W. J. Usery Workplace Research Group Paper Series

Working Paper 2013-6-1
June 2013

Identifying Idiosyncratic Career Taste and Skill with Income Risk

Stephen H. Shore
Georgia State University

Daniel Barth
Hamilton College

Shane T. Jensen
University of Pennsylvania

Georgia State University
ANDREW YOUNG SCHOOL
OF POLICY STUDIES

Identifying idiosyncratic career taste and skill with income risk

STEPHEN H. SHORE[1][2][3]
Department of Risk Management and Insurance
The Robinson College of Business, Georgia State University
Atlanta, GA 30303
sshore@gsu.edu

DANIEL BARTH
Department of Economics
Hamilton College
Clinton, NY 13323
dbarth@hamilton.edu

SHANE T. JENSEN
Department of Statistics
The Wharton School, University of Pennsylvania
Philadelphia, PA 19104
stjensen@wharton.upenn.edu

June 6, 2013

Abstract

How important to well-being is choosing a career with the right fit? This question is difficult to answer because we observe individuals only in their chosen careers, not in the other (presumably inferior) options they did not choose. To overcome this problem, we use expected utility to cardinalize a logit 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 - 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 estimate the model using individual-specific measures of income volatility to proxy for income risk and survey questions about hypothetical income gambles to proxy for risk preference, both from the PSID. We separate idiosyncratic career taste from skill using the pay gap between high- and low-income risk people with high and low risk-aversion.

---

[1] We thank Ellie Pavlick for excellent research assistance. We thank seminar participants at Dartmouth College, Brown University, the University of Maryland – College Park, Kyoto University, the Bureau of Labor Statistics, the Risk Theory Society 2010 meeting, the SOLE/EALE 2010 meeting, the 2010 NBER Summer Institute, New York University, the University of North Carolina, and the University of Nevada Las Vegas. Previous versions of this draft have circulated as "The Human Capital Risk-Return Menu" and "A Structural Model of Career Choice."

[2] JEL: J24 (Human Capital; Skills; Occupational Choice; Labor Productivity)

[3] keywords: occupational choice, career choice, income risk, idiosyncratic taste and skill

1 Introduction and Motivation

The Road Not Taken

Two roads diverged in a wood, and I–

I took the one less traveled by,

And that has made all the difference.

–Frost (1920)

This paper aims to test the claim implicit in Frost's poem.[1] If we think of Frost's "roads" as career paths, how important is road choice to well-being? How important is a chosen road's idiosyncratic fit, of finding the road best suited to idiosyncratic tastes and skills? In the language of economists, how much money would an individual demand to accept another road? We use a model of road choice to estimate the distribution of road quality options.

The influence of idiosyncratic taste and skill on career choice is important for a few reasons. To the degree that idiosyncratic career taste affects career choice, workers who are forced to change careers may face substantial welfare costs beyond their well-studied forgone income (see Jacobson, LaLonde, and Sullivan (1993), Ruhm (1991), and Couch and Placzek (2010) as just a few examples). Similarly, government programs that provide training in a particular career (LaLonde (1995), Jacobson, LaLonde, and Sullivan (2005)) may be of limited value if the individuals assigned to that program lack interest in or aptitude for that career.

Such a model for career choice is challenging for three reasons. First and most obviously, we never observe the full choice set; we observe only the career that was chosen, but not the (presumably inferior) options that were not chosen.[2] Second,

---

[1] In fact, a large body of poetry criticism argues that Frost's intended meaning was not the literal and commonly believed one (Pritchard, 1984). Scholars note "Frost's decision to make his two roads not very much different from one another, for passing over one of them had the effect of wearing them 'really about the same.'"(Monteiro, 1988)

[2] This is merely a career choice application of the classic problem of measuring treatment effects (surveyed in Holland, 1986).

data on a chosen career's attributes may be limited. Third, career choice decisions depend substantively on non-pecuniary factors (e.g. working conditions, interest in the work, etc.) that economists typically don't observe and about which we have few interesting, quantitative hypotheses.

We overcome these problems using variables we do observe and about which we have well-developed models: a chosen career's income risk and a person's coefficient of relative risk aversion. Expected utility provides a quantitative theory of the dollar equivalent cost of a given level of income risk for a person with a given level of risk aversion. While risk-related variables may affect career choice, they are hardly the only drivers of such choice. We use them to identify the importance of other factors, including the unobservable ones that we model as idiosyncratic. This is possible if we assume that risk aversion is an attribute of a person (and therefore not affected by the career they choose) and income risk is an attribute of a career (and therefore not affected by the risk aversion of the person who chooses it).

We envision a model in which people are endowed with a preference for risk and level of overall ability, broadly conceived as risk-free, career-independent earning potential. Careers differ in income risk and typical, worker-independent pay. We allow individuals to have idiosyncratic taste for – or skill and correspondingly higher pay in – some careers over others. Income risk is resolved after the career is chosen, though workers know careers' risks when they choose a career. Both careers and individuals may differ in other attributes, which may be unobservable to the econometrician though all are observable to workers. We abstract from search frictions and bounded rationality concerns; workers observe and understand the full set of career options from which they can choose. In section 5.4 we show how our results may be reinterpreted as the welfare cost of search frictions as opposed to the importance of idiosyncratic taste and skill. We also abstract from intertemporal concerns; workers chose careers only once. Unfortunately, data does not exist which could identify a dynamic version of our model. We have neither the data to estimate the time-series of risk-aversion, nor a compelling way to identify risk-related career changes.

Figure 1 shows that, *ceteris paribus*, the relationship between income risk and risk-aversion will be weakly negatively monotonic; risk tolerant individuals will choose the riskiest careers (which will carry a compensating wage differential for income risk) while risk intolerant individuals will choose the safest careers.[3] The local risk-return trade-off (the marginal risk premium) is determined by the risk-aversion (slope of the risk-return indifference curve) of the marginal individual at that quantity of risk. There is already a large empirical and theoretical literature on this risk-return relationship.[4] We aim to see if this well-studied and well-understood risk-return relationship can be used to identify other parameters in a career choice model. In particular, we identify the importance of idiosyncratic taste and skill from the risk-based mismatch of risk-averse people into risky careers.

Idiosyncratic tastes and skills will lead some highly risk-averse people to choose careers with high income risk and some risk-tolerant people to choose careers with low income risk. Expected utility gives the welfare cost of deviating from an anticipated income risk choice for someone with a given risk preference. By observing the distribution of these deviations we can back out the dollar-equivalent value of the idiosyncratic tastes or skills that made these deviations optimal. The disper-

---

[3] Deleire and Levy (2004) present a similar figure in which people with heterogeneous preference for injury (not income risk) sort into safer and riskier (by probability of injury) jobs.

[4] Weiss (1972) and Pacios-Huerta (2003) identify the risk attributes associated with various demographic groups. Hartog, Plug, Diaz-Serrano, and Vieira (2003), Hartog and Vijverberg (2007), and Berkhout, Hartog, and Webbink (2006) estimate "risk-augmented Mincer (1974) equations" (surveyed in Hartog (2009)) to identify the empirical risk-return relationship; Berkhout, Hartog, and Webbink (2006) allow this relationship to vary with demographic characteristics. Flyer (2004) identifies the impact of income risk on occupational choice (while 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).

Figure 1: Sorting of the Risk Tolerant into Volatile Careers

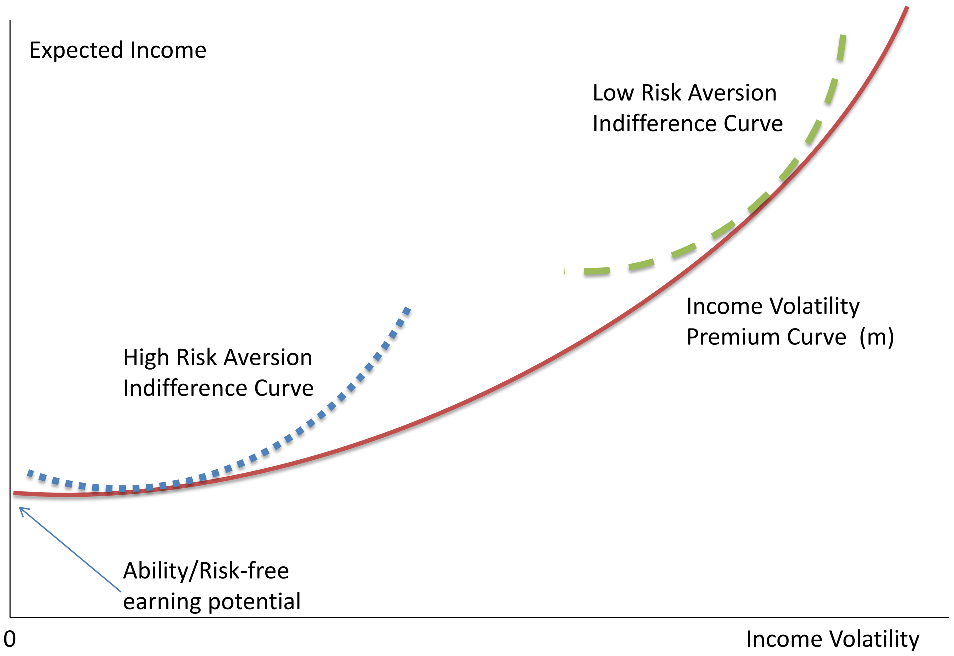


Figure 1 presents a stylized risk-return menu. The solid curve represents the menu of risk-return options, which has a positive slope reflecting the increased compensation for taking higher income risk. The dashed curve represents the indifference curve of a more risk-tolerant individual, while the dotted curve represents the indifference curve of a more risk-averse individual. Tangencies reflect the optimal decision of each individual for the given risk-return menu.

sion of idiosyncratic taste and skill is identified from shifts in the distribution of income risk as risk-aversion changes. The tighter the correlation between income risk and risk aversion, the lower the implied dispersion of idiosyncratic taste and skill. We estimate a lower bound of 36% of income on the standard deviation of these idiosyncratic factors.

The model provides a clean mapping from a feature of the data to an *a priori* seemingly unrelated parameter. The feature of the data is the degree to which the distribution of chosen income risk shifts with risk aversion; the parameter is the importance (formally, the standard deviation as expressed in log-income equivalent units) of idiosyncratic taste for or skill in one career over another. The model implies that a variety of nuisance parameters, such as the distribution of career options and their quality, are absorbed by the distribution of income risk chosen by risk neutral people; the key parameter of interest is identified only from the shift in that distribution as risk-aversion increases.

