Identifying idiosyncratic career taste and skill with income risk

Daniel Barth
Office of Financial Research, U.S. Department of the Treasury

Stephen H. Shore
The Robinson College of Business, Georgia State University

Shane T. Jensen
Department of Statistics, The Wharton School, University of Pennsylvania

How important to well-being is choosing a career with the right fit? We estimate a model of career choice in a setting where we observe the income risk of chosen careers and the risk aversion of the people who choose them. The key parameter of interest representing the importance of idiosyncratic taste and skill in career choice is identified from the shift in the distribution of income risk with risk aversion. We document that those who self-identify as risk tolerant are more likely to have volatile incomes. However, this correlation is far from perfect. The model gives this weak correlation an economic interpretation: idiosyncratic fit is an important determinant of career choice. We separate idiosyncratic career taste from skill using the pay gap between high- and low-income-risk people with high and low risk aversion.

Keywords. Occupational choice, career choice, income risk, idiosyncratic taste and skill.

JEL classification. J24.

1. Introduction and motivation

In this paper, we use a model of career choice rooted in expected utility to estimate the importance of idiosyncratic career taste and skill. We develop and estimate a model in

which people differ in their risk tolerance and careers differ in their riskiness. Individuals have idiosyncratic taste for, and skill and correspondingly higher pay in, some careers over others. They choose a career only once. When making this choice, workers observe and understand the attributes of each career option (including its riskiness), but income uncertainty is only resolved after choosing a career.

If income risk were the only salient feature of career choice, we should expect to see perfect sorting of the most risk-tolerant people into the riskiest careers, with compensating wage differentials inducing people to enter riskier careers. Empirically, we document that those who self-identify as risk tolerant are more likely to have volatile incomes, but this correlation is far from perfect. Our model of optimal career choice gives this correlation an economic interpretation. Both idiosyncratic taste for and skill in a particular career will lead an individual to deviate from perfect sorting, choosing instead a career that is a particularly good fit. Expected utility quantifies the benefits of choosing a career with a better fit, which must be large enough to compensate the individual for risk.

Our paper builds on a large body of work on the economics of occupational choice. Particularly relevant to our setup is the two-sector model developed by Roy (1951) and later extended by Heckman and Honore (1990). In these two-sector models, workers differ in sector-specific ability. While the original Roy model has no nonpecuniary sector preference, more recent work generalizing the Roy model allows for nonpecuniary differences. (Cunha, Heckman, and Navarro (2005), Cunha and Heckman (2008)). All jobs within each coarse sector are assumed to be the same. Sector-specific ability is identified from the choice of sector and the cross-sectional distribution of wages. Keane and Wolpin (1997) is another important paper on sector choice. In their model, individuals are endowed with an innate ability in each of five sectors and choose their sector in each period. This choice will depend on an individual's sector-specific ability, their sector-specific experience, nonpecuniary taste for sectors, and sector-wide shocks that change over time but are common to all individuals in a sector.[1] Changes in the choice of sector as well as income changes within sector and over time are used to infer the distribution of sector-specific ability and sector-wide shocks. Unlike the small number of coarse sectors considered by these models, our paper assumes a continuum of career options. These papers use the dispersion of wages for identification, which implies they must rule out idiosyncratic taste for a particular sector as well as within-sector differences in job attributes. We identify idiosyncratic career fit from the joint distribution of income risk and risk aversion—but not the distribution of the level of income—and therefore do not need to make these restrictions. Instead, we must make the strong assumption that risk aversion is an attribute of individuals whereas income risk is an attribute of the jobs they choose. Our model is also unable to accommodate the dynamics considered in Keane and Wolpin (1997).

Expected utility allows us to use these risk measures to cardinalize a logit model of career choice, so that all estimates can be expressed in terms of their (log) certainty equivalents. Logit models have long been used in occupational choice literature to study

---

[1]In the extended model, nonpecuniary sector tastes are common to all individuals and do not vary with time or individual characteristics, except for the military sector in which the nonpecuniary reward varies with age.

the relative importance of covariates on choice (Boskin (2004), Field (2009)). Keane and Wolpin (1997) provide an alternative—and very different—cardinalization within a model of optimal educational investment and subsequent choice from broad career categories. Their model explores how schooling and work affect the development of occupation-specific skills, and makes it possible to explore the effects of college subsidies on behavior.[2]

There is already a large empirical and theoretical literature on the relationship between income risk and career choice.[3] We aim to use this well studied and well understood risk–return relationship to identify a seemingly unrelated parameter, the dispersion of idiosyncratic career fit (combining the taste for and skill in particular careers). This dispersion in fit is identified from shifts in the distribution of income risk as risk aversion changes. The tighter is the correlation between income risk and risk aversion, the lower is the implied dispersion of idiosyncratic taste and skill. Using self-reported risk tolerance and estimates of income volatility—which we use to proxy for income risk—from the Panel Study of Income Dynamics (PSID) data, we estimate a lower bound of 36% of income on the standard deviation of these idiosyncratic factors.[4]

Our model shows how to identify the joint importance of idiosyncratic taste *and* skill from the joint distribution of income risk and risk aversion. We can separate taste *from* skill using data on income levels. When a risk-averse person chooses a career with substantial income risk, on average he must be compensated in some way for this risk. Such compensation could be in the form of higher idiosyncratic skill in this career (and therefore higher pay) or higher idiosyncratic taste for this career (and therefore higher enjoyment). To the degree that idiosyncratic skill dominates idiosyncratic taste, we should see risk-averse people with high income risk earning more than risk-averse people with low income risk. By comparing this high-income-risk versus low-income-risk pay gap for those with high and low risk aversion, we can difference out market-wide compensating differentials for income risk and isolate the fraction of idiosyncratic career fit due to taste.

[2]Lazear (2004) provides a model of skill development for entrepreneurs, whose endowment of a general set of skills can be augmented by schooling. Our model abstracts from the schooling decisions considered in Keane and Wolpin (1997) and Lazear (2004) so as to focus on the direct link that a simpler model implies between the importance of idiosyncratic fit and data on income risk and risk tolerance.

[3]Weiss (1972) and Pacios-Huerta (2003) identify the risk attributes associated with various demographic groups. Hartog et al. (2003), Hartog and Vijverberg (2007), and Berkhout, Hartog, and Webbink (2010) estimate "risk-augmented Mincer (1974) equations" (surveyed in Hartog (2011)) to identify the empirical risk–return relationship; Berkhout, Hartog, and Webbink (2010) allow this relationship to vary with demographic characteristics. Flyer (2004) identifies the impact of income risk on occupational choice (and Harris and Weiss (1984) provide a theory of such a choice); Diaz-Serrano and Hartog (2006) do the same for educational choice (with Levhari and Weiss (1974) and Krebs (2003) providing theory on this). Fuchs-Schündeln and Schündeln (2005) use occupational choice as an instrument for income risk. Dillon (2016) estimates a model of occupational choice with compensating wage differentials for income risk.

[4]Meghir and Pistaferri (2004) and Alvarez, Browning, and Ejrnaes (2010) show that income volatility differs across individuals. Jensen and Shore (2011, 2015) estimate the distribution of ex ante, individual-specific volatilities in the population. The 1996 PSID includes a measure of self-reported risk tolerance, elicited from a survey asking the individual if he/she would take a series of hypothetical income gambles (Barsky et al. (1997), Sahm (2012), Kimball, Sahm, and Shapiro (2008, 2009).

The rest of the paper is organized as follows: Section 2 presents the model; Section 3 discusses the data used to estimate the model; Section 4 presents the estimation strategy; Section 5 offers estimation results; Section 6 presents our policy counterfactuals; and Section 7 concludes the paper. Replication files are available in a supplementary file on the journal website, http://qeconomics.org/supp/424/code_and_data.zip.

2. Model

We present a model in which individuals choose from a set of career options. Each career option has a quantity of income risk, a typical pay for that career, and other non-pecuniary attributes. Each individual has a preference for income risk, an overall ability (which affects pay in all careers equally), and other attributes. There is a distribution of career options and a distribution of people in the population. In addition to these innate traits of careers and individuals, there are traits specific to an individual in a given career. Some individuals have an idiosyncratic taste for some careers over others, and some individuals are idiosyncratically better (have higher productivity and therefore higher pay) in some careers than others. An individual's ability in a specific career will reflect the overall ability he/she brings to all careers and his/her idiosyncratic ability in that specific career. From the set of career options, each individual makes a one-time, irrevocable choice of the best career. Then the career-specific income shock is realized.

2.1 Setup

2.1.1 *Careers*   Career options are indexed by $c \in \{1, \ldots, N_C\}$. Career c has four attributes, $X_c^C \equiv \{\sigma_c^2, y_c^C, x_c^{CO}, x_c^{CU}\}$, where σ_c^2 is a measure of the income risk in career c,[5] y_c^C is the career-specific component of log pay in career c, $x_c^C \equiv [x_c^{CO}; x_c^{CU}]$ is a vector of covariates or attributes of career c, x_c^{CO} are the attributes observable to the econometrician and to workers, and x_c^{CU} denotes the set of attributes observable to workers but not to the econometrician. The industry in which a career resides or the average hours worked by employees are examples of typically observed career attributes (contained in x_c^{CO}), whereas the noisiness of a career is a typically unobserved career attribute (contained in x_c^{CU}).

2.1.2 *People*   People are indexed by $i \in \{1, \ldots, N_I\}$. Person i has four attributes, $X_i^I \equiv \{\gamma_i, y_i^I, x_i^{IO}, x_i^{IU}\}$, where γ_i is a measure of risk aversion for person i, y_i^I is the person-specific component of log pay (general ability or productivity) for person i, $x_i^I \equiv [x_i^{IO}; x_i^{IU}]$ is a vector of attributes of person i, x_i^{IO} is the set of attributes observable both to the econometrician and to workers in the model, and x_i^{IU} is the set of attributes observable to workers in the model but not to the econometrician. Math skill is an example of a typically unobserved individual attribute (contained in x_i^{IU}), whereas age, gender, race, and education are typically observed attributes (contained in x_i^{IO}).

---

[5]The assumption that income risk is associated with a career, not an individual, is a strong one. Jacobs, Hartog, and Vijverberg (2009) discusses the biases associated with making this assumption in reduced-form risk–return estimation.

2.1.3 *Individual-career-specific fit*　We assume that some careers are a better fit for some people than others. The fit of person i in career c is characterized by two attributes, $X_{i,c}^{\varepsilon} \equiv \{y_{i,c}^{\varepsilon}, l_{i,c}^{\varepsilon}\}$, where $y_{i,c}^{\varepsilon}$ is the individual-career-specific component of log pay (idiosyncratic productivity) of person i in career c and $l_{i,c}^{\varepsilon}$ is an individual-career-specific measure of idiosyncratic enjoyment of person i in career c.

We require that $X_{i,c}^{\varepsilon}$ and $X_{i,c'}^{\varepsilon}$ be identically distributed and independent of one another when $c \neq c'$, and also that $X_{i,c}^{\varepsilon}$ be independent of X_i^I and X_c^C. Independence across c is the standard "independence of irrelevant alternatives" assumption present in multinomial logit settings. Independence across i is also required for inference when we estimate the model on data.

2.1.4 *Preferences*　The model that follows assumes risk-averse, expected utility maximizing individuals who care about *stochastic* income Y and career enjoyment L. Individuals have Cobb–Douglas preferences over Y and L, and expected utility preferences over the composite, v. Individual i in career c has an expected utility of

$$Eu(i,c) = E\left[\frac{v_{i,c}^{1-\gamma_i}}{1-\gamma_i}\right], \tag{1}$$

$$v_{i,c} \equiv Y_{i,c}^{1-\alpha} L_{i,c}^{\alpha}, \tag{2}$$

$$\ln Y_{i,c} \equiv y_c^C + y_i^I + y^x(x_i^I, x_c^C) + y_{i,c}^{\varepsilon} + \sigma_c \xi - \frac{1}{2}\sigma_c^2, \tag{3}$$

$$\ln L_{i,c} \equiv l^x(x_i^I, x_c^C) + l_{i,c}^{\varepsilon}. \tag{4}$$

Composite felicity $v_{i,c}$ is a Cobb–Douglas function of income $Y_{i,c}$ and career enjoyment $L_{i,c}$. The relative importance of income and career enjoyment is determined by α. We impose an elasticity of substitution of 1 and do not allow for heterogeneity in α.[6]

For simplicity, we assume a one-period model in which income Y is merely equal to consumption. Log income in equation (3) is the sum of the career-specific component of pay (y^C), including a premium for size, risk, or non-pecuniary attributes, the individual-specific component of pay or ability (y^I), the effect of the interaction of individual- and career-specific covariates on pay ($y^x(x_i^I, x_c^C)$), the individual-career-specific component of pay ($y_{i,c}^{\varepsilon}$), the individual's career-specific productivity, and the realization of a stochastic income shock ($\sigma_c \xi - \frac{1}{2}\sigma_c^2$). The random variable ξ is modeled as a standard normal variable, so that $\sigma_c \xi - \frac{1}{2}\sigma_c^2$ has an exponentiated expectation equal to 1. Log enjoyment in equation (4) is the sum of the effect of the interaction of individual- and career-specific covariates on enjoyment ($l^x(x_i^I, x_c^C)$) and individual-career-specific enjoyment (l^{ε}).

The model explicitly assumes that individuals never switch careers. Although this is surely a restrictive assumption, we offer two caveats. First, in our model a *career* is defined as an income path, which is broader than the traditional definition based on

---

[6]The Cobb–Douglas structure in equation (2) (with an elasticity of substitution of 1) allows for a clean analytic solution in which α need not be estimated. Absent data on career enjoyment, there is no straightforward way to estimate the elasticity of substitution or α directly.

specific task compositions or industry codes. Our model requires that the "career" an individual chooses has time-invariant values for income risk, expected pay, and idiosyncratic fit. It does *not* require that individuals remain in the same industry or occupation code. Second, many individuals do not switch careers even in the narrow sense. The rate of job-to-job transitions using the Shimer (2005) lower bound estimate is less than 10% per year (as calculated by Mukoyama (2014)), and during some periods is close to zero. Further, job-to-job transition rates are likely to underestimate career changes, because many individuals will switch jobs within the same career. Finally, job-to-job transitions have declined steadily since the early 2000s (Mukoyama (2014)). The benefit of this restriction on career switching is that it allows us to present a parsimonious model, identified from the joint distribution of risk aversion and income risk.

