"Worker Overconfidence: Field Evidence and Implications for Employee Turnover and Firm Profits": Online Appendix

Mitchell Hoffman and Stephen Burks

The Online Appendix is organized as follows. Appendix A provides additional discussion and analysis on belief survey non-response, as well as on other issues. Appendix B provides more information on the Firm B field experiment. Appendix C uses a one-period version of the structural model to briefly show that differential overconfidence lowers the probability of quitting. Appendix D provides omitted derivations from the structural model. Appendices E-F collect additional figures and tables. Appendix G provides further discussion on measuring productivity.

For brevity in typesetting, Appendices D-F appear below. A complete version of the Online Appendix containing all Appendices and the Appendix References appears within the replication package.

D Structural Model and Estimation Details

Estimation Sample. For the sample for the structural analysis, we start with our baseline data subset sample of 895 drivers. Next, we drop any drivers who are ever seen working at non-piece rate trucking jobs at Firm A where they are paid based on their activities or on salary (e.g., this drops drivers who ever go to work themselves as driver trainers at the training schools). We also drop a small number of drivers with a missing individual characteristic, leaving an estimation sample of 699 drivers.

Probability of Staying. Let $\Lambda(x) = \frac{exp(x)}{1+exp(x)}$ and let fixed non-pecuniary taste for the job be $\alpha + X\overline{\alpha}$, where α has a mass point distribution and $\overline{\alpha}$ are the utility coefficients associated with different worker characteristics. At time T, the probability of staying, given the state variables, is:[27]

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

To evaluate this probability, we need to calculate both $E^b(y_T|y_1,...,y_{T-1})$ and $E^b(V(\mathbf{x})|\mathbf{x}_T)$. The former depends on $y_1,...,y_{T-1}$, which would imply that the state space has dimensionality of order K^{T-1} when y_t is discretized with K values. The key to avoiding a very high dimensional problem is that in a normal learning model (both a model with standard beliefs and our generalized learning model), the worker's expectation of future productivity depends only on his prior and his de-trended average of past productivity. That is, *the average of past productivity is a sufficient statistic for the sequence* $y_1,...,y_{t-1}$ (DeGroot, 1970).

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

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

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

$$\begin{aligned} E^b(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) &= E^b_{y_t} E^b_{\varepsilon|y_t}(V_{t+1}(\mathbf{x}_{t+1})|\mathbf{x}_t) && (9) \\ &= E^b_{y_t} E^b_{\varepsilon}(max\{\overline{V}^S_{t+1}(\mathbf{x}_{t+1}) + \varepsilon^S_{t+1}, \overline{V}^Q_{t+1} + \varepsilon^Q_{t+1}\}|\mathbf{x}_t) && (10) \\ &= \int \tau \log\left(\exp(\frac{\overline{V}^S_{t+1}(\mathbf{x}_{t+1})}{\tau}) + \exp(\frac{\overline{V}^Q_{t+1}}{\tau})\right) f^b(y_t|y_1,...,y_{t-1})dy_t && (11) \\ &= \sum_k \tau \log\left(\exp(\frac{\overline{V}^S_{t+1}(\mathbf{x}_{t+1})}{\tau}) + \exp(\frac{\overline{V}^Q_{t+1}}{\tau})\right) P^b(y_t^k|y_1,...,y_{t-1}) && (12) \end{aligned}$$

(9) expresses that the value function depends on future miles and future idiosyncratic shocks. (10) uses the definition of V, that V_s^Q is independent of the state variable, and that the idiosyncratic shocks are independent of miles. (11) integrates out y_t, which are not observed when the driver makes his period t decision, but are observed in the future, as well as integrates out the idiosyncratic shocks. (12) follows because, in implementation, miles will be discretized into K possible values.

[27] After time T, quitting is governed by the asymptotic value functions in (5) using $E^b(\cdot)$ instead of $E(\cdot)$.

The perceived transition probability $P^b(y_t^k|y_1,...,y_{t-1})$ can be easily shown to depend only on $\overline{y}_{t-1}$, and is expressed below in (19). Related derivations can be found in Rust (1987) and Stange (2012).

