence” and say more about the term below.

To evaluate the importance of overconfidence for worker welfare, firm profits, and optimal training contracts, and to perform policy counterfactuals, we develop a structural model of worker turnover under training contracts, building on [Jovanovic \(1979\)](#). After training, workers learn about their underlying productivity through weekly productivity realizations, and decide when if ever to quit. However, we do not impose that workers are fully rational. Workers may hold biased priors, or learn

faster or slower than predicted by Bayes’ rule, thereby nesting the standard model as a special case. Using our rich subjective belief data for identification, we estimate that workers exhibit significant bias, both in means and in variances. Our model fits many aspects of the data quite well, both in-sample (under one contractual regime) and out-of-sample (under other training contracts), whereas a standard model performs far worse.

We use our model to perform counterfactuals. In the main counterfactual, we show that eliminating worker overconfidence would moderately increase worker welfare by 1.7% (because workers make better decisions), but would substantially reduce firm profits by \$8,000 per truck, actually making profits per worker negative. This baseline counterfactual takes contracts as constant, but, in reality, firms may adjust their contracts in response to changes in worker beliefs. Repeating our counterfactual when firms optimally set contracts given worker beliefs, we find that the impacts of debiasing are qualitatively similar, but slightly larger. Overconfidence has a large impact on the optimal contract, with greater overconfidence associated with a larger optimal penalty. In a second counterfactual, we also show that banning training contracts could actually plausibly increase worker welfare in this market, given the significant distortion from their biased beliefs.

Our paper contributes to several literatures. First, there is a large literature on why do firms pay for general training. Answers include credit constraints; asymmetric information about worker ability ([Acemoglu and Pischke, 1998](#); [Autor, 2001](#)) or about the quantity of training received; and wage compression (e.g., due to unionization or minimum wages) ([Acemoglu and Pischke, 1999b](#)); see [Acemoglu and Pischke \(1999a\)](#) for a survey. We provide some of the first evidence that formal contracts play an important role in making training profitable for firms, as well as illustrating the importance of worker beliefs for the profitability of investment in training.

Second, our paper relates to the literature on overconfidence. [Moore and Healy \(2008\)](#) provide an excellent survey of recent work and divide overconfidence into three types: absolute overconfidence, relative overconfidence, and over-precision. Our structural estimates indicate both significant absolute overconfidence and over-precision, but our paper’s largest focus is on absolute overconfidence, which we refer to hereafter simply as overconfidence. While overconfidence has been documented by psychologists for nearly half a century, a large share of what we know about overconfidence is based on short-term laboratory tasks, e.g., trivia games; thus, much less is known about how overconfidence evolves over time, particularly in the longer-run. We contribute by providing weekly field evidence on overconfidence and its evolution over two years, by far the longest high-frequency evidence on overconfidence we are aware of in any academic discipline, and provide so in a natural workplace setting. Second, while a number of labor studies estimate structural learning models, ours is one of the first to do so allowing for potentially biased beliefs. More generally, we contribute to a small but growing literature using subjective beliefs to estimate different types of structural models—see, e.g., [Wiswall and Zafar \(2012\)](#) for a noteworthy recent example—as well as to an even smaller literature combining structural estimation with author-conducted field experiments.

Third, our paper shows that firms may benefit from behavioral biases of their workers. Although a number of field studies analyze how firms may benefit from customer biases, ours is the first (to our knowledge) to analyze how firms may benefit from biases of their workers. We estimate that worker

overconfidence is critical for the profitability of investment in training—even with training contracts, if workers were not overconfident, we estimate that training would otherwise not be profitable.

The main data elements in the paper have never been analyzed previously. Other papers by one of the authors (Burks) have also examined data from a subset of drivers at Firm A, consisting of some of the worker hired over an 8-month period at 1 of Firm A’s several training schools.¹ Our paper uses data from all newly trained workers at Firm A over an 8-year period; comprised of tens of thousands of workers and millions of worker-weeks, this dataset is orders of magnitude larger and has never been analyzed previously. Both the large size of the data and the cross-training school contractual variation are essential for estimating the impact of training contracts. In addition, the subjective beliefs data from Firm A that play a critical role in our reduced-form and structural analysis, collected on drivers in the subset above, have never been analyzed previously. Finally, the experiments and data-gathering from Firm B were conducted for this paper and have never been used in any other research.

Section 2 provides background on training contracts and trucking. Section 3 describes the data and analyzes the impact of the contractual changes. Section 4 analyzes subjective productivity belief data, both from Firm A (without incentives) and from Firm B (with randomized financial incentives). Section 5 develops the model and then structurally estimates it. Section 6 performs counterfactual simulations on different contracts and beliefs, including eliminating worker overconfidence, and presents the parallel information experiment on reducing overconfidence at Firm B. Section 7 concludes.

2 Background on Training Contracts and on Trucking

Theoretical Preliminaries. In Becker’s (1964) model of human capital investment, workers finance their own training through reduced wages during the training period. However, if workers have limited credit and training lasts for a short period of time (e.g. a few weeks, as in trucking), workers will not be able to finance training through a reduced training period wage without making their wage highly negative. As we show in a one-period model in Appendix B, this situation can be remedied by having the worker take on a training contract, helping the firm recover training costs after quits and potentially also reducing quitting. The Appendix B model is a stylized one-period version of the dynamic structural model (developed later in Section 5) and allows us to make analytical predictions.

¹Burks et al. (2009) examine whether worker cognitive skills predict experimental measures of worker preferences, worker strategies in experimental games, and worker retention. Anderson et al. (2012) compare measures of social preferences between truckers, students, and non-trucker adults. Rustichini et al. (2012) examine whether measures of trainee personality predict experimental measures of worker preferences, worker strategies in experimental games, health behavior, worker retention, and worker accidents. Finally, Burks et al. (2012), analyzes new truckers predicting their quintile on an IQ test to test between different theories of relative overconfidence (people tending to overestimate how well they do compared to other people). Our paper differs in that we study absolute overconfidence instead of relative overconfidence; we study beliefs about productivity instead of about performance on an IQ test; and we study beliefs over time instead of at a single point in time. In addition, Burks et al. (2012) is focused on testing between different theories of the *causes* of relative overconfidence across people, whereas our paper focuses on the *consequences* of absolute overconfidence for worker behavior and contract design. Although the papers deal with quite different issues, we view the contributions as complementary. Burks et al. (2008) describe the collection of the data from Firm A in detail.

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. Thus, turnover will be unaffected by a training contract, as a training contract is merely a “property right” held by firms over the quit decision.² In our 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. By assuming in our model that workers have private information and that there is no renegotiation, we can show that allowing for training contracts increases training profits and reduces turnover (Proposition 1 or P1).

One concern about training contracts is that they may harm workers if workers do not have accurate beliefs. If workers are overconfident about their expected post-training ability relative to their outside option, they may be willing to take on a high-penalty training contract, and if overconfidence persists, they will be more likely to stay after training. Though harmful to workers, this may increase training profits for firms who seek to reduce post-training quitting. In Appendix B, we prove that overconfidence increases training profits (P3-P4). Further, we prove that the greater the level of worker overconfidence, the higher the optimal training contract penalty (P5).

Background on the U.S. Trucking Industry. Truckdriving is a very large occupation of 3.2 million U.S. workers, 1.8 million of which operate heavy trucks such as those used by the firms we study (BLS, 2010). Firms A and B are in the long-haul truckload segment of the for-hire trucking industry, which is the largest employment setting for this occupation. An important distinction is between long-haul and short-haul trucking. Long-haul truckload drivers are usually paid by the mile (a piece rate) (Belzer, 2000) and drive long distances from home. In contrast, short-haul truckload drivers generally spend fewer nights away from home and are not usually paid by the mile.³

The main training for heavy truckdrivers is that needed to obtain a commercial driver’s license (CDL). Most new drivers take a formal CDL training course, and in some states it is required by law (BLS, 2010). CDL training can be obtained at truck driving schools run by trucking companies, at private truck driving schools, and at some community colleges. CDL training courses usually last 2-4 weeks, and consist of classroom lectures, simulator driving, and actual behind-the-wheel truck

²To 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. 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. Finally, we note that even if the conditions of the Coase Theorem held, changing training contracts could influence quit rates by affecting worker selection—Section A.4, however, finds little evidence of selection effects.

³We highlight a few more institutional details. Truckload is the segment that hauls full trailer loads. Truckload has low rates of unionization and high turnover, and most drivers do not own their own trucks. Around 10% of trucks 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). All the drivers we study are non-union company drivers. In terms of pay, a long-haul company driver may receive about 30 cents per mile. For an analysis of productivity in trucking, see Hubbard (2003).

driving. The market price for CDL training at private training schools varies, but is often several thousand dollars. In phone surveys we did with the 30 largest truckload firms, 16 of 30 firms report operating their own training school at some point from 2001-2010. When firms provided training, they almost always used a training contract: Only two companies (including Firm A) sometimes provided CDL training without using a training contract. See Appendix J for more on the survey.

Production. Truckload drivers haul full loads between a wide variety of points around the US. Per the federal Hours-of-Service Regulations, truckdrivers are legally restricted from working more than 60 hours per week. While our data do not contain driver hours, Firm A told us that most of their drivers work approximately 60 hours per week. At Firm A, loads are assigned by a dispatching team at company headquarters (instead of the driver’s supervisor), and assigned primarily according to proximity. Once a load is finished, a driver may start a new one.

Productivity in long-haul trucking is measured in miles per week. There are significant cross-driver differences in average productivity, as well as substantial idiosyncratic variation in productivity within drivers. When asked the reason for significant cross-driver differences, Firm A managers emphasized several factors including speed, skill at avoiding traffic, route planning (miles are calculated by geographic distance between two points, not distance traveled), not getting lost, and coordinating with people to unload the truck. For example, drivers who arrive late to a location may have to wait long periods of time for their truck to be unloaded, which can be highly detrimental to weekly miles. As for the sources of week-to-week variation, managers emphasized weather, traffic, variable loading/unloading time, and unlucky load assignments. Thus, weekly miles, our measure of productivity, reflect both driver performance and effort, as well as factors that drivers do not control and may be difficult to predict *ex ante*. See Appendix D for more on measuring productivity.

3 Reduced Form Effects of the Contract Changes

To examine the effects of training contracts, we analyze two large contract changes at Firm A, a leading long-haul trucking firm. 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. To preserve Firm A’s anonymity, we omit certain details and do not provide the exact number of total workers.

3.1 Contract Changes

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 the contract was to help increase retention, with a secondary motivation being to help recover costs. To implement the contracts, it was necessary to have the contracts certified by the states where the training schools are located. (Managers explained to us that training schools are considered private colleges, and training contracts are counted as a form of loan contract.) This certification process took different amounts of time

in different states. The contract was approved quickly in a few training schools and was in use by mid 2002. At another training school, the contract was not approved until the end of 2002, and in a couple schools the contracts were never used as the certification process dragged on indefinitely. According to the Director of Driver Training, differences in time for state approval were most likely idiosyncracies of the state bureaucracy, and not related to the type of impact the contracts might have. 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. (According to several Firm A managers, the contract also covered fires so as to prevent workers who wanted to quit from trying to get fired; however, according to these managers, the firm did not intentionally fire workers to collect training penalties.)

After several years of the 12-month contract, management decided to switch to an 18-month contract. According to a Senior Vice President, the interest in changing the contract stemmed from a desire to retain new drivers for longer. 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 about \$65 for each week of service.⁴ Both the adoption of the 12-month and 18-month contracts were made without additional changes to driver pay.⁵ Both the changes to the 12-month and 18-month contracts applied only to new drivers; drivers who had already signed on with no contract or the 12-month contract did not have their contracts altered.

Enforcement. Although it may seem difficult to enforce training contracts with truckers, Firm A made strenuous enforcement efforts and collected 25-30% of the penalties owed. New drivers signed a written contract specifying penalties if they left within some period of time. No initial bond posted. Upon early exit, drivers were contacted by Firm A to pay the amount due. If drivers did not pay promptly, they were referred to one of multiple collection agencies. Firm A collected roughly 20 percent of the amount owed before referring accounts to collections. The collection agencies then collected an additional 5 to 10 percent.⁶

3.2 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 tens of thousands of drivers from January 2002 to December 2009. We focus on new inexperienced drivers who are trained by Firm A. Drivers are paid by the mile, with small payments for other tasks (e.g., helping unload a truck). The per mile piece rate increases with driver tenure.

⁴Of the \$65 per week that was deducted from the worker's initial quit penalty, the worker had around \$12.50 each week deducted from his 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.

⁵Firm A pays slightly different mileage rates depending on driver regions. It increased its overall pay schedule twice during the period for which we have data, once, in early 2004, and a second time, in late 2007.

⁶Comprehensive driver-level collection data are not available. The 25-30% collection rate is based on company records from late 2004. While a 25-30% collection rate may seem on face to be low, Firm A made strenuous enforcement efforts. Drivers who did not pay received numerous strongly worded letters and phone calls from the company before being referred to collections. For drivers who remained delinquent, credit agencies were notified.

For each driver, there is also basic demographic information. We restrict our 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.

Very detailed data are available for a subset of 895 new drivers who completed training at one of the firm’s training schools in late 2005 and 2006. We refer to drivers in this group as the data subset, and we focus the structural analysis on this group. We collected subjective belief data about next week’s productivity only for drivers in the data subset.

Panel A of Table 1 shows that most driver characteristics are similar across the three contractual regimes: No contract, 12-month contract, and the 18-month contract.⁷ Panel B 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. The drivers have very low average credit scores. Of the 88% of drivers with credit scores (12% do not have a sufficient credit history to have a credit score), the mean and median credit score are 586 and 564, respectively, compared to a median of 723 for the U.S. general population at the time of data collection; further, 53% of drivers have a credit score below 600, compared to only 15% of the U.S. general population.⁸

One limitation of personnel data is we generally do not see where drivers go when they terminate, so to remedy this, we did an “exit survey” by mail for drivers in the data subset. In drivers leaving the firm, only 12% report moving to another long-haul trucking firm. Roughly 50% of drivers moved to a non-trucking job or unemployment, and the remainder report moving to non-long haul trucking firms (e.g. local delivery drivers). See Appendix A.2 for more on this.

3.3 Do Training Contracts Affect Quitting?

We show that training contracts significantly reduce quitting. We first plot survival curves under the different contractual regimes in Figure 1. Survival is significantly higher for both the 12-month and 18-month contracts relative to the no contract regime. Figure 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.⁹

To estimate with control variables, we estimate Cox proportional hazard models of quitting:

$$\log(h_{it\tau}) = \alpha_t + \beta_1 * 12MC_i + \beta_2 * 18MC_i + \beta_3 * UE_{i\tau} + \beta_4 \bar{y}_{it} + \gamma_\tau + I^{yearhire} + I^{school} + X_i \lambda \quad (1)$$

⁷There are some slight differences, primarily, that the share of Hispanic drivers is slightly lower under the 12-month and 18-month regimes relative to the no-contract regimes. However, this appears to reflect inter-school demographic differences (since contracts were in place in different schools for different lengths of time) as opposed to differences over time in the share Hispanic across contracts within a training school.

⁸The credit score is the FICO-98 and ranges from 300 to 850. Firm A purchased credit scores for drivers in the data subset. What credit score constitutes a “subprime” borrower varies by lender, but the cutoff is often 620 or 640. Thus, the majority of drivers in the sample would be considered subprime borrowers. Drivers are especially over-represented among those with very low credit scores, with 43% having scores below 550 compared to only 7% of the U.S. population. The credit score statistics are from the “Report to the Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit,” issued by the Board of Governors of the Federal Reserve System in August 2007.

⁹Firm A managers suggested that the small bumps in quitting at 52 weeks under the no contract and 18-month contract regimes may result from workers postponing quitting until then to be able to say that they worked for a full year at their last employer when applying for other jobs.

where $h_{it\tau}$ is the quit hazard of driver i with t weeks of tenure in year τ ; α_t is the log baseline hazard; $12MC_i$ and $18MC_i$ are dummies for i having the 12-month or 18-month contracts; $UE_{i\tau}$ is the unemployment rate in i 's state of residence at time τ ; $\bar{y}_{it}$ is average productivity to date; γ_τ is a time fixed effect; $I^{yearhire}$ is a year of hire fixed effect; I^{school} is a school fixed effect; and X_i are individual covariates. Training contracts vary at the school-week of hire level. The main coefficients of interest are β_1 and β_2 , the effects of the two contracts on the quit hazard. β_1 and β_2 are identified by changes in training contracts across workers hired at different dates at a given training school.

Results are in Table 2. In column 1, the coefficients on the 12-month and 18-month contract variables are -0.18 and -0.11, respectively, meaning the contracts decreased quitting by 18 and 11 percent.¹⁰ A percentage point increase in state unemployment is estimated to reduce quitting by 5.6%. Thus, the impact of the 12-month (18-month) training contracts is similar to a 3 (2) percentage point increase in the unemployment rate. The estimates of β_1 and β_2 remain similar when average productivity and demographics are controlled for. The interaction effects in column 6 indicate that 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, we ran linear probability models of quitting on interactions of the contracts with worker quarter of tenure. The estimates are shown in Figure 3. Under the 12-month contract, 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. A complementary approach to identification is to analyze the quit patterns of new drivers hired before and after the contract changes using an event study. For the transition from no contract to the 12-month contract, we 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, we 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.¹¹ As seen in Figure 4, for the transition to the 12-month contract, the probability

¹⁰Throughout the reduced form analysis on the full Firm A dataset, standard errors are clustered at the training school-week of hire level (the level of variation for the training contracts). Doing so allows for arbitrary correlation of the error within training school classes (drivers attending the same training school at the same time).

¹¹ For the transition from the 12-month to the 18-month contract, the event study is estimated using:

$$Quit5378_{isq} = \sum_{j=\underline{T}}^{\bar{T}} \theta_j D_{sq}^j + I^{yearhire} + I^{school} + X_i \lambda + \epsilon_{isq} \quad (2)$$

where $Quit5378_{isq}$ is a dummy for quitting in weeks 53-78 (conditional on having stayed through week 52) by driver i from school s with quarter of hire q and ϵ_{isq} is an error. D_{sq}^j is a dummy for whether those in quarter of hire q at training school s are j periods from the introduction of the 18-month contract; formally, $D_{sq}^j = \mathbf{1}(q - e_s = j)$, where e_s is the quarter when school s adopted the 18-month contract. We normalize $\theta_{-1} = 0$. For the switch from no contract to the 12-month (18-month) contracts, we assume that $\underline{T} = -3$ and $\bar{T} = 9$ ($\underline{T} = -7$ and $\bar{T} = 4$). We "bin up" the end points by including dummies for the event time being less than $\underline{T}$ or greater than $\bar{T}$. Additionally, we restrict the

of quitting during weeks 46-52 drops by roughly 5 percentage points. For the transition to the 18-month contract, there is a decrease of over 20 percentage points in the probability of quitting in weeks 53-78 occurring at the time of the contract change. Given that an average of 53% of workers quit during weeks 53-78, this effects represents a sizeable decrease in response to the contract.

3.3.1 Threats to Identification

We discuss additional threats to identification. We describe why we believe them to be unlikely to affect our main results.

Worker Sorting Between Schools. One potential confound to identifying the impact of training contracts on quitting would be worker sorting between schools. For example, a worker who believed he had a high chance of quitting might prefer to attend a training school that did 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.¹²

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 we 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 we will overestimate the effect. As noted earlier, managers at Firm A thought this unlikely to be a concern, believing the timing of training contract approval to be idiosyncratic and likely reflective of differences in the speed of state bureaucracy. In the data, training contract adoption is not predicted by state unemployment rates (which may be correlated with unobserved factors affecting quitting), as seen in Table E3. In addition, the results in Table 2 are similar when training school-specific year-of-hire trends are included (results available on request), suggesting contract change impacts are not merely the result of pre-existing trends in quitting.

Concurrent Policy Changes at Firm A or Other Firms. Contract changes at training schools were not accompanied by other changes in Firm A policy, such as changes in applicant screening or worker benefits, according to Firm A managers.

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 that worker tenure did not affect contract enforcement.

estimation sample in the event study to drivers who eventually exit the company.

¹²Of 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.

3.4 Understanding the Quit Impacts: 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 could also affect the *selection* of workers who choose to work at the firm, e.g. quit penalties might deter workers with low productivity or low taste for the job from signing up. (Note that workers selecting into the firm is different from the previously considered possibility of workers sorting between schools.) We provide three tests suggesting that the effect of the training contracts on quitting operated primarily through incentives. For brevity, the tests are described in Appendix A.4.

3.5 Worker Quitting and Learning about Productivity

In the structural model, we model quitting under training contracts as a product of worker learning, and provide reduced-form evidence for this assumption here. We assume that workers are initially uncertain about their productivity at the job and become more likely to quit if they learn their underlying productivity is low. One implication of learning about productivity is that at every level of tenure, workers who have higher productivity to date should be less likely to quit. Figure E1 confirms this by comparing the average miles per week of drivers who quit that week versus drivers who make it to that week and do not quit. Quitting drivers have lower average miles in prior weeks than non-quitting drivers. Selection on productivity is examined with controls in Table 2. As can be seen, an increase in past average productivity significantly reduces the hazard of quitting.

4 Worker Productivity Beliefs

Worker productivity beliefs are key in theories of turnover. We examine whether incorporating subjective beliefs can help better predict productivity and turnover under training contracts, as well as provide evidence of overconfidence. We first present our main results from two years of non-incentivized belief data from Firm A, and then present the Firm B incentivized data to show the results are robust to incentives.