2.1.5 *Career value*   Plugging equations (2), (3), and (4) into equation (1), evaluating the expectation, and transforming yields a log income certainty equivalent measure of the value of career c to person i:

$$
\begin{aligned}
V(i,c) &\equiv \frac{\ln\big((1-\gamma_i)Eu\big)}{(1-\alpha)(1-\gamma_i)} \\
&= y_i^I + y_c^C + y^x\big(x_i^I, x_c^C\big) + y_{i,c}^\varepsilon + \frac{\alpha}{1-\alpha}\big(l_{i,c}^\varepsilon + l^x\big(x_i^I, x_c^C\big)\big) - \frac{1}{2}(\alpha + \gamma_i - \alpha\gamma_i)\sigma_c^2.
\end{aligned}
\tag{5}
$$

Individuals will choose the career with the highest V. Note that person-specific ability (y^I) has no impact on the career chosen; it merely shifts the value of all careers equally. There is no way to separate large α from large $l_{i,c}^\varepsilon$ in equation (5), which informs the transformations

$$
\widetilde{l^x\big(x_i^I, x_c^C\big)} \equiv \frac{\alpha}{1-\alpha} l^x\big(x_i^I, x_c^C\big); \qquad \widetilde{l_{i,c}^\varepsilon} \equiv \frac{\alpha}{1-\alpha} l_{i,c}^\varepsilon.
\tag{6}
$$

We also make a transformation to risk aversion,

$$
\widetilde{\gamma}_i \equiv \alpha + \gamma_i - \alpha\gamma_i,
\tag{7}
$$

to match the object estimated in the PSID. The term γ is the curvature of the utility function with respect to felicity, v. By contrast, hypothetical gambles in the PSID seek to estimate the curvature of the utility function with respect to income, which in our model is given by $\widetilde{\gamma} = \alpha + \gamma_i - \alpha\gamma_i$.

We can then rewrite the value of each career as

$$
V(i,c) = y_i^I + y_c^C + y^x\big(x_i^I, x_c^C\big) + \widetilde{l^x\big(x_i^I, x_c^C\big)} - \frac{1}{2}\widetilde{\gamma}_i\sigma_c^2 + y_{i,c}^\varepsilon + \widetilde{l_{i,c}^\varepsilon}.
\tag{8}
$$

If we group pecuniary and nonpecuniary idiosyncratic terms and also group pecuniary and nonpecuniary covariate terms as

$$
\varepsilon_{i,c} \equiv y_{i,c}^\varepsilon + \widetilde{l_{i,c}^\varepsilon} \quad \text{and} \quad v\big(x_i^I, x_c^C\big) \equiv y^x\big(x_i^I, x_c^C\big) + \widetilde{l^x\big(x_i^I, x_c^C\big)},
\tag{9}
$$

we arrive at our final expression for the log dollar-value certainty equivalent of a career:

$$V(i, c) = y_i^I + y_c^C + v(x_i^I, x_c^C) - \frac{1}{2}\widetilde{\gamma}_i \sigma_c^2 + \varepsilon_{i,c}. \tag{10}$$

Equation (10) gives career choice a standard, random utility, multinomial choice structure (McFadden (1974)). Individuals choose the career that gives them the highest utility, which depends on career attributes that affect everyone equally (y^C), career attributes that affect different individuals differently (observable and unobservable covariates $v(x_i^I, x_c^C)$ and the utility cost of income risk σ_c^2, which depends on risk aversion $\widetilde{\gamma}_i$), and an error term ($\varepsilon_{i,c} \equiv y_{i,c}^\varepsilon + \widetilde{l}_{i,c}^\varepsilon$). What is unique here is that our economic model provides a cardinalization, so that when a career's value is expressed in terms of log-certainty equivalent income, the coefficient on $\sigma_c^2 \times \widetilde{\gamma}_i$ is $(-\frac{1}{2})$. This cardinalization means that coefficient estimates and the standard deviation of the error term now have an absolute, log-income-equivalent meaning.

2.2 *Stylized model without idiosyncratic career preference*

We begin by considering a model without idiosyncratic career taste, skill, or covariates, so that $\varepsilon_{i,c}$ and $v(x_i^I, x_c^C)$ are zero. All individuals with the same $\widetilde{\gamma}$ are indifferent among any options they choose with positive probability. This implies a weakly (negatively) monotonic relationship between risk aversion and income risk choice. In this case, we should never see a more risk-tolerant person choosing less income risk. We consider a continuum of careers on some range of σ_c^2, which have full support in the sense that all careers are chosen by someone. Let $\widetilde{\gamma}(\sigma_c^2)$ be the risk aversion of the person who chooses income risk σ_c^2.[7]

At an interior optimum, the individual's first order condition requires that[8]

$$\frac{dy^C}{d\sigma_c^2} = \frac{1}{2}\widetilde{\gamma}. \tag{11}$$

If equation (11) must hold for each $\{\sigma_c^2, \widetilde{\gamma}(\sigma_c^2)\}$ pair and we know the risk aversion of the marginal individual for each σ_c^2, then we can trace out y^C as

$$y_c^C = y_0^C + \frac{1}{2}\int_0^{\sigma_c^2} \widetilde{\gamma}(x)\,dx. \tag{12}$$

Here, y_0^C is the log pay for a risk-free career. Note the strong assumptions needed here, namely that all individuals face the same risk–return menu (up to an ability intercept, which can differ across individuals). A graphical depiction of this menu is given in Figure 1.

---

[7]Because of the full support assumption, each σ_c^2 is chosen by someone and therefore maps to a $\tilde{\gamma}$, though a measure zero set of σ_c^2 values may map to multiple $\tilde{\gamma}$ values. The fact that the number of such points is of measure zero means that the values we use here do not affect the risk–return menu.

[8]Obtained by differentiating expected utility in equation (10) with respect to σ_c^2, setting equal to zero, and rearranging terms.
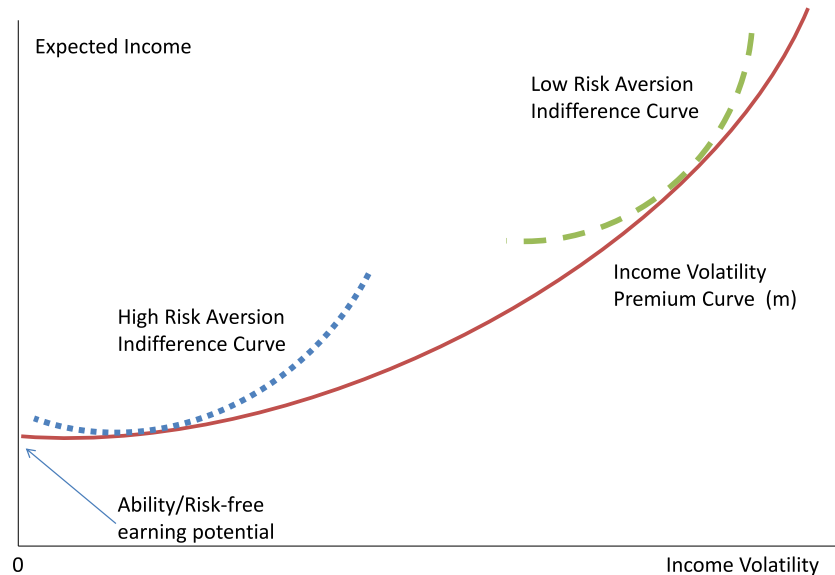
FIGURE 1. Sorting of the risk tolerant into volatile careers. This figure presents a stylized risk–re-
turn menu. The solid curve represents the menu of risk–return options, which has a positive
slope that reflects the increased compensation for taking higher income risk. The dashed curve
represents the indifference curve of a more risk-tolerant individual, and the dotted curve rep-
resents the indifference curve of a more risk-averse individual. Tangencies reflect the optimal
decision of each individual for the given risk–return menu.

2.3 *The distribution of chosen careers*

The multinomial choice framework we introduce becomes tractable when we work with
extreme value errors, which gives equation (10) a logit structure. Let r refer to the set of
careers for person i in a rectangle on the $\{y^C + v(x_i^I, x^C), \sigma^2\}$ plane; let s_r be the share
of careers that fall in region r. We assume that the number of careers in each region r is
large enough that $\max_{c \in r} \varepsilon_{i,c}$ has an extreme value distribution (with scale parameter β)
or that each $\varepsilon_{i,c}$ has an extreme value distribution (with scale parameter β) to begin
with.[9] When $\varepsilon_{i,c}$ has an extreme value distribution, $\text{var}(\varepsilon_{i,c}) = \beta^2 \pi^2/6$; when its maxi-
mum does, $\text{var}(\varepsilon_{i,c}) \propto \beta^2 \pi^2/6$.[10] Consider the choice among careers c in range r. Taking

---

[9]In this case, we require that the cumulative distribution function (cdf) of $\varepsilon_{i,c}$ be twice differentiable
de Haan and Ferreira (2006). The normal and exponential distributions are examples of such distributions.
Coupled with the independence assumptions from Section 2.1.3, this implies that the maximum of $\varepsilon_{i,c}$ has
an extreme value distribution (of Type I, Gumbel).

Note that we make a homoskedasticity assumption here, that all individuals have the same scale parame-
ter, β. We have no way to separate extreme-value errors with a common scale parameter (homoskedasticity)
from some other distribution with heterogeneous scale parameters (heteroskedasticity). Both could imply
the same unconditional distribution of errors across individuals.

[10]There are two technical advantages to an extreme-value approach. First, increasing the number of
careers affects only the location parameter μ, shifting the whole distribution up while leaving its shape
(governed by parameter β) unchanged. As a result, we can normalize out μ, so that we need not take a
position on the total number of careers N_C (an idea without precise meaning) to identify the model. Second,

the size of the rectangle to zero, within-range differences between careers c in X^C will be trivially small. As a result, if the individual chooses a career from within range r, it will be the one with the highest $\varepsilon_{i,c}$.

Given the extreme value distribution, the expected value of the chosen career is

$$
\begin{aligned}
\overline{V(X_i^I)} &\equiv E\big[V(i,r) \mid V(i,r) > V(i,q), \forall q \neq r\big] \\
&= \mu + \beta\gamma_{em} + y_i^I + \beta\ln\bigg(\sum_q s_q e^{(y_q^C + v(x_i^I, x_q^C) - \frac{1}{2}\widetilde{\gamma}_i \sigma_q^2)/\beta}\bigg).\text{[11]}
\end{aligned}
\tag{13}
$$

The probability that an individual's preferred career will lie in range r is

$$
\text{prob}\big(V(i,r) \geq V(i,q), \forall q \neq r\big) \propto s_r e^{(y_r^C + v(x_i^I, x_r^C) - \frac{1}{2}\widetilde{\gamma}_i \sigma_r^2)/\beta}.
\tag{14}
$$

The full derivations of equations (13) and (14) are provided in Appendix A.1. We rewrite equation (14) by taking the size of each range to zero to characterize the distribution of chosen careers[12]

$$
f\big(X^C \mid X^I\big) \propto f\big(X^C \mid \widetilde{\gamma} = 0, X^I\big) e^{-\frac{1}{2}\widetilde{\gamma}\sigma^2/\beta},
\tag{15}
$$

$$
f\big(X^C \mid \widetilde{\gamma} = 0, X^I\big) \propto f^C\big(X^C\big) e^{(y^C + v(x^I, x^C))/\beta}.
\tag{16}
$$

We do not observe all elements of X^C or X^I, and therefore must integrate out unobservables. This requires that we make several assumptions about the unobservable attributes of careers and individuals. To integrate out unobservable attributes of careers, we require that $\widetilde{\gamma}$ does not affect the expected payoff of some risk levels more than others, so that $E[e^{(y^C + v(x^I, x^C))/\beta} \mid X^I, \sigma^2, x^{CO}]$ does not vary with $\widetilde{\gamma}$. If we take the example from Section 2.1.1 of noisiness as an unobservable career attribute, workers' distaste for noisiness across careers with different income risk levels must be unaffected by their risk aversion. To integrate out unobservable attributes of individuals, we make the assumption that the expected value of careers at various income risk levels cannot be differentially affected by individual unobservables for different levels of risk aversion. In other words, $E[e^{((y^C + v(x^I, x^C))/\beta)} \mid \sigma^2, x^{CO}, x^{IO}]$ should not vary with $\widetilde{\gamma}$. In the example from Section 2.1.2 of math aptitude as an unobservable individual attribute, the benefits of math aptitude by career income risk cannot depend on risk aversion. Appendix A.2 presents these assumptions formally, and shows how we use them to integrate out individual- and career-specific unobservables.

There are a variety of reasons to think that these assumptions might be violated. For example, worker conscientiousness may be a confounding factor. Conscientiousness is

---

results are not dependent on a particular parametric shape for the distribution of individual-career-specific shocks, $\varepsilon_{i,c}$.

[11]Here μ and β are the location and scale parameters of the extreme value distribution, and $\gamma_{em} \approx 0.577$ refers to the Euler–Mascheroni constant. Summation takes place over all rectangles q on the $\{y^C + v(x_i^I, x^C), \sigma^2\}$ plane.

[12]In the limit, sums become integrals and the share of career options in each range (s_r) becomes the distribution of career options ($f^C(X^C)$).

unobservable and may be positively correlated with risk aversion. Conscientious work-
ers might make any career less risky. In this case, estimates of β would be biased down-
ward as we would observe a *stronger* correlation between risk aversion and income risk
in the data.

