The Human Capital Risk-Return Menu

STEPHEN H. SHORE
Department of Economics
Johns Hopkins University
Baltimore, MD 21218
shore@jhu.edu

DANIEL BARTH
Department of Economics
Johns Hopkins University
Baltimore, MD 21218
dbarth4@jhu.edu

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

March 14th, 2010
PRELIMINARY
COMMENTS WELCOME

Abstract

We develop and estimate a structural model in which individuals vary in risk tolerance, ability and idiosyncratic skill in (and taste for) various “careers.” Individuals choose from a set of careers that differ in income risk, typical pay, and other attributes. Expected utility normalizes a logit framework, giving a certainty equivalent interpretation to the parameters of the career choice decision we estimate. Given data on the joint distribution of income volatility and risk tolerance – the former our proxy for income risk from Jensen and Shore (2009a,b), the latter from survey questions about hypothetical income gambles in the PSID – we back out the implied average income risk premium and the importance of idiosyncratic preference for and skill in specific careers. We separate idiosyncratic preference from idiosyncratic skill using the pay gap between high- and low-income risk people with high and low risk-aversion. We identify the change in the risk-return menu that would generate the observed change in the risk distribution since 1970 in the U.S.

1We thank Ellie Pavlick for excellent research assistance.
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)

In making costly investments in a specific career, occupation, or life course, people may consider a wide variety of attributes that go along with that career. This paper aims to develop and estimate a tractable and unified model of such a choice. This is challenging for an economist to attempt for two reasons. First, career choice decisions are largely based on non-pecuniary factors (e.g. flexible hours, working with other people, physically demanding, etc.) we typically don’t observe and about which we have few interesting hypotheses. Second, economists never see exactly what careers people choose or how they feel about them. This makes Frost’s claim (above) untestable without a model of the counter-factual. Without a model of road choice in which we estimate the distribution of roads, how can we know how different they are, if the “road less traveled” would have made any difference?

We overcome these problems using variables we observe and about which economists have well-developed models: a chosen career’s income risk and a person’s risk aversion. While these risk-related variables may affect career choice, they are hardly the most important drivers of such choice. However, we use them to infer the importance of more important but unobservable factors, which we model as idiosyncratic. What is the magnitude of the idiosyncratic taste for and

\[ \text{In fact, a large body of poetry criticism argues that Frost’s intended meaning was not the literal and commonly believed one (Pritchard, 1984). Scholars note “Frost’s decision to make his two roads not very much different from one another, for passing over one of them had the effect of wearing them ‘really about the same.’ ” (Monteiro, 1988) } \]
skill in various careers? How much would two randomly chosen careers differ (in a certainly equivalent, willingness-to-pay sense) in how much a particular individual would like them and how good he would be at them?

We envision a model in which people are endowed with a preference for risk and level of overall ability, broadly conceived as risk-free earning potential. Careers differ in income risk and typical pay. Both careers and individuals may differ in other attributes, which may be unobservable to the econometrician though all are observable to workers.

Ceteris paribus, the relationship between income risk and risk-aversion should 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 (as shown in Figure 1). The local risk-return trade-off (the marginal risk premium) at a given quantity of risk is determined by the risk-aversion (slope of risk-return indifference curve) of the marginal individual, who finds it optimal to choose this quantity of risk.

There is already a large empirical and theoretical literature on this risk-return relationship. While we can estimate this relationship using our novel individual-specific estimates of income volatility alongside PSID survey responses on risk aversion, this is not our primary goal. First, we aim to understand how changes in the risk-return menu can explain changes in the distribution of income risk we observe in the U.S. since 1970. Second, we aim to see if this well-studied and well-understood risk-return relationship can be used to identify other parameters in a

---

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.

Individuals have idiosyncratic taste for some careers over others and idiosyncratic aptitude (and therefore income) in some careers over others. This idiosyncratic taste or aptitude will lead some highly risk-averse people to choose careers with high income risk (that they happen to love more or be better at than careers with less risk); similarly, it will lead some risk-tolerant people to choose careers with low income risk. Figure 2 shows two individuals with the same risk preference who choose different quantities of risk because they happen to excel in (and therefore earn more with) careers with different quantities of risk. The insight of this paper is that the risk-return model in the previous paragraph allows us to identify the variance of these idiosyncratic factors. We know the welfare cost of
deviating from an anticipated income risk choice; by observing the distribution of such deviations we can back out the economic magnitude of the idiosyncratic shocks that made these deviations optimal. We estimate a lower bound of 50% of income on the standard deviation of these values.

