Training Contracts, Employee Turnover, and the Returns from Firm-sponsored General Training*

Mitchell Hoffman†
Stephen V. Burks‡

November 2016

Abstract

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 plausibly exogenous contractual variation from a leading trucking firm and a variety of empirical approaches (diff-in-diff, event study, structural estimation), we show that training contracts significantly reduce post-training quitting, increase training profits, and reduce worker welfare. The impacts of the contracts seem to be driven by making a given worker less likely to quit, as opposed to selecting better workers.

JEL Classifications: J24, D03, M53, J41
Keywords: Training contract; firm-sponsored general training; organizations

* A previous paper appeared as “Training Contracts, Worker Overconfidence, and the Return from Firm-Sponsored General Training.” That previous paper has been divided in two, with the present paper focusing on the impact of training contracts. The companion paper (Hoffman and Burks [2016]) focuses on worker overconfidence and uses a different dataset (though also based on workers from Firm A). Some portions of text may be similar across the two papers. We are deeply indebted to David Card, Stefano DellaVigna, John Morgan, and Steve Tadelis for their advice and encouragement. For especially detailed comments, we also thank Ben Handel, Ben Hermalin, Ken Judd, Pat Kline, Botond Koszegi, Don Moore, Matthew Rabin, and Kenneth Train, as well as numerous seminar participants. We especially thank the many trucking industry managers and drivers who shared their insights with us. We thank managers at Firms A for sharing their data and for facilitating on-site data collection. We thank Mark Gergen, Anthony Kraus, and Gregory Lemmer for guidance in understanding relevant legal issues. Graham Beattie, Christina Chew, Dan Ershov, Sandrena Frischer, Will Kuffel, Amol Lingmurkar, Kristjan Sigurdson, and Irina Titova provided outstanding research assistance. Hoffman acknowledges financial support from the National Science Foundation IGERT Fellowship, the Kauffman Foundation, and the Social Science and Humanities Research Council of Canada. Burks and the Truckers & Turnover Project acknowledge financial support from Firm A, the MacArthur Foundation, the Sloan Foundation, the Trucking Industry Program at Georgia Tech, and University of Minnesota, Morris.

†University of Toronto Rotman School of Management and NBER; mitchell.hoffman@rotman.utoronto.ca
‡University of Minnesota, Morris and IZA; svburks@morris.umn.edu
1 Introduction

Firm-sponsored training is common in many countries and is a central means by which workers accumulate human capital (e.g., [Acemoglu and Pischke 1998, 1999a; Autor 2001]). However, it has been recognized since Pigou (1912) that general training is subject to a “hold-up” problem: firms may be reluctant 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 for policy, given concerns that US workers receive less training than workers in countries with lower turnover, and for economic theory.¹

To discourage workers from quitting after training, firms often use training contracts: the firm pays for training and workers are fined for premature quitting. Training contracts of this form are used for many jobs (e.g., truckers, policemen, nurses, pilots, and federal employees, to name a few), but have received limited attention from economists.² How do training contracts affect quitting and training profits, and why? We demonstrate that training contracts significantly reduce post-training quitting and increase the profitability of general training. Furthermore, the impact of training contracts appears to occur by making worker less likely to quit instead of by selecting better workers.

We show that training contracts reduce quitting using plausibly exogenous contractual variation from the trucking industry. Although it is one specific industry, trucking is ideal for our study because it is large (see Section 3.1 for details), training contracts are widely used, and productivity (miles driven per week) is easy to measure. At a leading firm, which we call Firm A, training was initially provided with no contractual obligation. In an effort to increase retention, the firm introduced a training contract fining workers for quitting in their first 12 months. Different states approved the contract for use at different speeds, leading the contract to be phased into different training schools.

¹See e.g., Acemoglu and Pischke 1998, 1999a; Autor 2001; Barron et al. 1999; Cappelli 2004 for evidence that firms often pay for general training, both nominally and in terms of incidence, and for general discussion of firm-sponsored general training. US firms appear to spend less on training than firms in other countries (Lynch 1993; Brunello and Medio 2001), but firm training expenditures are still substantial, estimated in 1994 at $50 billion annually for private-sector formal training programs (Baron and Kreps 1999).

²We discuss related economics work in footnote 3. For our examples of workers with training contracts, as well as discussion of relevant legal issues, see the law articles by Kraus 1993, 2008. Other workers with training contracts include firefighters, pilots, mechanics, salesmen, paramedics, electricians, accountants, teachers, flight attendants, bank workers, repairmen, firm-sponsored MBAs, and social workers. Courts in the US and abroad have generally ruled that training contracts are legally permissible, arguing they serve the public good by promoting investment in training.
at different times. Several years later, contracts were changed a second time, also in a staggered fashion.

Exploiting the staggered contractual roll-out in a “diff-in-diff,” we estimate that the two contracts reduced quitting by about 20% and 15%, respectively, relative to no contract. Event study analysis shows that the contracts led to sharp changes in quitting in the quarter after the change, as well as no evidence of pre-trends. While these results may seem intuitive, they violate a basic prediction of the Coase Theorem where negotiation through side payments leads to efficient turnover irrespective of the quit penalty in place (Lazear 1990).

Having estimated the causal impacts of the contracts, we seek to understand the mechanisms. As emphasized by Lazear (2000) and others, contracts may have affect employee behavior in two broad fashions: they may affect behavior (“incentives”) or they may affect sorting (“selection”). In our setting, the evidence is most consistent with incentives being the driver of the impacts we observe, as the workers brought in under the training contracts do not appear superior in non-quitting characteristics or other dimensions of behavior.

Last, we use a structural model of turnover developed in Hoffman and Burks (2016) to analyze how training contracts affect worker welfare and firm profits, as well as to understand to what extent observed behavior across the contracts is consistent with a basic model. After receiving training, workers make a decision every week whether to quit, gradually learning about their underlying productivity by observing their weekly miles (similar to Jovanovic (1979)). If individuals quit, they are fined according to a training contract, which specifies penalties at different tenure levels. We show that the 12-month and 12-month training contracts led to similar increases in profits relative to not having a contract. Further, both contracts moderately decreased worker welfare.

The central contribution of the paper is to show that training contracts significantly reduce quitting, estimating the effects using plausibly exogenous intra-firm contractual variation. There is a large literature on why firms pay for general training—we provide some of the first evidence (as well as the most causally definitive evidence to date) that formal contracts play an important role in making training profitable for firms.3

---

3 Acemoglu and Pischke (1999a) review the literature on firm-sponsored general training. There are many other reasons besides credit constraints that firms may pay for general training including informational asymmetries, search frictions, and labor market institutions that compress the wage distribution. The part of the training literature most related to our paper is that on tuition reimbursement. Employer-provided tuition reimbursement programs are quite common in the US, being offered in as many as 85% of medium and large firms (Cappelli 2004). In a sample of MBA
A secondary contribution of our paper is to analyze worker behavior under training contracts using a structural model, using the model to estimate impacts on worker welfare and profits. The rich contractual variation also serves to provide natural out-of-sample variation for validating the structural model. This is fairly rare in the literature, although it is something that scholars increasingly advocate when possible (Keane and Wolpin, 2007; Todd and Wolpin, 2006).

The paper proceeds as follows. Section 2 provides theoretical discussion on how training contracts may affect worker behavior. Section 3 provides background information on the trucking industry and describes the data. Section 4 provides reduced-form estimates of the contracts. Section 5 provides structural estimates. Section 6 concludes.