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

$$
\begin{aligned}
L_i &= \int L(d_{i1},...,d_{it},y_{i1},...,y_{it},b_{i1},...,b_{it}|\alpha,\eta_b) f(\alpha,\eta_b)\, d\alpha d\eta_b && (13)\\
&= \int \{L(d_{i1},...,d_{it}|y_{i1},...,y_{it},b_{i1},...,b_{it},\alpha,\eta_b) * L(b_{i1},...,b_{it}|y_{i1},...,y_{it},\alpha,\eta_b) && (14)\\
&\quad * L(y_{i1},...,y_{it}|\alpha,\eta_b) f(\alpha,\eta_b)\, d\alpha d\eta_b\} \\
&= \left[\int L(d_{i1},...,d_{it}|y_{i1},...,y_{it},\alpha,\eta_b) * L(b_{i1},...,b_{it}|y_{i1},...,y_{it},\eta_b) f(\alpha,\eta_b)\, d\alpha d\eta_b\right] L(y_{i1},...,y_{it}) && (15)\\
&= \left[\int \prod_{s=1}^{t} L(d_{is}|d_{i1},...,d_{is-1},y_{i1},...,y_{it},\alpha,\eta_b) * \prod_{s=1}^{t} L(b_{is}|b_{i1},...,b_{is-1},y_{i1},...,y_{it},\eta_b) f(\alpha,\eta_b)\, d\alpha d\eta_b\right] && (16)\\
&\quad * L(y_{i1},...,y_{it}) \\
&= \left[\int \prod_{s=1}^{t} L(d_{is}|y_{i1},...,y_{is-1},\alpha,\eta_b) \prod_{s=1}^{t} L(b_{is}|y_{i1}...,y_{is-1},\eta_b) f(\alpha,\eta_b)\, d\alpha d\eta_b\right] \left(\prod_{s=1}^{t} L(y_{is}|y_{i1}...,y_{is-1})\right) && (17)\\
&\equiv \left[\int L_i^1(\alpha,\eta_b) L_i^3(\eta_b) f(\alpha,\eta_b)\, d\alpha d\eta_b\right] L_i^2 && (18)
\end{aligned}
$$

Equations (13) and (14) follow by the law of total probability. (15) holds because productivity is unaffected by the taste and overconfidence heterogeneity and because beliefs are unaffected by the taste heterogeneity. (16) follows because (a) future miles are not observed when a worker decides to quit and (b) quit decisions are independent of reported subjective beliefs conditional on the overconfidence unobserved heterogeneity. (17) follows because (a) since the ε shocks are iid, the decision to quit is conditionally independent of all prior decisions to quit (given the miles realizations and the unobserved heterogeneity) and (b) reported subjective beliefs are conditionally independent of past reported subjective beliefs conditional on productivity and the belief heterogeneity. In (18), we define the part of the likelihood due to the quitting decisions as $L_i^1(\alpha,\eta_b)$, the part due to the miles realizations as L_i^2, and the part due to subjective beliefs as $L_i^3(\eta_b)$.

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

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

with

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

where $\gamma_s = \frac{s\sigma_0^2}{s\sigma_0^2+\sigma_y^2}$, $\Omega_s = \frac{\sigma_0^2\sigma_y^2}{s\sigma_0^2+\sigma_y^2} + \sigma_y^2$, and $\gamma_s^b = \frac{s\sigma_0^2}{s\sigma_0^2+\widetilde{\sigma_y}^2}$.[28]

The overall likelihood is computed, first, by integrating over the unobserved heterogeneity for each individual's likelihood, and then by taking the product over all people. Since the unobserved heterogeneity is mass-point distributed, the integral becomes a sum.

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

Perceived Transitions between Miles. In solving the dynamic programming problem, expected future mileage is governed by a perceived transition matrix. As mentioned above, we discretize productivity into K values. In our baseline estimation, we let productivity range in increments of 300 from 100 to 4,000 miles per week (that is, $K = 14$). Perceived transitions between miles states are given by:

$$
P^b(y_s^k|y_1, ..., y_{s-1}) = \Phi\left(\frac{y_s^k + .5 * kstep - E^b(y_s^k|y_1, ..., y_{s-1})}{\sqrt{\Omega_{s-1}^b}}\right) - \Phi\left(\frac{y_s^k - .5 * kstep - E^b(y_s^k|y_1, ..., y_{s-1})}{\sqrt{\Omega_{s-1}^b}}\right) \quad (19)
$$

where $\Omega_s^b = \frac{\sigma_0^2\widetilde{\sigma_y}^2}{s\sigma_0^2+\widetilde{\sigma_y}^2} + \widetilde{\sigma_y}^2$, y_s^k is the value of y_s at the kth grid point, and where $kstep$ is the distance between grid points. See Stange (2012) for a similar formula. Our estimates are similar using a finer grid with increments of 100 from 100 to 4,000 miles ($K = 40$) as seen in column 7 of Table F1. We can also derive perceived transition probabilities between levels of average productivity to date.

$$
\begin{aligned}
P^b(\overline{y}_s^k|\overline{y}_{s-1}) = \Phi&\left(\frac{s\left(\overline{y}_s^k + .5 * kstep\right) - (s-1)\,\overline{y}_{s-1} - E^b\left(y_s|y_1, ..., y_{s-1}\right)}{\sqrt{\Omega_{s-1}^b}}\right) \\
-\Phi&\left(\frac{s\left(\overline{y}_s^k - .5 * kstep\right) - (s-1)\,\overline{y}_{s-1} - E^b\left(y_s|y_1, ..., y_{s-1}\right)}{\sqrt{\Omega_{s-1}^b}}\right)
\end{aligned}
$$

For the parts of the likelihood on miles (L_i^2) or on subjective beliefs (L_i^3), we use the mileage data in continuous form instead of discretized.[29]

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

[29] That is, we assume that perceived transition probabilities are based on miles in a discrete form, whereas actual miles and beliefs are not. This can be justified on the grounds that perceived transition probabilities may be conceptually difficult for drivers, and may be naturally thought of according to a discrete grid.