Non-incentivized belief data from Firm A. Every week on Wednesday, we asked drivers in the data subset to predict their miles for the following pay week (Sunday-Saturday, starting on the Sunday in 4 days). Drivers responded by typing an answer to the below question, which we sent over the Qualcomm message system in their truck: *About how many paid miles do you expect to run during your next pay week?* We interpret this question as asking drivers for their subjective mean.¹³

¹³Another possible interpretation is that it is asking drivers for the median of their subjective mile distribution for 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 may seem unlikely to matter for the reduced-form or structural estimation. Instead of asking for a point-estimate, another method of belief elicitation would to have asked truckdrivers for their subjective productivity distribution at every week (e.g. “what is the chance you will run between X and Y miles next week,” varying X and Y to span the whole distribution), as has been done in the pioneering work of Charles Manski and colleagues (Manski, 2004). We chose our approach of asking for point estimates because of our desire to reduce survey time burden for the two-year weekly

There are several potential concerns with using our beliefs question to predict behavior and study overconfidence. First, researchers might worry that beliefs are stated to please others, e.g. drivers exaggerate their productivity beliefs to please their boss. However, in our setting, drivers were informed repeatedly that their responses and participation were never to be shared with the company. That is, driver supervisors would never even know whether a worker participated in the survey, let alone what his responses were. Second, no incentives were used to incentivize accurate belief responses. However, as we discuss below in our field experiment with Firm B, we find no evidence that beliefs are different when workers are rewarded for accurate beliefs. Finally, there is substantial non-response: the average response rate to the weekly beliefs survey is 28%. Using Inverse Probability Weighting and other methods later to address non-random response, we show our main structural results are very similar, suggesting bias from non-response is limited.¹⁴

In Table 3, we show that subjective productivity beliefs help predict productivity over and beyond other predictors. Specifically, we consider regressions of the form:

$$y_{i,t} = \alpha + \beta b_{i,t-1} + \gamma \bar{y}_{i,t-1} + X_i \delta + \epsilon_{i,t} \quad (3)$$

where $y_{i,t}$ is driver i 's productivity in his t^{th} week with the company; $b_{i,t-1}$ is his subjective belief about his productivity in week t stated in week $t-1$; $\bar{y}_{i,t-1}$ is lagged average productivity to date; X_i are controls; and $\epsilon_{i,t}$ is an error. Column 2 estimates $\hat{\beta} = 0.15$, meaning, controlling for observable characteristics, a driver whose expectation is 100 miles higher than another driver will end up driving an average of 15 miles more. Once average productivity to date or individual fixed effects are added, the coefficient drops to between 0.06 and 0.08. That is, *within person*, subjective beliefs have some predictive power, but less so. Overall, the results suggest that subjective beliefs have informational content, being somewhat predictive across individuals, and mildly predictive within individuals. The relatively low coefficients likely reflect attenuation bias due to measurement error in subjective beliefs.

Table 4 shows that quitting decisions reflect subjective beliefs outside of predictors in a standard Bayesian model. We estimate Cox proportional hazard models of quitting of the form:

$$\log(h_{i,t}) = \alpha_t + \beta b_{i,t} + \gamma \bar{y}_{i,t} + X_i \delta, \quad (4)$$

Average productivity to date, $\bar{y}_{i,t}$ is a sufficient statistic for beliefs about productivity in a standard Bayesian normal learning model. However, a 100 mile increase in subjective miles predicts a 6 percent decrease in the probability a worker quits. The true effects are likely higher, with observed estimates biased downward due to measurement error. The effect is robust to the inclusion of different controls including education, demographics, and work type. The coefficient on beliefs does not change very much across the different specifications.¹⁵ The finding that having higher productivity beliefs is associated with a lower chance of quitting is quite robust. To show that the finding is not driven by outliers, Table F4 repeats Table 4 using a dummy for productivity beliefs being above the median

study and because of concern that eliciting distributions might be difficult for less-educated drivers.

¹⁴The response rate seems low, but is comparable to that in many nongovernmental surveys. See Appendix A.1 for discussion and analysis related to non-response.

¹⁵Burks et al. (2009) and Rustichini et al. (2012) show that measures of driver cognitive ability and personality are correlated with some lab experimental and workplace behaviors. Including these measures in Tables 3 or 4 has almost no impact on our estimates of β and γ .

(as opposed to a continuous measure) and finds sizeable impacts. Our results are also robust to using lagged productivity beliefs (Table F5) or a worker’s average productivity belief to date (Table F6), the latter which aims to measure beliefs more of as a stable worker characteristic.¹⁶ In contrast to a “standard” setup where workers hold the same beliefs given their productivity signals, it appears that workers’ heterogeneous subjective beliefs predict quitting.

Although subjective beliefs are predictive, Figure 5 shows that average beliefs consistently exceed actual productivity. Productivity and productivity beliefs are collapsed by week of tenure and then smoothed using a local polynomial regression. Beliefs exceed actual productivity both in means (Panel A) and medians (Panel B). Workers are initially overconfident by roughly 500 miles per week, about 25% of their average productivity. This percentage declines over time, though it is quite persistent and decreases slowly. Even after 100 weeks, workers still overpredict by around 150 miles per week. Two concerns with these figures are (i) The sample is changing over time (due to quits) and (ii) The productivity line is based on all workers whereas the belief line is based on workers who respond to the survey in a given week. To address (i), we restrict the sample to workers who are there for most of the sample period (at least 75 weeks) in Panel C. The overall pattern of overconfidence is similar, though standard errors are larger. To address (ii), we re-make the picture in Panel D dropping the 39% of workers who never respond to the survey. We also restrict to workers who are there at least 75 weeks, and again look at medians instead of means (so as to verify that our results are not being driven by outliers). The picture is again similar.¹⁷

The average results mask that beliefs are sensible in several ways, and that there is a lot of heterogeneity within and across drivers. First, drivers update beliefs in a manner consistent with aspects of Bayesian updating. Increases in average productivity to date are associated with substantial increases in future subjective beliefs, both across and within drivers, and the weight on average productivity to date increases with tenure (see Table F7). Second, though on average workers are overconfident in almost every week, weeks where driver predictions are higher than next week’s actual miles constitute only 65% of the data; drivers *underpredict* in 35% of driver-weeks, so it is not the case that each individual driver is overpredicting his miles every week. Third, drivers differ substantially in the share of the time they overpredict their miles, and in the level of their over- or

¹⁶These two checks help assuage the concern of reverse causality; that is, one concern is that people who expect to quit in the future might believe that they will slow down and drive fewer miles. These two checks yield the same result. In additional unreported robustness checks with squared beliefs or dividing beliefs into various quantiles, the relationship between beliefs and quitting is relatively linear. Despite the different checks and despite the fact that the coefficient on beliefs does not change much by including observable variables, it is possible that there may be selection on unobserved variables. Thus, we interpret the results here as evidence that overconfidence correlates with fewer quits instead of that overconfidence causes fewer quits. In addition to these checks, one might also think to include some version of (Beliefs - Productivity) instead of Beliefs as a regressor. However, as we discuss after Prop. 2 in Appendix B, it is the level of a person’s perceived inside and outside options that affects quitting in theory, not overprediction. Further, including (Beliefs-Productivity) as a regressor imposes the restriction that coefficients on Beliefs and Productivity are the same.

¹⁷In Panels C and D, we stop at 75 weeks instead of the full sample to increase sample size. However, results are similar if we restrict to workers who are there for 100 weeks). More generally, we have made the basic graph comparing productivity and productivity beliefs a number of different ways, varying means vs. medians, restricting to workers with different tenure levels, and restricting based on survey response (all subjects, excluding subjects who never respond, and restricting only to weeks where both subject productivity and productivity belief are available). The basic path of overprediction is similar across these different specifications. In Figure F2, we also plot Beliefs - Productivity as a function of tenure, and observe a similar pattern of slowly decreasing overconfidence.

under-prediction. Figure F3 shows 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.

There are several interpretations of our result that workers tend to systematically over-predict their productivity and that this over-prediction takes many weeks to decline. For example, workers may be stating *aspirational* productivity levels instead of their true expectations. Or instead, workers may choose to report their expected mileage in a week supposing that “everything goes well” and there are no unexpected delays or hiccups. For both these explanations, one might imagine that misprediction could be eliminated if only workers were incentivized to state the mean of their subjective productivity distribution. Alternatively, the overprediction may reflect workers’ true beliefs and instead reflect a persistent behavioral bias that would be difficult to eliminate with an incentive; overprediction may fade only slowly, given both substantial idiosyncratic variation in miles and given potential variance bias. To distinguish between these explanations, we turn to incentivized data.

Incentivized belief data from Firm B. To distinguish between these different explanations and to overcome other concerns with non-incentivized data (e.g. that non-incentivized subjects do not “think hard” enough about their forecasts), we randomized financial incentives for truckdrivers at another large long-haul trucking company, Firm B, to accurately guess about their productivity. 272 workers were randomly assigned to guess without financial incentives or to receive up to \$10 per week for guessing about their productivity. Subjects did this for 2-6 weeks before being randomly re-assigned to another treatment: Control (nothing changes), increased incentive (up to \$50 per week), or debiasing. Even though the mapping between miles and earnings is straightforward (as truckers are paid per mile), we also had subjects provide incentivized guesses about their earnings (to address the possibility that for some reason workers are overconfident about their mileage, but not overconfident about their earnings). Workers were incentivized in their guesses using a quadratic scoring rule; see Appendix E for further information.

Table 5 shows that neither the \$10 incentive nor the \$50 incentive had a significant impact on productivity beliefs or earnings beliefs. Given that the standard errors are moderately sized, we are unable to conclude that the experiment had no effect on worker beliefs. For example, using column 1 of Table 5, the 95% confidence interval on the impact of the \$10 incentive was -128 miles to +68 miles, and the 95% confidence interval on the impact of the \$50 incentive was -169 miles to +182 miles. Given a mean overconfidence level of 245 miles, we are able to reject the hypothesis that *all* the observed overconfidence would disappear if workers were given either \$10 or \$50 incentives. In addition, Table F3 shows that there are no statistically significant differences as to whether productivity beliefs are more predictive of actual productivity when they are financially incentivized.

5 Model and Structural Estimation

Our reduced-form analysis suggests that training contracts reduce quitting, that workers overpredict their productivity, and that greater belief bias is correlated with a lower chance of quitting. We now develop a structural model of quitting and belief formation under training contracts to help

understand these results and to do counterfactuals (i.e. different training contracts and changing beliefs). With the estimated model, we can look at how much training contracts and overconfidence affect training profits and welfare, and look how overconfidence affects the optimal training contract.

The model is a discrete time extension of the model in Jovanovic (1979) allowing for biased beliefs. A worker decides each week whether to quit his job. The problem is an optimal stopping problem; once he quits, he cannot return.¹⁸ Quitting is the only decision to make – in particular, there is no effort decision.¹⁹ Workers have different underlying productivities, but productivity is initially unknown, both to the worker and the firm. The worker is forward-looking in his quitting decision and each week’s miles provides him a noisy signal from which he learns about his underlying productivity. However, workers may be subject to belief biases. The worker’s priors need not be accurate, e.g. he may believe that the job is on average quite lucrative. 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 reflecting productivity beliefs, quitting decisions will also reflect a driver’s underlying taste for the job or career (e.g. how much a driver dislikes being away from home) as well as idiosyncratic shocks (e.g. a fight with the boss). In the model, both the piece rate and training contracts are taken as given; both are endogenized later in Section 6 when we consider optimal contract design for the firm.²⁰

5.1 Model Setup

The time horizon is infinite and given in weeks 1, 2, Workers have baseline productivity η , which is distributed $N(\eta_0, \sigma_0^2)$. Workers are paid by a piece rate, w_t , that depends on their tenure. Workers know the piece rate-tenure profile, and believe that this profile will not be changed by the company at some future date.²¹ A worker’s weekly miles, y_t , are distributed $N(a(t) + \eta, \sigma_y^2)$, and weekly earnings are thus $Y_t = w_t y_t$. $a(t)$ is a known learn-by-doing process, which we specify below. The worker’s outside option is r_t and also depends on his tenure. 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

¹⁸While workers who quit Firm A are allowed to re-apply, very few ever return. Of workers starting in 2002-2003 who quit in their first two years, less than 5% return by the end of 2009.

¹⁹Effort decisions are not included in most related structural learning models, e.g. Arcidiacono (2004) and Stange (2012). Our data do not contain exogenous variation in the piece rate that would be needed to plausibly identify the cost of effort function. We speculate, however, that including effort in the model would not qualitatively affect our main conclusions or would actually strengthen them. For example, suppose that there was complementarity between effort and perceived ability. Under this assumption, our main counterfactual of eliminating worker overconfidence in Section 6 would have an additional downside for firms of reducing worker effort. One potential modification of our conclusions would come if overconfidence was useful for agents in setting goals or overcoming self-control problems (Benabou and Tirole, 2002). To the extent that effort and overcoming self-control via overconfidence were important, this would seem likely to reduce the calculated worker welfare gain from debiasing in Section 6.

²⁰In the model we estimate, the firm makes no decisions. In addition to taking the piece rate and training contract as fixed, the firm is assumed not to fire workers. Firing is 3 times less common in the data than quitting, and ignoring firing enormously simplifies the model by preventing us from having to estimate a dynamic game. In Section 6, we add two steps before workers start to make their quitting decisions: (1) Firms offer $(\bar{k}, \bar{w})$, which the workers choose whether to accept; (2) Training occurs, during which workers’ taste and belief bias are realized.

²¹Assumptions of this form are standard in structural labor and personnel economics, and allows us to avoid having to specify beliefs over possible future firm policy changes. We believe the assumption is reasonable in our setting, given that the pay schedule is only changed twice over an 8-year period. Specifically, about 73% of new drivers never experience a change in the overall pay schedule.

having observed their past miles $y_1, y_2, \dots, y_{t-1}$, but not their current week miles, y_t . Workers and firms are assumed to be risk-neutral and to have a discount factor given by δ .²²

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

$$V_t(\mathbf{x}_t) = \max_{d_t, d_{t+1}, \dots} E_t \left(\sum_{s=t}^{\infty} \delta^{t-s} u_t(d_s, \mathbf{x}_s) | d_t, \mathbf{x}_t \right). \quad (5)$$

where $\mathbf{x}_t$ is the vector of state variables ($\mathbf{x}_t$ includes past miles, $y_1, \dots, y_{t-1}$, and is detailed further below). (5) can also be written as a Bellman Equation: $V_t(\mathbf{x}_t) = \max_{d_t} E_t(u_t(d_t, \mathbf{x}_t) + \delta V_{t+1}(\mathbf{x}_{t+1}) | d_t, \mathbf{x}_t)$.

The per-period utility from staying at the job is equal to the sum of the worker's non-pecuniary taste for the job, earnings, and an idiosyncratic shock:

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

where α is the worker's non-pecuniary taste for the job, and 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 . We assume that there is unobserved heterogeneity in non-pecuniary taste for the job, α , with α drawn from a mass-point distribution (Heckman and Singer, 1984).

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 the penalty for quitting at tenure t . The utility from quitting is the fine,²³ plus the discounted value of his outside option, plus an idiosyncratic shock:

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

where ε_t^Q is an iid unobserved idiosyncratic error with the same distribution as ε_t^S . It is useful to define the two choice-specific value functions. Let $V_t^S \equiv E_t(u_t(1, \mathbf{x}_t) + \delta V_t(\mathbf{x}_{t+1}) | 1, \mathbf{x}_t)$ and $V_t^Q \equiv E_t(u_t(0, \mathbf{x}_t) + \delta V_t(\mathbf{x}_{t+1}) | 0, \mathbf{x}_t)$ be the value functions for staying and quitting, respectively. Plugging in for $u_t(1, \mathbf{x}_t)$ and $u_t(0, \mathbf{x}_t)$, the choice-specific value functions can be written as:

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

and the Bellman Equation can be re-written as $V_t(\mathbf{x}_t) = \max_{d_t \in \{0,1\}} (V_t^S(\mathbf{x}_t), V_t^Q(\mathbf{x}_t))$.

Agents gradually learn their productivity as more and more productivity signals are observed. Thus, after a sufficiently large number of periods, T , the value function can be approximated by the

²²Risk neutrality is assumed in many dynamic learning models, e.g. Crawford and Shum (2005), Nagypal (2007), Stange (2012), and Goettler and Clay (2011), 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.

²³Even though only 25-30% of the penalties owed were collected, as described in Section 3.1, we assume that drivers act as if the utility cost of quitting is equivalent to the utility loss from paying the contract penalty. We believe this assumption is not unreasonable. Firm A was very firm with new drivers about its intention to collect money owed upon a quit. After a quit, drivers who did not pay faced aggressive collection contacts by both Firm A and collection agencies, as well as the reporting of delinquency to credit agencies. As a robustness check, we have also estimated the model assuming drivers as if the utility loss from quitting is 0.3 times the penalty. The estimates are similar, but the model fit is actually slightly worse.

following asymptotic value functions:

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

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

$$E(y_t|y_1, \dots, y_{t-1}) = \frac{\sigma_y^2}{(t-1)\sigma_0^2 + \sigma_y^2} \eta_0 + \frac{(t-1)\sigma_0^2}{(t-1)\sigma_0^2 + \sigma_y^2} \frac{\sum_{s=1}^{t-1} y_s - a(s)}{s-1} + a(t) \quad (6)$$

This expression is a weighted sum of the agent's prior and his demeaned average productivity to date. As t increases, the agent will eventually shift all the weight from his prior to his average productivity signals. We augment the standard learning model in two ways. First, we 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 η is drawn from a distribution $N(\eta_0 + \eta_b, \sigma_0^2)$. Second, we allow for agents to have a perception of signal noise that may be different from the true signal noise in the population: Workers perceive the standard deviation of weekly productivity signals to be $\widetilde{\sigma}_y$ instead of σ_y . With these two assumptions, an agent's subjective expectation of his productivity, which we denote by E^b (where b stands for belief) is:

$$E^b(y_t|y_1, \dots, y_{t-1}) = \frac{\widetilde{\sigma}_y^2}{(t-1)\sigma_0^2 + \widetilde{\sigma}_y^2} (\eta_0 + \eta_b) + \frac{(t-1)\sigma_0^2}{(t-1)\sigma_0^2 + \widetilde{\sigma}_y^2} \frac{\sum_{s=1}^{t-1} y_s - a(s)}{s-1} + a(t) \quad (7)$$

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

We allow that workers' reported subjective beliefs include some measurement error, as accurately reporting one's beliefs about productivity may be unusual or unfamiliar for a worker. We assume that reported beliefs equal underlying subjective beliefs plus a normally distributed error. The reported subjective belief, b_{it} , of driver i at tenure week t is distributed: $b_{it} \sim N(E^b(y_{it}|y_{i1}, \dots, y_{it-1}), \sigma_b^2)$.

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

1. Workers form beliefs b_t given past earnings $y_1, y_2, \dots, y_{t-1}$.
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 with the learning by doing function $a(t) = 2a_1 * (\Lambda(a_2 t) - .5)$, where $\Lambda(x) = \frac{e(x)}{1+e(x)}$ and t is worker tenure in weeks. $a(t)$ depends only on tenure; thus, the speed of learning by doing does not depend on the number of miles driven or on the ability of the driver. Workers fully anticipate the path of $a(t)$.²⁴

²⁴The logistic 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, as well as the empirical results on tenure and productivity in Shaw and Lazear (2008). Here, a_1 is the total amount by which productivity increases and a_2 indicates the speed of learning by doing. We believe our assumption that workers fully anticipate the learning by doing process is reasonable in our setting. In interviews, drivers and managers often referred to a steep "learning curve" for rookie drivers.

In addition, we 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 about 5 weeks, during which drivers receive a flat salary of about \$375 per week and during which their productivity is not recorded in the payroll data. We account for the possibility that drivers who quit during these 5 weeks may lack some skills required to be a productive truckdriver. Specifically, we assume the reservation wage over time is $r_t = r - \frac{6 - \min\{t, 6\}}{5} s_0$. We fix r using outside data, while s_0 , the value of skills gained while working with a senior driver, is a parameter to estimate.²⁵

Solving the Model. The state variables consist of past miles, the piece rate, the training contract, taste heterogeneity, a person’s level of overconfidence, a vector of observable additional characteristics (X), and the idiosyncratic shocks: $\mathbf{x}_t = (y_1, \dots, y_{t-1}, \mathbf{w}, \mathbf{k}, \alpha, \eta_b, X, \epsilon)$. We allow for heterogeneity in taste for the job and in overconfidence. To solve the model, we first solve for the asymptotic value functions (after all learning has taken place) using value function iteration. With the asymptotic value functions in hand, backward recursion can then be applied to solve the dynamic programming problem. We provide further details in Appendix C.

5.2 Discussion of Model Assumptions

Outside option. In our model, the outside option, r_t , depends on tenure, but not productivity. This feature differs from many models of firm-sponsored general training where the worker is paid the same share of his marginal product at both his inside and outside option (though less than his full marginal product at both), e.g. Acemoglu and Pischke (1999b). We believe our assumption is realistic in our context given that only 12% of workers who exit report moving to another long-haul trucking job, with the vast majority moving to another type of work (see Section 3.2). Having high ability in long-haul trucking does not necessarily imply that one will have high ability in non-trucking jobs or even in short-haul trucking.²⁶ In addition, our assumption about the outside option is consistent with our earlier finding that, all else equal, workers with higher productivity to date are substantially less likely to quit. (That is, if workers had the same productivity in their inside and outside options and were paid the same share of their marginal product at each, then high and low ability workers would be equally likely to quit, which, as seen in Table 2, is not the case.)

Beliefs. While our model allows for non-standard belief formation, these features are estimated from data instead of imposed. The model does not assume that people have overconfident priors or learn more slowly than would be predicted by Bayes’ Rule, but rather these features are identified via the belief data (see Section 5.3); our model nests the standard model as a special case. Several aspects

²⁵In addition, we assume that drivers do not begin learning from their productivity realizations until after 5 weeks of tenure. In addition to allowing for skill accumulation during the first 5 weeks, we alternatively estimate the model allowing for continuous skill accumulation as driver tenure increases. Specifically, we assume that $r_t = r + 2\theta_1 * (\Lambda(\theta_2 t) - .5)$, where $\Lambda(x) = \frac{e(x)}{1+e(x)}$, and θ_1 and θ_2 are parameters to estimate.