2 Theoretical Discussion

In Becker's (1964) model of training, workers finance their own training through reduced wages during the training period. However, if workers have limited credit and training is brief (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 model in Appendix B, this situation can be remedied with 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 in Hoffman and Burks (2016) and allows us to make analytical predictions.

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, occurring when 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.

To see why, consider the case where it is socially optimal for the worker to quit, but disadvantageous for the firm. Without a training contract, the worker will quit; the firm will try to “bribe” students. Manchester (2010) found that 87% received tuition assistance, with 42% of those obligated to come back to the firm for 12 or more months after completing the MBA. For a recent theory paper on bonding and turnover, which includes analysis of training contracts, see Peterson (2010). Dustmann and Schoenberg (2012) analyze a different commitment issue in training: the ability of firms to commit to providing a certain quality of training. We also contribute to the broader empirical contracts literature; Chiappori and Salanie (2003) argue that theory on contracts has outstripped empirics, reflecting that plausibly exogenous variation in contracts (such as in our paper) is rare.

4For exceptions, see, e.g., Card and Hyslop (2005), Todd and Wolpin (2006), and Duflo et al. (2012).
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 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.
Assuming that workers have private information and that there is no renegotiation, allowing for
training contracts increases training profits and reduces turnover (Proposition 1 or P1).

3 Background on our Setting and Data Description

3.1 Background

Truckdriving in the US. Truckdriving is a common occupation, with roughly 1.8 million US
workers operating heavy trucks such as those used by Firm A [BLS 2010]. Firm A is in the long-
distance truckload segment of the for-hire trucking industry, which is the largest employment setting
for this occupation. An important distinction exists 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 are not usually paid by the mile and
typically spend fewer nights away from home.  

The main training for heavy truckdrivers is that needed to obtain a commercial driver’s license
(CDL). Most new truckers take a formal CDL training course, and in some states it is required by
law [BLS 2010]. CDL training can be obtained at various venues, including truck driving schools

[5]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 4.2, however, finds little evidence of selection effects.

[6]We highlight a few more institutional details. Truckload is the segment that hauls full trailer loads. Truckload
has employee turnover rates, often over 100% per year [Burks et al. 2008], as well as low unionization, 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. For an analysis of productivity in trucking,
see Hubbard [2003].
run by trucking companies, private truck driving schools, and some community colleges. At Firm A, the CDL training drivers received lasted about 2-3 weeks, and included classroom lectures, simulator driving, and actual behind-the-wheel truck 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, about half the firms report providing CDL training at some point from 2001-2010. See Appendix D.2 for more on the survey.

**Production.** Truckload drivers haul full loads between a wide variety of locations. While our data do not contain driver hours, drivers are constrained by the federal legal limit of about 60 hrs/week, and managers informed us that drivers often work up to the limit. Firm A loads are assigned via a central dispatching system and are assigned primarily by proximity (as well as hours left up to the federal limit). 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. According to managers at the firm, productivity differences across drivers reflect various factors, including speed, skill at avoiding traffic, route planning (miles are calculated according to a pre-specified distance between two points, not by distance traveled), not getting lost, and coordinating with people to unload the truck. As for the sources of week-to-week variation, managers emphasized weather, traffic, variable loading/unloading time, and disadvantageous load assignments. Thus, driver miles reflect both driver performance and effort, as well as factors that drivers do not control and may be difficult to predict *ex ante.* See [Hoffman and Burks (2016)] for more on measuring productivity.

**Contract Changes.** To examine the effects of training contracts, we analyze two large contract changes at Firm A, a leading trucking firm. At the start of our data period, Firm A provided CDL training to thousands of new drivers per year at no cost at several geographically dispersed training schools. There was no contractual obligation. We omit certain details from our descriptions to preserve Firm A’s anonymity.

Around late 2000, management proposed implementing a training contract. The primary...
motivation was to increase retention, with a secondary motivation being to help recover costs. To implement the contracts, contracts had to be certified by the states where the training schools were located. This certification process took different amounts of time in different states. In three training schools, the contract was certified in 2001 and was put in use in early spring 2002. At another training school, certification did not occur until late spring 2002 and the contract was not used until fall 2002, and in one school the contract was never used due to certification issues. Managers told us that cross-state differences in time for state certification seemed idiosyncratic and were unlikely to be related to the type of impact the contracts might have. The quit penalty 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 contract applied for both quits and fires.  

After several years of the 12-month contract, in an effort to further increase retention, management decided to switch to an 18-month contract. The initial penalty for quitting was increased to roughly $5,000, but gradually decreased with worker tenure—the penalty was reduced by about $65 per week of service. Again, the contract was phased in gradually. Adoption of the 12-month and 18-month contracts were made without additional changes to driver pay. 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. Appendix A.1 gives further background on the contract changes.

**Enforcement.** While it may seem hard to enforce training contracts with truckers, Firm A made strenuous enforcement efforts. Drivers signed a written contract specifying penalties for early exit. No bond was posted. Upon early exit, drivers were contacted by Firm A to pay the amount due. If drivers did not pay, they were often referred to one of multiple collection agencies. For drivers who remained delinquent, credit agencies were notified. Although comprehensive driver-level collection data are not available, the available data and conversations with managers suggest that approximately 30% of quit penalties were collected (details in Appendix A.2).

---

8 According to 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.

9 Of the roughly $65 per week that was deducted from the worker’s initial quit penalty, the worker had around $13 each week deducted from his pay check. After two years with the company, the driver would receive a bonus payment roughly equal to the total amount deducted from his pay check in the first 18 months.

10 Unlike the first contract change, the order of locations for the second contract change appears to have been more actively chosen by the firm. Appendix A.1 discusses why we do not believe it to be a source of bias.
3.2 Data

The data from Firm A are ideal for analyzing the impact of training contracts due to the large sample size and high frequency of observation. We focus on new inexperienced drivers who are trained by Firm A. The data contain weekly miles and earnings, as well as basic demographics, for thousands of new drivers for 2002-2009. 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. We restrict our sample to drivers from 5 training schools, excluding several schools where the training provided differed and/or the precise contract change dates were not available. We refer to this sample as the “full sample.”

Several other papers by one or both of the authors have analyzed Firm A data on a subset of roughly 1,000 new drivers trained at one of the firm’s training schools in late 2005-2006. However, the full 8-year Firm A dataset on all new workers, is much larger than the data subset; our paper is the first to analyze the full dataset and the dataset is critical for identifying the impact of training contracts (given the cross-training school contractual variation).

Table 1 presents sample means. The sample is primarily male and is majority white. The majority of workers have the 12-month contract, but there are still sizable shares with non contract and the 12-month contract.

In the full sample of drivers, we do not observe driver credit scores. However, in the subset of drivers studied by Hoffman and Burks (2016), there is information on credit scores. The average credit score is 586 and the median credit score is 564—these are quite low relative to the US median credit score of 723 (median at time of data collection).

Figure 1 compares quit hazards under the three contractual regimes. Quit rates are high (sometimes over 1% per week depending on tenure). For drivers with the 12-month contract, there is a spike in quitting at the 52-week mark. There are also smaller bumps at the 52-week mark under the no contract and 18-month regimes. Firm 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.