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

χ^2 **test.** χ^2 is calculated as $\sum_t$ (the number of drivers at risk during week of tenure t) * [(actual quit hazard(t) - predicted quiz hazard(t))^2 / predicted quiz hazard(t)].

12-Month Contract in Structural Model. The quit penalties under the training contracts varied slightly by training school at the firm. Furthermore, if drivers could not pay the money owed upon a quit, a significant interest rate may also have been assessed. For the structural estimation, we assume a penalty of $3,750 for the 12-month contract.

Zero Mile Weeks. The data contain a significant number of zero mile weeks for drivers. These often reflect weeks where the driver is not working. These weeks are not counted toward the miles component of the likelihood, and average miles to date (in terms of the quit decision) is given by the prior week's average miles to date. Also excluded from the likelihood are a small number of driver-weeks with predictions of 0 miles (estimates are similar when they are instead included).

Compensation and Additional Bonuses. At Firm A, Drivers may receive small quarterly bonuses (based on customer/shipper satisfaction, good fuel economy, and other factors).[30] In addition, for low-mileage loads, drivers may receive "premiums" in cents per mile above their regular cents per mile. For computational simplicity, we ignore all bonuses and premiums in our analysis. Further, at some points in the past, the firm has provided a guaranteed minimum earnings level for new inexperienced drivers when starting out (e.g., up through week 12), and we ignore this as well. For the piece rate-tenure profile in the structural model, we use data from an internal firm document in 2004. It provides the profile for the region where the training school in the data subset is located. We use the profile for the most common work type. Although pay per mile continues to increase after three years of tenure, for simplicity in our model, we assume that pay per mile for three years and after is the same as the rate before the three-year mark.

Maximization. For the inner loop in the Rust (1987) procedure, we use a tight tolerance of 1e-15. For our baseline estimates, we maximize the likelihood function using "fminunc" in Matlab. We use a Quasi-Newton algorithm and a function tolerance of 2e-5. We verified that another Matlab algorithm, "fminsearch" (Nelder-Mead), yields convergence to the same parameter estimates. For the estimates with learning by doing, we first maximize using "fminunc" and then use the estimates as starting values for performing "fminsearch" (doing "fminsearch" with a function tolerance of 1e-5). We perform the maximization while restricting that the levels of the taste unobserved heterogeneity mass points (α_1, α_2, α_3) are greater than or equal to -$375 for the baseline case and greater than or equal to -$2,000 for the case with learning by doing.[31]

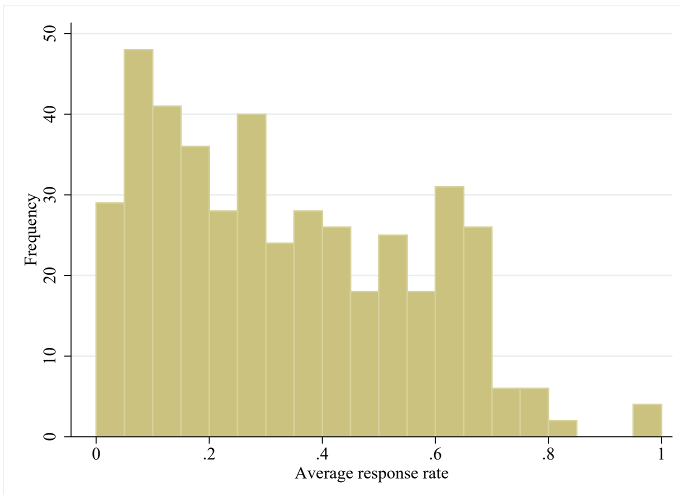
[30] At some points in the past, new inexperienced drivers only became eligible to receive a quarterly bonus after one year of tenure.

[31] If this restriction is not made, for the models with belief bias, maximization will sometimes yield parameter estimates where one of the taste mass points tends toward $-\infty$. In the baseline model (column 2 of Table 4), this can lead to a point with very low α for one unobserved type and high τ. In the model with learning by doing (column 2 of Table 5), it can lead to a point with very low α for one unobserved type and very high θ_1. We view such parameter estimates as less economically plausible, leading us to impose a restriction on the taste mass points. However, even for such parameter estimates, the key belief parameters, as well as the impact of debiasing on worker retention, firm profits, and worker welfare, are qualitatively similar to those in the main results.