This framework provides an economic interpretation to 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 is not a single number for all individuals; there is a distribution of volatility values across individuals at a point in time. Jensen and Shore (2009a,b) provide a method to estimate individual-specific measures of income volatility or the distribution of these volatilities. 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 large gambles (Barsky, Juster, Kimball, and Shapiro, 1997; Sahm, 2007; Kimball, Sahm, and Shapiro, 2008, 2009). We merge these risk-aversion values with Jensen and Shore (2009a,b)’s 1997 estimates of individuals’ volatilities in the PSID. Unsurprisingly, Jensen and Shore (2009a,b) find that individuals who self-identify as risk tolerant 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 levels of income volatility. Our results are identified from shifts in the distribution of income volatility in the PSID (conditional on risk-aversion) as risk-aversion changes.

To identify the model with idiosyncratic taste and skill, we rely on two key assumptions. First, we must assume independence of individual-career-specific values; they must be i.i.d. across individuals within a career and independent of a career’s risk within an individual. While these independence assumptions are strong, they have testable implications for the degree of heteroskedasticity in the
relationship between risk tolerance and chosen risk. Second, we assume that the number of career options is large (going to infinity), or alternatively that underlying shocks have an extreme value distribution. Consequently, the maximum of these idiosyncratic values can be characterized by an extreme value distribution. Extreme value theory does not require that we know the number of careers in the economy. Since the exact definition of what constitutes a career isn’t well defined let alone observable, this is fortunate.

These assumptions (i.i.d. plus extreme value distribution) imply a multinomial 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 a model, the multinomial logit setting can be identified only up to a normalization: doubling the utility from all careers (doubling all coefficients and the

\footnote{Relaxing this assumption is possible, but inherits the challenges present more generally in heteroskedastic logit settings (Bhat, 1995).}
error term) has no affect on career choice. Our model provides a normalization with an economic meaning so that all estimates can be expressed in terms of their certainty equivalents.67

The second assumption (an infinite number of careers) implies a continuous logit structure. In past work, this has been used in spacial or urban economics to study home location choice (Ben-Akiva, Litinas, and Tsunokawa, 1985). There are effectively an infinite number of possible homes along a two-dimensional plane; while researchers don’t really observe which home each person chooses (in the sense of knowing all its attributes or modeling the choice of a particular home), they do observe its location. In our setting, there are effectively an infinite number of careers. While we don’t know exactly what career a person chooses or many of its attributes, we observe where that career is located along an income risk “line.”

Naturally, our normalized logit setting inherits the restrictions on unobservable career and individual attributes of a reduced-form logit model. To restrict career-specific unobservables, we must assume that career-specific unobservables that are not independent of income risk must have the same affect on career choice.

Keane and Wolpin (1997) provides an alternative – and very different – normalizing model of optimal educational investment and subsequent choice from five broad career categories.7

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.

Dynamic data can be used to overcome the problem that ability may be correlated with pay; pay changes when workers change occupations 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”.

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.
regardless of risk aversion. To restrict individual-specific unobservables, we must assume that individual-specific unobservables that affect a career’s value to an individual should not vary based on the income risk of that career.

This model is static in the sense that there is only a single income shock, though we extend the model to the case of realistic income dynamics with optimal multi-period consumption/saving decision.

The strong restriction on the model is that the distribution of risk choices will have the same “shape” regardless of risk-aversion. As risk-aversion increases, the distribution of risk choices will shift away from risky choices with an exponential structure we make explicit. The key insight of the model is that the magnitude of this shift identifies the variance of individual-career-specific taste and skill (together), measured in certainty equivalent units.

The model allows us to use the joint distribution of income risk and risk-aversion to estimate the variance of idiosyncratic taste and skill together. We can separate these using income data. 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 productivity shocks are larger than idiosyncratic taste shocks, 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. Since we observe a similar pay gap for those with high and low risk-aversion, we infer that nearly all idiosyncratic variation is in career taste and not career skill.