4 Analysis

4.1 Do Training Contracts Affect Quitting?

**Diff-in-Diff.** Table 2 shows using Cox proportional hazard models that training contracts significantly reduced quitting. The regressors of interest are dummies for the 12-month and 18-month contracts, which vary at the school-week of hire level. We also include current quarter-year fixed effects, quarter-year of hire fixed effects, training school dummies, the annual state unemployment rate, a driver’s average productivity to date, and other controls. The training contract dummies are identified by changes in training contracts across workers hired at different dates at a given training school. Throughout 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).

In column 1, the coefficients on the 12-month and 18-month contract variables are -0.212 and -0.150, respectively, meaning the contracts decreased quitting by about 21% and 15%. A percentage point increase in state unemployment is estimated to reduce quitting by about 3.5%. Thus, the impact of the training contracts on quitting is roughly the same as a 4-6 percentage point increase in the unemployment rate. The estimates remain similar when average productivity and demographics are controlled for.\(^{14}\)

To examine further how the effects of the contracts varied with tenure, we estimate a linear probability model of quitting on interactions of the contracts with worker quarter of tenure. Estimates are shown in Figure 2. Under the 12-month contract, relative to no contract, the largest negative impacts on quitting are observed in weeks 27-52, with quitting increasing in weeks 53-65; this is probably workers postponing quitting until their 12-month contract is over. Under the 18-month

\(^{14}\)We have repeated the results in Table 2 with month-year of hire and current month-year fixed effects (instead of the quarter-year of hire and current quarter-year fixed effects). Our findings our roust, with actually slightly larger impacts of the contracts.
contract, relative to no contract, negative impacts on quitting are especially strong in quarters 5-6.\textsuperscript{15}

**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 the 4th quarter of tenure (weeks 40-52). 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 to the 18-month contract, we analyze quitting in the 5th and 6th quarters of tenure (weeks 53-78). Those under the 12-month contract may have waited for the year to end, whereas these weeks are still under contract for the 18-month contract. For the transition from the 12-month to the 18-month contract, the event study is estimated using:

\[
\text{Quit}_{isqt} = \sum_{j=T}^{T} \theta_j D_{sq}^j + X_{it} \lambda + \epsilon_{isqt}
\]

where \(\text{Quit}_{isqt}\) is a dummy for quitting in week \(t\) by driver \(i\) from school \(s\) with quarter of hire \(q\) and \(\epsilon_{isqt}\) 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, \(D_{sq}^j = 1(q - e_s = j)\), where \(e_s\) is the quarter when school \(s\) adopted the 18-month contract. \(T\) and \(T\) define the range of event time under study. \(X_{it}\) is a vector of individual controls. We restrict the estimation sample in the event study to drivers who eventually quit the company. We normalize \(\theta_{-1} = 0\).

As seen in Figure 3 for the transition to the 12-month contract, the probability of quitting in a week between weeks 40-52 drops by roughly 0.7 percentage points. For the transition to the 18-month contract, there is a substantial decrease in quitting during weeks 53-78 (varying between around 0.5 to 1 percentage point). These event study impacts are quite sizable relative to the average quit rates shown in Figure 1. Of course, the event study impacts are not overall treatment effects of the contracts, and instead represent changes in behavior at particular tenure ranges where the contracts might seem especially potent (e.g., before the contracts expire).

\textsuperscript{15}That is, for the 18-month contract, drivers appear reluctant to quit in quarters 5 and 6, even though (unlike the 12-month contract) the penalty is pro-rated and gradually decreasing. We do not fully understand this, but we speculate this could reflect a (perhaps psychic) value to some drivers of “finishing the contract.”
4.2 Incentives or Selection?

A decrease in quitting from training contracts may result through incentive and/or selection effects. If a worker is penalized for quitting, he may become less likely to quit, i.e., an incentive effect. However, adding a training contract 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. We perform three tests suggesting that the effect of the training contracts on quitting operated primarily through incentives.

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, Appendix Table 3 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 4, the contracts appear to modestly increase productivity (by around 20-30 miles per week), but it is only statistically significant in column 4. 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. Panel (b) shows little evidence for this, with 1 out of 10 coefficients are statistically significant at 5% significance.

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. As can be seen in column 3 of Table 2, the contract dummy coefficients are

\footnote{We 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. In these results, we find no evidence that the contracts induced positive selection. We did this by repeating the last column of Panel of Table 4 while restricting to when drivers are in their first 6 months.}
only modestly smaller in magnitude once productivity is controlled for. Thus, this test provides only limited evidence to support the idea that the contract induced significant positive selection.\(^{17}\)

These tests provide support for some selection due to training contracts, but the overall effects seem fairly limited. Given the strong evidence of selection effects of contracts in other personnel settings (e.g., Lazear (2000)), why does positive selection here seem relatively 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.

4.3 Threats to Identification

We discuss potential threats to identification and why they are unlikely to affect our results.

**Endogenous Contract Changes.** Our estimation assumes that 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 training contracts where future quitting was expected to be lower, then we will overestimate the effect. As noted earlier, our conversations with managers suggest that contract roll-out seems unlikely to be driven by such factors. 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 5. Also, the estimated reductions in quitting are actually a bit larger once training school-specific linear quarter-year of hire trends are included (column 5 of Table 2), suggesting the estimated reductions are not the result of pre-existing trends in quitting.

**Worker Sorting Between Schools.** Another potential confound to identifying the impact of

\(^{17}\)This test is inspired by the test for selection in Lazear (2000), who tests for selection by analyzing whether the coefficient on the piece-rate dummy changes once individual fixed effects are added. He finds that the coefficient on the contract dummy falls by half, leading him to conclude that selection explains half the treatment effect of the contract. Our test is significantly more indirect, given that we cannot observe the same individual under multiple contractual regimes. Accounting for selection by adding individual fixed effects is also used in other papers such as Lafontaine and Shaw (2016), who study the phenomenon of serial entrepreneurship.
training contracts on quitting would be worker sorting between schools. (Note that workers sorting between schools is different from the above-mentioned possibility of workers selecting into the firm.) 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 generally attend a training school based on their state of home residence. Specifically, 93% of drivers in the full sample attend the modal training school in their state.\footnote{This is generally the training school closest to their state. The percentage attending the modal training school in their state is 89\% if we look at all training schools instead of the 5 training schools in the full sample. Drivers who don’t attend their state’s modal training school often live in states roughly equidistant from two training schools.}

**Concurrent Firm Policy Changes.** According to managers, contract changes at training schools were not accompanied by other changes in Firm A policy, such as changes in applicant screening or worker benefits.

**Tenure-Varying Contract Enforcement.** Under the 12-month contract, quitting after 51 weeks leads to the same penalty as quitting 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 said that worker tenure did not affect contract enforcement.

### 5 Structural Simulations of the Different Training Contracts

The analysis so far has considered reduced-form impacts of the contracts. In this section, we use a structural model developed in [Hoffman and Burks (2016)](HoffmanFirm2016) to calculate worker welfare and firm profits under the different contracts. A main advantage of the structural approach is that it enables one to analyze worker welfare (and firm profits) while accounting for option value in learning about ability, for unobserved heterogeneity in taste for the job, and for idiosyncratic taste shocks. The welfare consequences of “locking in” a worker via a training contract may seem to depend importantly on these features; for example, locking in a worker who is roughly indifferent between his inside and outside options may have little worker welfare consequences, but may be much more important when there are large differences.
For brevity, we lay out the model in detail in Appendix C. Described verbally, workers solve an “optimal stopping problem” of when (if ever) to quit the firm. Workers have an underlying productivity, which is initially unknown to them, but drawn from a known distribution. Every week, drivers observe their miles, which is a noisy realization of their underlying productivity, and this enables them to learn about their productivity. The learning process is a generalization of a purely rational version of Bayes’ Rule, thereby accommodating a broader range of updating behavior than would be imposed by strict rationality (this updating is identified by using subjective forecasts made by drivers about their productivity in Hoffman and Burks (2016)). Workers are compensated by a piece rate that depends on their tenure. Workers are risk-neutral. If the worker stays with the firm, they receive their earnings at the firm plus a taste shock. In contrast, if a worker exits the firm, they receive their outside option.

