

This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: The Analysis of Firms and Employees: Quantitative and Qualitative Approaches

Volume Author/Editor: Stefan Bender, Julia Lane, Kathryn Shaw, Fredrik Andersson, and Till von Wachter, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 978-0-226-04287-9; 0-226-04287-1

Volume URL: <http://www.nber.org/books/bend08-1>

Conference Date: September 29-30, 2006

Publication Date: October 2008

Chapter Title: Using Behavioral Economic Field Experiments at a Firm: The Context and Design of the Truckers and Turnover Project

Chapter Author: Stephen V. Burks, Jeffrey Carpenter, Lorenz Götte, Kristen Monaco, Kay Porter, Aldo Rustichini

Chapter URL: <http://www.nber.org/chapters/c9112>

Chapter pages in book: (45 - 106)

Using Behavioral Economic Field Experiments at a Firm

The Context and Design of the Truckers and Turnover Project

Stephen V. Burks, Jeffrey Carpenter, Lorenz Götte, Kristen Monaco, Kay Porter, and Aldo Rustichini

2.1 Introduction

The Truckers and Turnover Project is a statistical case study of a single large trucking firm and its driver employees. The cooperating firm operates in the largest segment of the for-hire trucking industry in the United States, the “full truckload” (TL) segment, in which approximately 800,000 people

Stephen V. Burks is an associate professor of economics and management at the University of Minnesota, Morris. Jeffrey Carpenter is an associate professor of economics at Middlebury College. Lorenz Götte is a senior economist in the Research Center for Behavioral Economics and Decisionmaking at the Federal Reserve Bank of Boston. Kristen Monaco is a professor of economics at California State University, Long Beach. Kay Porter is a business research manager at the cooperating firm. Aldo Rustichini is a professor of economics at the University of Minnesota.

Earlier versions of this paper have been issued as NBER Working Paper no. 12976 (March 2007), IZA Discussion Paper no. 2789 (May 2007), and, also, under a different title—“Adding Behavioral Economics Field Experiments to the Industry Studies Toolkit: Predicting Truck Driver Job Exits in a High Turnover Setting”—as a Sloan Industry Studies Working Paper (2007). This paper was presented at the Sloan Industry Studies Annual Research Conference, held in Boston, MA, April 25 to 27, 2007, and the authors gratefully acknowledge the support of the Alfred P. Sloan Foundation for the conference. It was also presented at the Conference on the Analysis of Firms and Employees (CAFE), held September 29 to 30, 2006, in Nuremberg, Germany, and the authors gratefully acknowledge the financial support provided to the Conference by the Institute for Employment Research (IAB), the Data Access Center (FDZ-BA/IAB), the Deutsche Forschungsgemeinschaft (German Research Foundation), their research network “Flexibility in Heterogeneous Labour Markets,” the Alfred P. Sloan Foundation, and the National Science Foundation. The authors gratefully acknowledge generous financial support for the Truckers and Turnover Project research from the John D. and Catherine T. MacArthur Foundation’s Research Network on the Nature and Origin of Preferences; the Alfred P. Sloan Foundation; the Trucking Industry Program at the Georgia Institute of Technology (one of the industry studies programs initiated by the Sloan Foundation); the University of Minnesota, Morris; the Federal Reserve Bank of Boston; and from the cooperating motor carrier. We also especially thank the managers and employees of the cooperating carrier, whose involvement and active support made the project possible. The de-

are employed, according to the 2002 Economic Census. The TL segment has a high turnover labor market for its main employee group, tractor-trailer drivers, and the project is designed to address a number of academic and business questions that arise in this setting.

One major part of the project matches proprietary personnel and operational data to new data collected by the researchers to create a two-year panel study of a large subset of new hires. The most distinctive innovation of this project component is the data collection process, which combines traditional survey instruments with behavioral economics experiments. The survey data include information on demographics, risk and loss aversion, time preference, planning, nonverbal IQ, and the MPQ personality profile. The data collected by behavioral economics experiments include risk and loss aversion, time preferences (discount rates), backward induction, patience, and the preference for cooperation in a social dilemma setting. Subjects will be followed over two years of their work lives. Among the major design goals are to discover the extent to which the survey and experimental measures are correlated, and whether and how much predictive power, with respect to key on-the-job outcome variables, is added by the behavioral measures.

The panel study of new hires is being carried out against the backdrop of a second research component, the development of a more conventional in-depth statistical case study of the cooperating firm and its employees. This component involves constructing large historical data sets from fragmented legacy IT sources and using them to create multivariate models of turnover and productivity. Two main emphases are on the use of survival analysis to model the flow of new employees into and out of employment, and on the correct estimation of the tenure-productivity curve for new hires, accounting for the selection effects of the high turnover.

The project is designed to last three and a half years, with the first half-year for set up, and then a year for the initial intensive data collection in the panel study of new hires, in parallel with the construction of the data sets for the statistical case study and the initial generation of modeling from these data. Then there will be two years of lower-intensity work while follow-up data is collected from the participants in the panel study of new hires.

The balance of the chapter is structured as follows. Section 2.2 sets the context by describing the U.S. trucking industry and the role of the TL seg-

signers of the field experiments used in the project thank Catherine Eckel and Kate Johnson for sharing protocol and design details from field experimental work in Mexico and for offering helpful advice on our design issues. We also thank Urs Fischbacher, the developer of the z-Tree software used in the economic experiments, for rushing development of a new version with features we needed. The views expressed herein are those of the authors, and not of the Federal Reserve System, nor the Federal Reserve Bank of Boston, nor of any of the employers of any of the authors, nor of the project sponsors.

ment within it. Section 2.3 discusses the nature of the labor market for TL drivers and why it has had a high turnover equilibrium for about twenty-five years. Section 2.4 discusses the nature of the research relationship with the cooperating firm and how it was constructed. Section 2.5 discusses the statistical analysis of historical operational and human resource data from the firm. It has two main subparts: Section 2.5.1 exhibits preliminary findings on turnover, and section 2.5.2 does the same for productivity. Section 2.6 describes the design of the panel study of new hires and has four main subparts. Section 2.6.1 describes the context of the project's use of behavioral economic field experiments. Section 2.6.2 covers the process by which new students are trained as tractor-trailer drivers, and section 2.6.3 discusses the schedule for the data collection effort at the training school. Section 2.6.4 lists and describes the five data collection activities (three experiments and two survey-type instruments) that take place during the first two-hour session of each data collection event, while section 2.6.5 does the same for the six activities (three experiments and three survey-type instruments) during the second two-hour section of each data collection event. Section 2.7 reflects on the implications of the project for the relevant research communities and public policy. Appendix A lists the project team during the first two project years, and appendix B provides a list and time line for the main data elements being collected by the project.

2.2 The U.S. Trucking Industry

2.2.1 Segments within the Industry

To a casual observer, one truck looks much like another, but in fact, the operations that provide trucking services in the United States are meaningfully differentiated from each other on several dimensions. At the broadest level, trucking operations are broken into private carriage versus for-hire carriage, based on a legal relationship: whether the carrier also owns the freight (private carriage) or is hauling it for another party (for-hire carriage).¹ In recent years, for-hire carriers, one of which is the focus of the present study, have typically operated about one-third of the heavy trucks in the overall U.S. fleet, but about three-fifths of the total miles run by such vehicles (Burks, Monaco, and Myers-Kuykindall 2004a).²

For-hire trucking is itself further broken into a number of distinct seg-

1. Private carriers are firms primarily in nontrucking lines of business who provide trucking services internally as support functions to their primary business operations. Examples might be deliveries of food by a retail grocery chain to its stores in trucks it also owns or pickups of parts for assembly at an auto plant by the auto manufacturer's freight vehicles.

2. Heavy freight vehicles are defined here as having a gross vehicle weight (GVW) of more than 26,000 pounds, the level at which weight alone is sufficient to require the driver to hold a commercial driver's license (CDL).

ments, separated along three cross-cutting dimensions. Within each segment, interfirm competition is significant, but across segments it may be muted, or in some cases even absent. The 2002 quinquennial Economic Census, because of its use of the relatively new North American Industrial Classification System (NAICS), which is based on production process characteristics, gives a good overview of the structure of the for-hire trucking industry at this level of segmentation. For-hire truck transportation as a whole, NAICS category 484, generated \$165.56 billion in revenue in 2002, or about 1.56 percent of that year's gross domestic product (GDP).³

The first broad scale distinction within for-hire trucking is between firms that use general purpose equipment (i.e., standard enclosed van trailers) to handle general commodities and those that use specialized equipment to handle special commodities (examples of the latter would be refrigerated vans, flatbeds, tank trailers, and various other types of specialized equipment). According to the Economic Census, in 2002, general freight operations generated \$111.60 billion annual revenue (67.4 percent of the total), and specialized freight had \$54.01 billion annual revenue (32.6 percent of the total). A second cross-cutting broad scale distinction is between firms that make long distance intercity hauls and those that specialize in operations in and around a particular metropolitan area. In 2002, the Economic Census reports \$120.21 billion in annual revenue for long-distance trucking (72.6 percent of the total) and \$45.35 billion for local hauls (27.4 percent).

A third cross-cutting broad scale distinction is based on the size of the typical shipment hauled, and this dimension on which firms differ is of particular relevance to the present study. It is easiest to understand this distinction by considering full-truckload service in contrast to the other two, less-than-truckload (LTL) and parcel service. At one end of the spectrum are firms like the one providing data for the current study. The archetypal TL carrier sends a driver with a tractor-trailer to a shipper's dock to fill up the trailer with a load, typically weighing from 10,000 to 48,000 pounds.⁴ The driver takes the loaded trailer wherever in the United States the shipment is destined and unloads at the consignee's dock. The driver is then dispatched empty, possibly after waiting for a while, to the next location where a full load is available for pick up. Full-truckload carriers may use specialized equipment for special commodities, but if they haul general commodities, they use general purpose equipment to maximize the chance of backhauls.⁵

By contrast, both parcel and LTL firms aggregate large numbers of in-

3. Calculation is by the authors; total GDP is from the U.S. Department of Commerce, Bureau of Economic Analysis: <http://www.bea.gov/>.

4. The variation is because some less-dense freight exhausts a trailer's volume at low weight levels, while higher-density freight hits the weight limit before the volume limit.

5. That is, this is to maximize the chance of picking up a return load from near the point at which a first one is delivered.

dividual shipments collected at local terminals by local drivers into full trailer loads and move them between terminals on fixed route systems. Parcel carriers handle very small shipments (each piece typically being no larger than 150 pounds, with the average nearer to 50 pounds), and LTL carriers aggregate medium-sized shipments (widely varying, but with average size around 1,000 pounds). The Economic Census does not group parcel service firms with the for-hire trucking industry, but with air freight carriers. However, it does capture LTL and TL firms within trucking. In 2002, the TL segment dominated the general freight portion of (nonparcel) for-hire trucking, with 67.9 percent of the total employment and 83.8 percent of the total revenue. If the segments of specialized freight that are primarily TL by shipment size are added to the mix,⁶ then TL's share of the total employment of 1.137 million jumps to 72.8 percent, and its share of the total revenue of \$124.50 billion rises to 77.1 percent.

2.2.2 Differences in the Type of Competition within Segments

The differences across the segments in the operational routines needed affect the form and intensity of competition within each segment. Specifically, in the parcel and LTL segments, the need for a fixed network of freight rehandling terminals creates an entry barrier.⁷ While competition among parcel and LTL carriers is frequently strong, it generally takes place among incumbents. This is evidenced by the numbers of firms in the long-distance parcel and LTL segments. In parcel, there are really only four firms with full national coverage (UPS, FedEx, DHL, and the USPS).⁸ There are more LTL firms, but the number is still small. The 2002 Economic Census identifies eighty-nine long-distance general freight LTL firms with five or more establishments, which is the minimum number of terminals needed to give significant geographic scope; there are only fifty-seven firms with ten or more.

But in TL there are essentially no entry barriers. Because TL carriers do not normally rehandle freight once it is loaded, they do not typically require terminals, nor regular route patterns, for cost-competitive operations. So a one-truck carrier can cover the entire nation, and in doing so is competitive, on a load-by-load basis, with most of the services offered by

6. Essentially, this means adding all specialized freight except household goods moving.

7. A brand new LTL carrier that wants to serve more than a single metropolitan area must create and operate a network that is of minimum size necessary to attract sufficient traffic from shippers with differing destination demands, relative to the total shipment flow densities in the geographic area it wishes to serve. But such networks exhibit strong economies of density (a combination of both scale and scope economies)—at low volumes, the average costs are high, but they fall rapidly as volume increases. The expenses of running such a network until a large enough market share is obtained to make the new network cost competitive with those of incumbent carriers are nonrecoverable (or “sunk”) if the firm exits. And the existence of a sunk cost of entry is the classic definition of an entry barrier.

8. Local parcel service is easier to enter, and there are many firms of small geographic scope.

one of the TL-segment's giants. When more complex service coordination is the key factor in market penetration, small firms can subcontract to third-party logistics providers.⁹ And in fact, there is a continual flow into, and out of, the TL segment, mostly by firms operating at small to medium scales. In TL, the 2002 Economic Census identified 25,831 long-distance general freight firms.¹⁰ The market concentration levels in these two segments also show the differing nature of competition. In LTL, the 2002 Economic Census puts the revenue share of the top four long-distance general freight LTL firms at 36.3 percent, while it calculates the share of the top four long-distance general freight TL firms to be only 14.7 percent.

The implication of these facts is that most of TL service is what business analysts call a "commodity business" and what economists call "perfectly competitive." As a result, the firms "at the margin," whose choices set prices for the whole market, in TL are often not the big players, exploiting economics of scale, but may instead be the small firms in the competitive fringe of the industry segment. Their pricing is, in turn, driven significantly by the wages drivers in such firms are willing to accept. Small firms generally face more modest wage expectations from their employees than do large ones, and they also have the benefit of more personal relationships between owners, managers, and drivers. And owner-operators, who make up a significant subset of the small firms, can always choose to pay themselves less in order to get started in the business. Large firms can choose to pay a modest premium above the level set by such firms because they may have cost efficiencies in other areas, and they may be able to maintain a small price premium due to offering customers a number of different services in an integrated fashion, but if they raise their wages too high, they will make their costs uncompetitive. This industry structure sets the context for the derived demand for truck drivers in TL freight and the consequent nature of the labor market for TL drivers.

2.3 The Labor Market for TL Drivers

2.3.1 Segmented Labor Markets Emerge

The American Trucking Associations' (ATA) quarterly turnover report typically shows the average turnover rate at large TL motor carriers to be in excess of 100 percent per year (ATA Economic and Statistics Group 2005).¹¹ Driver turnover among these carriers is an economically signifi-

9. Because a TL carrier can subcontract actual movements in a spot market to owner-operators, it is possible for a firm to enter TL for-hire carriage initially with zero trucks.

10. Unlike the case of LTL, because TL firms don't have freight terminal networks, single establishment firms can be of national geographic scope, but, in fact, 997 of these had more than one establishment, which is still more firms than in the LTL segment.

11. The ATA is a federation of several separate trucking associations.

cant phenomenon—truckload carriers make up the largest segment of for-hire motor carriage by employment, with approximately 600,000 drivers working at any given time (U.S. Census Bureau 2004).¹² This segment of the universe of for-hire trucking firms emerged into its present form after the economic deregulation of 1980, which transformed the structure of the trucking industry. Before deregulation, the nature of the entry barriers created by government policies resulted in lots of TL output by firms using the LTL-type organization of production, with a fixed network of freight handling terminals (Belzer 1995; Burks 1999). But in the postderegulation period, carriers specialized quite strongly in one or another specific shipment size, from the smallest (parcel), through middle-sized shipments (LTL), to the largest ones (TL) (Corsi and Stowers 1991; Belzer 1995; Burks, Monaco, and Myers-Kuykindall 2004b).

As the TL industry segment emerged, so did a parallel segmentation of the labor market for truck drivers (Belzer 1995; Burks 1999).¹³ Drivers wanting to enter employment at parcel and LTL carriers generally found job queues,¹⁴ while the labor market for TL driving jobs began exhibiting high rates of turnover. In fact, the labor market in the TL segment has essentially been in a high turnover equilibrium since soon after the end of the recessions of 1981–1982.¹⁵

2.3.2 The TL Driver's Job

To understand this situation, we start with a short description of the human capital investment needed to become a driver and then discuss the working conditions encountered by the typical driver. Driving a tractor-trailer requires training for, and passing, the state-administered written and driving tests for a commercial driver's license (CDL). Typically a high school equivalent level of literacy is required, and training begins with at least two weeks mixed between classroom work and in-truck practice. This is usually followed by a few days to as much as a few weeks of initial driving experience, which is often obtained with an experienced driver riding in the cab as a coach, while the trainee is still driving on a "learner's permit," before he or she has taken the final test for the CDL. While the CDL test is administered separately by each state, as of 1991 they do so under

12. The calculation is this: in the 2002 Economic Census, TL firms have 72.8 percent of the total employment of 1.137 million workers in (nonparcel) trucking, and the usual rule of thumb is that about 75 percent of the labor force employed by a TL firm is made up of drivers, the balance being made up of sales, customer service, administrative, and managerial employees.

13. In fact, the argument of the second cited work is that the labor market segmentation was itself a significant driver of the parallel industry segmentation.

14. This was especially true at unionized carriers, but was also true to some degree at nonunion ones.

15. It is an indication of the institutionalization of the high turnover secondary labor market equilibrium in TL trucking that the ATA has published its turnover report continuously since 1996.

Federal standards for what must be included. It comprises both written and driving portions, and the minimum legal age at which it may be taken is twenty-one. Trucking firms generally considered a driver to be satisfactorily experienced after a year of work, so the level of human capital required places the job somewhere between unskilled and skilled, and it is best labeled as “semiskilled.”

Once a driver is licensed, the key problem in retention is generally perceived to be the working conditions faced by a tractor-trailer operator in the archetypal long-haul, randomly dispatched, forty-eight-state service provided by most TL firms. In addition to the stresses of handling a big rig among swarms of cars, many drivers have very long weekly work hours on an irregular schedule. In one published survey of long-haul drivers, 21.9 percent reported working seventy plus hours each week, and two out of three drivers reported working sixty plus hour weeks (Stephenson and Fox 1996). Other surveys report similar findings (Belman and Monaco 2001). A survey of long-haul drivers in the Midwest found the median driver worked sixty-five hours, with 25 percent reporting eighty or more hours. In a twenty-four-hour period, the median hours worked was 11, median hours driving was 8.5, and median hours in nondriving work was 2 (Belman, Monaco, and Brooks 2005). These hours contrast to those in two industries in which there are occupations with similar human capital requirements, manufacturing and construction, which had average work weeks of 40.8 and 38.3 hours in 2004, respectively (Bureau of Labor Statistics 2002).