After integrating out unobservables, equation (15) becomes equation (17), which is
the key structural equation that we take to the data:

$$f\big(\sigma^2 \mid \widetilde{\gamma}, x^{IO}, x^{CO}\big) \propto f\big(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO}\big)e^{-\frac{1}{2}\widetilde{\gamma}\sigma^2/\beta}, \tag{17}$$

where

$$\begin{aligned}
&f\big(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO}\big) \\
&\quad \propto f^C\big(\sigma^2|x^{CO}\big)\frac{f^C\big(x^{CO}\big)}{f\big(x^{CO}\big)}E\big[e^{(y^C+v(x^I,x^C))/\beta} \mid \sigma^2, x^{IO}, x^{CO}, \widetilde{\gamma} = 0\big].
\end{aligned} \tag{18}$$

The critical insight from equation (17) is that the distribution of risk choices made
by risk-averse people $f(\sigma^2|\widetilde{\gamma})$ is completely determined by the distribution of choices
made by risk-neutral people $f(\sigma^2|\widetilde{\gamma} = 0)$ and a risk shift determined by a single parame-
ter β. Although there are more sophisticated models of occupational choice (e.g., Keane
and Wolpin (1997)), the key advantage of our model is that it implies a simple and intu-
itive structure. Each conditional distribution $f(\sigma^2|\widetilde{\gamma})$ for a given $\widetilde{\gamma}$ is merely an exponen-
tial shift of another such conditional distribution for another $\widetilde{\gamma}$. The degree of that shift
is governed by β, which is proportional to the standard deviation of the idiosyncratic
individual-specific-career taste and skill shocks. For large shocks (high β), the shift is
modest and conditional distributions look more similar to one another (and more simi-
lar to the distribution of careers, f^C). For small shocks (low β), the shift is more substan-
tial and conditional distributions for high and low $\widetilde{\gamma}$ become more different (and each
becomes more concentrated around the "best" choice for that $\widetilde{\gamma}$).

Note that this model is highly overidentified when we observe the joint distribution
of σ^2 and $\widetilde{\gamma}$. The model is agnostic about the risk distribution chosen by risk-neutral
people $(f(\sigma^2 \mid \widetilde{\gamma} = 0))$.[13] However, $f(\sigma^2 \mid \widetilde{\gamma} = 0)$ and a single parameter (β) completely
determine the risk distribution $f(\sigma^2 \mid \widetilde{\gamma})$ for all $\widetilde{\gamma}$.

---

[13]The model implies that a risk-neutral person's career choices (with distribution $f(X^C \mid \widetilde{\gamma} = 0, X^I)$, as
shown in equation (16)) will be proportional to the frequency of career options $(f^C(X^C))$. *Ceteris paribus*,
a risk-neutral person will be twice as likely to choose a career with a given set of attributes if twice as many
careers have those attributes. A risk-neutral person is also more likely to choose careers with a higher
career-specific component of pay and enjoyment $(y^C + v(x^I, x^C))$. These career-specific attributes domi-
nate career frequency when idiosyncratic career taste and skill are relatively unimportant ($\beta \to 0$). Without
idiosyncratic career fit, risk-neutral people will merely choose the career with the highest $y^C + v(x^I, x^C)$;
the distribution of risk choices will be extremely tight around the best choice. However, as the importance
of idiosyncratic career fit increases ($\beta \to \infty$), careers are chosen only in proportion to their frequency; the
distribution of choices becomes as diffuse as the distribution of career options f^C. We should be unsur-
prised to see that individual-specific ability (y^I) does not affect career choice as it increases the benefit of
all careers equally.

TABLE 1. Summary statistics.

	Mean	St. Dev.	Min	Max
Age (years)	42.4	8.0	27	60
Education (years)	14.1	2.3	0	17
Married	85.4%	.	.	.
Black	3.2%	.	.	.
Annual income (2005 $s)	$56,284	$55,194	0	$753,042
Family size	3.2	1.3	1	9

Note: This table summarizes data from the 1,490 male household heads in the sample in 1996. Each observation is weighted by its PSID supplied sample weight. The variable "black" is calculated as of 1997.

3. DATA

Our data are the core sample of the Panel Study of Income Dynamics (PSID). The PSID was designed as a nationally representative panel of U.S. households (Hill (1991)); it provides annual or biennial labor income spanning the years 1968–2005. Restricting ourselves to male household heads aged 22–60[14] gives us 52,181 observations on 3,041 individuals with 17 years of recorded data per individual on average. Furthermore, we restrict the sample to individuals with income (and therefore volatility) values in 1991–1996, with risk-tolerance responses recorded in the 1996 wave, and nonzero population weights.[15] There are 1,490 individuals who meet this criteria. Summary statistics about the demographics of this group in 1996 are shown in Table 1.

3.1 *Risk aversion*

In 1996, the PSID included a series of survey questions that aimed to elicit estimates of risk tolerance. Respondents were asked a series of questions about hypothetical income gambles. The first such question was, "[Y]ou are given the opportunity to take a new, and equally good, job with a 50–50 chance that it will double your income and spending power. But there is a 50–50 chance that it will cut your income and spending power by a third. Would you take the new job?" If the respondent answered "yes," she was asked the same question again though she faced the risk that her income would be cut by one-half instead of one-third; if she answered "no," the question was again the same but she faced the risk that her income would be cut by only one-fifth. For those people who answered yes or no to both questions, one additional question was asked with an income cut of three-quarters or one-tenth, respectively. Based on the responses to these questions, individuals were placed into one of four risk-tolerances bins.

In our model, estimated risk tolerance corresponds to the value $1/\widetilde{\gamma}$, not $1/\gamma$. This is because the hypothetical gambles in the PSID seek to estimate the curvature of the

[14]Age restrictions are standard in the income dynamics literature, though exact age ranges vary slightly: Gottschalk and Moffitt (2002) (22–59), Meghir and Pistaferri (2004) (25–55), and Abowd and Card (1989) (21–64).

[15]Individuals who entered the sample through marriage are assigned a zero weight in the PSID. We keep these individuals in the sample by assigning them their spouses' weights.

utility function with respect to income, which in our model is given by $\widetilde{\gamma} = \alpha + \gamma_i - \alpha\gamma_i$. Further, the risk-tolerance estimates in the PSID are especially well suited for our model, as they hold all non-income considerations fixed, meaning the effects of risky career income can be separated from the effects of career enjoyment.

3.2 *Income volatility*

Using data from the PSID, we calculate two "off-the-shelf" measures of income volatility. Jensen and Shore develop a methodology to estimate nonparametrically the distribution of volatility of *excess* log income—the residual from a regression to predict the natural log of labor income. This regression is weighted by PSID-provided sample weights, normalized so that the average weight in each year is the same. We use the following elements as covariates in this regression: a cubic in age for each level of educational attainment (none, elementary, junior high, some high school, high school, some college, college, graduate school); the presence and number of infants, young children, and older children in the household; the total number of family members in the household; and dummy variables for each calendar year. Including calendar year dummy variables eliminates the need to convert nominal income to real income explicitly.

Some other papers have dropped observations with missing and zero income (Gottschalk and Moffitt (2002)) or modeled unemployment explicitly (Pistaferri (2002)), but neither route is available to us because the method in Jensen and Shore is not designed to handle missing data or zero values for income. Instead, Jensen and Shore use hot-deck imputation for missing values when calculating volatility. Zero values are bottom-coded at the equivalent of half-time, federal minimum wage real income. Aside from using their volatility values, we do not explicitly use bootstrapped income data. We follow Jensen and Shore in using top- and bottom-coded incomes.[16]

Jensen and Shore estimate the parameters of a standard process for income dynamics, similar to those in Carroll and Samwick (1997) and Meghir and Pistaferri (2004). Excess log income $y_{i,t}$ for individual i at time t—the residual from a regression to predict log income with covariates, defined on the previous page—is modeled as the sum of permanent income, transitory income, and error $e_{i,t}$:

$$y_{i,t} = \underbrace{\sum_{k=1}^{t-3} \omega_{i,k} + \sum_{k=t-2}^{t} \phi_{\omega,t-k} \cdot \omega_{i,k}}_{\text{Permanent income}} + \underbrace{\sum_{k=t-2}^{t} \phi_{\varepsilon,t-k} \cdot \varepsilon_{i,k}}_{\text{Transitory income}} + e_{i,t}. \tag{19}$$

Permanent income is the weighted sum of past permanent shocks $\omega_{i,k}$ to income. Transitory income is the weighted sum of recent transitory shocks $\varepsilon_{i,k}$ to income.[17] The permanent shock, transitory shock, and error term are assumed to be normally distributed

---

[16]For details on how missing and zero values are handled, as well as specifics on income top and bottom codes, see Jensen and Shore (2015).

[17]In this framework, permanent shocks come into effect over three periods and transitory shocks fade completely after three periods, giving us three permanent weight parameters ($\phi_{\varepsilon,0}, \phi_{\omega,1}, \phi_{\omega,2}$) and three transitory weight parameters ($\phi_{\varepsilon,0}, \phi_{\varepsilon,1}, \phi_{\varepsilon,2}$). We refer to these weights ϕ collectively as the income process parameters, which will need to be estimated in our model. Jensen and Shore posit flat prior distribu-

as well as independent of one another over time and across individuals. The permanent shocks $\omega_{i,t}$ have mean zero and variance $\sigma^2_{\omega,i,t} \equiv \mathrm{E}[\omega^2_{i,t}]$; the transitory shocks $\varepsilon_{i,t}$ have mean zero and variance $\sigma^2_{\varepsilon,i,t} \equiv \mathrm{E}[\varepsilon^2_{i,t}]$. Finally, there is a "noise variance," which refers to the variance of measurement error $\mathrm{E}[e^2_{i,t}]$ that is assumed constant across individuals and over time.

Jensen and Shore (2011, 2015) develop a Markovian hierarchical Dirichlet process (MHDP) prior to estimate the distribution of ex ante expected volatility.[18] We use the estimates of the ex ante expected permanent volatility distribution—the distribution of $\sigma^2_{\omega,i,t}$ in equation (19)—and take the average of these estimates over the years 1991–1996 as our final measure of income volatility, σ_c in equation (3); while it abstracts from predictable covariate-driven variation, the log income from our occupational choice model in equation (3) corresponds to the excess log income we analyze in our statistical model. Note that many occupational choice models deal with a selection problem, in which the unconditional distribution of volatility is not equal to the distribution that would be observed if individuals were assigned to occupations at random. Because we assume that volatility is an attribute of an occupation not a person (see Section 2.1.1 for details), selection does affect the volatility distribution in our model by assumption, so that the unconditional distribution and the distribution obtained from random matching are assumed to be the same.

3.3 *Empirical evidence of sorting*

What is important for sensible estimates of β is a minimal level of sorting of the more risk-tolerant individuals into the more risky careers. Table 2 shows the extent of this sorting by presenting the joint distribution of income volatility and risk aversion.[19] In this table, σ^2 values are divided into 10 bins, which correspond approximately to the 1st, 5th,

---

tions for each weight parameter (i.e., $p(\phi) \propto 1$). However, so as to give meaning to the magnitude of our transitory shocks, we normalize the weights placed on transitory shocks to sum to 1 ($\sum_k \phi_{\varepsilon,k} = 1$). This is a slight generalization of an income process in which permanent shocks enter immediately ($\phi_\omega = 1$), transitory shocks fade immediately ($\phi_{\varepsilon,t} = 1; \phi_{\varepsilon,t-k} = 0$ for $k > 0$), and the noise error ($e_{i,t}$, included only for technical reasons related to the estimation procedure) can be dropped as equivalent to the transitory shock:

$$y_{i,t} = \sum_{k=1}^{t-3} \omega_{i,k} + \varepsilon_{i,t}.$$

[18]Our Markovian hierarchical Dirichlet process allows income volatility to vary between individuals and within individuals across time, while still sharing information over time and across individuals. This distribution of income volatility within individuals and across the population is estimated by the Markov chain Monte Carlo method. Specifically, we implemented a Gibbs sampler that iterates between sampling the global parameters of the model (e.g., residual variance, weights on permanent and transitory income shocks), the latent permanent and transitory income shocks for each individual in each year, and the volatility parameters for each individual in each year that underlie those permanent and transitory income shocks. An individual's volatility in a given year is sampled from a pool of volatility values shared across the population, but with additional preference given to previously observed volatility values for that individual.

[19]The Jensen and Shore (2011, 2015) method estimated separately by risk-tolerance bin yields similar sorting, as do estimates of income risk obtained using the Meghir and Pistaferri (2004) method.

Table 2. Estimated distribution of income volatility by self-reported risk aversion $\widehat{\sigma}^2$ from Jensen and Shore.

$\widehat{\sigma}^2$ Percentile		$\widehat{\sigma}^2 >$	$\widehat{\sigma}^2 \leq$	Raw $\widehat{\sigma}^2$ distribution	$\widehat{\sigma}^2$ Distribution conditional on $\widetilde{\gamma}$			
					$[0, 2)$	$[2, 3.84)$	$[3.84, 7.52)$	$[7.52, \infty)$
min	1st	0.160^2	0.164^2	1.01%	1.20%	1.12%	0.50%	1.08%
1st	5th	0.164^2	0.170^2	3.07%	3.68%	4.10%	2.04%	2.82%
5th	10th	0.170^2	0.173^2	5.01%	4.79%	4.33%	4.58%	5.54%
10th	30th	0.173^2	0.178^2	22.03%	22.91%	22.68%	23.00%	20.98%
30th	50th	0.178^2	0.180^2	25.54%	20.60%	26.96%	22.95%	28.45%
50th	70th	0.180^2	0.182^2	17.06%	13.94%	15.72%	19.13%	18.21%
70th	90th	0.182^2	0.217^2	16.30%	16.39%	17.20%	17.45%	15.46%
90th	95th	0.217^2	0.345^2	5.11%	8.73%	5.14%	4.50%	3.60%
95th	99th	0.345^2	1.000^2	4.23%	6.40%	2.76%	4.69%	3.53%
99th	max	1.000^2	1.000^2	0.64%	1.36%	0.00%	1.15%	0.33%
Number of observations				1,490	320	241	267	662
Percentage of observations				100%	21.50%	16.17%	17.90%	44.43%

Note: This table shows the distribution of σ^2 estimates. The σ^2 estimates are the average of 1991–1996 estimates of permanent volatility. Volatility estimates are from Jensen and Shore and are top-coded at 1. The $\widetilde{\gamma}$ ranges are from the coarsely binned responses to the 1996 risk-tolerance supplement to the PSID. Both the raw (rounded) number of observations and the percentage of observations in each range represent PSID sample-weighted observations.