²⁶As discussed in Section 2, in contrast to long-haul drivers who are usually paid per mile, short-haul drivers are usually paid by the hour. For short-haul drivers, relationship-management skills (for managing customer and client relationships) are more important than being able to do a lot of miles quickly while far from home. Some workers will be better at long-haul whereas others will be better at short-haul.

of our generalized normal learning model receive support from the results in Section 4. Differences in subjective beliefs are moderately predictive of differences in productivity across workers, but only mildly predictive within workers. This finding is consistent (broadly) with our modeling assumption that workers do not have private information about their underlying productivity. Subjective beliefs do, however, affect quitting decisions in our model, which is consistent with our earlier empirical finding that, all else equal, workers with greater belief bias are less likely to quit.

The strongest assumption in the model is that workers are not overconfident about their outside option despite potentially being overconfident about their current job ability. However, for overconfidence to “lock in” workers after training, this assumption is stronger than necessary. Instead, for overconfidence to reduce quitting, the worker must be more overconfident about his current job earnings than his outside option (see Proposition 2 in the Appendix); that is, he must exhibit *differential overconfidence*. If the strong assumption of no outside overconfidence fails, but workers are still differentially overconfident, overconfidence will still theoretically reduce quitting, but less so than if the strong assumption held.

While the assumption of differential overconfidence is difficult to test, we present several pieces of evidence and arguments on why it seems reasonable in our setting. First, we collected data on workers’ *perceived outside options*. Drivers in the data subset were asked what their earnings would have been had they not started work with Firm A. We compare drivers’ response to this question to what “similar-looking” people earned in the March 2006 Current Population Survey (CPS). As Figure F4 shows, the perceived outside option workers would have earned had they not gone through training does not appear to be higher than what people like them in the CPS are earning. (Also, there is little correlation between a worker’s perceived inside option and his perceived outside option; see the notes of Figure F4 for more on this data.) Second, while current earnings are proportional to ability in long-haul 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.²⁷ Third, 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.²⁸ Fourth, we can view the results in Table 4 as supportive of differential overconfidence.²⁹ Fifth, the assumption is consistent with

²⁷Performance 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 like trucking 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.

²⁸Van den Steen (2004) provides a “winner’s curse” argument on how self-selection can promote belief biases. Consider a worker choosing among several jobs. For each job, he receives a noisy signal about his productivity there. The agent will naturally choose the job with the highest signal, and will be overconfident there relative to other jobs. While it is easy to imagine that workers may have different beliefs when choosing between long-haul trucking jobs and non-long-haul trucking jobs, it is also quite possible that workers might expect to have different productivity at different long-haul trucking firms. For example, one driver may have a lot of experience driving around the South and would be less productive at a firm where most of the routes were in the Northeast. Based on where the routes are, it may be harder for a driver to get home regularly, and having to spend a lot of time and effort getting home could negatively affect productivity.

²⁹In Proposition 2 in Appendix B, we prove that more overconfident workers will be less likely to quit if and only if they are more overconfident about the inside than the outside option. In Table 4, we showed that, all else equal, more overconfident workers are less likely to quit.

psychological and behavioral economics evidence on overconfidence and on cognitive dissonance.³⁰

Micro-foundations of Overconfidence. Several theoretical micro-foundations for overconfidence have been proposed in the literature, including evolutionary advantages, self-signaling, and social-signaling. Since our model and estimation do not depend on the source of the overconfidence, we remain agnostic about the source of the overconfidence; our contribution is to document overconfidence and explore the implications of overconfidence for behavior and welfare, not to understand its foundation. We do, however, assume that agents do not receive psychological utility from their beliefs, consistent with our Section 4 experimental finding that incentives do not appear to reduce overconfidence. If agents received psychological utility from beliefs, we might expect them to trade off incentives for accuracy with the utility value from stating high beliefs (unless, of course, the personal benefits from stating optimistic beliefs are so strong that agents are unwilling to trade-off moderate-sized incentives to reduce their optimism). (To the extent that the assumption is incorrect and truckers do receive substantial psychological benefits from their beliefs, this would cause us to overstate the benefit to workers of eliminating overconfidence.) Also, we remain agnostic whether workers are overconfident about their own skills (e.g. “I have great endurance on the road”) vs. whether they are overoptimistic about external events (e.g. “traffic will be better next week”).

5.3 Estimation and Identification

The model is estimated by maximum likelihood. In Appendix C, we derive the likelihood function and describe the estimation procedure. Although the parameters are jointly identified, we can discuss the main data features that allow us to identify particular 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 learning by doing parameters, b_1 and b_2 , are identified by how much productivity goes up (b_1) and how quickly (b_2). 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. The continuous skill gain parameters, θ_1 and θ_2 are identified by how much quitting changes with tenure given the increase in measured productivity.

Taste heterogeneity. The taste for job 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 beliefs data are critical for identifying the belief parameters.

³⁰Moore and Swift (2010) show that different measures of overconfidence are only weakly related to each other and to various individual characteristics. Turning to cognitive dissonance, cognitive dissonance theory (Festinger, 1957) suggests that agents may be averse to beliefs that are inconsistent with having made good decisions. For a worker who has invested substantial time and effort (and has additionally incurred a financial obligation to stay) to train with Firm A, it may be mentally difficult to believe one is a bad match with Firm A. In these models, e.g. Mayraz (2011), a worker may come to believe that being with Firm A gives him the highest earnings.

Prior mean bias (overconfidence), η_b , is identified by the difference between believed and actual productivity, particularly at lower tenure levels. The believed standard deviation of productivity shocks, $\widetilde{\sigma}_y$, determines the subjective speed of learning in the model. The larger $\widetilde{\sigma}_y$ is, the slower that agents’ initial overprediction will disappear. The standard deviation of beliefs, σ_b , is identified by noise in beliefs unrelated to information in model-predicted subjective expectations. An increase in σ_b leads to greater week-to-week fluctuations in beliefs unrelated to actual productivity.³¹

Scale parameter. 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.

5.4 Implementation

To solve the dynamic programming problem numerically, we discretize miles into increments of 100 from 100 to 4,000 miles per week. The outside option, r , is taken to be the median full-time earnings from the 2006 March CPS of workers like the “median” driver (35-year old males with a high school degree), which is \$32,000 per year. We convert this to a weekly wage of \$640. The weekly discount factor, δ , is assumed rather than estimated. In our preferred specification, we take $\delta = 0.9957$, corresponding to an annual discount factor of 0.8.³² In our baseline estimates, we do not include demographic covariates or heterogeneity in overconfidence.³³ After presenting our baseline estimates below, we discuss how they are robust to alternative versions of these assumptions.

We estimate the model using workers in the data subset, all of whom have a 12-month contract, as these are the only workers for which we collected subjective beliefs at Firm A. We focus on 699 workers with complete data (see Appendix C). As in several papers combining field experiments and structural estimation (e.g. Card and Hyslop, 2005; Todd and Wolpin, 2006; Duflo et al., 2012), our estimation uses workers under one regime (the 12-month contract), and then we do out-of-sample predictions using the other regimes (no contract and 18-month contracts). (We note for our paper that the contractual conditions are plausibly exogenous instead of randomly assigned.)

³¹Heterogeneity in η_b is identified from differences across people in the extent of productivity overprediction. We also note that η_b and $\widetilde{\sigma}_y$, in addition to affecting subjective beliefs, will also affect quitting. For example, the faster that agents begin to rely on their average productivity to date in making quit decisions (that is, the faster their quitting decisions reflect “learning”), the smaller that $\widetilde{\sigma}_y$ will be.

³²It is standard in most dynamic structural models to assume rather than estimate the discount factor, as the discount factor is usually weakly identified. We have experimented with a broad range of discount factors in sensitivity analysis. Model fit appears best for annual discount factors in the range of 0.80. However, assuming an annual discount factor of 0.90 or 0.95 yields quite similar estimates (see Table H1). A discount factor of 0 yields a substantially worse fit, evidence that workers in our context are forward looking. An annual discount factor of 0.80 is “low,” but is comparable or higher than discount factors used or estimated in other models analyzing dynamic choices of blue-collar or low-income workers (e.g. Paserman, 2008; Fang and Silverman, 2009; Warner and Pleeter, 2001).

³³We have also estimated models with covariates, for example, allowing taste for the job, α , to depend on gender, education, race, and age. However, including covariates has little effect on model fit and on the estimates of the other parameters. See also Card and Hyslop (2005) for an example of a dynamic model where covariates are excluded because they do not significantly improve model fit.

5.5 Structural Results

Table 6 displays the main structural estimates and indicates substantial mean bias and variance bias. As a benchmark, column 1 provides estimates assuming no mean bias in their productivity expectation. Column 2 allows for mean bias, estimating a mean bias of 589 miles (or roughly 30% of 2,025 miles, the estimated mean of the true productivity distribution). The productivity parameters in column 2 also differ from those in column 1 and seem more reasonable in size.³⁴ In terms of variance bias, the believed standard deviation of productivity shocks is roughly 2.5 times higher than the actual standard deviation of productivity shocks. This implies that workers update beliefs considerably slower than would be predicted by Bayes' Rule. Recall that the weight agents place on their signals relative to their prior is $\frac{t\sigma_0^2}{t\sigma_0^2 + \sigma_y^2}$. 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.³⁵

There is also significant heterogeneity in non-pecuniary taste for the job. Both models suggest three types of workers: Workers with a strong distaste for the job, workers with moderate distaste, and workers with a moderate positive taste. Given that workers are kept away from home for several weeks at a time, that many workers dislike this is unsurprising.

Table 7 takes the baseline model and adds learning by doing and continuous skill accumulation. The fit is better in all three specifications than in Table 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 707 miles. The estimated taste heterogeneity is also somewhat different.

In Figure 6, we simulate the careers of 40,000 drivers to show that the model fits the data quite well. The quit hazard-tenure profile, the productivity-tenure profile, and the beliefs-tenure profile observed in the data are plotted using an Epanechnikov kernel. 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.

The models with mean bias (columns 2 in Tables 6 and 7) fit the data much better than the models without mean bias (columns 1), according to a likelihood ratio test ($p < 0.01$). To analyze model fit on the quitting data alone, we compare the observed number of drivers quitting (O_t) with the number predicted from the model (E_t) using a chi-squared test. The chi-squared statistic is $\sum_t \frac{(E_t - O_t)^2}{E_t}$. Inference is conducted using T degrees of freedom, where T is the maximum number of weeks a driver is observed. For column 1 of Table 6, the chi-squared statistic is 520.2, whereas for columns 2, the chi-squared statistic is 214.0 ($p < 0.01$).³⁶ Thus, the fit in terms of quitting is

³⁴For 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.

³⁵This finding is consistent with the well-known psychological phenomenon of “conservatism” (Edwards, 1968), where agents update less than a rational agent would after receiving new information. For recent evidence of conservatism in different economic contexts, see Mobius et al. (2010) and Eil and Rao (2011). For theory on conservative updating, see Schwarzstein (2012).

³⁶Chi-squared tests are often used to assess the fit of dynamic models, e.g. Keane and Wolpin (1997) and Card and Hyslop (2005). Comparing the chi-squared statistic in column 1 of Table 6 with that in column 2 after one parameter

considerably better in the model with belief bias than without it.

Robustness. Our main estimates are quite robust to different assumptions, as we show in Table H1. We increase the discount factor, use inverse probability weighting to address non-response, correct for measurement error in productivity, Winsorize high subjective beliefs, allow for learning to occur over 200 weeks instead of 130, and increase the outside option. Increasing the discount factor and using inverse probability weighting do not significantly change the estimates. Winsorizing subjective beliefs at 4,000 miles decreases the mean bias term by 5% to 560 miles, suggesting that very high beliefs are accounting for only a small share of the overall mean bias. Most of the other modifications do not much change the estimates.³⁷ Our results are also robust to allowing for heterogeneity in overconfidence.³⁸

One can imagine several alternative non-standard economic forces that affect a worker’s taste for the job, e.g. a feeling of commitment to stay with the firm providing training. However, time-invariant shifters of taste for the job are already being accounted for via the taste heterogeneity parameters. In contrast, overconfidence provides a time-varying impact on the value of staying that appears to fit the data well.

Out-of-Sample Fit. In addition to predicting well in sample, our model also makes reasonable out-of-sample predictions, which is a more demanding test for the model. Although the fit is imperfect, Figure H4 shows that the model, estimated using workers under the 12-month contract, can predict basic retention patterns under the no contract (panel a) and 18-month contract (panel b) regimes. The χ^2 statistics for model-predicted quit behavior are 120.3 (no contract) and 341.3 (18-month contract). Our preferred model with mean bias predicts substantially better out-of-sample than the model without mean bias, for which the χ^2 statistics are 171.7 (no contract) and 471.7 (18-month contract). Panel c of Figure H4 simulates survival under the three contracts and matches basic data patterns in Figure 1. The 12-month contract regime has the highest survival at the one year mark, but is exceeded thereafter by the 18-month contract. The model also stacks up well at some tenure levels against the quasi-experimental estimates from Table 2, which control for observable differences between workers under the 3 regimes. For example, in panel c of Figure H4, the simulated 12-month and 18-month contracts reduce the share of people who have quit at 60 weeks by 17.8% and 11.8%, respectively (or 12 and 8 percentage points). These are close to our baseline

is added, the difference in chi-squared is highly significant ($\chi^2_{df=1} = 520.2 - 214.0 \rightarrow p < 0.01$). (As caveated though 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.) The model with mean bias also fits other aspects of the data besides the means in Figure 6, such as the distribution of beliefs and productivity over tenure (see Figure H1). To see the poor model fit for the model without mean bias, see Figure H2.

³⁷We have also performed additional robustness checks that we omit for ease of exposition. We have (1) Estimated with finer and coarser discretizations of miles (as suggested by Rust (1987)); (2) Eliminated high subjective beliefs instead of Winsorizing them; and (3) Assume the taste heterogeneity is normally distributed instead of mass point distributed. The estimates are generally robust to these checks. Making the miles grid more fine or coarse does not much affect the estimates unless the grid is sufficiently coarse. Eliminating subjective beliefs decreases the mean belief bias to 506 miles instead of 589 miles in the baseline and 560 miles when Winsorized.

³⁸We have estimated the model allowing for mass point heterogeneity in η_b . Repeating our later counterfactuals using these estimates, our conclusions are quite similar. With two mass points in overconfidence, eliminating overconfidence increases worker welfare by 1.5%, but increases profits by \$7,000 per truck, similar to as in Table 9, and optimal contracts are very similar to those in Table 10. Results with overconfidence heterogeneity are available on request.

estimates in column 1 of Table 2, where we estimated 17.9% and 10.8% reductions in the quit hazard.

6 Counterfactual Simulations

We use our structural estimates to quantify the importance of training contracts and biased beliefs, as well as to analyze potential firm or government policies. All counterfactuals use our baseline estimates in column 2 of Table 6.

Profits and Welfare Under the Three Contracts. We show that Firm A increased profits, but decreased worker welfare through its two contract changes by simulating worker behavior under the three observed training contracts. We use two measures of firm profits: profits per worker and profits per truck (see Bojilov (2011) for similar profit formulas in a different context). Average profits per worker is defined as average profits brought in from a worker during his time with the company (the profit stream ends when the worker quits), plus training contract penalties, minus training costs:

$$\pi = \sum_{t=1}^{\infty} \delta^{t-1} (1 - Q_t) ((P - w_t - mc)y_t - FC) + \sum_{t=1}^{\infty} \delta^{t-1} \theta k_t q_t - TC \quad (8)$$

where q_t is a dummy for quitting in week t , $Q_t = \sum_{s=1}^t q_s$ is a dummy for 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. Based on consultation with Firm A managers, we assume that $P = \$1.80/\text{mile}$, $mc = \$1.16/\text{mile}$, $\theta = 0.3$, $FC = \$600/\text{week}$, and $TC = \$2,500$. We equate the firm's weekly discount factor to the worker's, $\delta = 0.9957$, so as to avoid having results being driven by differences in discount factors; our conclusions are unchanged if we assume higher discount factors for both worker and firm.

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 we set equal to the cost of the training. Profits per truck is given by: $\pi = \sum_{t=1}^{\infty} \delta^{t-1} [(1 - Q_t) ((P - w_t - mc)y_t - FC) + Q_t (-TC + \theta k_t + \pi)]$ which can be re-arranged to yield

$$\pi = \frac{\sum_{t=1}^{\infty} \delta^{t-1} [(1 - Q_t) ((P - w_t - mc)y_t - FC) + Q_t (-TC + \theta k_t)]}{1 - \sum_{t=1}^{\infty} \delta^{t-1} Q_t} \quad (9)$$

Since workers have biased beliefs, average *experienced utility* will differ from a worker's *ex ante expected utility*. Worker welfare is measured using experienced utility by summing earnings, taste for trucking, and idiosyncratic shocks.³⁹ We simulate the entire data-generating process, simulating 3,000 workers for up to 1,300 weeks each. As seen in Table 8, both measures of profits are higher

³⁹Measuring welfare is an important issue in behavioral economic models (Bernheim and Rangel, 2009). In our case, agents do not have non-standard preferences, only non-standard beliefs. Our focus on experienced utility is shared by the empirical work of Grubb and Osborne (2011) and the theoretical framework in Mullainathan et al. (2012).

under the 12-month contract than no contract, and higher under the 18-month contract than the 12-month contract. Profits increase with the contractual changes, though worker welfare is reduced.⁴⁰

Debiasing: Reducing Worker Overconfidence. To examine quantitatively how overconfidence affects quitting, worker welfare, profits, and contractual choices, we simulate eliminating worker overconfidence, which we also refer to, following the psychology literature, as “debiasing.” We also consider eliminating overconfidence by one-half, recognizing that debiasing may be incomplete in practice. As seen in Table 9, worker welfare moderately increases since worker quitting decisions become less distorted by overconfidence. However, worker retention and profits substantially decline. Profits per worker decline by about \$2,000, as due to the increase in quitting, the firm has less time to make profits from a given worker. Profits per truck decline by \$8,925 since the firm must spend more training new drivers to replace the ones who quit. Our counterfactual serves first to quantify how much overconfidence affects training profits and worker welfare in this market. To the extent that debiasing is feasible, as we investigate below, it may also be viewed as a counterfactual public policy experiment, e.g. a government information campaign that reduces overconfidence.⁴¹

Debiasing with Optimal Contractual Responses. Our last counterfactual took training contracts as fixed. We relax this here and allow firms to optimally adjust contracts in response to debiasing. We show that debiasing leads firms to substantially decrease optimal quit penalties, but that the overall conclusions from the baseline debiasing counterfactual remain quite similar.

In the optimal contracting problem we consider, the firm sets the piece rate and quit penalty for every week $1, 2, \dots, \infty$. Although the trucking industry as a whole is closer to perfect competition than monopsony, the largest firms tend to dominate the market for firm-sponsored training, as we document in Appendix J. Depending on where a potential truckdriver lives, it may be difficult to find multiple firms that are willing to provide firm-sponsored training at a local training facility. We therefore model the problem as a single firm setting wages and quit penalties subject to a perceived individual rationality constraint for the worker. Rather than solving analytically, we attack the problem using computational methods. Because this is a computationally difficult problem, we 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. Adding the decision of the firm, the full timing of the model is: (1) Firms offer $(\bar{k}, \bar{w})$, which the

⁴⁰In addition, throughout the counterfactuals, we focus separately on profits and worker welfare, and do not analyze total welfare. While we found our conclusions on profits and worker welfare to be very robust to different assumptions, we found total welfare to depend more closely on particular assumptions made. That our counterfactuals do not lead to unambiguous changes in total welfare is not surprising, given our model allows for multiple market failures, including private information about taste for the job and shocks, biased beliefs, quitting externalities, and monopsony power in the training market.

⁴¹In column 4 of Table 9, as a point of comparison to a counterfactual of eliminating overconfidence, we also consider a counterfactual where learning is “turned off” and underlying ability, η , is revealed to workers immediately. (This is equivalent to setting $\widetilde{\sigma}_\eta = \sigma_\eta = 0$.) Eliminating uncertainty is a main counterfactual explored in structural learning models with standard beliefs (e.g. Arcidiacono, 2004; Crawford and Shum, 2005; Stange, 2012). In our model with biased beliefs, revealing ability increases worker welfare, but substantially reduces worker retention and profits relative to the baseline. This is because revealing ability eliminates the impact of overconfidence (prior mean bias), as a worker who knows his ability perfectly will ignore his prior. Column 4 suggests that, in our setting, the slow speed of worker learning about productivity is beneficial to firms.

workers choose whether to accept; (2) Training occurs, during which workers’ taste and belief bias are realized; and (3) Work begins. The firm solves the problem:

$$\max_{\bar{k}, \bar{w}} E\pi(\bar{k}, \bar{w}) \text{ s.t. } E^bV(\bar{k}, \bar{w}) \geq \bar{u} \quad (10)$$

where $E\pi(\bar{k}, \bar{w})$ is the firm’s expected profit per worker and $E^bV(\bar{k}, \bar{w})$ 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. We set $\bar{u}$ equal to the perceived utility a worker would have gotten in each of the counterfactuals in Table 9. We maximize using a grid search.

Column 1 of Table 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.⁴² When overconfidence or learning are removed, the optimal quit penalty decreases. 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 increases; in addition, higher piece rates are more appealing to workers who are more overconfident. 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. See Appendix A.5 for more on this.

