NBER WORKING PAPER SERIES

OVERCONFIDENCE AND OCCUPATIONAL CHOICE

Edward P. Lazear

Working Paper 21921
http://www.nber.org/papers/w21921

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2016

I am grateful to Vivienne Groves, David Kreps, and Ali Yurukoglu for helpful comments. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Overconfidence and Occupational Choice
Edward P. Lazear
NBER Working Paper No. 21921
January 2016
JEL No. D02,J0,M5,M50,M51

ABSTRACT

A statistical theory of overconfidence is proposed and applied to the issue of occupational choice. Individuals who can choose whether to engage in an activity or not must estimate their performance. The estimates have error and that error has positive expectation among those who engage in the activity. As a result, an unbiased ex ante estimate of performance in an occupatoin results in an ex post biased estimate of ability among those enter. The statistical theory of overconfidence provides a number of testable implications, most significant of which is that overconfidence should be more prevalent in occupations where estimates of ability are noisier. This and other implications are tested and found to hold using the Current Population Survey and Panel Study of Income Dynamics data.

Edward P. Lazear
Graduate School of Business
Stanford University
Stanford, CA 94305
and Hoover Institution
and also NBER
lazear@stanford.edu

There is a growing body of literature that suggests that individuals are overconfident of their abilities to perform a given task.[1] Sometimes, this takes the form of those who are in an occupation or activity believing that they can perform the required tasks at a level of performance that exceeds reality. For example, it has been suggested that some, including CEOs, have more faith in their firm or in those that they acquire than is warranted.[2] Additionally, some have interpreted this as evidence of irrationality, reflecting a psychological defect that afflicts at least some significant subgroups of the population.[3]

An alternative view presented here here is that there is a statistical explanation for overconfidence observed in some real world settings. The idea is simple. Individuals estimate their effectiveness when they enter a profession or engage in a specific activity. Those who choose to enter are not a random draw of the population. In particular, they tend to be the most able at that activity, but they are also the ones who have received a better-than-average signal of their expected performance. The estimate used on which they base their behavior is the best unbiased estimate of their performance, but the average estimation error among the group that enters an activity or occupation will be positive, leading to an overstatement of their actual performance.[4] Even individuals who know this fully can do no better by altering their decision rule.

The advantage of the statistical view of over-confidence is that it provides a formal definition of overconfidence and consequently provides implications that are testable and absent from irrational views of overconfidence. One, emphasized in the empirical analysis below, is that when estimates of performance are noisier, there will be more average overconfidence than when the estimates are more precise. This shows up in occupational switching that should more prevalent in occupations where ability is more difficult to predict. Additionally, over a worker's career, the rational theory implies that overconfidence becomes less of a factor because the estimates of ability contain more signal and less noise as the number of observations rise.

Although the statistical argument cannot account for all observations of overconfidence like those in laboratory environments where individuals are constrained to participate in theexperiment and cannot opt out, the environment considered here may be the most consequential for economic life. Choosing an occupation and obtaining the skills for it affect lifetime wealth as much as any decision that the typical individual makes. Therefore,

[1]There is a large literature in both psychology and economics. Early examples in the economics literature include Camerer and Lovallo (1999) and Thaler (2000).

[2]See, for example, Kent Daniel, David Hirshleifer and Avanidhar Subrahmanyam (1998), Malmendier and Tate (2005, 2008),

[3]The literature focuses attention on difference in overconfidence by gender. See, for example, Barber and Odean (2001), Gneezy, Niederle, and Rustichini (2003), and Croson and Gneezy (2009).

[4]See Lazear (2004) for a related argument in a different context. Perhaps closer, and containing a number of similar theoretical results, is Van Den Steen (2004).

understanding the decision process behind occupational choice is worthwhile.

Other implications of the model are:

1. The likelihood of entering the wrong occupation and the cost of doing so are inversely related. When measurement error is small, large mistakes are less likely to be made.
2. As noise in productivity estimation increases, the proportion choosing each occupation moves toward equality. This distorts wages relative to the perfect information world.
3. Measurement error and overconfidence aside, workers still move to occupations in which they have a comparative advantage as is the case in standard matching models.
4. As time in career lengthens, the productivity-estimation error becomes more likely to reflect signal (differences in productivity across occupations that are not picked up in obsevables) than noise. As a result, the importance of the error in predicting occupational switching should increase with tenure.

All of the predictions are found to hold using the Current Population Survey (CPS) or the Panel Study of Income Dynamics (PSID).

1. Model

Initially, consider a one-period structure with two occupations, *A* and *B*. An individual must make a choice between the two and that choice is informed by an estimate of the individual's ability in each occupation as

(1) $\hat{q}_A = q_A + \varepsilon_A$ and $\hat{q}_B = q_B + \varepsilon_B$

where q_A and q_B are the worker's true ability in occupations *A* and *B* and ε_A and ε_B reflect estimation error, assumed have expectation zero and to be independent of each other and of true productivity, q_A and $q_{B.}$

Thus, $\hat{q}_A$ and $\hat{q}_B$ are unbiased estimates of the true ability.

Similarly, an estimate of the difference between ability in *A* and ability in *B* is given by

(2) $\hat{q}_A - \hat{q}_B = \delta + v$

where δ is defined as q_A - q_B and v is defined as ε_A - ε_B . Also define the resulting density functions for δ and v as $h(\delta)$ and $f(v)$, respectively. The assumptions above imply that δ and v have expectation equal to zero and are independent of one another.

Given only this information, the choice of occupation must be made on the basis of (1) so an individual chooses occupation *A* whenever

$\hat{q}_A > \hat{q}_B$

or, using (2), whenever

$-\delta < \nu$.

An error in occupational choice is made if the individual chooses occupation *A* when he should choose *B*, which is to say that $\hat{q}_A - \hat{q}_B$ and δ<0, or conversely, if the individual chooses occupation *B* when he should choose *A*, which is to say that $\hat{q}_B - \hat{q}_A$ and δ>0. To mistakenly choose A, it must be true that $\delta<0$ and that $\nu>-\delta$.

Two properties follow immediately from this structure.

Lemma 1: The cost of a mistake moves with |δ|.

Proof: Immediate. The loss of choosing *A* over *B* is definitionally q_B - q_A, which is defined as -δ. Since $\delta < 0$, a small loss requires that δ be close to zero, which means that its absolute value is also small. |||

Proposition 1: The probability of making an error and the cost of the error are inversely related.

Proof: Consider the case of entering *A* incorrectly, which occurs when $\delta<0$ and $\hat{q}_A > \hat{q}_B$. Since the setup is completely symmetric, the proof for the opposite error of choosing *B* when *A* should have been chosen is analogous.

Given Lemma 1, it is sufficient to show that the probability of making an error, conditional on δ<0 increases in δ. To see that, note

$$\Pr(\hat{q}_A > \hat{q}_B \mid \delta < 0) = \frac{\Pr(\hat{q}_A > \hat{q}_B \cap \delta < 0)}{\Pr(\delta < 0)}$$

$$= \frac{\Pr(-\delta < \nu \cap \delta < 0)}{\Pr(\delta < 0)}$$

$$= \frac{\int_{-\infty}^{0}\int_{-\delta}^{\infty} f(\nu)h(\delta)d\nu d\delta}{\Pr(\delta < 0)}$$

Then

$$\frac{\partial}{\partial\delta} = \frac{\int_{-\infty}^{0} F(-\delta)h(\delta)\,d\delta}{\Pr(\delta < 0)} > 0$$

since both *F(-δ)*and *h(δ)* are positive. |||

Figure 1 describes the situation. At point 1, the individual mistakenly chooses occupation *A* because $\hat{q}_A - \hat{q}_B$. The mistake is a bad one. The large negative δ signifies that the value of choosing *B* is much greater than that of choosing *A*. At the same time, making a mistake of this magnitude is unlikely. In order to obtain a high value of $\hat{q}_A - \hat{q}_B$ when δ is highly negative, it is necessary that the measurement error component, v, be very large (positive), which happens with low probability. Point 2 has the opposite characteristic. The mistake is not a bad one because q_A is close to q_B (δ is close to zero), but the likelihood of making an error of this sort is large because the measurement error needed to make such a mistake, v, is small.

1.1 Overconfidence

Overconfidence is the most important result of the analysis and follows directly from the structure laid out above. Overconfidence takes the form of over-estimating one's ability relative to true ability. If overconfidence is simply defined as having an estimate of ability that exceeds true ability, that phenomenon definitionally translates to drawing a positive measurement error.

Let us begin by showing that those who choose occupation *A* are statistically overconfident of their ability in *A* and statistically under-confident of their ability in *B*. Conversely, those who choose *B* are statistically overconfident of their ability in *B* and statistically under-confident of their ability in *A*

Definition 1: An individual is statistically overconfidence of ability when the expectation of his estimate of performance in an activity exceeds the true expectation of performance.

The formal definition of overconfidence is that, on average, among those who choose *A*, the unbiased estimate of ability in *A* exceeds true ability in *A*, or that

(3) $E(\hat{q}_A > | \; \hat{q}_A - \hat{q}_B > 0) > q_A$

which is equivalent to

(4) $E(\varepsilon_A \,|\, \hat{q}_A - \hat{q}_B > 0) > 0$.

Analogously, for those who choose *B*, overconfidence means that

(5) $E(\varepsilon_B \,|\, \hat{q}_B - \hat{q}_A > 0) > 0$.

Choosing occupation *A* means that $\hat{q}_A - \hat{q}_B > 0$, which in turn implies

$$\varepsilon_A > \varepsilon_B - \delta \; .$$

Proposition 2: On average, those who are found in an occupation are statistically overconfident of their ability in that occupation.