10th, 30th, 50th, 70th, 90th, 95th, and 99th percentiles of the σ^2 distribution. Individuals with low risk aversion (stated $\widetilde{\gamma}$ between 0 and 2) are more likely to have the highest volatility values; individuals with high risk aversion (stated $\widetilde{\gamma}$ above 7.52) are less likely to have the highest volatility values.

Table 3 offers additional reduced-form evidence on the relationship between income volatility and risk aversion. That table includes three panels that differ in the functional form for income volatility used as the dependent variable; each panel presents three regressions that differ in their control variables. The relationship between income volatility and risk aversion is negative throughout. The significance and magnitude of this relationship are relatively invariant to the control variables we include. This relationship is statistically insignificant when the dependent variable is the level of income volatility, top-coded at $\sigma^2 = 1$; this specification is dominated by observations with extremely high values for σ^2. When the dependent variable is the volatility level top-coded at the 95th percentile for income volatility $\sigma^2 \leq 0.345^2$ or the log of volatility, the results become statistically significant.

4. Estimation

If we could observe the joint distribution of data $\{\sigma^2, \widetilde{\gamma}, x^{IO}, x^{CO}\}$, then equation (17) is straightforward to estimate by maximum likelihood. We need only choose a parametric (or nonparametric) structure for $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$, and estimate its parameters along with β. Table 2 shows the nonparametric approach we pursue, splitting σ^2 into 10 ranges. We assign each range a σ^2 value equal to the within-range weighted average, with each σ^2 observation weighted by its PSID-supplied sample weight. Covariates

TABLE 3. Relationship between income volatility (1991–1996) and risk aversion (1996).

Dep. Var.	Income risk level (top-code at $\sigma^2 = 1.000$)			Income risk level (top-code at $\sigma^2 = 0.345^2$)			Income risk log (top-code at $\sigma^2 = 1.000$)		
$E[\widetilde{\gamma} \mid \text{bin}]$	−0.002	−0.003	−0.002	−0.001***	−0.001***	−0.001***	−0.018**	−0.020**	−0.018**
	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)	(0.009)	(0.009)	(0.009)
Age		0.001**	0.001***		0.000	0.000		0.005**	0.006**
		(0.001)	(0.001)		(0.000)	(0.000)		(0.002)	(0.003)
Controls	No	No	Yes	No	No	Yes	No	No	Yes
R^2	0.001	0.006	0.031	0.006	0.007	0.034	0.003	0.007	0.037
No. of obs.	1,490	1,490	1,490	1,490	1,490	1,490	1,490	1,490	1,490

Note: This table shows the ordinary least squares (OLS) regressions to predict individual-specific measures of income risk with the self-reported risk-aversion bin. The σ^2 estimates are the average of 1991–1996 estimates of permanent volatility from Jensen and Shore. The term $E[\widetilde{\gamma} \mid \text{bin}]$ refers to the expected value of risk aversion conditional on the risk-aversion bin, which we estimate using the signal-noise structure identified in Kimball, Sahm, and Shapiro (2009). The variable age refers to the individual's age in years, and controls include occupation, family, and demographic characteristics. The asterisk (*) indicates significance at the 10% level, the double asterisks (**) indicate significance at the 5% level, and the triple asterisks (***) indicate significance at the 1% level.

aside, we need only estimate β and nine probabilities: the probability that a risk-neutral person will land in each of the 10 volatility bins.

The complication is that we do not observe $\widetilde{\gamma}$ exactly; we see only into which of four coarse bins $\widetilde{\gamma}$ falls. Furthermore, there is measurement error in $\widetilde{\gamma}$, so that the true value for $\widetilde{\gamma}$ may not even fall in the range of its bin. We adopt the classical measurement error structure proposed in Kimball, Sahm, and Shapiro (2009) to model the distribution of $\widetilde{\gamma}$ in the PSID given that we observe it with error, and even then, only in bins. In particular, Kimball, Sahm, and Shapiro estimate the following structure for $\widetilde{\gamma}$:

$$\widehat{\ln(1/\widetilde{\gamma})} = \ln(1/\widetilde{\gamma}) + e, \tag{20}$$

$$\begin{bmatrix} \ln(1/\widetilde{\gamma}) \\ e \end{bmatrix} \sim N\left(\begin{bmatrix} -1.05 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.76 & 0 \\ 0 & 1.69 \end{bmatrix} \right). \tag{21}$$

We observe true log risk tolerance $(\ln(1/\widetilde{\gamma}))$ plus noise (e), placed into bins, so that a given observation lies in a given bin if $\widehat{\ln(1/\widetilde{\gamma})} > \underline{\text{bin}}$ and $\widehat{\ln(1/\widetilde{\gamma})} < \overline{\text{bin}}$, where $\underline{\text{bin}}$ and $\overline{\text{bin}}$ are the lower and upper bounds of the bins, respectively. Again, Table 2 shows these ranges and the fraction of observed data that falls into each.[20]

We can then identify the relationship between our data ($f(\sigma_i^2 \mid \widehat{\ln(1/\widetilde{\gamma})} \, \text{bin}_i)$) and the object we wish to estimate ($f(\sigma^2 \mid \widetilde{\gamma})$ from equation (17)):

$$
\begin{aligned}
&f\big(\sigma_i^2 \mid \widehat{\ln(1/\widetilde{\gamma})} \, \text{bin}_i\big) \\
&\quad \propto \int_{\ln(1/\widetilde{\gamma})} f\big(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO}\big) e^{-\frac{1}{2}\widetilde{\gamma}\sigma_i^2/\beta} \\
&\qquad \times f_{\ln(1/\widetilde{\gamma})}\big(\ln(1/\widetilde{\gamma}) \mid \widehat{\ln(1/\widetilde{\gamma})} \, \text{bin}_i\big) \, d\ln(1/\widetilde{\gamma}),
\end{aligned} \tag{22}
$$

[20]We approximate this distribution with a 38 element grid, assigning a probability that $\widetilde{\gamma}$ will be each of the following values: $\{0.5, 1.25, 2, 2.5, 3, 3.4, 3.8, 4.5, 5.5, \ldots, 9.5, 10, 10.5, 11, 12, \ldots, 34\}$.

$$f_{\ln(1/\widetilde{\gamma})}\big(\ln(1/\widetilde{\gamma}) \mid \widehat{\ln(1/\widetilde{\gamma})}\,\mathrm{bin}_i\big) = f_{\ln(1/\widetilde{\gamma})}\big(\ln(1/\widetilde{\gamma})\big)\frac{\mathrm{pr}\big(\widehat{\ln(1/\widetilde{\gamma})}\,\mathrm{bin}_i \mid \widetilde{\gamma}\big)}{\mathrm{pr}\big(\widehat{\ln(1/\widetilde{\gamma})}\,\mathrm{bin}_i\big)}. \tag{23}$$

Given the distribution of true variation and classical measurement error estimated by Kimball, Sahm, and Shapiro, it is trivial to calculate $f_{\ln(1/\widetilde{\gamma})}(\ln(1/\widetilde{\gamma}) \mid \widehat{\ln(1/\widetilde{\gamma})}\,\mathrm{bin}_i)$ and $\mathrm{pr}(\widehat{\ln(1/\widetilde{\gamma})}\,\mathrm{bin}_i \mid \widetilde{\gamma})$ for each $\widetilde{\gamma}$ in our grid for each of the four risk-aversion bins; $\mathrm{pr}(\widehat{\ln(1/\widetilde{\gamma})}\,\mathrm{bin}_i)$ is similarly easy to calculate for each of the four risk-aversion bins.

Armed with this distribution of $\widetilde{\gamma}$, we estimate $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$ and β by maximizing the likelihood of observing our data:

$$\max_{\{f(\sigma^2|\widetilde{\gamma}=0,x^{IO},x^{CO}),\ \beta\}} \prod_i f\big(\sigma_i^2 \mid \widehat{\ln(1/\widetilde{\gamma})}\,\mathrm{bin}_i\big). \tag{24}$$

We have 10 parameters to estimate: the 9 free probabilities that comprise $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$ and β, which we search for iteratively.[21] First, we guess values of $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$ and β. Next, we calculate $f(\sigma^2 \mid \widetilde{\gamma}, x^{IO}, x^{CO})$ for each observed value of σ_i^2 and each value of $\widetilde{\gamma}$ on our grid. Next, we calculate $f(\sigma^2 \mid \widetilde{\gamma}, x^{IO}, x^{CO})$ for each of the 10 grid values of σ^2 and each of the 4 coarse bins for $\widetilde{\gamma}$ by integrating over each value of $\widetilde{\gamma}$ possible in each bin. This gives the likelihood of each $(\sigma_i^2, \widehat{\ln(1/\widetilde{\gamma})}\,\mathrm{bin}_i)$ observation lying in one of the $10 \times 4 = 40$ possible ranges we observe in Table 2. We then compute the likelihood of observing the data in Table 2. We search over $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$ and β to find values that maximize the likelihood.

5. Results

Equation (17) show the key model parameters we estimate in Section 5.1: β and $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$. The parameter β (proportional to $\mathrm{var}(\varepsilon_{i,c})$) measures the importance of idiosyncratic taste and skill from the shift in the distribution of income risk as risk aversion increases; $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$ is the distribution of income risk chosen by risk-neutral people, which shifts with covariates (θ). In Section 5.2, we present the risk–return menu implied by the β we estimate in Section 5.1 under different assumptions about the elasticity of demand for careers. In Section 5.3, we present results from the regressions implied by equation (64), designed to separate the relative importance of idiosyncratic taste from idiosyncratic skill.

5.1 *Parameter estimates*

Table 4 shows the coefficient estimates from equation (17) using two different estimation methods. The first method produces our main results, which are used in Section 5.2 and in our counterfactual analysis below. This approach assumes the Jensen–Shore volatility values are estimated with certainty. These results are reported in the left three columns of Table 4. The β value estimated without additional controls is 0.649, so

---

[21]Although there are 10 values of $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$, they must sum to 1, which leaves only 9 free parameters to estimate.

TABLE 4. Parameter estimates.

Parameter	Average σ^2			Bootstrap σ^2		
$\widehat{\beta}$	0.649	0.553	0.588	3.075	2.617	2.876
$\beta <$	2.749	2.027	2.302	18.107	25.931	16.898
$\beta >$	0.279	0.246	0.258	-12.385	0.759	0.801
Lowest	0.8%	1.0%	1.1%	2.5%	3.0%	3.2%
Income	2.6%	3.1%	3.2%	3.2%	3.9%	4.1%
Risk	4.4%	5.1%	5.2%	3.6%	4.4%	4.6%
$\uparrow$	19.5%	22.5%	22.9%	19.8%	21.9%	22.6%
$f(\sigma^2\|\gamma=0)$	22.8%	26.1%	26.5%	19.1%	20.2%	20.3%
	15.4%	17.5%	17.7%	20.1%	20.4%	20.4%
$\downarrow$	15.1%	16.8%	16.9%	20.1%	19.6%	19.4%
Highest	6.0%	5.2%	4.9%	4.6%	4.3%	4.2%
Income	10.7%	2.5%	1.7%	5.1%	1.8%	1.3%
Risk	2.7%	0.2%	0.1%	1.5%	0.2%	0.1%
Age $\times\sigma^2$	.	0.08	0.08	.	0.06	0.06
Edu. $\times\sigma^2$	.	0.02	0.17	.	0.01	0.11
Race $\times\sigma^2$	No	Yes	Yes	No	Yes	Yes
Occ. $\times\sigma^2$	No	No	Yes	No	No	Yes

Note: This table displays the estimates of β (including a 90% confidence interval), $f(\sigma^2|\widetilde{\gamma}=0)$, and θ from equations (17) and (18). The point estimates of β correspond to the variance of idiosyncratic taste and skill shocks. The term $f(\sigma^2|\widetilde{\gamma}=0)$ is the probability that a risk-neutral individual populates each of the 10 σ^2 bins. The vector θ represents the coefficient estimates of these controls.

that the standard deviation of idiosyncratic career values is $64.9\% \times \pi/\sqrt{6}$ of income (in log points). We obtain a 90% confidence interval for $\hat{\beta}$ using a likelihood ratio test.[22] We find upper and lower bounds of $\beta < 2.749$ and $\beta > 0.279$, respectively. Although these estimates are large, we view the lower bound on β as entirely plausible; it implies a dispersion of idiosyncratic taste or skill of around 36% of income. Said differently, a 1 standard deviation decrease in career "enjoyment" is equivalent to a pay decrease of 36%. Adding controls for age, race, education, and occupation changes the point estimate and upper/lower bounds of β only slightly.[23]

The second approach explicitly incorporates model uncertainty in the Jensen–Shore volatility estimates. We use 100 bootstrapped Jensen–Shore volatility samples and estimate the model for each sample. For each sample, just as before, σ^2 is split into 10 bins based on the distribution percentiles show in Table 2.[24] This gives us 100 sets of β

[22]Specifically, we calculate a restricted likelihood value by solving the model for $f(\sigma^2 \mid \widetilde{\gamma}=0, x^{IO}, x^{CO})$ conditional on a fixed value of β. We then search for the smallest and largest fixed β values that allow us to reject that the restricted model is correct using the two-sided likelihood ratio test.

[23]Results using the Meghir and Pistaferri (2004) method of income volatility are broadly similar. Unsurprisingly, β estimates generated by the Meghir–Pistaferri moments are higher. This is consistent with the attenuation bias in estimates of $1/\beta$ we would expect given the more dispersed Meghir–Pistaferri volatility estimates, which measure realized rather than expected volatility.