Of course, the extent to which a structural model is useful for such an exercise depends critically on whether it provides a reasonable account of the data. To do so, we simulate worker behavior under the 3 different contractual regimes. Appendix Figure A1 shows that our model provides reasonable out-of-sample predictions. Further details on the out-of-sample predictiveness are provided in Appendix A.6.

Profits are computed as production profits, plus penalties from the training contracts, minus firm costs from training. For a worker staying $T$ periods at the firm before quitting, profits are given by:

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

(2)

Here, $\delta$ is the discount factor, $P$ is the price charged by the firm for a mile of shipping, $mc$ denotes the non-wage marginal cost per mile (such as fuel costs and truck wear), $y_t$ is the driver’s productivity (miles in a week), $FC$ denotes fixed costs per week (e.g., back office support for drivers), $q_t$ is a dummy variable for quitting in a given week, $\theta$ is the share of the training contract penalty collected by the firm, and $TC$ represents the training cost per worker. We use the same parameters and use the same simulation procedure as in Hoffman and Burks (2016).

Table 6 shows that profits are higher with the two training contracts (compared to no contract), but that worker welfare is lower. Interestingly, even though worker quitting behavior follows different patterns under the 12-month and 18-month patterns, worker welfare and firm profits are fairly similar.
under the two different contracts.\textsuperscript{19}

For the firm we study, profits were increased by adopting a training contract versus having no contract. However, additional attempts at optimization toward picking a better contract did not appear to produce sizable gains. Some readers may be tempted to ask the question of why might the firm may not have been optimizing in the first place. A simple explanation is that the use of training contracts may be thought of as a management practice, for which technological adoption is not immediate, as explored in Bloom et al. (2016).

6 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 role of commonly-used contracts that fine workers for quitting after training.

Using plausibly exogenous contractual variation created by the staggered introduction of training contracts across training schools within a firm, we show that implementing a training contract reduced quitting by around 15-20\%. The impacts on quitting are evidence across different research designs, including difference-in-difference and event study analysis. These impacts appear to be driven by making employees less likely to quit, as opposed to selecting better employees, though we caveat that our analysis faces the challenge individuals are observed only under one contract, so we cannot apply some common tests for selection vs. incentives exploiting within-person differences.

Although we focus on a single industry, training contracts are used in many jobs, both other blue-collar jobs (e.g., mechanics and electricians) and high-skill jobs (e.g., pilots, accountants, and stockbrokers). Future work should examine whether the impacts of training contracts are similar. For example, one might imagine that workers in high-skill jobs may be more adept at renegotiating contracts compared to workers in lower-skill jobs. As such, training contracts may be thought to have less of an impact in those settings.

Another important question about training contracts is, how might they affect the beliefs

\textsuperscript{19}As in Hoffman and Burks (2016), we focus separately on profits and worker welfare, as opposed to analyzing 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.
of workers and selection on other dimensions (such as cognitive and non-cognitive ability)? Oyer and Schaefer (2005) argue that stock option grants may draw in employees and managers who are more optimistic about the performance of a firm. In a related vein, if individuals are overconfident about their earning capacity at a firm relative to their outside option, they may be more likely to sign training contracts compared to individuals who are less overconfident. In our companion paper Hoffman and Burks (2016), we collected a lot of information on a subset of workers (including productivity expectations), but all of these workers had the 12-month contract. In future work, it would be fascinating for researchers to examine empirically how contracts affect selection of beliefs, perhaps by eliciting different beliefs under different contractual regimes.

References


**Figure 1:** Training Contracts and the Hazard of Quitting

Notes: These figures plot the quitting hazard under the 3 contractual regimes, using drivers in the Firm A full sample. It focuses only on quits (fires are ignored). 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. In each panel, the x-axis is driver tenure in weeks.
Figure 2: 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 an OLS regression of quitting (0 or 1) for a driver in a given week on training contract-quarter of tenure interactions and controls. The controls are the same as in column 4 of Table 2 except we also include week of tenure dummies (in place of the baseline hazard function included in the Cox model in Table 2). Standard errors are clustered at the school-week of hire level. The two figures are based on one regression, with panel (a) plotting interactions of the 12-month contract and different quarters of tenure, and with panel (b) doing the same with respect to the 18-month contract.
Figure 3: Event Studies: The Impact of Training Contracts on Quitting, Comparing Before and After the Contract Changes (with 95% Confidence Intervals)

Notes: The solid line denotes the coefficient estimate, with the dotted lines denoting the 95% confidence interval. Panel (a) analyzes quitting in weeks 40-52 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. The individual controls are the same as in column 3 of Table 2. For panel (a), the plotted coefficient for “-2” is an indicator for event time equal to “-3” or to “-2.” We combine them together to increase power. Beyond the event time coefficients plotted, we also include a dummy for event time greater than or equal to 8. For panel (b), beyond the event time dummies plotted, we also include a dummy for event time -8 or less, as well as a dummy for event time of 5 or greater. 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. Panel (b) includes one training school that transitioned directly from no contract to the 18-month contract, but the figure is very similar if that training school is removed. Standard errors are clustered at the school-week of hire level.
Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.09</td>
</tr>
<tr>
<td>Black</td>
<td>0.19</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.04</td>
</tr>
<tr>
<td>Age</td>
<td>37</td>
</tr>
<tr>
<td>Married</td>
<td>0.38</td>
</tr>
<tr>
<td>No contract</td>
<td>0.10</td>
</tr>
<tr>
<td>12-month contract</td>
<td>0.71</td>
</tr>
<tr>
<td>18-month contract</td>
<td>0.19</td>
</tr>
<tr>
<td>Number of workers</td>
<td>N</td>
</tr>
</tbody>
</table>

Notes: The sample is drawn from trained drivers at Firm A from 2002 to 2009. The exact number of drivers, $N$, is withheld to protect the confidentiality of the firm, $N >> 5,000$. See Appendix A.3 for more details on data and sample construction.
Table 2: Impact of the Training Contracts on Quitting – Cox Model, Diff-in-Diff

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12m contract</td>
<td>-0.212***</td>
<td>-0.211***</td>
<td>-0.198***</td>
<td>-0.197***</td>
<td>-0.235***</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.052)</td>
<td>(0.055)</td>
<td>(0.055)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>18m contract</td>
<td>-0.150**</td>
<td>-0.160**</td>
<td>-0.145*</td>
<td>-0.137*</td>
<td>-0.189**</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.073)</td>
<td>(0.075)</td>
<td>(0.076)</td>
<td>(0.088)</td>
</tr>
<tr>
<td>State unemployment rate</td>
<td>-0.027***</td>
<td>-0.037***</td>
<td>-0.036***</td>
<td>-0.036***</td>
<td>-0.036***</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Average miles to date</td>
<td>-0.062***</td>
<td>-0.063***</td>
<td>-0.063***</td>
<td>-0.063***</td>
<td>-0.063***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Quarter-Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Quarter-Year of hire FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Training School FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Demographic Controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>School-specific time trends</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
</tbody>
</table>