A related issue is that long-haul drivers are often away from home for multiple weeks at a time, with little predictability about the date of return. In the same survey previously mentioned, only 20.7 percent of TL drivers reported that they were home almost every day, while 28.7 percent of drivers in the same study reported being home less often than once every two weeks (Stephenson and Fox 1996). In the survey of drivers from the Midwest, the median long-haul driver had last been home four days prior to the interview, though one-quarter had been away from home ten days or longer (Belman, Monaco, and Brooks 2005). A less tangible issue is that both drivers and firms like to think of CDL holders as professionals, in command of a big rig and responsible for its safe operation. But trucking is a service business, and a primary job function of the driver is to make shippers and receivers happy. The implications vary by customer shipping or receiving location, but this can place drivers somewhat lower than they might expect on the supply chain status hierarchy.

Of course, not every driver in TL operations faces the same conditions. The foregoing description applies to those “running the system,” or being randomly dispatched across the forty-eight U.S. states. Some TL operations are dedicated to the service of particular large customers, and drivers in these operations have a more restricted set of pickup and delivery loca-

tions; more regular schedules, on average; and generally enjoy more time at home, as well. And some TL operations move freight between cities via trailer-on-flat-car (TOFC) or container-on-flat-car (COFC) intermodal methods. Drivers in these operations usually have regional or local runs to and from intermodal facilities and are often home nightly, or nearly so.

Given these facts, a labor economist would expect to observe a “compensating differential” built into the wages of TL drivers that have the worst conditions. In other words, other things equal, TL firms should offer long-haul randomly dispatched drivers a higher earnings level than stay-at-home jobs requiring similar human capital to compensate for their poorer working conditions. But dissatisfaction over wage compensation levels is frequently cited as a leading reason for TL driver turnover (Cox 2004).

2.3.3 Buying “Effective Labor”

Perhaps a better way to think of the firm’s decision problem, which captures the nature of the driver labor market and the TL driver’s job, is to consider the nature of “effective labor” in this context. For a TL firm, this is the application of labor services to move trucks to and from geographically specific customer locations on the particular time schedule desired by the firm. There are three main factors that go into the cost of effective labor in this setting. One is the cost of recruiting and training new drivers to replace those who leave, to account for the lower productivity of inexperienced drivers, and also to account for any growth in business. A second is the cost of paying compensating differentials to drivers with the worst conditions to slow driver exits. The third is the operational cost of making driver working conditions better. In response to stochastic customer demands, the most efficient allocation of equipment frequently calls for irregular schedules and little time at the driver’s home terminal. When this is the case, making schedules more regular and increasing the driver’s time at home is costly.

The key point is that these three cost factors can, to a significant degree, be traded off against each other, with higher expenditure in one area lowering the expenditure in another. The firm’s goal can then be construed in the standard manner: it is to find the cost-minimizing mix of these factors. Historically, the best thinking among many competing TL firms appears to be that spending more on recruiting and training is a cheaper way to get the needed units of effective labor than paying more to raise compensating wage differentials or improve schedules.¹⁶

A stable equilibrium characterized by high turnover rates defines what labor economists call a “secondary labor market” (Cain 1976; Dickens and

16. There is actually another cost factor in “effective labor” that is nonnegligible, the costs of accidents, which inexperienced drivers have at a higher rate than do experienced ones. We do not address that cost in this paper.

Lang 1993).¹⁷ The persistence of the secondary labor market for drivers in TL trucking since sometime in the early 1980s has occasioned much discussion in the trucking industry trade press over the years, as well as a number of academic studies (examples include Casey 1987; Griffin, Rodriguez, and Lantz 1992; Stephenson and Fox 1996; Griffin and Kalnbach 2002; Beadle 2004). Through the ATA, the industry has commissioned significant analytic efforts to understand the management issues raised by a high turnover business model and the long-term demographic trends affecting the viability of the model (Gallup Organization 1997; ATA Economic and Statistics Group 2005). The major findings suggest that firms are aware of the trade-offs among the components of effective labor and that within this framework firms adjust to changes in the conditions of the demand for, and supply of, effective labor. It appears that as a result, the labor market as a whole also adjusts, perhaps with some lags, to such changes.

A major study done by consultants at Global Insight for the ATA links the supply of truck drivers to the supply of labor for semiskilled jobs in construction because this type of work often represents the next best opportunity for likely truckers. The labor demands in these two industries are driven by significantly different macroeconomic factors. During the 1990s, when the derived demand for drivers was high, there was a modest premium—truckers' earnings were an average of 6 to 7 percent above a position demanding similar levels of human capital in construction. The downturn of the economy in 2000 to 2001 created slack in the trucking labor market, but the arrival of low interest rates kept the derived demand in construction relatively stronger. As a result, for a few years, the average long-haul driver could expect to make less than if employed in the construction industry. By 2004, the gap had narrowed, with long haul drivers' earnings 1.5 percent below that of construction workers (Global Insight, Inc. 2005). These facts suggest that wage levels do adjust over time to changes in the balance of labor supply and labor demand, but the persistence of the high turnover numbers shows that the levels of compensating differential being offered are not sufficient to lower turnover to the levels typical in other blue-collar jobs.¹⁸

It is well documented that the flows into and out of industry (as well as related movements of dissatisfied drivers between firms) represent a substantial cost to firms. A study by the Upper Great Plains Transportation Institute found in 1998 that replacing one dry van TL driver conservatively costs \$8,234, and the industrywide cost total was estimated at nearly \$2.8 billion in 1998 dollars (Rodriguez et al. 2000). The study's authors sug-

17. Correspondingly, the ATA typically reports turnover rates at LTL firms to be in the 10 percent to 20 percent range, which makes them roughly equivalent in turnover to nontrucking jobs requiring similar amounts of human capital.

18. The Global Insight study used government data that does not distinguish TL from LTL among drivers for firms in long-distance trucking, but TL drivers make up the predominant share of the categories they analyze.

gested that this estimate is conservative, but it gives an idea of the magnitude of the turnover costs that TL firms must balance against the alternative costs of raising wages or adjusting operational and dispatching decisions in order to lower turnover.

One might well ask whether firms have fully explored the possibilities for trade-offs among the three factors behind the cost of effective TL labor. Most firms are operating with high turnover costs and relatively lower costs for compensating differentials and operational adjustments that improve driver lifestyles. Is it possible that some large discrete shift along the frontier could move a firm out of a “local cost minimum” in this region to a different local minimum that might be lower in total costs?

In fact, J.B. Hunt, then the second largest firm in the industry, engaged in a highly publicized experiment with switching from a business model with high turnover and modest wage costs to one with higher wage costs but lower turnover in 1996. It took the portion of its workforce facing the worst conditions (long and irregular dispatches) and raised wages by 35 percent, while at the same time closing down its driver training schools (Cullen 1996; Isidore 1997). The net result was a cut in both turnover and accident rates by approximately one-half (Belzer, Rodriguez, and Sedo 2002). However, the long-run net financial benefits were not as clear (Waxler 1997); most of the other large firms in the industry, including the one providing data for the present study, continue to train many of their new drivers from scratch, and nearly all TL firms use the high turnover–modest-pay-premium model.

The long-run dynamics of driver labor supply and demand are made more complex by the growth of the long-haul TL industry. Between 2004 and 2014, Global Insight projects it will grow at a rate of 2.2 percent, which translates into an additional 320,000 heavy-duty long-haul new jobs. This statistic does not include the number of drivers needed to replace those who will retire during this time; the industry will need to find an estimated 219,000 additional drivers to replace the one in five drivers who are fifty-five years old or older and are approaching retirement. Concurrently with an increase in demand for drivers, the growth rate of the overall U.S. labor force will slow from 1.4 percent to .5 percent between 2005 and 2014 (Global Insight, Inc. 2005). Another challenging trend for the industry is that to date, Hispanics, who comprise the fastest growing segment in the workforce, represent a lower percentage of drivers than they do of the overall labor supply. It is possible that the conjunction of these factors means that a secular trend toward higher prices for trucking labor has begun. This, in turn, could shift the nature of the trade-offs that firms face among the components of effective TL labor, and—along with fuel price trends and the limitations on the growth of labor productivity in trucking (Boyer and Burks 2007)—it could also dampen the long-run growth prospects of the industry (Reiskin 2006).

2.4 Working with the Cooperating Firm

The cooperating trucking firm is a large company of national geographic scope, with divisions that operate in several of the segments of TL trucking, including long-haul random dispatch service, dedicated carriage for large customers, and intermodal services. By revenue and employment it is among the top one hundred firms in TL. The firm began as a family-owned enterprise in the regulatory era, although it has grown through multiple acquisitions as well as internal expansions, and the original family has not been centrally involved in top management for some time.

Under family control, the management culture was stable and effective, but was also, by design, relatively inward looking. It was based on long-term employment relationships with managerial and administrative ranks filled with “trucking people,” whose careers tended to be built within this single firm. A significant portion of the management started as front-line driver supervisors or, in some cases, as drivers and then worked their way up. Managers at the firm tended to learn their skills on the job and did not see much need to look elsewhere, except to service vendors who could provide expertise relevant to particular practical business problems, such as targeted marketing surveys.

During the period between deregulation and the end of the twentieth century, the firm made many major and critical strategic moves, some of which were quite daring. But the decisions leading to these moves were primarily based on the vision and judgment calls of the trucking people in top managerial positions. There was little thought of broad strategic planning in the formal sense. Early in the new millennium, a new CEO, who had significant formal training in management-related areas, directed the first exercise in formal strategic planning in the firm’s history, following a process recipe provided by a major consultancy. This exercise began to increase the interest within the firm in planning as a useful activity and also increased interest in establishing the analytic foundations for planning work.

The University of Minnesota, Morris, faculty began to work with the firm starting in the fall of 2002, initially on a single pilot project in the form of faculty-guided analysis by an advanced undergraduate student. The project was successful and laid the foundation for an expanding series of faculty-guided research projects over the next two years on a variety of topics. These projects operated on a gift-exchange basis: the faculty and students contributed their time as teaching and learning functions and the firm paid out-of-pocket expenses and provided access (under appropriate confidentiality restrictions) to proprietary business data. The core of the process involved selecting topics of both business and academic interest and for which advanced undergraduates could provide analyses of business use, as well as generating course-level academic output. By the third year of such projects, about twenty students supervised by six different faculty

members had done small projects on several continuing topics, from the analysis of exit interviews, to some initial turnover and productivity analyses, to work on the recruitment and retention of Hispanic employees.

Within the firm, the linchpin of the process was a senior executive who had joined the firm from the outside and who had significant prior experience working fruitfully with academics. He was promoted to responsibility for a number of the aspects of human resources and driver training and moved into his new role just as the firm as a whole was opening up internally to the importance of strategic analysis. From this initial contact, UMM came to work with several other executives, at similar or higher levels of authority and responsibility, on specific projects.

On the UMM side, the linchpin was an industry studies connection: the initial supervising faculty member (Burks) worked with the Sloan-funded Trucking Industry Program as a doctoral student and as a postdoctoral fellow.¹⁹ This added academic depth and polish to trucking industry institutional knowledge he had originally begun acquiring in his youth, when he worked as a tractor-trailer driver during the era of deregulation, between two bouts in graduate school. Burks's background, along with a passion for all things trucking-related, gave him credibility with executives and allowed him to guide the UMM side of the relationship so that useful business deliverables always accompanied the academic results of interest to faculty and students.

On the basis of the relationship constructed through the student projects, Burks and a second UMM researcher, biostatistician Jon Anderson, developed a small project contractually sponsored by the firm for the summer of 2004. This project began exploring the historical data retained by the firm for strategic purposes, including the analysis of the determinants of driver productivity and turnover. The larger scale design of the Truckers and Turnover Project was developed from the starting point provided by the results of this project. Burks, who devoted a sabbatical year to the project, is the principal organizer, and he has been joined in creating and developing the substantive content of the project by the coauthors of the present chapter, as well as by a number of other colleagues, who are based at several other institutions.²⁰

2.5 Research Component One: Statistical Case Study of Historical Data

Research Component One is a statistical case study of some of the historical personnel and operations data of the cooperating trucking firm. There are three interrelated parts to this component. The first is building

19. Burks was a doctoral student at the University of Massachusetts at Amherst; the Trucking Industry Program (TIP) was then located at the University of Michigan and is now hosted by the Georgia Institute of Technology.

20. A complete list of coinvestigators appears in appendix A.

the data sets needed for analysis, the second is analyzing turnover, and the third is analyzing driver productivity. The goal of the first part is to take the many different data and report outputs produced by the fragmented legacy information technology (IT) resources at the firm and construct from them data sets that permit useful strategic and tactical analyses. Because the firm's IT investments began in the early mainframe era, and those investments were focused primarily on solving succeeding generations of practical business problems, the data storage and reporting functions at the firm do not lend themselves easily to strategic use. Data set assembly, documentation, and validation are consuming, and will continue to consume, a very large part of the project's resources.

The goal of the second part is to use survival analysis to map the differences in turnover by driver group, to use hazard functions to explore the different time paths of exits by driver group, and to use Cox proportional hazard multivariate regression to analyze the interaction between the various factors that can affect exits. The goal of the third part is to use panel data multivariate regression models to map the tenure-productivity curve of new drivers as they gain experience, using a fixed effects variant to make a first-order adjustment for the impact of selection on the tenure-productivity relationship. Once the panel data model is sufficiently robust, the estimated fixed effects will then be further dissected statistically.

A key (proprietary) business deliverable from this part of the project will be the assembly of the results of the turnover and productivity models to create an "expected net value of human capital" model for the investment in recruiting and training various types of drivers, who are utilized in various types of operational settings at the firm. Central academic results are expected to be generated from both the turnover and productivity models. Additionally, the analysis of Research Component Two, the panel study of new hires, will be integrated with the results of the analyses from the statistical case study. We next briefly describe the challenges and sketch a few pilot findings from the turnover and productivity analyses.

2.5.1 Initial Work on Turnover

The proprietary human resource data set used for initial turnover analysis was constructed from three distinct initial data files, which share the feature that each record provides information on one driver during one calendar week. The constituent files covered different calendar periods, so we utilize the calendar window during which all three overlap, September 1, 2001, through March 31, 2005. The first file, Weekly Hires, consists of some of the data elements recorded about a driver during the week he or she is hired. Drivers who are rehired during the calendar window have more than one line in this file. The second file, Weekly Separations, contains information recorded about a driver during the week that he or she separates from the firm. Drivers who are rehired and who, as a result, also

separate more than once during our calendar window have more than one line in this file. The third file, Weekly Employment, consists of one observation in each week for each driver employed during that week. Combining all three data sets gives a complete picture, week by week, of flows in, flows out, and who is currently working for the firm.

However, there are some important limitations in these data and a resulting major problem with analyzing them. The Weekly Hire and Weekly Separations data files contain a number of useful variables, including several key breakout variables, such as the driver's division (e.g., dedicated, intermodal, system) and what kind of prior training or experience the driver had when they joined the firm.²¹ Unfortunately, the Weekly Employment data file is missing these key variables. This means that at the present initial stage of the analysis we don't have this information on the drivers who do not experience either a hire or a separation event during our calendar window. And our information is incomplete for drivers who experience only a hire or only a separation event. In particular, the division to which the driver is assigned is known prospectively at the time of the hire event. But it changes later for many drivers, and we only have the updated information in the separation event record for that subset that does depart.

To partially compensate for these problems, we take the following steps. Breakout variables that are of interest in the present study are carried forward to all observations on a given driver, from that driver's hiring observation. This gives us reasonably accurate information on the previous trucking industry training or experience of each driver (because this is not information that changes with tenure). It also tells us which division of the firm's operations a new driver is expected to be assigned to at the time of hire. Because the data on the type of work assignment is so noisy after this process, and because we would only be able to update it for those who exit, we do not pursue specific findings about the impact of the type of work on retention in the present analysis.²²

A further implication of the data limitations is that we restrict ourselves in this initial work to the subset of drivers for which we observe a hiring event during our calendar window because we do not have either hire or separation observations for long-time incumbent employees and so are missing their key breakout variable values. Given an industry context in

21. Not included, on the other hand, are items such as age, gender, level of formal education, or ethnic category.

22. We experimented with the following procedure. We flowed the values from the separation observation backward, to all prior observations of that particular driver, for the variable recording division to which the driver is assigned—for those drivers who have an observed separation only. (This overwrote the forward-flowed divisional assignment data from the time of hire for those separated drivers for whom we observe the hire event.) This gives us improved information on those who separated, but at the cost that noise is differentially left in the observations on those who do not separate. The results were not credible, so we abandoned this part of the analysis until further information can be added to the data set.

which there are large inflows all the time, however, this subgroup is of significant independent interest, irrespective of what might be found if a more inclusive group could be analyzed. Also, because we are not confident that we can correctly identify all the characteristics of second or later spells of employment, we here only examine the first spell of employment during our calendar window, for those drivers who have more than one observed hiring event.²³ These restrictions still leave us with a lot of data: we analyze a set of more than 500,000 observations covering more than 5,000 distinct individual drivers, observed during the period from September 1, 2001, through March 31, 2005.²⁴

Our procedure will be to first examine the survival curve for the entire set of drivers we consider here, along with the associated hazard function, which exhibits the time path of exit risk that gives rise to the survival curve. Then we will separate out the survival curves for discrete subgroups of interest and test for differences between them, and we will also examine the hazard functions for each subgroup for useful insights. It should be noted that our analysis does not distinguish between the possible different reasons for separation. In particular, of the separation events that we observe, 76.4 percent are voluntary quits, while 23.6 percent are discharges for cause, but our survival curves and hazard functions include both.²⁵

Descriptive Results for All First-Hire-Event Employment Spells

We begin by examining the survival pattern for the first observed employment spell of all drivers having a hire-event during our calendar window. Figure 2.1 displays the central results. The vertical axis indicates the percentage of the population initially entering employment that remains after each amount of time on the job, shown on the horizontal axis in weeks from the start of employment.

Some key qualitative facts emerge from this picture. First, turnover rates do look extremely high. At 10.1 weeks, 25 percent of the population is gone, 50 percent have left by 29.1 weeks (the median survival time), and 75 percent have departed by 75 weeks. Second, there is a leveling off of departures in the second six months on the job, followed by an acceleration at the end of the first year. This is consistent with the fact that most of the trainees observed here who undergo the firm's full training program sign a

23. This does not prevent us from examining rehires, as a significant number of the first spells we observe are of rehired drivers.