[24]Whereas the 10 σ^2 bins in our primary estimation correspond only approximately to the reported quantiles, in the bootstrap estimation we construct the 10 bins to exactly match the σ^2 quantiles. We construct the 10 σ^2 bins separately for each bootstrapped sample.

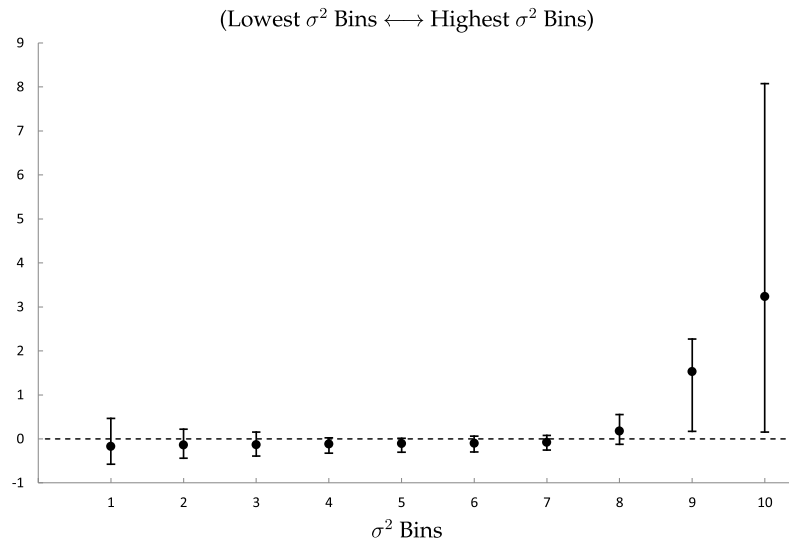(Lowest σ^2 Bins $\longleftrightarrow$ Highest σ^2 Bins)

FIGURE 2. Over-/underrepresentation of risk-neutral individuals by σ^2 bins. This figure presents estimates of $f(\sigma|\widetilde{\gamma}=0)$. These are normalized by dividing by the unconditional population distribution across bins and subtracting 1. This shows the degree to which risk-neutral individuals are estimated to overweight or underweight this bin relative the population as a whole. This panel shows 95% confidence intervals from a likelihood ratio test (where only this probability but no other parameters are restricted).

and $f(\sigma^2 \mid \widetilde{\gamma}=0, x^{IO}, x^{CO})$ estimates. In this case, point estimates are determined by the median value of each parameter,[25] and confidence intervals are determined by the 5th and 95th percentiles of the 100-sample parameter distributions. The results from this method are reported in the right three columns of Table 4.

While the estimates of $\widehat{\beta}$ obtained using bootstrapped income volatility samples are larger and significantly more dispersed than those obtained under the assumption of model certainty, this is completely unsurprising. The relationship between income risk and risk aversion is simply weaker within bootstrapped samples than when income risk is estimated across samples. A weaker correlation between income risk and risk aversion will push estimates of $\widehat{\beta}$ toward infinity. This is most easily seen through the difference in the estimated upper bounds of $\widehat{\beta}$. However, note that once covariates are included, the lower bound of $\widehat{\beta}$ is broadly similar to the estimated value of $\widehat{\beta}$ under the first approach.

Along with estimates of β, Table 4 shows the estimates of $f(\sigma^2 \mid \widetilde{\gamma}=0, x^{IO}, x^{CO})$ and θ. Figure 2 depicts the estimated (scaled) distribution of $f(\sigma^2 \mid \widetilde{\gamma}=0, x^{IO}, x^{CO})$ under the assumption that the Jensen–Shore volatility values are estimated with certainty. This is equal to $f(\sigma^2 \mid \widetilde{\gamma}=0)$ when the model is estimated without additional covariates. Figure 2 shows the degree to which risk-neutral individuals are estimated to overweight or underweight this bin relative to the population as a whole. For each σ^2 bin we obtain a 95% confidence interval by finding the highest and lowest values of

---

[25]The median is defined as the value of the 50th entry of the sorted distribution, rather than the average of the 50th and 51st.
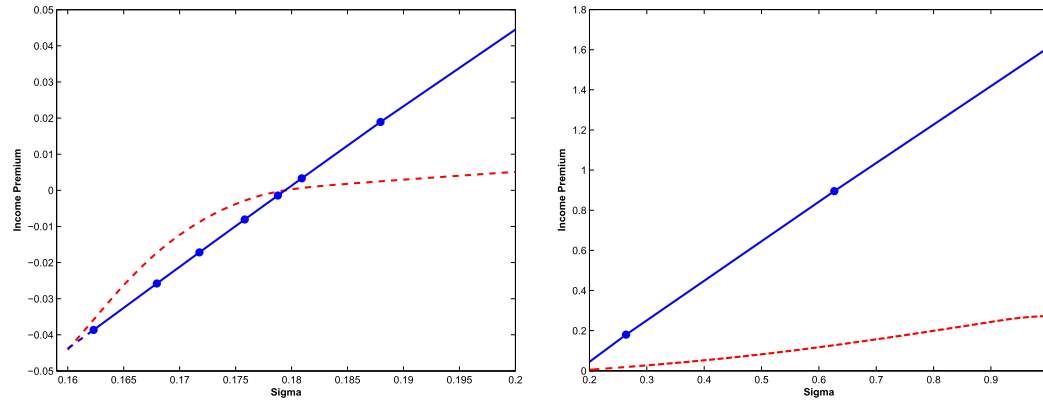
FIGURE 3. Average risk premia by income risk bin. The figure shows the estimated income premium y^C at the midpoint of each σ^2 bin. The two panels display the full range of σ values on different vertical axis scales. The dashed curve reflects the perfect-sorting case where $\beta = 0$. The solid curve reflects the required risk premium needed to rationalize the data, under the assumption that career supply is inelastic so that pay adjusts so that the income risk distribution of career options equals the income risk distribution of chosen careers, when equation (17) is estimated without covariates.

$f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$ such that the restricted model fails to reject the likelihood ratio test that the restricted $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$ value is correct.

5.2 *The risk–return menu*

Further, equation (18) shows that the distribution of σ^2 choices by risk-neutral people may reflect the distribution of career options f^C or the relative value of those options $(E[e^{(y^C + v(x^I, x^C))^{\frac{1}{\beta}}} \mid \sigma^2])$. There is no way to differentiate these two cases without a model of wage adjustment. At one extreme, we can assume that the demand for workers in each career option is completely inelastic, so that wages adjust until the unconditional distribution of chosen careers $f(\sigma^2)$ is equal to the distribution of career options (f^C). In this case, we implicitly observe f^C, and can identify $(y^C + v(x^I, x^C))$, the income premium needed to fill all careers at each level of volatility. Assuming no heterogeneity conditional on σ^2, from equation (18) we have $e^{y^C + v(x^I, x^C)} = (\frac{f(\sigma^2 \mid \widetilde{\gamma} = 0)}{f^C})^{\beta}$. Given our estimates of $f(\sigma^2 \mid \widetilde{\gamma} = 0)$, we can trace out the implied risk–return menu, the income premium needed to fill all careers at each volatility bin. Estimates of this risk–return menu are shown in Figure 3. Note the substantial risk premium required to fill the high-income-risk bins. This is consistent with the idea that important idiosyncratic taste or skill in various careers implies that the marginal person choosing a risky career is not very risk tolerant, and must be offered a significant risk compensation (either in pay or enjoyment) to fill this risky career.

At the other extreme, we can assume that demand for workers in each career is completely elastic, so that the value of each career is the same in expectation. In this case, careers are filled in proportion to their frequency, so that careers with twice as many

slots are twice as likely to be chosen by the risk-neutral individual. In this case, the risk–return menu is simply a horizontal line. The distribution of risk-neutral choices in this scenario is shown in Figure 2.

5.3 *Idiosyncratic taste or skill?*

Equation (17) provides a way to estimate β from the degree to which the conditional distribution of risk choices shifts with risk aversion (recall that β measures the standard deviation of $y_{i,c}^{\varepsilon} + \widetilde{l}_{i,c}^{\varepsilon}$). Without additional information, we cannot separate the relative importance of individual-specific shocks to skill in specific careers ($y_{i,c}^{\varepsilon}$) from individual-specific shocks to taste for (enjoyment of) those careers $\widetilde{l}_{i,c}^{\varepsilon}$.

However, we can separate these two effects using income data. If people dislike risk, they must be compensated in some way for taking more of it. The more risk averse a person is, the greater such compensation must be. This compensation could come in the form of higher pay or more career enjoyment. Risk-averse people will only choose risky jobs if they love them or are very productive in them (thereby earning particularly high pay). In a world in which most idiosyncratic variation is in enjoyment, we will see risk-averse people compensated by choosing risky jobs they particularly enjoy. In a world in which most idiosyncratic variation is in career-specific skill, we will see risk-averse people compensated by choosing jobs at which they particularly excel and therefore earn higher pay. We should not see this pattern among the risk neutral. The relative importance of idiosyncratic skill and taste can therefore be identified from the degree to which observed compensating wage differentials for risk increase with risk aversion.

The model formalizes this intuition in Appendix A.3, showing a linear relationship between expected log pay and the interaction of income risk and risk aversion that reveals the proportion of idiosyncratic fit attributable to skill. The resulting equation is

$$E\big[\log \mathrm{pay}_{i,c} \mid V_{i,c} \geq V_{i,c'} \ \forall c'\big]$$

$$= \underbrace{(\mu + \beta\gamma_{em})\left(1 - \frac{\mathrm{var}\big(\widetilde{l}_{i,c}^{\varepsilon}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l}_{i,c}^{\varepsilon}\big)}\right)}_{\text{Constant independent of individuals or careers}}$$

$$+ \underbrace{y_i^I + \beta\ln\left(\sum_q s_q e^{(y_q^C + v(x_i^I, x_q^C) - \frac{1}{2}\widetilde{\gamma}_i \sigma_q^2)/\beta}\right)\left(1 - \frac{\mathrm{var}\big(\widetilde{l}_{i,c}^{\varepsilon}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l}_{i,c}^{\varepsilon}\big)}\right)}_{\text{Individual-specific attributes including risk tolerance}} \tag{25}$$

$$+ \underbrace{y_c^C \frac{\mathrm{var}\big(\widetilde{l}_{i,c}^{\varepsilon}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l}_{i,c}^{\varepsilon}\big)}}_{\text{Career-specific attributes}} \underbrace{- v\big(x_i^I, x_c^C\big)\left(1 - \frac{\mathrm{var}\big(\widetilde{l}_{i,c}^{\varepsilon}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l}_{i,c}^{\varepsilon}\big)}\right)}_{\text{Observable nonrisk attributes}}$$

$$+ \underbrace{\frac{1}{2} \times \widetilde{\gamma}_i \times \sigma_c^2\left(1 - \frac{\mathrm{var}\big(\widetilde{l}_{i,c}^{\varepsilon}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l}_{i,c}^{\varepsilon}\big)}\right)}_{\text{Risk tolerance and income risk interaction}}.$$

The first line in this log pay equation depends on neither individual attributes (X^I) nor chosen career attributes (X^C). The second line depends on individual attributes (X^I, specifically y_i^I and $\widetilde{\gamma}_i$) but not chosen career attributes. The third line depends on chosen career attributes (specifically y_c^C and x_c^{CO} but not σ_c^2) and nonrisk observables (x_i^{IO}). The final line depends on both an individual attribute ($\widetilde{\gamma}_i$) and a career attribute (σ^2). This line shows that the relative importance of idiosyncratic skill and taste can be identified from the degree to which observed compensating wage differentials for risk increases with risk aversion.

Equation (25) suggests that a simple regression can be used to recover the relative importance of taste shocks ($\widetilde{l}_{i,c}^{\varepsilon}$) compared with all shocks ($y_{i,c}^{\varepsilon} + \widetilde{l}_{i,c}^{\varepsilon}$). The regression predicts pay with the following controls: a constant, individual-specific controls (including risk aversion), career-attribute controls (adding a measure of income risk), and the interaction of observable nonrisk individual- and career-specific controls. As shown in the fourth line of equation (25), it also includes the interaction between risk aversion ($\widetilde{\gamma}_i$) and income risk (σ_c^2). Assuming that individual-specific unobservables that affect pay are uncorrelated with income risk, the coefficient on this interaction identifies the relative importance of career taste in idiosyncratic fit, $\frac{1}{2} \times (1 - \mathrm{var}(\widetilde{l}_{i,c}^{\varepsilon})/\mathrm{var}(y_{i,c}^{\varepsilon} + \widetilde{l}_{i,c}^{\varepsilon}))$.

The parameter of interest in equation (64)—the coefficient on $\frac{1}{2} \times \widetilde{\gamma}_i \times \sigma_c^2$—is estimated by an OLS regression. If this coefficient is 0, the variation in career choice is exclusively in idiosyncratic taste; if this coefficient is 1, it is exclusively in skill; intermediate values indicate the presence of both idiosyncratic taste and skill. The intuition here is that risk-averse people demand a larger "compensation" to enter high-risk careers. As a result, the gap in compensation between high- and low-risk careers will be greatest for those with the highest risk aversion. If we do not observe a pay gap, this compensation must be in the form of idiosyncratic taste (loving your job).

Table 5 shows the results from regressing pay on $\widetilde{\gamma}_i$, σ_c^2, their interaction, and covariates.[26] Point estimates for the coefficient on $\frac{1}{2} \times \widetilde{\gamma}_i \times \sigma_c^2$ range from 0.14 to 0.26 in various specifications. When nonlinear functions of σ_c^2 are included as controls, the estimate has small enough standard errors that we can rule out a coefficient of zero. Results with this specification indicate that idiosyncratic skill and the increased compensation that follows explains up to half (and idiosyncratic taste explains at least half) of the idiosyncratic factors driving career choice. Finally, note the similarity between this regression and the risk-augmented Mincer equations from Hartog (2011), which provide a consistency check on our particular sample.