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. It focuses only on quits (fires are ignored). 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.) Beyond the controls listed, all regressions include work type controls, as well as dummies for the roughly 2-3 week “transition periods” in between contract phase-in at a school and the first group of trainees graduating CDL training with the new training contract at the school. Demographic controls are controls for gender, race, marital status, and age. Average miles to date is a driver’s average weekly productivity to date over positive mile weeks and is given in terms of hundreds of miles driven per week; its value is set to 0 if the driver has not had any positive mile weeks to that date. The “school-specific time trends” are training school-specific linear time trends in quarter-year of hire. The coefficients can be interpreted as approximate percentage changes in quitting, e.g., column (1) indicates the contracts reduced quitting by about 21 and 15 percent. The exact *M* is withheld to protect firm confidentiality, *M* >> 100,000. * significant at 10%; ** significant at 5%; *** significant at 1%.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12m contract</td>
<td>-0.049</td>
<td>-0.050</td>
<td>-0.040</td>
<td>-0.048</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(0.098)</td>
<td>(0.098)</td>
<td>(0.097)</td>
</tr>
<tr>
<td>18m contract</td>
<td>0.031</td>
<td>0.033</td>
<td>0.058</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td>(0.121)</td>
<td>(0.120)</td>
<td>(0.120)</td>
</tr>
<tr>
<td>Unemployment controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Productivity controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Demographic controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>0.99M</td>
<td>0.99M</td>
<td>0.99M</td>
<td>0.99M</td>
</tr>
</tbody>
</table>

Notes: This table mirrors columns 1-4 of Table 2 (with the same controls in column 1 here as in column 1 in Table 2 etc.). The difference is that the failure event in the Cox proportional hazard model is getting fired instead of quitting. The estimates show no evidence that the contracts affected firing. Also, when the firing hazards are plotted, they look similar under the three contract regimes. The exact $M$ is withheld to protect firm confidentiality, $M >> 100,000$. * significant at 10%; ** significant at 5%; *** significant at 1%
Table 4: Training Contracts have Limited Selection Effects

<table>
<thead>
<tr>
<th>Panel A: Selection on Productivity</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample:</td>
<td>All Weeks</td>
<td>Trim 5/95%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>12m contract</td>
<td>27.90</td>
<td>29.43</td>
<td>31.44</td>
<td>32.25*</td>
</tr>
<tr>
<td>18m contract</td>
<td>18.60</td>
<td>17.59</td>
<td>31.23</td>
<td>29.50</td>
</tr>
<tr>
<td></td>
<td>(29.78)</td>
<td>(30.12)</td>
<td>(22.60)</td>
<td>(22.20)</td>
</tr>
<tr>
<td>Demographic Controls</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>0.98M</td>
<td>0.98M</td>
<td>0.75M</td>
<td>0.75M</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Selection on Characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep. Var:</td>
<td>Black</td>
<td>Hispanic</td>
<td>Female</td>
<td>Married</td>
</tr>
<tr>
<td>12m contract</td>
<td>-0.004</td>
<td>0.025***</td>
<td>-0.006</td>
<td>-0.003</td>
</tr>
<tr>
<td>(0.015)</td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.018)</td>
<td>(0.401)</td>
</tr>
<tr>
<td>18m contract</td>
<td>0.007</td>
<td>0.011</td>
<td>0.006</td>
<td>0.022</td>
</tr>
<tr>
<td>(0.021)</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.024)</td>
<td>(0.513)</td>
</tr>
<tr>
<td>Observations</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>0.73N</td>
</tr>
</tbody>
</table>

Notes: Standard errors clustered by school-week of hire in parentheses. 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). In columns 1 and 3 of Panel A, the controls are the same as in column 2 of Table 2. Demographic controls are controls for gender, race dummies, marital status, and driver age. Panel B reports OLS regressions of driver characteristics on training contract dummies and controls. In Panel B, an observation is a driver. The exact M (driver weeks) and N (drivers) are withheld to protect firm confidentiality, M >> 100,000, N >> 5,000. The regressions include quarter-year of hire fixed effects, work type controls, school controls, and the annual state unemployment rate at time of hire. * significant at 10%; ** significant at 5%; *** significant at 1%
<table>
<thead>
<tr>
<th>Dep var:</th>
<th>Has 12m contract</th>
<th>Has 18m contract</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>State unemployment rate</td>
<td>0.002</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.81</td>
<td>0.84</td>
</tr>
<tr>
<td>Observations</td>
<td>0.81N</td>
<td>0.90N</td>
</tr>
</tbody>
</table>

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. An observation is a driver. 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). 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. Both regressions include quarter-year of hire fixed effects, training school fixed effects, and work type controls. The exact $N$ is withheld to protect firm confidentiality. * significant at 10%; ** significant at 5%; *** significant at 1%
Table 6: Profits and Welfare Under Different Contracts

<table>
<thead>
<tr>
<th>Contractual regime:</th>
<th>No contract</th>
<th>12 month</th>
<th>18 month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profits per worker</td>
<td>$3,123</td>
<td>$4,907</td>
<td>$4,968</td>
</tr>
<tr>
<td>Welfare per worker</td>
<td>$57,933</td>
<td>$54,840</td>
<td>$54,478</td>
</tr>
</tbody>
</table>

Notes: This table presents profits per worker and welfare per worker under the three different training contracts used by Firm A. Profits and welfare are described in Section 5 of the main text, and are detailed further in Hoffman and Burks (2016). The model simulated here corresponds to column 2 in the baseline structural estimates table of Hoffman and Burks (2016).
Web Appendix

Appendix A provides additional discussion and analysis on various issues. Appendix B analyzes a one-period version of the structural model and derives analytical results. Appendix C provides the structural model, which is simulated in the main text. Appendix D discusses three miscellaneous issues (measuring productivity, the prevalence of firm-sponsored training in trucking, and the law as it relates to training contracts).

A Additional Discussion and Results

A.1 Impact of Training Contracts: Further Background, Robustness, and Event Study Analysis

12-month contract, further background. While the contract was certified for different training schools at different times, it should be noted that there is not a deterministic relationship between date of certification and when it was phased in to the different states’ training centers. Specifically, for the schools in our sample for which the contract was certified, there was a lag of about 4-12 months after certification before the contract came into effect. To examine empirically whether there could be some relation between contract impacts and time between certification and phase-in, we included an interaction term of the 12-month contract with time from certification to phase-in. We found no systematic difference in the impact of the contract according to this difference (i.e., the interaction term was insignificant and close to 0).

In our sample of 5 schools, one school never received the contract (certification was sought, but never received). A manager suggested the state where this school was located had a pro-employee orientation.

18-month contract, further background. As was the case for the 12-month contract, adoption of the 18-month contract was also staggered. However, unlike the 12-month contract, our understanding is that differences in adoption timing for the 18-month contract were more actively chosen by the firm. The contract was first brought into one training school before being brought into other training schools at a later date. However, conversations with senior managers suggest that the difference in timing was unlikely to be related to the impact of the 18-month contract. Managers related that the first training school to receive the 18-month contract happened to be located near where the Director of Training was based on the time, thereby suggesting that bringing the contract there first (as opposed to the other training school locations) may have been mostly an issue of convenience. Unemployment was higher at the end of the data period when the 18-month contract is brought in—we account for this with cohort and time fixed effects, as well as controlling for the annual state unemployment rate.