Following Knittel and Metaxoglou (2014), we perform a number of checks on our optimization procedure for all our four main models in Tables 4 and 5. First, for the identified optima, we checked that our exit code indicates successful convergence; that $\|g\|_\infty$ and $g'H^{-1}g$ are small, where g is the gradient and H is the Hessian; and that H is positive-definite. Second, in line with Knittel and Metaxoglou (2014) and DellaVigna et al. (2017), for each model, we randomly generate a variety of starting values. We use uniform distributions for each parameter, drawing values over roughly economically plausible ranges. We verify that our reported parameter values yield the best fit out of the various estimates achieved.

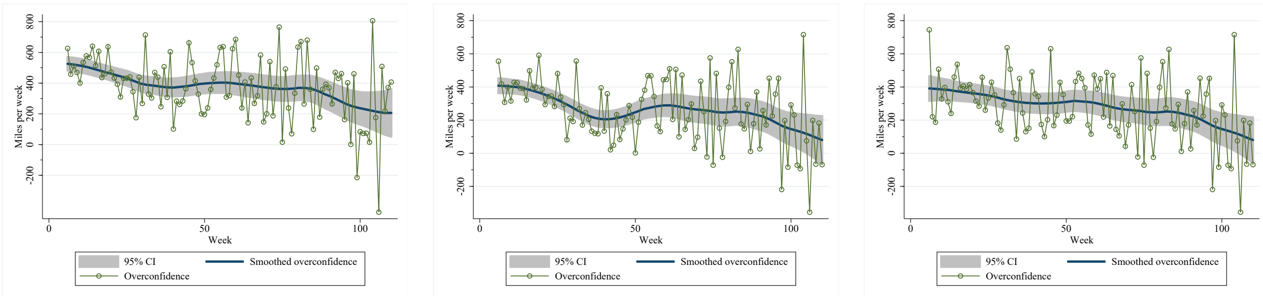
E Further Reduced-Form Results

Figure E1: Heterogeneity in Response Rates to the Firm A Subjective Productivity Beliefs Survey



Notes: This figure plots the distribution of driver-level average response rate to the survey (averaged over a driver's weeks in the data), excluding drivers who never respond. On the y-axis is the number of drivers in each bin. Observations are excluded from the sample if weekly miles are 0, if weekly predicted miles are 0, or if the driver is ever observed in the dataset receiving activity-based pay or salary pay instead of being paid by the mile.

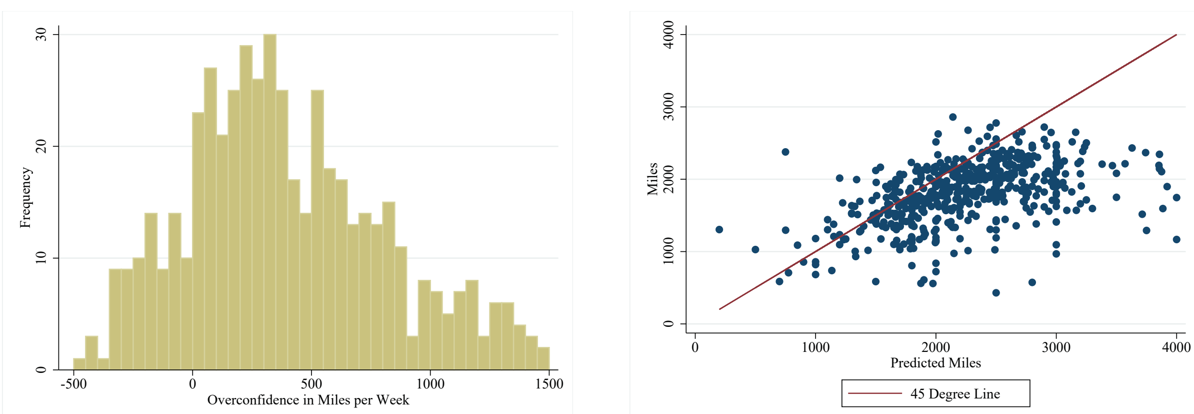
Figure E2: Tenure and Overconfidence (Productivity Beliefs *minus* Productivity)