Proof: By definition 1, overconfidence then requires that $E(\hat{q}_A > | \; \hat{q}_A - \hat{q}_B > 0) > q_A$ or equivalently, that

$$E(\varepsilon_A \,|\varepsilon_A > \varepsilon_B - \delta) > 0 \; .$$

This conditional expectation must be positive because the unconditional expectation of ε_A equals zero. |||

Individuals who choose occupation *A* may do so for one of two reasons. They may actually be better at *A* than they are at *B*, i.e., $\delta>0$. Alternatively, they may be worse at *A* than they are at *B,* but in estimating their ability in the two occupations, measurement error favors *A* sufficiently to lead them to conclude that they are actually better at *A.* Those who choose *A* include, disproportionately, those who got positive measurement error of performance in *A,* which means that $\hat{q}_A > q_A$. This is the definition of overconfidence. An individual's unbiased estimate of true ability to perform in occupation *A* is too high among those who choose *A.*

The condition

$$\hat{q}_A - \hat{q}_B > 0$$

also implies that

$$\varepsilon_B < \varepsilon_A + \delta$$

so

$$E(\varepsilon_B \mid \hat{q}_A - \hat{q}_B > 0) = E(\varepsilon_B \mid \varepsilon_B < \varepsilon_A + \delta) < 0$$

again, because the unconditional expectation of ε_B equals zero.

This is statistical under-confidence. Those who choose occupation *A* are under-confident of their ability in occupation *B.* The same logic holds. Those who choose *A* may truly be poor at occupation *B.* Alternatively, they may be better at *B* than they are at *A*, but in estimating their ability in the two occupations, measurement error favored *A* sufficiently to lead them to conclude that they were actually better at *A.* Among those who choose *A* those who disproportionately who got negative measurement error of performance in *B*, which means that $\hat{q}_B < q_B$.

Obviously, there is nothing specific that formally distinguishes occupation *A* from *B.* Consequently, *A* and *B* can be reversed to yield the result that those who choose *B* are overconfident of their ability in *B* and under-confident of their ability in *A.*[5]

1.2 Overconfidence and Occupational Choice

Is there anything an individual can do with the knowledge that those in occupation *A* are overconfident of their ability in *A*? The analysis seems akin to the winner's curse literature where knowledgeable bidders shade their bids to take into account that if they win the auction, they will have, on average, overbid.[6] The answer is no. Even though individuals choosing occupations know that the average individual in occupation *A* overestimates his ability in *A*, two facts remain.

The unbiased estimate of the difference between ability in *A* and *B* for any given individual remains $\hat{q}_A - \hat{q}_B$. Workers who receive a positive signal of *A* relative to *B* understand that part of that signal reflects measurement error and that they are unlikely to do as

---

[5]This result is found earlier in Van den Steen (2004). He emphasizes the role of choice in affecting the amount of overoptimism. Overoptimism results only because agents can choose between options so those who choose an particular option tend to have a high estimate of its value.

[6]The literature on winner's curse traces back to Wilson (1977) , but see also the earlier paper by Brown (1974), which examines bias in assessing the value of an investment project and uses a statistical argument similar to the one emphasized here.

well in *A* as the estimate will tell them. But moving to *B* is not superior. It is also true that those who obtain estimates that suggest they are better in *B* are also likely to be somewhat disappointed. Still, $E(q_A - q_B \mid \hat{q}_A - \hat{q}_B > 0)$ is positive, albeit not as large as $E(\hat{q}_A - \hat{q}_B \mid \hat{q}_A - \hat{q}_B > 0)$ because the latter also includes the positive expected measurement error among those who select *A*. The formal analysis that precedes this sections shows that, but an example is instructive.

Suppose that q_A and q_B are independent and drawn from a normal distribution with mean equal to one-hundred and standard deviation equal to ten. The errors, ε_A and ε_B are independent and also drawn from normal distributions with mean zero and standard deviation equal to ten.

The worker chooses the occupation in which he has the highest estimated output on the basis of the sign of $\hat{q}_A - \hat{q}_B$. A simulation of 10 million draws yields the results shown in table 1. Note that the actual q_A among those who choose *A* is lower than the estimated $\hat{q}_A$ by the size of the average error conditional on choosing A. Among those who choose *A*, the estimated expected output is higher than the realization. But there is no way to use this information to improve the choice. Specifically, setting $\hat{q}_A = \hat{q}_A - x$ for any non-zero value of x results in lower expected output.[7] The worker understands that the average output among those who choose *A* is lower than would be expected based on $\hat{q}_A$ alone, but there is nothing that can be done to improve the choice. In the example shown in table 1, penalizing $\hat{q}_A$ by 3.99, which is the expected value of the error, results in fewer choices of occupation *A*, but also an expected output of 103.91 in stead of 103.99. Any non-zero penalty reduces expected output below the maximum of 103.99.

---

[7]This is true irrespective of the distributions of q_A and q_B. They can be correlated, asymmetric, or both. It is always true that expected output is maximized by choosing occupation A whenever the unbiased estimate of q_A exceeds the unbiased estimate of q_B.

Table 1

Variable	**Mean**
q_A	100.00
q_B	100.00
$\hat{q}_A$	100.00
$\hat{q}_B$	100.00
$\hat{q}_A \mid \hat{q}_A > \hat{q}_B$	107.98
$\hat{q}_B \mid \hat{q}_A > \hat{q}_B$	92.03
$\varepsilon_A \mid \hat{q}_A > \hat{q}_B$	3.99
$q_A \mid \hat{q}_A > \hat{q}_B$	103.99

1.3 Measurement Error and the Probability of Occupational Error

The overall goal, in addition to formulating a statistical view of overconfidence, is to provide implications, especially those that differentiate the statistical theory from others. Perhaps the most direct implication of the statistical overconfidence approach that does not follow from other approaches is that overconfidence varies with the signal-to-noise ratio. Intuitively, there should be no overconfidence in occupations where there is no measurement error and overconfidence should be prevalent in occupations in which there is a great deal of measurement error. It follows that the amount of overconfidence and under-confidence increases in the variance of the error.

Generally, however, the interest is on actions that are taken by workers, given measurement error, and how those actions depend on the error distribution. For example, it is useful to derive the probability of switching out of an occupation because of overconfidence and how that probability varies with observable factors.

To start, note that in (4), the conditional expectation of ε_A depends on the underlying distribution of ε_A. Here it is shown that the probability of a worker in occupation *A* being overconfident of ability in *A* is higher with more disperse distributions of the error ε_A for any given distributions of ε_B and δ.

Let the distribution of ε_A be given by $f_i(\varepsilon_A)$. Consider two density functions, $f_1(\varepsilon_A)$ and $f_2(\varepsilon_A)$ as shown in figure 2. Specifically, the two distributions are assumed to have a single crossing point at x=x* for the region where $\varepsilon_A < 0$, and the more disperse distribution, $F_1(x)$, is defined as the distribution that has the higher density value, $f_1(x)>f_2(x)$, for x<x*. Finally, let the distributions both have the property that $F_1(0) = F_2(0)$. The densities need not be symmetric. They are required neither to have any particular shape nor to come from the same family of distributions.

Definition 2: Given two distributions, $F_1(x)$ and $F_2(x)$ with $F_1(0) = F_2(0)$ and a single crossing point at x* in the region where x<0, F_1 is defined to have greater spread than F_2 if $f_1(x) > f_2(x)$ for x<x*.

Definition 3: The probability of overconfidence among those who choose occupation *A* is

$$\Pr(\varepsilon_A > 0 \mid \hat{q}_A > \hat{q}_B).$$

The probability of over confidence can be written as

(6) $\Pr(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta)$,

which equals

(7) $\Pr(\varepsilon_A > 0 \cap \varepsilon_A > \varepsilon_B - \delta) / \Pr(\varepsilon_A > \varepsilon_B - \delta)$

The expression in (6) can be broken up into two parts as

(8) $$\Pr(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta) = \Pr(\varepsilon_A > 0 \mid \varepsilon_B - \delta \geq 0 \cap \varepsilon_A > \varepsilon_B - \delta) \Pr(\varepsilon_B - \delta \geq 0) + \Pr(\varepsilon_A > 0 \mid \varepsilon_B - \delta < 0 \cap \varepsilon_A > \varepsilon_B - \delta) \Pr(\varepsilon_B - \delta < 0)$$

If $\varepsilon_B - \delta > 0$, then the first term of (8) is independent of the distribution of ε_A because the conditional probability in (6) equals 1 in that case, always. If ε_A is greater than a positive number, it is greater than zero, i.e, the numerator and denominator of (7) have the same value.

Proposition 3: The probability of overconfidence increases in the spread of the distribution of the estimation error, ε.

Proof:

First note the following:

(9) a. $\Pr(\varepsilon_A > 0) = 1 - F_i(0)$

b. $\Pr(\varepsilon_A > \varepsilon_B - \delta) = 1 - F_i(\varepsilon_B - \delta)$

Consider the two distribution functions, F_1 and F_2 defined above. Because ε_A, ε_B and δ are independent, $\Pr(\varepsilon_B - \delta \geq 0)$ is not affected by the distribution of ε_A. Additionally, because $\Pr(\varepsilon_A > 0 \mid \varepsilon_B - \delta \geq 0 \cap \varepsilon_A > \varepsilon_B - \delta) = 1$, the likelihood of overconfidence, $\Pr(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta)$ depends only on how the second term in (8) is affected by the distribution of ε_A.