Is debiasing feasible? Testing the counterfactual using a field experiment with Firm B. We have shown that eliminating overconfidence would have significant consequences for firm profits and worker welfare. Is such a counterfactual feasible? Psychologists have long been interested in whether overconfidence and other behavioral biases can be eliminated, focusing primarily on laboratory settings.⁴³ After discussion with psychologists on different methods of debiasing, we chose to inform the workers at Firm B about our findings on overconfidence with the workers at Firm A. Workers were either administered the debiasing treatment, where they received information about our findings about overconfidence, a suggestion to reflect on past predictions, and information about their average prediction in their first several weeks of the experiment; or the control treatment, where they received information about their average prediction in the first several weeks of the experiment. Our debiasing treatment is deliberately somewhat heavy-handed, as we wanted to avoid a treatment that seemed too weak to affect anyone’s beliefs. At the same time, however, we wanted our debiasing not to require a lot of individual information (as would, say, an alternative treatment of providing individual-level feedback on over-prediction), both for logistical ease and in recognition that an actual

⁴²There are at least two explanations for the computed optimal quit penalty being higher than the one used in practice. First, it may reflect our assumption that the enforceability and collection of quit penalties is independent of the size of the penalties—e.g., in practice, it may be harder for firms to enforce larger penalties. Second, the observed 12-month contract may have actually been suboptimally low, which the firm gradually learned, making the firm gradually move to the 18-month contract with a higher quit penalty.

⁴³Fischhoff (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). In economics, the only paper we are aware of that explicitly attempts to debias overconfidence is Larkin and Leider (2012), though there is growing work on debiasing in other contexts (e.g. debiasing misperceptions about financial investments or schooling decisions).

debiasing policy may not be able to provide extensive individual information.

In Table 5, we see that debiasing moderately reduces miles beliefs and substantially reduces earnings beliefs (though the difference between the two impacts is not statistically significant when both are put in standardized units). Debiasing reduces miles beliefs by 113 miles (column 2) and earnings beliefs by \$86 (column 5) in our preferred specifications with controls. However, effects vary substantially by the week relative to debiasing. In the week of debiasing, miles beliefs decline by 207 miles and earnings beliefs decline by \$101. Given that the average overprediction with the Firm B data is 252 miles (that is, beliefs exceed actual productivity by 252 miles per week), this implies that the experiment eliminated nearly 80% of miles overconfidence in the first week. Given an average overprediction in earnings of \$101, the experiment eliminated nearly all of average earnings overconfidence in the first week. The coefficients decline as more weeks pass. For the earnings predictions, the impacts of debiasing remain statistically significant (even after 4-6 weeks), whereas for mileage predictions, the impacts remain sizeable, but not statistically significant.

While it is possible that the experimental impacts on beliefs could be driven by an “experimental demand” effect, as is the case for many lab and field experimental findings, we note that the debiasing reductions are similar whether or not a financial incentive is being randomly given for guessing about productivity. For miles predictions, the reduction is 105 (standard error=88) miles with an incentive and 123 (se=97) miles without an incentive, when we repeat column 2 of Table 5, splitting the sample by incentive for guessing or not. For earnings predictions, the reduction is \$92 (se=\$29) with an incentive and \$73 (se=\$46) without an incentive. If the debiasing impacts occurred merely because subjects wanted to tell the surveyors “what they wanted to hear,” one would think that this may not occur when subjects are incentivized to guess correctly.

Table G4 shows that there were, however, no significant impacts of experimental debiasing on actual quitting, surveyed intention to search for a new job, or surveyed job satisfaction. While we designed the experiment to estimate impacts on beliefs in Table 5 with reasonable precision, the impacts on non-belief outcomes are much less precisely estimated. The 95% confidence interval for the impact on quitting is -18 to +14 percentage points. In our counterfactual simulation, eliminating 50% of overconfidence increases quitting by 12 percentage points at week 9 of tenure. Given our experimental 95% confidence interval for the impact of debiasing on quitting includes our counterfactually simulated impact, we cannot rule out that our experiment would have achieved a similar impact on quitting as the one in our counterfactual simulation.⁴⁴

“Banning” Training Contracts. If workers have accurate beliefs about their expected pro-

⁴⁴We look at the impact of debiasing at 9 weeks in our counterfactual simulation since this time period corresponds most closely to debiasing in the randomized experiment. In the experiment, workers are tracked after debiasing for about 2 months to calculate their quitting percentage. In addition, we look at 50% debiasing since our randomized experiment did not permanently eliminate all of worker overconfidence and is best thought of as reducing some of worker overconfidence. One reason why our experiment may have failed to eliminate all overconfidence permanently is that it was a one-time intervention. Given that the experiment’s impacts on beliefs appear to fade somewhat after a few weeks, it is not particularly surprising that the experiment did not affect real outcomes like quitting. To more permanently eliminate worker overconfidence, it may instead be necessary to provide debiasing information on a more frequent basis. Finally, while the experimental impacts seeming to diminish over time could potentially reflect experimental demand effects, they seem more likely to us to be a manifestation of limited memory.

ductivity, then restricting the type of contracts workers and firms are allowed to sign seems unlikely to be welfare-enhancing. 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. We model a ban in training contracts as a restriction on the firm’s optimization problem. Instead of (10), the firm solves the restricted problem:

$$\max_{\bar{w}} E\pi(\bar{k} = 0, \bar{w}) \text{ s.t. } E^b V(\bar{k} = 0, \bar{w}) \geq \bar{u} \quad (11)$$

In this formulation, we ask: If firms continued to provide training after being outlawed from using quit penalties and could only adjust piece rates, how would this affect profits and welfare? We emphasize that this exercise is not the same thing as a true ban, since a ban would likely affect whether firms are willing to train. With this large caveat in mind, we examine the results in the last two columns of Table 10. Worker welfare is 4.3% higher when training contracts are banned compared to when the firm optimally sets training contracts unrestricted. We also examine a combined policy of banning training contracts and 50% debiasing. Worker welfare is 2.9% 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.^{45,46}

7 Conclusion

Given significant concern about under-investment, understanding what makes training profitable for firms is critical both for economic theory and for policy. This paper explores the impact on the profitability of training of the interaction of (1) Commonly-used contracts that fine workers for quitting after training, with (2) Workers’ biased beliefs about their post-training productivity.

Using exogenous contractual variation, we 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

⁴⁵Instead of assuming that banning training contracts would have no impact on the level of training, we can repeat the counterfactual making the opposite extreme assumption that banning training contracts would shut down the market for firm-sponsored general training. This would not eliminate all new truckdrivers, but would limit new truckdrivers to workers who could afford to pay for training at private training schools or community colleges. Such a policy would likely reduce welfare in a model with standard workers, but could potentially be welfare-improving here, given worker overconfidence. Using our estimates, such a counterfactual would reduce worker welfare by 1% (from \$151,016 to \$149,462). If we call a contract “exploitative” if the worker’s average ex-post utility from the contract is lower than his outside option, following [Eliaz and Spiegel \(2006\)](#), this suggests that training contracts are not exploitative. In reality, a ban on training contracts would fall somewhere in between shutting down and having no impact on the quantity of firm-sponsored training.

⁴⁶We have studied optimal training contracts and banning training contracts under the assumption a single monopolistic firm. We believe this is a reasonable assumption, given that firm-sponsored training in trucking is concentrated in the largest firms. However, it is worth examining how competition would affect our 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.

home state unemployment rate. Data on workers' subjective productivity beliefs predict actual productivity and quitting, but also indicate a significant and persistent pattern of overconfidence whereby workers believe they will achieve higher miles than they actually attain. Relying on relatively weak identifying assumptions, these results suggest that training contracts reduce quitting, that workers overpredict their productivity, and that greater belief bias is correlated with a lower probability of quitting.

To quantify the impact of training contracts and biased beliefs for profits and welfare, we make stronger assumptions, structurally estimating a quitting model with potentially biased beliefs. We show that both training contracts and overconfidence are critical for the profitability of training. If firms could not use training contracts or worker overconfidence was eliminated, firm profits from training would fall substantially, sometimes from positive to negative. Further, eliminating worker overconfidence significantly affects the optimal training contract.

Our results parallel several studies in behavioral industrial organization indicating that firms may profitably exploit consumers' biases, focusing instead on the behavioral biases of workers. An important difference in our setting, besides the identity of the parties involved, is the possibility of a baseline market failure of underinvestment. In a second-best world with underinvestment in general training, the existence of worker overconfidence makes training and training contracts more profitable, potentially increasing the quantity of firm training. Unfortunately, because our data are from only two firms, we are unable to calculate how overconfidence affects the share of firms willing to train; thus, we are unable to weigh this benefit against the distortion from overconfidence on worker decision-making. Thus, we caution against strong policy interpretations of our counterfactual result that debiasing would increase worker welfare.

Although we focus on a single industry, 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 stockbrokers). Future work should examine whether the impacts of training contracts in these jobs appear to be similar. While nearly exclusively piece rate compensation is not shared by most other jobs, piece rate compensation is not necessary for overconfidence to make workers more likely to sign training contracts and less likely to quit after training. For example, workers may be over-optimistic about some other aspect of the job, such as the probability of being promoted. While we focus on firm training, overconfidence can help entice or "lock in" workers even if there is no training involved, and may be relevant for other types of human capital investment, e.g. going to college, or for occupational choice in general.⁴⁷

Worker overconfidence may be important for many aspects of optimal job design and compensation. For example, when firms can choose to pay flat wages or piece rates, paying a piece rate may be appealing if workers are overconfident since overconfident workers perceive they may earn more than

⁴⁷See [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. David Segal, "Law Students Lose the Grant Game as Schools Win," *New York Times*, April 2011 and Liz Goodwin, "Law grads sue school, say degree is 'indentured servitude'," *Yahoo News*, August 2011. In a different application for workers, the impact of overconfidence about job-finding on optimal unemployment insurance, see [Spinnewjin \(2010\)](#). For work on overoptimism and stock options for nonexecutive workers, see, e.g., [Oyer and Schaefer \(2005\)](#).

they actually will (Larkin and Leider, 2012). Future work should continue to analyze the interplay between worker biases and firms' contractual choices.

References

- Acemoglu, Daron 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*, 1999a, 109 (453), F112–42.
- and —, "The Structure of Wages and Investment in General Training," *Journal of Political Economy*, 1999b, 107 (3), 539–572.
- Anderson, Jon, Stephen Burks, Jeffrey Carpenter, Lorenz Gtete, Karsten Maurer, Daniele Nosenzo, Ruth Potter, Kim Rocha, and Aldo Rustichini, "Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples," *Experimental Economics*, 2012, *Forthcoming*, 1–20.
- 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.
- 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.
- Becker, Gary, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, Chicago, IL: University of Chicago Press, 1964.
- Belzer, Michael, *Sweatshops on Wheels: Winners and Losers in Trucking Deregulation*, New York, NY: Oxford University Press, 2000.
- Benabou, Roland and Jean Tirole, "Self-Confidence And Personal Motivation," *Quarterly Journal of Economics*, 2002, 117 (3), 871–915.
- Bernheim, B. Douglas and Antonio Rangel, "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *Quarterly Journal of Economics*, 2009, 124 (1), 51–104.
- BLS, "Bureau of Labor Statistics Occupational Outlook Handbook, 2010-11: Truck Drivers and Driver/Sales Workers," 2010.
- Bojilov, Raicho, "Incentives to Work or Incentives to Quit?," 2011. Mimeo.
- 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," *Review of Economic Studies*, 2012, *Forthcoming*.
- , —, —, Kristen Monaco, Kay Porter, and Aldo Rustichini, "Using Behavioral Economic Field Experiments at a Firm: The Context and Design of the Truckers and Turnover Project," in "The Analysis of Firms and Employees: Quantitative and Qualitative Approaches" 2008.
- Card, David and Dean R. Hyslop, "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers," *Econometrica*, 2005, 73 (6), 1723–1770.
- Ching, Andrew, Tulin Erdem, and Michael Keane, "Learning Models: An Assessment of Progress, Challenges and New Developments," 2011. Mimeo.
- 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.
- DeGroot, Morris, *Optimal Statistical Decisions*, New York, NY: Mcgraw-Hill College, 1970.
- Duflo, Esther, Rema Hanna, and Stephen Ryan, "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, 2012, 102 (4), 1241–78.
- Edwards, Ward, "Conservatism in Human Information Processing," in B. Kleinmuntz, ed., *Formal Representation of Human Judgement*, 1968.
- Eil, David and Justin Rao, "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 2011, *Forthcoming*.
- Eliasz, Kfir and Ran Spiegler, "Contracting with Diversely Naive Agents," *Review of Economic Studies*, 2006, 73 (3), 689–714.
- Fang, Hanming and Dan Silverman, "Time-Inconsistency And Welfare Program Participation: Evidence From The NLSY," *International Economic Review*, 2009, 50 (4), 1043–1077.
- Festinger, Leon, *A Theory of Cognitive Dissonance*, Stanford, CA: Stanford University Press, 1957.
- Fischhoff, 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.
- Goettler, Ronald and Karen Clay, "Tariff Choice with Consumer Learning and Switching Costs," *Journal of Marketing Research*, 2011, 48 (4), 633–652.
- Grubb, Michael and Matthew Osborne, "Cellular Service Demand: Tariff Choice, Usage Uncertainty, Biased Beliefs, and Learning," 2011. Mimeo.
- 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.

- Hubbard, Thomas N.**, “Information, Decisions, and Productivity: On-Board Computers and Capacity Utilization in Trucking,” *American Economic Review*, 2003, 93 (4), 1328–1353.
- 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.
- Keane, Michael P. and Kenneth I. Wolpin**, “The Career Decisions of Young Men,” *Journal of Political Economy*, 1997, 105 (3), 473–522.
- Larkin, Ian and Stephen Leider**, “Incentive Schemes, Sorting and Behavioral Biases of Employees: Experimental Evidence,” *American Economic Journal: Microeconomics*, 2012, 4 (2), 184–214.
- 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.
- Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent**, “Performance Pay and Wage Inequality,” *Quarterly Journal of Economics*, 2009, 124 (1), 1–49.
- Manski, Charles F.**, “Measuring Expectations,” *Econometrica*, 2004, 72 (5), 1329–1376.
- Mayraz, Guy**, “Priors and Desires: a Model of Payoff Dependent Beliefs,” 2011. Mimeo.
- Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya Rosenblat**, “Managing Self-Confidence: Theory and Experimental Evidence,” 2010. Mimeo.
- Moore, Don A. and Paul J. Healy**, “The Trouble With Overconfidence,” *Psychological Review*, 2008, 115 (2), 502 – 517.
- Moore, Don and Samuel Swift**, “The Three Faces of Overconfidence in Organizations,” in Rolf Van Dick and J Keith Murnighan, eds., *Social Psychology and Organizations*, 2010.
- Mullainathan, Sendhil, Joshua Schwarzstein, and William J. Congdon**, “A Reduced-Form Approach to Behavioral Public Finance,” *Annual Review of Economics*, 2012, 4, 511–540.
- Nagypal, Eva**, “Learning by Doing vs. Learning About Match Quality: Can We Tell Them Apart,” *Review of Economic Studies*, 2007, 74 (1), pp. 537–566.
- 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.
- Paserman, M.Daniele**, “Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation,” *Economic Journal*, 2008, 118 (531), 1418–1452.
- Rust, John**, “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 1987, 55 (5), 999–1033.
- Rustichini, Aldo, Colin DeYoung, Jon Anderson, and Stephen Burks**, “Toward the Integration of Personality Theory and Decision Theory in the Explanation of Economic Behavior,” Mimeo, University of Minnesota 2012.
- Sanna, Lawrence J., Norbert Schwarz, and Shevaun L. Stocker**, “When Debiassing Backfires: Accessible Content and Accessibility Experiences in Debiassing Hindsight,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2002, 28 (3), 497 – 502.
- Schwarzstein, Joshua**, “Selective Attention and Learning,” 2012. Mimeo.
- Shaw, Kathryn and Edward P. Lazear**, “Tenure and Output,” *Labour Economics*, 2008, 15 (4), 704–723.
- Spinnewjin, Johannes**, “Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs,” 2010. Mimeo, LSE.
- Stange, Kevin M.**, “An Empirical Investigation of the Option Value of College Enrollment,” *American Economic Journal: Applied Economics*, 2012, 4 (1), 49–84.
- Stinebrickner, Ralph and Todd Stinebrickner**, “Learning about academic ability and the college drop-out decision,” 2011. Mimeo.
- Todd, Petra E. and Kenneth I. Wolpin**, “Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility,” *American Economic Review*, 2006, 96 (5), 1384–1417.
- Van den Steen, Eric**, “Rational Overoptimism (and Other Biases),” *American Economic Review*, 2004, 94 (4), 1141–1151.
- Warner, John T. and Saul Pleeter**, “The Personal Discount Rate: Evidence from Military Downsizing Programs,” *American Economic Review*, 2001, 91 (1), 33–53.
- Wiswall, Matthew and Basit Zafar**, “Determinants of College Major Choices: Identification from an Information Experiment,” Mimeo, New York Federal Reserve Bank 2012.

Figure 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 2002 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 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, $N \gg 5,000$. These figures plot the quitting hazard at any week under the 3 contractual regimes, using all drivers starting at Firm A between 2002 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 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 (12). 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 age). Standard errors 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 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. The solid line denotes the coefficient estimate, with the dotted lines denoting the 95% confidence interval. Panel (a) analyzes quitting in weeks 46-52 before and after the change to the 12-month contract whereas panel (b) 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 training 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 age), as in Equation (2). The number of quarters before and after the event varies between the two contracts due to limits on the number of quarters of data available before or after contract changes. Standard errors clustered at the school-week of hire level.

Figure 5: Overconfidence: Comparing Subjective Productivity Forecasts with Actual Worker Productivity (as a Function of Worker Tenure)

(a) Means

(b) Medians

(c) Means, Restrict to Workers Who Stay at Least 75 Weeks

(d) Medians, Restrict to Workers Who Stay at Least 75 Weeks & Respond to Belief Survey

Notes: These figures analyze actual and believed productivity for the 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 panels (a) and (c), the productivity and belief data are collapsed into weekly means before local polynomial smoothing. In panels (b) and (d), the productivity and belief data are collapsed into weekly medians before local polynomial smoothing. In panel (d), we restrict to the 61% of workers who ever respond to the survey. The figure is similar if we restrict miles to weeks where the driver responds to the survey, though standard errors are larger.

Although it may seem surprising how slowly overconfidence diminishes, it can be rationalized in the structural model given the substantial idiosyncratic variation in miles in the data, combined with a significant dose of conservatism / variance bias. See Figure F2 for overprediction as a function of driver tenure.

Figure 6: Structural Model: Model Fit

Notes: These figures compare model-simulated data against the actual data to assess model fit. We plot the survival curve, the quit hazard, the mean miles-tenure profile, and the mean beliefs-tenure profile. The model simulated corresponds to Column 2 in Table 7. We simulate the entire data-generating process for 40,000 drivers. The data are from 699 drivers with the 12-month contract, all of whom are from the same training school and hired in late 2005 or 2006. The paths of quits, earnings, and beliefs over tenure are plotted using an Epanechnikov kernel, with bandwidths of 6 weeks, 5 weeks, and 10 weeks, respectively.

Table 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
Drivers	0.09 <i>N</i>	0.73 <i>N</i>	0.18 <i>N</i>

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

Panel C: Beliefs in Data Subset

Percentile	Full Sample		Estimation Sample	
	Miles	Miles Beliefs	Miles	Miles Beliefs
10%	897	1500	1019	1500
25%	1367	1800	1475	2000
50%	1883	2300	1972	2500
75%	2427	2750	2506	2800
90%	2942	3000	3005	3050
Mean	1908	2323	1998	2423

Notes: Panel A provides summary statistics for all trained drivers at Firm A from 2002 to 2009. 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. Panels B and C restrict to drivers in the data subset. Data subset drivers are from the same training school and were hired in late 2005 or 2006. Panel B presents summary statistics about driver characteristics. Panel C presents quantiles and means on productivity and productivity beliefs, both for the full sample and for the 699 drivers with complete data that we use in the structural estimation (see Appendix C). Though mean miles and beliefs are higher for the estimation sample, the difference between the two is similar in both samples. In both samples, median mile beliefs lie at roughly the 70th percentile of miles. Though slightly less clean, estimating the structural model using the full sample yields very similar conclusions as to estimating on the estimation sample.

Table 2: Impact of the Training Contracts on Quitting – Cox Model, Diff-in-Diff

Sample of drivers:	All	All	All	Has demogs	Has demogs	Has demogs
	(1)	(2)	(3)	(4)	(5)	(6)
12m contract	-0.179*** (0.046)	-0.173*** (0.046)	-0.180*** (0.048)	-0.180*** (0.051)	-0.165*** (0.051)	
18m contract	-0.108* (0.062)	-0.117* (0.062)	-0.138** (0.064)	-0.129** (0.066)	-0.110* (0.067)	
State unemployment rate		-0.055*** (0.011)	-0.067*** (0.012)	-0.059*** (0.013)	-0.060*** (0.013)	-0.065*** (0.013)
Avg miles to date			-0.053*** (0.003)	-0.045*** (0.003)	-0.043*** (0.004)	-0.044*** (0.004)
12m contract * (wks≤52)						-0.287*** (0.062)
12m contract * (52<wks≤78)						0.085 (0.078)
12m contract * (wks>78)						0.028 (0.087)
18m contract * (wks≤52)						-0.033 (0.076)
18m contract * (52<wks≤78)						-0.957*** (0.109)
18m contract * (wks>78)						-0.129 (0.127)
Time FE (yr)	Yes	Yes	Yes	Yes	Yes	Yes
Year of Hire FE	Yes	Yes	Yes	Yes	Yes	Yes
Training School FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	No	No	Yes	Yes
Observations	<i>M</i>	<i>M</i>	<i>M</i>	0.87 <i>M</i>	0.87 <i>M</i>	0.87 <i>M</i>

Notes: An observation is a driver-week. The regressions are Cox proportional hazard models, with coefficients reported and standard errors clustered by school-week of hire in parentheses. Driver tenure is controlled for non-parametrically. State unemployment is the annual unemployment rate in a driver's state of residence in the current year. (We have also tried including the annual state unemployment rate in a driver's year of hire, but its impact is insignificant once the current year annual state unemployment rate is included.) 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 (4) differs from column (3) in that it restricts the sample to be the same as in column (5). The coefficients can be interpreted as approximate percentage changes in quitting, e.g. column (1) indicates the contracts reduced quitting by 18 and 11 percent. The approximation is less accurate for large percentages; the coefficient of -0.957 in column (6) corresponds to an odds ratio of 0.38, indicating a 62% reduction in quitting. The exact *M* is withheld to protect firm confidentiality, $M \gg 100,000$. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: Do Productivity Beliefs Predict Productivity? OLS Regressions

	(1)	(2)	(3)	(4)	(5)
L. Predicted miles	0.194*** (0.023)	0.147*** (0.019)	0.064*** (0.015)	0.062*** (0.015)	0.078*** (0.022)
L. Avg miles to date			0.708*** (0.038)	0.689*** (0.038)	
Demographic Controls	No	Yes	No	Yes	No
Education Controls	No	Yes	No	Yes	No
Work Type Controls	No	Yes	No	Yes	No
Subject FE	No	No	No	No	Yes
Observations	8,443	8,443	8,429	8,429	8,443
R-squared	0.069	0.127	0.190	0.194	0.291

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 regressions include week of tenure dummies. 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 4: Do Productivity Beliefs Predict Quitting?