(a) Means (b) Medians (c) Medians, ≥75 Wks, Responder

Notes: This figure analyzes the evolution of average driver overconfidence as a function of driver tenure. Overprediction, defined as productivity beliefs *minus* realized productivity, is collapsed (across all drivers) by week of tenure. Week t on the graph is corresponded to the driver's prediction in week t about productivity in week $t+1$, as well as to the driver's actual productivity in week $t+1$. The dots correspond to the collapsed means or medians. The smoothed curve is plotted using a local polynomial regression and a bandwidth of 7 weeks. In panel (a) beliefs minus actual productivity across drivers is collapsed into weekly means before local polynomial smoothing. In panels (b) and (c), beliefs minus actual productivity across drivers is collapsed into weekly medians before smoothing. In panel (c), we restrict to workers who stay at least 75 weeks and who respond to the beliefs survey that week. Results are similar if instead we look at workers who ever respond. Observations are excluded from the sample if weekly miles (one week ahead) are 0, if weekly predicted miles are 0, or if the driver is ever observed in the dataset receiving activity-based pay or salary pay instead of being paid by the mile. By looking at overprediction (instead of productivity and beliefs separately), we restrict to realized mile observations where there is a corresponding prediction. We restrict attention to weeks of tenure between 6 and 110 (early weeks involve training and the sample becomes relatively scant after around two years).

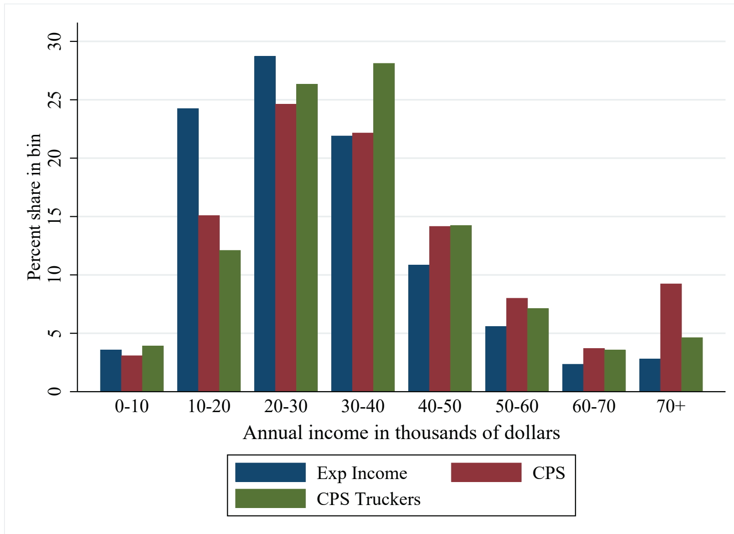
Figure E3: Distribution of Overconfidence Across Drivers



(a) Dist of Individual Overconfidence (b) Predicted Miles and Actual Miles

Notes: This figure presents reduced-form evidence on the distribution of overconfidence across drivers. Panel (a) plots a histogram of driver-level overconfidence, where overconfidence is defined as the difference between average beliefs and average productivity. In panel (b), each driver is represented by a dot located at their average productivity and average beliefs. For both panels, we calculate average productivity by averaging over all the driver's weeks (excluding weeks with 0 miles), and we calculate average beliefs by averaging over all the driver's weeks (excluding weeks with predictions of 0 miles). In panel (a), we restrict attention to drivers with driver-level overconfidence between -500 and 1,500 miles. In Panel (b), the figure is made while dropping any drivers for whom average beliefs or average productivity is greater than 4,000 miles.

Figure E4: Are Workers Overconfident About their Outside Option? A Comparison of Firm A Workers' Believed Outside Option with Earnings of Similar Workers in the CPS



Notes: This figure analyzes worker beliefs about their outside option. During driver training, workers at Firm A were asked "Which range best describes the annual earnings you would normally have expected from your usual jobs (regular and part-time together), if you had not started driver training with [Firm A], and your usual jobs had continued without interruption?" Answers were given in eight intervals: $\$0 - \$10,000$, $\$10,000 - \$20,000$, $\$20,000 - \$30,000$, $\$30,000 - \$40,000$, $\$40,000 - \$50,000$, $\$50,000 - \$60,000$, $\$60,000 - \$70,000$, $\$70,000+$. "Exp Income" is the expected income answer to this question, which is present for drivers in our data. The CPS comparison data are from the 2007 March CPS (also known as the Annual Social and Economic Supplement to the CPS). "CPS" is the income and earnings for 35-year old male workers with a high school degree who worked full-time last year and had positive income and earnings. "CPS Truckers" is the same as "CPS" except it is for "Driver/sales workers and truck drivers" ("Occ"=9130) and uses the age range of 30-40 instead of 35. *Provided we can compare our truckers to the workers in the CPS, there is no evidence that drivers overestimate their outside option.* **Further, in a weekly regression of perceived outside option in dollars on driver beliefs about their inside option in dollars & Table 3 full controls, the coefficient on beliefs about the inside option is only** 0.07 **(**$p - val = 0.16$**), suggesting that perceived inside and outside options are weakly correlated.**

Table E1: Do Productivity Beliefs Predict Quitting? Robustness Check Comparing Above-Median and Below-Median Beliefs