Write the conditional probability overconfidence among those who choose *A* as

$\Pr_i(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta)$ i = 1, 2

where the subscript i denotes whether ε_A is distributed as $F_1(\varepsilon_A)$ or $F_2(\varepsilon_A)$. Then a sufficient condition for $\Pr_1(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta)$ to exceed $\Pr_2(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta)$ is that

(10) $\Pr_1(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta) - \Pr_2(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta) > 0 \quad \forall \ \varepsilon_B - \delta < 0$.

Equations (7) and (9a,b) imply that

(11) $\Pr_1(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta) - \Pr_2(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta) = \dfrac{1 - F_1(0)}{1 - F_1(\varepsilon_B - \delta)} - \dfrac{1 - F_2(0)}{1 - F_2(\varepsilon_B - \delta)}$

Because, $F_1(0) = F_2(0)$, the sign of (11) depends only on the denominator. If
$F_1(\varepsilon_B - \delta) > F_2(\varepsilon_B - \delta)$,
then the sign of (11) is positive .

For $\varepsilon_B - \delta \leq x^*$ as defined above and shown in figure 2, it is clear that
$F_1(\varepsilon_B - \delta) > F_2(\varepsilon_B - \delta)$ because $f_1(x) > f_2(x)$ for $x<x^*$.

For $x^* < \varepsilon_B - \delta < 0$,

$$F_i(\varepsilon_B - \delta) = F_i(0) - \int_{\varepsilon_B - \delta}^{0} f_i(x)dx$$

But since $F_1(0) = F_2(0)$ and $f_1(x) < f_2(x)$ for $x^*<x<0$, $F_1(\varepsilon_B - \delta) > F_2(\varepsilon_B - \delta)$. Taken together, this implies that

$F_1(\varepsilon_B - \delta) > F_2(\varepsilon_B - \delta) \quad \forall \ \varepsilon_B - \delta < 0$,

which completes the proof that $\Pr_1(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta) - \Pr_2(\varepsilon_A > 0 \mid \varepsilon_A > \varepsilon_B - \delta) > 0$. |||

This provides an important empirical implication. Specifically, the likelihood of overconfidence is higher in occupations that have higher spread. As will be shown below in section 6, coupled with a reasonable model of learning about ability, this result also implies that the probability of switching out of an occupation rises in the variance in measurement error.

1.4 Other Views of Overconfidence

Statistical overconfidence is defined as having estimates of productivity in the various occupations that are unbiased, that is,

$$E(\hat{q}_A) = q_A$$

and

$$E(\hat{q}_B) = q_B$$

This is equivalent to saying that the expectations of the errors, ε_A and ε_B, are zero. Although this results in overconfidence in each occupation among those who go into the occupation, among the entire population, estimates of productivity are unbiased.

A natural way to think about irrational overconfidence is that the expectations of the errors are positive, that is, people overestimate their productivity levels or formally, that

$$E(\varepsilon_i) = \omega$$

where $\omega>0$ is the overconfidence parameter and is stable over time. This formalization of irrational overconfidence has come to be known as the “Lake Wobegon effect,” an early example of which is in automobile driving, where the majority of drivers view themselves as superior to the median.[8]

---

[8]See Svenson, Ola (February 1981). "Are We All Less Risky and More Skillful Than Our Fellow Drivers?". Acta Psychologica 47 (2): 143–148. doi:10.1016/0001-6918(81)90005-6 and
Iain *A.* McCormick; Frank H. Walkey; Dianne E. Green (June 1986). "Comparative Perceptions of Driver Ability: A Confirmation and Expansion". Accident Analysis & Prevention 18 (3):

The automobile driving example, and others, seems to reflect stability of ω. Most drivers have been driving for a significant amount of time, yet the Lake Wobegon effect persists, even among experienced drivers.

In a multi-period context, an alternative view of irrationality might better be defined as using the “wrong” learning model. Rather than having ω>0, agents may update in a non-Bayesian way. This is defined more formally in section 1.6, below.

The desire is to determine empirically whether and in which situations the concept of statistical overconfidence better explains the data than does that of irrational overconfidence. If $\hat{q}_A$ could be distinguished from q_A, then there would be a test of irrational overconfidence. Indeed, that is exactly what laboratory experiments sometimes do, by obtaining ex ante predictions of productivity and then observing the actual productivity. This is more difficult to do in the real world, but there are some examples. Hoffman and Burks (2015) has shown that truckers tend continually to overestimate their work productivity. Also, discussed earlier is a financial literature that examines overconfidence in investment behavior and firm decision making.

Actual observations of choice over occupations may not be helpful. Specifically, if the overestimate of ability were the same in both occupations, then the choice of occupation would not be affected because the $\hat{q}_A - \hat{q}_B$ would difference out the common overconfidence component, ω. In some ways, that is encouraging because it means that as long as overconfidence, even of the irrational type, is neutral across activities, choices are unaffected and behavior remains invariant to it. Not only would a worker choose the right occupation, but even overconfidence about something like automobile driving ability might be harmless. A driver who assumes that he is better than average may be appropriately cautious because he assumes that other drivers are worse than average and must compensate for their poor driving.

We may not be so lucky, however. It is possible that some activities are less susceptible to overconfidence than others. For example, if people estimated the value of their productivity in boring, everyday activities accurately, but overestimated their likely success in novel, exciting opportunities, they would short change the mundane, while trying and failing in too many flashy activities.

Is there any natural way to select one occupation or type of person as more susceptible to overconfidence than another? The literature that is based on lab experiments does that. As discussed earlier, there are a number of studies that conclude that men are more likely to be overconfident than are women. Even here, however, it is necessary to specify which activities are at play and how the relevant expectation is defined. Suppose, for example, that as a result of discrimination, men are generally more successful than women. When a new activity is encountered, even unbiased estimation would lead men to estimate their productivity to be higher than would women, even if in that specific activity it turns out, ex post that women do as well as men. Of course, this view could not hold if, say, men habitually overestimate their performance and women habitually underestimate their performance.

1.5 Occupational Distribution and Equilibrium Wages

205–208. doi:10.1016/0001-4575(86)90004-7.

The theory of statistical overconfidence implies that mistakes are made. Some individuals who are better suited to *B* enter *A* and some who are better suited to *A* enter *B* . What does this imply about the number of individuals in occupation *A* relative to *B* and how does this affect the equilibrium wages?

First, it is shown that mismatches may occur in equilibrium and that too many may be in one occupation over another, even with symmetric measurement error. (Recall that irrational overconfidence is defined as asymmetric measurement error.)

Let there be N individuals. A worker should choose *A* over *B* when $q_A > q_B$ or, equivalently, when $\delta>0$. Therefore, the socially efficient number of workers in *A* and *B* are N[1-H(0)] and N H(0), respectively.[9] Since the worker estimates productivity in each field imperfectly and chooses *A* when and only when

$$\hat{q}_A > \hat{q}_B$$

or equivalently, when and only when

(12) $\delta > -\nu$,

the actual number of individuals in occupation *A* is

$$N\int_{-\infty}^{\infty}\int_{-\nu}^{\infty} h(\delta)\, f(\nu)\, d\delta\, d\nu$$

or

$$\textit{Actual Number in A} = N\int_{-\infty}^{\infty}[1-H(-\nu)]\, f(\nu)\, d\nu$$

which, in general does not equal N[1-H(0)]. As a consequence, the number of workers in the occupations does not necessarily equal the socially appropriate number.

To get a sense of how this varies with the situation, consider the following example. Suppose that 1-H(0) = .9 so that most of the working population should be in occupation *A.* Suppose that f(ν) is a symmetric density function with mean zero and variance σ^2_ν . As σ^2_ν goes to infinity, whether condition (12) holds or not depends only on ν. Further, because the distribution of ν is symmetric, the probability that (12) holds goes to ½ . Conversely, as σ^2_ν goes to zero, whether condition (12) holds or not depends only on δ so the actual number in *A* equals the socially optimal number in *A.*

This example illustrates two points. First, the proportions in the occupations need not, and in general will not, equal the socially optimal proportions. Second, the deviation from social optimality depends on the variance in the δ distribution compared with the variance in the ν distribution. As measurement error becomes more important, the likelihood of having correct

[9]Having the correct the number of workers in each occupation does not guarantee efficiency because the wrong workers could be in each occupation, even if their numbers offset appropriately.

proportions in the occupations declines.[10]

If the supply of labor to the various occupations differs from that which would hold under efficiency, then it is also true that wages will differ from those that would result in a perfect information competitive equilibrium. The higher is σ^2_v , the more likely is there oversupply in the occupation that should have few than half of the population. As a result, wages are depressed relative to the perfect information competitive equilibrium in the occupation that should have fewer than ½ of the workers.

An empirical implications follows. Suppose that there are K groups in the population and that each group accounts for share α_k of the workforce. For example, k might index three education groups: college or more, high school, and less-than-high school. Define α_{jk} as the share of industry j comprised of group k, where j=*A,B.* Also denote δ_{ik} and v_{ik} as the differences in productivity and in measurement error, respectively, for individual *i* in group *k*. Then

Proposition 4: For any variance in δ_{ik} and for any symmetric distribution of $v_{ik,}$ as the variance in v_{ik} goes to infinity, the proportion of group *k* that choose occupation *A* goes to ½.