This framework provides an application for the recent literature on heterogeneity in income volatility. Income volatility is frequently used as a proxy for income risk. Meghir and Pistaferri (2004) and Alvarez, Browning, and Ejrnaes (2001) show that income volatility differs across individuals. Jensen and Shore (2009a,b) 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 they would take a series of hypothetical income gambles (Barsky, Juster, Kimball, and Shapiro, 1997; Sahm, 2007; Kimball, Sahm, and Shapiro, 2008, 2009). We merge these risk-aversion values with estimates of individuals' volatilities in the PSID, both from Jensen and Shore (2009a,b) and also from Meghir and Pistaferri (2004). Individuals who self-identify as risk tolerant tend to have more volatile income streams. At the same time, we observe a non-degenerate joint distribution of income volatility and risk tolerance; conditional on observed risk tolerance, individuals choose a wide variety of income volatilities.

Our parametric and statistical assumptions imply a logit structure (McFadden, 1974). Logit models have long been used in the reduced-form occupational choice literature to study the relative importance of covariates on choice from a finite list of observed careers (Boskin, 2004; Field, 2009). Without an economic model, the multinomial logit setting can be identified only up to a normalization: doubling the utility from all careers (doubling all coefficients and the error term) has no effect on career choice. Our model provides a cardinalization of a logit model of career risk choice, so that all estimates can be expressed in terms of their (log) certainty equivalents.[56] Because we assume a continuum of careers, our model has the continuous logit structure previously used to study home location choice (Ben-Akiva, Litinas, and Tsunokawa, 1985). In the housing application, there is effectively a continuum of homes located on a two-dimensional plane; while researchers may observe few home attributes, they do observe the home's location in this plane. In our setting, we don't observe the exact chosen career, but we observe where that career is located along an income risk "line."[7]

Naturally, our cardinalization inherits the econometric limitations of any logit model; anything that will bias logit coefficient estimates will be a problem for us. Most significantly, we require that risk aversion is associated with a person, and income risk is associated with a career. In truth, income risk may depend on unobservable individual attributes as well, and might be correlated with risk-aversion. In this case, to the degree that risk-averse people will make any career less risky, our estimates of the variance of idiosyncratic factors will be biased downward and

---

[5] Keane and Wolpin (1997) provides an alternative – and very different – cardinalization within a model of optimal educational investment and subsequent choice from five broad career categories.

[6] The common reduced-form alternative is to use pay as one of the covariates that influences occupational choice, giving other coefficients a dollar-equivalent interpretation (Willis and Rosen, 1979; Robertson and Symons, 1990; Siow, 1984). This alternative assumes away the problem that we do not observe occupational pay, but rather the pay of those who choose an occupation. High pay in a given occupation may reflect not just high pay for that occupation but also high ability (or idiosyncratic individual-career-specific skill) among those who choose that occupation. In some sense, this assumption is the first-moment analog to the one we make in this paper; we assume that income risk (the second moment of pay) is associated with a career and not the person that chooses this career; to use pay as a numeraire, we must assume that expected pay (the first moment of pay) is associated with a career and not the person who chooses that career.

Dynamic data can be used to overcome the problem that ability may be correlated with pay; pay changes resulting from occupational changes can be used to estimate career-specific effects holding overall individual ability fixed (Stinebrickner, 2001). This approach does not tackle the problem that idiosyncratic individual-career-specific skill may change at such transitions, with some careers systematically receiving workers for whom that career is a better "fit".

[7] While most papers on occupational choice rely on choice from a finite and specified list of occupational options, some papers allow the number of occupations to be very large, using occupational categories primarily to identify the attributes of chosen careers. For example, Deleire and Levy (2004) use 46 occupational codes to map their occupational attribute of interest (injury and fatality risk) to individuals who choose those occupations. In our case, panel data allow us to obtain individual-specific estimates of our career attribute of interest (income volatility as a measure of income risk), so we can examine career attribute choice without explicitly observing the chosen career.

can be viewed as a lower bound. Similarly, we might imagine that risky careers lead individuals to become more risk tolerant; again, this biases estimates of the importance of idiosyncratic factors downward.

The 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.

The rest of the paper is organized as follows: Section 2 presents the model; Section 3 discusses the data used in estimating the model; Section 4 presents the estimation strategy; Section 5 offers estimation results; and Section 6 concludes the paper.

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. 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 $c \in \{1, ..., N_C\}$. Careers have four attributes, $X^C \equiv \{\sigma^2, y^C, x^{CO}, x^{CU}\}$; $X_c^C \equiv \{\sigma_c^2, y_c^C, x_c^{CO}, x_c^{CU}\}$ is the set of attributes for career c. σ_c^2 is a measure of the income risk in career c.[8] y_c^C is a career-specific measure of log pay in career c. $x_c^C \equiv \left[x_c^{CO}; x_c^{CU}\right]$ is a vector of covariates or attributes of career c; x_c^{CO} are the attributes observable to the econometrician and to workers; x_c^{CU} are 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}), while the noisiness of a career is a typically unobserved career attribute (contained in x_c^{CU}). We discuss the identifying restrictions on x^{CU} on page 17.

Later, we will consider a continuum of atomistic careers, so that $N_C \to \infty$. In this case, $f^C(X^C)$ is the distribution of career attributes, taken over the set of *possible* careers. Naturally, in equilibrium some careers may be chosen more than others, so that the distribution of career *options* will typically not be the distribution of *chosen* careers.

---

[8] 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.

2.1.2 People

People are indexed by $i \in \{1, ..., N_I\}$. People have four attributes, $X^I \equiv \{\gamma, y^I, x^{IO}; x^{IU}\}$; $X_i^I \equiv \{\gamma_i, y_i^I, x_i^{IO}, x_i^{IU}\}$ is the set of attributes for person i and $f^I(X^I)$ is the distribution of attributes in the population. γ_i is a measure of risk-aversion for person i. y_i^I is a person-specific measure of log pay (general ability or productivity) for person i. $x_i^I \equiv \left[x_i^{IO}; x_i^{IU}\right]$ 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; 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}). We discuss the identifying restrictions on x^{IU} on page 17.

2.1.3 Individual-Career-Specific Fit

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

$f_{i,c}^\varepsilon(X^\varepsilon)$ is the joint distribution of $X_{i,c}^\varepsilon \equiv \{y_{i,c}^\varepsilon, l_{i,c}^\varepsilon\}$. 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 when $c \neq c'$ 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 (for leisure). 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 one and do not allow for heterogeneity in α.

For simplicity, we assume a one-period model in which income Y is merely equal to consumption.[9] Log income in equation (3) is the sum of: career-specific pay (y^C), including a premium for size, risk, or non-pecuniary attributes; individual-specific pay or ability (y^I); the effect of the interaction of individual- and career-specific covariates on pay ($y^x(x_i^I, x_c^C)$); individual-career-specific 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 one. Log enjoy-

---

[9] It is straightforward extend this to a multi-period setting in which a one-time career decision affects income dynamics, and consumption and saving respond optimally to income shocks. This richer structure loses the clean analytic framework presented below, but it is easy to implement numerically. We omit it here for parsimony.

ment 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). Covariates affect the pay and enjoyment of individual i in career c as follows:

$$y^x(x_i^I, x_c^C) \equiv \iota' \left(\theta^y \cdot \left(x_i^I x_c^{C\prime}\right)\right) \iota;\ l^x(x_i^I, x_c^C) \equiv \iota' \left(\theta^l \cdot \left(x_i^I x_c^{C\prime}\right)\right) \iota.\text{[10]} \tag{5}$$

The model explicitly assumes that individuals never switch careers. This is surely a restrictive assumption. However, this restriction 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:

$$\begin{aligned} V(i,c) \equiv \frac{\ln\left(\left(1-\gamma_i\right) Eu\right)}{\left(1-\alpha\right)\left(1-\gamma_i\right)} &= y_i^I + y_c^C + y^x(x_i^I, x_c^C) + y_{i,c}^\varepsilon \\ &+ \frac{\alpha}{1-\alpha}\left(l_{i,c}^\varepsilon + l^x(x_i^I, x_c^C)\right) - \frac{1}{2}(\alpha + \gamma_i - \alpha\gamma_i)\sigma_c^2. \end{aligned} \tag{6}$$

---

[10] The effect of covariates on pay and enjoyment depends on coefficient matrices, θ^y and θ^l:

$$\theta^y = \begin{bmatrix} \theta^y_{OO} & \theta^y_{OU} \\ \theta^y_{UO} & \theta^y_{UU} \end{bmatrix}; \theta^l = \begin{bmatrix} \theta^l_{OO} & \theta^l_{OU} \\ \theta^l_{UO} & \theta^l_{UU} \end{bmatrix}.$$

θ^y and θ^l are matrices of size $N_{IO} + N_{IU}$ by $N_{CO} + N_{CU}$, where N_{CO} and N_{CU} are the number of observable and unobservable career attributes, respectively, and N_{IO} and N_{IU} are the number of observable and unobservable individual attributes, respectively. To explain the notation here, θ^y_{OU} is a N_{IO} by N_{CU} matrix that gives the impact on pay of the interaction of observable individual attributes with unobservable career attributes. Note that "$\cdot$" indicates pair-wise multiplication, and ι is a vector of ones so that we merely sum up all elements of the matrix of effects, $\theta^y \cdot \left(x_i^I x_c^{C\prime}\right)$ or $\theta^l \cdot \left(x_i^I x_c^{C\prime}\right)$.

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 $l^{\varepsilon}_{i,c}$ in equation (6), which informs the following transformations,

$$l^x\widetilde{(x^I_i, x^C_c)} \equiv \frac{\alpha}{1-\alpha} l^x(x^I_i, x^C_c); \widetilde{l^{\varepsilon}_{i,c}} \equiv \frac{\alpha}{1-\alpha} l^{\varepsilon}_{i,c}. \tag{7}$$

We also make a transformation to risk aversion,

$$\widetilde{\gamma_i} \equiv \alpha + \gamma_i - \alpha\gamma_i \tag{8}$$

to match the object estimated in the PSID (as discussed on page 22).