Robustness. We have repeated the results in Table 2 with month-year of hire and current month-year fixed effects (instead of the quarter-year of hire and current quarter-year fixed effects). Our findings our roust, with actually slightly larger impacts of the contracts.

1This could have reflected a desire by the firm to try to bring the contract in at the same time while facing the constraint that the states were taking different amounts of time to certify the contracts.
A.2 Contract Enforcement

As discussed in the text, the available data and conversations with managers suggest that roughly 30% of quit penalties were collected. In terms of data on collections, our data are limited to company records on collections from late 2003. In late 2003, the firm was collecting around 20% of quit penalties, though this was early in our sample period, and the collection rate was likely increasing over the sample period. The head of the company’s collections department estimated the total collection rate to be 30%. We also had conversations on the subject of collection rates with a Senior Vice President and the Director of Training. One believed the collection rate to be about 33%, whereas the other believed the collection rate to be about 19%.

Managers also explained to us that training schools are considered private colleges, and training contracts may be counted as a form of loan contract.

In terms of how the collection rate matters for the paper’s results, the results in Table 6 are qualitatively robust to other values of $\theta$.

A.3 More Details on Data and Sample Construction

Full Dataset. The data are constructed directly from the various personnel files of the firm (for details, see Burks et al. [2008]). When race, gender, or marital status is missing, its value is set to the excluded category. Thus, for race, the categories are Black, Hispanic, and Other (including White, Other race, and Race missing). For gender, the categories are Female and Non-female (including Male and Gender missing). For marital status, the categories are Married and Non-married. When age is missing as a control variable, missing values are set to mean age. When a driver is missing the annual state unemployment (because they are missing state of residence), we include a dummy variable to indicate it being missing.

As described in the main text, the dataset is restricted to 5 schools. These 5 schools comprise about 96% of brand-new inexperienced drivers in our sample period. Of the additional schools, there are two of them that we know never received training contracts. One of them was a 3rd party school that the firm contracted with, and the other is a school that primarily focused on providing non-CDL training to drivers who already had CDLs. We exclude these two training schools because the training provided differed (either it was not provided by the firm or it was delivered by people who focused primarily on providing training to other types of drivers), but our results are robust to including these schools. The other schools have small numbers of drivers and we lack information on contract details. Of the 5 schools, 4 are very similar to one another, whereas one is slightly different (our results become even stronger if this 5th school is excluded).

We also eliminate a small number of cases where a driver re-joins the firm having quit previously. Drivers are observed from hire until termination or the end of 2009. The firm stopped hiring new inexperienced drivers at the end of 2008 (due to the economic slowdown), but all drivers are still observed until the end of 2009.

Teams. A small share of driver-weeks involve drivers working in two-person teams (e.g., one person drives while the other sleeps). In the data subset studied in Hoffman and Burks [2010], about 13% of driver-weeks involve a driver working with another driver. For team drivers, the firm equally divides total miles driven among the two drivers in the payroll data provided to us. We control for whether a driver is a team driver as part of the work type controls.

---

2This number is conditional on driver school code being non-missing. Including drivers where school code is missing, these 5 schools comprise about 91% of brand-new inexperienced drivers in our sample period.
A.4 Worker Credit Scores

As described in Section 3.2, Firm A drivers have very low average credit scores. Firm A purchased credit scores for drivers in the data subset. The credit score is the FICO-98 and ranges from 300 to 850. 53% of drivers have a credit score below 600, compared to only 15% of the US general population. 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 US 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 Federal Reserve Board of Governors (August 2007).

A.5 Exit Survey

There were 8 possible responses in the survey: Over-the-Road long haul, Over-the-Road regional, Driving locally, Nondriving job, Unemployed, Disabled, Retired, or Other. We provide our numbers in the text ignoring the 7% of responses given as Disabled, Retired, or Other. We ignore these categories because they may be different from other types of exits, but the percentages in the text are similar if these categories are included. Regional drivers deliver loads in a particular region. Like long-haul drivers, regional drivers are also usually paid by the mile, so ability to get miles may transfer from long-haul to regional. Still, drivers who are best at long-haul need not be the same one who are best at regional work.

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 25%. However, whether a driver responded to the exit survey is uncorrelated with average productivity and most demographics, suggesting the results from the exit survey are unlikely to be biased by non-response. The one significant predictor of response is that older drivers were more likely to respond, with an additional 10 years at age of hire associated with a 6 percentage point increase in the probability of responding. We do not think this will bias our findings, as age is not significantly correlated with whether a driver reports moving to a long-haul job (or to either a long-haul or regional job).

Our purpose in the exit survey is to examine the share of drivers who are leaving for some different type of work. One limitation we face in using the exit survey for this purpose is that while most drivers in our data period at Firm A are doing long-haul work, a small share are doing work that is more correctly thought of as regional. Because of this, the share of workers who are moving to the same type of work is probably higher than 12%. However, based on understanding of Firm A, a considerable majority of Firm A drivers in our data are leaving for other types of work, and we control for work type in the various regressions.

A.6 Out-of-sample Fit

In addition to predicting well in sample, the structural model from Hoffman and Burks (2016) also makes reasonable out-of-sample predictions, which is a more demanding test for the model. Although the fit is imperfect, Figure A1 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 $\chi^2$ statistics for model-predicted quit behavior are 120 (no contract) and 341 (18-month contract). Our preferred model with mean bias predicts substantially better out-of-
sample than the model without mean bias, for which the $\chi^2$ statistics are 172 (no contract) and 472 (18-month contract).\(^3\)

Panel c simulates survival under the three contracts, yielding differences in quitting across contracts that stack up well against the quasi-experimental estimates from Table 2 (which control for observable differences). For example, relative to no contract, the simulated 12-month and 18-month contracts reduce the share of people who have quit at 60 weeks by around 18% and 14% (under no contract, the share who have quit by 60 weeks is about 70%). These lie close to the range of estimates for quit reductions estimated in Table 2, where we estimate reductions of 20-24% and 14-19% for the two contracts. While the impact of the 12-month contract is a little low relative to Table 2 at 60 weeks, the simulated 12-month contract reduces quitting by 26% at 20 weeks.

**Figure A1: Out-of-Sample Prediction**

![Graphs](image)

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. The out-of-sample prediction for the 18-month contract is imperfect—many drivers seem to postpone quitting until after 18 months, even though there is no discontinuous drop in quit penalty at 18 months. As discussed in Appendix footnote 15, we speculate this might reflect a (perhaps psychic) value to some drivers of “finishing the contract.”

### A.7 Additional Information on the Counterfactual Simulations

We further discuss our definition of profits.