Proof: For an individual *i* in group *k* to choose industry *A* over *B*, it must be true that

$$\hat{q}_{A_{ik}} > \hat{q}_{B_{ik}}$$

or that

$$\delta_{ik} > -\nu_{ik}$$

For any finite variance in δ_{ik}, as the variance in ν_{ik} goes to infinity, the difference between $\hat{q}_{A_{ik}}$ and $\hat{q}_{B_{ik}}$ depends only on the noise term, ν_{ik}. Because the noise is distributed symmetrically, v_{ik} is positive half the time, resulting in the choice of occupation *A* over *B* half of the time. |||

The market equilibrium is affected by the variance in ν. In the extreme case above, where
σ^2_ν approaches infinity for one group, half of that entire group enters occupation A and half enters B, even if the appropriate proportions, based on δ, are far from one-half. As before, this means oversupply to one industry and under-supply to the other, which results in too much labor and too low a wage in one industry and the reverse in the other.[11] The general implication is that

[10]Of course, it could be coincidentally that the proportions that are generated mimic the socially optimal ones. For example, suppose that 1-H(0) = .5 and that ν is distributed symmetrically around zero. As the variance in ν goes to infinity, it is certain that the actual proportion in *A* is the socially optimal proportion. However, the individuals who are in *A* are not the correct ones to be in A because measurement error means that individuals are assigned to occupations based on randomness rather than on δ_{ab} , the latter being necessary for efficiency.

[11]A variant on this is to assume that the worker has some prior information on δ and does Bayesian updating after obtaining information on $\hat{q}_A$. This will change the “one-half” implication quantitatively but not qualitatively. In the extreme case, where the prior is

the noisier is the estimate of productivity in an industry for a particular group, the closer the proportion of that industry comprised of the group will be to complete random choice. To test this empirically, it is necessary to be able to measure something that relates to the group-specific signal-to-noise ratio and then to observe how the group weights in the industries reflect that noise.

1.6 Multi-period Structure

It is desirable to have an empirically verifiable measure of overconfidence that goes beyond asking people about their confidence or observing their behavior in an experimental setting. Some researchers have tried to assess overconfidence by examining behavior, for example, by looking at the investment behavior of CEOs who buy assets in their own firms.[12] The view here is that overconfidence is a statistical phenomenon, reflecting a choice that is optimal given ex ante information, but one that should also be remedied over time as more information becomes available. The larger the amount of statistical over-confidence, the more corrective action that should be taken.

One implication seems intuitive and is shown formally in this section. Proposition 3 stated that the probability of overconfidence increases in the spread of the estimation error distribution. A worker can learn with work experience that his occupational choice was a mistake. For example, if an individual chooses occupation *A*, then he learns about his ability in occupation *A* and updates his estimate of δ over time. This implies that switches out of an occupation, conditional on having chosen it in the first place, should depend on the amount of measurement error that existed initially and how quickly that measurement error disappears with experience.

To derive this implication formally, it is necessary to extend the model beyond one period. To focus on the importance of noise and learning, the model is stripped down to its bare

---

degenerate and the variance in v is positive, there would be no tendency at all toward 1/2 for higher variance groups. Specifically, if a worker was certain of the value of δ, that would be the value assumed, irrespective of the estimate that is obtained from the $\hat{q}_A$. At the other extreme, if the prior were completely diffuse and the worker had no ex ante information on δ, only the observations on $\hat{q}_j$ would be relevant. In general, some weight would be attached to the prior and to the observed value in coming up with a posterior and although the weight on the observed value would fall with increased variance in v, there is no general statement that can be made about the nature of updating without specific assumptions about the prior.

[12]In addition to finance literature already cited, see Brunnermeier and Parker (2005) on overconfidence in portfolio choice. The finance context is somewhat different from the analysis of occupational choice because the purchase of stock is like bidding on a common value asset and has the winner's curse property that others' information should be used in drawing inferences about the value of stock. In the case of occupational choice, the individual's ability is person-specific and idiosyncratic. As a result, others' behavior shed no light on the worker's personal decision.

essentials. Assume that a worker gets a signal on $\hat{q}_A - \hat{q}_B$ at the beginning of period 0, on which he bases his choice between the two occupations. To make things simple, assume that $\varepsilon_B = 0$ and that the only measurement error that exists comes from mis-estimating ability in occupation *A*. Stated differently, ν is simply ε_A.[13] Assume further that the worker learns perfectly his true δ after working in an occupation for one period.[14]

The probability of switching from occupation *A* to occupation *B* after period zero is simply the probability that the individual chose occupation *A* by mistake, given that he chose A. This conditional probability can be written

$$\Pr(\text{switch from } A \text{ to } B \text{ after period zero}) = \Pr(\nu > -\delta \cap \delta < 0) / \Pr(\hat{q}_A > \hat{q}_B)$$

There are two reasons that a worker could have chosen occupation *A* in period zero. Either he belongs there, which requires δ>0, or he does not belong there, but measurement error was sufficient to induce him to choose *A* mistakenly. It is the latter that induces a switch. Thus, the expression above for a switch becomes

$$\Pr(\text{switch from } A \text{ to } B \text{ after period zero}) = \Pr(\nu > -\delta \cap \delta < 0) / \Pr(\nu > -\delta)$$

which can be written as

$$\Pr(\text{switch from } A \text{ to } B \text{ after period zero}) = \Pr(\nu > -\delta \cap \delta < 0) / [\Pr(\nu > -\delta \cap \delta < 0) + \Pr(\nu > -\delta \cap \delta \geq 0)]$$

Since ν is now identical to ε_A, the F_i notation used above for ε_A now can be used to reflect the distribution functions for ν. Thus, F_1 is the distribution of the two with the higher spread, which means that the importance of measurement error should be greater with F_1 than with F_2 as shown in figure 2. Then,

Proposition 5: For any given distribution of δ, the probability of switching out of occupation *A* and into occupation *B* increases with the spread in the distribution of ν.

Proof: Given that the density function on ν associated with distribution i is $f_i(\nu)$,

---

[13] This bears some resemblance to the two-armed-bandit structure originally formulated by Herman Robbins, "Some Aspects of the Sequential Design of Experiments," *Bulletin of the Amer. Math. Soc.* 58, (1952), 527-35. When the worker knows that he is not good at *A*, it pays for him to take a chance on another occupation.

[14] An alternative assumption that does not qualitatively change results is that he learns only his output in the occupation in which he works during period zero.

Pr(switch from *A* to *B* after period zero) = $\dfrac{\int_{-\infty}^{0}\int_{-\delta}^{\infty} f_i(\nu)h(\delta)d\nu d\delta}{\int_{-\infty}^{0}\int_{-\delta}^{\infty} f_i(\nu)h(\delta)d\nu d\delta + \int_{0}^{\infty}\int_{-\delta}^{\infty} f_i(\nu)h(\delta)d\nu d\delta}$

or

Pr(switch from *A* to *B* after period zero) = $\dfrac{\int_{-\infty}^{0}(1-F_i(-\delta))h(\delta)d\delta}{\int_{-\infty}^{0}(1-F_i(-\delta))h(\delta)d\delta + \int_{0}^{\infty}(1-F_i(-\delta))h(\delta)d\delta}$

Consider how the value of this expression changes as spread in the error increases. When $\delta < 0$ so that $-\delta > 0$, $F_2(-\delta) > F_1(-\delta)$. Conversely, when $\delta>0$ so that $-\delta<0$, $F_1(-\delta) > F_2(-\delta)$. The combination of the two implies that the numerator rises relative to the denominator when moving from F_2 to F_1 because the first term of the denominator equals the numerator and the second is larger for i=1 than for i=2. Thus, the probability of a switch from *A* to *B* is higher with F_1 than it is with F_2 . |||

The intuition of this result is made clear by considering two extremes. Suppose there was no measurement error at all at the beginning of period zero. Then all who choose *A* do so because $\delta>0$. There are no mistakes. At the other extreme, let the variance in measurement error go to infinity. Then (as discussed in 1.5, above), half would choose A and those who chose A would be choosing completely on the basis of noise. The true δ would have no impact on the initial choice because the signal-to-noise ratio is zero. After one period of experience in A, the true δ would be learned and H(0) of those currently in A (the proportion who should actually be in B) would switch out of A.

Note the empirical implication is not merely that the more noise there is at the outset, the more switches. Imbedded in this structure is that noise in occupation *A* is reduced relative to noise in *B* as the worker acquires experience in the occupation. Still, other things equal, occupations in which wages spread more with experience should also be evidenced by more occupational switching.

The rate at which learning about δ and ν occurs also affects the switch probability. To see this, consider two occupations that are equally noisy, meaning that the variance in ε_i is the same in both occupations. If experience reduces the variance more quickly in one occupation than in the other, then switches should occur to a greater extent in that occupation with the greater reduction in variance. Occupations in which there is no learning about productivity do not experience switching even if the initial variance in noise is large. One way to estimate the rate of learning is to examine the pattern of wage variation with experience. When estimates of productivity are poor, everyone is paid the same amount. As learning about productivity occurs, wages move toward productivity and spread out. Consequently, more switching should occur in occupations that have a larger increase in wage variation with experience because those are the ones in which learning about productivity is likely to be largest. Of course, other factors could confound this prediction. For example, one occupation may experience higher wage variation with experience simply because experience is more important in that occupation, coupled with the fact that workers learn at different rates. But this explanation can be examined by holding constant the slope of the experience-earnings profile.

2. Empirical Implications

In this section, the implications of the theory are used to discuss results that have already been produced by the previous literature.

The primary goal is to distinguish statistical overconfidence from irrational overconfidence, as defined in section 1.4. The results of sections 1.3 and 1.6 yield some implications. Here, those implications are tested using data on occupational switching from the CPS January 2012 , from the CPS March 2012, and from the PSID.