	(1)	(2)	(3)	(4)	(5)
Predicted miles	-0.059*** (0.018)			-0.059*** (0.020)	-0.067*** (0.021)
Avg miles to date		-0.079*** (0.010)	-0.112*** (0.039)	-0.002 (0.036)	-0.062 (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. Column 3 differs from column 2 in that it restricts to the sample of driver-weeks for which there is a corresponding belief expectation. 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 5: Do Incentives for Accuracy or Information Reduce Worker Overconfidence? The Field Experiment with Firm B

Dep var:	Miles Prediction (in miles)			Earnings Prediction (in \$)		
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives for accuracy (up to \$10/wk)	-30.0 (49.1)	-52.8 (45.7)	-53.2 (45.7)	-31.2 (22.4)	-35.1 (22.1)	-35.2 (22.2)
Larger incentives (up to \$50/wk)	6.4 (87.8)	-25.2 (82.9)	-22.7 (83.2)	8.2 (41.9)	8.3 (41.0)	8.5 (41.1)
Debiasing	-96.0 (70.7)	-112.9* (65.4)		-68.3** (30.7)	-85.8*** (29.1)	
Debiasing X 0wk post-treat			-207.0** (80.2)			-100.7*** (34.4)
Debiasing X 1wk post-treat			-91.3 (89.4)			-71.6** (31.0)
Debiasing X 2wk post-treat			-111.3 (82.0)			-92.2*** (33.8)
Debiasing X 3wk post-treat			-82.5 (95.2)			-73.4* (39.0)
Debiasing X 4-6wks post-treat			-62.0 (73.0)			-84.9** (35.8)
Overprediction in sample	252 mi	252 mi	252 mi	\$101	\$101	\$101
Joint signifiance of the two incentive treatments (p-value)	0.827	0.499	0.497	0.369	0.276	0.276
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Observations	1,096	1,071	1,071	1,085	1,060	1,060
R-squared	0.16	0.21	0.21	0.06	0.09	0.09
Mean dep var	2316	2315	2315	765.2	765.8	765.8
Subjects (clusters)	254	243	243	248	236	236

Notes: OLS regressions with standard errors clustered by driver in parentheses. An observation is a worker-week. All regressions include a dummy for receiving the no debiasing treatment, dummies for the number of days not worked in a week, and worker tenure in weeks. The demographic controls are gender, age, trucking experience, and region of home residence. To limit the effect of outliers, we trim the lower and upper 5% on each dependent variable. Results without trimming are also similar, though the estimates are less precise. In addition, the results are very similar if instead of 5% trimming we simply drop obvious outliers individually. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Baseline Structural Estimates

		No Bias (1)	Belief Bias (2)
	<u>Productivity and Skill Parameters</u>		
η_0	Mean of prior productivity dist	2464 (18)	2025 (17)
σ_0	Std dev of prior productivity dist	475 (18)	286 (11)
σ_y	Std dev of productivity shocks	707 (3.6)	706 (3.6)
s_0	Skill accumulation in first 5 weeks	14.9 (3.1)	8.6 (4.4)
	<u>Taste UH Parameters</u>		
μ_1	Mass point 1 of taste UH	-249 (8)	-260 (14)
μ_2	Mass point 2 of taste UH	-107 (20)	-135 (13)
μ_3	Mass point 3 of taste UH	139 (43)	135 (41)
p_1	Probability type 1	0.55 (0.04)	0.34 (0.07)
p_1	Probability type 2	0.24 (0.03)	0.43 (0.06)
	<u>Belief Parameters</u>		
η_b	Belief bias		589 (28)
$\widetilde{\sigma}_y$	Believed std dev of productivity shocks	3653 (158)	1888 (136)
σ_b	Std dev in belefs	300 (1.4)	298 (1.4)
	<u>Scale Parameter</u>		
τ	Scale param of idiosyncratic shock	1620 (178)	2208 (350)
	Log-likelihood	-91064	-90865
	Number of workers	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. 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 7: Structural Estimates with Learning by Doing and Skill Accumulation

		No Bias (1)	Belief Bias (2)
	<u>Productivity and Skill Parameters</u>		
η_0	Mean of prior productivity dist	2401 (12)	1838 (23)
σ_0	Std dev of prior productivity dist	508 (19)	274 (10)
σ_y	Std dev of productivity shocks	706 (1.5)	706 (1.6)
b_1	Learning by doing level	154 (12)	242 (15)
b_2	Learning by doing speed	0.08 (0.01)	0.11 (0.01)
θ_1	Skill accumulation level	56 (39)	118 (55)
θ_2	Skill accumulation speed	0.92 (0.33)	0.03 (0.01)
	<u>Taste UH Parameters</u>		
μ_1	Mass point 1 of taste UH	-219 (42)	-420 (79)
μ_2	Mass point 2 of taste UH	-77 (44)	-162 (56)
μ_3	Mass point 3 of taste UH	167 (57)	176 (62)
p_1	Probability type 1	0.54 (0.07)	0.60 (0.04)
p_1	Probability type 2	0.25 (0.05)	0.19 (0.04)
	<u>Belief Parameters</u>		
η_b	Belief bias		707 (24)
$\widetilde{\sigma}_y$	Believed std dev of productivity shocks	3420 (138)	1329 (63)
σ_b	Std dev in belefs	299 (0.3)	297 (0.3)
	<u>Scale Parameter</u>		
τ	Scale param of idiosyncratic shock	1733 (166)	1808 (201)
	Log-likelihood	-91020	-90767
	Number of workers	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. 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 8: Profits and Welfare Under Different Contracts

	No contract	12 month	18 month
Profits per worker	\$1,696	\$3,022	\$3,070
Profits per truck	-\$222	\$4,095	\$4,243
Welfare per worker	\$154,086	\$151,016	\$150,763

Notes: This table presents profits per worker and profits per truck under the three different training contracts used by Firm A. Profits per worker and profits per firm are defined in Section 6 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%. A weekly discount factor of 0.9957 is assumed for workers and firms, corresponding to an annual discount factor of 0.8. The model simulated here corresponds to column 2 of Table 6.

Table 9: Counterfactual Simulations, No Contractual Response

	Baseline	50% debias	100% debias	Reveal ability
Retention at 20 wks	0.79	0.61	0.39	0.32
Retention at 40 wks	0.56	0.41	0.27	0.26
Retention at 60 wks	0.43	0.33	0.24	0.24
Welfare per worker	\$151,016	\$152,551	\$153,522	\$154,007
Profits per worker	\$3,022	\$1,909	\$748	\$1,434
Profits per truck	\$4,095	\$408	-\$4,830	-\$2,182
Ability at 20 wks	2,040	2,048	2,052	2,120
Ability at 40 wks	2,072	2,077	2,059	2,133
Ability at 60 wks	2,094	2,087	2,062	2,131

Notes: This table 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 η_b 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 simulated here corresponds to column 2 of Table 6.

Table 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	\$10,800	\$9,200	\$0	\$400	NA	NA
Wage factor	1.009	1.027	0.910	0.908	0.987	0.963
Retention at 20 wks	0.96	0.81	0.28	0.23	0.66	0.37
Retention at 40 wks	0.90	0.72	0.20	0.17	0.37	0.23
Retention at 60 wks	0.66	0.53	0.18	0.16	0.28	0.21
Welfare per worker	\$147,396	\$149,807	\$153,822	\$153,812	\$153,655	\$154,160
Profits per worker	\$3,748	\$2,670	\$1,338	\$1,866	\$2,122	\$1,567
Profits per truck	\$5,684	\$2,974	-\$3,966	-\$1,381	\$1,136	-\$1,935
Ability at 20 wks	2,028	2,038	2,031	2,118	2,046	2,048
Ability at 40 wks	2,037	2,051	2,066	2,157	2,079	2,056
Ability at 60 wks	2,062	2,076	2,085	2,159	2,088	2,061

Notes: This table reports the results of the counterfactual simulations described in the text, while assuming that training contracts are optimally adjusted in response. The “wage factor” is the scale factor, $\bar{w}$, in Equation 10. Under the 50% debias and 100% debias counterfactuals, worker overconfidence is reduced by 50% or 100% (by reducing η_b 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. The ban training contracts counterfactual is different than just setting the quit penalty to 0 because the firm also optimally adjusts the wage factor. 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 simulated here corresponds to column 2 of Table 6.

FOR ONLINE PUBLICATION ONLY

Web Appendix

This Web Appendix contains additional material for “Training Contracts, Worker Overconfidence, and the Provision of Firm-Sponsored General Training.” Section [B](#) provides a one-period version of paper’s structural model that can be used to make analytical predictions. Section [C](#) provides omitted derivations from the structural model. Section [D](#) provides further discussion of our productivity measure, including the issue of measurement error. Section [E](#) provides more information about the field experiment at Firm B. Section [F](#) provides additional results on the reduced-form impact of training contracts, accompanying Section [3](#) in the main text. Section [G](#) gives more results on worker subjective beliefs about their productivity, accompanying Section [4](#) in the main text. Section [H](#) provides additional tables on the results of the field experiment. Section [I](#) contains robustness checks for our structural estimates and counterfactuals. Section [J](#) provides further information on our interviews with HR managers on the extent of training and training contracts in trucking. Section [K](#) describes a few relevant legal issues on training contracts.

A Additional Discussion and Results

A.1 Non-response to the Firm A Beliefs Survey

As discussed in Section [4](#), the overall response rate to the Firm A beliefs survey was 28%, computed by averaging over all drivers and weeks in the estimation sample.¹ This rate may seem low, but is comparable to that in many nongovernmental surveys. For example, [Card et al. \(2012\)](#) find a response rate of 20% in a survey of UC Berkeley employees. We suspect one reason our survey’s response rate was low is because the survey was entirely voluntary, with no communication at all regarding the survey from supervisors. Although we may have gotten higher participation if supervisors had encouraged workers to participate, we deliberately conducted the survey this way so as to eliminate any desirability pressure from supervisors (which, had we done so, might have contaminated worker response).

As seen in Figure [F1](#), the survey response rate is fairly constant over driver tenure, despite workers leaving the sample by quitting. 61% of data subset workers respond to at least one survey. Drivers who are white, male, older, and have higher schooling are more likely to respond each week, as seen in Table [F1](#).

To address non-random response, we use Inverse Probability Weighting. There are two stages in estimation with Inverse Probability Weighting. In the first stage, we fit a probit model of whether a driver responds to the survey as a function of time-invariant observables (race, gender, and age and marital status when started work with Firm A). None of these covariates are otherwise used in the estimation of the structural model. In the second stage, we use the inverse of the predicted values to weight each driver’s contribution to the likelihood, and perform maximum likelihood. The identifying assumption for inverse probability weighting is that survey response is missing at random conditional on the observable characteristics used in the first stage probit model. As seen in Table [H1](#), our main structural results are quite robust to inverse probability weighting.

It is possible, still, that unobserved characteristics could affect non-response. However, a large advantage of our data is that we have a great deal of additional information about people that could

¹A previous version of the paper reported a response rate of 21%; this number was based on a larger data sample that included weeks where the driver was not working.

potentially affect their response beyond these covariates, including cognitive ability, personality, and experimental preferences. For example, one might think that people with low cognitive ability, a disagreeable personality, or those who are impatient might be less likely to participate. This allows us to turn facts which are generally unobservable into observables. Our results are robust to this.

While it is still possible that additional unobserved characteristics may affect non-response, we believe non-response is unlikely to explain our main conclusions at least on worker overconfidence, given our reduced-form results in Figure 5 indicating similar patterns of overprediction and learning with all subjects and conditional on survey response. For standard errors, we ignore estimation error from the first stage probit; doing so leads to conservative standard errors (Wooldridge, 2002).

A.2 Exit Survey

One limitation of personnel data is we generally do not see where drivers go when they terminate, so to remedy this, we did an exit survey by mail for drivers in the data subset. In drivers leaving the firm, only 12% report moving to another long-haul trucking firm. Roughly 50% of drivers moved to a non-trucking job or unemployment, and the remainder report moving to non-long haul trucking firms (e.g. local delivery drivers). One concern about doing an exit survey is that workers may lie about where they went next. For our survey, however, it was repeatedly emphasized to drivers that their responses were anonymous and would never be seen by the company, presumably eliminating incentives to lie. Another concern is that drivers who respond to the exit survey may be non-representative, as the response rate on the exit survey was only about 30%. However, whether a driver responded to the exit survey is uncorrelated with most observables, including average worker productivity and basic demographic variables, suggesting the results from the exit survey are unlikely to be biased by non-response.

A.3 Impact of Training Contracts, Robustness

We have repeated the results in Table 2 with different types of controls, including month of hire fixed effects, week of hire fixed effects, and school by year of hire fixed effects. The main results are similar across most of these specifications—the impacts are slightly larger with month of hire or week of hire fixed effects, and are slightly smaller with school by year of hire fixed effects. To examine further how the effects of the contracts varied with tenure, we ran linear probability models of quitting on interactions of the contracts with worker quarter of tenure:

$$Q_{it\tau} = \sum_{r=1}^8 \beta_{1r} * 12MC_i * TQ_r + \sum_{r=1}^8 \beta_{2r} * 18MC_i * TQ_r + \beta_3 * UE_{it} + \beta_4 \bar{y}_{it} + \eta_t + \gamma_\tau + I^{yearhire} + I^{school} + X_i \lambda + \epsilon_{it\tau} \quad (12)$$

where $Q_{it\tau}$ is a dummy for quitting and TQ_r is a dummy for a driver being in the r th quarter of tenure (e.g. $r = 1$ means the driver has 1-13 weeks of tenure, $r = 2$ means 14-26 weeks of tenure, etc.). The estimates are shown in Figure 3.

A.4 Understanding the Quit Impacts: 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, e.g. quit penalties might deter workers with low productivity or low taste for the job from signing up. (Note that workers selecting into the firm is different from the previously considered possibility of workers sorting between schools.)

In three tests, we provide evidence that the effect of the training contracts on quitting operated primarily through incentives. Because the main purpose of the paper is to study the interaction of training contracts with overconfidence, as opposed to separating incentive from selection effects, the tests are described here (for brevity).

One informal test for selection is to see whether training contracts affect firing. If training contracts deterred low-quality drivers from working at Firm A, one would expect a decrease in the rate of firing. However, Table E2 shows no statistically significant evidence that the contracts affected firing, though the estimates are less precise than those on quitting.

A second test is to examine whether selection occurred on various observable characteristics, the most obvious being productivity: Did adding a training contract lead more productive workers to begin working at the firm? As seen in Table E1, 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 taste for the job or tendency to quit) are more likely to choose to work for the firm once training contracts are in place. 1 out of 10 coefficients are statistically significant at 5%, so there is little evidence for this.²

Our 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 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.³

These tests provide support for some selection due to training contracts, but the effects seem limited. Given the strong evidence of selection effects of contracts in other personnel settings (e.g. Lazear, 2000), why does positive selection here seem small? One possibility is that for some reason the contract may not have been salient to drivers when they signed up for the job. However, this seems unlikely to be the case, as a discussion of training contracts was a mandatory part of interviews at Firm A, according to managers. What seems a more likely possibility to us is that workers lack private information about their productivity when signing up. Unlike in Lazear (2000), the workers here are new to long-haul trucking. Long-haul trucking is very different from most other jobs and it may be difficult to predict how good one will be at it.

A.5 Table 10 Counterfactuals

In the counterfactuals in Table 10, we 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

²We display our test for selection using all our data to maximize power. A concern, however, is that because more productive people are less likely to quit, the impact of the training contracts on quitting could confound their impact on selection. To address this issue, we re-did our analysis restricting to drivers in their first 6 months. As with the main results, we found no evidence that the contracts induced positive selection. Repeating column 6 of Table E1 while restricting to when drivers are in their first 6 months, the coefficient on the 12-month contract is -10 miles (standard error=13) and the coefficient on the 18-month contract is -19 miles(se=16).

³This 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. Our test is significantly more indirect, given that we cannot observe the same individual under multiple contractual regimes.

approach we have also explored is to use the same $\bar{u}$ in each counterfactual. Optimal profits are often lower than in Table 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.

B One Period Model

In this section, we present a model of training contracts and turnover to accompany the discussion in Section 2 and derive analytical predictions. We show that both allowing firms to use training contracts and the existence of worker productivity overconfidence reduce quitting and increase the profitability of training. Further, we show that overconfidence increases the optimal training contract. The model is a 1-period version of our structural model in Section 5, abstracting away from dynamics and worker heterogeneity.⁴ The model is fairly similar to that in Peterson (2010), with one difference being that we assume monopsony in the market for training whereas Peterson (2010) assumes perfect competition. Both our papers allow for competition in the post-training labor market.

Consider a firm which trains its workers. Workers are risk-neutral and have an initial productivity of zero at the firm and a pre-training 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. We assume that ε has a distribution function F and has support over the entire real line. Defining $\Psi(x) = 1 - F(x)$, we assume that the function $x \mapsto x\Psi^{-1}(x)$ is concave. This property is satisfied by many distributions including the uniform and normal distributions. If the firm chooses to train, it also chooses a piece rate w to pay, so the worker's total earnings are $W = w\eta$. 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 $\bar{W}$ minus any training contract k . We assume only that the outside option after training is greater than or equal to the worker's outside option before training ($\bar{W} \geq r$). The case of $\bar{W} = r$ corresponds to the worker choosing whether to go to another occupation whereas $\bar{W} = \eta$ may correspond to the worker choosing to leave for another firm within the same occupation (if training is portable across firms, but occupation-specific). We can also think of $\bar{W} = r$ as specific training and $\bar{W} = \eta$ as fully general training. Recognizing that enforcement of the contract may be imperfect, we allow that only a share $\theta \in [0, 1]$ of the contract is collected (the firm collects θk). (Our results hold for the case of perfect enforcement, $\theta = 1$, unless stated otherwise.) We assume, however, that the training contract is experienced fully by the worker (that is, the worker suffers a utility loss of k after quitting). That is, although the worker is initially credit-constrained, we assume after training that the worker can be penalized, either by making payments to the firm (which the worker can do with either increased income or credit access obtained because of training) or other means (e.g. damage to the worker's credit score or utility loss from pestering).⁵ The timing of the model is as follows:

⁴However, the model here is also more general in that it uses a more general distribution for the taste shocks. In addition, the model is an optimal contracting problem. Thus, we show that allowing training contracts and the existence of overconfidence increase training profits, even when firms are optimally setting contracts. Many models of firm-sponsored general training have two periods, see the review in Acemoglu and Pischke (1999a). Our 1-period model takes the 2-period training model and collapses the first period of training to a single point in time. That is, we make the assumption that training occurs instantaneously, which is consistent with the brief nature of training in trucking.

⁵We believe our assumption, that drivers act as if the utility cost of quitting is equivalent to the utility loss from paying the contract penalty, is reasonable in our setting; see the footnote in Section 5.1 for more on this. Our model conclusions should still hold if the worker's utility cost for quitting is less; a fine of \$2 of which the firm collects 25% and the worker pays 50% in utiles operates the same as a fine of \$1 of which the firm collects 50% and the worker pays 100% in utiles. By the term *training profits*, we mean the profits a firm receives from hiring and training one worker (vs. not hiring and training the worker). Although we talk here of the *profitability* of general training, we could

1. The firm chooses whether to train, and if so, sets the worker's earnings, W , and the level of the training contract, k .
2. The worker decides whether or not to accept the contract, and training occurs.
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:

$$\max_{w,k} (1 - F(-W - k + \bar{W})) * (\eta - W) + F(-W - k + \bar{W})\theta k - c$$

$$E \max (W + \varepsilon, \bar{W} - k) \geq r$$

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

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

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

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

If $\lambda = 0$, an interior solution exists only if $(1 - F(-W - k + \bar{W})) = -\theta F(-W - k + \bar{W})$, which is not possible for $\theta > 0$.⁶ Because the IR constraint cannot hold with equality while setting $k = 0$ (since $\bar{W} \geq r$ and ε has full support over the real line), $k = 0$ cannot be optimal.

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

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

$$= P(\eta - \bar{W}) + k(P + (1 - P)\theta) + P\Psi^{-1}(P).$$

Note that the IR constraint can be written as $r \leq \mathbb{E}max\{\bar{W} - \Psi^{-1}(P) - k + \varepsilon, \bar{W} - k\}$ or as $r \leq \mathbb{E}max\{\varepsilon - \Psi^{-1}(P), 0\} + \bar{W} - k$. 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 $k = k(P)$.