We can then re-write the value of each career as:

$$V(i,c) = y^I_i + y^C_c + y^x(x^I_i, x^C_c) + l^x\widetilde{(x^I_i, x^C_c)} - \frac{1}{2}\widetilde{\gamma_i}\sigma^2_c + y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}}. \tag{9}$$

If we group pecuniary and non-pecuniary idiosyncratic terms and also group pecuniary and non-pecuniary covariate terms as

$$\varepsilon_{i,c} \equiv y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}} \text{ and } v\left(x^I_i, x^C_c\right) \equiv y^x(x^I_i, x^C_c) + l^x\widetilde{(x^I_i, x^C_c)}, \tag{10}$$

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

$$V(i,c) = y^I_i + y^C_c + v\left(x^I_i, x^C_c\right) - \frac{1}{2}\widetilde{\gamma_i}\sigma^2_c + \varepsilon_{i,c}. \tag{11}$$

Equation (11) gives career choice a standard, random utility, multinomial logit 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 $v\left(x_i^I, x_c^C\right)$ 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 $v\left(x_i^I, x_c^C\right)$ 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.[11]

At an interior optimum, the individual's first order condition requires that[12]

$$\frac{dy^C}{d\sigma_c^2} = \frac{1}{2}\widetilde{\gamma}. \tag{12}$$

If equation (12) must hold for each $\{\sigma_c^2, \widetilde{\gamma}(\sigma_c^2)\}$ pair and we know the risk aversion

---

[11] 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.

[12] Obtained by differentiating expected utility in equation (11) with respect to σ_c^2, setting equal to zero, and rearranging terms.

of the marginal individual for each σ_c^2, then we can trace out y^C as

$$y_c^C = y_0^C + \frac{1}{2}\int_0^{\sigma_c^2} \widetilde{\gamma}(x)dx. \tag{13}$$

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.

2.3 Incorporating Idiosyncratic Career Preference

The stylized model in Section 2.2 has a homogeneous risk-return menu. Consequently, we should never observe an individual with higher risk aversion choosing higher income risk. This is wildly at odds with the data, which shows substantial heterogeneity in the volatility associated with individuals with the same survey-based estimate of risk aversion. We model this heterogeneity in equation (11), where $v\left(x_i^I, x_c^C\right)$ is the effect of the interaction of individual- and career-specific covariates on pay and enjoyment (for person i in career c), and $X_{i,c}^{\varepsilon} \equiv \{y_{i,c}^{\varepsilon}, \widetilde{l_{i,c}^{\varepsilon}}\}$ are idiosyncratic individual-career-specific productivity and taste shocks, respectively.

The logit structure we introduce becomes tractable when we work with extreme value errors. Such errors can be obtained under either of two assumptions. Under the first, we require that the underlying idiosyncratic terms ($\varepsilon_{i,c}$) have an extreme value distribution (of Type I, Gumbel). Under the second, we assume that the number of careers N_C be large in the sense that we can use extreme value theory to describe the best career. In this case, we require that the 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). In either case, our aim is to estimate the scale parameter β that governs the distribution either of an extreme-valued $\varepsilon_{i,c}$ or the extreme-valued maxima. When $\varepsilon_{i,c}$ has an extreme value distribution $var(\varepsilon_{i,c}) = \beta^2\pi^2/6$; when its maximum does, $var(\varepsilon_{i,c}) \propto \beta^2\pi^2/6$.[13]

Let r refer to the set of careers for person i in a rectangle on the $\left\{y^C + v\left(x_i^I, x^C\right), \sigma^2\right\}$ 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. Consider the choice among careers c in range r. Taking 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

$$\begin{aligned} \overline{V(X_i^I)} &\equiv E[V(i,r) \mid V(i,r) > V(i,q), \forall q \neq r] \\ &= \mu + \beta\gamma_{em} + y_i^I + \beta\ln(\sum_q s_q e^{(y_q^C + v\left(x_i^I, x_q^C\right) - \frac{1}{2}\widetilde{\gamma}_i\sigma_q^2)/\beta}).^{14} \end{aligned} \tag{14}$$

Note the expected value of a chosen career does not depend on X^C, so that learning the attributes of a chosen career provides no information about expected well-

---

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

[14] 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 $\left\{y^C + v\left(x_i^I, x^C\right), \sigma^2\right\}$ plane.

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

$$prob(V(i,r) \geq V(i,q), \forall q \neq r) \propto s_r e^{(y_r^C + v\left(x_i^I, x_r^C\right) - \frac{1}{2}\widetilde{\gamma}_i \sigma_r^2)/\beta}. \tag{15}$$

The full derivations of equations (14) and (15) are provided in Section A.1 of the Appendix. We re-write equation (15) by taking the size of each range to zero, so that the sums become integrals and s_r becomes $f^C(X^C)$:

$$f(X^C \mid X^I) \propto f(X^C \mid \widetilde{\gamma} = 0, X^I) e^{-\frac{1}{2}\widetilde{\gamma}\sigma^2/\beta}; \tag{16}$$

$$f(X^C \mid \widetilde{\gamma} = 0, X^I) \propto f^C(X^C) e^{(y^C + v\left(x^I, x^C\right))/\beta}. \tag{17}$$

The model implies that a risk-neutral person (equation (17)) will choose careers proportional to their frequency $f^C(X^C)$. *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 higher career-specific pay and enjoyment ($y^C + v\left(x^I, x^C\right)$). These career-specific attributes dominate career frequency when idiosyncratic career taste and skill are relatively unimportant ($\beta \rightarrow 0$). Without idiosyncratic career fit, risk-neutral people will merely choose the career with the highest $y^C + v\left(x^I, x^C\right)$; the distribution of risk choices will be extremely tight around the "best" choice. However, as the importance of idiosyncratic career fit increases ($\beta \rightarrow \infty$), careers are chosen only in proportion to their frequency; the distribution of choices becomes as diffuse as the distribution of careers f^C. We should be unsurprised to see that individual-specific ability (y^I) does not affect career choice as it increases the benefit of all careers equally.

Of course, we don't observe all elements of X^C or X^I. As with any logit model, our estimates may be biased unless we make the assumptions needed to integrate out individual- and career-specific unobservables. We begin by integrating out unobservable components of careers.[15] To do so, we require that $\widetilde{\gamma}$ does not affect the expected payoff of some risk levels more than others, so that

$$E\left[e^{(y^C+v\left(x^I,x^C\right))/\beta} \mid X^I, \sigma^2, x^{CO}\right] \tag{18}$$

does not vary with $\widetilde{\gamma}$. If we take the example from page 8 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.

Then, we integrate out individual-specific unobservables. To do so, we must 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,

$$E\left[e^{(\left(y^C+v\left(x^I,x^C\right)\right)/\beta)}|\sigma^2, x^{CO}, x^{IO}\right] \tag{19}$$

should not vary with $\widetilde{\gamma}$. In the example from page 9 of math aptitude as an unobservable individual attribute, the benefits of math aptitude by career income risk cannot depend on risk-aversion. Section A.2 of the Appendix presents these assumptions formally, and shows how we use them to integrate out individual- and career-specific unobservables.

In this case, equations (16) and (17) become:

$$f(\sigma^2 \mid \widetilde{\gamma}, x^{IO}, x^{CO}) \propto f(\sigma^2 \mid \widetilde{\gamma}=0, x^{IO}, x^{CO}) e^{-\frac{1}{2}\widetilde{\gamma}\sigma^2/\beta}; \tag{20}$$

$$f(\sigma^2 \mid \widetilde{\gamma}=0, x^{IO}, x^{CO}) \propto f^C(\sigma^2|x^{CO}) \frac{f^C(x^{CO})}{f\left(x^{CO}\right)} E\left[e^{(y^C+v\left(x^I,x^C\right))/\beta} \mid \sigma^2, x^{IO}, x^{CO}, \widetilde{\gamma}=0\right]. \tag{21}$$

---

[15] We do not observe career-specific pay (y^C) (only total pay which includes individual-specific ability, y^I, and an idiosyncratic productivity shock to the chosen career, $y^{\varepsilon}_{i,c}$) or unobservable career-specific attributes (X^{CU}). We integrate these out to obtain the marginal distribution of observable career choices.

The critical insight from equations (16) or (20) is that the distribution of risk choices for risk-averse people $f(\sigma^2|\widetilde{\gamma})$ is completely determined by the distribution for risk-neutral people $f(\sigma^2|\widetilde{\gamma} = 0)$ and a single parameter β. Each conditional distribution $f(\sigma^2|\widetilde{\gamma})$ for a given $\widetilde{\gamma}$ is merely an exponential 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 similar to the distribution of careers, f^C). For small shocks (low β), the shift is more substantial and conditional distributions for high and low $\widetilde{\gamma}$ become more different (and each becomes more concentrated around the "best" choice for that $\widetilde{\gamma}$). When idiosyncratic shocks are large, the distribution of risk choices by risk-neutral people will reflect primarily the distribution of career options ($f^C(\sigma^2 \mid x^{CO})$); when idiosyncratic shocks are small, the distribution of risk choices by risk-neutral people will reflect primarily which risk values have careers with the highest expected (pecuniary and non-pecuniary) value.

Note that this model is highly over-identified 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 ($f(\sigma^2 \mid \widetilde{\gamma} = 0)$). 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 $\tilde{\gamma}$.

2.4 Idiosyncratic Career Taste vs. Idiosyncratic Career Skill

Equation (20) 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 $y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}}$). Without additional information, we cannot separate the relative importance of individual-specific shocks to skill in specific careers ($y^{\varepsilon}_{i,c}$) from individual-specific shocks to taste for (enjoyment of) those careers $\widetilde{l^{\varepsilon}_{i,c}}$.