24. The precise number of drivers and observations is suppressed for confidentiality reasons.

25. The primary statistical methodology is survival analysis. Standard descriptive and analytical methods are problematic when the key dependent variable (here, the length of job tenure) is a time period, as ongoing spells observed at any given point in time are censored: they continue for an unknown further period. Instead, a conditional probability approach is needed to correctly take into account the statistical information contained in censored observations (Kiefer 1988; Cleves, Gould, and Gutierrez 2004)

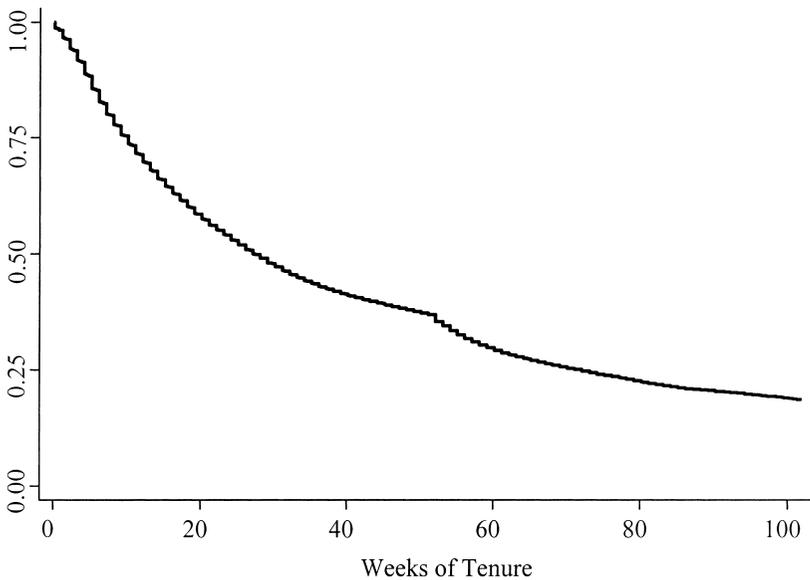


Fig. 2.1 Kaplan-Meier survival curve: Estimates the percentage remaining from this set of drivers at each week of tenure

contract to pay back about half the cost of training (several thousand dollars) if they do not complete a year of service after training. Plus, the job options within trucking are more plentiful for drivers with a year of experience. The surprise, in fact, is that so many new drivers leave before the first year is up. Clearly, these departures cause both the firm and the drivers to incur real costs.

Further insights may be obtained by examining the hazard function for this group of drivers. (See figure 2.2.) The vertical axis indicates the probability of leaving during any particular week shown on the horizontal axis, given that the driver made it to the beginning of the week.²⁶ Here the differences in risk of departure are shown more clearly. Exit risk is highest at about six to eight weeks, which is approximately when new trainees first pull a load by themselves, without the assistance of an instructor-driver in the cab. Once drivers make it past this stage, exit risk declines sharply until the one-year mark is reached, when separation risk spikes to almost the

26. Or, to be slightly more careful, the vertical axis shows a “departure rate” because it is the conditional probability just described, divided by the number of analysis-time units contained in each unit on the horizontal axis. In our case the denominator is 1, so the rate is also a simple conditional probability. Formally, the hazard function is defined to be the ratio of the density of employment duration to the employment duration survival function, or $h(x) = f(x)/S(x)$.

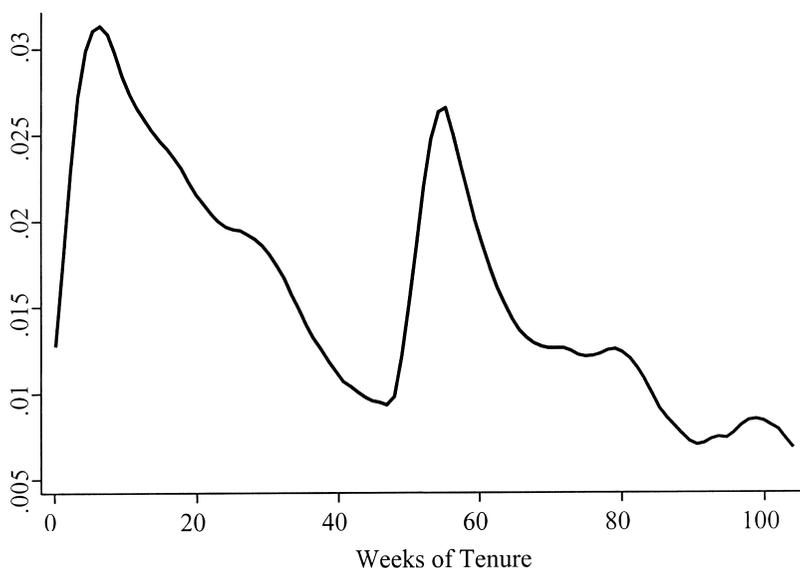


Fig. 2.2 Smoothed hazard function: Estimates the rate of departure from this set of drivers at each week of tenure, conditional on survival to the beginning of the week

same level as at the beginning. Drivers who make it to the end of two years are essentially self-selected to have a high likelihood of turning out to be longer-term employees.

Descriptive Results by Level of Previous Experience or Training

Drivers who are hired by the cooperating firm arrive with different levels of prior training and prior experience. In figures 2.3 and 2.4 and table 2.1, the differing performance of these subgroups with respect to retention gives rise to separate survival curves and hazard functions. The best retention is exhibited by the small group (4 percent of the total) of rehires. This can be observed from the fact that their survival curve is well above the curves of the other subgroups and is quantified in table 2.1. We can see in the table that rehires have the longest time period of any group at which 75 percent still remain (almost four months), and at which 50 percent still remain (over five years). Rehires also have a retention period for 25 percent of the starting population that is so long that it cannot be meaningfully calculated in our data. This is not surprising—rehires are the self-selected subset of drivers who are not only experienced drivers, but who have worked at least once already at the cooperating firm. Having explored other opportunities, they now choose to return to this firm as their best current option.

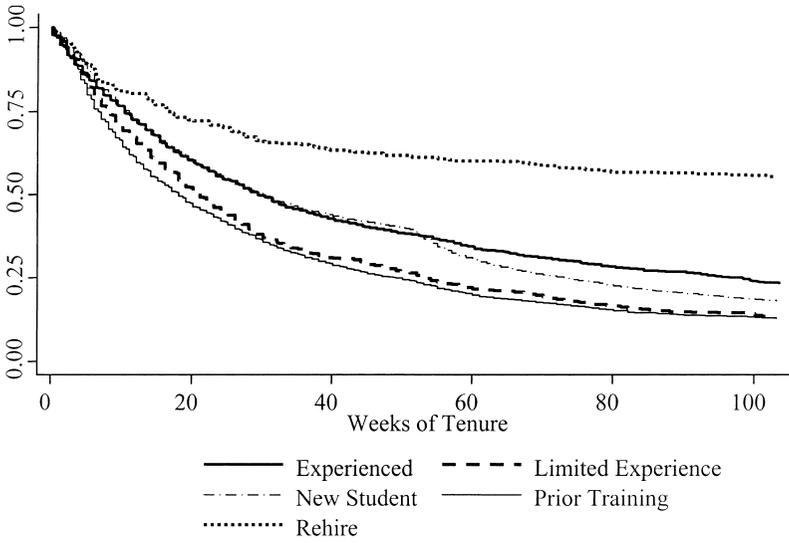


Fig. 2.3 Kaplan-Meier survival curves by type of student: Estimates the percentage remaining from each subset at each week of tenure

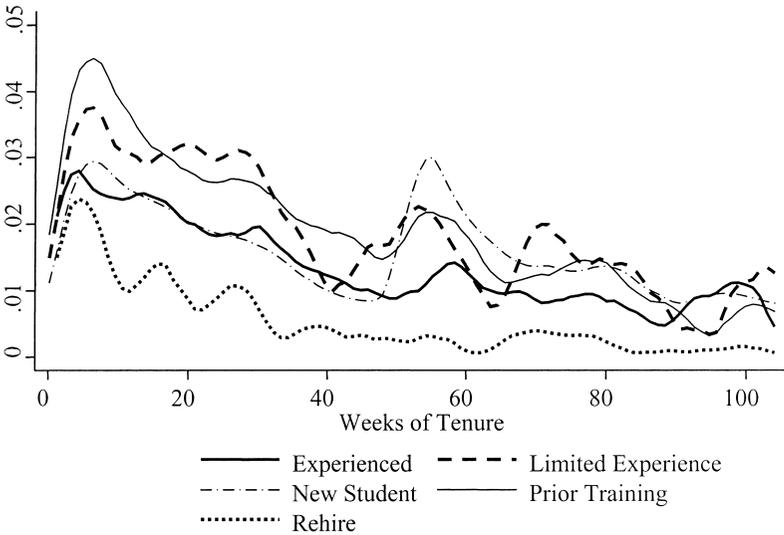


Fig. 2.4 Smoothed hazard functions by type of student: Estimates the rate of departure from each subset of drivers, conditional on survival to the beginning of the week

Table 2.1 Weeks of job tenure by type of student

Drivers for whom a “hire event” is observed (N > 5,000)	Estimated job tenure (weeks)			
	Percent of drivers	75% of drivers remaining	50% of drivers remaining	25% of drivers remaining
All drivers	100	10.1	27.4	72.1
Rehire	4	16.6	284.7	n.a. ^a
Experienced	8	10.4	29.4	98.3
New students	73	11.1	30.1	73.1
Limited experience	3	8.1	21.1	53.1
Prior training	14	6.7	18.1	49.1

^aRehires have a retention period for 25% of the starting population that is so long that it cannot be meaningfully calculated in the data.

The hazard function for these drivers is distinctive as well. It shows a modest spike in exit probability early, with falling exit risk thereafter, and also a very distinct periodicity during the first year, which likely reflects the incentive effects of the firm’s quarterly bonus system. Rehires are eligible for the firm’s quarterly bonus immediately upon starting work and also have experience with the incentive provided by the particular bonus system offered by the firm. The periodicity in the rehire hazard function suggests drivers in this group who may consider leaving during the first year are likely to wait until they have completed a quarter and have qualified for the bonus before separating. Also noteworthy, and sensible, is that there is no “first-year-effect” spike in the rehire hazard rate—this effect in the aggregate hazard function is entirely due to the behavior of other subgroups.

Next consider experienced drivers. These are students who have significant levels of over-the-road tractor-trailer experience with other employers before coming to the cooperating firm. Like rehires, they only have to take a refresher training course that takes a few days, instead of the multiple-week basic training course all other drivers new to the firm are required to pass. Their retention performance is not as good as that of the rehires, but it is still well above that of the lowest groups, with 75th, 50th, and 25th percentile retention periods of 10.4, 29.4, and 98.3 weeks, respectively. Their hazard function shows the usual pattern of an early peak, with later declines, and appears to have a muted version of the periodicity seen in rehires. This would make sense, as experienced drivers are eligible for the bonus system immediately, but don’t have as much experience with its incentives as rehires.

The next item to note is akin to Sherlock Holmes’s famous observation about the mysterious behavior of the dog in the night. The dog didn’t bark when it should have, and correspondingly one would expect new students with no prior background of any kind in trucking to have different (and in

particular, poorer) retention performance than experienced drivers.²⁷ But in these data, both new students who are learning the industry from scratch and experienced drivers who are new to the cooperating firm have closely similar retention behavior for nearly the first entire year of employment. New students actually do slightly better than experienced drivers near the end of the first year. At that point, their hazard function spikes very sharply, and their performance drops below that of experienced drivers. This is likely associated with the facts that their training contracts are completed and they then have enough experience to easily switch trucking firms if they desire. Because new students are by far the largest group (73 percent) of drivers for whom we observe a hire event, their behavior is very important in determining that of the entire aggregate driver population. Thus, the size of their initial aggregate spike in exit risk, as well as that after a year of service, both strongly shape the aggregate survival curve and hazard function.

As it turns out, a Chi-square statistical test of the significance of the difference in overall survival performance between new drivers and those with experience at firms other than the one providing the data shows that experienced drivers do better overall, at the 5 percent significance level ($p = .018$). But, as table 2.1 shows, the effect is all driven by the one-year exits of new drivers, and the magnitude of the effect is much smaller than the difference between either of these groups and rehires.²⁸ For instance, 50 percent of the rehire group is estimated to still be at work for the cooperating firm 5.48 years after the hire event we observe, while for drivers with experience at other firms, it is only 6.8 months, and for new students it is essentially the same, at 6.9 months.²⁹ At longer durations of employment, we see a modest difference: 25 percent of the drivers with experience at other firms still remain at 22.6 months, while it is only 16.9 months for the same proportion of new drivers.

Last, consider the retention performance of the two final groups: drivers with some prior experience and those with some prior training. Both these groups are identified by the driver recruiting staff at the cooperating firm as having some background in trucking, but not enough to qualify the student to take only the short training course for fully experienced drivers. To extend the previous allusion, here is a dog barking loudly—these two groups do quite badly, by comparison to students wholly new to trucking. The job tenure lengths for the retention of the 75th, 50th, and 25th percentiles of

27. The mysterious behavior (in “The Silver Blaze”) was that the dog did not bark when someone removed a valuable race horse from the barn, which was a clue to the thief’s identity.

28. The pairwise differences between rehires and new drivers, and between rehires and experienced drivers, are both significant—the Chi-square p -values for Type 1 error are zero to four decimal places.

29. The base time unit for the statistical analysis is weeks, so months are everywhere calculated as weeks divided by 4.33.

students with limited driving experience is 1.87, 4.94, and 12.25 months, respectively. This tells us that only 25 percent make it to the completion of their one-year-service-after-training employment contract; the other 75 percent are incurring a multithousand dollar debt in order to leave early.³⁰ Students with only some prior training, but no prior experience, do even worse, with retention periods for the 75th, 50th, and 25th percentiles of only 1.58, 4.18, and 11.33 months, respectively. So less than one-quarter of these students complete their training contracts. (The difference between these two groups is significant by the Chi-square test, at the 5 percent level [p -value of .045].)³¹

Why should these students be at the bottom of the performance ranking when normally prior training or experience would be expected to improve retention? A reasonable hypothesis is that it has to do with the distinctive characteristics of a high-turnover, secondary labor market. In this type of market, there is always demand for drivers at some job or other. So someone with prior experience of any kind, as well as the graduates of any of the many commercial driver training schools, can get some job, as long as they have a CDL. It may not be a very desirable job, but it is possible to accumulate experience if one is willing to put up with some of the poorer working conditions available in an industry segment known for having poor conditions on average. In this context, coming to the cooperating firm and being willing to assume the debt contract that accompanies the full training program is a bad signal. There may be many specific reasons outside a prospective driver's control that lead to such a decision. For example, the student could have experienced some kind of family event that stopped his or her prior training before the CDL exam or caused him or her to quit a prior job quickly. But, on average, students with some prior training or some prior experience are likely either to be job switchers who just couldn't do better for the time being, but who will be looking to leave as soon as possible, or to be job candidates who were unsuccessful at someone else's training course, or were otherwise judged inadequate by other firms. Either of these reasons means the student is more likely to depart.

2.5.2 Pilot Work on Productivity

The pilot work on productivity utilized a different set of data files from the cooperating firm than did the turnover work described in the preceding section. We began with two data files, one containing basic information (especially hire date and separation date, if any) on all the drivers who had separated during the period of one year (for example, in some of the pilot

30. Except for those who are hired by a rival firm that is willing to pay off their indebtedness—something which is known to occur in this labor market.

31. The pairwise differences between either of these groups and any of those with better retention performance is highly significant—the Chi-square p -values for Type 1 error are zero to four decimal places.

work we used 2003), and the second, extracted at the end of that year, containing similar information on all currently employed drivers. Then two separate additional files containing demographic information, and racial and ethnic identity from voluntary Equal Employment Opportunity Commission (EEOC) employee disclosure forms, were added.

Merging these using the internal employee number (driver number) as an identifier immediately caused problems. It turned out that driver numbers are not unique, but are recycled on a regular basis, so we had to delete some duplicate cases that really represented different drivers.³² “Hire date,” a key variable for survival analysis also turned out to be problematic. As one might expect in a high turnover setting, a small but significant number of drivers become reemployed, some having as many as four or five successive employment spells. The problem was that drivers gone for less time than some threshold (six months at one point, but varied over time) kept their original hire date, while those gone longer were assigned a new one. The latter fact made it impossible to distinguish rehires from new drivers with recycled driver numbers.

To do a productivity analysis, the key addition to the records already described was information from the firm’s payroll records, which provide a week-by-week compilation of the items added to (or deducted from) each employee’s pay, with each such transaction constituting a line of data. The taxes and fringe benefit co-pays were in a separate data source to which we did not have access, but even so the initial files had as many as forty-four transactions per driver per pay period, with more than one million lines of data per file. We proceeded to document the different variables that contained coded information about the driver’s work assignment and pay structure, consulting subject-matter experts at the firm regularly. Each variable could take on multiple values, the meanings of which to some degree changed over time as operational needs changed. In addition, we began to document all the meanings of the values of the key variable specifying what type of transaction each line of the payroll file contained. There were several hundred distinct values of this variable, including values denoting several different types of mileage pay, dozens of types of lump sum pay for specific tasks, dozens of types of pay advances and pay deductions, and so on.

After documentation, we next “rolled up” the payroll file. We sorted the file by driver and pay-week and then accumulated all the transaction-level information we were interested in having on a weekly basis into new variables so that the last transaction in each driver-pay-week record contained cumulative information for the week. The kinds of information in the re-

32. For the pilot work, we did not want the responsibility of making use of social security numbers, although a secure method for making use of the relevant identification information has been developed for later work.

sulting records included such key items as the total (paid) miles, and the amount paid for them, and the total number of dispatches. Also included was information on various kinds of ancillary activities when they generated a pay transaction, such as paid customer stops, pay supplements for very short runs, paid maintenance delays, and so on. The payroll data thus provides a very rich set of information about what each driver does during each week.