2.1 More noise implies more occupational switching

The most compelling empirical predictions result from propositions 3 and 5, which relate the likelihood of switching occupations to the distribution of the noise terms. Proposition 3 states that the probability of overconfidence increases in the spread of the distribution of estimation error. Proposition 5 extends that to a dynamic context and says the probability of switching out of an occupation and into another increases in the spread of the estimation errors, given the distribution of true comparative advantage, δ.

It is necessary to find an empirical analogue of estimation error. First, write

(13) $\hat{q}_{ji} = x_{ji}\,\beta + \varepsilon_{ji}$

for individual i in occupation j, where x_{ji} are the values of the observables for that individual and β are the coefficients that apply to those explanatory variables for occupation j.

In competition, workers are paid their observed output so the worker's wage, w_{ji}, is equal to $\hat{q}_{ji}$. Then, an estimate of ε_{ji} is simply

(14) $w_{ij} - x_{ji}\beta$

The empirical implication is that switches out of occupation j should be more common for any given distribution of δ when occupation j has a higher spread in ε, measured by the standard deviation of the residual in the wage equation.

2.2 Individuals should switch out of occupations in which they have comparative disadvantage

The time-varying analogue of (13) is

(15) $\hat{q}_{jit} = z_{jit}\,\Gamma + \alpha_i + \theta_{ji} + \xi_{jit}$

where $\hat{q}_{it}$ is the worker's estimated output at time t, z_{jit} are observables at time t and include occupation dummies, α_i is person-specific ability component that is not captured by the observables, and θ_{ji} is the match effect of person i in occupation j, i.e., the person-specific comparative advantage component. Finally, ξ_{jit} is the error that may reflect measurement, luck or other factors.

The primary prediction is that workers should tend to switch out of an occupation when they are less well-suited to it. Some of this is predictable and built into the observables. For example, the part-time and summer jobs that are most appropriate for college students are not the ones that form the basis of most careers and this should be related to age, experience, and education levels that vary as the individual progresses through school.

Much of the occupational switching occurs over time as workers learn their comparative advantages. In the two occupation case, that was reflected in δ. In the multi-occupation case, there is an θ_{ji} that refers to individual i's ability in occupation j, taking out the average effect of being in occupation j and the general ability effect for individual i.

An individual who has a high output, reflected in a high wage, in period t, might have that wage because he has favorable characteristics (high $z\Gamma$), because he is more able than others (high α), because he in a well-paying occupation (high coefficient on his occupation dummy), or because he is well-suited to that particular occupation (high θ_{ji}). To the extent that each of these components can be identified, they may affect the likelihood of switching occupations. A standard proportional hazard approach with the appropriate variables included can shed light on this issue.[15]

This also relates to the implications of proposition 1, which states that the cost of the error and the probability of an error occurring are inversely related.When the cost of the error is high, i.e., when $|\delta|$ is large, the individual is unlikely to choose the wrong occupation because it is rare that $\nu>-\delta$. Because those who have strong comparative advantages in the occupation will likely have chosen correctly initially, these individuals are more likely to stay in that occupation.

The error, $\xi_{jit,}$ in (15) can be broken up further to yield empirical implications. Note that the terms that comprise the estimate of productivity contain person, occupation, and person-occupation effects, but they do not contain a firm-worker match effect. That is incorporated into ξ_{jit}. Even if ξ_{jit} could be separated perfectly from $z_{jit}\Gamma$, α_j, and θ_{ji}, the retrieved ξ_{jit} would still contain a component that relates to true productivity, namely the firm-worker match effect, rather than estimation or measurement error. As the worker remains longer on the job, it is standard logic that the firm-worker match component gets estimated more precisely and the measurement error component becomes less important. This implies that ξ_{jit} increasing measures

[15]Among the earliest work on occupational investment and its importance in earnings determination is Shaw (1984), which finds that occupation-based models are much better at explaining earnings variation than are those based simply on work experience. Shaw (1987) builds a structural model to estimate the transferability of skills across occupations and estimates it using the National Longitudinal Survey of Young Men. Farber (1999) provides an excellent survey of the early literature on job mobility, especially that which focuses on the role of specific human capital (which could be occupation specific). Farber divides those studies into those based on worker heterogeneity and specific human capital, those based on worker tenure, which may proxy these factors, and those that examine the earnings experience of displaced workers. Somewhat related is Neal (1995), which examines the effect of industry (as opposed to occupational) switching. Neal shows that the cost of switching industries is related to work experience and tenure. These studies point more directly to the cost of losing occupation or industry skills when a worker is displaced involuntarily rather than the benefit of moving to a more appropriate occupation voluntarily when a worker learns about his or her comparative advantage.

match effects that relate to productivity rather than just noise. Variations in ξ_{jit} for new workers are unlikely to result in different separation behavior because these variations reflect pure noise. But for senior workers, cross-occupation variation in ξ_{jit} is more likely to reflect firm-worker match effects being more important in some occupations than in others. They should result in action. Specifically, occupational variation in ξ_{jit} should be more predictive of turnover behavior for more senior workers than for junior workers.

3. Data Description

Three different datasets are used because no one dataset includes all the relevant information. For the primary test that examines the relation between occupational switching and the amount of noise in the occupation-specific estimate of productivity, two CPS months are used. The wage data are from the 2012 March CPS, which gives detailed data on the hourly wage rate and on a person's demographic characteristics. The March CPS is used to estimate the occupation-specific variation in ε, which is the key independent variable used to predict occupational switching. There are about 132, 877 working individuals in the wage data, a number that is reduced to 115,039 to obtain complete information on all individuals. The wage function in equation (14) is estimated using these data.

Unfortunately, the March data do not provide detail on job switching. To determine occupational switching, it is necessary to use the January data. Although January data lack detailed wage information, they do contain better detail on the prior job history. The occupational breakdowns are the same in both months so that the variation in ε obtained from the March wave can be used to predict the amount of occupation-specific noise that was present in the January respondent's previous-year occupation. Matching the occupation that the January 2012 respondent held the previous year with the estimated occupation-specific variance in ε obtained from the March 2012 data provides a way to predict which individuals would have been more likely to switch occupations in the year prior to January 2012.

More specifically, the January 2012 CPS data contain 39,992 observations on individuals whose current job and job one year earlier are reported. It is possible in these data to determine whether an individual was employed in the same job, and specifically, in the same occupation last year as he or she is today. The inter-year occupational stability is the variable of key interest and that can be determined on an individual or occupational basis. Since one can compute whether a worker switched occupations during the year prior to January 2012, the switch rate can be computed for each of 22 occupations in the data and those rates can be related to the spread in the ε distribution that is obtainable from the March, 2012 CPS data. Since the variation in the key variable, ε, occurs at the occupational level, the appropriate unit of analysis is the occupation. As a result, the 39,992 observations are used to compute the occupational averages for switching rates and the independent variables for the 22 occupations. The final data then contains 22 observations where each observation reflects the mean values for that occupation. Table 2, Panels A, B, and C provide summary statistics for the CPS data sets.

Wage data on the prior year's job is not accurately reported in the CPS data and the worker's prior wage for those who leave their jobs and switch occupation is key for the purposes of testing the importance of comparative advantage, δ, in occupational switching. The estimation of (15) requires a serious panel dataset. The Panel Study of Income Dynamics (PSID) is used for this purpose. This is a well-known data set and for the purposes here, a maximum of 37,891 observations over all years on 7743 individuals are valid. The statistics on the relevant variables are summarized in table 2, Panel D.

4. Results

4.1 Occupational Switching Increases in the Spread of the Error Distribution

Recall the primary implication is that the amount of switching out of an occupation and into another because of statistical overconfidence rises in the spread of the distribution of measurement error. In the context of the model, the larger the spread in ε, given the distribution of δ, the more occupational switching. The distribution of ε varies by occupation so the model described by equation (14) is estimated using the March, 2012 CPS data. This is done in two ways. First, the same equation is assumed to hold across all occupations. Equation (14) is estimated as written where the vector of β is assumed to be the same across all occupations. The variables that are included to explain wages are the typical ones: age, education, and gender. This is estimated on slightly over 115,000 observations where the dependent variable is the hourly wage rate (stated or imputed by CPS). The results of this first stage are typical. The coefficients on age, education and male are positive and the r-squared for the entire regression is .23. An alternative version allows the β vector to be occupation-specific, which results in a different set of estimated ε. Allowing full interaction by occupation boosts the r-squared to .31.

Given the estimated ε for each worker in the sample, the standard deviation of the ε is computed within each occupation. That is the key variable of interest, the hypothesis being that the larger the standard deviation in ε, the larger is the occupational turnover rate. As stated above, the within-occupation standard deviation of ε is estimated in two ways. The first assumes the same wage equation holds for the full sample. The second allows different coefficients on each explanatory variable for every occupation.

4.1.1 Occupational Switching is Related to Residual Variation in Wages

The main results are contained in table 3. Note first that the dependent variable is the proportion who change occupations during the year. The unit of analysis is the occupation so there are 22 observations in all and each variable described in panel B of table 3 relates to the average within-occupation value that the variable takes on across the 22 occupations. For example, the average age is 39.4. This is the within-occupation mean age, averaged (unweighted) across the 22 occupations.

Figure 3 shows the kernel density of the dependent variable. It is well-behaved and lies within the interior of the interval, taking on a modal value of about .11, meaning that about 11% of the sample changed occupations during the previous year. This number seems consistent with other data. In particular, the JOLTS (Job Openings and Labor Turnover Survey) suggest about 50 million hires and separations during 2012 when employment was about 140 million workers, which implies a total turnover rate of over 1/3. But only a fraction of those (about 1/3 again, if the CPS number is accurate) change occupations.