However, we can separate these two effects using income data. Observed log pay (ignoring the mean-zero income shock ξ) is:

$$\text{log pay}_{i,c} \equiv y^I_i + y^C_c + y^x(x^I_i, x^C_c) + y^{\varepsilon}_{i,c}. \tag{22}$$

Combining equations (11) and (22) yields:

$$\text{log pay}_{i,c} = V_{i,c} - l^x\widetilde{(x^I_i, x^C_c)} + \frac{1}{2}\widetilde{\gamma}_i\sigma^2_c - \widetilde{l^{\varepsilon}_{i,c}}. \tag{23}$$

We can then take the expectation of log pay conditional on career c having the highest $V_{i,c}$ from equation (14):

$$\begin{aligned} E[\text{log pay}_{i,c} \mid V_{i,c} \geq V_{i,c'}\, \forall c'] &= \overline{V(X^I)} - l^x\widetilde{(x^I_i, x^C_c)} + \frac{1}{2}\widetilde{\gamma}_i\sigma^2_c - E[\widetilde{l^{\varepsilon}_{i,c}} \mid V_{i,c} > V_{i,c'}\, \forall c'] \\ &= \mu + \beta\gamma_{em} + y^I_i + \beta\ln(\sum_q s_q e^{(y^C_q + v\left(x^I_i, x^C_q\right) - \frac{1}{2}\widetilde{\gamma}_i\sigma^2_q)/\beta}) \\ &\quad - l^x\widetilde{(x^I_i, x^C_c)} + \frac{1}{2}\widetilde{\gamma}_i\sigma^2_c - E[\widetilde{l^{\varepsilon}_{i,c}} \mid V_{i,c} > V_{i,c'}\, \forall c']. \end{aligned} \tag{24}$$

Next, we take the expectation of $V_{i,c}$ from equation (11) conditional on career c having the highest $V_{i,c}$:

$$\begin{aligned} E[V(i,c) \mid V_{i,c} \geq V_{i,c'}\, \forall c'] = \overline{V(X^I)} = y^I_i + y^C_c + v\left(x^I_i, x^C_c\right) - \frac{1}{2}\widetilde{\gamma}_i\sigma^2_c \\ + E[y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}} \mid V_{i,c} \geq V_{i,c'}\, \forall c']. \end{aligned} \tag{25}$$

Plugging equation (14) into equation (25) and re-arranging terms yields:

$$\begin{aligned} &E[y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}} \mid V_{i,c} \geq V_{i,c'}\, \forall c'] \\ &\quad = \mu + \beta\gamma_{em} + \beta\ln(\sum_q s_q e^{(y^C_q + v\left(x^I_i, x^C_q\right) - \frac{1}{2}\widetilde{\gamma}_i\sigma^2_q)/\beta}) - y^C_c - v\left(x^I_i, x^C_c\right) + \frac{1}{2}\widetilde{\gamma}_i\sigma^2_c. \end{aligned} \tag{26}$$

By assuming joint normality of $y^{\varepsilon}_{i,c}$ and $\widetilde{l^{\varepsilon}_{i,c}}$, so that the signal extraction problem is linear, $E[y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}} \mid V_{i,c} \geq V_{i,c'} \, \forall c']$ from equation (26) identifies $E[\widetilde{l^{\varepsilon}_{i,c}} \mid V_{i,c} > V_{i,c'} \, \forall c']$ in equation (24):

$$E[\widetilde{l^{\varepsilon}_{i,c}} \mid V_{i,c} \geq V_{i,c'} \, \forall c'] = \frac{var(\widetilde{l^{\varepsilon}_{i,c}})}{var(y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}})}(\mu + \beta\gamma_{em} + \beta \ln(\sum_q s_q e^{(y^C_q + v\left(x^I_i, x^C_q\right) - \frac{1}{2}\widetilde{\gamma}_i \sigma^2_q)/\beta}) \\ - y^C_c - v\left(x^I_i, x^C_c\right) + \frac{1}{2}\widetilde{\gamma}_i \sigma^2_c). \qquad (27)$$

Plugging equation (27) into equation (24) yields:

$$E[\log \text{pay}_{i,c} \mid V_{i,c} \geq V_{i,c'} \, \forall c'] = (\mu + \beta\gamma_{em})(1 - \frac{var(\widetilde{l^{\varepsilon}_{i,c}})}{var(y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}})}) \\ + y^I_i + \beta \ln(\sum_q s_q e^{(y^C_q + v\left(x^I_i, x^C_q\right) - \frac{1}{2}\widetilde{\gamma}_i \sigma^2_q)/\beta})(1 - \frac{var(\widetilde{l^{\varepsilon}_{i,c}})}{var(y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}})}) \\ + y^C_c \frac{var(\widetilde{l^{\varepsilon}_{i,c}})}{var(y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}})} - v\left(x^I_i, x^C_c\right)(1 - \frac{var(\widetilde{l^{\varepsilon}_{i,c}})}{var(y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}})}) \\ + \frac{1}{2}\widetilde{\gamma}_i \sigma^2_c (1 - \frac{var(\widetilde{l^{\varepsilon}_{i,c}})}{var(y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}})}). \qquad (28)$$

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 γ_i) but not chosen career attributes. The third line depends on chosen career attributes (specifically y^C_c and x^{CO}_c but not σ^2_c) and individual observables (x^{IO}_i). The final line depends on both an individual attribute (γ_i) and a career attribute (σ^2).

Equation (28) suggests that a simple regression can be used to recover the relative importance of taste shocks ($\widetilde{l^{\varepsilon}_{i,c}}$) compared with all shocks ($y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}}$). The regression predicts pay with the following controls: a constant; individual-specific controls (including risk aversion); career-attribute controls (particularly a measure of income risk); and, the interaction of individual- and career-specific controls (besides risk aversion and income risk). As shown in the fourth line of equation (28), it also includes the interaction between risk aversion ($\widetilde{\gamma}_i$) and income risk (σ_c^2); the coefficient on this interaction identifies $\frac{1}{2} \times (1 - var(\widetilde{l^{\varepsilon}_{i,c}})/var(y^{\varepsilon}_{i,c} + \widetilde{l^{\varepsilon}_{i,c}}))$.

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 ability or productivity, 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.

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 to 2005. Restricting ourselves to male household heads aged 22 through 60[16] 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 through 1996, with risk tolerance responses recorded in the 1996 wave, and non-zero population weights[17]. There are 1,490

---

[16] 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).

[17] 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.

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

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.

individuals meeting 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 which 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 to both questions they were asked or no to both 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 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, since 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 non-parametrically 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 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.

While some other papers have dropped observations with missing and zero income (Gottschalk and Moffitt, 2002) or modeled unemployment explicitly (Pistaferri, 2002), neither route is available to us because the method in Jensen and Shore is not designed to handle missing data or zeros. Instead, Jensen and Shore fill in hot-deck imputed missing values when calculating volatility. Aside from using their volatility values, we do not explicitly use bootstrapped income data. We follow Jensen and Shore in using top- and bottom-codes. The income dynamics used by Jensen and Shore to estimate income volatility are quite standard, characterizing the evolution of excess log income for individual i over time t (Carroll and Samwick, 1997; Meghir and Pistaferri, 2004). Excess log income $y_{i,t}$ is modeled as the sum of permanent income, transitory income, and error $e_{i,t}$.

$$y_{i,t} = \underbrace{\sum_{k=1}^{t-3} \omega_{i,k} + \sum_{k=t-2}^{t} \phi_{\omega,t-k} \cdot \omega_{i,k}}_{\text{Permanent income}} + \underbrace{\sum_{k=t-2}^{t} \phi_{\varepsilon,t-k} \cdot \varepsilon_{i,k}}_{\text{Transitory income}} + e_{i,t}. \tag{29}$$

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.[18]

The permanent shock, transitory shock, and error term are assumed to be normally distributed 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 $\sigma^2_{\varepsilon,i,t} \equiv \mathrm{E}[\varepsilon^2_{i,t}]$:

$$\begin{pmatrix} \omega_{it} \\ \varepsilon_{i,t} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\omega,i,t} & 0 \\ 0 & \sigma^2_{\varepsilon,i,t} \end{bmatrix} \right). \tag{30}$$

We refer to $\sigma^2_{i,t} \equiv (\sigma^2_{\varepsilon,i,t}, \sigma^2_{\omega,i,t})$ jointly as the volatility parameters. Finally, we have "noise variance"which refers to the variance of measurement error $\gamma^2 \equiv \mathrm{E}[e^2_{i,t}]$ that is constant across individuals and over time.

Jensen and Shore (2009a,b) develop a Markovian hierarchical Dirichlet Process (MHDP) prior that they use to estimate the distribution of ex-ante expected volatil-

---

[18] In this framework model, permanent shocks come into effect over three periods and transitory shocks fade completely after three periods, giving us three permanent weight parameters $(\phi_{\omega,0}, \phi_{\omega,1}, \phi_{\omega,2})$ and three transitory weight parameters $(\phi_{\varepsilon,0}, \phi_{\varepsilon,1}, \phi_{\varepsilon,2})$. We refer to these weights $\boldsymbol{\phi}$ collectively as the income process parameters, which will need to be estimated in our model. Jensen and Shore posit flat prior distributions for each weight parameter (i.e. $p(\phi) \propto 1$). However, in order to give meaning to the magnitude of our transitory shocks, we normalize the weights placed on transitory shocks to sum to one ($\sum_k \phi_{\varepsilon,k} = 1$).

ity. We use the estimates of the ex-ante expected permanent volatility distribution from their paper, and take the average of these estimates over the years 1991-1996 as our final measure of income volatility.

The Jensen and Shore (2009a,b) method for estimating income volatility is one of a few possible estimation methods. For example, we could use the moment proposed by Meghir and Pistaferri (2004),

$$\sigma^2_{\omega,i,t} = E[(y_{i,t} - y_{i,t-1}) \times (y_{i,t+m} - y_{i,t-1-n})], \tag{31}$$

where we choose $m = 2$ and $n = 2$ respectively. We could also use the Jensen and Shore (2009a,b) method with income volatility estimated separately by risk tolerance bin.

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 based on two different income volatility estimation methods: the original Jensen and Shore (2009a,b) method and the Meghir and Pistaferri (2004) method.[19] It is clear from Table 2 that there is sorting of the more risk tolerant into riskier careers, regardless of the income volatility estimation method. Because this sorting is present regardless of the method employed, we use only the Jensen and Shore (2009a,b) estimates in Section 4.

---

[19] The Jensen and Shore (2009a,b) method estimated separately by risk tolerance bin yields similar results, and is therefore omitted from Table 2.

Table 2: Volatility Distributions by Specifications

Volatility Distributions by Risk Aversion and Specification

	MHDP				Meghir-Pistaferri			
%-tile	$[0.2)$	$[2, 3.84)$	$[3.84, 7.52)$	$[7.52, \infty)$	$[0.2)$	$[2, 3.84)$	$[3.84, 7.52)$	$[7.52, \infty)$
min	0.0257	0.0260	0.0264	0.0261	-1.0000	-1.0000	-0.9593	-1.0000
1st	0.0268	0.0265	0.0276	0.0268	-0.3929	-0.5386	-0.5447	-0.7107
5th	0.0290	0.0287	0.0296	0.0293	-0.1447	-0.3016	-0.2560	-0.2299
10th	0.0299	0.0300	0.0305	0.0300	-0.0763	-0.0531	-0.0626	-0.0441
25th	0.0311	0.0313	0.0313	0.0313	-0.0048	-0.0001	-0.0061	-0.0032
50th	0.0322	0.0320	0.0323	0.0322	0.0155	0.0124	0.0095	0.0092
75th	0.0346	0.0331	0.0332	0.0330	0.1196	0.0568	0.0324	0.0448
90th	0.0799	0.0377	0.0472	0.0390	0.4374	0.1839	0.1947	0.1979
95th	0.3289	0.0828	0.1446	0.0795	0.7874	0.3122	0.3987	0.4196
99th	1.0000	0.2647	1.0000	0.7426	1.0000	0.8469	1.0000	1.0000
max	1.0000	0.6804	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
mean	0.0713	0.0413	0.0628	0.0541	0.0957	0.0328	0.0365	0.0370
st. dev	0.1534	0.0532	0.1439	0.1145	0.2788	0.2017	0.2210	0.2273

Table 2 displays income volatility distribution percentiles by risk aversion bin, based on two different income volatility estimation methods: the Jensen and Shore (2009a,b) methodology and the Meghir and Pistaferri (2004) method described above. Jensen-Shore volatility values are averaged over the years 1991-1996, while Meghir-Pistaferri volatility values are averaged over 1991-1995. All (averaged) volatility values are top coded at 1, while Meghir-Pistaferri values are bottom coded at -1 as well. Each observation is weighted by its PSID supplied sample weights.