We can also use income data to identify the joint distribution of overall in-
dividual earning potential and risk-aversion. While we observe pay, this is the sum of three terms: individual earning potential in any career, the chosen career’s typical pay (including a compensating differential for risk), and the individual’s idiosyncratic ability in their chosen career. We can subtract our model’s estimates of the last two terms from observed pay to obtain individual-specific estimates of overall individual earning potential. These estimates have a low correlation with risk-aversion.

2 Model

We present a model of career choice over risky careers. Individuals choose from a set of career option. Each career option comes with a degree of income risk, typical pay for that career, and other non-pecuniary attributes. Each individual comes with a preference for income risk, overall ability (typical pay for that individual), 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; also, some individuals are idiosyncratically better (more productive, 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, career-specific income risk is realized. We then observe the distribution of career risk chosen for a variety of levels of risk aversion, as well as the average pay by income risk and risk-aversion.
2.1 Setup

2.1.1 Careers

Career options are indexed by \(c \in \{1, \ldots, N_C\}\). Careers have four attributes, \(X_C \equiv \{\sigma^2, y^C, x^{CO}, x^{CU}\}\); \(X^C_c \equiv \{\sigma^2_c, y^C_c, x^{CO}_c, x^{CU}_c\}\) is the set of attributes for career \(c\). \(\sigma^2\) is a measure of the income risk in career \(c\). \(y^C_c\) is a career-specific measure of log pay in career \(c\). \(x^{CO}_c\) are the attributes observable to the econometrician and to workers; \(x^{CU}_c\) are the set of attributes observable to workers but not to the econometrician. \(N_{CO}\) and \(N_{CU}\) are the number of observable and unobservable attributes, respectively.

Later, we will want to allow 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 will be chosen more than others, so that \(f^C\) will typically not be the distribution of the attributes of chosen careers.

2.1.2 People

People are indexed by \(i \in \{1, \ldots, N_I\}\). People have four attributes, \(X_I \equiv \{\gamma, y^I, x^{IO}, x^{IU}\}\); \(X^I_i \equiv \{\gamma_i, y^I_i, x^{IO}_i, x^{IU}_i\}\) is the set of attributes for person \(i\). \(\gamma_i\) 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 \(i\). \(x^I_i \equiv [x^{IO}_i; x^{IU}_i]\) is a vector of covariate attributes of person \(i\); \(x^{IO}_i\) are the set of attributes observable both to the econometrician and to workers in the model; \(x^{IU}_i\) are the set of attributes observable to workers in the model but not to the econometrician. \(N_{IO}\) and \(N_{IU}\) are the number of observable and unobservable attributes, respectively.

---

9 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.
We will assume an infinite number of people, so that \( N \rightarrow \infty \). In this case, \( f^I(X^I) \) is the distribution of peoples’ attributes. Naturally, we will only observe a subset of these people in our data.

### 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, \( X^\varepsilon \equiv \{y^\varepsilon, l^\varepsilon\} \). \( X^\varepsilon_{i,c} \equiv \{y^\varepsilon_{i,c}, l^\varepsilon_{i,c}\} \) is the fit for person \( i \) in career \( c \). \( y^\varepsilon_{i,c} \) is an individual-career-specific measure of log pay (idiosyncratic productivity) of person \( i \) in career \( c \). \( l^\varepsilon_{i,c} \) is an individual-career-specific measure of idiosyncratic enjoyment of person \( i \) in career \( c \).

\[ f^\varepsilon_{i,c}(X^\varepsilon) \] is the joint distribution of \( X^\varepsilon_{i,c} \). We require that \( X^\varepsilon_{i,c} \) and \( X^\varepsilon_{i',c'} \) be identically distributed and independent of one another when \( c \neq c' \). Independence when \( c \neq c' \) is the standard “independence of irrelevant alternatives” assumption present in multinomial logit settings. Independence when \( i \neq i' \) is also required for inference when we estimate the model on data.