	(1)	(2)	(3)	(4)	(5)
Predicted miles are above their median level (0 or 1)	-0.762***			-0.689**	-0.940***
	(0.274)			(0.305)	(0.338)
Avg miles to date		-0.081***	-0.118***	-0.023	-0.078*
		(0.013)	(0.038)	(0.035)	(0.042)
Demographic Controls	No	Yes	Yes	No	Yes
Work Type Controls	No	Yes	Yes	No	Yes
Observations	8,509	33,374	8,509	8,509	8,509

Notes: This table is similar to Table 3. It differs in that the main regressor is a dummy for whether predicted miles is above its median level (instead of predicted miles in continuous form). The odds-ratios are 0.47, 0.50, and 0.39 for columns 1, 4, and 5, respectively, indicating reductions in quitting of 53%, 50%, and 61% from having above-median subjective beliefs vs. below median beliefs. * significant at 10%; ** significant at 5%; *** significant at 1%

Table E2: Do Productivity Beliefs Predict Quitting? Robustness Check with Lagged Values

	(1)	(2)	(3)	(4)	(5)	(6)
L. Predicted Miles	-0.028*	-0.036*			-0.030	-0.032
	(0.017)	(0.019)			(0.019)	(0.020)
L. Avg miles to date			-0.053***	-0.039	0.006	-0.020
			(0.013)	(0.039)	(0.034)	(0.041)
Demographic Controls	No	Yes	Yes	Yes	No	Yes
Work Type Controls	No	Yes	Yes	Yes	No	Yes
Observations	8,343	8,343	32,649	8,343	8,343	8,343

Notes: This table is a robustness check to Table 3, where predicted miles and average miles to date are lagged (instead of un-lagged). We also add an additional column, column 2, which is the column 1 specification plus additional controls. In columns 1, 2, 4, 5 and 6, the sample is restricted to observations with non-missing lagged average miles to date, positive lagged miles beliefs, and lagged miles beliefs less than or equal to 5,000 miles. * significant at 10%; ** significant at 5%; *** significant at 1%

Table E3: Do Productivity Beliefs Predict Quitting? Robustness Check with a Person's Average Subjective Belief to Date

	(1)	(2)	(3)	(4)	(5)	(6)
Avg predicted miles to date	-0.031*	-0.054**			-0.031*	-0.038
	(0.016)	(0.025)			(0.017)	(0.027)
L. Avg miles to date			-0.053***	-0.075**	-0.001	-0.052
			(0.013)	(0.034)	(0.032)	(0.039)
Demographic Controls	No	Yes	Yes	Yes	No	Yes
Work Type Controls	No	Yes	Yes	Yes	No	Yes
Observations	8,493	8,493	32,649	8,493	8,493	8,493

Notes: This table is a robustness check to Table 3, where predicted miles is replaced by average predicted miles to date. We also add an additional column, column 2, which is the column 1 specification plus additional controls. In columns 1, 2, 4, 5 and 6, the sample is restricted to observations with non-missing miles, non-missing lagged average miles to date, non-missing mile beliefs, and positive average mile beliefs to date.

Table E4: Do Workers Update their Subjective Productivity Beliefs?

	(1)	(2)	(3)	(4)	(5)
L. Avg miles to date	0.878***	0.622***	0.507***		0.403***
	(0.083)	(0.076)	(0.078)		(0.067)
Tenure X L. Avg miles to date			0.0032**		
			(0.0016)		
L^2. Avg miles to date				0.551***	
				(0.073)	
L. Miles				0.086***	
				(0.015)	
Demographic Controls	No	Yes	Yes	Yes	No
Work Type Controls	No	Yes	Yes	Yes	No
Individual FE	No	No	No	No	Yes
Observations	8,624	8,624	8,624	8,317	8,624
R-squared	0.162	0.335	0.337	0.337	0.614

Notes: This table presents OLS regressions of subjective productivity beliefs on lagged average productivity to date. Standard errors clustered by worker in parentheses. Columns 1-2 show that workers increase their subjective beliefs in response to increases in lagged average productivity to date, as predicted in a normal learning model. Column 3 shows that, as predicted in a normal learning model, workers increase the weight on lagged average productivity to date as worker tenure increases. Column 4 shows that, while agents weigh recent productivity shocks, they place most of the weight on accumulated average productivity to date. Column 5 confirms that updating occurs within worker. All columns include week of tenure dummies. Demographic controls are as in Table 2. * significant at 10%; ** significant at 5%; *** significant at 1%