6. Counterfactuals

The model suggests at least three interesting counterfactuals that we pursue in this section. In the first, we calculate the value of worker training based on the fraction of careers affected and the observed average pay increase associated with training. In the second, we ask how much we would have to compensate individuals to choose careers in the

---

[26]Note that $\widetilde{\gamma}_i$ refers to $E[\widetilde{\gamma}_i \mid \widehat{\widetilde{\gamma}\mathrm{bin}}]$, which is based on the distribution of $\widetilde{\gamma}_i$ and measurement error proposed by Kimball, Sahm, and Shapiro (2009).

TABLE 5. Impact of income risk and risk aversion on income.

Dependent variable:	Log average income: Jensen–Shore				
σ^2	−0.749***	−0.823***	−1.091**	−1.309***	−1.579
	(0.166)	(0.153)	(0.464)	(0.418)	(1.606)
$(\sigma^2)^2$					1.373
					(1.143)
$\ln(\sigma^2)$					−0.234
					(0.152)
$\widetilde{\gamma} = $ 2nd lowest			−0.002	−0.005	−0.017
			(0.058)	(0.052)	(0.052)
$\widetilde{\gamma} = $ 2nd highest			0.105*	0.150***	0.133***
			(0.056)	(0.050)	(0.050)
$\widetilde{\gamma} = $ highest			0.003	0.045	0.023
			(0.054)	(0.049)	(0.049)
$\frac{1}{2} \times \sigma^2 \times \widetilde{\gamma}$			0.142	0.203	0.255**
			(0.173)	(0.151)	(0.127)
Age	No	Yes	No	Yes	Yes
Race	No	Yes	No	Yes	Yes
Family size	No	Yes	No	Yes	Yes
Education	No	Yes	No	yes	Yes
R^2	0.018	0.203	0.023	0.212	0.229
Number of Obs.	1,484	1,484	1,484	1,484	1,484

Note: All results are for OLS regressions weighted by PSID-provided sample weights. "Age" indicates whether a linear age control was included; "Family size" indicates whether linear controls for total family size, presence and number of babies, young children, and older children were included; "Race" indicates whether white, black, and other race controls were included; "Education" indicates whether a linear years of schooling variable was included. The full sample includes 1,490 observations, but six of these have an income of zero throughout, and consequently a missing log income. The σ^2 refers to the average of Jensen and Shore's estimates of permanent income volatility from 1991 to 1996. The dependent variable is the log of average income, averaged over the period 1991–1996. Standard errors are givien in parentheses: The asterisk (*) indicates significant at the 10% level; the double asterisk (**), significant at the 5% level; the triple asterisk (***) significant at the 1% level.

same proportion as risk-neutral workers. In the third, we ask how much better off people would be if their chosen careers were also socially optimal in perfectly matching the most risk-tolerant people to the riskiest jobs.

6.1 *The value of worker training*

Offering workers training in new skills and occupations has been a popular way for both federal and state agencies to address low-income and displaced worker unemployment and underemployment. A common feature of these programs is that they offer training in some fields (e.g., those in high demand such as energy or health care), but not others.[27] A concern raised about such programs is that workers displaced from one occupation may be poorly suited for the careers in which training is being subsidized.

---

[27]The entirety of the industries represented by the President's 2008 Community-Based Job Training Grant Awardees can be broadly categorized as health care, energy, manufacturing/construction/logistics, transportation/aviation, biotech/nanotech, and information technology (http://www.doleta.gov/grants/sga/ DOL-SGA-DFA-PY-07-01-List_of_grantees.pdf, as accessed on 12/12/2014). Washington state offers addi-

Our model provides a way to quantify that concern. In particular, our model provides a quantitative relationship between three features of the training program: the size of the subsidy available, the proportion of careers available for that subsidy, and the average benefit of the program (as well as the proportion of people who find the program beneficial).

Training programs typically subsidize job training in specific careers, which is equivalent to increasing the expected lifetime earnings in those careers net of training costs. Let S be the subsidy provided to favored careers, expressed as the log increase in lifetime income net of training costs. Let ρ represent the share of careers eligible for the subsidy. We assume that whether a career is subsidized is independent of its risk and of workers' idiosyncratic taste and skill in that career. Further, we ignore general equilibrium effects, in which subsidies may change the salaries offered in various careers.

Given this setup, the fraction of workers who would choose a subsidized career is

$$\pi = \frac{\rho e^{\frac{S}{\beta}}}{\rho e^{\frac{S}{\beta}} + (1 - \rho)}. \tag{26}$$

Workers fall into three groups: (i) those who would choose a subsidized career even absent the subsidy (a share ρ of the population); (ii) those who do not choose subsidized careers with or without a subsidy; and (iii) those who would not choose subsided careers absent the subsidy, but who are induced to take up these careers because of the subsidy (a share $\pi - \rho$ of the population).

Consider a training program that increases expected pay by 10% ($S = 0.1$) for 10% of careers ($\rho = 0.1$). For workers in group (i)—who constitute 10% of the population and would have chosen a subsidized career anyways—the subsidy obviously increases their income by 10%. Workers in group (ii)—who do not choose favored careers even with the subsidy—constitute 88.53% of the population. This group would not participate in the training program, and therefore does not benefit from the subsidy. Workers in group (iii)—who are induced to take up favored careers because of the subsidy—make up $\pi - \rho = 1.47\%$ of the population. This is equivalent to 1.63% of workers who would not have chosen favored careers absent the subsidy ($\frac{\pi - \rho}{1 - \rho}$). This low take-up rate is implied by the high importance of idiosyncratic taste and skill (β) estimated in this paper. The model shows that workers find most careers unappealing, so subsidizing a small subset of careers will induce relatively few workers to enter those likely unappealing careers. Figure 4 shows how the proportion of individuals induced to change careers due to the subsidy ($\frac{\pi - \rho}{1 - \rho}$) varies with S and ρ. For workers in group (iii), the benefit of the subsidy is 4.90% of lifetime income, so their welfare is increased by roughly half the size of the subsidy they receive.[28] This suggests that government programs that provide

---

tional unemployment insurance benefits to laid-off or displaced workers who pursue career training, but only for training in pre-approved fields that are in "high demand" in the individual's labor market (http://www.esd.wa.gov/uibenefits/specialservices/training/acceptable-training-programs.php, as accessed on 12/12/2014).

[28]The expected increase in log earnings net of training is

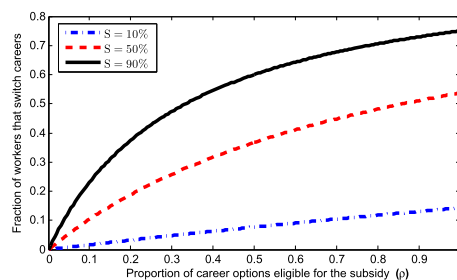$$\Delta\mu = \beta \ln(\rho e^{S/\beta} + 1 - \rho).$$

FIGURE 4. Proportion of workers who switch careers due to the subsidy. This figure shows the fraction of workers who are induced to switch careers by a job-training subsidy ($\frac{\pi-\rho}{1-\rho}$), as calculated from equation (26) with $\beta = 0.649$, for various values of the subsidy (S) and the proportion of careers eligible for the subsidy (ρ).

training in a particular career (LaLonde (1995), Jacobson, LaLonde, and Sullivan (2005)) may be of limited value because some of the individuals who choose to participate in the program will have limited interest in or aptitude for those careers.

To the degree that idiosyncratic career taste affects career choice, displaced workers may face substantial welfare costs beyond their well studied forgone income, due to the loss of the high idiosyncratic enjoyment associated with their previous career (see Jacobson, LaLonde, and Sullivan (1993), Ruhm (1991), and Couch and Placzek (2010) as just a few examples). Our model and this counterfactual exercise suggests that these concerns are quantitatively important.

6.2 *The certainty-equivalent cost of risk*

Risk-neutral individuals choose the best careers ignoring risk, choosing jobs with the best combination of idiosyncratic fit and other attributes. The distribution of their choices shows what any worker would choose if income risk were eliminated from all jobs or if risk were not considered when choosing careers. Coupled with our model, this insight allows us to calculate how much risk "distorts" career choices. Equivalently, how much would we have to compensate risk-averse individuals for choosing the best careers ignoring risk?

The model provides an estimate of the risk distribution of careers chosen by risk-neutral and risk-averse individuals. It is straightforward to calculate and compare the welfare costs of risk faced by a risk-averse individual choosing both the risk distribution they would choose taking risk into account and the risk distribution a risk-neutral

---

In this case, our model implies that the expected increase in average wages is $\Delta\mu = 1.07\%$. However, a full 1.00% of this 1.07% increase comes directly from individuals who were already choosing training-affected careers (a 10% wage increase for the 10% who already choose these careers). The additional 0.07% increase is due to the career switchers, who comprise $\pi = 1.47\%$ of the population. This means the average wage increase for the switchers is $0.07/0.0147 = 0.0490$, or 4.90%. The benefits for this group must range between 0% and 10%, so that a large estimated value of β implies that the distribution of benefits within this range is roughly uniform and that the average value is approximately the midpoint.
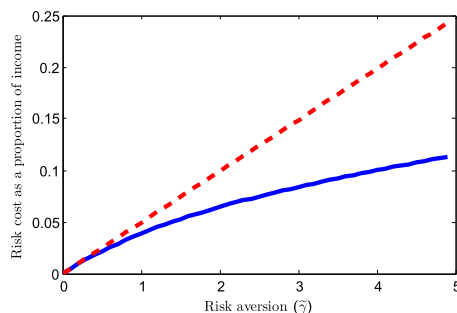
FIGURE 5. Log-income certainty equivalent of income risk. This figure shows the expected risk cost for risk-averse workers when careers are chosen optimally, and under the counterfactual that careers are chosen in the same proportion as for risk-neutral workers. The blue-solid curve represents the average risk cost under optimal choices; the red-dashed line represents the average risk cost of risk-neutral choices for risk-averse workers. The horizontal axis is worker risk-aversion. The vertical axis is the fraction of worker pay.

person would choose.[29] The difference between these two welfare costs represents the amount an individual would need to be compensated to choose the best career ignoring risk.

Figure 5 plots, for each level of risk aversion, the average welfare cost of income risk both when income risk is ignored when choosing careers and when it is taken into consideration. The vertical distance between the curves is the fraction of income needed to compensate individual i, with risk aversion $\widetilde{\gamma}_i$, for choosing careers without regard for his/her risk. Figure 5 shows that the expected cost of risk is linear in risk aversion for risk-neutral choices, but that the risk cost under optimal choices grows more slowly, as risk-averse individuals substitute more heavily away from painful income risk as risk aversion grows. For modest levels of risk aversion, the welfare cost of ignoring risk is quite low. Idiosyncratic taste and skill are much more important drivers of career choice than risk for workers with moderate risk aversion, so ignoring risk entirely does not incur large additional costs.

6.3 *The value of perfect sorting*

Absent idiosyncratic taste for and skill in different careers, we would expect perfect sorting of the most risk-tolerant people into the riskiest careers, induced to enter those careers by a compensating wage differential for income risk. Idiosyncratic fit leads to deviations from perfect sorting, and in this sense provides a socially suboptimal allocation of

---

[29]First, we calculate the average cost of risk for risk-averse individuals, conditional on $\widetilde{\gamma}$, at their optimal distribution of career choices. To do so, we follow our previous structure and divide income risk into 10 ranges, and denote by σ_j^2 the weighted-average value within each range. The average cost of risk for individual i with risk aversion $\widetilde{\gamma}_i$ is then simply $\sum_j f(\sigma_j^2 | \widetilde{\gamma}) \times (-\frac{1}{2}\widetilde{\gamma}\sigma_j^2)$, where $f(\sigma_j^2 | \widetilde{\gamma})$ is derived from our structural estimates of $f(\sigma_i^2 | \widetilde{\gamma} = 0)$ and β. Next, we calculate the hypothetical average cost for risk-averse workers who choose careers in the same proportion as risk-neutral workers: $\sum_j f(\sigma_j^2 | \widetilde{\gamma} = 0) \times (-\frac{1}{2}\widetilde{\gamma}\sigma_j^2)$.

risk in the population. How costly is this deviation? How much better off would people be if their chosen career, ceteris paribus, lied on the perfect-sorting risk line?

We estimate that the average observed welfare cost of risk ($\frac{1}{N}\sum_i \frac{1}{2} \times \tilde{\gamma}_i \times \sigma_c^2$) is 0.140. On average, individuals would be willing to give up approximately 14% of their income to eliminate income risk. Under the counterfactual of perfect sorting, this welfare cost falls to 0.067. The difference, approximately 7.3% of income, can be viewed as the potential welfare gain associated with eliminating the mismatch of risk-averse people into risky careers.

7. Conclusion

This paper has documented that those who self-identify as risk tolerant are more likely to have volatile incomes, but that this correlation is far from perfect. Our model of optimal career choice gives this correlation an economic interpretation: an individual's perceived idiosyncratic taste for and/or skill in a career varies dramatically from one career to another. The presence of a modest income gap between high- and low-risk careers for more risk-averse people—relative to more risk-tolerant ones—indicates that both idiosyncratic skill and idiosyncratic taste are meaningful determinants of occupational choice.

The results presented here have important implications for on-the-job training, and more generally for investment in human capital. Individuals choose the career with the best fit, the career that jointly maximizes their enjoyment of and skill in that career. Our counterfactual shows that career-specific training may be of limited value because some of the individuals who choose to participate in the program will have limited interest in or aptitude for those careers. Our parameter estimates suggest that these concerns are quantitatively important.