However, the payroll file records what drivers are actually paid for, which is in general a subset of what they actually do. So, for instance, the first pickup stop and first delivery stop on each loaded dispatch are not separately compensated. Extra pickup or delivery stops are paid when they occur on long-distance random dispatch loads, but only some of the time when they are on a scheduled run dedicated to a particular customer that is engineered to have multiple stops. Most drivers are primarily compensated by the mile, and these drivers are paid miles for all their dispatches, which normally includes loaded miles, plus miles pulling an empty trailer, repositioning for a new load, and also any bobtail miles (i.e., without a trailer). However, drivers generally run more miles than those for which they are paid. Paid miles are based on a least-distance routing algorithm, which is historically standard in the industry but which undercounts the actual miles by several percent (recent guesstimates by managers at our firm for the average undercount range from 4 percent to 6 percent).³³ Despite these limitations, the payroll data provide a very useful starting point for the productivity analysis.³⁴

Descriptive Productivity Results for Inexperienced Long-Haul Random Dispatch Drivers

We began our pilot work with a subset of drivers for the years 2002 and 2003. The subset is those drivers who were inexperienced at hire (i.e., those who had to take the full training course offered by the firm), who were assigned to drive solo (as opposed to in a team) on long-haul random dispatch runs, and who were in their 5th week to 156th week of tenure with the firm.³⁵ This gave us more than 100,000 pay-week observations on more than

33. This is, in part, because the standard algorithms are to and from standard reference points, and given the circuitry of the road network, this undercounts actual miles on average. It is also because drivers are responsible for selecting a practical route for a large loaded tractor-trailer, which is often more circuitous than the least-distance version. In addition, drivers may choose to deviate for other reasons (for example, to run on a turnpike where the salt trucks will be out at night in the Pennsylvania mountains in winter, as opposed to a non-toll highway on which such services are more uncertain), as long as they don't exceed certain percentage standards for excess miles and meet delivery schedules.

34. For later work, it is expected supplemental information will be added from a separate operational events data set also maintained by the firm. It is not the place to start because it has its own limitations and also because it is about an order of magnitude larger than the payroll data set.

35. Drivers begin receiving mileage pay when they first pull a load on their own, without a trainer in the truck with them, and the earliest this occurs is about the fifth week.

2,000 drivers. Examining the key dependent variable, miles per week, we observed very high variance (see figure 2.5). In particular, there were negative values and also very high positive values. The former turned out to be due to mistaken pay being charged back against a driver's earnings and the latter to a small number of drivers from the firm's early days who were permitted to accumulate vacation earnings over several years and were being paid upon retirement. We decided to trim the extremes and had to choose whether to leave in zero-miles weeks or use only positive-miles ones and what upper bound to use.

The actual maximum number of miles that a solo driver could legally run during this period, given state speed limits and Federal Hours of Service Regulations for operators of commercial vehicles, was about 4,000 per week. But during at least part of this period, until the practice was ended, some drivers at the firm were paid for their runs only after they submitted completed paperwork for each dispatch. This meant that if they held their paperwork they could have one (or even two) weeks in a row with zero paid miles and then a week with very high miles. We decided to trim only the negative values, leaving zero-miles weeks in, and also trimmed values over 6,500 after looking at the distribution of the upper tail.

Further examination showed that almost 20 percent of our observations were of zero-miles pay weeks. So we first trimmed out all the pay-week ob-

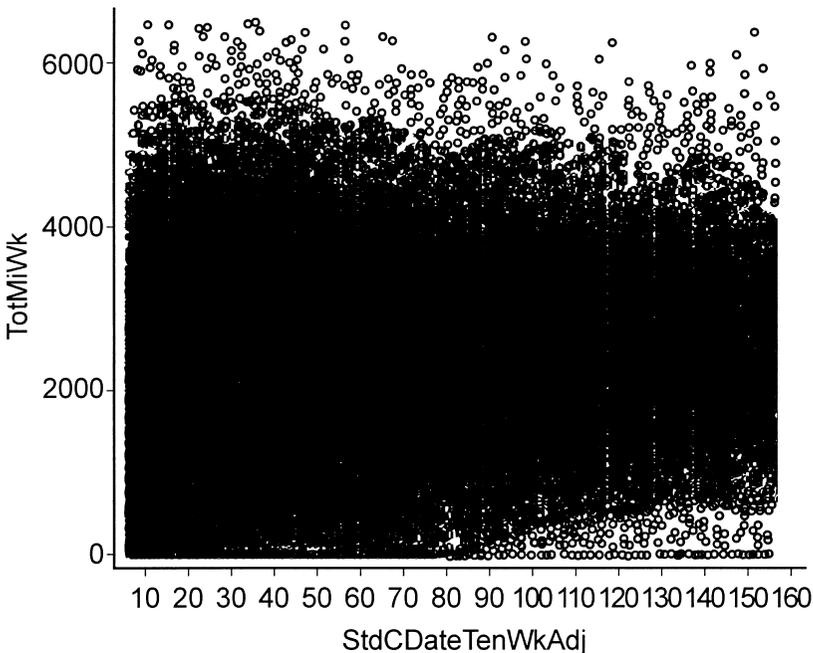


Fig. 2.5 Miles per week by week of driver tenure

servations that were associated with any payroll transaction that could exogenously cause the driver to either miss work or be paid on a nonmileage basis. This included pay weeks with disability pay, vacation or holiday pay, salary (sometimes paid to driver-trainers), lump-sum training pay, and the like. Then we discovered that the payroll system was generating dummy paychecks for drivers who had separated from the firm, for several weeks after separation when the driver left owing money, for example, for things like cash pay advances or purchases at a store at a company terminal. When all of these cases were trimmed out, we reduced the number of weeks with zero miles substantially, but 6 percent of our observations remained with zero miles per week.

Figure 2.6 exhibits a simple descriptive version of the tenure-productivity curve for this subset of drivers. Even after all the trimming, the remaining weeks with zero miles affect the mean values quite significantly. Without zero-miles weeks, the initial increase to full productivity is achieved at about nine months, whereas with zero-miles weeks it is nearer to a year. There is a sharp drop in the curve at one year of tenure with zero miles included. This is undoubtedly related in some way to the fact that drivers with one year of experience can more easily switch firms and also

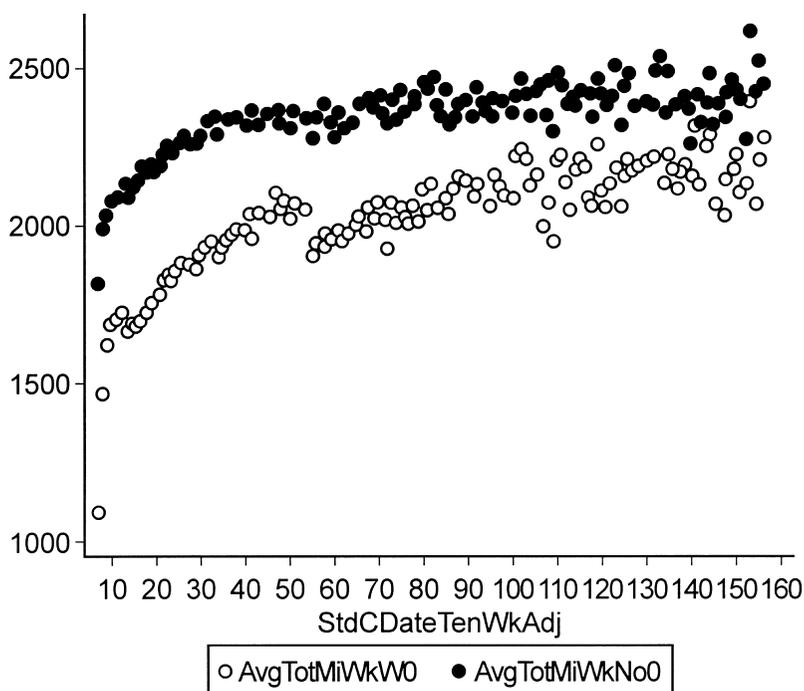


Fig. 2.6 Average miles per week, by week of tenure, with and without zero-miles weeks

to the fact that most of the drivers in this subset, all of whom have taken the firm's training, assume a debt of several thousand dollars for its cost, which is forgiven at the end of a year of service after training is completed.³⁶ The balance of the pilot analysis keeps all the zero-miles weeks in the data set, but a goal of the full statistical case study is to dig deeper into this phenomenon and develop better evidence on whether they all should be included when analyzing productivity.

The Impact of Selection on the Tenure-Productivity Curve

A major goal of the statistical investigation of productivity is to analyze the true causal effect of increasing tenure on expected miles per week (the treatment effect), while accounting for any impact the high turnover rate might have (the selection effect). A priori, a reasonable hypothesis would be that drivers with lower productive capacity would be more likely to leave at any given level of tenure. In order to test this hypothesis, we start by running a fixed effects panel data ordinary least squares (OLS) regression model, with total miles per week as the dependent variable. We use all the independent variables from the payroll data set that plausibly measure exogenous factors that affect productivity. These include driver tenure (a linear term and as many higher-order terms as prove significant) and the number of dispatches (linear term plus those higher-order terms that prove significant). We also include variables such as the number of short-haul pay supplements, the number of paid maintenance delays, as well as dummy variables for each week of the calendar year (to capture any time-period effects or time trends), and a dummy variable for each terminal at which drivers are based (to capture any geographic effects of the home base). We use robust standard errors. The pilot version of this model has an adjusted R^2 of .66.

The fixed effects model constrains the coefficients on all the independent variables to be the same across all drivers, but permits each driver to have his or her own regression plane intercept, or constant. This constant, or fixed effect, which is estimated by the regression, may be thought of as a measure, specific to the model and the data, of the degree of "job match" between the driver and his employment at the firm. In the context of the model, it is the number of miles which the driver "brings to the job each week" (which can be positive or negative), according to the model estimate. Allowing this specific flexibility in the regression model provides a statistical adjustment for the relative speed with which drivers of high and low job match turn over, in predicting the miles each driver will operate per week. So we take the predicted values from the model and compute the average of these values over each week of tenure on the job. These averages, when

36. New drivers also earn their first week of paid vacation at this point, but that cannot be the reason for the drop in the averages, as those weeks have been removed from the data.

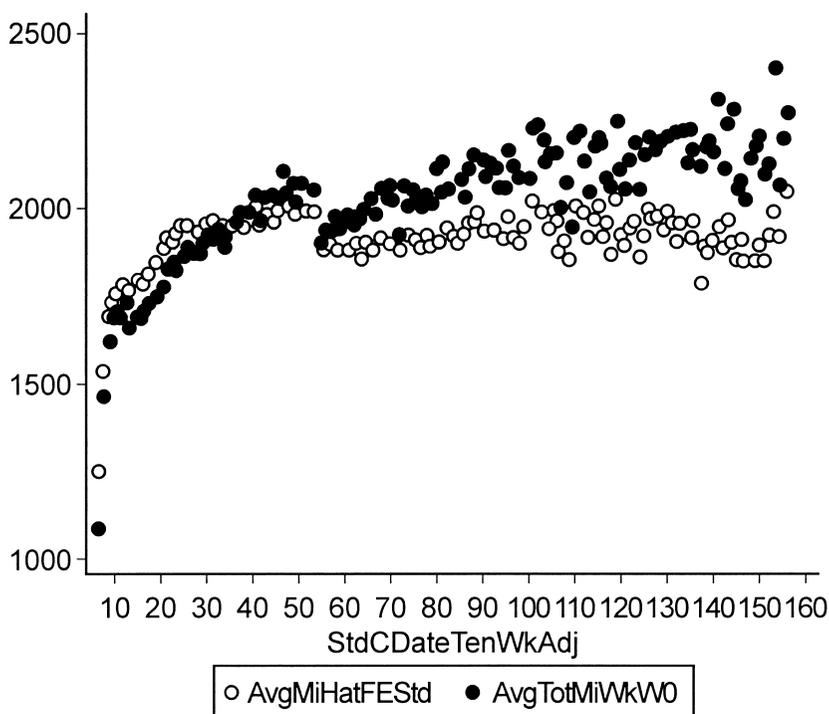


Fig. 2.7 The tenure-productivity curve with (AvgMiHatFEStd) and without (AvgTotMiWkWO) the fixed effects correction for selection

graphed, produce a “selection-corrected” tenure-productivity curve. In figure 2.7, we compare this new tenure-productivity curve with the simple descriptive version of the same curve we exhibited in figure 2.6.³⁷

When both curves are level and the selection-corrected tenure-productivity curve is below the old curve, the graph suggests that the true effect of tenure on productivity is smaller than it appears in the simple descriptive case. This implies that drivers with poor job matches are leaving differentially faster than those with good ones, which is in accord with our hypothesis. By contrast, when the selection-corrected tenure-productivity curve is either rising more rapidly than the old curve, or is above it when they are level, it says that the true effect of tenure on productivity may be larger than it appears in the simple descriptive case. This could imply that drivers with good job matches are leaving differentially faster than those with bad ones. The pilot results shown in figure 2.7 clearly show that our initial hypothesis is true from about the ninth or tenth month on. But fig-

37. Both curves are for the case in which zero-miles weeks that cannot be specifically explained away are retained in the data set.

ure 2.7 is ambiguous about whether this is also true early in the tenure of new drivers, when the firm has its highest rates of separations. A major goal of the full statistical case study is to clarify these pilot results.

2.6 Research Component Two: Panel Study of New Hires

Research Component Two is a study of 1,069 new driver-trainees who were among those recruited by the firm to start their education at a specific training school operated by the firm. The basic design of the panel study is quite straightforward in conception, although it is quite labor intensive and costly to carry out. A large amount of data is being collected on each driver trainee, starting with an initial contact while each was in the first phase of training, and then continuing with follow-up data collections over two years of the trainee's work life at the firm, or until the trainee exits the firm, whichever comes first.

The follow-up data collections include these elements: (1) a follow-up paper survey for the driver mailed to his or her home every six months, for two years, as long as they stay employed by the firm; (2) an exit survey for the driver mailed to the driver's home soon after their separation, if it occurs during the first two years; (3) a weekly survey with two questions sent over the satellite unit to the driver's truck;³⁸ (4) an initial survey (and consent form) for the driver's spouse or significant other, asking about family/work-life issues, and mailed to the driver's family soon after the driver entered the study; (5) a follow-up survey to the driver's family mailed every six months up to two years; and (6) an exit survey for the driver's family mailed upon driver separation, if it occurs during the first two years. As subjects are informed as they enter the study, a cash gift of \$5 is included in each survey mailing, with the goal of increasing the response rate. Finally, the drivers' on-the-job performance data will be collected as part of future updating of the master data files for the turnover and productivity studies of Research Component Two.

2.6.1 The Use of Behavioral Economic Experiments

A central project design goal is to perform a multivariate statistical analysis of the relationship between all the factors that are being measured and the success on the job of the trainees, where employee success is measured first by the length of time they are retained by the firm and, second, by their productivity on the job.³⁹ While the researchers (and the firm's managers) start with a number of hypotheses about what might matter in predicting each of these outcome variables, this research component is essentially a

38. The two questions are: "How happy are you with your job right now?" (Likert scale response), and "How many miles do you expect to run next week?"

39. The analysis has the potential to be extended to include safety performance, but that is not part of the present project.

large-scale empirical investigation, designed to let the data tell us which factors matter statistically and which do not. In this regard, it is particularly exciting that we have the opportunity to employ both traditional measurement instruments and a selected set of behavioral economic experiments.

There are many reasons one might have for doing behavioral economic experiments in the field. But one cross-cutting categorization is whether the investigation is primarily about treatment effects or about differences in individual characteristics. A treatment effect is exactly what it sounds like: analogous to studying the aggregate differences in the symptoms of ill patients who randomly received a specific medical treatment, as compared to those who randomly did not, an economic treatment effect is the difference in aggregate behavior across two variants of the experimental setting.⁴⁰ An example is the difference in average transaction prices between traders in a pit market and those in a double-auction market for the same commodity and with the same valuations and costs.

However, simple behavioral economic experiments can also be thought of as measurement tools for the characteristics, including the preferences, of individual subjects, as argued in Camerer and Fehr (2004). This is the approach that fits the specifics of the institutional setting of the Project. We anticipate that the results we accumulate will provide significant evidence on the relative utility of conventional and experimental measures of individual characteristics in predicting on-the-job outcomes and in complementing or substituting for each other.

One methodological point should be noted. We plan to look at the relationships among the various measures we are collecting on each subject, as there is little evidence in the literature on this topic for many of our measures, let alone evidence using the subject population from which we are drawing our participants. If this were our only goal, it would be important to vary the order in which the different measures are implemented during the initial data collection, as it is quite possible that order effects could be important for some of these relationships. Order variations are a standard feature of many experimental economic designs when individual subjects take part in more than one experiment.

However, because the central design goal of our project is to examine the predictive power of the various measures with respect to individual on-the-job outcomes, a countervailing methodological need is to present—as nearly as possible—exactly the same set of stimuli to each subject so that the relationship to the outcome variables is always the same. In addition, the complexity of administering the data collection process at the field site made it essentially impossible to vary the order, once we found a sequence

40. Of course, one of the key features which makes experimental evidence the scientific gold standard is that random assignment to treatment and nontreatment groups means that any causal factors not directly controlled for in an effective experimental design wash out of the results because they equally affect both groups.

of data collection events that fit the schedule. For these reasons, the sequence of measurements in the initial data collection was not varied.

The initial data collection process took place on a Saturday in the middle of the first two weeks of the training process for new driver trainees. We next describe the training process to provide context for understanding where the initial data collection fits in the new driver's work life.

2.6.2 The Driver Training Process

The first two weeks of training for every inexperienced driver take place at one of the driver training schools maintained by the firm. The initial training includes a large dose of classroom work in which students are introduced to the firm and learn essential facts about the equipment they'll be operating, the regulations governing commercial vehicle operators and operations, map reading and course-plotting, and the safety rules and procedures specific to the firm. It also alternates hands-on training time between truck simulators and an actual Class 8 tractor-trailer.⁴¹ Trainees first drive bobtail (tractor with no trailer), and then with an empty and a loaded trailer, on the school's property. Once the trainee has qualified for a CDL learning permit and is judged ready by the instructors, he or she quickly begins driving on the actual roads and highways surrounding the school, with a driver-trainer in the right seat of the tractor.

The simulators have two purposes. One is to speed familiarization with the basic features and operational characteristics of the Class 8 tractor-trailer. An example is learning how to correctly shift a ten-speed transmission, which requires frequent double-clutching, coordinated at first by conscious attention to engine RPMs in comparison to road speed, especially for downshifts. Another basic feature drivers are first exposed to in the simulator is how to maneuver around corners on city streets with a rig that is about 65 feet (just under 20 meters) long, and that includes a trailer that is 53-feet (16.2 meters) long. A second main function of the simulators is to give drivers practice at responding to dangerous settings that could never be practiced in real life, such as how to avoid going off the road in response to a blow-out on a steering axle tire, first on dry pavement, and then on glare ice on a freeway in the middle of car traffic.