#### A.7.1 Profits

We make a number of simplifications in our calculation of profits. In particular, beyond not including firing decisions, we ignore a number of components of profits, including vacancy costs, hiring costs (including recruiting costs and any hiring bonuses), employee referral bonuses, trucking accident costs, non-mileage driver pay (including driver bonuses), and driver benefits. Instead, we simply make an assumption on the overall fixed cost per week. It would be difficult and taxing for the model to try to model all of these different components, some of which we have only limited data on. Not separately modeling these different components should not affect the conclusions of our

---

\(^3\)One reason why out-of-sample fit is imperfect in panels a and b, which is not really a strike against the model, is that the comparisons with other contractual conditions have no additional controls. Thus, we believe that the greatest focus in Figure A1 should be on panel c, which provides simulated differences between the contracts, and we can compare these to the quasi-experimental estimates in Table 2 (recall that Table 2 has many controls).
counterfactual analyses unless these interact in some way with the counterfactuals. The general conclusions of the counterfactuals, however, seem robust to different assumptions. We also calculated profits per worker per week for our counterfactuals (instead of profits per worker) and the conclusions are robust.

A.8 Other Papers Using the Firm A Data Subset

As mentioned in Section 3.2, several other papers have used data from the Firm A data subset (also called the “New Hire Panel” in the other papers) to study various topics. Burks et al. (2009) examine whether worker cognitive skills predict experimental measures of worker preferences, worker strategies in experimental games, and worker retention. 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. Anderson et al. (2013) compare measures of social preferences between truckers, students, and non-trucker adults.

In an unrelated paper written after ours, we combined the entire Firm A data with data from 8 other firms to study differences across workers in terms of whether they were hired through a referral from an incumbent employee (Burks et al., 2015). Whether a driver was hired via referral is uncorrelated with both training contract status and average overconfidence, so excluding referral status from our present analysis should have little impact on our findings.

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 allowing firms to use training contracts reduces quitting and increase the profitability of training. The model is a 1-period version of our structural model in Hoffman and Burks (2016), abstracting away from dynamics and worker heterogeneity. 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 \( \eta \), where \( \eta > r \). Workers have some non-pecuniary taste for the job \( \varepsilon \), which they learn after training. We assume that \( \varepsilon \) 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 \( W \) 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 \( (W \geq r) \). The case of \( W = r \) corresponds to the worker choosing whether to go to another occupation whereas \( W = \eta \) may correspond to the worker choosing to leave for another firm within the same occupation (if training is portable across firms, however, 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.
but occupation-specific). We can also think of $W = r$ as specific training and $W = \eta$ 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 $\theta k$). (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:

1. The firm chooses whether to train, and if so, sets the worker’s piece rate, $w$, and the level of the training contract, $k$.
2. The worker decides whether or not to accept the contract (of $w$, $k$, and receiving training), and training occurs.
3. The taste shock $\varepsilon$ is realized and the worker decides whether to quit.
4. Payoffs are realized.

In this baseline model, workers have rational beliefs about their post-training productivity. Specifically, workers believe their post-training productivity will be $\eta$. With rational beliefs, it is the same for the firm to set realized earnings, $W$, as it is for the firm to set the piece rate, $w$. The firm’s problem can be written as:

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

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

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

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

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

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

---

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 discussion in Hoffman and Burks (2016) for more discussion 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 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.
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 $r \leq E_{max}(\bar{W} - \Psi^{-1}(P) - k + \varepsilon, \bar{W} - k)$ or as $r \leq \bar{W} - k + E_{max}(\varepsilon - \Psi^{-1}(P), 0)$. 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

$$\eta - \bar{W} + \Psi^{-1}(P) + P\Psi^{-1}\ell(P) = 0.$$  \hspace{1cm} (3)

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

$$\eta - \bar{W} + \Psi^{-1}(P) + P\Psi^{-1}\ell(P) + k'(P) \times (P + (1 - P)\theta) + k(1 - \theta) = 0.$$ \hspace{1cm} (4)

Given that $k'(P) \times (P + (1 - P)\theta) + k(1 - \theta)$ is positive for all $P$ and that $\Psi^{-1}(P) + P\Psi^{-1}\ell(P)$ is decreasing in $P$ (by the concavity of $P\Psi^{-1}(P)$), it follows that the $P$ that solves Equation (4) (where the firm optimally sets $k$) is greater than the $P$ that solves Equation (3) (where $k = 0$). \hfill \blacksquare

### C Structural Model

Below is the exposition of the structural model in [Hoffman and Burks (2016)](#). The text here is currently the exact same as in [Hoffman and Burks (2016)](#), but please see [Hoffman and Burks (2016)](#) for a justification of the different assumptions made, particularly the assumptions about non-standard beliefs.

The time horizon is infinite and given in weeks 1, 2, ... . Workers have baseline productivity $\eta$, which is distributed $N(\eta_0, \sigma^2_0)$. 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.\(^7\) A worker’s weekly miles, $y_t$, are distributed $N(a(t) + \eta, \sigma^2_y)$,\(^8\) 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 $t$ having observed their past miles $y_1, y_2, ..., y_{t-1}$ 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 $\delta$.\(^9\)

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

$$V_t(x_t) = \max_{d_t, d_{t+1}, ...} E_t \left( \sum_{s=t}^{\infty} \delta^{s-t} u_s (d_s, x_s) | d_t, x_t \right).$$ \hspace{1cm} (5)

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

---

\(^7\)Assumptions 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 it is not common for the firm to make large changes in the pay schedule.

\(^8\)Assuming that signals are normally distributed is standard in structural learning models (see the survey by [Ching et al. (2013)](#)). Visually, the distribution of signals (miles) among all workers has a bell shape centered close to around 2,000 miles, suggesting this assumption is reasonable (and that the distribution is closer to normal than to log-normal or uniform).

\(^9\)Risk neutrality is assumed in many dynamic learning models (e.g., [Crawford and Shum (2005)](#), [Nagypal (2007)](#), [Stange (2012)](#), [Goettler and Clay (2011)](#), though not in all (for examples with risk aversion, see the survey by [Ching et al. (2013)](#)). [Coscelli and Shum (2004)](#) show that risk parameters are not identified in certain classes of learning models.
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:

\[ u_t(1, x_t) = \alpha + w_t y_t + \varepsilon_t^S, \]

where \( \alpha \) is the worker’s non-pecuniary taste for the job, and \( \varepsilon_t^S \) is an i.i.d. idiosyncratic error unobserved to the econometrician (but observed by the worker) with an Extreme Value-Type 1 distribution and scale parameter \( \tau \). Since workers likely differ unobservedly in taste for the job, we assume there is unobserved heterogeneity in non-pecuniary taste for the job, \( \alpha \), with a 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 \( k \) denote the training contract, with \( k_t \) the penalty for quitting at tenure \( t \). The utility from quitting is the fine, plus the discounted value of his outside option, plus an idiosyncratic shock:

\[ u_t(0, x_t) = -k_t + \frac{r_t}{1 - \delta} + \varepsilon_t^Q, \]

where \( \varepsilon_t^Q \) is an i.i.d. unobserved idiosyncratic error with the same distribution as \( \varepsilon_t^S \). Let \( V_t^{S} \equiv E_t( u_t(1, x_t) + \delta V_t(\mathbf{x}_{t+1}) | \mathbf{x}_t \) and \( V_t^{Q} \equiv E_t( u_t(0, x_t) + \delta V_t(\mathbf{x}_{t+1}) | 0, \mathbf{x}_t \) be the choice-specific value functions for staying and quitting, respectively. Plugging in for \( u_t(1, x_t) \) and \( u_t(0, x_t) \), the choice-specific value functions are given by:

\[ V_t^Q = -k_t + \frac{r_t}{1 - \delta} + \varepsilon_t^Q \equiv V_t^Q + \varepsilon_t^Q \]

\[ V_t^S = \alpha + E_t(w_t y_t | \mathbf{x}_t) + \delta E(V_{t+1}(\mathbf{x}_{t+1}) | \mathbf{x}_t) + \varepsilon_t^S \equiv V_t^S + \varepsilon_t^S, \]