Table 3 presents the joint distribution of volatility and risk aversion. In that table, σ^2 values are divided into 10 bins, corresponding to the 1^{st}, 5^{th}, 10^{th}, 30^{th}, 50^{th}, 70^{th}, 90^{th}, 95^{th}, and 99^{th} percentiles of the σ^2 distribution. Again it is clear that individuals with high risk aversion are less likely to have the highest volatility values. The distribution of σ^2 values is also shown in the left panel of Figure 2. The right panel shows the proportion of the data in each of the 10 σ^2 bins.

Table 3: Estimated distribution of income volatility by self-reported risk-aversion

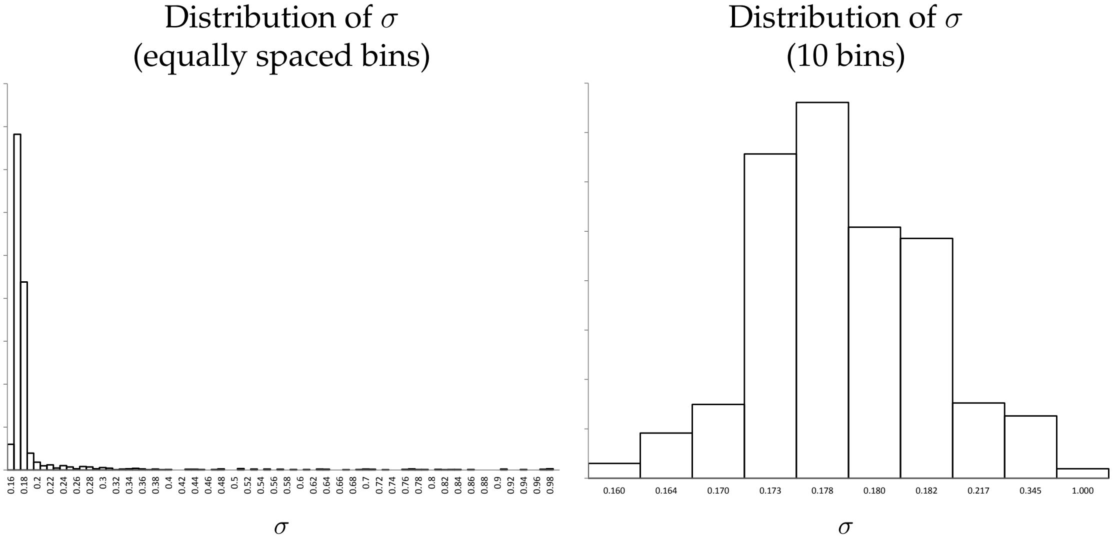
$\widehat{\sigma}^2$ from Jensen and Shore

$\widehat{\sigma}^2$ 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 %
# of Observations				1,490	320	241	267	662
$\%$ of Observations				100%	21.50%	16.17%	17.90%	44.43%

Table 3 shows the distribution of σ^2 estimates. σ^2 estimates are the average of 1991 to 1996 estimates of permanent volatility. Volatility estimates are from Jensen and Shore and are top-coded at 1. $\widetilde{\gamma}$ 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.

Regressions presented in Table 4 offer additional reduced-form evidence for the negative relationship between income risk and risk aversion. The coefficient on estimated risk-aversion is negative and significant for both σ^2 and the log of σ^2 when occupation, family, and demographic covariates are included. The coefficient on the log of income risk is significant regardless of additional controls. Note that while a negative relationship between income risk and risk aversion exists, it is not particularly strong — a one-unit higher level of risk aversion is associated with an $\approx 11\%$ lower level of income risk. This statistically significant but economically modest sorting of the more risk tolerant individuals into riskier careers is precisely the feature of the data that identifies our model.

Figure 2: Jensen-Shore Permanent Income Volatility Bins



Distribution of σ (equally spaced bins)

Distribution of σ (10 bins)

The left panel presents the distribution of 1991-1996 σ estimates from Jensen and Shore, a histogram of the standard deviation of permanent income changes. The right panel shows the distribution of income volatility by the 10 coarse bins used in the non-parametric estimation of β and $f(\sigma|\widetilde{\gamma}=0)$.

Table 4: Relationship between Income Volatility ('91-'96) and Risk Aversion ('96)

	Jensen-Shore					
Dep. Var.	Income Risk Level			Income Risk Log		
$E[\widetilde{\gamma} \mid \text{bin}]$	-0.002	-0.003	-0.015*	-0.018**	-0.020**	-0.116**
	(0.002)	(0.002)	(0.008)	(0.009)	(0.009)	(0.049)
age		0.001**	0.001***		0.005**	0.006**
		0.000	(0.001)		(0.002)	(0.003)
controls	no	no	yes	no	no	yes
R^2	0.001	0.006	0.036	0.003	0.007	0.044
# of Obs.	1,490	1,490	1,490	1,490	1,490	1,490

Table 4 shows the OLS regressions to predict individual-specific measures of income risk with self-reported risk-aversion bin. $E[\widetilde{\gamma} \mid \text{bin}]$ refers to the expected value of risk aversion conditional on risk aversion bin, which we estimate using the signal-noise structure identified in Kimball, Sahm, and Shapiro. The variable age refers to the individual's age in years, and controls include occupation, family, and demographic characteristics. $*$ Indicates significance at the 10% level, $**$ indicates significance at the 5% level, and $***$ indicates significance at the 1% level.

Table 5: Distribution of income risk and risk-aversion by broad occupational category

		Jensen- Shore σ^2		$\widetilde{\gamma}$ bin			
occupation	obs	mean	st.dev	least ← risk averse → most			
prof./tech.	339	0.044	0.079	20 %	21 %	18 %	40 %
managers	343	0.069	0.158	27 %	13 %	15 %	45 %
clerical	221	0.047	0.094	21 %	17 %	14 %	48 %
craftsmen	284	0.059	0.133	16 %	15 %	23 %	46 %
operators	157	0.050	0.089	17 %	13 %	18 %	52 %
laborers	52	0.043	0.057	15 %	19 %	17 %	48 %
farmers	36	0.141	0.240	17 %	17 %	8 %	58 %
n/a	58	0.063	0.102	31 %	21 %	24 %	24 %
overall	1,490	0.057	0.122	21 %	16 %	18 %	45 %

Table 5 shows the distribution of self-reported risk preference by one-digit occupational categories. Risk preferences are recorded in 1996 and occupation categories are recorded in 1991. The n/a category includes non-responses.

Table 5 shows (1991-1996) income volatility and (1996) risk-aversion data by (1991) “one-digit” occupational category. Note that, while income volatility varies across occupations, the correlation between occupational income volatility and occupational risk tolerance is quite low.

4 Estimation

If we could observe the joint distribution of data $\left\{\sigma^2, \widetilde{\gamma}, x^{IO}, x^{CO}\right\}$, then estimation of equations (20) and (21) by maximum likelihood is straightforward. 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 by maximum likelihood. Table 3 shows the non-parametric 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 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 $\widetilde{\gamma}$:

$$\widehat{\ln(1/\widetilde{\gamma})} = \ln(1/\widetilde{\gamma}) + e \tag{32}$$

$$\begin{bmatrix} \ln(1/\widetilde{\gamma}) \\ e \end{bmatrix} \sim N\left(\begin{bmatrix} -1.05 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.76 & 0 \\ 0 & 1.69 \end{bmatrix}\right) \tag{33}$$

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 3 shows these ranges and the fraction of observed data that falls into each.[20]

We can then identify the relationship between our data ($f(\sigma^2 \mid \widehat{\ln(1/\widetilde{\gamma})}\text{ bin})$) and the object we wish to estimate ($f(\sigma^2 \mid \widetilde{\gamma})$ from equation (21)):

$$f\left(\sigma^2 \mid \widehat{\ln(1/\widetilde{\gamma})}\text{ bin}\right) = \int\limits_{\ln(1/\widetilde{\gamma})} f\left(\sigma^2 \mid \widetilde{\gamma}\right) f_{\ln(1/\widetilde{\gamma})}\left(\ln\left(1/\widetilde{\gamma}\right) \mid \widehat{\ln(1/\widetilde{\gamma})}\text{ bin}\right) d\ln\left(1/\widetilde{\gamma}\right) \tag{34}$$

$$f_{\ln(1/\widetilde{\gamma})}\left(\ln\left(1/\widetilde{\gamma}\right) \mid \widehat{\ln(1/\widetilde{\gamma})}\text{ bin}\right) = f_{\ln(1/\widetilde{\gamma})}\left(\ln\left(1/\widetilde{\gamma}\right)\right) \frac{\text{pr}\left(\widehat{\ln(1/\widetilde{\gamma})}\text{ bin} \mid \widetilde{\gamma}\right)}{\text{pr}\left(\widehat{\ln(1/\widetilde{\gamma})}\text{ bin}\right)} \tag{35}$$

Given the distribution of true variation and classical measurement error estimated

---

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

by Kimball, Sahm, and Shapiro, it is trivial to calculate $f_{\ln(1/\widetilde{\gamma})}\left(\ln\left(1/\widetilde{\gamma}\right) \mid \widehat{\ln(1/\widetilde{\gamma})}\text{ bin}\right)$ and pr $\left(\widehat{\ln(1/\widetilde{\gamma})}\text{ bin} \mid \widetilde{\gamma}\right)$ for each $\widetilde{\gamma}$ in our grid for each of the four risk-aversion bins; pr $\left(\widehat{\ln(1/\widetilde{\gamma})}\text{ bin}\right)$ is similarly easy to calculate for each of the four risk-aversion bins.