Appendix

A.1 *The expected value of a chosen career*

If x is an extreme value distributed (Type 1, Gumbel) random variable, with location parameter μ and scale parameter β, the cdf is given by $F(x \mid \mu, \beta) = e^{-e^{-(x-\mu)/\beta}}$. Imagine that x is formed as the maximum of a collection of independent and identically distributed (i.i.d.) extreme value distributed random variables. Let r be a subset containing a share s_r of this collection. Then the maximum value within r will have the cdf

$$F_r(x \mid \mu_r, \beta_r) = \left(F(x \mid \mu, \beta)\right)^{s_r} = \left(e^{-e^{-(x-\mu)/\beta}}\right)^{s_r} \tag{27}$$

$$= e^{-e^{-(x-\mu-\beta \ln(s_r)/\beta}} = F\left(x \mid \mu + \beta \ln(s_r), \beta\right). \tag{28}$$

The expected value of the maximum of x is $\mu + \beta\gamma_{em}$, which implies the subset r has an expected maximum value of $\mu + \beta \ln(s_r) + \beta\gamma_{em}$. Given that $\varepsilon_{i,c}$ has an extreme value

distribution in each s_r (as defined in Section 2.3), this informs the transformation

$$W(i,r) \equiv \max_{c \in r} V(i,c) = \max_{c \in r} \left[y_i^I + y_c^C + v(x_i^I, x_c^C) - \frac{1}{2} \widetilde{\gamma}_i \sigma_c^2 + \varepsilon_{i,c} \right] \tag{29}$$

$$= y_i^I + y_c^C + v(x_i^I, x_c^C) - \frac{1}{2} \widetilde{\gamma}_i \sigma_c^2 + \max_{c \in r} \varepsilon_{i,c} \tag{30}$$

$$= y_i^I + y_c^C + v(x_i^I, x_c^C) - \frac{1}{2} \widetilde{\gamma}_i \sigma_r^2 + EV(\mu + \beta \ln(s_r), \beta) \tag{31}$$

$$= \mu + \beta \gamma_{em} + y_i^I + \beta \ln \left(s_r e^{(y_c^C + v(x_i^I, x_c^C) - \frac{1}{2} \widetilde{\gamma}_i \sigma_r^2)/\beta} \right)$$
$$+ EV(-\beta \gamma_{em}, \beta),\,^{[30]} \tag{32}$$

where equation (32) follows from pulling $\beta \ln(s_r)$ out of the expectation, adding/subtracting $\beta \gamma_{em}$, and combining terms.

With equation (32) in hand, we can compute analytically the probability that a given range s_r will produce the maximum value. In particular, define

$$Z_r \equiv a + \beta \ln(s_r) + EV(-\beta \gamma_{em}, \beta), \tag{33}$$

$$s_r = \mathrm{prob}(Z_r > Z_s, \forall s \neq r); \qquad \sum_r s_r = 1. \tag{34}$$

Combining equations (32) and (34) gives the probability that an individual's preferred career will lie in range r:

$$\mathrm{prob}\big(W(i,r) > W(i,q), \forall q \neq r\big) = \frac{s_r e^{(y_r^C + v(x_i^I, x_r^C) - \frac{1}{2} \widetilde{\gamma}_i \sigma_r^2)/\beta}}{\sum_q s_q e^{(y_q^C + v(x_i^I, x_q^C) - \frac{1}{2} \widetilde{\gamma}_i \sigma_q^2)/\beta}}. \tag{35}$$

The probability that a given range will have the highest value (equation (35)) is nothing more than the probability density function (pdf), the joint distribution of attributes X^C of careers chosen given i:

$$f(X^C \mid i) \equiv \mathrm{prob}\big(W(i,r) > W(i,q), \forall q \neq r\big).$$

We rewrite equation (35) by taking the size of each range to zero, so that the sums become integrals and s_r becomes $f^C(X^C)$:

$$f(X^C \mid X^I) = \frac{f(X^C \mid \widetilde{\gamma} = 0, X^I) e^{-\frac{1}{2} \frac{1}{\beta} \widetilde{\gamma} \sigma^2}}{\iiiint_{X_q^C} f(X_q^C \mid \widetilde{\gamma} = 0, X^I) e^{-\frac{1}{2} \frac{1}{\beta} \widetilde{\gamma} \sigma_q^2} \, dX_q^C}, \tag{36}$$

---

[30]Note that the extreme value distribution (EV) in equation (32) has mean zero. As C increases, μ increases by $\ln(\ln(N_C))$. We envision a limiting setting in which for all c, y^C falls at this same rate. Therefore, $\lim_{N_C \to \infty} \mu + y_r^C$ converges to a constant. As the number of careers increases, the average quality of a randomly chosen career falls to keep the expected quality of the best career constant.

$$f\big(X^C \mid \widetilde{\gamma} = 0, X^I\big) = \frac{f^C(X^C)e^{((y^C + v(x^I, x^C))/\beta)}}{\iiiint_{X_q^C} f^C(X_q^C)e^{((y^C + v(x^I, x^C))/\beta)}\, dX_q^C}. \tag{37}$$

The result is equations (15) and (16).

A.2 *Integrating out unobservables*

The main career attribute of interest is σ_c^2. We want to express the distribution of chosen income risk in terms of only observables. We begin by restating equations (15) and (16) as

$$f\big(X^C \mid X^I\big) = \frac{f(X^C | \widetilde{\gamma} = 0, X^I)e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}}{\iiiint_{X_q^C} f(X_q^C | \widetilde{\gamma} = 0, X^I)e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma_q^2}\, dX_q^C}, \tag{38}$$

$$f\big(X^C \mid \widetilde{\gamma} = 0, X^I\big) = \frac{f^C(X^C)e^{((y^C + v(x^I, x^C))/\beta)}}{\iiiint_{X_q^C} f^C(X_q^C)e^{((y^C + v(x^I, x^C))/\beta)}\, dX_q^C}, \tag{39}$$

which taken together imply that

$$f\big(X^C \mid X^I\big) = k_1\big(X^I\big)f^C\big(X^C\big)e^{((y^C + v(x^I, x^C))/\beta)}e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}, \tag{40}$$

$$k_1\big(X^I\big) \equiv \frac{1}{\iiiint_{X_q^C} f^C(X^C)e^{((y^C + v(x^I, x^C))/\beta)}e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}\, dX_q^C}. \tag{41}$$

We then integrate equation (40) over y^C and x^{CU} to obtain the marginal distribution

$$f\big(\sigma^2, x^{CO} \mid X^I\big) \tag{42}$$
$$= \iint k_1\big(X^I\big)f^C\big(X^C\big)e^{((y^C + v(x^I, x^C))/\beta)}e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}\, dy^C\, dx^{CU}$$

$$= k_1\big(X^I\big)f^C\big(\sigma^2, x^{CO}\big) \iint f^C\big(y^C, x^{CU} | \sigma^2, x^{CO}\big)e^{((y^C + v(x^I, x^C))/\beta)}$$
$$\times e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}\, dy^C\, dx^{CU}. \tag{43}$$

Equation (43) results from pulling k_1 and $f^C(\sigma^2, x^{CO})$ out of the integral, because they do not depend on y^C or x^{CU}. We can then write the double integral in equation (43) as an expectation over y^C and x^{CU},

$$E\big[e^{((y^C + v(x^I, x^C))/\beta)}e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} | \sigma^2, x^{CO}, X^I\big], \tag{44}$$

in which case equation (43) becomes

$$f\big(\sigma^2, x^{CO} \mid X^I\big) \propto f^C\big(\sigma^2, x^{CO}\big)E\big[e^{((y^C + v(x^I, x^C))/\beta)}e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} | \sigma^2, x^{CO}, X^I\big]. \tag{45}$$

Conditioning on x^{CO} using Bayes' rule, equation (45) becomes

$$f\big(\sigma^2 \mid x^{CO}, X^I\big) \propto \frac{f^C\big(\sigma^2, x^{CO}\big)}{f\big(x^{CO}\big)} E\big[e^{((y^C+v(x^I,x^C))/\beta)}e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}\mid\sigma^2, x^{CO}, X^I\big]$$

$$\Rightarrow \quad f\big(\sigma^2 \mid x^{CO}, X^I\big) \propto f^C\big(\sigma^2|x^{CO}\big)\frac{f^C\big(x^{CO}\big)}{f\big(x^{CO}\big)} \tag{46}$$

$$\times E\big[e^{((y^C+v(x^I,x^C))/\beta)}e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}\mid\sigma^2, x^{CO}, X^I\big].$$

So far, we have transformed equations (38) and (39) into equation (46) without additional assumptions. Next, we want to write equation (46) as a shift of choices made by risk-neutral individuals, meaning we want to take $e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}$ out of the expectation. To do so, we must make the assumption

$$E\big[e^{((y^C+v(x^I,x^C))/\beta)}e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}\mid\sigma^2, x^{CO}, X^I\big]$$

$$= E\big[e^{((y^C+v(x^I,x^C))/\beta)}\mid\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\big]e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}. \tag{47}$$

If σ^2 and $e^{((y^C+v(x^I,x^C))/\beta)}$ are correlated (so that risky jobs have more or less appealing other attributes), this must be equally true for all γ. Plugging the assumption from equation (47) into equation (46) yields the distribution of risk choices for risk-neutral individuals (imposing $\widetilde{\gamma}=0$) and for risk-averse individuals relative to risk-neutral individuals:

$$f\big(\sigma^2 \mid x^{CO}, X^I, \widetilde{\gamma}=0\big) \propto f^C\big(\sigma^2|x^{CO}\big)\frac{f^C\big(x^{CO}\big)}{f\big(x^{CO}\big)}$$

$$\times E\big[e^{((y^C+v(x^I,x^C))/\beta)}\mid\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\big], \tag{48}$$

$$f\big(\sigma^2 \mid x^{CO}, X^I, \widetilde{\gamma}\big) \propto f\big(\sigma^2 \mid x^{CO}, X^I, \widetilde{\gamma}=0\big)e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}. \tag{49}$$

To this point, we have integrated out career-specific unobservables, transforming equations (15) and (16) into equations (48) and (49). Next, we integrate out unobservable individual attributes. To do so, we separate X^I into its constituent parts, rewriting equation (49) as

$$f\big(\sigma^2 \mid x^{CO}, y^I, \widetilde{\gamma}, x^{IO}, x^{IU}\big)$$

$$\propto f^C\big(\sigma^2|x^{CO}\big)\frac{f^C\big(x^{CO}\big)}{f\big(x^{CO}\big)} E\big[e^{((y^C+v(x^I,x^C))/\beta)}\mid\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\big]e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}. \tag{50}$$

Using Bayes' rule, we transform equation (50) into the joint distribution of σ^2 and x^{IU} (dropping y^I because it affects all careers equally):

$$
\begin{aligned}
&f\big(\sigma^2, x^{IU} \mid x^{CO}, \widetilde{\gamma}, x^{IO}\big) \\
&\qquad \propto f^C\big(\sigma^2 | x^{CO}\big) \frac{f^C\big(x^{CO}\big)}{f\big(x^{CO}\big)} f\big(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}\big) \\
&\qquad\quad \times E\big[e^{((y^C + v(x^I, x^C))/\beta)} | \sigma^2, x^{CO}, X^I, \widetilde{\gamma} = 0\big] e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2},
\end{aligned}
\tag{51}
$$

or, equivalently,

$$
\begin{aligned}
&f\big(\sigma^2, x^{IU} \mid x^{CO}, \widetilde{\gamma}, x^{IO}\big) \\
&\quad = k_2\big(x^{CO}, \widetilde{\gamma}, x^{IO}\big) f^C\big(\sigma^2 | x^{CO}\big) \frac{f^C\big(x^{CO}\big)}{f\big(x^{CO}\big)} f\big(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}\big) \\
&\qquad \times E\big[e^{((y^C + v(x^I, x^C))/\beta)} | \sigma^2, x^{CO}, X^I, \widetilde{\gamma} = 0\big] e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}; \\
&k_2\big(x^{CO}, \widetilde{\gamma}, x^{IO}\big) \\
&\quad \equiv \left(\iint \left[\begin{array}{c} f^C\big(\sigma^2 | x^{CO}\big) \frac{f^C\big(x^{CO}\big)}{f\big(x^{CO}\big)} f\big(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}\big) \\ \times E\big[e^{((y^C + v(x^I, x^C))/\beta)} | \sigma^2, x^{CO}, X^I, \widetilde{\gamma} = 0\big] e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} \end{array} \right] d\sigma^2 \, dx^{IU} \right)^{-1}.
\end{aligned}
\tag{52}
$$

We then integrate over individual unobservables (x^{IU}):

$$
\begin{aligned}
&f\big(\sigma^2 \mid x^{CO}, \widetilde{\gamma}, x^{IO}\big) \\
&\quad = \int k_2\big(x^{CO}, \widetilde{\gamma}, x^{IO}\big) f^C\big(\sigma^2 | x^{CO}\big) \frac{f^C\big(x^{CO}\big)}{f\big(x^{CO}\big)} \\
&\qquad \times f\big(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}\big) E\big[e^{((y^C + v(x^I, x^C))/\beta)} | \sigma^2, x^{CO}, X^I, \widetilde{\gamma} = 0\big] \\
&\qquad \times e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} \, dx^{IU} \\
&\Rightarrow \quad f\big(\sigma^2 \mid x^{CO}, \widetilde{\gamma}, x^{IO}\big) \\
&\qquad = k_2\big(x^{CO}, \widetilde{\gamma}, x^{IO}\big) f^C\big(\sigma^2 | x^{CO}\big) \frac{f^C\big(x^{CO}\big)}{f\big(x^{CO}\big)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} \\
&\qquad\quad \times \int f\big(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}\big) \\
&\qquad\quad \times E\big[e^{((y^C + v(x^I, x^C))/\beta)} | \sigma^2, x^{CO}, X^I, \widetilde{\gamma} = 0\big] dx^{IU}.
\end{aligned}
\tag{53}
$$
$$
\tag{54}
$$

So as to write the integral over x^{IU} as a part of expectation, we need to impose the assumption

$$\int f(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}) E[e^{((y^C + v(x^I, x^C))/\beta)} \mid \sigma^2, x^{CO}, X^I, \widetilde{\gamma} = 0] \, dx^{IU}$$

$$= E[e^{((y^C + v(x^I, x^C))/\beta)} \mid \sigma^2, x^{CO}, x^{IO}, \widetilde{\gamma} = 0]. \tag{55}$$

Assumption (55) means that the expected value of careers at various income risk levels cannot be differentially affected by individual unobservables for different levels of risk aversion. In this case, we arrive at the final expression for $f(\sigma^2 \mid x^{CO}, \widetilde{\gamma}, x^{IO})$:

$$f(\sigma^2 \mid x^{CO}, \widetilde{\gamma} = 0, x^{IO}) \propto f^C(\sigma^2 \mid x^{CO}) \frac{f^C(x^{CO})}{f(x^{CO})}$$

$$\times E[e^{((y^C + v(x^I, x^C))/\beta)} \mid \sigma^2, x^{CO}, x^{IO}, \widetilde{\gamma} = 0], \tag{56}$$

$$f(\sigma^2 \mid x^{CO}, \widetilde{\gamma}, x^{IO}) \propto f(\sigma^2 \mid x^{CO}, \widetilde{\gamma} = 0, x^{IO}) e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}. \tag{57}$$

Equations (56) and (57) are identical to equations (17) and (18).