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

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 following asymptotic value functions:

\[ V_t^Q = \frac{rT}{1 - \delta} + \varepsilon_t^Q \equiv V_t^Q + \varepsilon_t^Q \]

\[ V_t^S = \alpha + w_T \eta_t + \delta E(V(\mathbf{x}' | \mathbf{x}) + \varepsilon_t^S \equiv V_t^S + \varepsilon_t^S \]

\[ V(\mathbf{x}) = \max_{d \in \{0, 1\}} \left( V_t^S(\mathbf{x}), V_t^Q(\mathbf{x}) \right) \]

**Belief Formation.** In a standard normal learning model, a worker’s beliefs about his period \( t \) productivity equals the weighted sum of his prior and his demeaned average productivity to date:

\[ E(y_t | y_1, ..., y_{t-1}) = \frac{\sigma_y^2}{(t - 1) \sigma_y^2 + \sigma_0^2} \eta_{t0} + \frac{(t - 1) \sigma_0^2}{(t - 1) \sigma_y^2 + \sigma_0^2} \sum_{s=1}^{t-1} y_s - a(s) + a(t) \]  \[ (6) \]

As \( t \) increases, the agent eventually shifts 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, \( \eta_t \), is drawn from a distribution \( N(\eta_0, \sigma_0^2) \), agents believe \( \eta_t \) is drawn from a distribution \( N(\eta_0 + \eta_0, \sigma_0^2) \). Second, we allow for agents to have a perception of signal noise that may be different from the true signal noise: workers perceive the standard deviation of weekly productivity signals to be \( \tilde{\sigma}_y \) instead of \( \sigma_y \). With these two assumptions,

---

10 Even though only a portion of the penalties owed were collected, as described in Section ??, 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 reasonable. 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 experimented with estimating versions of the model assuming drivers act as if the utility loss from quitting is 0.3 times the penalty. Model fit tended to be less good. Indeed, our preferred model still fails to fully match the quitting spike at one year, as seen in Figure ??.
an agent’s subjective expectation of his productivity, denoted by $E^b$ (where $b$ stands for belief), is:

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

If $\eta_b$ 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 $\sigma_y$.

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: $b_{it} \sim N(E^b(y_{it}|y_1, ..., y_{t-1}), \sigma_b^2)$.

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

1. Workers form beliefs $b_t$ given past miles $y_1, y_2, ..., y_{t-1}$.
2. $\varepsilon_s^S$ and $\varepsilon_s^Q$ 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 $a(t) = 2a_1*(\Lambda(a_2t) - .5)$, 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 $a(t)$.

We also account for skill accumulation following CDL training. After CDL training at Firm A, drivers do “on-the-job training” which includes driving with an experienced driver riding along. We use a length of 5 weeks for on-the-job training. We account for the possibility that drivers may gain valuable skills during this time: we assume the outside option 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 from on-the-job training, is estimated. (Besides allowing for skill accumulation during the first 5 weeks, we alternatively estimate the model allowing for continuous skill accumulation: $r_t = r + 2\theta_1*(\Lambda(\theta_2t) - .5)$, where $\Lambda(x) = \frac{e(x)}{1+e(x)}$, and $\theta_1$ and $\theta_2$ are parameters to estimate.)

**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: $x_t = (y_1, ..., y_{t-1}, w, k, \alpha, \eta_b, X, \varepsilon)$. The model can allow for heterogeneity in taste and in the job and/or 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.

**D Miscellaneous**

**D.1 Measuring Productivity**

Firm A drivers are mostly paid by the miles. Drivers also get small additional payments for non-miles related tasks (e.g., loading and unloading, going through customs, scales weighing, working

---

11The 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, managers often referred to a steep “learning curve” for rookie drivers.

12During this time, drivers often are paid by flat salary instead of by mile. We use a flat salary of $375 per week during on-the-job training. We also assume drivers do not begin learning about their productivity until after 5 weeks.
on trailers, and training other drivers). Some drivers are paid based on their activities or on salary instead of being paid by the mile (e.g., drivers who work as full-time instructors at the training schools).

Beyond tenure with the firm, the driver’s rate per mile increases with experience outside the firm. However, the drivers we study are new to the industry, so the distinction is not relevant.

Truckers in firms like Firm A are allowed to work 70 hours over 8 days, per the US federal hours-of-service regulations. This translates to a federal legal limit of roughly 60 hours per calendar week. See http://www.fmcsa.dot.gov/rules-regulations/topics/hos/index.htm, accessed in October 2010.

To our understanding, desirable loads are not systematically assigned 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. Firm A is a leading firm with a large number of available 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. Miles are imperfectly observed each week because miles are only recorded once a driver reaches his destination. Hoffman and Burks (2016) further detail the source of the measurement error, explain how we can correct for it, and shows it has little impact on the estimates and conclusions in that paper.

D.2 Industry Survey on CDL Training

We did an industry survey of large trucking firms. We find that other firms provide training under training contracts (like Firm A), but that training is concentrated among the largest firms.

We conducted phone interviews with the 20 largest dry-van and 10 largest refrigerated trucking companies in the US. At each company, we asked to speak to someone who was familiar with the details of driver training, usually the director of human resources, the director of training, or a driver recruiter. We collected panel data on each firm’s training practices from 2001-2010. For each year, we asked whether the firm provided CDL training and whether a training contract was used. Our findings from the survey are:

1. 16 of 30 firms report providing CDL training at some point from 2001-2010. This could be from operating their own training school or from a partnership with a 3rd party school where the firm paid for the worker’s training.

2. Larger firms are more likely to train. A 1 log-point increase in firm revenue is associated with a 22 percentage point increased probability of training.

---

13 We note that even if there were various forms of systematic assignment of loads to drivers, 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.

14 The list of companies was obtained from Transport Topics (2009), a leading industry trade journal. Using the journal, we picked the largest 30 companies (20 in dry-van, 10 in refrigerated) after excluding several firms that were (1) Primarily comprised of drivers that owned their own trucks (owner-operators), (2) Primarily providers of logistic or staffing services instead of trucking services, or (3) Canadian companies. We were able to interview someone at all 30 companies, though some companies required multiple follow-up phone calls.

15 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 were not able to provide longer-term information. In addition, the survey is based on recollections of the person we spoke to.

16 In addition or as an alternative to providing CDL training, many companies offer tuition reimbursement programs, where drivers can pay for their training on their own, and receive payments from the company over time.

17 Specifically, this is from a linear probability model regressing whether a firm offered CDL training in a given year.
3. 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.

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

D.3 Legal Issues About Training Contracts

We describe how training contracts of different forms are generally legal in the US. We draw primarily on the law articles by Kraus (1993, 2008). 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.\(^{18}\) Training contracts of many different lengths are allowed and observed. 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. In addition, 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 US, which is different than is the case for other mobility-restricting labor contracts like non-compete agreements, where enforcement varies by state (e.g., Lavetti et al., 2013). Kraus (1993, 2008) also report instances of training contracts in the UK and Iceland.

Appendix References


\(^{18}\) 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.

on log 2008 revenue and a control for dry-van vs. refrigerated status. The coefficient on log 2008 revenue is 0.22 (\(p = 0.04\)).