The theory predicts that occupational switching should be greater in occupations in which the spread in ε is high. The spread in ε is measured by the occupation-specific standard deviation of the residual from the wage regression. But the theoretical prediction assumes a given distribution of δ. The distribution of δ may vary with observable factors, like age, education and gender. The fact that the wage regression includes these factors does not eliminate the across-occupation variations in the distribution of δ. Worse, the distribution of δ may be correlated with the distribution of ε across occupations. For example, as mentioned earlier, occupations in which human capital is more important may also be occupations with higher variation in wages, inter-person productivity variations and productivity estimation error. Specifically, the variation in the residual in the wage regression picks up errors in productivity estimation error as well as inter-person variations in true productivity. Formally, the residual defined in (14) contains not only measurement error, but another person-specific effect. The δ component relates to occupation fixed effects for a given individual i, which requires observing that individual in multiple jobs in all occupations. Later, mobility reflecting comparative advantage reflected in δ is analyzed using panel data, but in this section, using the occupational data, δ cannot be separated from measurement or estimation error, which is what ε is supposed to reflect. However, many variables that might be expected to affected comparative advantage across occupations (like education) are observable. Table 3 allows specifications where those variables are included and allowed to affect occupational switching. These take out some, although not all, of the cross-individual differences in comparative advantage.

Table 3, and columns 3 and 6 especially, provides strong support for the prediction that occupational switching is positively related to the spread in the ε distribution and that comparative advantage in the occupations, reflected in the distribution of δ, varies with education and age. The difference between the columns 3 and 6 is that the wage regression used to estimate the residuals that form the basis for calculating the standard deviation of ε conforms to two different forms. In column 3, the coefficients in the wage equation that generates the estimated ε are assumed to be the same across occupations. In column 6, the coefficients are allowed to vary across occupations, meaning that 22x4 or 88 separate coefficients are estimated in the wage regression. The sample of 115,039 observations is used to estimate both forms.

As is apparent, the standard deviation of epsilon is strongly correlated with the likelihood that a worker shifts occupations. Theory implies that statistical overconfidence in initial occupational choice is more pronounced in high standard-deviation-of-epsilon occupations. The larger is the residual in the wage regression, the more important is ε and the larger is the variation in epsilon, the larger is the likelihood of switching out of the occupation. This is what is found in table 3. All specifications that control for observable factors that affect comparative advantage support this prediction.

Note that the results are quite strong and statistically significant, somewhat surprising given only 22 observations. Of course, each observation carries significant weight because it reflects average values across all individuals within that occupation and the initial sample from which these 22 observations are created contain over 39,000 individuals. The magnitude is also impressive. The standard deviation in the wage residual itself has a standard deviation across occupations of 3.5, so a one standard deviation change in the value of that variable is associated with a .013 increase in average occupational switch,[16] which is over 10% of the overall probability of an occupational switch.

Column 8 reports results based on a first-stage wage regression that deletes outliers.

[16]This is based on the estimates in column 6, but the coefficients differ little across specifications with all variables included.

Some individuals in the initial sample of about 115,000 observations had hourly wage rates that appear extreme. It is impossible to know whether these observations are accurate and provide information or whether they just reflect errors, coding or other. To check for robustness, the wage residuals were re-estimated using an sample that deleted outliers, defined as those with hourly wage rates deviated from the occupation mean wage by more than three occupation-specific standard deviations of the wage. This eliminated 1258 observations. The wage regression was re-estimated both constraining coefficients to be the same across occupations and allowing the coefficients of the wage regression to differ for each occupation. The results are similar so table 3 column 8 reports the full-interaction version.

Once again, the results in table 3, column 8 support the theoretical prediction. Occupations that have more variance in their wage residuals also have higher average rates of occupational switching. The coefficient on the standard deviation of ε is almost twice as large when outliers are excluded.

Age and education have the expected effects on occupational switching, again reflecting comparative advantage. As an individual ages, more of the person's skills become occupation specific. It is very unlikely that an airline pilot who has been flying for twenty-five years is going to discover that his comparative advantage lies elsewhere. The skills developed over that period make it highly unlikely that he would be worth more in another occupation. Similarly, the more education a person has, the more occupation specific it is likely to be. First grade is completely general, whereas a Ph.D. in chemical engineering is quite specific to working in that field. The higher the level of education, the greater the expected value of δ for those in their initial placement.

4.1.2 Placebo

It is possible that other factors associated with variance in occupational wages could explain the fact that higher estimation-error occupations experience more switches. The most directly implied by the theory is comparative advantage. If there are costs of switching occupations, then the larger is the spread in the δ distribution, the more likely it is that a worker who finds himself in the wrong occupation will be willing to switch. This is examined directly below using the PSID data.

It is possible, however, to shed light on the specific implication that it is estimation error, in addition to or perhaps more than other factors, that account for differences in occupational switching. Variation in the wage can be broken up into the explained and unexplained part. The entire theory of statistical overconfidence is based on the unoservables, which is why the focus has been on the standard deviation of the the wage residuals. Although the observables such as age and education may themselves affect occupational switching (as just discussed), there is no reason why predicted variation in wages should affect occupational switching because the predicted variation should be known in advance and incorporated into the initial selection of occupations.

Occupations with high variation in the predicted component of wages are not necessarily ones in which overconfidence should be more prevalent. It is only the unobserved components that the theory predicts relate to overconfidence and therefore to the likelihood of switching out of an occupation. As a result, predicted variation acts as a placebo in this analysis. Only unexplained variation in wages, not predicted variation, should be correlated with occupational switching.

The results of the analysis are contained in table A1. No matter which of the versions of wage estimation is performed (outliers included: coefficients constrained to be the same and coefficients allowed to differ by occupation; outliers excluded: coefficients constrained to be the same and coefficients allowed to differ by occupation), the occupation-specific standard deviation of the fitted values never enters significantly in the occupational switching regressions. The t-ratios range from a low of -.01 to a high of 1.5, depending on the specification. Additionally, with a couple of exceptions, age, education and male are insignificant as well in these specifications. Note also that in column 5 of table A1, both the fitted and residual values from the wage equation are independent variables. The residual remains significant and the fitted values remain insignificant. Furthermore, the R-squared jumps for .38 (in column 2) to .59, suggesting the importance of the residuals in explaining occupational switching. It is the unobserved part of productivity, not the predictable part that is correlated with occupational switching. This is supportive of the statistical view of overconfidence.

4.2 Occupational Switching Declines in Comparative Advantage in Initial Occupation

The PSID is used to examine whether occupational switching follows the predictions of section 4.2, above. First, it is necessary to obtain estimates of θ_{ji} as defined in (15) for each individual because this is the variable that reflects person *i*'s comparative advantage in occupation *j*. To isolate θ_{ji} from α_i, it is necessary to observe the individual in multiple occupations. The occupation dummies pick up the average effect of occupation on output, here assumed to be reflected in the observed wage. It is also possible to take out the person effect, α_i, because a given individual is observed working in multiple periods. Deviating observations from the person mean eliminates the α_i and also all time-invariant observables. The θ_{ji} are identified by observing a given individual in multiple occupations throughout his or her career. Since the PSID follows respondents over prolonged time periods, a large fraction of the sample has switched occupations, which allows identification of the θ_{ji}. The interpretation of the effect of θ_{ji} in a proportional hazard model is that when an individual is in a high θ_{ji} occupation, he is less likely to switch out of it than when in a low θ_{ji} occupation.

Consider , for example, two individuals, one of whom has been employed in two occupations, the other has been employed in three occupations over their careers. The theory would predict this pattern under the following circumstances. Both individuals' first occupations are not ones for which they are well-suited. The first individual is a good match with his second occupation, meaning that he has a high value of θ. Because of that, he stays in that job and it becomes an absorbing state. The second individual has had three jobs. She is poorly suited to the first, meaning a low θ, poorly suited to the second, meaning a low θ again, but well-suited to the third. She switches out of the first two jobs and never switches out of the third job.

The estimation occurs in two steps. First, a panel wage regression is estimated. It includes occupational dummies, year dummies and person effects. Occupation-person effects, i.e., the match effect that is the $\theta_{ji,}$ is calculated as the average residual for a specific person within each occupation that the individual has held during the entire panel. It is the measure of comparative advantage. After person, occupation and match effects are taken out, the residual contains only the ξ_{jit} as defined in (15).

Theory predicts that the higher the $\theta_{ji,}$ the lower the likelihood that a worker will switch out of that occupation. This is comparative advantage. The results of estimating a proportionate hazard model are contained in table 4. Included in each model is tenure (in weeks). It is well-known that the hazard rate should decline with time on the job, as implied by both human capital and matching theory, which is what is incorporated in the discussion of learning in section 1.6.

The occupation mixed effect shows up as predicted in both specifications. Individuals with higher estimated values of θ are less likely to switch out of that occupation. The effects are stronger when observable factors age, sex, race and education are held constant.

As already discussed, a primary implication of the analysis is that occupations with higher wage residual variation should be those where overconfidence is greatest and where occupational switching is most pronounced. This shows up in the PSID analysis as it did in the CPS. The larger is the occupation-specific standard deviation of the residual, ξ, the more likely is occupational switching. The coefficient in column 2 on the standard deviation of ξ supports this view.