A.3 *Separating idiosyncratic taste from skill*

Observed log pay (ignoring the mean-zero income shock ξ) is

$$\log \text{pay}_{i,c} \equiv y_i^I + y_c^C + y^x(x_i^I, x_c^C) + y_{i,c}^\varepsilon. \tag{58}$$

Combining equations (10) and (58) yields

$$\log \text{pay}_{i,c} = V_{i,c} - l^x\widetilde{(x_i^I, x_c^C)} + \frac{1}{2}\widetilde{\gamma}_i \sigma_c^2 - \widetilde{l_{i,c}^\varepsilon}. \tag{59}$$

We can then take the expectation of log pay conditional on career c having the highest $V_{i,c}$ from equation (13):

$$E[\log \text{pay}_{i,c} \mid V_{i,c} \geq V_{i,c'} \, \forall c']$$

$$= \overline{V(X^I)} - l^x\widetilde{(x_i^I, x_c^C)} + \frac{1}{2}\widetilde{\gamma}_i \sigma_c^2 - E[\widetilde{l_{i,c}^\varepsilon} \mid V_{i,c} > V_{i,c'} \, \forall c']$$

$$= \mu + \beta\gamma_{em} + y_i^I + \beta \ln\left(\sum_q s_q e^{(y_q^C + v(x_i^I, x_q^C) - \frac{1}{2}\widetilde{\gamma}_i \sigma_q^2)/\beta}\right)$$

$$- l^x\widetilde{(x_i^I, x_c^C)} + \frac{1}{2}\widetilde{\gamma}_i \sigma_c^2 - E[\widetilde{l_{i,c}^\varepsilon} \mid V_{i,c} > V_{i,c'} \, \forall c']. \tag{60}$$

Next, we take the expectation of $V_{i,c}$ from equation (10) conditional on career c having the highest $V_{i,c}$:

$$E[V(i,c) \mid V_{i,c} \geq V_{i,c'} \, \forall c'] = \overline{V(X^I)} = y_i^I + y_c^C + v(x_i^I, x_c^C) - \frac{1}{2}\widetilde{\gamma}_i \sigma_c^2$$

$$+ E[y_{i,c}^\varepsilon + \widetilde{l_{i,c}^\varepsilon} \mid V_{i,c} \geq V_{i,c'} \, \forall c']. \tag{61}$$

Plugging equation (13) into equation (61) and rearranging terms yields

$$
E\big[y_{i,c}^{\varepsilon} + \widetilde{l_{i,c}^{\varepsilon}} \mid V_{i,c} \geq V_{i,c'}\ \forall c'\big]
$$

$$
= \mu + \beta\gamma_{em} + \beta\ln\!\left(\sum_q s_q e^{(y_q^C + v(x_i^I, x_q^C) - \frac{1}{2}\widetilde{\gamma}_i\sigma_q^2)/\beta}\right) \tag{62}
$$

$$
- y_c^C - v\big(x_i^I, x_c^C\big) + \tfrac{1}{2}\widetilde{\gamma}_i\sigma_c^2.
$$

By assuming joint normality of $y_{i,c}^{\varepsilon}$ and $\widetilde{l_{i,c}^{\varepsilon}}$, so that the signal extraction problem is linear, $E[y_{i,c}^{\varepsilon} + \widetilde{l_{i,c}^{\varepsilon}} \mid V_{i,c} \geq V_{i,c'}\ \forall c']$ from equation (62) identifies $E[\widetilde{l_{i,c}^{\varepsilon}} \mid V_{i,c} > V_{i,c'}\ \forall c']$ in equation (60):

$$
E\big[\widetilde{l_{i,c}^{\varepsilon}} \mid V_{i,c} \geq V_{i,c'}\ \forall c'\big]
$$

$$
= \frac{\mathrm{var}\big(\widetilde{l_{i,c}^{\varepsilon}}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l_{i,c}^{\varepsilon}}\big)}\left(\mu + \beta\gamma_{em} + \beta\ln\!\left(\sum_q s_q e^{(y_q^C + v(x_i^I, x_q^C) - \frac{1}{2}\widetilde{\gamma}_i\sigma_q^2)/\beta}\right)\right. \tag{63}
$$

$$
\left. - y_c^C - v\big(x_i^I, x_c^C\big) + \tfrac{1}{2}\widetilde{\gamma}_i\sigma_c^2\right).
$$

Plugging equation (63) into equation (60) yields

$$
E\big[\log \mathrm{pay}_{i,c} \mid V_{i,c} \geq V_{i,c'}\ \forall c'\big]
$$

$$
= (\mu + \beta\gamma_{em})\left(1 - \frac{\mathrm{var}\big(\widetilde{l_{i,c}^{\varepsilon}}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l_{i,c}^{\varepsilon}}\big)}\right)
$$

$$
+ y_i^I + \beta\ln\!\left(\sum_q s_q e^{(y_q^C + v(x_i^I, x_q^C) - \frac{1}{2}\widetilde{\gamma}_i\sigma_q^2)/\beta}\right)\left(1 - \frac{\mathrm{var}\big(\widetilde{l_{i,c}^{\varepsilon}}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l_{i,c}^{\varepsilon}}\big)}\right) \tag{64}
$$

$$
+ y_c^C \frac{\mathrm{var}\big(\widetilde{l_{i,c}^{\varepsilon}}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l_{i,c}^{\varepsilon}}\big)} - v\big(x_i^I, x_c^C\big)\left(1 - \frac{\mathrm{var}\big(\widetilde{l_{i,c}^{\varepsilon}}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l_{i,c}^{\varepsilon}}\big)}\right)
$$

$$
+ \tfrac{1}{2}\widetilde{\gamma}_i\sigma_c^2\left(1 - \frac{\mathrm{var}\big(\widetilde{l_{i,c}^{\varepsilon}}\big)}{\mathrm{var}\big(y_{i,c}^{\varepsilon} + \widetilde{l_{i,c}^{\varepsilon}}\big)}\right).
$$

References

Abowd, J. and D. Card (1989), "On the covariance structure of earnings and hours changes." *Econometrica*, 57 (2), 411–445. [563]

Alvarez, J., M. Browning, and M. Ejrnaes (2010), "Modelling income processes with lots of heterogeneity." *Review of Economic Studies*, 77 (4), 1353–1381. [555]

Barsky, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro (1997), "Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study." *Quarterly Journal of Economics*, 112 (2), 537–579. [555]

Berkhout, P., J. Hartog, and D. Webbink (2010), "Compensation for earnings risk under worker heterogeneity." *Southern Economic Journal*, 76 (3), 762–790. [555]

Boskin, M. (2004), "A conditional logit model of occupational choice." *Journal of Political Economy*, 82 (2), 389–398. [555]

Carroll, C. D. and A. A. Samwick (1997), "The nature of precautionary wealth." *Journal of Monetary Economics*, 40 (1), 41–71. [564]

Couch, K. A. and D. W. Placzek (2010), "Earnings losses of displaced workers revisited." *American Economic Review*, 100 (1), 572–589. [576]

Cunha, F., J. Heckman, and S. Navarro (2005), "The 2004 Hicks lecture: Separating uncertainty from heterogeneity in life cycle earnings." *Oxford Economic Papers*, 57 (2), 191–261. [554]

Cunha, F. and J. Heckman (2008), "A new framework for the analysis of inequality." *Macroeconomic Dynamics*, 12 (S2), 315–354. [554]

de Haan, L. and A. Ferreira (2006), *Extreme Value Theory: An Introduction*. Springer. [560]

Diaz-Serrano, L. and J. Hartog (2006), "Is there a risk-return trade-off in educational choices? Evidence from Spain." *Investigaciones Economicas*, 30 (2), 353–380. [555]

Dillon, E. (2016), "Risk and return tradeoffs in lifetime earnings." Report, Arizona State University. [555]

Field, E. (2009), "Educational debt burden and career choice: Evidence from a financial aid experiment at NYU law school." *American Economic Journal: Applied Economics*, 1 (1), 1–21. [555]

Flyer, F. A. (2004), "The influence of higher moments of earnings distributions on career decisions." *Journal of Labor Economics*, 15 (4), 689–713. [555]

Fuchs-Schündeln, N. and M. Schündeln (2005), "Precautionary savings and self-selection: Evidence from the German reunification 'experiment'." *The Quarterly Journal of Economics*, 120 (3), 1085–1120. [555]

Gottschalk, P. and R. Moffitt (2002), "Trends in the transitory variance of earnings in the United States." *Economic Journal*, 112, 68–73. [563, 564]

Harris, M. and Y. Weiss (1984), "Job matching with finite horizon and risk aversion." *Journal of Political Economy*, 92 (4), 758–779. [555]

Hartog, J. (2011), "A risk augmented mincer earnings equation? Taking stockvol." In *Research in Labor Economics*, Vol. 33, Emerald Publishing. [555, 573]

Hartog, J., E. Plug, L. Diaz-Serrano, and J. Vieira (2003), "Risk compensation in wages, a replication." *Empirical Economics*, 28 (3), 639–647. [555]

Hartog, J. and W. Vijverberg (2007), "On compensation for risk aversion and skewness affection in wages." *Labour Economics*, 14 (6), 938–956. [555]

Heckman, J. and B. Honore (1990), "The empirical content of the Roy model." *Econometrica*, 58 (5), 1121–1149. [554]

Hill, M. S. (1991), *The Panel Study of Income Dynamics: A User's Guide*, Vol. 2. Sage Publications, Inc. [563]

Jacobs, B., J. Hartog, and W. Vijverberg (2009), "Self-selection bias in estimated wage premiums for earnings risk." *Empirical Economics*, 37 (2), 271–286. [556]

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993), "Earnings losses of displaced workers." *American Economic Review*, 83 (4), 685–709. [576]

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (2005), "Estimating the returns to community college schooling for displaced workers." *Journal of Econometrics*, 125 (1–2), 271–304. [576]

Jensen, S. T. and S. H. Shore (2011), "Semi-parametric Bayesian modeling of income volatility heterogeneity and dynamics." *Journal of American Statistical Association*, 106, 1280–1290. [555, 565]

Jensen, S. T. and S. H. Shore (2015), "Changes in the distribution of earnings volatility." *Journal of Human Resources*, 50 (3), 811–836. [555, 564, 565, 566, 567, 574]

Keane, M. and K. Wolpin (1997), "The career decisions of young men." *Journal of Political Economy*, 105 (3), 473–522. [554, 555, 562]

Kimball, M. S., C. R. Sahm, and M. D. Shapiro (2008), "Imputing risk tolerance from survey responses." *Journal of the American Statistical Association*, 103, 1028–1038. [555]

Kimball, M. S., C. R. Sahm, and M. D. Shapiro (2009), "Risk preferences in the PSID: Individual imputations and family covariation." *American Economic Review Papers and Proceedings*, 99, 363–368. [555, 567, 568, 573]

Krebs, T. (2003), "Human capital risk and economic growth." *Quarterly Journal of Economics*, 118 (2), 709–744. [555]

LaLonde, R. J. (1995), "The promise of public sector-sponsored training programs." *Journal of Economic Perspectives*, 9 (2), 149–168. [576]

Lazear, E. (2004), "Balanced skills and entrepreneurship." *American Economic Review*, 94 (2), 208–2011. [555]

Levhari, D. and Y. Weiss (1974), "The effect of risk on the investment in human capital." *American Economic Review*, 64 (6), 950–963. [555]

McFadden, D. (1974), "Conditional logit analysis of qualitative choice bea." In *Frontiers in Econometrics* (P. Zarembka, ed.), 105–142. [559]

Meghir, C. and L. Pistaferri (2004), "Income variance dynamics and heterogeneity." *Econometrica*, 72 (1), 1–32. [555, 563, 564, 565, 569]

Mincer, J. (1974), *Schooling, Experience and Earnings*. Columbia University Press. [555]

Mukoyama, T. (2014), "The cyclicality of job-to-job transitions and its implications for aggregate productivity." *Journal of Economic Dynamics and Control*, 39, 1–17. [558]

Pacios-Huerta, I. (2003), "An empirical analysis of the risk properties of human capital returns." *American Economic Review*, 93 (3), 948–964. [555]

Pistaferri, L. (2002), "Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply." *Journal of Labor Economics*, 21 (3). [564]

Roy, A. (1951), "Some thoughts on the distribution of earnings." *Oxford Economic Papers, New Series*, 3 (2), 135–146. [554]

Ruhm, C. J. (1991), "Are workers permanently scarred by job displacements?" *American Economic Review*, 81 (1), 319–324. [576]

Sahm, C. (2012), "How much does risk tolerance change?" *Quarterly Journal of Finance*, 2 (4). [555]

Shimer, R. (2005), "The cyclicality of hires, separations, and job-to-job transitions." *Federal Reserve Bank of St. Louis Review*, 87 (4), 493–507. [558]

Weiss, Y. (1972), "The risk element in occupational and educational choices." *Journal of Political Economy*, 80 (6), 1203–1213. [555]