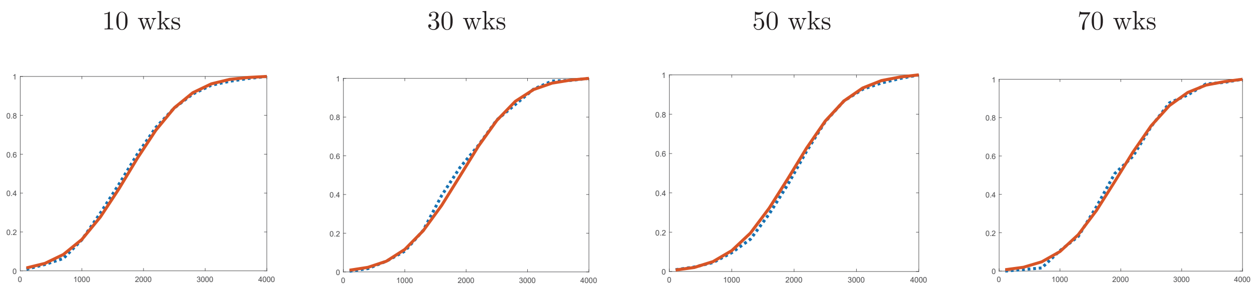
Table E5: Do Productivity Beliefs Predict Productivity? OLS Regressions at Firm B

	(1)	(2)	(3)	(4)	(5)	(6)
L. Pred miles	0.299***	0.298***	0.264***	0.140***	0.146***	0.056
	(0.051)	(0.053)	(0.064)	(0.052)	(0.053)	(0.053)
L. Avg miles to date				0.571***	0.583***	
				(0.077)	(0.075)	
$10 Incentive			-1.955			
			(2.775)			
$10 Incentive X L. Pred miles			0.085			
			(0.116)			
$50 Incentive			-4.476			
			(6.860)			
$50 Incentive X L. Pred miles			0.174			
			(0.309)			
Demographic Controls	No	Yes	Yes	No	Yes	No
Subject FE	No	No	No	No	No	Yes
Observations	803	803	803	695	695	803

Notes: The dependent variable is miles driven per week (in hundreds). An observation is a driver-week. Standard errors clustered by driver in parentheses. All regressions include worker tenure in weeks, dummies for the number of days not worked in a week, and dummies for a worker's week in the study. The demographic controls are gender, age, trucking experience, and region of home residence. These drivers are all from Firm B where we collected subjective productivity forecasts similar to as at Firm A, but randomizing financial incentives for accurate guessing to some workers. We present the data here to show, as at Firm A, that productivity beliefs are moderately predictive of actual productivity across workers, but only weakly so within workers. This finding is consistent with our model in Section 4. In addition, we see that there are no statistically significant differences as to whether productivity beliefs are more productive of actual productivity when they are financially incentivized. * significant at 10%; ** significant at 5%; *** significant at 1%

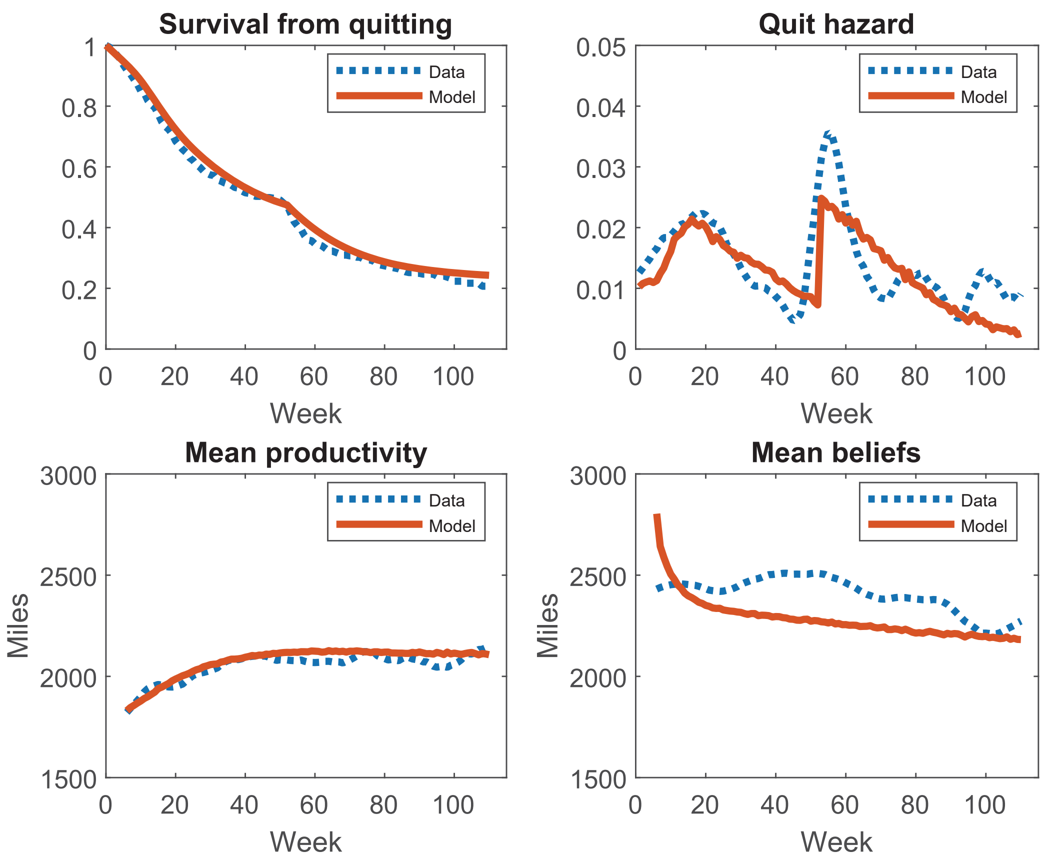
F Further Structural Results

Figure F1: Other Aspects of Model Fit: Tenure and the Distribution of Productivity