Armed with this distribution of $\widetilde{\gamma}$, we search for maximum likelihood estimates of $f(\sigma^2 \mid \widetilde{\gamma} = 0)$ and β iteratively. First, we guess values of $f(\sigma^2 \mid \widetilde{\gamma} = 0)$ and β. Next, we calculate $f(\sigma^2 \mid \widetilde{\gamma})$ for each value of σ^2 and $\widetilde{\gamma}$ on our grid. Next, we calculate $f(\sigma^2 \mid \widetilde{\gamma})$ for each of the 10 grid values of σ^2 and each of the four coarse bins for $\widetilde{\gamma}$ by integrating over each value of $\widetilde{\gamma}$ possible in each bin. This gives the likelihood of an observation lying in one of the $10 \times 4 = 40$ possible ranges we observe in Table 3. We then compute the likelihood of observing the data in Table 3. We search over $f(\sigma^2 \mid \widetilde{\gamma} = 0)$ and β to find values which maximize the likelihood.

5 Results

Equations (20) and (21) 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 $var(\varepsilon_{i,c})$) 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 (28), designed to separate the relative importance of idiosyncratic taste from idiosyncratic skill. In Section 5.4, we reinterpret our results in the context of search frictions.

5.1 Parameter Estimates

Table 6: Parameter Estimates

Parameter	Average σ^2			Bootstrap σ^2		
$\widehat{\beta}$	0.649	0.553	0.588	3.075	2.617	2.876
β <	2.749	2.027	2.302	18.107	25.931	16.898
β >	0.279	0.246	0.258	-12.384	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 %
↑	19.5 %	22.5 %	22.9 %	19.8 %	21.9 %	22.6 %
f($\sigma^2 \vert \gamma = 0$)	22.8 %	26.1 %	26.5 %	19.1 %	20.2 %	20.3 %
	15.4 %	17.5 %	17.7 %	20.1 %	20.4 %	20.4 %
↓	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 %	5.2 %	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

Table 6 displays the estimates of β (including a 90% confidence interval), $f(\sigma^2|\widetilde{\gamma} = 0)$, and θ from equations (20) and (21). The point estimates of β correspond to the variance of idiosyncratic taste and skill shocks. $f(\sigma^2|\widetilde{\gamma} = 0)$ 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.

Table 6 shows the coefficient estimates from equations (20) and (21) using two different estimation methods. The first method assumes the Jensen-Shore volatility values are estimated with certainty. Results from this approach are reported in the left three columns of Table 6. The β value estimated without additional controls is 0.649, so 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.[21] 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 one 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.[22]

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 3. This gives us 100 sets of β 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, and confidence intervals are determined by the 5^{th} and 95^{th} percentiles of the 100-sample parameter distributions. The results from this method are reported in the right three columns of Table 6.

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 $\widehat{\beta}$ towards infinity. This is most easily seen through the difference in the estimated upper bounds of $\widehat{\beta}$.

---

[21] 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 beta values that allow us to reject that the restricted model is correct using the two-sided likelihood ratio test.

[22] Results using the Meghir Pistaferri 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.

However, note that once covariates are included, the lower bound of $\widehat{\beta}$ is broadly similar the estimated value of $\widehat{\beta}$ under the first approach.

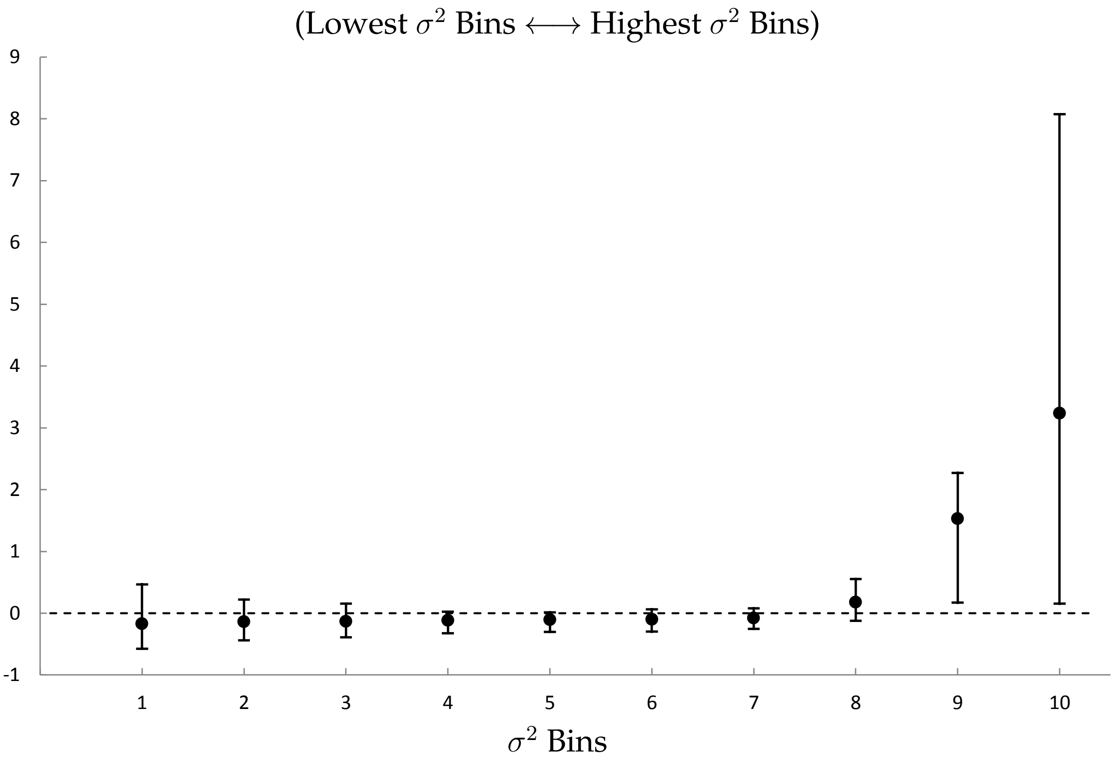
Note that if risk-averse individuals make their own income streams less risky, estimates of β will be biased downward; we would observe a *stronger* correlation between risk aversion and income risk in the data. This implies that our estimate of the lower bound on β is smaller than the true lower bound.

Along with estimates of β, Table 6 shows the estimates of $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$ and θ. Figure 3 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 3 shows the degree to which risk-neutral individuals are estimated to over-weight or under-weight this bin relative the population as a whole. For each σ^2 bin we obtain a 95% confidence interval by finding the highest and lowest values of $f(\sigma^2 \mid \widetilde{\gamma} = 0, x^{IO}, x^{CO})$ 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 (21) 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 ($E\left[e^{\left(y^C+v\left(x^I,x^C\right)\right)^{\frac{1}{\beta}}} \mid \sigma^2\right]$). 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 ($y^C + v\left(x^I, x^C\right)$), the income premium needed to fill all careers at each level

Figure 3: Over/Under Representation 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 value in the right panel of Figure 2 and subtracting one. This shows the degree to which risk-neutral individuals are estimated to over-weight or under-weight 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).

of volatility. Assuming no heterogeneity conditional on σ^2, from equation (21) we have $e^{y^C+v\left(x^I,x^C\right)} = (\frac{f(\sigma^2|\widetilde{\gamma}=0)}{f^C})^{\beta}$. 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 4. 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 are shown in Figure 3.

Figure 4: Average Risk Premia by Income Risk Bin

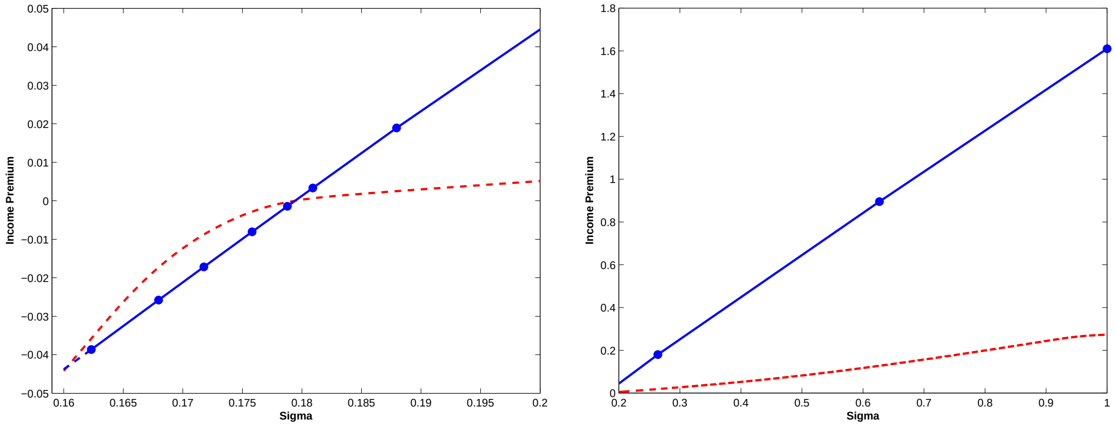


Figure 4 shows the estimated income premium y^C at the midpoint of each σ^2 bin. The two panels display the full range of σ values on different vertical axis scales. The dashed curve reflects the perfect-sorting case from equation (13). The solid-curve reflects the required risk-premium needed to rationalize 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 (20) is estimated without covariates.

5.3 Idiosyncratic Taste or Skill?

Next, we use income data to decompose $\varepsilon_{i,c}$ into idiosyncratic career skill ($y^{\varepsilon}_{i,c}$) and taste ($\widetilde{l^{\varepsilon}_{i,c}}$). Estimation by OLS of equation (28) identifies $(1-var(\widetilde{l^{\varepsilon}_{i,c}})/var(y^{\varepsilon}_{i,c}+\widetilde{l^{\varepsilon}_{i,c}}))$, the coefficient on $\frac{1}{2} \times \widetilde{\gamma}_i \times \sigma^2_c$. 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 7 shows the results from regressing pay on $\widetilde{\gamma}_i$, σ^2_c, their interaction, and covariates.[23] When non-linear functions of σ^2_c are included as controls, the coefficient estimate on $\frac{1}{2}\times\widetilde{\gamma}_i\times\sigma^2_c$ has standard errors small enough that we may rule out a coefficient of zero. Results with this specification indicate that idiosyncratic skill and the increased compensation that follows has at least some role in driving career choice. Finally, note the similarity between this regression and the risk-augmented Mincer equations from Hartog (2009), which provides a consistency check on our particular sample.

5.4 Reinterpreting Our Results in the Context of Search

Recall that in Section 2.2 we presented an illustration without idiosyncratic taste and skill which implied perfect sorting of the most risk-averse people into the safest careers. In Section 2.3 we interpreted deviations from this perfect sorting as an indication of the presence of idiosyncratic taste for or skill in various careers. An alternative interpretation is that deviations from perfect sorting reflect the presence

---

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

of search frictions (see Lucas and Prescott (1974), Mortensen and Pissarides (1994), Pissarides (2000), Rogerson, Shimer, and Wright (2005) and Galenianos, Kircher, and Virag (forthcoming) as just a few examples with search frictions). Even if individuals do not have idiosyncratic taste for or skill in various careers we may still see deviations from perfect sorting if individuals do not have access to the full range of income risk options.