Trainees who complete the initial two-week training process are officially hired by the firm on their date of completion, after passing a basic skills qualification test. Once hired, the driver goes back to his or her home base at one of the firm's terminal locations. While the content of the CDL

41. By definition, Class 8 vehicles that can operate on the National Highway System (NHS) have a maximum GVW of between 33,000 pounds and 80,000 pounds. The units utilized by most truckload carriers, including the cooperating firm, are at the top of this range, and have a maximum GVW of 80,000 pounds. (The NHS is a large subset of all U.S. highways designated by the U.S. Department of Transportation (USDOT), on which federal size and weight standards prevail.)

exam is federally regulated, the exact rules governing driver training vary from state to state.⁴² So depending on their proficiency level and on the state in which they will be based, trainees may also take the CDL exam at the training school, or they may do so later, at a location nearer their home base. Whether they have the CDL yet or are still using a learner's permit, all new drivers have a second training phase, during which the trainee works for between one and three more weeks, actually hauling freight from their home base, but with a certified driver trainer in the right seat of the tractor. When the trainer judges the trainee ready (and he or she has acquired a CDL), the new driver is assigned a tractor of his or her own and goes to work on his or her own.

As is mentioned in previous sections, it is typical for the TL carriers that train all or most of their new drivers to offer the training on a credit contract. The contract specifies that the trainees assume a debt of several thousand dollars, which is approximately the market value of the training. According to the contract, the firm provides both phases of the training, including access to the classroom, instructors, and trucks, plus needed classroom supplies. For the first phase of training, transportation to and from the training school, a hotel room while at training, and lunches during the training day are also covered. The trainee owes no payments on the credit extended as long as they stay employed by the firm, and the debt is fully discharged if the trainee completes a specific period of service after the point they begin working on their own (either one year, at many firms, including this one, or eighteen months at some others). Trainees who do not complete the period of service, however, become legally liable for repayment of the amount of credit extended.

2.6.3 The Initial Data Collection Process

The initial data collection process was piloted in October through December, 2005, and went to full operation on twenty-two Saturdays during January through August, 2006. The school starts a class into the initial two-week training process (described in the preceding section) every week on Sunday (thus, two classes are in residence at any given time). So trainees in target classes received a one-page flier advertising the study as part of their orientation on their first day at school. The flier explained the opportunity to take part in the University of Minnesota study the next weekend and that trainees who volunteered could expect to earn cash by participating. At the beginning of the class day on the Saturday of data collection, Burks conducted an informed consent process in which he introduced himself as a former driver turned university researcher and explained the goals and

42. The CDL comes in three categories, "Class C" through "Class A," and for freight vehicles are differentiated primarily by the ascending maximum GVW of the trucks the bearer can operate. The Class A CDL is required to operate Class 8 vehicles on public roads.

procedures of the study to the potential subjects. The study is governed by the standards of the Human Subjects Protection Committee of the University of Minnesota's Institutional Review Board (IRB), so a formal consent document was used which spelled out the risks and benefits of taking part in the study.⁴³

The data collection at the training school was set up as two, two-hour-long blocks, spent doing tasks with the researchers, either on computers or with paper and pencil, with a short break in between. Training classes range from thirty-five to seventy students, and the largest group that could be accommodated at one time for data collection was thirty-two subjects, so each class was broken into two groups. The first group worked with the researchers from early to late morning, and the second group from late morning to midafternoon. On data collection days, the class day was structured so that those potential subjects who chose not to take part did not have extra training available, but instead could spend the extra time in the break room. The buses that transport students to and from the hotel bring everyone at one time in the early morning and take everyone back at one time in the afternoon. Given the monetary compensation being offered, the relatively low opportunity cost of taking part, and the credible guarantee of confidentiality from the University, 91 percent of the trainees offered the opportunity chose to join the study. Initial data was collected from 1,069 participants, with clean information on 1,036 of these.⁴⁴

Except for one instrument (the Educational Testing Service's test of Quantitative Literacy, labeled "Numeracy" in our schedule), all the instruments were administered on a wireless network provided by the cooperating firm of refurbished Dell notebook computers, with a newer Dell notebook as the master controlling computer. The software used was z-Tree (Fischbacher 2007), the toolbox for constructing computer-administered economic experiments and surveys developed at the Institute for Empirical Economics at the University of Zürich.⁴⁵ The data collection took place in a temporary computerized experimental economics laboratory that the

43. This document included the following key facts: (1) the kind of data collection activities involved, both initially and later, (2) that subjects would receive \$20 in initial cash "thank you gifts" for taking part, and have the opportunity to earn substantially more, (3) that the money paid to them was coming not from the firm but from two nonprofit foundations, (4) that under university rules for protecting research participants, the new data collected about them would be kept confidential from everyone except academic investigators approved by the Institutional Review Board (IRB), including specifically that it would never be available to their managers at the trucking firm, and (5) that participation was completely voluntary.

44. Due to a programming glitch, one item, Hit 15 Points, has valid information for only 893 subjects.

45. Because we are running a licensed adaptation of a nonverbal IQ instrument that requires the display of a large number of graphics files, each of which is a scanned image of a page with a pattern-matching task, we ran a beta version of z-Tree that was extended to handle graphics files. Our special thanks to developer Urs Fischbacher for providing this new version in time for our project.

research team set up in one of the classrooms at the training school on the Friday before the data collection dates on Saturdays, and which was then removed at the end of the Saturday event. The team hand-built a set of cloth dividers hung from frames adapted from portable garment racks, which were installed between all the computer stations in order to separate subjects during the data collection. Because the data collection sessions exceeded the endurance of notebook PC batteries, extra power outlets were installed in the walls and a set of portable extension cords were deployed as part of the laboratory set up. The wireless network was part of a virtual local area network (VLAN) set up by the firm's IT support personnel under a data security agreement approved by the University of Minnesota Human Subjects Committee. The VLAN separated the computers on which the confidential data was being collected from the rest of the firm's network and provided secure data storage until the initial data collection was complete, and the project moved back to the UMM campus.

Table 2.2 shows the time sequence and schedule for the actual data collection event of February 25, 2006. This class of trainees was relatively large, at fifty-nine students. Fifty-one out of fifty-six, or 91 percent, chose to take part.

While it was theoretically possible for a subject to end up with only the initial \$20 "thank you gifts," we always paid at least a dollar or two more. The actual payouts depended in part on the use of stochastic devices (plastic bowls filled with colored or numbered poker chips), as described in the following account of each data collection activity. The average total earnings for the four-hour period of the initial data collection were \$53, with a minimum of \$21, and a maximum of \$168. By design, we wanted most subjects to be pleased about their initial interaction with the researchers, given the follow-up contacts with them which are called for by the design.

The latter fact—that we needed to gain not only the immediate but also the longer-term cooperation of our subjects—also affected the protocol for subject behavior during the four-and-one-half-hour sequence of data events. With undergraduate students in a university experimental economics laboratory and with a protocol that takes significantly less time than the project's, it is not hard to ask subjects to be quiet, to not speak to their neighbors, to not do other tasks on the computers, and not to read, write, or use cell phones or PDAs. In order to provide a neutral setting, this is the behavior requested—and normally received from—typical student subjects in economic experiments, unless the treatment being investigated is one involving changing some aspect of this, such as allowing specific types of communication. We established expectations of our driver-trainee participants that were similar to this, but found that to receive willing compliance we had to allow for participants who finished ahead of others on tasks that took a long time to get up and leave the lab setting temporarily. We did this explicitly during the Numeracy instrument and during the MPQ in-

Table 2.2 Data collection activity time allocations: February 25, 2006

Actual time	Scheduled time	Total time	Informed consent process
7:13 59	7:10		Start informed consent process How many people in the room?
7:29	7:30	0:16	Finish informed consent process
Session 1a			
7:30	7:30	0:09	Check-in
7:39	7:40	0:04	Information
7:43	7:42	0:19	<i>Activity 1: Prisoner's Dilemma</i> Computer generated: Pay on beliefs for what percentage of people will send \$5 as person 1, how much person 2 sends if person 1 sends \$0, and how much person 2 sends if person 1 sends \$5, \$1 each
8:02	8:09	0:39	<i>Activity 2: Multidimensional Personality Questionnaire</i> No payment
8:41	8:47	0:12	<i>Activity 3: Risk/Loss Aversion</i> 1 question, #23; <i>green</i> by drawing out of a bowl; everyone paid for their selection
8:53	8:58	0:27	<i>Activity 4: Demographics</i> No payment
	9:16		<i>Activity 5: Big Red Button</i> Computer generated Release participants
9:09			First person left at this time
9:20	9:26		Last person left at this time
Session 2a			
9:37	9:40	0:08	Check-in
9:45	9:45	0:02	Information
9:47	9:47	0:09	<i>Activity 1: Time Preferences</i> 2 subjects, #12 and #19; 1 question, #18
9:56	9:57	0:47	<i>Activity 2: Nonverbal IQ</i> Start time of test 10:02 2 subjects, #8 and #21; pay \$1 for each correct answer; pay on "pre" and "post" beliefs, \$2 each
10:43	10:30	0:25	<i>Activity 3: Numeracy (Quantitative Literacy)</i> Start time of test 10:44 2 subjects, #8 and #18; pay \$2 for each correct answer; pay on "pre" and "post" beliefs, \$2 each
11:08	10:55	0:08	<i>Activity 4: Ambiguity Aversion</i> 1 question, #7; <i>blue</i> by drawing out of a bowl; everyone paid for their selection
11:16	11:05	0:12	<i>Activity 5: Hit 15 Points</i> Computer generated
11:28	11:28	0:12	<i>Activity 6: Risk, Impatience, and Cooperation Survey</i> No payment Release Participants
11:35			First person left at this time
11:40	11:40		Last person left at this time

(continued)

Table 2.2 (continued)

Actual time	Scheduled time	Total time	Informed consent process
Session 1b			
12:15	12:30	0:10	Check-in
12:25	12:40	0:03	Information
12:28	12:42	0:23	<i>Activity 1: Prisoner's Dilemma</i> Computer generated; pay on beliefs for what percentage of people will send \$5 as person 1, how much person 2 sends if person 1 sends \$0, and how much person 2 sends if person 1 sends \$5, \$1 each
12:51	1:09	0:35	<i>Activity 2: Multidimensional Personality Questionnaire</i> No payment
1:34	1:47	0:13	<i>Activity 3: Risk/Loss Aversion</i> 1 question, #21; blue by drawing out of a bowl; everyone paid for their selection
1:47	1:58	0:32	<i>Activity 4: Demographics</i> No payment
	2:16		<i>Activity 5: Big Red Button</i> Computer generated Release participants
2:04			First person left at this time
2:19	2:26		Last person left at this time
Session 2b			
2:25	2:40	0:09	Check-in
2:34	2:45	0:02	Information
2:36	2:47	0:11	<i>Activity 1: Time Preferences</i> 2 subjects, #17 and #16; 1 question, #7
2:47	2:57	0:39	<i>Activity 2: Nonverbal IQ</i> Start time of test 2:53 2 subjects, #17 and #22; pay \$1 for each correct answer; pay on "pre" and "post" beliefs, \$2 each
3:26	3:30	0:28	<i>Activity 3: Numeracy (Quantitative Literacy)</i> Start time of test 3:30 2 subjects, #17 and #16; pay \$2 for each correct answer; pay on "pre" and "post" beliefs, \$2 each
3:54	3:55	0:11	<i>Activity 4: Ambiguity Aversion</i> 1 question, #22; blue by drawing out of a bowl; everyone paid for their selection
4:05	4:05	0:16	<i>Activity 5: Hit 15 Points</i> Computer generated
4:21	4:28	0:16	<i>Activity 6: Risk, Impatience, and Cooperation Survey</i> No payment Release participants
4:29			First person left at this time
4:37	4:40		Last person left at this time

strument, when differences in reading speed made the differences in completion time especially large, and we also allowed participants to read after completion if they were done early during the MPQ.

We next turn to a brief description of each item in the sequence of data collection events.

2.6.4 Session One Data Collection Events

The first two-hour block of data collection activities consists of three behavioral economic experiments and two more conventional measures.

Prisoner's Dilemma

Our version of this experiment is a sequential and strategic one. Person 1 (the first mover) and Person 2 (the second mover) each are allocated \$5. Person 1 can send either \$0 or \$5 to Person 2, and Person 2 can respond by sending \$0, \$1, \$2, \$3, \$4, or \$5 back. All funds sent are doubled by the researchers. Each subject provides their complete strategy in the game: they make both an unconditional decision for the first-mover role and a conditional one for the second-mover role (first, how to respond to being sent \$0 and, second, how to respond to being sent \$5, doubled to \$10.) Subjects are randomly matched and their role selected by the computer after their decisions. This is a variant of the task used in Burks, Carpenter, and Götte (2006).

Before each decision screen, subjects are also asked how they think other participants in the room will act in this experiment. The first question is “What percent of the participants do you think will send their \$5 as Person 1?” and pays \$1 if the subject is correct within plus or minus 5 percent. The second and third questions are “If Person 1 does not send/does send, what is the average that participants in this room will send back?” and pays \$1 for each question if the subject is within plus or minus \$0.25 of the actual average.

There will be four main data items of potential interest from this experiment: first-mover choices, the estimate of first-mover behavior of others, second-mover conditional choices, and the estimate of the second-mover choices of others. We intend to analyze what other factors about the participants predict their choices in this experiment, as well as use some aspects of their choices as predictors of other experimental responses, and also of some kinds of on-the-job behavior. A potential workplace application is in predicting on-the-job cooperation. The TL driver's job is relatively individualized, but there are two instances in which the data store from Research Component One may in the future be augmented to permit the construction of an on-the-job measure of cooperation.⁴⁶

One of the benefits of the strategic form of this experiment is that the

46. See appendix B, items 4.1 and 4.2.

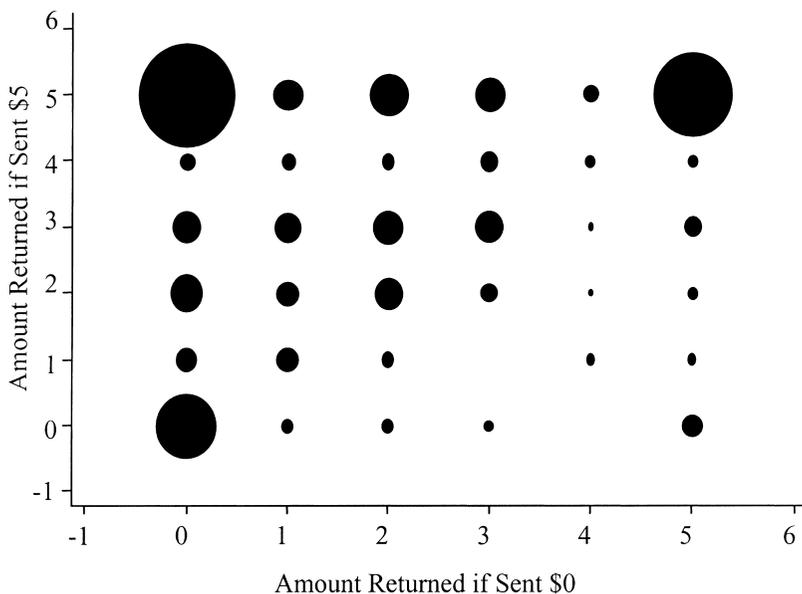


Fig. 2.8 Second-mover strategies in the sequential prisoner's dilemma

second-mover choices provide a fairly clear typing of the respondents, in terms of who is an egoist (never sends anything to the first mover), who is a conditional cooperator (sends back exactly what the first mover sent), and who is an altruist (always sends everything), with some gradations possible for intermediate dollar responses. As figure 2.8 shows, participant behavior is varied in this experiment, but “pure” versions of these three strategies predominate among second movers (shown by the large circles, which are sized proportionally to the number of respondent choices they capture). The three pure strategies make up 61 percent of the total responses, and most of the rest are intermediate responses. Those responses below and to the right of the 45-degree line fall in the “wingnut” category; these participants may have misunderstood the instructions, a hypothesis we can test by correlating these responses with some of our measures of cognitive performance.

Multidimensional Personality Questionnaire (MPQ)

The Multidimensional Personality Questionnaire (Patrick, Curtin, and Tellegan 2002), otherwise known as the MPQ, is the second measurement instrument. This is a standard personality profile that consists of eleven different scales that represent primary trait dimensions: well-being, social potency, achievement, social closeness, stress reaction, alienation, aggression, control, harm avoidance, traditionalism, and absorption. The short

version used in the study has 154 multiple-choice questions. Almost all of the 154 questions have the same four possible answers: “Always True,” “Mostly True,” “Mostly False,” and “Always False.” There is no separate payment for this survey.

A principal use of this scale will be to generate a set of control variables in regression modeling of on-the-job outcomes. We also expect to look at how the traits defined by the MPQ relate to answers to the other survey instruments and to behavior in the experiments. As figure 2.9 shows, the trainee drivers are similar overall to the benchmark population presented in Patrick, Curtin, and Tellegan (2002), which is promising given that our primary goal is to use the variation across individuals in the components of the MPQ. Some initial hypotheses, for instance, are that high “achievement” orientation should be correlated with job success as a TL driver, especially given the incentive-pay method faced by these drivers. One noteworthy point is the apparent difference from the population in the factor “unlikely virtues.” The questions in this factor are set up so that only someone who is “too good to be true” will have a high score, and the authors of the instrument intend it as a check on the truthfulness of the subject’s re-

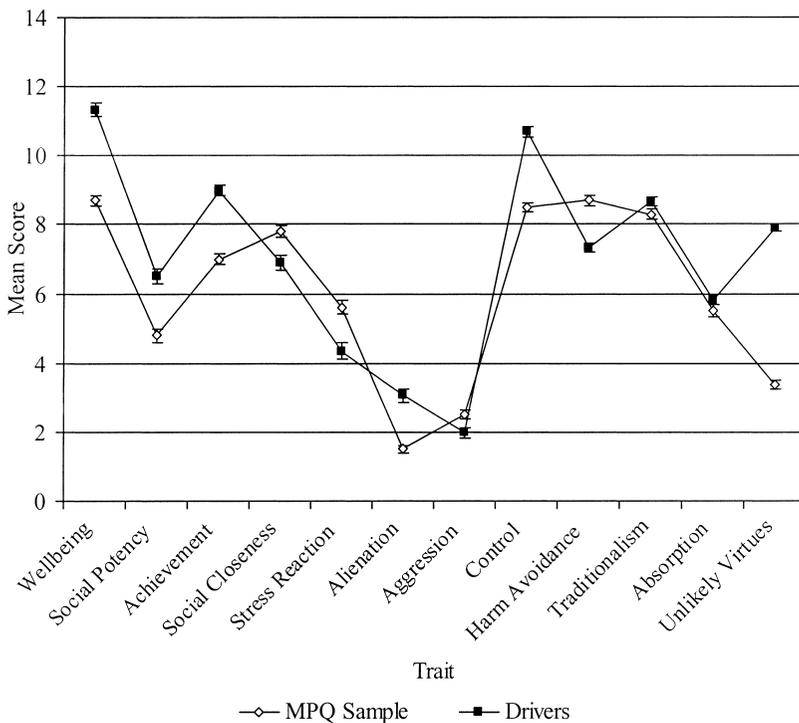


Fig. 2.9 Multidimensional Personality Questionnaire profile of subjects compared to MPQ reference population profile

sponses. This MPQ factor offers the prospect that we may be able to control for this effect to some degree in multivariate analysis by using the unlikely virtues score as a regressor.