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

$$\eta - \bar{W} + \Psi^{-1}(P) + P\Psi^{-1}'(P) = 0. \quad (13)$$

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

$$\eta - \bar{W} + \Psi^{-1}(P) + P\Psi^{-1}'(P) + k'(P) \times (P + (1 - P)\theta) + k(1 - \theta) = 0. \quad (14)$$

Given that $k'(P) \times (P + (1 - P)\theta) + k(1 - \theta)$ is positive for all P and that $\Psi^{-1}(P) + P\Psi^{-1}'(P)$ is decreasing in P (by the concavity of $P\Psi^{-1}(P)$), it follows that the P that solves Equation (14) (where the firm optimally sets k) is greater than the P that solves Equation (13) (where $k = 0$). ■

Next, we examine the impact of productivity overconfidence on quitting. We temporarily ignore the issue of optimal contracting, focusing instead on the behavior of two workers, both of whom are receiving the same wage and the same training contract. If a worker is overconfident, we let $B(W)$ denote his belief about his earnings in the inside option, and $B(\bar{W})$ be his belief about his earnings in the outside option.

also frame the model to deliver results on the *probability* of training. Specifically, instead of assuming a fixed cost of training, c , we could assume that c was a random variable.

⁶The boundary solutions for the unconstrained problem are to have W go to minus infinity and k go to positive infinity faster (all workers stay and get paid minus infinity) or to have W go to minus infinity and have k go to positive infinity slower (all workers quit and the firm collects infinity from them). Both of these solutions violate the IR constraint.

Proposition 2 *Consider two workers with the same ability, training contract, and piece rate, one worker who is overconfident and one who is not. Then the overconfident worker will be less likely to quit than the rational worker if and only if he is more overconfident about his inside than his outside option.*

Proof. The probability of staying for the rational worker is $1 - F(\bar{W} - W)$, whereas it is $1 - F(B(\bar{W}) - B(W))$ for the overconfident worker. The probability of staying is higher for the overconfident worker when $B(W) - B(\bar{W}) > W - \bar{W}$ or when $B(W) - W > B(\bar{W}) - \bar{W}$. ■

We test this proposition in Table 4. Specifically, we regress quitting on worker subjective beliefs and average productivity to date. We want to use Beliefs instead of (Beliefs - Productivity) as the main regressor since the probability of staying, $1 - F(B(\bar{W}) - B(W))$, depends only on beliefs, and not the difference between beliefs and productivity. Empirically, we find that workers with higher beliefs are less likely to quit.⁷

In the next propositions, we examine the impact of overconfidence when firms optimally set contracts. 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. We first consider overconfidence that persists after the worker goes through training. After this, we will consider overconfidence that goes away after training via learning. In a change of notation from Proposition 2, we use η' to denote the worker's belief about his post-training productivity at Firm A, so his believed post-training Firm A earnings are $w\eta'$. Further, we use $\bar{W}$ to denote the worker's *believed* outside option earnings (instead of the actual outside earnings). The firm's problem can then be written as:

$$\arg \max_{w,k} (1 - F(-w\eta' - k + \bar{W})) * (1 - w)\eta + F(-w\eta' - k + \bar{W})\theta k - c$$

$$E\max(w\eta' + \epsilon, \bar{W} - k) \geq r$$

Proposition 3 *An increase in permanent overconfidence leads to higher training profits, 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 $w > \frac{d\bar{W}}{d\eta'}$, where w solves $\eta - w\eta = G(-w\eta + \bar{W})$; then an increase in η' increase the profitability of training.*

Proof. Note that the outside option $\bar{W}$ 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 $\eta' > \eta$, setting a wage of w^* is still feasible (as $E\max(w\eta' + \epsilon, \bar{W}) > E\max(w\eta + \epsilon, \bar{W})$). Further, given the rational wage w^* , an increase in productivity beliefs increases profits so long as $\text{sgn}\left(\frac{\partial(1 - F(-w\eta' - k + \bar{W}))}{\partial\eta'}\right) = \text{sgn}\left(-f * \left(-w + \frac{d\bar{W}}{d\eta'}\right)\right) = \text{sgn}\left(w - \frac{d\bar{W}}{d\eta'}\right) = +$. 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

⁷Instead of comparing an overconfident worker with a rational worker in Proposition 2, we could alternatively examine the impact on retention of slightly raising a worker's overconfidence (that is, his belief about his inside option). This will increase retention if and only if $\frac{\partial B(W)}{\partial B(\bar{W})} < 1$. So, when the problem is re-phrased this way, the required assumption is not differential overconfidence, but rather that beliefs about the outside option rise less than one-for-one with beliefs about the inside option. The assumption that outside beliefs rise less than one for one with inside beliefs is closely related to the assumption of differential overconfidence, and also seems quite plausible in our setting, for many of the same reasons that we give in Section 5.2.

productivity will be $\eta' > \eta$ and increasing beliefs from η to η' increases profits under (w^*, k^*) so long as $w > \frac{d\bar{W}}{d\eta'}$ 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, we 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 4 *An increase in non-permanent overconfidence leads to higher training profits 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. ■

We now analyze the effect of overconfidence on the optimal level of the training contract. An increase in overconfidence can be thought of as having “income” and “substitution” effects for the optimal contract. When the employee becomes more overconfident, he has become “richer” in terms of expected utility; thus, the firm can either increase the quit penalty or decrease the wages while still satisfying the IR constraint. This is the income effect. For the substitution effect, we note that the worker’s indifference curve in quit penalty-wage space is upward sloping; that is, the worker would receive the same utility from a contract with a high wage and high training contract, and a contract with a low wage and a low training contract. Becoming more overconfident about his ability makes the worker like higher wages more (because wage and ability are multiplicative) and dislike quit penalties less (because he thinks he will be less likely to quit). This flattens his indifference curve, making him prefer more both higher quit penalties and a higher wage. The income and substitution effects move in the same direction for the quit penalty.

Proposition 5 *Suppose that the share of the quit penalty collected is not “too large” (formally, $\theta < \eta * \min\{\frac{1}{k}, \frac{1}{\eta'}\}$). Then, an increase in permanent overconfidence leads to an increase in the optimal quit penalty.*

Proof. Let $\alpha = \frac{\eta'}{\eta}$. The IR constraint is given by $\int_{\bar{W}-k-w\alpha\eta}^{\infty} (w\alpha\eta + \varepsilon) f(\varepsilon) d\varepsilon + (\bar{W} - k) F(\bar{W} - k - w\alpha\eta) =$

r . Differentiating with respect to k and applying Leibnitz’s Rule, we get that $\frac{\partial w}{\partial k} = \frac{F}{\alpha\eta(1-F)}$. We plug this into the first order constraint $f(1 + \alpha\eta\frac{\partial w}{\partial k})(1 - w)\eta - (1 - F)\eta\frac{\partial w}{\partial k} - f(1 + \alpha\eta\frac{\partial w}{\partial k})\theta k + \theta F = 0$ to obtain the condition that: $\alpha((1 - w)\eta - \theta k)H - F + \alpha\theta F = 0$.

Next we apply the Implicit Function Theorem to solve for $\frac{\partial k}{\partial \alpha}$.

$$M_1 = \alpha((1 - w)\eta - \theta k)H - F + \alpha\theta F = 0$$

$$M_2 = \int_{\bar{W}-k-w\alpha\eta}^{\infty} (w\alpha\eta + \varepsilon) f(\varepsilon) d\varepsilon + (\bar{W} - k) F(\bar{W} - k - w\alpha\eta) - r = 0$$

$$\begin{pmatrix} \frac{\partial M_1}{\partial k} & \frac{\partial M_1}{\partial w} \\ \frac{\partial M_2}{\partial k} & \frac{\partial M_2}{\partial w} \end{pmatrix} \begin{pmatrix} \frac{\partial k}{\partial \alpha} \\ \frac{\partial w}{\partial \alpha} \end{pmatrix} = - \begin{pmatrix} \frac{\partial M_1}{\partial \alpha} \\ \frac{\partial M_2}{\partial \alpha} \end{pmatrix}$$

$$\begin{pmatrix} \frac{\partial k}{\partial \alpha} \\ \frac{\partial w}{\partial \alpha} \end{pmatrix} = - \left(\begin{array}{c|c} -\alpha\eta(1 - w)H' - \alpha\theta H & -\alpha\eta H + \alpha\eta f \\ \hline +\alpha\theta k H' + f & -\alpha^2\eta H'((1 - w)\eta - \theta k) \end{array} \right)^{-1} \begin{pmatrix} ((1 - w)\eta - \theta k)H + w\eta f \\ \hline -\alpha w((1 - w)\eta - \theta k)\eta H' \end{pmatrix}$$

This gives us that

$$\frac{\partial k}{\partial \alpha} = H(\eta - k\theta) \frac{1 - F}{-f + FH + H\theta\alpha + \alpha\eta H' - FH\theta\alpha - k\theta\alpha H' - w\alpha\eta H'}$$

We note that the numerator will be positive whenever $\theta < \frac{\eta}{k}$. Further, note that the denominator can be re-written as $H'(\alpha\eta(1-w) - \alpha\theta k) + \alpha\theta H + H(F(1-\theta\alpha)) - f$. However, in order for k to maximize profits, we must have that $\frac{\partial^2 \pi}{\partial k^2} < 0$. Given that $M_1 = \alpha \frac{\partial \pi}{\partial k}$, if $\frac{\partial^2 \pi}{\partial k^2} < 0$, it then follows that $\frac{\partial M_1}{\partial k} < 0$, which then implies that $H'(\alpha\eta(1-w) - \alpha\theta k) + \alpha\theta H - f > 0$. Thus, a sufficient condition for the denominator to be positive is for $H(F(1-\theta\alpha)) > 0$, which occurs when $\theta < \frac{1}{\alpha} = \frac{\eta}{\eta'}$. ■

While our proof assumed that $\theta < \eta * \min\{\frac{1}{k}, \frac{1}{\eta'}\}$, we suspect that the proposition may hold in general. In numerical simulations we have run, the proposition appears to hold for all $\theta \in [0, 1]$. By assuming particular functional forms for ϵ , we can sometimes derive an analytic expression for the optimal training contract.

Example 6 Suppose that ϵ has a logistic distribution. Further, suppose that none of the quit penalties are collected ($\theta = 0$) and let worker ability be equal to one ($\eta = 1$). Then the optimal quit penalty is $k = \bar{W} - \ln(\exp(r) - \exp(\eta' - 1))$, which is increasing in worker overconfidence.

C Structural Model and Estimation Details

Estimation Sample. For the structural analysis, we start with the 895 data subset drivers who completed training. We drop any drivers who are ever seen working at non-piece rate trucking jobs at Firm A 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). We also drop a few drivers missing any of several of the key covariates, leaving an estimation sample of 699 drivers.

Probability of Staying. Let $\Lambda(x) = \frac{\exp(x)}{1+\exp(x)}$. At time T , the probability of staying, given the state variables, is:

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

To evaluate this probability, we need to calculate both $E^b(y_T | y_1, \dots, y_{T-1})$ and $E^b(V(\mathbf{x}) | \mathbf{x}_T)$. The former depends on $y_1, \dots, y_{T-1}$, 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 (both a model with standard beliefs and our generalized learning model), the worker's expectation of future productivity depends only on his prior and his demeaned average of past productivity. That is, *the demeaned average of past productivity is a sufficient statistic for the entire sequence $y_1, \dots, y_{t-1}$* (DeGroot, 1970).

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

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

Calculating $E^b(V_{t+1}(\mathbf{x}_{t+1}) | \mathbf{x}_t)$ requires integrating expectations of both future miles and idiosyn-

cratic shocks:

$$E^b(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) = E_{y_t}^b E_{\epsilon|y_t}^b(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) \quad (15)$$

$$= E_{y_t}^b E_{\epsilon}^b(\max\{\bar{V}_{t+1}^S(\mathbf{x}_{t+1}) + \epsilon_{t+1}^S, \bar{V}_{t+1}^Q + \epsilon_{t+1}^Q\}|\mathbf{x}_t) \quad (16)$$

$$= \int \tau \log \left(\exp\left(\frac{\bar{V}_{t+1}^S(\mathbf{x}_{t+1})}{\tau}\right) + \exp\left(\frac{\bar{V}_{t+1}^Q}{\tau}\right) \right) f^b(y_t|y_1, \dots, y_{t-1}) dy_t \quad (17)$$

$$= \sum_k \tau \log \left(\exp\left(\frac{\bar{V}_{t+1}^S(\mathbf{x}_{t+1})}{\tau}\right) + \exp\left(\frac{\bar{V}_{t+1}^Q}{\tau}\right) \right) \Pr(y_t^k|y_1, \dots, y_{t-1}). \quad (18)$$

(15) expresses that the value function depends on future miles and future idiosyncratic shocks. (16) uses the definition of V , that V_s^Q is independent of the state variable, and that the idiosyncratic shocks are independent of miles. (17) integrates out y_t , which are not observed when the driver makes his period t decision, but are observed in the future, as well as integrates out the idiosyncratic shocks. (18) follows because, in implementation, miles will be discretized into K possible values. The probability $\Pr(y_t^k|y_1, \dots, y_{t-1})$ can be easily shown to depend only on $\bar{y}_{t-1}$, and is expressed below in (25). As long as average miles to date by time $t+1$, $\bar{y}_t$, is a sufficient statistic for x_{t+1} , then it follows that average miles by time t , $\bar{y}_{t-1}$, is a sufficient statistic for x_t . Related derivations can be found in Rust (1987) and Stange (2012).

Likelihood Function. Let $L_i = L(d_{i1}, \dots, d_{it}, y_{i1}, \dots, y_{it}, b_{i1}, \dots, b_{it})$ be the likelihood of driver i for an observed sequence of quitting decisions, miles realizations, and subjective beliefs. We show how to derive the likelihood function.

$$L_i = \int L(d_{i1}, \dots, d_{it}, y_{i1}, \dots, y_{it}, b_{i1}, \dots, b_{it}|\alpha, \eta_b) f(\alpha, \eta_b) d\alpha d\eta_b \quad (19)$$

$$= \int \{L(d_{i1}, \dots, d_{it}|y_{i1}, \dots, y_{it}, b_{i1}, \dots, b_{it}, \alpha, \eta_b) * L(b_{i1}, \dots, b_{it}|y_{i1}, \dots, y_{it}, \alpha, \eta_b) * L(y_{i1}, \dots, y_{it}|\alpha, \eta_b)\} f(\alpha, \eta_b) d\alpha d\eta_b \quad (20)$$

$$= \left[\int L(d_{i1}, \dots, d_{it}|y_{i1}, \dots, y_{it}, \alpha, \eta_b) * L(b_{i1}, \dots, b_{it}|y_{i1}, \dots, y_{it}, \eta_b) f(\alpha, \eta_b) d\alpha d\eta_b \right] L(y_{i1}, \dots, y_{it}) \quad (21)$$

$$= \left[\int \prod_{s=1}^t L(d_{is}|d_{i1}, \dots, d_{is-1}, y_{i1}, \dots, y_{it}, \alpha, \eta_b) * \prod_{s=1}^t L(b_{is}|b_{i1}, \dots, b_{is-1}, y_{i1}, \dots, y_{it}, \eta_b) f(\alpha, \eta_b) d\alpha d\eta_b \right] L(y_{i1}, \dots, y_{it}) \quad (22)$$

$$= \left[\int \prod_{s=1}^t L(d_{is}|y_{i1}, \dots, y_{is-1}, \alpha, \eta_b) \prod_{s=1}^t L(b_{is}|y_{i1}, \dots, y_{is-1}, \eta_b) f(\alpha, \eta_b) d\alpha d\eta_b \right] \left(\prod_{s=1}^t L(y_{is}|y_{i1}, \dots, y_{is-1}, \alpha, \eta_b) \right) \quad (23)$$

$$\equiv \left\{ \int L_i^1(\alpha, \eta_b) L_i^3(\eta_b) f(\alpha, \eta_b) d\alpha d\eta_b \right\} L_i^2 \quad (24)$$

Equations (19) and (20) follow by the law of total probability. (21) holds because productivity is unaffected by the taste and overconfidence heterogeneity and because beliefs are unaffected by the taste heterogeneity. (22) follows because (a) future miles 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. (23) follows because (a) since the ϵ shocks are iid, the decision to quit is conditionally independent of all prior decisions to quit (given the miles 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 (24), we define the part of the likelihood due to the quitting decisions as $L_i^1(\alpha, \eta_b)$, the part due to the miles realizations as L_i^2 , and the part due to subjective beliefs as $L_i^3(\eta_b)$.

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

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

with

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

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 $\gamma_s^b = \frac{s\sigma_0^2}{s\sigma_0^2 + \sigma_y^2}$.⁸

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.

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

Transitions Between Miles. As mentioned in the main text, we discretize productivity into K values. In our baseline estimation, we let productivity range in increments of 100 from 100 to 4,000 miles per week, *i.e.* $K = 40$. Transitions between miles states are given by:

$$\Pr(y_s^k | y_1, \dots, y_{s-1}) = \Phi \left(\frac{y_s^k + .5 * kstep - E^b(y_s^k | y_1, \dots, y_{s-1})}{\sqrt{\Omega_{s-1}}} \right) - \Phi \left(\frac{y_s^k - .5 * kstep - E^b(y_s^k | y_1, \dots, y_{s-1})}{\sqrt{\Omega_{s-1}}} \right) \quad (25)$$

where $kstep$ is the distance between miles realizations. See [Rust \(1996\)](#) and [Stange \(2012\)](#) for similar formulas.

Estimation Procedure. The model is estimated by maximum likelihood using an extension of the canonical nested fixed point algorithm ([Rust, 1987](#)). For every parameter guess, we first use value function iteration to solve for the asymptotic value functions (V_S and V_Q). We these in hand, we use backwards recursion to solve for the choice-specific value function V_t^S and V_t^Q for $t = 1, \dots, T$.

12-Month and 18-Month Contracts in Structural Model. The quit penalties under the training contracts varied slightly by school. For the structural estimation, we assume a penalty of \$3,750 for the 12-month contract. For the 18-month contract, we assume a penalty of \$5,250 that is reduced linearly over 78 weeks to 0.

Zero Mile Weeks. 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 miles component

⁸This follows by applying the standard formula for the conditional density for a multivariate normal distribution: $X_1 | (X_2 = x_2) \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

of the likelihood, and average miles to date (in terms of the quit decision) is given by the prior week’s average miles to date.

D Measuring Productivity

Drivers at Firm A, like most long-haul trucking firms, are paid by the mile. 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 on salary instead of by the mile (e.g. drivers who work full time as driver trainers at the training schools).⁹

Per the federal Hours-of-Service Regulations, truckdrivers are legally restricted from working more than 60 hours per week. Instead of 60 hours in 7 days, drivers can alternatively work up to 70 hours over eight days. See <http://www.fmcsa.dot.gov/rules-regulations/topics/hos/index.htm>, accessed in October 2010.

As reported in the main text, trucking loads at Firm A are assigned by a company-wide central dispatcher instead of the driver’s boss and are assigned primarily according to proximity. According to Firm A managers, load assignment is not based on driver ability (e.g. assignment of more desirable loads to good drivers). In addition, there is no scope for boss-worker favoritism, since the driver’s boss, with whom he interacts with over the week, does not assign him loads.¹⁰ Firm A is a leading firm and it is rare there is a shortage of loads. During the time period we study, the firm had basic on-board computers (Hubbard, 2003), but drivers were responsible for all route planning and time management.

In addition, there is a small amount of measurement error in our productivity measure, miles per week. We describe the source of the measurement error, explain how we correct for it, and show it has little impact on our estimates and conclusions.

Miles are imperfectly observed each week because drivers miles are only recorded once a driver reaches his destination. If the driver is in the middle of a load at the end of the week, the miles on that load performed during the concluding week will be counted toward miles on the week just beginning. This issue arises in the mileage data for Firm A and Firm B. Without information on loads, it is difficult to correct for measurement error.

To correct for measurement error and address how much measurement error affects our results, we requested and analyzed new *load-level data* from Firm A, covering most drivers over a 9-week period. With the new data, we assign half of the mileage from a driver’s first load each week to the current week and half to the previous week for any loads spilling over weeks.¹¹ It should be noted that even for the newer load-level data, we are still not observing actual miles within the week-time exactly. However, we can come much closer to a driver’s true productivity in a given week.

Using the new load-level data, we develop a simple algorithm to correct week-level data for measurement error. The load-level data provides “true miles” in a week. We create week-level data, as in the main Firm A dataset, by aggregating loads by week and adding the measurement error.

⁹In a later paper (Burks et al., 2013), we have combined the Firm A data with data from other firms to study differences across workers in terms of whether they were referred for the job by incumbent employees. Training contract status and overconfidence are both uncorrelated with whether drivers are referred.

¹⁰We note that even if there were non-random assignment of drivers to loads, this would not affect the main message or conclusions of the paper, only the interpretation of what drivers are overconfident about. Whether drivers are overconfident about how quick will be at delivering loads or whether they are overconfident about what type of loads they will be assigned, they will still be more likely to sign training contracts and less likely to quit after training, if they are overconfident relative to their outside option.

¹¹While most of the Firm A data are from payroll records, the load level data are created from operational records.

The basic idea for the algorithm is when we observe a low-mileage week followed by a high-mileage week, we transfer some miles from the high to the low mileage week because a small portion of the difference is likely measurement error.

Formally, note that the observed miles in week-level data, y_t^m , is equal to:

$$y_t^m = y_t + \alpha_{t-1}A(t-1, t) - \alpha_t A(t, t+1)$$

where y_t is true miles; $A(t-1, t)$ is the number of miles from a load that started in week $t-1$ and ended on week t ; and α_{t-1} is the share of those miles that were completed in week $t-1$. That is, observed miles are true miles, plus spillover miles from the past week to the current week, minus spillover miles from the current week to the next week. We re-arrange this equation to get $y_t = y_t^m - \alpha_{t-1}A(t-1, t) + \alpha_t A(t, t+1)$. To empirically implement our best guess of y , we consider the regression equation:

$$y_{i,t} = y_{i,t}^m - \beta(y_{i,t}^m - y_{i,t-1}^m) + \beta(y_{i,t+1}^m - y_{i,t}^m) + \epsilon_{i,t}$$

We wish to find the β that leads to the smallest sum of square deviation between our “corrected” productivity measure, the right-hand side, and the “true” productivity measure, the left-hand side. By moving $y_{i,t}^m$ to the left-hand side, we can estimate this equation by OLS, obtaining an estimate $\hat{\beta} = 0.091$, with a standard error of 0.001.