In the data, the average value of $\frac{1}{2} \times \tilde{\gamma}_i \times \sigma_c^2$ is 0.1399, so that on average individuals would be willing to give up $\approx 14\%$ of their income (in log terms) to eliminate income risk. Under the counterfactual of perfect sorting, the average value is 0.0668 (6.68% of income in log terms). The difference, 0.0731 (7.31% of income in log terms), can be viewed as the potential welfare gain associated with eliminating the mismatch of risk-averse people into risky careers. In the context of a model with search, 7.31% is the welfare gain associated with eliminating search frictions.

One desirable feature of our model is its robustness to randomly incomplete menus. When $\varepsilon_{i,c}$ has an extreme value distribution (as discussed in Section 2.3), estimates of β will be unaffected by individuals observing only a random subset of the full risk-return menu, so long as for every risk value available under the full menu there is a career in the restricted menu with a risk value to which it is arbitrarily close. When $\varepsilon_{i,c}$ are not extreme value distributed, we also require that the number of options in this restricted menu goes to infinity. Fewer draws from the risk-return menu affect only the expectation of the optimal career, not the degree of risk-based mismatch. For search frictions to affect estimates of β, such frictions must make entire income risk ranges unavailable to some individuals.

Table 7: 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
	(0.142)	(0.129)	(0.376)	(0.341)	(1.482)
$(\sigma^2)^2$					1.373
					(1.047)
$\ln(\sigma^2)$					-0.234*
					(0.128)
$\widetilde{\gamma} = 2^{\text{nd}}$ lowest			-0.002	-0.005	-0.017
			(0.058)	(0.053)	(0.052)
$\widetilde{\gamma} = 2^{\text{nd}}$ highest			0.105*	0.150***	0.133***
			(0.057)	(0.052)	(0.052)
$\widetilde{\gamma} =$ highest			0.003	0.045	0.023
			(0.051)	(0.046)	(0.046)
$\frac{1}{2} \times \sigma^2 \times \widetilde{\gamma}$			0.142	0.203	0.255*
			(0.147)	(0.133)	(0.132)
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
$\#ofObs.$	1,484	1,484	1,484	1,484	1,484

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. While the full sample includes 1,490 observations, 6 of these have an income of zero throughout, and consequently a missing log income. 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 to 1996. Standard errors in parenthesis: *significant at 10% level ; **significant at 5% level ; ***significant at 1% level.

6 Conclusion

This paper has documented that those who self-identify as risk-averse are more likely to have volatile incomes, but that this correlation, while negative, is far from -1. 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 an income gap between high- and low-risk careers for more risk-averse people – relative to more risk-tolerant ones – indicates that some of this variation is idiosyncratic skill in one career over another, not just idiosyncratic taste for one career over another.

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 which jointly maximizes their enjoyment of and skill in that career. Training individuals in careers they would not otherwise choose will have little, if any, benefit, as any randomly chosen career is likely to be far worse than the best one.

A 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 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:

$$F_r(x \mid \mu_r, \beta_r) = (F(x \mid \mu, \beta))^{s_r} = (e^{-e^{-(x-\mu)/\beta}})^{s_r} \tag{36}$$

$$= e^{-e^{-(x-\mu-\beta\ln(s_r))/\beta}} = F(x \mid \mu + \beta\ln(s_r), \beta). \tag{37}$$

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 on page 15), this informs the following transformation:

$$W(i,r) \equiv \max_{c\in r} V(i,c) = \max_{c\in r}\,[y_i^I + y_c^C + v\left(x_i^I, x_c^C\right) - \frac{1}{2}\widetilde{\gamma}_i\sigma_c^2 + \varepsilon_{i,c}]; \tag{38}$$

$$= y_i^I + y_c^C + v\left(x_i^I, x_c^C\right) - \frac{1}{2}\widetilde{\gamma}_i\sigma_c^2 + \max_{c\in r}\varepsilon_{i,c}; \tag{39}$$

$$W(i,r) = y_i^I + y_c^C + v\left(x_i^I, x_c^C\right) - \frac{1}{2}\widetilde{\gamma}_i\sigma_r^2 + EV(\mu + \beta\ln(s_r), \beta); \tag{40}$$

$$W(i,r) = \mu + \beta\gamma_{em} + y_i^I + \beta\ln(s_r e^{(y_c^C + v\left(x_i^I, x_c^C\right) - \frac{1}{2}\widetilde{\gamma}_i\sigma_r^2)/\beta}) + EV\left(-\beta\gamma_{em}, \beta\right),^{24} \tag{41}$$

where equation (41) follows from pulling $\beta\ln(s_r)$ out of the expectation, adding/subtracting $\beta\gamma_{em}$, and combining terms.

With equation (41) in hand, we can compute analytically the probability that a given range s_r will produce the maximum value. In particular, define:

$$Z_r \equiv a + \beta\ln\left(s_r\right) + EV\left(-\beta\gamma_{em}, \beta\right) \tag{42}$$

$$s_r = prob\left(Z_r > Z_s, \forall s \neq r\right); \sum_r s_r = 1 \tag{43}$$

Combining equations (41) and (43) gives the probability that an individuals

---

[24] Note that the extreme value distribution (EV) in equation (41) 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, $\lim\limits_{N_C\to\infty} \mu + y_r^C$ 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.

preferred career will lie in range r:

$$prob(W(i,r) > W(i,q), \forall q \neq r) = \frac{s_r e^{(y_r^C + v\left(x_i^I, x_r^C\right) - \frac{1}{2}\widetilde{\gamma}_i \sigma_r^2)/\beta}}{\sum_q s_q e^{(y_q^C + v\left(x_i^I, x_q^C\right) - \frac{1}{2}\widetilde{\gamma}_i \sigma_q^2)/\beta}}. \tag{44}$$

The probability that a given range will have the highest value (equation 44) is nothing more than the pdf, the joint distribution of attributes X^C of careers chosen given i:

$$f(X^C \mid i) \equiv prob(W(i,r) > W(i,q), \forall q \neq r)$$

We re-write equation (44) by taking the size of each range to zero, so that the sums become integrals and s_r becomes $f^C(X^C)$:

$$f(X^C \mid X^I) = \frac{f(X^C | \widetilde{\gamma} = 0, X^I) e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}}{\iiiint_{X_q^C} f(X_q^C | \widetilde{\gamma} = 0, X^I) e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma_q^2} dX_q^C}, \tag{45}$$

$$f(X^C \mid \widetilde{\gamma} = 0, X^I) = \frac{f^C(X^C) e^{\left(\left(y^C + v\left(x^I, x^C\right)\right)/\beta\right)}}{\iiiint_{X_q^C} f^C(X_q^C) e^{((y^C + v(x^I, x^C))/\beta)} dX_q^C} \tag{46}$$

The result is equations (16) and (17) on page 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 (16) and (17) as:

$$f(X^C \mid X^I) = \frac{f(X^C | \widetilde{\gamma} = 0, X^I) e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}}{\iiiint_{X_q^C} f(X_q^C | \widetilde{\gamma} = 0, X^I) e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma_q^2} dX_q^C}, \tag{47}$$

$$f(X^C \mid \widetilde{\gamma} = 0, X^I) = \frac{f^C(X^C) e^{\left(\left(y^C + v\left(x^I, x^C\right)\right)/\beta\right)}}{\iiiint_{X_q^C} f^C(X_q^C) e^{((y^C + v(x^I, x^C))/\beta)} dX_q^C}, \tag{48}$$

which taken together imply that:

$$f(X^C \mid X^I) = k_1\left(X^I\right) f^C(X^C) e^{\left(\left(y^C + v\left(x^I, x^C\right)\right)/\beta\right)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}; \tag{49}$$

$$k_1\left(X^I\right) \equiv \frac{1}{\iiiint_{X_q^C} f^C(X^C) e^{((y^C + v(x^I, x^C))/\beta)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} dX_q^C}. \tag{50}$$

We then integrate equation (49) over y^C and x^{CU} to obtain the marginal distribution:

$$f(\sigma^2, x^{CO} \mid X^I) = \iint k_1\left(X^I\right) f^C(X^C) e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} dy^C dx^{CU} \tag{51}$$

$$= k_1\left(X^I\right) f^C(\sigma^2, x^{CO}) \iint f^C(y^C, x^{CU}|\sigma^2, x^{CO}) e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} dy^C dx^{CU}. \tag{52}$$

Equation (52) results from pulling k_1 and $f^C(\sigma^2, x^{CO})$ out of the integral, since they do not depend on y^C or x^{CU}. We can then write the double integral in equation (52) as an expectation over y^C and x^{CU}:

$$E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} |\sigma^2, x^{CO}, X^I\right], \tag{53}$$

in which case equation (52) becomes:

$$f(\sigma^2, x^{CO} \mid X^I) \propto f^C(\sigma^2, x^{CO}) E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} |\sigma^2, x^{CO}, X^I\right]. \tag{54}$$

Conditioning on x^{CO} using Bayes' rule, equation (54) becomes:

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

$$\Rightarrow f(\sigma^2 \mid x^{CO}, X^I) \propto f^C(\sigma^2|x^{CO}) \frac{f^C(x^{CO})}{f\left(x^{CO}\right)} E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} |\sigma^2, x^{CO}, X^I\right]. \tag{55}$$

So far, we have transformed equations (47) and (48) into equation 55 without additional assumptions. Next, we want to write equation (55) 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 following assumption:

$$\begin{aligned} &E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} |\sigma^2, x^{CO}, X^I\right] \\ =&E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} |\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\right] e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}. \end{aligned} \tag{56}$$

If σ^2 and $e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)}$ 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 (56) into equation (55) yields the distribution of risk choices for risk-neutral individuals (imposing $\widetilde{\gamma} = 0$) and for risk-averse individuals relative to risk-neutral individuals:

$$\begin{aligned} f(\sigma^2 \mid x^{CO}, X^I, \widetilde{\gamma}=0) &\propto f^C(\sigma^2|x^{CO}) \frac{f^C(x^{CO})}{f\left(x^{CO}\right)} \\ &\quad \times E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)}|\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\right], \end{aligned} \tag{57}$$

$$f(\sigma^2 \mid x^{CO}, X^I, \widetilde{\gamma}) \propto f(\sigma^2 \mid x^{CO}, X^I, \widetilde{\gamma}=0) e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} \tag{58}$$