Risk/Loss Aversion

In the risk/loss aversion experiment, there are twenty-four questions that are divided into four blocks of six questions each (see table 2.3). There are two possible choices for each question, an amount of money received with certainty, and a 50/50 gamble that pays a higher dollar amount if “your color” is chosen and a lower dollar amount if the “other color” is chosen. To avoid any hint of experimenter control over the outcome, subjects choose which color (blue or green) is theirs for the random outcomes as their last response.

Each of the four blocks of six questions follows the same format. In each block, the amount for “your color” and the “other color” do not change, and the amount of the “for sure” option increases by \$0.50 per question. The design is intended to identify where the subject crosses over from the “for sure” option to the “gamble” option, relative to where a risk-neutral person would do so. After all subjects have made their decisions, one poker

Table 2.3 Risk/loss aversion

Question no.	For sure	Your color	Other color
1	Win \$2.00	Win \$10.00	Win \$2.00
2	Win \$3.00	Win \$10.00	Win \$2.00
3	Win \$4.00	Win \$10.00	Win \$2.00
4	Win \$5.00	Win \$10.00	Win \$2.00
5	Win \$6.00	Win \$10.00	Win \$2.00
6	Win \$7.00	Win \$10.00	Win \$2.00
7	Win \$0.00	Win \$5.00	Lose \$1.00
8	Win \$0.50	Win \$5.00	Lose \$1.00
9	Win \$1.00	Win \$5.00	Lose \$1.00
10	Win \$1.50	Win \$5.00	Lose \$1.00
11	Win \$2.00	Win \$5.00	Lose \$1.00
12	Win \$2.50	Win \$5.00	Lose \$1.00
13	Lose \$2.50	Win \$1.00	Lose \$5.00
14	Lose \$2.00	Win \$1.00	Lose \$5.00
15	Lose \$1.50	Win \$1.00	Lose \$5.00
16	Lose \$1.00	Win \$1.00	Lose \$5.00
17	Lose \$0.50	Win \$1.00	Lose \$5.00
18	Lose \$0.00	Win \$1.00	Lose \$5.00
19	Win \$1.00	Win \$5.00	Win \$1.00
20	Win \$1.50	Win \$5.00	Win \$1.00
21	Win \$2.00	Win \$5.00	Win \$1.00
22	Win \$2.50	Win \$5.00	Win \$1.00
23	Win \$3.00	Win \$5.00	Win \$1.00
24	Win \$3.50	Win \$5.00	Win \$1.00

chip is drawn by a participant that identifies which question (one through twenty-four) everyone will be paid on, and then a second participant draws a colored chip to determine whether the winning color, for subjects who chose the gamble on the selected question, is green or blue.

The choice blocks overlap the zero-dollar point in order to allow inference about the likelihood that the implied utility function has different slopes in the positive and negative domains. But there is no set of choices only in the negative domain because the project is relying on the long-term responsiveness of participants, and pilot work showed that a choice block fully in the negative domain was perceived as too unfriendly a task.

Figure 2.10 shows the distribution of the total number of risky choices for our subject pool and shows that there is significant variation across participants, which is desirable for the purpose of using the variations across individuals to predict other behavior. Risk aversion and/or loss aversion are potentially key explanatory factors in TL firm turnover because nearly all of each driver's pay is based on piece rates (cents per mile for a fixed number of miles for each dispatch). Paychecks can vary quite substantially from one week to the next, and, thus, there is a certain level of short-run financial risk that goes with the job. The question for the study is to what extent this form of experimental measurement of financial risk is predictive of the impact of pay fluctuations on driver turnover, as compared to things such as survey responses and credit score.

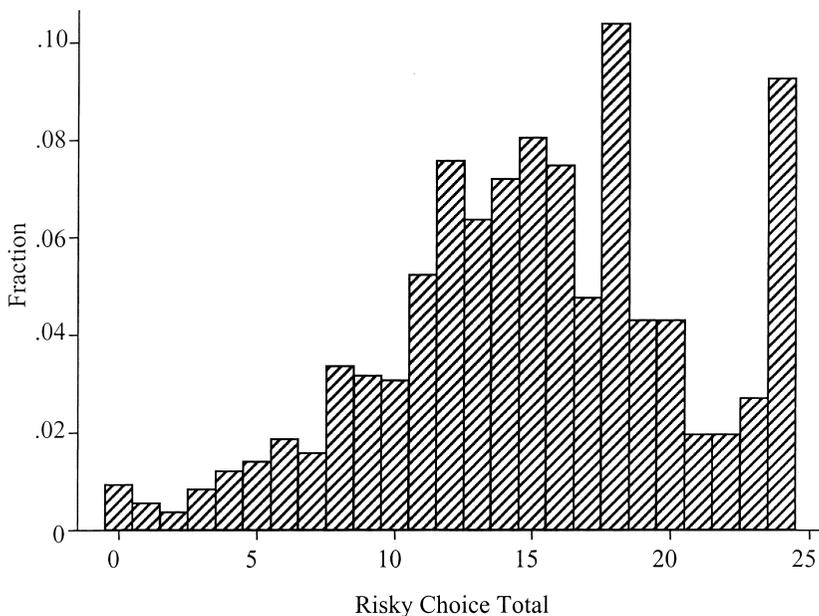


Fig. 2.10 The total number of risky choices

Demographics

This instrument is a compilation of thirty-seven demographic questions. The types of questions vary, but the areas that are covered include education level, languages, marital status, previous job experience, family lifestyle, racial or ethnic identity, country of birth, how many people are in the household, household income, and networking done while at training. These questions are mostly multiple choice, with a few numerical answer questions as well. There is no separate payment for completing this demographic questionnaire.

Figure 2.11 shows the distribution of ages of the trainee drivers, as compared to the distribution among blue-collar workers, as computed from the March 2005 Current Population Survey. The minimum age to acquire the CDL is twenty-one, so there are no new trainees that are less than this minimum age. However, in general the trainee pool is similar in distribution to the entire population of blue-collar workers in the United States, albeit with a modest shift toward the age range of twenty-one to thirty.

Figure 2.12 shows the distribution of the highest education level completed among the trainee drivers, with the distribution among blue-collar

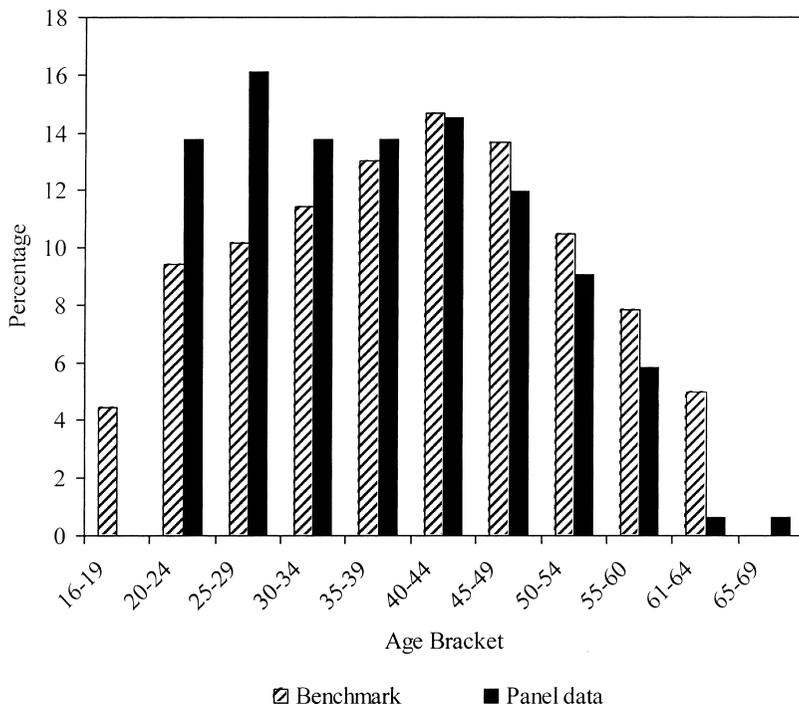


Fig. 2.11 Distribution of participant age with benchmark

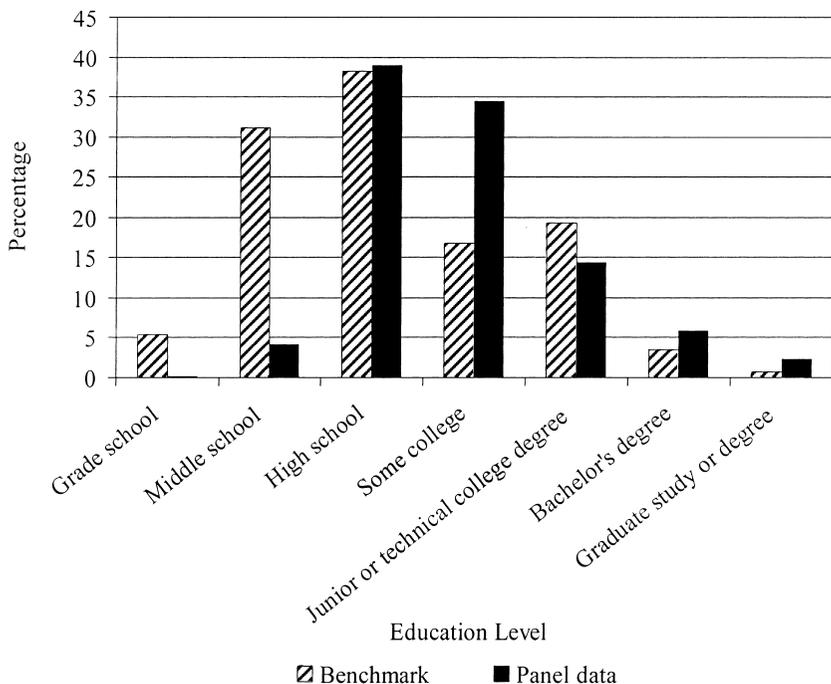


Fig. 2.12 Participant education levels with benchmark

workers in the United States as a comparison. The minimum literacy requirements for the TL driver job are undoubtedly the reason there are very few trainees with very low educational attainment, compared to the blue-collar workforce in the United States. It is interesting that the trainee group has a much higher proportion of individuals with some college, but possibly a lower proportion with a junior college or technical degree, as compared to the population of blue-collar workers. Given that someone who has acquired some level of useful education in college is overqualified for this job, this pattern is consistent with the view that completing a two-year degree is indeed correlated with job market success as fewer of these individuals are in our subject pool. A speculation about the apparently slightly higher proportion of four-year-degree holders in the trainee group, as compared to the benchmark, is that these are folks with other options who are attracted (at least initially) by the TL-driver lifestyle.

The Red Button

This experiment is designed to measure short-term impatience. It is the last data collection event in the first of the two, two-hour sessions, and it is followed by ten to twenty minutes of break time, during which participants can use the restroom and stretch their legs and get something from the

vending machines if they wish. It is directly preceded by the demographic questionnaire, to which it is linked. This means that, unlike most of the other instruments, subjects do not have to wait until everyone else completes the prior activity, but instead each participant can start the red button task as soon as he or she has individually completed the demographic questionnaire. This setup is designed to minimize peer effects in the red button task as each subject will start the task at a different time.

The red button activity takes exactly ten minutes, and the task is to wait quietly until a timer on the task screen counts down from 600 seconds to zero. As described in section 2.6.3, during each session subjects are requested in general to either work on the task at hand or wait quietly. We ask that they do not to talk to their neighbors, nor engage in any other activity, such as pulling out something to read. During the red button task, these rules are quite strictly enforced. Participants are paid \$5 if they complete the task by waiting the full ten minutes. However, subjects may choose to end this experiment earlier if they wish. Each click of a button below the countdown clock on the computer screen, which is labeled “Reduce Waiting Time,” will cut the individual’s waiting time—but each click also costs them \$1. The first click reduces the waiting time by five minutes; the second click reduces the waiting time by three more minutes, and the third click reduces the waiting time by an additional two minutes. Once the timer says zero, no matter how long the subject has waited, then he or she is free to get up and leave the room for break time. As figure 2.13 shows, 29 percent of the participants clicked at least once. This is actually lower than we aimed for; our initial calibration runs may have led us to set the opportunity cost of the first click too high.

Patience is relevant to quit decisions, and being impatient can lead to poor job performance when a significant part of the job involves being able to wait on customer docks and then hurry up when customers are ready. Being patient in traffic is also a relevant job skill. We expect to investigate the extent to which red button choices will predict impatience on the job, as measured by quit decisions, controlling for prior earnings and prior waiting time during the earlier data collection activities.

2.6.5 Session Two Data Collection Events

The second two-hour block of data collection activities contains three behavioral economic experiments and three more conventional measures.

Time Preferences

In this experiment, there are twenty-eight questions that are divided into four blocks of seven questions each. There are two possible choices for each question: a smaller amount of money paid sooner, and a larger amount of money paid later. Each of the four blocks of seven questions follows the same format. The amount for the higher payoff at a later date is al-

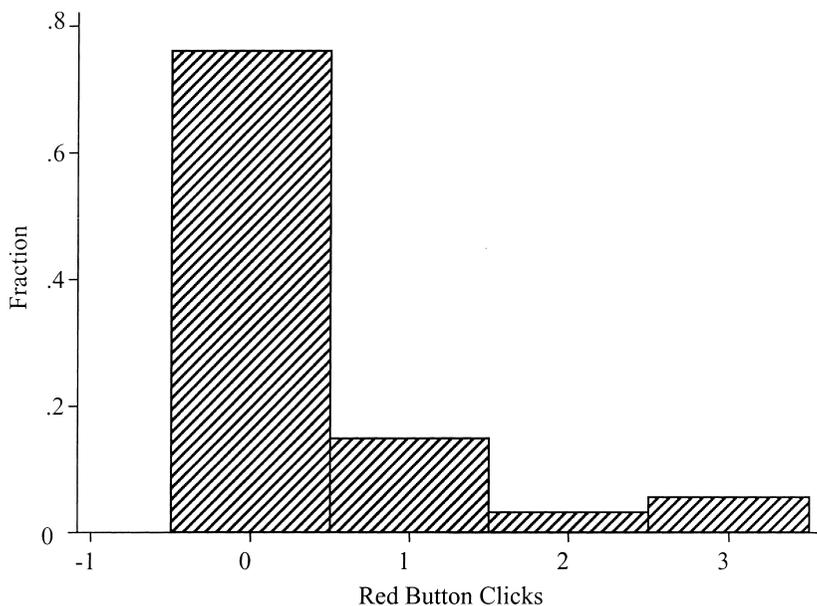


Fig. 2.13 Clicks to reduce waiting time at a cost

ways \$80, and the amount for the lower payoff at an earlier time begins at \$75 and decreases by \$5 intervals to \$45. The point at which a subject switches from the later payoff to the earlier one (if a switch is made) provides an implicit point estimate of the subject's discount rate over that time horizon.

The time frames are (1) today (Saturday) versus Monday; (2) today versus next Thursday; (3) Monday versus Monday plus one week, and (4) Monday versus Monday plus four weeks. The two matched pairs of time frames are designed to allow subjects that have different discount rates for choices with and without a front-end delay, that is, with and without an initial delay for both payments.⁴⁷ After all subjects have made their decisions, a participant draws one poker chip to select which of the twenty-eight questions will be activated, and then two poker chips are chosen the same way which identify two of the subjects in the session who will be paid for their choices on that question. Payments are, of course, made on the dates requested, either in person, or with an official bank check mailed in a University of Minnesota envelope.

As figure 2.14 shows, the three modes of the distribution of future

47. The contrast between choices in which one option is an immediate payment versus those in which both payments involve a delay will permit us to observe "impatience for immediate results," often described in the literature as quasi-hyperbolic discounting, if it should appear.

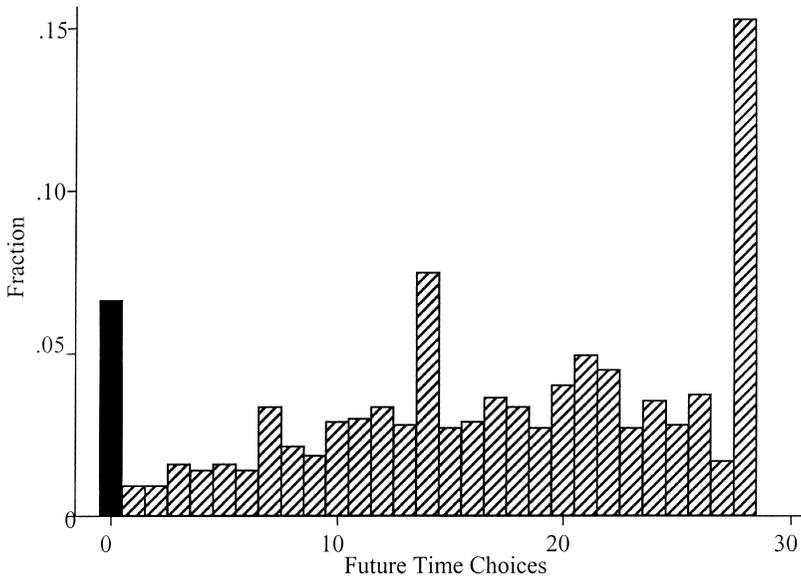


Fig. 2.14 Total future choices

choices are at zero, fourteen, and all, with a median of seventeen future choices out of a possible twenty-eight. There is a wide dispersion in individual responses. We expect to examine the relationship between this measure and such other measures as nonverbal IQ, risk/loss aversion, and impatience. And we will investigate the extent to which it adds predictive power to our statistical models of quits.

Nonverbal IQ

The IQ instrument used is a computerized adaptation of the Standard Progressive Matrices by J.C. Raven (Raven, Raven, and Court 2000); the authors created the adaptation under license from The Psychological Corporation.⁴⁸ Each question is presented as a graphic image. On top is a large rectangular box containing some kind of a pattern with a piece missing out of the lower right hand corner. On the bottom are six (or eight) possible pieces that could be used to complete the image on top. Each section starts with easy images and gets progressively more difficult.