Having developed the algorithm with the load-level data, we can now apply the algorithm to the main Firm A data and do a robustness check on the impact of controlling for measurement error. Re-doing the results on whether productivity beliefs predict productivity in Table F2, we find very similar results with the measurement error correction as before without it. The structural estimates are also quite robust to correcting for measurement error; see Table H1.

While the paper focuses at times on production differences between short-haul and long-haul trucking, we note that Firms A and B are both part of the Truckload segment of the trucking industry (long distance truckload comprises NAICS categories 484121 and 484230). See Burks et al. (2011) for more discussion on the trucking industry and differences between the Truckload and Less than Truckload segments.

E Field Experiments with Firm B, Further Information

The goals of the field experiment were to (1) See if the unincentivized results on productivity from Firm A were robust to incentives for accurate guessing, and (2) Analyze whether our main counterfactual of eliminating overconfidence is feasible. Our results are that incentives do not affect our results and that reducing overconfidence does indeed seem feasible.

Firm B is a large long-haul trucking firm. Unlike Firm A, Firm B does not train its own drivers, so all drivers in the study had already received a commercial driver’s license. These workers did not have training contracts. However, given that our two goals (seeing whether incentives and information affect overconfidence) do not directly involve training contracts, this should not be a concern.

We contacted all workers who had started at the company in the last few months. The experiment was conducted via weekly phone surveys. 272 workers participated in the experiment. Phone calls were made by one of the authors (Hoffman) and by undergraduate research assistants.

The first randomization in the experiment was whether or not a worker would receive incentives for accurately guessing about his mileage or earnings. Of the 272 workers, 134 were in the \$10 incentive group and 138 were in the No Incentive group. A picture of the experimental design is presented in Figure G1. There is a slight difference between the numbers in the two groups because for logistical reasons, workers were randomized to the \$10 incentive or control groups before they

agreed to participate in the experiment.¹²

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.¹³ Under a quadratic scoring rule, the agent's payoff is given by $A - B(x - b)^2$, 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, with $B > 0$. This mechanism is incentive-compatible if the subject is risk-neutral.¹⁴ 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. To preserve incentive-compatibility, we paid drivers for *either* their prediction of earnings or miles, with which one determined randomly. The \$10 bonus 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.¹⁵ All subjects were paid a \$5 participation fee for each survey taken.

Significant care was taken so that the incentives for accuracy would be understood by drivers. We told drivers that they would be rewarded for their accuracy and explained the payment amounts through examples. We explained to drivers that the reward rule was incentive-compatible, that is, "that you maximize your reward by stating your true beliefs." If drivers had further questions or wanted to know more, we walked them through additional examples and provided them with the exact mathematical formula for the reward. Our approach of telling subjects the quadratic scoring rule is incentive-compatible follows Radzevick and Moore (2011) and Hoffman (2012). The experimental instructions and survey wording are given below in Section E.2.

Workers guessed about their productivity and earnings for 2-6 weeks. The number of weeks for workers in Stage I varied based on the week on which workers were first contacted and the number of weeks for which we were unable to contact them for an interview. While there is substantial subject attrition through the experiment, there are no significant differences between the different experimental groups in attrition.

¹²Specifically, we did the randomization before workers agreed to participate because the phone interviewers explained the incentive system to workers in the incentive group immediately after the worker accepted.

¹³Several recent papers using quadratic scoring rules include Holt and Smith (2009), Radzevick and Moore (2011), and Hoffman (2012). While there is an active debate about when they should be used and in what form (Hossain and Okui, 2010; Offerman et al., 2009; Schlag and van der Weeley, 2009), quadratic scoring rules are a standard and established tool in experimental economics (Selten, 1998).

¹⁴To 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 :

$$\begin{aligned} & \operatorname{argmax}_b \int A - B(x - b)^2 f(x) dx \\ &= \operatorname{argmin}_b \int (x - b)^2 f(x) dx \\ & \text{FOC} : \int \frac{d}{dx} (x - b)^2 f(x) dx = 0 \\ & b = \int x f(x) dx \end{aligned}$$

¹⁵For 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. To examine the robustness of the results to larger incentives, we also used larger incentives of up to \$50 per week for a smaller number of drivers, as we describe below. The incentive system was designed in consultation with firm managers, who believed that both the small and larger incentive systems would be salient for drivers.

After 2-6 weeks, workers were assigned to receive debiasing (75 workers), no debiasing (73 workers), or a larger incentives treatment (26 workers). In the larger incentives treatment, drivers were paid up to \$50 per guess and faced sharper penalties for mistakes. These drivers were paid according to the rules $50 - 200 * ((x_m - b_m)/1000)^2$ and $50 - 800 * ((x_e - b_e)/1000)^2$.

The debiasing treatment consisted of telling workers at Firm B about the existence of overconfidence in the workers at Firm A, as well as reminding the Firm B workers of their average prediction so far to date. The no debiasing treatment consisted of simply reminding drivers of their average prediction so far to date.

Drivers selected for debiasing were read the following script. (After the first paragraph, drivers were asked if they had any questions or comments.)

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 your 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.

Drivers selected not to receive debiasing were simply told:

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

E.1 Regression Analysis with Field Experiments

Table 5 shows the effect of incentives and information on people's beliefs. It shows regressions of the form:

$$b_{it} = \alpha_0 + \alpha_1 10INCENT_{it} + \alpha_2 50INCENT_{it} + \alpha_3 DEBIAS_{it} + \alpha_4 NODEBIAS_{it} + \beta t + X_{it}\delta + \epsilon_{it}$$

where b_{it} is agent i 's subjective belief at tenure t ; $10INCENT_{it}$ is a dummy for having up to a \$10 incentive for guessing about productivity in week t ; $50INCENT_{it}$ is a dummy for having up to a \$50 incentive for guessing about productivity in week t ; $DEBIAS_{it}$ is a dummy for having received the debiasing treatment at or before week of tenure t ; $NODEBIAS_{it}$ is a dummy for having received the no debiasing treatment at or before week of tenure t ; X_{it} is other control variables; and ϵ_{it} is an error. The coefficients α_1 and α_2 are the impact of financial incentives for accuracy on worker beliefs. The coefficients α_3 and α_4 are the impact of information on worker beliefs. In Table 5, we estimate that the incentives have little impact, but that the debiasing did significantly reduce subjects' overprediction. The effects of debiasing have higher statistical significance for earnings than for miles, but the difference in the impacts between earnings and miles is not statistically significant when both are put in standardized units.¹⁶

We show the main results are robust in two robustness checks where we focus on either debiasing or incentives for accuracy. In Table G2, we restrict the sample to weeks where drivers have already

¹⁶To compare impacts on miles vs. impacts on earnings, we standardized the dependent variable and re-ran the regressions in Table 6. The overall impact on miles was -0.14 standard deviations (se=0.08 s.d.) and the overall impact on earnings was -0.20 s.d. (se=0.07 s.d.). After 4-6 weeks, the impact on miles was -0.08 s.d. (se=0.09 s.d.) whereas the impact on earnings was -0.20 s.d. (se=0.09 s.d.).

received the Debiasing or No Debiasing treatment in the current or a past work. In Table G3, we restrict to workers in Stage I of the experiment, where they are receiving either incentives or no incentives for accurate guessing. In this sample, as well, we again see no evidence of the incentives on mileage predictions, earnings predictions, or overconfidence (mileage prediction minus actual productivity that week)

E.2 Experiment Wording

E.2.1 Incentivized Version, \$10

[For subsequent surveys] Hi this is [FULL NAME] from the University of California, with the trucking survey. Might you like to participate again?

[First survey] In the next two questions, we're going to ask you to estimate your miles and earnings for next week, if you're willing. We're going to give you a small reward (in addition to the \$5) for predicting accurately. For example, if you run exactly the number of miles you predict, you get \$10. For each mile you're off, the reward will go down, with larger reductions the further you are off. If you're off by 500 miles, you get \$7.50. And if you're off by 1,000 or more miles, you get \$0. We'll use a similar reward system for your prediction on how much you will earn. This might sound complicated, but this system has been used in other research, and is specially designed so that you maximize your reward by stating your true beliefs. We'll pay you either for your miles or your earnings guess, with which one chosen randomly. Does this make sense? [If not, explain to them. Also, go through payment system on back if want to know more.]

[For subsequent surveys] Do you happen to remember how the reward system works, where you get rewarded for guessing close to the actual number of miles you run? [If not, refresh their memory.]

- How many miles do you expect to run next week, that is, from Tuesday until next Tuesday?
- How many dollars do you expect to earn before taxes next week, that is, from Tuesday until next Tuesday?

E.2.1.1 Further Information on Payment System to Give Respondents [This information was given to respondents when they had further questions about the quadratic scoring rule.]

Here are some further examples of how you will be paid for the miles prediction:

Distance Between Actual and Predicted Miles	Your Payment
Your guess equals the actual	\$10
Your guess is 250 miles from the actual	\$9.38
Your guess is 500 miles from the actual	\$7.50
Your guess is 750 miles from the actual	\$4.38
Your guess is 1,000 or more miles from the actual	\$0.00

Specifically, your payment will be given by the equation $\text{Payment} = \$10 - \$10 \times (\text{Actual Miles in Thousands} - \text{Predicted Miles in Thousands})^2$.

Here are some further examples of how you will be paid for the earnings prediction:

Distance Between Actual and Predicted Miles	Your Payment
Your guess equals the actual	\$10
Your guess is 100 dollars from the actual	\$9.60
Your guess is 200 dollars from the actual	\$8.40
Your guess is 300 dollars from the actual	\$6.40
Your guess is 400 dollars from the actual	\$3.60
Your guess is 500 or more dollars from the actual	\$0.00

Specifically, your payment will be given by the equation $\text{Payment} = \$10 - \$40 \times (\text{Actual Earnings in Thousands} - \text{Predicted Earnings in Thousands})^2$.

E.2.2 Unincentivized Version

[For subsequent surveys] Hi this is [FULL NAME] from the University of California, with the trucking survey. Might you like to participate again?

- How many miles do you expect to run next week, that is, from Tuesday until next Tuesday?
- How many dollars do you expect to earn before taxes next week, that is, from Tuesday until next Tuesday?

E.2.3 Increasing the Incentive to \$50 per Week

This week, we're going to do something a little different. You will earn up to \$50 for predicting accurately instead of \$10. For example, if you run exactly the number of miles you predict, you get \$50. For each mile you're off, the reward will go down, with larger reductions the further you are off. If you're off by 250 miles, you get \$37.50. And if you're off by 500 or more miles, you get \$0. We'll use a similar reward system for your prediction on how much you will earn. This might sound complicated, but this system has been used in other research, and is specially designed so that you maximize your reward by stating your true beliefs. We'll pay you either for your miles or your earnings guess, with which one chosen randomly. Does this make sense? [If not, explain to them. Also, go through payment system on back if want to know more.]

E.2.3.1 Further Information on Payment System to Give Respondents, Incentive up to \$50 per week [This information was given to respondents when they had further questions about the quadratic scoring rule.] Here are some further examples of how you will be paid for the miles prediction:

Distance Between Actual and Predicted Miles	Your Payment
Your guess equals the actual	\$50.00
Your guess is 125 miles from the actual	\$46.88
Your guess is 250 miles from the actual	\$37.50
Your guess is 375 miles from the actual	\$21.88
Your guess is 500 or more miles from the actual	\$0.00

Specifically, your payment will be given by the equation $\text{Payment} = \$50 - \$200 \times (\text{Actual Miles in Thousands} - \text{Predicted Miles in Thousands})^2$, so long as it is greater than 0.

Here are some further examples of how you will be paid for the earnings prediction:

Distance Between Actual and Predicted Earnings	Your Payment
Your guess equals the actual	\$50
Your guess is 50 dollars from the actual	\$48
Your guess is 100 dollars from the actual	\$42
Your guess is 150 dollars from the actual	\$32
Your guess is 200 dollars from the actual	\$18
Your guess is 250 or more dollars from the actual	\$0

Specifically, your payment will be given by the equation $\text{Payment} = \$50 - \$800 \times (\text{Actual Earnings in Thousands} - \text{Predicted Earnings in Thousands})^2$, so long as it is greater than 0.

E.3 Additional Questions asked in the Weeks After Debiasing or No Debiasing

- **Job Search.** Taking everything into consideration, how likely is it you will make a genuine effort to find a new job within the next 6 months?
 - Not at all likely

- Somewhat likely
 - Very likely
- **Job Satisfaction.** All in all, how satisfied are you with your job?
 - Not at all satisfied
 - Not too satisfied
 - Somewhat satisfied
 - Very satisfied

F Training Contract Impacts and Learning about Productivity, Further Results

Table E1: Selection Effects of Training Contracts

<u>Panel A: Selection on Productivity</u>						
Sample:	All Weeks		Exclude 0 Mile Weeks		Trim 5/95%	
	(1)	(2)	(3)	(4)	(5)	(6)
12m contract	0.17 (28.59)	-7.01 (28.55)	24.69 (23.95)	21.1 (23.63)	23.74 (18.68)	22.01 (18.58)
18m contract	-21.56 (31.01)	-25.44 (31.29)	11.78 (28.72)	8.31 (28.60)	13.62 (22.06)	10.37 (22.12)
Demog controls	No	Yes	No	Yes	No	Yes
Mean Dep. Var:	1661	1690	1963	1956	1951	1947
Observations	0.99 <i>M</i>	0.86 <i>M</i>	0.84 <i>M</i>	0.74 <i>M</i>	0.75 <i>M</i>	0.67 <i>M</i>
R-squared	0.24	0.22	0.05	0.05	0.04	0.04
<u>Panel B: Selection on Other Characteristics</u>						
	(1)	(2)	(3)	(4)	(5)	
Dep. Var:	Black	Hispanic	Female	Married	Age	
12m contract	-0.009 (0.014)	0.021** (0.010)	0.012 (0.010)	-0.016 (0.017)	0.111 (0.399)	
18m contract	0.006 (0.020)	0.014 (0.013)	0.023 (0.015)	-0.005 (0.024)	0.574 (0.541)	
Observations	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	0.72 <i>N</i>	
R-squared	0.03	0.03	0.01	0.01	0.01	

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, $M \gg 100,000$. All regressions include time fixed effects (for each month), year of hire fixed effects, 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 E2: Impact of the Training Contracts on Firing – Cox Model, Diff-in-Diff

Sample of drivers:	All	All	All	Has demogs	Has demogs	Has demogs
	(1)	(2)	(3)	(4)	(5)	(6)
12m contract	-0.031 (0.074)	-0.030 (0.075)	-0.063 (0.074)	0.001 (0.076)	-0.007 (0.076)	
18m contract	0.055 (0.102)	0.054 (0.102)	-0.002 (0.101)	0.013 (0.102)	0.006 (0.102)	
State unemployment rate		-0.005 (0.017)	-0.013 (0.018)	-0.023 (0.021)	-0.027 (0.021)	-0.028 (0.021)
Avg miles to date			-0.049*** (0.004)	-0.055*** (0.005)	-0.052*** (0.005)	-0.052*** (0.005)
12m contract * (wks≤52)						-0.016 (0.085)
12m contract * (52<wks≤78)						-0.162 (0.165)
12m contract * (wks>78)						0.138 (0.151)
18m contract * (wks≤52)						0.006 (0.113)
18m contract * (52<wks≤78)						-0.044 (0.186)
18m contract * (wks>78)						0.020 (0.212)
Time FE (yr)	Yes	Yes	Yes	Yes	Yes	Yes
Year of Hire FE	Yes	Yes	Yes	Yes	Yes	Yes
Training School FE	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	No	No	No	Yes	Yes
Observations	<i>M</i>	<i>M</i>	<i>M</i>	0.87 <i>M</i>	0.87 <i>M</i>	0.87 <i>M</i>

Notes: An observation is a driver-week. The regressions are Cox proportional hazard models, with coefficients reported. Standard errors clustered by school-week of hire in parentheses. Driver tenure is controlled for non-parametrically. State unemployment is the annual unemployment rate in a driver's state of residence. 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 (4) differs from column (3) in that it restricts the sample to be the same as in column (5). The exact *M* is withheld to protect firm confidentiality, *M* >> 100,000. * significant at 10%; ** significant at 5%; *** significant at 1%

In addition, the visual firing hazards looking very similar under the three contractual regimes. These were included in a prior version of the paper, though omitted here for brevity, and are available on request.

Figure E1: Conditional on Week of Tenure, Workers who Quit that Week Have Lower Average Productivity to Date Than Workers Who Stay

Notes: This figure analyzes the average miles to date in each week of all new drivers from 2002 to 2009 at Firm A. In each week, it compares the average miles of drivers who quit with those of drivers who do not quit that week. For example, it can be used to compare the average miles per week (from weeks 1-19) of drivers who quit in the 20th week with the average miles per week (from weeks 1-19) of drivers who survive to the 20th week, but do not quit. Zero mile weeks are excluded. The figures are plotted using a local polynomial regression with an Epanechnikov kernel. A bandwidth of 3 weeks is used for the non-quitting and a bandwidth of 3 weeks is used for the quitting line.

Table E3: Does the Unemployment Rate Predict Training Contract Changes?

Dep var:	Has 12m contract (1)	Has 18m contract (2)
State unemployment rate	-0.023 (0.031)	-0.018 (0.015)
Time FE (yr)	Yes	Yes
Year of Hire FE	Yes	Yes
Training School FE	Yes	Yes
R-squared	0.21	0.71
Observations	0.83M	0.89M

Notes: This table examines whether state unemployment rates predict whether drivers have training contracts or not, in an effort to examine whether training contract changes may have been correlated with labor market conditions. Column 1 analyzes drivers with either no contract or the 12-month contract. Column 2 analyzes drivers with either the 12-month contract or the 18-month contract. Standard errors in parentheses are clustered by driver's state of residence. Each column is an OLS linear probability model where the dependent variable is whether a driver has a 12-month contract (versus no contract) and whether a driver has an 18-month contract (versus the 12-month contract). The exact M is withheld to protect firm confidentiality, $M \gg 100,000$. * significant at 10%; ** significant at 5%; *** significant at 1%

G Worker Beliefs, Further Results

Figure F1: Response Rate to the Firm A Subjective Productivity Beliefs Survey, by Driver Tenure

Notes: This graph shows the average response to the beliefs survey time. The y-axis plots the number of drivers responding to the survey divided by the number of remaining drivers with the company. As can be seen, the response rate is fairly constant as driver tenure accumulates at between, usually between 20% and 30%.

Figure F2: Tenure and Overconfidence (Productivity Beliefs *minus* Productivity)

(a) Means

(b) Medians

(c) Medians, ≥ 75 Wks, Responder

Notes: This figure analyzes the evolution of average driver overconfidence as a function of driver tenure in the data subset. Overprediction, defined as productivity beliefs *minus* realized productivity, is collapsed (across all drivers) by week of tenure. The dots correspond to the collapsed means. The smoothed curve is plotted using a lowess regression with a bandwidth of 0.5. In panel (a) beliefs minus actual productivity across drivers is collapsed into weekly means before local polynomial smoothing. In panels (b) and (c), beliefs minus actual productivity across drivers is collapsed into weekly medians before smoothing. In panel (c), we restrict to workers who stay at least 75 weeks and who respond to the beliefs survey that week. Results are similar if instead we look at workers who ever respond. Across these figures (and several others not reported with different samples), we see a slow pattern of gradual reduction in overprediction.

Figure F3: Distribution of Overconfidence Across Drivers

(a) Dist of Individual Overconfidence

(b) Predicted Miles and Actual Miles

Notes: This figure analyzes overconfidence among drivers in the data subset. It presents reduced-form evidence on the distribution of overconfidence across drivers. Panel (a) 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 panel (b), each driver is represented by a dot located at their average productivity and beliefs (averaged over all weeks the driver is observed).

Figure F4: 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 are 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). *Provided we can compare our truckers to the workers in the CPS, there is no evidence that drivers overestimate their outside option.* In addition, in a regression of perceived outside option in dollars on driver beliefs about their inside option in dollars and full controls, we get a coefficient of only 0.07 ($p = 0.16$) on beliefs about the inside option, suggesting that perceived inside and outside options are weakly correlated.

Table F1: How Do Driver Characteristics Predict Survey Response and Overconfidence?