Furthermore, the more subtle point that relates to cross-occupational difference in the within-occupation variance of ξ_{jit} can also be tested. Recall that the logic is that for young workers, variation in ξ_{jit} has a lower signal-to-noise ratio than for more senior workers. As time progresses in a career, wage variation, even that in the residual, is less likely to reflect noise. It is quite possible when a worker is new that he is misjudged by management. But as time in the job passes, the wage is more likely to reflect true productivity. As a result, occupation-specific variations in ξ should be more likely to reflect comparative advantage in senior workers than in junior workers. This implies that the importance of the occupation-specific standard deviation of ξ in inducing an occupational switch should grow with tenure holding the pure tenure effect constant. Once again, this is borne out by the positive, albeit very small, effect of the interaction between the standard deviation of ξ and tenure on turnover probability. Occupations with high variation in ξ_{jit} see more switching and shorter employment durations, particularly when that variation is for senior workers.

The age factor is also consistent with a comparative advantage interpretation. Older workers are less likely to switch occupations, which again makes sense. After a worker has been in the workforce for many years, it is expected that she would be better suited to the occupation in which she resides than to others. Males, non-whites, and the less educated are more likely to change occupations rapidly.

5. Another Application

The statistical explanation of overconfidence suggests that people who purchase items may not use them as much as would be expected, given that they purchased the good in the first place. A personal example comes to mind. Each spring, I must decide whether to purchase a season pass at my most skied mountain for the following year. I must estimate my work schedule, the winter snow and my desire to go skiing eight months in advance. Some years I buy the pass and some years, I do not. Part of the reason for buying it is convenience. A pass avoids ticket window queues, etc. But even taking convenience into account, there are years when I regret having bought the pass because I use it too infrequently.

Of course, there is little problem with having decision-makers err some of the time. Any model with imperfect forecasting would yield this result. Different here, though, is among the subset of individuals who actually buy the good, regret is more common than not, which appears like bias in the purchasing decision. The model above can be reinterpreted to yield that result.

To make things simple, suppose that there are no common-value components, so in the case of the ski pass, weather, which is common, is not an issue and it is only unforseen travel schedules that differ across potential buyers. Then an individual's best estimate of future use can be characterized by the form used in (1) or

$$\hat{q}_A = q_A + \varepsilon_A$$

with a different interpretation than was provided earlier. Here, think of $\hat{q}_A$ as the estimated value of the product, in the example, the season ski pass to the individual in question when making the decision on whether or not to buy. The cost of the product can be thought of as q_B. Then, the individual should buy the product if $\hat{q}_A > q_B$.

This is a special case of the condition that determined the choice of occupation *A* over *B*. Here, ε_B is zero, but none of the results relies on a non-zero value for ε_B. Thus, by Proposition 2, those who purchase the good are, on average, overconfident of its value. In the case of the ski pass, those who buy the pass overestimate its value relative to cost. Too many who buy the season pass overestimate (on average) the value of the season pass relative to its cost. This has exactly the winner's curse flavor, but the result here relies on the idiosyncratic value of the good to the individual rather than the common value. Just as was the case of occupational choice, there is nothing that the individual can do to improve the decision rule. The difference between $\hat{q}_A$ and q_B is the best unbiased estimate of the net value of the good.

As mentioned earlier, not all of the results on overconfidence can be explained away in this manner. The statistical theory cannot explain laboratory experiment results where individuals estimate their performance and the average for the entire sample is an overestimate of true performance. But it is consistent with ex ante choices that turn out ex post to have been bad ones, even for the average individual in the group.

6. Conclusion

It is not surprising that decision-makers do not always make correct decisions. In a world filled with uncertainty, it is necessary to act on the information that is available at the time a decision is made and sometimes that information turns out to be incorrect. More surprising is that individuals who make decisions rationally using the appropriate rules of statistical decision-making may, on average, be incorrect, generally overestimating the value of a particular decision.

The analysis here relates primarily to occupational choice. Individuals who choose to go into a particular occupation are, on average, overconfident of their ability in that occupation, with the typical individual having productivity in the occupation that exceeds initial estimates. This is consistent with a purely statistical phenomenon. Conditioning on those who choose to enter an occupation selects on individuals who have a true comparative advantage in the occupation and on individuals who have positive measurement error, meaning that they overestimate their abilities. There is nothing that the decision-maker can do to remedy this problem because there is no common information that can be used to undo the error. The ex ante unbiased estimate is the one used.

A number of implications are derived and tested using the CPS and PSID data. It is found that occupations that have higher potential measurement error, as proxied by the occupation-specific standard deviation of the residual in the wage equation, exit those occupations more frequently. This is exactly what the theory predicts. Overconfidence tends to occur in occupations where estimation error is high. Workers who incorrectly enter the wrong occupation, remedy the situation by moving to another after they learn that they have made mistakes.

It is also found that workers move in response to their true comparative advantages. The PSID panel data permit estimation of occupation-specific match effects. Those with negative match values are more likely to exit the occupation.

These results and others suggest that occupational choice can be explained well by statistical theory, despite observations that appear to contradict unbiased decision-making.

References

Barber, Brad and Terrance Odean. “Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment,” *Quarterly Journal of Economics,* 116, 1, (Feb 2001), pp. 261-92.

Brown, Keith. “A Note on the Apparent Bias of Net Revenue Estimates for Capital Investment Projects,” *Journal of Financial Economics,* 29,4 (September 1974) pp. 1215-16.

Brunnermeir, Markus K. and Jonathan A. Parker. “Optimal Expectations,” *American Economic Review*, 95 (4), Sept. 2005.

Camerer, Colin and Dan Lavallo. “Overconfidence and Excess Entry: An Experimental Approach,” *The American Economic Review* ol. 89, No. 1 (Mar., 1999), pp. 306-318.

Croson, Racjhel and Uri Gneezy. “Gender Difference in Preferences,” *Journal of Economic Literature,* 47, 2, (2009), pp. 448-74.

Farber, Henry. “Mobility and stability: The dynamics of job change in labor markets,” in *Handbook of Labor Economics,* 3,B: 1999, pp. 2439-2483.

Gneezy, Uri, Muriel Niederle and Aldo Rustichini. “Performance in Competitive Environments: Gender Differences,” *Quarterly Journal of Economics,* 118, 3, (Aug 2003) pp. 1049-74.

Hoffman, Mitchel and Stephen V. Burks. “Training Contracts, Worker Overconfidence, and the Return from Firm-Sponsonred General Training,” Working Paper, University of Toronto, Oct. 2015.

Kent Daniel, David Hirshleifer and Avanidhar Subrahmanyam “ Investor Psychology and Security Market Under- and Overreactions” *Journal of Finance,* LIII, 6, (Dec 1998), pp. 1839-85.

Lazear, Edward P. “The Peter Principle: A Theory of Decline,” *Journal of Political Economy* 112:1, pt. 2 (2004): S141-S163.

Malmendier, Ulrike and Geoffrey Tate. “CEO Overconfidence and Corporate Investment, *Journal of Finance,* LX, 6, (December 2005), pp. 2661-2700.

Malmendier, Ulrike and Geoffrey Tate. “Who makes acquisitions? CEO overconfidence and the market's reaction,” *Journal of Financial Economics,* 89, 1 (July 2008), pp. 20-43.

Mitchell Hoffman and Stephen Burks. "Training Contracts, Worker Overconfidence, and the Return to Firm-Sponsored General Training" University of Toronto (2015).

Neal, Derek. "Industry-Specific Human Capital: Evidence from Displaced Workers," *Journal of Labor Economics*, 13, 4, (1995), pp. 653-77.

Shaw, Kathryn. "A Formulation of the Earnings Function Using the Concept of Occupational Investment," *Journal of Human Resources*, 19,3 (Summer 1984): 319-340.

Shaw, Kathryn L. "Occupational Change, Employer Change, and the Transferability of Skills," *Southern Economic Journal*, 53, 3: (Jan. 1987) 702-719.

Thaler, Richard. "From Homo Economicus to Homo Sapiens," *Journal of Economic Perspectives*, 14, 1 (Winter 2000), pp. 133-41.

Wilson, Robert "A Bidding Model of Perfect Competition," *Review of Economic Studies,* 1977, 44, 511-18.

Van den Steen, Eric. "Rational Overoptimism (And Other Biases). *American Economic Review,* 94, 4 (sept. 2004), pp. 1141-51.