The original instrument consists of five sections, each containing twelve questions. Our version is not fully comparable to the standard results for the instrument because we are actively administering only the last four sec-

48. Pilot work was done with two other instruments, but one had irresolvable licensing issues, and the other proved unsuitable for our setting. The first several administrations of the Raven's instrument were done with paper and pencil, while we arranged the license for the computerized adaptation (used by permission of Harcourt Assessment).

tions, due to time constraints. The original is primarily benchmarked in an untimed format, and we cut our subjects off after thirty-one minutes, having given a warning three minutes earlier.⁴⁹ The simplest way to match our data with the untimed reference benchmarks is to note that everyone in our subject pool finished the fourth section before the announcement of a time limit. Our correction is then simply to multiply our subjects' scores on sections two, three, and four by 5/3 to scale them up to the same range as the score for untimed subjects doing all five sections.⁵⁰

After both verbal and written instructions and two practice questions, subjects fill out a "confidence question" that asks them how they think they will do as compared to other subjects in the room, by placing themselves in the correct quintile of the distribution of scores. When the Raven's task has been completed, the same confidence question is asked again. Subjects are paid an additional \$2 for placing themselves in the correct quintile. In addition, two subjects are randomly chosen to be paid \$1 per correct answer, for total possible earnings of \$48 each for their question answers. Starting at about subject 200, we also began asking whether subjects want to find out their own score and the group average when they receive their payout.⁵¹

Figure 2.15 shows approximately how our panel compares to a standard benchmark population (citizens of Iowa City, IA, in the late 1990s.) On a scale of 0 to 60, our subject pool has a higher density between 36 and 44, and a lower one from 54 on up. The median for our subjects is about two points lower than that for the benchmark population. The distribution looks reasonable compared to groups of workers in other countries against which the regular version of the instrument has been benchmarked. The primary use of this measure in the project analysis will be as a control variable in predicting on-the-job outcomes. But we also expect to look at the relationships between this indicator of nonverbal IQ and the other participant characteristics that we measure. We will also examine how IQ, confidence about one's performance both before and after the task, and the desire for full information about one's performance are related.

Numeracy (Quantitative Literacy)

This instrument is part of the test of adult quantitative literacy from the Educational Testing Service. The full instrument consists of two sections, of which only the first section was used here, due to time constraints. The section is made up of twelve questions and subjects are given exactly twenty

49. There is one published benchmark from a French subject pool using a thirty-minute time limit, but including all five sections.

50. This approach produces unbiased estimates at the cost of greater variance.

51. The two selected to be paid for correct answers will learn their score from their payoff, so they will just get the average as new information, but these two subjects are selected after this question is asked.

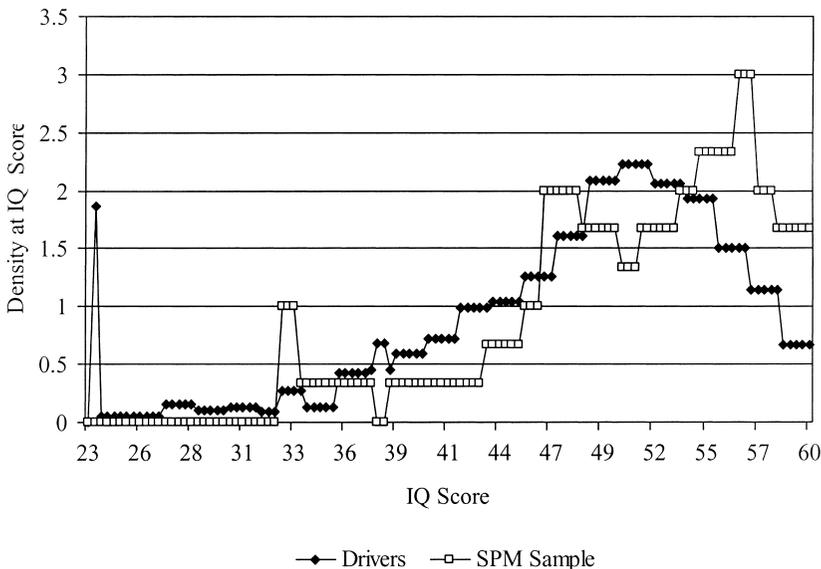


Fig. 2.15 Density of (adapted) nonverbal IQ instrument scores for drivers compared to density of benchmark population scores

minutes to complete the test. The test requires subjects to be able to add, subtract, compare numbers, compute a percentage, fill out a form, and to be able to read and understand a short problem, among other things.

As with the nonverbal IQ, after instructions and a brief practice question, subjects fill out a “confidence question” that asks them how they think they will do as compared to other subjects in the room, by quintiles. When the numeracy task has been completed, the same confidence question is asked again. Subjects are paid an additional \$2 for placing themselves in the correct quintile. Two subjects are randomly chosen to be paid \$2 per correct answer, for total possible earnings of \$24 each for their question answers. In addition, at about subject 200 we began asking whether subjects want to find out their own score and the group average when they receive their payment.⁵²

The distribution is not directly comparable to the distribution for the full instrument (i.e., when both sections are administered), but we appear to be getting good dispersion in performance (see figure 2.16). The mean score is 71 percent. The primary use of this in the project analysis will be as a control variable in predicting on-the-job outcomes. The job of TL driver re-

52. As with the nonverbal IQ indicator, the two selected to be paid for correct answers will learn their score from their payoff, so they will just get the average as new information, but these two subjects are selected after this question is asked.

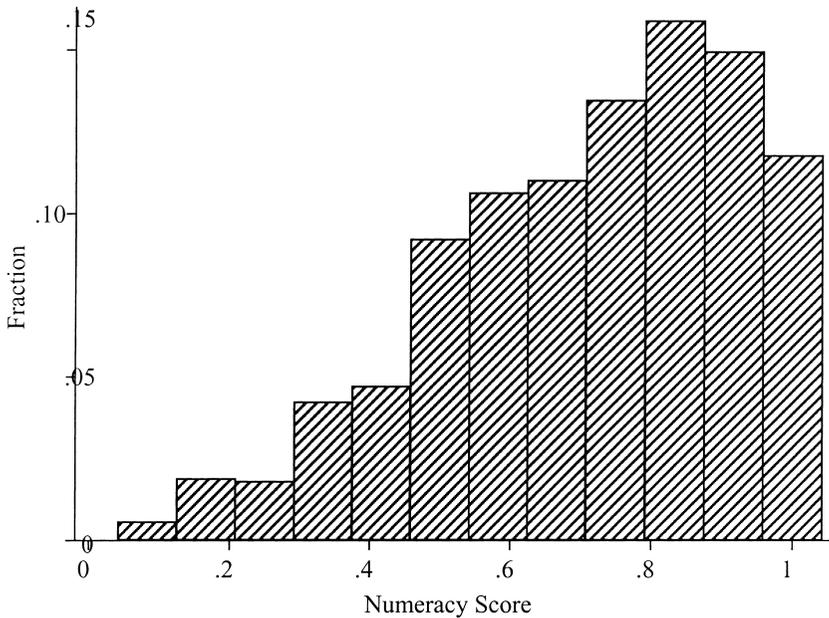


Fig. 2.16 Distribution of numeracy scores

quires continual application of numeracy skills, for example, in map reading and route planning, or in calculating hours remaining and hours coming available at specific future times, under the hours of service regulations for commercial motor vehicle operators. But we also expect to look at the relationships between this indicator of numeracy and the other participant characteristics that we measure. We will also examine how numeracy, confidence about one's performance both before and after the task, and the desire for full information about one's performance are related.

Ambiguity Aversion

Our ambiguity aversion experiment is identical to the risk/loss aversion experiment described in section 2.6.4, except for one detail: it is no longer known to be a 50/50 gamble if the subject chooses the uncertain option. Two blue poker chips and two green poker chips are placed in the bowl from which the winning color is drawn, and then out of sight of the subjects, six more chips are added that can be all green, all blue, or any mixture thereof. As a result, subjects only know that there is at least a 20 percent chance that green will be drawn and at least a 20 percent chance that blue will be drawn. All other features of the experiment are unchanged: there are four panels of six choices each, with the dollar values for certain and lottery outcomes as before. All subjects choose the certain payoff or

the lottery in each question in each panel and then which color is theirs for the lottery outcomes.

Figure 2.17 shows results for the first panel, when the choice is between a fixed amount ranging from \$2.00 to \$7.00 or a lottery with outcomes of \$2 or \$10. The horizontal axis shows the total number of lotteries chosen in the initial risk/loss aversion task, while the vertical axis shows the same total for the ambiguous version. Participants who made the same choices in both cases would be on the 45-degree line. The size of the circles shows the number of participants at each node of the grid, and it is apparent that many subjects did change the number of lotteries they selected in widely varying ways. We will be looking to see which subjects choose more or fewer risky options and whether their shift, if any, is related to other experimental or survey measures and to on-the-job success.

Hit 15 Points

This is a backward induction, or planning, task in the form of a small game between subject and computer. The computer and the subject take turns adding points to the “points basket,” and during each turn, the subject or the computer must add either one, two, or three points to the points basket. The goal is to be the player to add exactly the fifteenth point. The number of points in the points basket at the beginning of the round varies, and the computer and participant take turns going first. The first round is

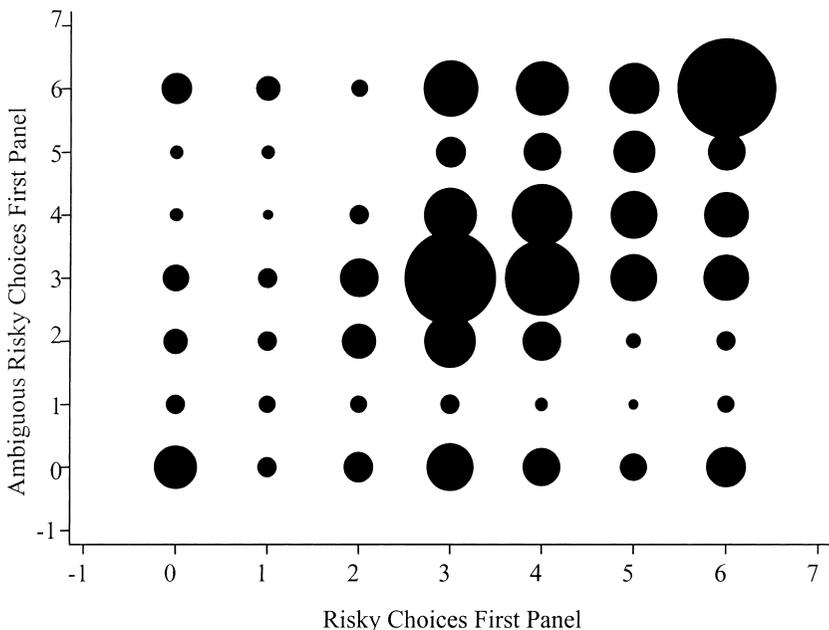


Fig. 2.17 Ambiguity and lottery choice

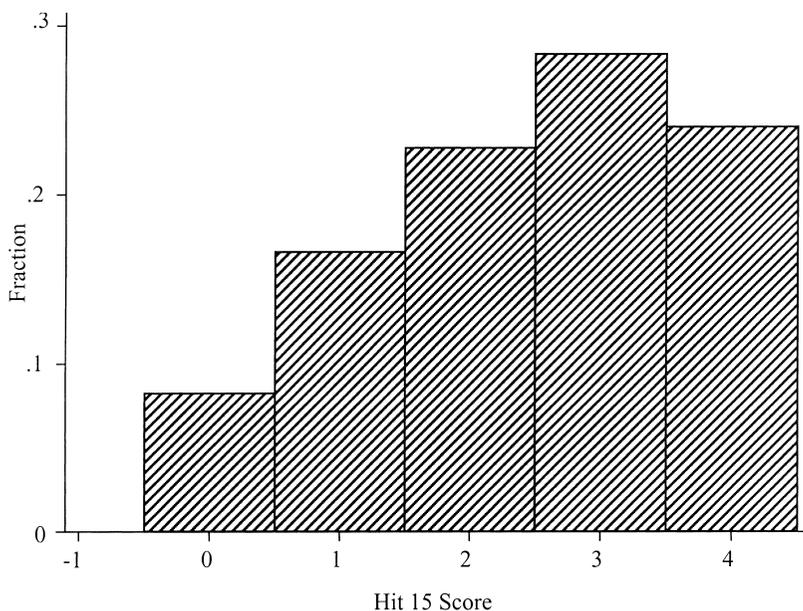


Fig. 2.18 Success in the Hit 15 Points game

set so as to give the subjects an example of how the first stage of backward induction works. Before each decision is made by the subject as to what number of points they want to add to the points basket, the subject is asked whether he or she will win or if the computer will win. The subjects are paid \$1 for each round that they win.

Figure 2.18 shows the distribution of scores on this task. The median falls between two and three correct rounds of the game.⁵³ Together with numeracy, we anticipate that this measure will be predictive of on-the-job success, and especially of high versus low productivity. The TL truck drivers have to do numerical backward induction every day, to calculate back from routing and delivery goals that are from a few hours to a few days ahead, in order to decide on the optimal course of action in the present. One regularly used application is in figuring out how many hours they need to get to their destination and how many hours they have available to drive under the regulations governing the hours of service for commercial vehicle operators.

Risk, Impatience, and Cooperation Survey

The last instrument used during the Saturday initial data collection events with the panel study participants is a list of attitude questions about

53. As mentioned earlier, the N is 893 for this task, as we discovered a programming error in the z-Tree code for the game after 176 subjects had already taken part.

risk preferences, patience and impatience, and cooperation. It was assembled from a selection of papers in the literature that propose various survey-response measures for these characteristics of subjects. We intend these to also be tried out as control variables in order to see if they are useful and, if so, whether they are substitutes or complements to the experimental measures in predicting behavior on other measures and on-the-job success.

2.6.6 Other Measures of Interest

The applicant information collected by the cooperating firm on trainees provides a few other data items of interest. There is an indicator of whether the participant is a smoker (whether the trainee requested a hotel room in which smoking is permitted during initial training). This may turn out to be related to risk attitudes. In addition, because the trainees apply for credit for the training contract, the firm is able to supply us with a credit score. We want to examine the relationships among our measures of risk/loss aversion, ambiguity aversion, time preference, and impatience, on the one hand, and the credit score, on the other. We also are interested in determining whether the credit score useful in predicting on-the-job success, and, if so, whether it is a complement or substitute for our other measures.

2.7 Some Implications of the Truckers and Turnover Project

The present chapter has presented the context and design of the Truckers and Turnover Project at a stage when much of the initial data collection has taken place, but while follow-up data is still being collected and before results have become available. The proof of this pudding will, of course, be in the eating, when data collection is complete and results are available. But there are three points that can be made about the project at this stage. The first two are about research methodology, and the third is about the policy significance of the subjects of the project.

2.7.1 For the Industry Studies Community

First, for the industry studies research community, the project stands as a kind of “proof of concept” for adding behavioral economic experiments to the already rich mix of data collection techniques it utilizes. The Sloan Foundation’s Web site states that the Foundation’s goal in supporting industry studies is “to build a strong community of industry studies scholars who contribute to academia, to industries, and to government.” Along with an interdisciplinary approach, the defining feature of this work according to the Foundation is that “industry studies scholars develop a deep understanding of particular industries by grounding their research in direct observation” (2007). The conventional tools for direct observation range from in-depth participant observation, to shorter structured site visits and interviews, to the collection of proprietary internal data for statistical analysis. To this list, the Truckers and Turnover Project prospectively

adds behavioral economic field experiments. The key, in addition to having an existing cooperative research relationship between academics and firm(s), is to design the behavioral experiments to fit the specific institutional and business context. The new data-collection techniques must both complement more traditional approaches and also must credibly add value, from both an academic and a business standpoint. In this regard, the essentially individualized production process of TL motor freight is especially suited for an initial exploration of the uses of experiments that measure individual characteristics of employees. But one can easily imagine more complex social dilemma experiments being applied in teamwork employment settings, for instance.⁵⁴

2.7.2 For the Experimental and Behavioral Economics Communities

Second, for the behavioral and experimental economics communities, the project also stands as a kind of “proof of concept” for the specific manner of applying behavioral experiments to the workplace used. As mentioned in section 2.6.1, there are a wide variety of scientific motives for conducting specific types of economic experiments, but one large cross-cutting categorization is whether the primary interest is in treatment effects or in measuring the characteristics of individual subjects. The latter is the purpose here, in keeping with the goal mentioned in the preceding paragraph about targeting the design to the specific institutional context.

The laboratory and the field are both relevant sources of empirical data to be used in the generation and evaluation of economic theory. But it is often difficult to translate the import of findings from studies in the fully controlled, abstract, and simplified setting of the lab directly to the uncontrolled and complex world of real employment relations and production processes. The approach here is close to what Harrison and List (2004) call a “framed field experiment,” in which the experiments are similar to those that might be done in a lab with student subjects, but are instead conducted in the field setting of the driver training school, with new driver-trainees as subjects so that the framing for the (relatively abstract) experimental tasks is provided by these contextual factors. We anticipate finding some new empirical regularities when we examine the ability of our experiments to predict on-the-job outcomes, and, if so, we will have provided significant evidence for their “external validity,” that is, for the ability of the lab measurements we use to capture individual characteristics that are important in practice in a specific real workplace setting.

2.7.3 The Larger Significance of Our Subject Pool

The last point we want to make is about the economic and human importance of the subjects of our study, driver-trainees in the truckload seg-

54. For some promising work along these lines, see the series of papers on fruit pickers in England by Bandiera, Barankay, and Rasul (2005, 2006).

ment of the motor freight industry. Much has been made in the last decade of the growth of the knowledge economy and of the role of workers who have the specific education and skills to prosper in it. But what about those who do not fit this description? In 2005, fully 40 percent of the U.S. workforce had a high school degree or less, and only 29 percent had a four-year college degree or more (Mishel, Bernstein, and Allegretto 2006, 153). As the United States continues its transition from a manufacturing economy to one in which service occupations and service industries dominate total employment, nonknowledge service workers will become an increasingly important part of the overall picture. The job of TL tractor-trailer driver is an archetypal example of nonknowledge work that is important in the knowledge economy.