	Survey Response (1)	Overconfidence (2)
Black	-0.043 (0.026)	-32.82 (95.75)
Hispanic	-0.170*** (0.023)	-246.91* (144.53)
Female	-0.061** (0.026)	-11.23 (131.23)
Married	0.026 (0.017)	-30.57 (58.56)
Age	0.004*** (0.001)	1.16 (2.16)
Years of Schooling	0.011** (0.005)	-20.53 (16.56)
Observations	40,396	8,628
R-squared	0.07	0.137

Notes: This table presents OLS regressions, where an observation is a driver-week. Column 1 is a linear probability model predicting whether a driver responds to the belief survey in a given week. Column 2 regresses Overconfidence, defined as miles predictions minus actual miles, on observable characteristics. Standard errors in parentheses clustered by driver. All regressions include weekly fixed effects and work type controls. 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 F2: Robustness Check on Do Productivity Beliefs Predict Productivity? OLS Regressions with Data Corrected for Measurement Error

	(1)	(2)	(3)	(4)	(5)
L. Predicted miles	0.193*** (0.022)	0.148*** (0.018)	0.063*** (0.013)	0.061*** (0.013)	0.079*** (0.021)
L. Avg miles to date			0.732*** (0.035)	0.714*** (0.034)	
Demographic Controls	No	Yes	No	Yes	No
Education Controls	No	Yes	No	Yes	No
Work Type Controls	No	Yes	No	Yes	No
Subject FE	No	No	No	No	Yes
Observations	8,061	8,061	8,049	8,049	8,061
R-squared	0.09	0.17	0.26	0.27	0.38

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 regressions include week of tenure dummies. 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 F3: Do Productivity Beliefs Predict Productivity? OLS Regressions at Firm B

	(1)	(2)	(3)	(4)	(5)	(6)
L. Pred miles	0.332*** (0.058)	0.328*** (0.058)	0.310*** (0.065)	0.163*** (0.053)	0.166*** (0.050)	0.036 (0.058)
L. Avg miles to date				0.605*** (0.080)	0.595*** (0.074)	
\$10 Incentive			-1.178 (3.024)			
\$10 Incentive X L. Avg miles to date			0.024 (0.124)			
\$50 Incentive			-2.358 (6.446)			
\$50 Incentive X L. Avg miles to date			0.169 (0.312)			
Demographic Controls	No	Yes	Yes	No	Yes	No
Subject FE	No	No	No	No	No	Yes
Observations	796	796	796	687	687	796
R-squared	0.074	0.103	0.106	0.253	0.271	0.547

Notes: The dependent variable is miles driven per week. An observation is a driver-week. Standard errors clustered by driver in parentheses. The demographic controls are gender, age, trucking experience, and region of home residence. These drivers are all from Firm B where we collected subjective productivity forecasts similar to as at Firm A, but randomizing financial incentives for accurate guessing to some workers. We present the data here to show, as at Firm A, that productivity beliefs are moderately predictive of actual productivity across workers, but only weakly so within workers. This finding is consistent with our model in Section 5. In addition, we see that there are no statistically significant differences as to whether productivity beliefs are more predictive of actual productivity when they are financially incentivized. * significant at 10%; ** significant at 5%; *** significant at 1%

Table F4: Do Productivity Beliefs Predict Quitting? Robustness Check Comparing Above-Median and Below-Median Beliefs

	(1)	(2)	(3)	(4)	(5)
Above median miles (0 or 1)	-0.765*** (0.274)			-0.717** (0.307)	-0.961*** (0.339)
Avg miles to date		-0.079*** (0.010)	-0.112*** (0.039)	-0.016 (0.037)	-0.068 (0.043)
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. Column 3 differs from column 2 in that it restricts to the sample of driver-weeks for which there is a corresponding belief expectation. All drivers are from the same training school and were hired in late 2005 or 2006. The odds-ratios are 0.47, 0.49, and 0.38 for columns 1, 4, and 5, respectively, indicating reductions in quitting of 53%, 51%, and 62% from having above-median subjective beliefs vs. below median beliefs. * significant at 10%; ** significant at 5%; *** significant at 1%

Table F5: Do Productivity Beliefs Predict Quitting? Robustness Check with Lagged Values

	(1)	(2)	(3)	(4)	(5)	(6)
L. Predicted miles	-0.028* (0.017)	-0.035* (0.019)			-0.030 (0.019)	-0.032 (0.020)
L. Avg miles to date			-0.061*** (0.010)	-0.040 (0.040)	0.008 (0.036)	-0.019 (0.042)
Demographic Controls	No	Yes	Yes	Yes	No	Yes
Education Controls	No	Yes	Yes	Yes	No	Yes
Work Type Controls	No	Yes	Yes	Yes	No	Yes
Observations	8,338	8,338	37,491	8,338	8,338	8,338

Notes: Column 4 differs from column 3 in that it restricts to the sample of driver-weeks for which there is a corresponding belief expectation. The notes are otherwise the same as in Table F4. Although the coefficients in columns 5 and 6 miss standard levels of statistical significance, they represent economically sizeable reductions in quitting, e.g. moving from the 50th (2,300) to the 90th percentile (3,000) in beliefs reduces quitting by 21%. * significant at 10%; ** significant at 5%; *** significant at 1%

Table F6: Do Productivity Beliefs Predict Quitting? Robustness Check with a Person's Average Subjective Belief to Date

	(1)	(2)	(3)	(4)	(5)	(6)
Avg predicted miles to date	-0.034** (0.016)	-0.055** (0.025)			-0.038** (0.019)	-0.046* (0.028)
L. Avg miles to date			-0.062*** (0.010)	-0.059* (0.034)	0.012 (0.034)	-0.029 (0.040)
Demographic Controls	No	Yes	Yes	Yes	No	Yes
Education Controls	No	Yes	Yes	Yes	No	Yes
Work Type Controls	No	Yes	Yes	Yes	No	Yes
Observations	8,632	8,632	39,388	8,632	8,632	8,632

Notes: Column 4 differs from column 3 in that it restricts to the sample of driver-weeks for which there is a corresponding belief expectation. The notes are otherwise the same as in Table F4. * significant at 10%; ** significant at 5%; *** significant at 1%

Table F7: Do Workers Update their Subjective Productivity Beliefs?

	(1)	(2)	(3)	(4)	(5)
L. Avg miles to date	0.684*** (0.083)	0.664*** (0.078)	0.572*** (0.082)		0.365*** (0.062)
Tenure X L. Avg miles to date			0.0025 (0.0016)		
L^2 . Avg miles to date				0.566*** (0.078)	
L. Miles				0.088*** (0.015)	
Demographic Controls	No	Yes	Yes	Yes	No
Education Controls	No	Yes	Yes	Yes	No
Individual FE	No	No	No	No	Yes
Observations	8,713	8,713	8,713	8,580	8,713
R-squared	0.331	0.337	0.337	0.329	0.629

Notes: This table plots OLS regressions of subjective beliefs on lagged average productivity to date. Standard errors clustered by worker in parentheses. Workers increase their subjective beliefs in response to increases in lagged average productivity to date, as predicted in a normal learning model. Also, as predicted in a normal learning model, workers increase the weight on lagged average productivity to date as worker tenure increases, although the interaction in column 3 is only marginally significant ($p = 0.13$). Column 4 shows that, while agents weigh recent productivity shocks, they place most of the weight on accumulated average productivity to date. Column 5 confirms that updating occurs within worker. All specifications include work type controls and week of tenure dummies. * significant at 10%; ** significant at 5%; *** significant at 1%

H Field Experiment, Further Results

Figure G1: Experimental Design

Notes: This figure presents the design for the field experiments, both the incentives for accurate guessing about productivity and the informational “debiasing” where we informed workers at Firm B about our findings on overconfidence with the workers from Firm A. Stage I and Stage II each lasted 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. The number of workers participating in Stage II is smaller than in Stage I due to attrition from the survey (where we were unable to have phone contact with a driver for an interview) and attrition from the company.

Table G1: Field Experiment at Firm B: Randomization Test

Panel A: Stage I	No Incentives	\$10 Incentive	t-test of (1) vs (2)
	(1)	(2)	(3)
Female	0.07	0.08	0.59
Age	41.9	41.2	0.64
Experience	8.40	7.52	0.43
West	0.43	0.43	0.93
South	0.35	0.40	0.35
Midwest	0.21	0.15	0.19
Northeast	0.01	0.01	0.98
Number of observations	138	134	

Panel B: Stage II	Debiasing	No Debiasing	\$50 Incentive	t-test of (1) vs (2)	t-test of (1) vs (3)	t-test of (2) vs (3)
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.01	0.13	0.12	0.01	0.02	0.88
Age	41.7	41.3	43.3	0.84	0.54	0.46
Experience	8.55	8.24	5.89	0.84	0.20	0.26
West	0.41	0.48	0.42	0.42	0.93	0.62
South	0.33	0.36	0.46	0.77	0.25	0.35
Midwest	0.25	0.15	0.08	0.12	0.06	0.34
Northeast	0.00	0.01	0.04	0.31	0.09	0.45
Number of observations	75	73	26			

Notes: This table analyzes the assignment of workers in the Firm B field experiment to the different experimental conditions. p-values are calculated using a two-sided t-test. The number of workers participating in Stage II is smaller than in Stage I due to attrition from the survey (where we were unable to have phone contact with a driver for an interview) and attrition from the company.

Table G2: Field Experiment at Firm B, Debiasing Robustness: Sample Restricted to Stage II of Experiment

Panel A: Impact on Mileage Prediction						
	0-6 weeks after debiasing (1)	Week of debiasing (2)	Week after debiasing (3)	2 weeks after debiasing (4)	3 weeks after debiasing (5)	4-6 weeks after debiasing (6)
Debiasing	-127.0** (59.9)	-135.1 (91.7)	-177.8* (104.3)	-56.5 (104.5)	-92.4 (122.7)	-170.6** (82.5)
Observations	474	110	84	77	67	136
R-squared	0.37	0.37	0.47	0.50	0.46	0.49

Panel B: Impact on Earnings Prediction						
	0-6 weeks after debiasing (1)	Week of debiasing (2)	Week after debiasing (3)	2 weeks after debiasing (4)	3 weeks after debiasing (5)	4-6 weeks after debiasing (6)
Debiasing	-87.3*** (26.4)	-118.8*** (40.5)	-87.7** (33.8)	-116.9** (44.7)	-26.8 (49.4)	-89.7** (40.1)
Observations	467	108	83	76	67	133
R-squared	0.27	0.31	0.43	0.41	0.41	0.27

Panel C: Impact on Prediction - Avg Pre-Debias Productivity						
	0-6 weeks after debiasing (1)	Week of debiasing (2)	Week after debiasing (3)	2 weeks after debiasing (4)	3 weeks after debiasing (5)	4-6 weeks after debiasing (6)
Debiasing	-196.7*** (52.1)	-151.2* (79.7)	-277.1*** (89.3)	-68.9 (106.8)	-105.6 (101.7)	-254.7*** (82.5)
Observations	447	105	79	68	65	130
R-squared	0.32	0.34	0.41	0.44	0.58	0.41

Notes: OLS regressions with standard errors clustered by driver in parentheses. An observation is a worker-week. All regressions include dummies for the number of days not worked in a week, worker tenure in weeks, and demographic controls (gender, age, trucking experience, and region of home residence). To limit the effect of outliers, we trim the lower and upper 5% on each dependent variable. * significant at 10%; ** significant at 5%; *** significant at 1%

Table G3: Field Experiment at Firm B, Incentive Robustness: Sample Restricted to Stage I of Experiment

	Miles Prediction		Earnings Prediction		Miles Overconfidence	
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives for accuracy (up to \$10/wk)	-9.7 (52.8)	-31.0 (49.6)	-20.1 (23.8)	-20.0 (23.9)	41.4 (69.1)	34.9 (67.9)
Demographic Controls	No	Yes	No	Yes	No	Yes
Observations	567	556	559	550	463	463
Mean dep var	2321	2321	754.2	755.1	304.3	304.3
R-squared	0.10	0.14	0.04	0.05	0.05	0.06
Subjects (clusters)	247	241	239	234	223	223

Notes: OLS regressions with standard errors clustered by driver in parentheses. An observation is a worker-week. All regressions include dummies for the number of days not worked in a week, and worker tenure in weeks. The demographic controls are gender, age, trucking experience, and region of home residence. To limit the effect of outliers, we trim the lower and upper 5% on each dependent variable. * significant at 10%; ** significant at 5%; *** significant at 1%

Table G4: Field Experiment: Effect of Debiasing on Intention to Search for a New Job and Job Satisfaction

Dep Var:	Intention to Search for a New Job (1-3)	Job Satisfaction (1-4)	Actual Quitting (0-1)
Method:	Ordered Probit (1)	Ordered Probit (2)	OLS (3)
Debiasing	-0.05 (0.26)	0.01 (0.22)	-0.02 (0.08)
Incentives for accuracy (up to \$10/wk)	-0.19 (0.26)	0.01 (0.22)	-0.04 (0.08)
Observations	99	319	115

Notes: This table reports the reports of the debiasing experiment with Firm B. Standard errors are clustered by driver in parentheses. The question about search intention was asked only once, coming 1-3 weeks after debiasing. The question about job satisfaction was asked in multiple weeks. The sample is restricted to people assigned to receive debiasing or not. Intention to Search for a New Job is the worker's intention to look for a new job during the next 6 months and is measured on a 1-3 Scale (Not at all likely, Somewhat likely, Very likely). Job Satisfaction is overall current job satisfaction and is measured on a 1-4 Scale (Not at all satisfied, Not too satisfied, Somewhat satisfied, Very satisfied). Actual Quitting is whether the worker ever quits during the time frame of the study. * significant at 10%; ** significant at 5%; *** significant at 1%

I Further Structural Results

Figure H1: Other Aspects of Model Fit: Tenure and the Distribution of Productivity

Notes: These figures plots the distribution of productivity at different tenure levels, both in the data and as simulated by the model. The data are from 699 drivers under the 12-month contract. The model is simulated using 40,000 simulated drivers. The figures show the model can well predict the changing distribution of worker productivity as tenure increases.

Figure H2: Model Fit: No Belief Bias

Notes: These figures compare model-simulated data against the actual data to assess model fit. We plot the survival curve, the quit hazard, the mean miles-tenure profile, and the mean beliefs-tenure profile. The model simulated corresponds to Column 1 in Table 7. We simulate the entire data-generating process for 40,000 drivers. The data are from 699 drivers with the 12-month contract, all of whom are from the same training school and hired in late 2005 or 2006. The paths of quits, earnings, and beliefs over tenure are plotted using an Epanechnikov kernel, with bandwidths of 6 weeks, 5 weeks, and 10 weeks, respectively.

Figure H3: Model Fit: Model Estimated With Overconfidence and Standard Learning

Notes: These figures compare model-simulated data against the actual data to assess model fit. We plot the survival curve, the quit hazard, the mean miles-tenure profile, and the mean beliefs-tenure profile. The model is similar to that in Column 2 in Table 7 except that it imposes standard learning, that is, where the perceived variance of the productivity signals equals the actual variance, $\tilde{\sigma}_y = \sigma_y$. We simulate the entire data-generating process for 40,000 drivers. The data are from 699 drivers with the 12-month contract, all of whom are from the same training school and hired in late 2005 or 2006. The paths of quits, earnings, and beliefs over tenure are plotted using an Epanechnikov kernel, with bandwidths of 6 weeks, 5 weeks, and 10 weeks, respectively.

Table H1: Robustness Tests of Alternative Model Specifications

	Baseline	Annual $\delta = .90$	IPW	Meas error corr	Winsorize beliefs at 4k mi	T=200	Higher outside option
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<u>Productivity and Skill Parameters</u>							
η_0	Mean of prior productivity dist	2025 (16.7)	2024 (16.5)	2024 (13)	1974 (15.4)	2025 (16.6)	2025 (16.7)
σ_0	Std dev of prior productivity dist	286 (10.8)	284 (10.7)	286 (8.4)	277 (8.6)	286 (10.8)	287 (10.8)
σ_y	Std dev of productivity shocks	706 (3.6)	706 (3.6)	708 (2.8)	619 (3.1)	707 (3.6)	706 (3.6)
s_0	Value of skills gained in wks 1-5	8.6 (4.4)	3.8 (2.8)	8.7 (3.5)	8.2 (4.3)	9.3 (5.6)	8.7 (4.4)
<u>Taste UH Parameters</u>							
μ_1	Mass point 1 of taste UH	-260 (14)	-325 (14)	-260 (11)	-267 (19)	-252 (21)	-255 (14)
μ_2	Mass point 2 of taste UH	-135 (12)	-166 (11)	-134 (10)	-146 (12)	-120 (15)	25 (13)
μ_3	Mass point 3 of taste UH	135 (41)	119 (40)	140 (33)	154 (48)	164 (45)	295 (36)
p_1	Probability type 1	0.34 (0.07)	0.31 (0.05)	0.34 (0.05)	0.27 (0.08)	0.31 (0.08)	0.35 (0.07)
p_2	Probability type 2	0.43 (0.06)	0.46 (0.05)	0.44 (0.05)	0.50 (0.07)	0.48 (0.08)	0.43 (0.06)
<u>Belief Parameters</u>							
η_b	Belief bias	589 (28)	591 (28)	590 (22)	638 (27)	560 (25)	588 (28)
$\widetilde{\sigma}_y$	Believed std dev of productivity shocks	1888 (136)	1847 (134)	1874 (105)	2004 (139)	1606 (98)	1899 (137)
σ_b	Std dev in belefs	298 (1.4)	298 (1.4)	299 (1.1)	299 (1.4)	259 (1.2)	298 (1.4)
<u>Scale Parameter</u>							
τ	Scale param of idiosyncratic shock	2208 (350)	2776 (462)	2220 (272)	2032 (272)	2781 (513)	2186 (346)
<hr/>							
Log-likelihood		-90865	-90869	-90978	-87626	-89080	-90865
Number of workers		699	699	699	699	699	699

Notes: This table presents a number of robustness checks for our main estimates. Standard errors in parentheses. Column 1 repeats the baseline estimates from column 2 of Table 6. Column 2 sets the discount factor equal to 0.9980, corresponding to an annual discount factor of 0.90; results are similar also with δ corresponding to an annual discount factor of 0.95. Column 3 uses inverse probability weighting to correct for survey non-response. Column 4 corrects for measurement error in productivity. Column 5 eliminates all the subjective belief observations where the stated belief is greater than 4,000 miles in a week. Column 6 increases the period during which learning about productivity may occur from 130 weeks to 200 weeks. Column 7 raises the outside option r by 25% from \$640 per week to \$800 per week.

Figure H4: Out-of-Sample Prediction

Note: This figure analyzes the ability of the structural model to make out-of-sample predictions. The model is estimated off of 699 drivers under the 12-month contract, with a weekly discount factor of 0.9957 (annual discount factor of 0.8). We simulate 40,000 drivers in each simulation. In panels (a) and (b), we compare model-simulated retention against actual driver retention for workers hired in 2006 facing either no contract (panel a) or the 18-month contract (panel b). In panel (c), we simulate retention under the three contracts, providing a simulation-based analogue of Figure 1.

J Survey on Training and Training Contracts

We conducted phone interviews during June 2010–November 2010 with the 20 largest dry-van and 10 largest refrigerated trucking companies in the U.S. The list of companies was obtained from [Transport Topics \(2009\)](#), a leading industry trade journal. We collected panel data on the type of training provided and on the training contracts used by each firm from 2001–2010. We picked the largest 30 companies excluding several firms that were (1) Primarily comprised of drivers that owned their own trucks (owner-operators) or (2) Provided exclusively logistics services instead of trucking services. We were able to interview someone at all 30 companies, though some companies required multiple follow-up phone calls. We collected panel data on the type of training provided and on the training contracts used by each firm over the last 10 years. At each company, we asked to speak to someone who is familiar with the details of driver training, usually the director of human resources, the director of training, or a driver recruiter. When the person we spoke with was not familiar with what the company had done in the past, we tried to conduct a second interview with someone who was. Despite this, however, some firms did not provide information further back than several years. We asked about whether the firm provided CDL training, whether it provided tuition reimbursement, and whether they provided “on-the-job” training to workers with CDLs, but limited driving experience. When firms provided CDL training, we asked whether a training contract was used. Time-permitting, we also asked questions about why firms chose to train, and their motivation for using different contracts.

Our main findings from the survey are as follows:

1. Larger firms are more likely to train. Doubling firm revenue is associated with a 16 percentage point increased probability of training.¹⁷
2. When asked why they choose to provide firm-sponsored training, firms often answer that they do so because it is often difficult to find enough qualified drivers.
3. There is significant variation across firms in the form of training contracts used, as noted below.

In addition to providing CDL training, 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 out of thirty never offered either firm-sponsored training or tuition reimbursement at some point from 2001–2010.

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

¹⁷In a linear probability model of whether a firm offered CDL training on log 2008 revenue, the coefficient on log 2008 revenue is 0.16 ($p = 0.06$).

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.

K Legal Issues About Training Contracts

[Kraus \(1993, 2008\)](#) detail how training contracts of many lengths are used. For example, in trucking, the duration of training contracts is often 6-24 months, whereas for police officers, contracts of 5 years are sometimes used. Although we 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 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 often allowed training contracts with seemingly large amounts owed. For example, 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. There are limits, however. *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 very large contract to be unenforceable. 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; see e.g. Judge Richard Easterbrook's opinion in *Heder v. City of Two Rivers*. Training contracts are enforced across the U.S., which is different than is the case for other mobility-restricting labor contracts like non-compete agreements (e.g. [Lavetti et al., 2013](#)).

Appendix References

- Burks, Stephen, Bo Cowgill, Mitchell Hoffman, and Michael Housman**, "The Facts About Referrals: Toward an Understanding of Employee Referral Status," 2013. Working Paper.
- , **Michael Belzer, Q. Kwan, S. Pratt, and S. Shackelford**, "Trucking 101: An Industry Primer," *Transportation Research Board*, 2011.
- Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez**, "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review*, 2012, 102 (6), 2981–3003.
- Hoffman, Mitchell**, "How is Information (Under-) Valued? Evidence from Framed Field Experiments," 2012. Mimeo.
- 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.
- Hubbard, Thomas N.**, "Information, Decisions, and Productivity: On-Board Computers and Capacity Utilization in Trucking," *American Economic Review*, 2003, 93 (4), 1328–1353.
- 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.
- Lavetti, Kurt, Carol Simon, and William White**, "Buying Loyalty: Theory and Evidence from Physicians," 2013. Mimeo.
- Lazear, Edward**, "Performance Pay and Productivity," *American Economic Review*, 2000, 90 (5), 1346–1361.
- 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," *Review of Economic Studies*, 2009, 76 (4), 1461–1489.
- Peterson, Jonathan**, "Employee Bonding and Turnover Efficiency," 2010. Mimeo, Cornell University.
- Radzevick, Joseph R. and Don A. Moore**, "Competing to Be Certain (But Wrong): Market Dynamics and Excessive Confidence in Judgment," *Management Science*, 2011, 57 (1), 93–106.
- 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.
- Schlag, Karl H. and Joel van der Weeley**, "Eliciting Probabilities, Means, Medians, Variances and Covariances without assuming Risk Neutrality," 2009. Mimeo, Universitat Pompeu Fabra.
- Selten, Reinhard**, "Axiomatic Characterization of the Quadratic Scoring Rule," *Experimental Economics*, 1998, 1 (1), 43–61.
- Stange, Kevin M.**, "An Empirical Investigation of the Option Value of College Enrollment," *American Economic Journal: Applied Economics*, 2012, 4 (1), 49–84.
- Transport Topics**, "Transport Topics: Top 100 For-Hire Carriers," 2009.

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