Figure 1

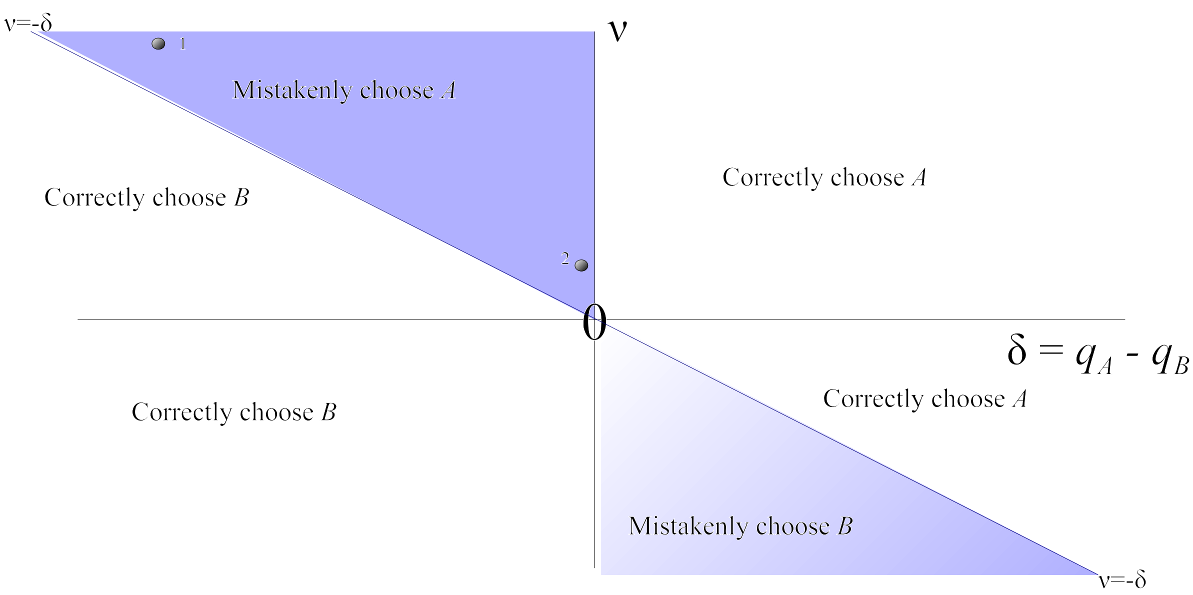


Figure 2

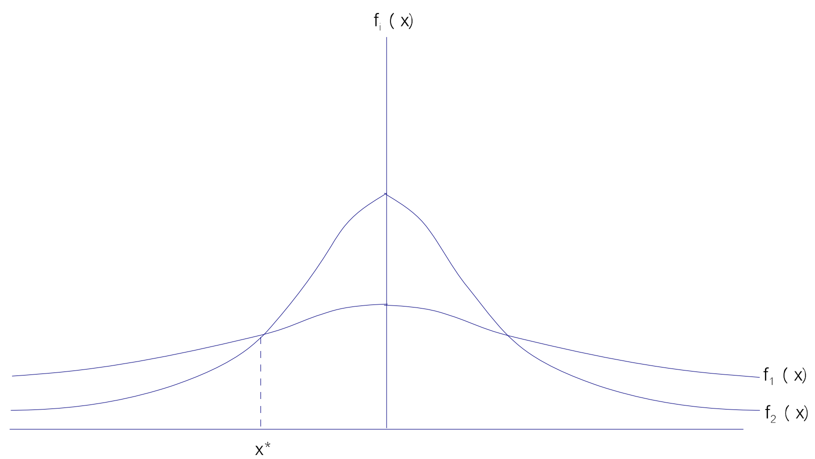
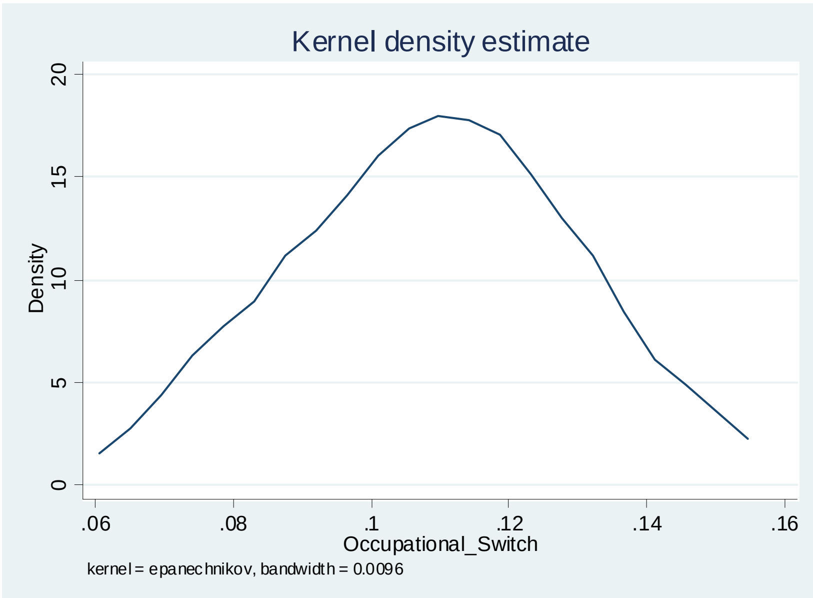


Figure 3

Kernel Density of Proportion Who Switch Occupation



Variable: Proportion who change occupation during the last year
Unit of analysis: Occupation
22 occupational observations

Table 2
Descriptive Statistics

Panel A:
Wage Data March 2012 use to estimate wage equation residuals

Variable	Mean	Standard Deviation	Additional Notes
Average Age	43.23	12.82	Ages 17 – 85
Average Education	40.90	2.53	codes used (31=less than first grade, ...45=professional school)
Male	.571	.494	
White	.83	1.24	
Usual Weekly Hours Worked	42.73	7.64	Full-time workers only (35 or more hours)
Weekly Earnings	9214.16	5047.68	
Hourly Wage	22.83	14.73	

Number of Observations=115,039

Panel B:
Raw Occupation Based Data from January, 2012

Variable	**Mean**	**Standard Deviation**	**Additional Notes**
Average Age	43.84	12.47	Ages 17 – 85
Average education	40.97	2.51	\codes used (31=less than first grade, ...45=professional school)
Race=White	.84		
Male	.56	.24	
Average Hourly Wage	17.29	4.04	
Tenure in Current Job	943.8	895.59	Tenure in Weeks
Different Work	.109	.312	

Number of Observations=39,992

Panel C:
Occupation Data created from January, 2012   22 observations

Variable	Mean	Standard Deviation	Additional Notes
Average Age	43.40	1.96	Ages 17 – 85
Average education	41.01	1.77	codes used (31=less than first grade, ...45=professional school)
Male	.58	.25	
Average Different Work	.11	.02	
Average Wage	24.30	8.10	

Number of Observations=22

Panel D: Panel Study of Income Dynamics

Variable	Mean	Standard Deviation	Additional Notes
Age	41.54	10.92	Ages 25-65
Male	.75	.43	
White	.66	.47	
Education	13.20	2.62	Education in years (1 = 1 year 2 = 2 years, etc)
Hourly Wage	14.99	17.35	
Weekly Hours	51.99	47.53	
Tenure	.93	1.56	Tenure in Years
Major Occupation Switch	.26	.44	
Detail Occupation Switch	.36	.48	

Number of Observations=37,891 on 7743 separate individuals

Table 3
Occupational Switching and Importance of Errors

Variable	**1** One wage equation	**2** One wage equation	**3** One wage equation	**4** One wage equation	**5** Occ. specific wage eq	**6** Occ. specific wage eq.	**7** Occ. specific wage eq.	**8** **Outliers removed** Occ. specific wage eq
Average Age	-.0033 (-1.66)		-.0056 (-3.15)	-.0059 (-2.96)		-.0056 (-3.22)	-.0057 (-2.86)	-.0054 (-3.02)
Average Education	-.0032 (-1.07)		-.0073 (-2.85)	-.0060 (-2.21)		-.0081 (-3.06)	-.0067 (-2.27)	-.0167 (-3.32)
Proportion Male	.0108 (0.69)		-.014 (-0.94)	-.0187 (-1.11)		-.0156 (-1.04)	-.0185 (-1.07)	-.0354 (-1.75)
Standard Deviation of Wage Residual (one equation)		000069 (0.58)	.0038 (3.20)	.0041 (2.58)				
Standard Deviation of Wage Residual (separate equation for each occupation)					.0049 (0.45)	.0037 (3.30)	.0036 (2.46)	.0068 (3.01)
R-squared	.31	.02	.57	.48	.01	.58	.47	.55
Weight	Obs per occup.	Obs per occup.	Obs per occup.	None	Obs per occup.	Obs per occup.	None	Obs per occup.

Number of observations = 22. t-ratios in parentheses

Table 4
Proportionate Hazard Model
Failure ≡ Occupational Switch

Variable	**1**	**2**
θ_{ji} (the person-occupation mixed effect)	.988 (-4.94)	.975 (-6.93)
Standard deviation of $\xi_{ji(t-1)}$ (the contemporaneous classic error after person, occupation and mixed effects are removed)		1.03 (6.10)
tenure (in weeks)	.999 (-4.29)	.998 (-3.08)
(Standard deviation of $\xi_{ji(t-1)}$) (tenure)		1.0002 (2.51)
age		.949 (-15.5)
male		1.27 (4.09)
white		.833 (-3.52)
education		.957 (-4.22)
χ^2	42.6	350.1

z values are in parentheses. 14,130 observations, with standard errors from clustering at the person level.

Appendix

Table A1
Placebo Using Fitted Values

Variable	**1** One set of coefficients	**2** Occupation-specific coefficients	**3** One set of coefficients Outliers removed	**4** Occupation-specific coefficients Outliers removed	**5** Occupation-specific coefficients
Average Age	-.0033 (-1.57)	-0.0045 (-2.10)	-.0032 (-1.52)	-.0049 (-2.22)	-.0058 (-3.16)
Average Education	-.0032 (-1.09)	-.0083 (-1.79)	-.0034 (-1.13)	-.0095 (-1.92)	-.0095 (-2.44)
Proportion Male	.0108 (0.67)	-.00041 (-0.02)	.0107 (0.66)	-.0054 (-0.29)	-.0178 (-1.12)
Standard Deviation of Fitted Values	-.000069 (-0.01)	.0044 (1.35)	-.0011 (-0.18)	.0052 (1.50)	.0015 (.050)
Standard Deviation of Wage Residual (separate equation for each occupation)					.0035 (2.85)
R-squared	.31	.38	.31	.39	.59
Weight	Obs per occupation	Obs per occupation	Obs per occupation	Obs per occupation	Obs per occupation

Number of observations = 22. t-ratios in parentheses.