The United States is a geographically large economy, so transportation and distribution have always been a key complement to U.S.-based production. And as the international trade in goods looms larger and larger in U.S. production and consumption, the more essential become the tasks associated with physically transporting traded goods and their ingredients here in the United States. Unlike some other types of services, the tasks involved in physically moving freight here are not subject to direct competition from abroad. According to the 2002 quinquennial Economic Census, trucking is the largest single industry in transportation services by employment, and the truckload segment of trucking had a headcount of approximately 800,000. This means that about 600,000 persons are TL tractor-trailer drivers at any one time.⁵⁵ Yet the ATA's survey results say that the typical turnover rate at TL firms is at least 50 percent per year.⁵⁶ If these figures are taken at face value, it means that several hundred thousand people train for and try out this job each year, only to leave it within a few months, probably having incurred a significant debt for training that most have little hope of repaying. Our pilot findings on turnover at the co-operating firm do not contradict this interpretation of the industry facts.

In this context a primary goal of the Truckers and Turnover Project is to use the best available means to measure the individual characteristics that make for success in this job. We will then ask whether the findings have actionable implications for business strategy and for public policy with respect to the labor market and education. Clearly, improving the match between job and employee by even a small amount would be a real improvement in economic circumstances for both firms and employees. It may even be that our findings will help in understanding how and to what extent the nature of these jobs can be improved, by quantifying more clearly the costs and benefits of lowering turnover.

55. This is by the rule of thumb that drivers make up about 75 percent of employment at a typical TL firm

56. The rate for large carriers typically averages over 100 percent, but most firms are small, and the rates for this group are usually between 50 percent and 100 percent per year.

Appendix A

The Truckers and Turnover Project Team

Project Organizer:

1. Stephen Burks, Division of Social Science, University of Minnesota, Morris; IZA; and Trucking Industry Program (Georgia Institute of Technology)

Project Coinvestigators (each colleague is collaborating on at least one component of the project, some on multiple components):

1. Jon Anderson, Division of Science and Math, University of Minnesota, Morris
2. Jeffrey Carpenter, Department of Economics, Middlebury College; IZA; and Norms and Preferences Research Network
3. Andrew Clark, Centre National de la Recherche Scientifique, PSE (France), and IZA
4. Lorenz Götte, Research Center for Behavioral Economics and Decisionmaking, Federal Reserve Bank of Boston, and IZA
5. Aldo Rustichini, Department of Economics, University of Minnesota, Twin Cities
6. Kristen Monaco, Department of Economics, California State University at Long Beach; and Trucking Industry Program (Georgia Institute of Technology)

On-site Research Team Members 2005 to 2006:

1. Kay Porter, Business Research Manager, Cooperating Firm
2. Adam Durand, Research Intern (fall 2005), University of Minnesota, Morris
3. William Leuthner, Research Intern (spring 2006), University of Minnesota, Morris
4. Erin Christenson, Research Intern (summer 2006), University of Minnesota, Morris

Appendix B

Summary List and Time Line for Data Items Collected

1. Project Component One: Statistical Case Study

Data items here come in separate data report files, each of which has different data definitions and data errors. These are joined into a master data file by the project team, in a labor-intensive data validation, cleaning, documentation, and merging process. Initial data set covers January 1, 2002, through December 31, 2005, by construction from the individual years. Updates are planned on semiannual or quarterly basis.

- 1.1. Payroll-based data. Rolled-up totals of activities for which drivers are paid, by driver and week. Some pay deductions, but not all that would be of economic interest, are included. Typical items of interest include paid miles, the number of units of mileage pay (approximately the number of dispatches), and the number of times supplementary pay was received for particular work event (examples: hand unloading, weighing the rig on a scale). Includes operations-related driver characteristics (e.g., the terminal at which the driver is based). Not included in the initial version: the identity of the direct supervisor. Created by project team from pay transaction files.
- 1.2. Demographic data. Marital status, gender, and (for those drivers choosing to self-identify to the firm) racial category. Not included, as not captured electronically by firm: educational attainment. Birth-dates are automatically included only in 2006 updates and later (see item 3.4).
- 1.3. Weekly list of drivers employed, with a limited set of driver characteristics. (Used as a check on payroll-based data because payment is not always made in the week activities took place.)
- 1.4. Weekly list of drivers hired in that week, with a limited set of driver characteristics.
- 1.5. Weekly list of drivers who separated in that week, with a limited set of driver characteristics.
- 1.6. Application records. Variables from the application process for those for whom a hire event is observed; has significant number of missing values. Typical items of interest: did applicant request a smoking hotel room when attending the training school, and how did applicant learn about the training opportunity?

2. Project Component Two: Panel Study of New Hires

- 2.1. Initial Intake Data on Driver Trainees. Collected from December 12, 2006, through August 8, 2007, on twenty-three Saturdays. 1,069

trainees entered the study, and there is good data on 1,036 of them, except that a bug in the initial version of Hit 15 Points (item 2.1.2.5) limits good data that includes this item to a subset of 893 persons.

2.1.1. First Data Collection Block.

- 2.1.1.1. Sequential strategic form Prisoner's Dilemma experiment. Choices as first and second mover, plus estimates of others' choices are collected.
- 2.1.1.2. Multidimensional Personality Questionnaire. Standard (short form of) instrument identifying eleven personality factors.
- 2.1.1.3. Risk/Loss Aversion experiment. Four panels of seven choices, each between a small monetary gamble and a small fixed sum.
- 2.1.1.4. Demographic profile. List of more than fifty questions compiled by project team, including information on educational attainment, labor market history, next best employment opportunity, and so on.
- 2.1.1.5. Red Button experiment. Subjects offered opportunity to leave for break early at a payoff cost.

2.1.2. Second Data Collection Block.

- 2.1.2.1. Time Preferences experiment. Four panels of six choices, each between a smaller amount at an earlier time and a larger amount at a later time. Two panels include "today" as the earlier time, and two panels have a front-end delay, with "Monday" (two days after the Saturday data collection) as the earlier time.
- 2.1.2.2. Nonverbal IQ: Sections two, three, and four of the original five sections of the Raven's Progressive Matrices. Participants are asked to rate their relative performance before and after the instrument, as well as whether they want to know their score and the group average upon receiving their final payout.
- 2.1.2.3. Quantitative Literacy: Section one of Form A of the Quantitative Literacy portion of the Educational Testing Service's (ETS) Tests of Applied Literacy Skills. (ETS provides standardized scoring only for subjects that do both sections of this test.) Participants are asked to rate their relative performance before and after the instrument, as well as whether they want to know their score and the group average upon receiving their final payout.
- 2.1.2.4. Ambiguity Aversion experiment. Repetition of Risk/Loss Aversion, except that we provide less information about the probability distribution in the gamble choices.
- 2.1.2.5. Hit 15 Points backward induction experiment.
- 2.1.2.6. Miscellaneous questions from the psychology and behavioral economics literature about impatience, risk, cooperation, and so on.

- 2.2. Follow-up Data from Trainees and their Families. The follow-up data collection period runs through August of 2008.
 - 2.2.1. Human Resources Data from Firm.
 - 2.2.1.1. Subset of Applicant Flow data, including whether trainee requested a smoking room at training.
 - 2.2.1.2. Credit score of the trainee (available because trainees sign a credit contract for training costs).
 - 2.2.1.3. If and when a driver exits, the exit date and a code for up to three reasons for the separation.
 - 2.2.2. Mail Surveys to New-Hire Panel Drivers.
 - 2.2.2.1. Continuing driver surveys. Forty-five questions about current opinions and recent work events, including some open-ended ones, sent to currently employed drivers between two and four weeks after the following tenure milestones: six months, twelve months, eighteen months, and twenty-four months.
 - 2.2.2.2. Exited driver surveys. Forty-four questions about current opinions, recent work events, and the driver's exit event, sent between two and four weeks after exit.
 - 2.2.3. Mail Surveys to Families of Trainees.
 - 2.2.3.1. Initial survey to families of new-hire panel drivers, to approximately match in timing the initial intake information from drivers. Not actually started until the beginning of April of 2006, at which time the families of all panel participants who had entered since December 2005, were surveyed. Thirty-five questions developed by project team, including some open-ended ones.
 - 2.2.3.2. Surveys to families of continuing drivers. Thirty-five questions about current opinions and recent family-work-life events, including some open-ended ones, sent to families of currently employed drivers between two and four weeks after the following tenure milestones: six months, twelve months, eighteen months, and twenty-four months.
 - 2.2.3.3. Surveys to families of exited drivers. Twenty-nine questions about current opinions, recent family-work-life events, and the driver's exit event, sent between two and four weeks after exit.
 - 2.2.4. Weekly two-question survey to currently employed drivers, in their trucks, via satellite link, on Wednesdays. Question 1: How satisfied are you with your job right now (1–5 Likert response scale)? Question 2: How many pay miles do you expect to run next week? These data will be documented, cleaned, and merged with the main panel study data set, at some point in the future.
 - 2.2.5. Time-at-home records. The firm's recording of when drivers request to be at home, and how well the firm does at fulfilling those requests, is accumulating in a data file that starts in early 2006, and

is expected to be updated through August 2008. These data will be documented, cleaned, and merged with the main panel study data set at some point in the future.

3. Control Groups for New Hire Panel Study

- 3.1. The initial intake data collected on driver trainees (all the items in 2.1) was also collected from 100 undergraduate students recruited at the University of Minnesota, Morris, in a campus computer lab, during the spring semester of 2007. The same protocol was used throughout, with the exception that the demographic profile questions were modified to be appropriate for student subjects.
- 3.2. The initial intake data collected on driver trainees (all the items in 2.1) is also planned to be collected from 100 adult subjects who are residents of Stevens County, MN, the rural county in which the University of Minnesota, Morris, campus is situated. The plan is to recruit subjects from the county who are not students, nor faculty, nor professional staff, during the summer and fall semesters of 2007. The same protocol will be used throughout, with the exception that the demographic profile questions will be modified to be appropriate for these subjects.

4. Miscellaneous Supplementary Data

- 4.1. Records of use of the satellite link in each tractor to report trailers needing maintenance. Available on all drivers by week from mid-2006; to be updated periodically. May be usable to generate an index of on-the-job cooperation with other drivers.⁵⁷
- 4.2. Records of the use of the satellite link in each tractor to respond to a particular message from the firm's central dispatch office. Available on all drivers by week from mid-2006, to be updated periodically. May be usable to generate an index of on-the-job cooperation with the firm.⁵⁸
- 4.3. Identification of direct supervisors by driver and work week. This

57. When a driver is directed to hook onto a trailer spotted at a customer location but the trailer turns out to be mechanically faulty and can't be used, it is a negative job event. This happens when the driver who previously had the trailer did not report it as bad, either because he did not realize it was or because he refused to take the time to do so. So when a driver does take the time to report as mechanically faulty a trailer he is dropping and will not immediately use further, thereby delaying the start of his next assignment, he can be said to incur a personal cost to provide a public good for drivers. The main issue in constructing an index is limiting confounds with unobservable factors, such as whether the driver is reporting the trailer as bad because he needs it repaired for his own immediate use.

58. Responding to the macro from the central office is a formal, but not a substantive, job requirement. Responding primarily helps anonymous staff members at the firm's central office, so if a driver delays proceeding with his current assignment to respond, it can be con-

data is potentially available in a supplementary file that may permit the historical data set used for the statistical case study to be updated. Updates from 2006 and later will include this information in data files already being amalgamated to make the statistical case study data set.

- 4.4. Birthdates. These are in the standard demographic data files on all drivers starting in 2006. Birthdates for drivers observed in earlier years reside in a supplementary file that must be documented and cleaned and merged with the 2002 to 2005 statistical case study data set.

5. Merging New Hire Panel and Statistical Case Study Data.

Beginning in the second half of 2008, when the statistical case study data files have been updated to include all of 2007 and at least some of 2008, it is intended to merge the two data sets together for those drivers who are in both data files. This will permit the regression models developed for the statistical case study to be run on the panel study drivers, with the ability to test what additional predictive power is added to those models by the new data items collected in the new hire panel portion of the study.

References

- Alfred P. Sloan Foundation. 2007. Industry studies. *Programs: Standard of living and economic performance*. http://www.sloan.org/programs/pg_standard.shtml.
- ATA Economic and Statistics Group. 2005. Employee turnover and workforce changes: Q4:2004. *Trucking Activity Report* 13:7.
- Bandiera, O., I. Barankay, and I. Rasul. 2005. Social preferences and the response to incentives: Evidence from personnel data. *Quarterly Journal of Economics* 120 (3): 917–62.
- . 2006. The evolution of cooperative norms: Evidence from a natural field experiment. *B.E. Journals in Economic Analysis and Policy: Advances in Economic Analysis and Policy* 6 (2): 1–26.
- Beadle, A. D. 2004. Building driver shortage. *Traffic World* 268 (42): 26–27.
- Belman, D. L. and K. Monaco. 2001. The effects of deregulation, de-unionization, technology, and human capital on the work and work lives of truck drivers. *Industrial and Labor Relations Review* 54 (2A): 502–24.
- Belman, D. L., K. Monaco, and T. Brooks. 2005. *Sailors of the concrete sea: A portrait of truck drivers' work and lives*. Lansing, MI: Michigan State University Press.
- Belzer, M. H. 1995. Collective bargaining after deregulation: Do the teamsters still count? *Industrial and Labor Relations Review* 48 (4): 636–55.
- Belzer, M. H., D. Rodriguez, and S. Sedo. 2002. Paying for safety: An economic

sidered a personally costly form of cooperation with management. The main issue in constructing an index is limiting confounds with unobservable factors, such as with variations in the level of exposure to such macro requests.

- analysis of the effect of compensation on truck driver safety. FMCSA Report. Washington, DC: Federal Motor Carrier Safety Administration.
- Boyer, K., and S. V. Burks. 2007. Stuck in the slow lane: Traffic composition and the measurement of labor productivity in the U.S. trucking industry. IZA Discussion Paper no. 2576. Bonn, Germany: Institute for the Study of Labor.
- Bureau of Labor Statistics. 2002. Historical hours and earnings—Average hours and earnings of production or non-supervisory workers on private non-farm payrolls by major industry sector, 1964 to date. U.S. Department of Labor. ftp://146.142.4.23/pub/suppl/empst.cese2.txt.
- Burks, S. V. 1999. The origins of parallel segmented labor and product markets: A reciprocity-based agency model with an application to motor freight. PhD diss., University of Massachusetts, Amherst.
- Burks, S. V., J. Carpenter, and L. Götte. 2006. Performance pay and the erosion of worker cooperation: Field experimental evidence. IZA Discussion Paper no. 2013. Bonn, Germany: Institute for the Study of Labor.
- Burks, S. V., K. Monaco, and J. Myers-Kuykindall. 2004a. The balance between private and for-hire carriage and trends in the use of large trucks (1977 to 1997). *Journal of the Transportation Research Forum* 43 (2): 159–71.
- . 2004b. Trends in the use of large trucks by truckload and less-than-truckload motor carriers in the 1990s. *Journal of the Transportation Research Forum* 43 (2): 173–87.
- Cain, G. G. 1976. The challenge of segmented labor market theories to orthodox theory: A survey. *Journal of Economic Literature* 11 (4): 1215–57.
- Camerer, C., and E. Fehr. 2004. Measuring social norms and preferences using experimental games: A guide for social scientists. In *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, ed. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis, 55–95. Oxford: Oxford University Press.
- Casey, J. F. 1987. An assessment of the truck driver shortage. *Transportation Executive Update* 1 (1): 20–23.
- Cleves, M. A., W. A. Gould, and R. G. Gutierrez. 2004. *An introduction to survival analysis using Stata, revised edition*. College Station, TX: Stata Press.
- Corsi, T., and J. R. Stowers. 1991. Effects of a deregulated environment on motor carriers: A systematic multi-segment analysis. *Transportation Journal* 30 (3): 4–28.
- Cox, K. 2004. TL driver turnover rate reaches new record high. *Transport Topics* (January 5):5, 35.
- Cullen, D. 1996. You get what you pay for: J. B. Hunt Transport Services Inc.'s driver compensation package. *Fleet Owner* 91 (11): 82–83.
- Dickens, W. T., and K. Lang. 1993. Labor market segmentation theory: Reconsidering the evidence. In *Labor economics: Problems in analyzing labor markets*, ed. W. Darity, Jr., 141–80. Norwell, MA: Kluwer Academic.
- Fischbacher, U. 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, forthcoming.
- Gallup Organization. 1997. Empty seats and musical chairs: Critical success factors in truck driver retention. Arlington, VA: ATA Foundation.
- Global Insight, Inc. 2005. *The U.S. truck driver shortage: Analysis and forecasts*. Arlington, VA: American Trucking Associations.
- Griffin, G., and L. Kalnbach. 2002. Predicting and classifying voluntary turnover decisions for truckload drivers. Report no. MPC-02-135. Fargo, ND: Upper Great Plains Transportation Institute, North Dakota State University.
- Griffin, G. C., J. M. Rodriguez, and B. M. Lantz. 1992. Job satisfaction in U.S.

- commercial drivers. Report no. DP-90. Fargo, ND: The Upper Great Plains Transportation Institute, North Dakota State University.
- Harrison, G. W., and J. A. List. 2004. Field experiments. *Journal of Economic Literature* 42 (4): 1009–55.
- Isidore, C. 1997. Hunt gambles on pay raise to keep drivers. *Journal of Commerce* 411 (28917): 1A.
- Kiefer, N. M. 1988. Economic duration data and hazard functions. *Journal of Economic Literature* 26:646–79.
- Mishel, L., J. Bernstein, and S. Allegretto. 2006. *The state of working America, 2006–2007*. Ithaca, NY: ILR Press.
- Patrick, C. J., J. J. Curtin, and A. Tellegan. 2002. Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment* 14 (2): 150–63.
- Raven, J., J. C. Raven, and J. H. Court. 2000. Section 3: The Standard Progressive Matrices. In *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Psychological Corporation.
- Reiskin, J. 2006. Shipper changes cut distances that truckload freight travels. *Transport Topics* (March 13):1, 49.
- Rodriguez, J., M. Kosir, B. Lantz, G. Griffin, and J. Glatt. 2000. The costs of truckload driver turnover. Report no. SP-146. Fargo, ND: Upper Great Plains Transportation Institute, North Dakota State University. 1–18.
- Stephenson, F. J., and R. J. Fox. 1996. Driver retention solutions: Strategies for for-hire truckload (TL) employee drivers. *Transportation Journal* 35 (4): 12–26.
- U.S. Census Bureau. 2004. *Truck transportation: 2002*. Washington, DC: Economics and Statistics Administration, U.S. Department of Commerce.
- Waxler, C. 1997. J. B. Hunt Transport shares may drop further. *Forbes* 159 (8): 382–83.