Notes: These figures plots the distribution of productivity at different tenure levels, both in the data and as simulated by the model (from column 2 of Table 5) with 200,000 simulated drivers.

Figure F2: Model Fit: Model Estimated With Overconfidence and Standard Learning



Notes: The notes are the same as for Figure 2 except the underlying model is different. The model is similar to that in Column 2 in Table 5 except that it imposes standard learning, that is, where the perceived variance of the productivity signals equals the actual variance, $\widetilde{\sigma_y} = \sigma_y$.

Table F1: Robustness Tests of Alternative Model Specifications

		Baseline	Annual $\delta = .90$	IPW	Winsorize beliefs at 4k mi	T=200	Higher outside option	Finer grid
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Productivity and Skill Parameters							
η_0	Mean of prior productivity dist	1993	1993	1990	1993	1993	1993	1993
		(15)	(15)	(15)	(15)	(15)	(15)	(15)
σ_0	Std dev of prior productivity dist	292	289	292	292	292	292	290
		(11)	(11)	(11)	(11)	(11)	(11)	(11)
σ_y	Std dev of productivity shocks	708	708	707	708	708	708	708
		(3.5)	(3.5)	(3.5)	(3.5)	(3.5)	(3.5)	(3.5)
s_0	Value of skilled gain wks 1-5	4.1	2.0	4.8	4.0	4.2	7.5	3.3
		(4.0)	(2.7)	(3.9)	(4.7)	(4.0)	(4.0)	(3.4)
	Taste UH Parameters							
μ_1	Mass point 1 of taste UH	-290	-367	-288	-279	-286	-130	-243
		(20)	(21)	(21)	(25)	(20)	(20)	(13)
μ_2	Mass point 2 of taste UH	-138	-170	-140	-122	-139	23	-127
		(12)	(12)	(13)	(14)	(12)	(12)	(21)
μ_3	Mass point 3 of taste UH	145	132	150	168	124	305	168
		(40)	(40)	(42)	(43)	(37)	(40)	(43)
p_1	Probability type 1	0.31	0.30	0.32	0.29	0.32	0.31	0.48
		(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
p_2	Probability type 2	0.46	0.48	0.46	0.49	0.46	0.46	0.31
		(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
	Belief Parameters							
η_b	Belief bias	674	683	661	614	673	674	685
		(32)	(33)	(32)	(26)	(32)	(32)	(30)
$\widetilde{\sigma_y}$	Believed std dev of productivity shocks	1673	1612	1714	1507	1683	1673	1618
		(128)	(121)	(136)	(95)	(130)	(128)	(113)
σ_b	Std dev in beliefs	877	877	870	662	877	877	877
		(8.0)	(8.0)	(8.0)	(6.0)	(8.0)	(8.0)	(8.0)
	Scalar Parameter							
τ	Scale param of idiosyncratic shock	2553	3356	2514	3004	2529	2554	2223
		(450)	(638)	(445)	(563)	(452)	(450)	(306)
	Log-likelihood	-94127	-94132	-93586	-92278	-94127	-94127	-94118
	Number of workers	699	699	699	699	699	699	699

Notes: This table presents a number of robustness checks for our main estimates. Standard errors are in parentheses and are calculated by inverting the Hessian. Column 1 repeats the baseline estimates from column 2 of Table 4. Column 2 sets the discount factor equal to 0.9980, corresponding to an annual discount factor of 0.90. Column 3 uses inverse probability weighting to correct for survey non-response (see Appendix A.1). Column 4 eliminates all the subjective belief observations where the stated belief is greater than 4,000 miles in a week. Column 5 increases the period during which learning about productivity may occur from 130 weeks to 200 weeks. Column 6 raises the outside option r by 25% from \$640 per week to \$800 per week. Column 7 uses a finer grid with increments of 100 miles from 100 miles to 4,000 miles. In columns 2, 3, and 5, we first estimate using "fmincon" in Matlab (imposing α_1, α_2, and α_3 are greater than or equal to -\$375) before running "fminunc" as described in Appendix D for the baseline models without learning by doing. Running "fminunc" at the end ensures that our method of calculating the Hessian (using "fminunc") is comparable across columns.