To this point, we have integrated out career-specific unobservables, transforming equations (16) and (17) into equations (57) and (58). Next, we integrate out unobservable individual attributes. To do so, we separate X^I into its constituent parts, re-writing equation (58) as:

$$\begin{aligned} &f(\sigma^2 \mid x^{CO}, y^I, \widetilde{\gamma}, x^{IO}, x^{IU}) \propto \\ &f^C(\sigma^2|x^{CO}) \frac{f^C(x^{CO})}{f\left(x^{CO}\right)} E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)}|\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\right] e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}. \end{aligned} \tag{59}$$

Using Bayes' rule, we transform equation (59) into the joint distribution of σ^2 and x^{IU} (dropping y^I since it affects all careers equally):

$$\begin{aligned} f(\sigma^2, x^{IU} \mid x^{CO}, \widetilde{\gamma}, x^{IO}) &\propto f^C(\sigma^2|x^{CO}) \frac{f^C(x^{CO})}{f\left(x^{CO}\right)} f(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}) \\ &\quad \times E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)}|\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\right] e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}, \end{aligned} \tag{60}$$

or equivalently:

$$\begin{aligned} f(\sigma^2, x^{IU} \mid x^{CO}, \widetilde{\gamma}, x^{IO}) &= k_2\left(x^{CO}, \widetilde{\gamma}, x^{IO}\right) f^C(\sigma^2|x^{CO}) \frac{f^C(x^{CO})}{f\left(x^{CO}\right)} f(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}) \\ &\quad \times E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)}|\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\right] e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}; \end{aligned} \tag{61}$$

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

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

$$\begin{aligned} &f(\sigma^2 \mid x^{CO}, \widetilde{\gamma}, x^{IO}) = \int k_2\left(x^{CO}, \widetilde{\gamma}, x^{IO}\right) f^C(\sigma^2|x^{CO}) \frac{f^C(x^{CO})}{f\left(x^{CO}\right)} \\ &\times f(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}) E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)}|\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\right] e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} dx^{IU} \end{aligned} \tag{62}$$

$$\begin{aligned} &\Rightarrow f(\sigma^2 \mid x^{CO}, \widetilde{\gamma}, x^{IO}) = k_2\left(x^{CO}, \widetilde{\gamma}, x^{IO}\right) f^C(\sigma^2|x^{CO}) \frac{f^C(x^{CO})}{f\left(x^{CO}\right)} e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2} \\ &\times \int f(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}) E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)}|\sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\right] dx^{IU} \end{aligned} \tag{63}$$

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

$$\begin{aligned}\int f(x^{IU} \mid \widetilde{\gamma}, x^{IO}, x^{CO}) E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} | \sigma^2, x^{CO}, X^I, \widetilde{\gamma}=0\right] dx^{IU} \\ = E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} | \sigma^2, x^{CO}, x^{IO}, \widetilde{\gamma}=0\right]. \qquad (64)\end{aligned}$$

Assumption (64) 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 $f(\sigma^2 \mid x^{CO}, \widetilde{\gamma}, x^{IO})$:

$$\begin{aligned}f(\sigma^2 \mid x^{CO}, \widetilde{\gamma}=0, x^{IO}) \propto\ & f^C(\sigma^2 | x^{CO}) \frac{f^C(x^{CO})}{f\left(x^{CO}\right)} \\ & \times E\left[e^{\left(\left(y^C+v\left(x^I,x^C\right)\right)/\beta\right)} | \sigma^2, x^{CO}, x^{IO}, \widetilde{\gamma}=0\right], \qquad (65) \\ f(\sigma^2 \mid x^{CO}, \widetilde{\gamma}, x^{IO}) \propto\ & f(\sigma^2 \mid x^{CO}, \widetilde{\gamma}=0, x^{IO}) e^{-\frac{1}{2}\frac{1}{\beta}\widetilde{\gamma}\sigma^2}. \qquad (66)\end{aligned}$$

Equations (65) and (66) are identical to equations (20) and (21) on page 17.

References

ABOWD, J., AND D. CARD (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–445.

ALVAREZ, J., M. BROWNING, AND M. EJRNAES (2001): "Modelling income processes with lots of heterogeneity," University of Copenhagen Working Paper.

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.

BEN-AKIVA, M., N. LITINAS, AND K. TSUNOKAWA (1985): "Continuous Spatial Choice: The Continuous Logit Model and Distribution of Trips and Urban Densities," *Transportation Research Part A: General*, 19A(2), 119–154.

BERKHOUT, P., J. HARTOG, AND D. WEBBINK (2006): "Compensation for Earnings Risk under Worker Heterogeneity," IZA Discussion Paper 2074.

BOSKIN, M. (2004): "A Conditional Logit Model of Occupational Choice," *Journal of Political Economy*, 82(2), 389–398.

CARROLL, C. D., AND A. A. SAMWICK (1997): "The Nature of Precautionary Wealth," *Journal of Monetary Economics*, 40(1), 41–71.

COUCH, K. A., AND D. W. PLACZEK (2010): "Earnings Losses of Displaced Workers Revisited," *American Economic Review*, 100(1), 572–589.

DE HAAN, L., AND A. FERREIRA (2006): *Extreme Value Theory: An Introduction*. Springer.

DELEIRE, T., AND H. LEVY (2004): "Worker Sorting and the Risk of Death on the Job," *Journal of Labor Economics*, 22(4), 925–953.

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.

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.

FLYER, F. A. (2004): "The Influence of Higher Moments of Earnings Distributions on Career Decisions," *Journal of Labor Economics*, 15(4), 689–713.

FROST, R. (1920): *The Road Not Taken*Henry Holt and Company.

GALENIANOS, M., P. KIRCHER, AND G. VIRAG (forthcoming): "Market Power and Efficiency in a Search Model," *International Economic Review*.

GOTTSCHALK, P., AND R. MOFFITT (2002): "Trends in the Transitory Variance of Earnings in the United States," *Economic Journal*, 112, C68–73.

HARRIS, M., AND Y. WEISS (1984): "Job Matching with Finite Horizon and Risk Aversion," *Journal of Political Economy*, 92(4), 758–779.

HARTOG, J. (2009): "A Risk Augmented Mincer Earnings Equation? Taking Stock," IZA Discussion Paper 4439.

HARTOG, J., E. PLUG, L. DIAZ-SERRANO, AND J. VIEIRA (2003): "Risk Compensation in Wages, a Replication," *Empirical Economics*, 28(3), 639–647.

HARTOG, J., AND W. VIJVERBERG (2007): "On Compensation for Risk Aversion and Skewness Affection in Wages," *Labour Economics*, 14(6), 938–956.

HILL, M. S. (1991): *The Panel Study of Income Dynamics: A User's Guide*, vol. 2. Sage Publications, Inc.

HOLLAND, P. W. (1986): "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81(396), 945–960.

JACOBS, B., J. HARTOG, AND W. VIJVERBERG (2009): "Self-Selection Bias in Estimated Wage Premiums for Earnings Risk," *Empirical Economics*, 37(2), 271–286.

JACOBSON, L. S., R. J. LALONDE, AND D. G. SULLIVAN (1993): "Earnings Losses of Displaced Workers," *American Economic Review*, 83(4), 685–709.

——— (2005): "Estimating the Returns to Community College Schooling for Displaced Workers," *Journal of Econometrics*, 125(1-2), 271–304.

JENSEN, S. T., AND S. H. SHORE (2009a): "The Evolution of the Distribution of Income Volitility," Working Paper, Johns Hopkins University.

——— (2009b): "Semi-Parametric Bayesian Modeling of Income Volatility Heterogeneity and Dynamics," Working Paper, Johns Hopkins University.

KEANE, M., AND K. WOLPIN (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

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.

——— (2009): "Risk Preferences in the PSID: Individual Imputations and Family Covariation," *American Economic Review Papers and Proceedings*, 99, 363–368.

KREBS, T. (2003): "Human Capital Risk and Economic Growth," *Quarterly Journal of Economics*, 118(2), 709–744.

LALONDE, R. J. (1995): "The Promise of Public Sector-Sponsored Training Programs," *Journal of Economic Perspectives*, 9(2), 149–168.

LEVHARI, D., AND Y. WEISS (1974): "The Effect of Risk on the Investment in Human Capital," *American Economic Review*, 64(6), 950–963.

LUCAS, R. E., AND E. C. PRESCOTT (1974): "Equilibrium Search and Unemployment," *Journal of Economic Theory*, (7), 188–209.

MCFADDEN, D. (1974): "Conditional Logit Analysis of Qualitative Choice Bea," *Frontiers in Econometrics, ed. by P. Zarembka*, pp. 105–142.

MEGHIR, C., AND L. PISTAFERRI (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

MINCER, J. (1974): *Schooling, Experience and Earnings*. Columbia University Press.

MONTEIRO, G. (1988): *Robert Frost and the New England Renaissance*. The University Press of Kentucky.

MORTENSEN, D. T., AND C. A. PISSARIDES (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61(3), 397–415.

PACIOS-HUERTA, I. (2003): "An Empirical Analysis of the Risk Properties of Human Capital Returns," *American Economic Review*, 93(3), 948–964.

PISSARIDES, C. A. (2000): *Equilibrium Unemployment Theory*. MIT Press, Cambridge, MA.

PISTAFERRI, L. (2002): "Anticipated and Unanticipated Wage Changes, Wage Risk, and Intertemporal Labor Supply," *Journal of Labor Economics*, 21(3).

PRITCHARD, W. (1984): *Frost: A Literary Life Reconsidered*. Oxford University Press.

ROBERTSON, D., AND J. SYMONS (1990): "The Occupational Choice of British Children," *The Economic Journal*, 100(402), 828–841.

ROGERSON, R., R. SHIMER, AND R. WRIGHT (2005): "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, 43(4), 959–988.

RUHM, C. J. (1991): "Are Workers Permanently Scarred by Job Displacements?," *American Economic Review*, 81(1), 319–324.

SAHM, C. (2007): "How Much Does Risk Tolerance Change?," Federal Reserve Board Working Paper.

SIOW, A. (1984): "Occupational Choice Under Uncertainty," *Econometrica*, 52(3), 631–645.

STINEBRICKNER, T. R. (2001): "Compensation Policies and Teacher Decisions," *International Economic Review*, 42(3), 751–779.

WEISS, Y. (1972): "The Risk Element in Occupational and Educational Choices," *Journal of Political Economy*, 80(6), 1203–1213.

WILLIS, R., AND S. ROSEN (1979): "Education and Self-Selection," *Journal of Political Economy*, 87(5), S7–S36.