We place very limited restrictions on the distribution of \( X^\varepsilon_{i,c} \). We do not require that \( y^\varepsilon_{i,c} \) and \( l^\varepsilon_{i,c} \) be independent of one another. We merely require that the cdf of \( y^\varepsilon_{i,c} + \frac{\alpha}{1-\alpha} l^\varepsilon_{i,c} \) (with a particular \( \alpha \in [0, 1) \) we define in Section 2.1.4) be twice differentiable (de Haan and Ferreira, 2006). The normal and exponential distributions are examples of such distributions. Coupled with the independence assumptions from the previous paragraph, this implies that \( \lim_{N \rightarrow \infty} \max_{c \in C} (y^\varepsilon_{i,c} + \frac{\alpha}{1-\alpha} l^\varepsilon_{i,c}) \) has an extreme value distribution (of Type I, Gumbel).

### 2.1.4 Preferences

The model that follows assumes a risk-averse expected utility maximizing individuals who cares about stochastic income \( Y \) and career enjoyment \( L \) (for leisure).

\(^{10}\)We also require that \( X^\varepsilon_{i,c} \) be independent of \( X^I_I \) and \( X^C_C \).
The individual has Cobb-Douglas preferences over \( Y \) and \( L \), and expected utility preferences over the composite, \( v \). Individual \( i \) in career \( c \) has expected utility

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

\( v \equiv Y_{i,c}^{1-\alpha} L_{i,c}^\alpha \)  

\[
\ln Y_{i,c,\xi} \equiv y_c + y_l + y_x(x_i, x_c) + y_{i,c} + \sigma_c \xi - \frac{1}{2}\sigma_c^2
\]

\[
\ln L_{i,c} \equiv \ell_x(x_i, x_c) + \ell_{i,c}
\]

Composite felicity (equation 2) \( (v) \) is a Cobb-Douglas function of income \( Y_{i,c,x} \) and career enjoyment \( L_{i,c} \). The relative importance of income and career enjoyment is determined by \( 1 - \alpha \). Note that we impose an elasticity of substitution of one and do not allow heterogeneity in \( \alpha \).

For now, we assume a one-period model in which income \( Y \) (in equation 3) is merely equal to consumption. Log income is the sum of: career-specific pay \( (y_c) \), including a premium for size, risk, or non-pecuniary attributes; individual-specific pay or overall productivity \( (y_l) \); the affect of the interaction of individual- and career-specific covariates on pay \( (y_x(x_i, x_c)) \); individual-career-specific pay \( (y_{i,c}) \), the individual’s career-specific productivity; and, the realization of a shock \( (\xi) \) with a standard normal distribution, adjusted so its exponentiated expectation is one. We extend this to a multi-period setting with income dynamics and optimal saving in Section 2.5. Log enjoyment is the sum of: the affect of the interaction of individual- and career-specific covariates on enjoyment \( (\ell_x(x_i, x_c)) \); and, individual-career-specific enjoyment \( (\ell_{i,c}) \).

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

\[
y_x(x_i, x_c) \equiv \ell_x'( \theta_y \cdot (x_i^l x_c^C)) \ell_i; \ell_x(x_i, x_c) \equiv \ell_x'( \theta_l \cdot (x_i^l x_c^C)) \ell_i
\]
The affect of covariates on pay or enjoyment depends on coefficient matrices, $\theta^y$ and $\theta^l$:

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

(6)

$\theta^y$ is a matrix of size $N_{IO} + N_{IU}$ by $N_{CO} + N_{CU}$. To explain the notation here, $\theta^y_{OU}$ 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 “·” indicates pair-wise multipication, and $\iota$ is a vector of ones so that we merely sum up all elements of the matrix of affects, $\theta^y \cdot (x_i^I x_c^C)$ or $\theta^l \cdot (x_i^I x_c^C)$.

### 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 $i$:

$$
V(i,c) \equiv \ln \left( \frac{(1 - \gamma_i) E u}{(1 - \alpha)(1 - \gamma_i)} \right) = y^I_i + y^C_c + y^x(x_i^I, x_c^C) + \varepsilon_{y,i,c} \\
+ \frac{\alpha}{1 - \alpha} \left( I^L_{i,c} + I^x(x_i^I, x_c^C) \right) - \frac{1}{2}(\alpha + \gamma_i - \alpha \gamma_i)\sigma^2_c
$$

(7)

Individuals will choose the career with the highest $V$. Note that the problem is set up so that person-specific ability ($y^I$) has no impact on the career chosen; it merely shift the value of all careers equally. In our data, we observe the curvature parameter of the utility function with respect to income or consumption. And while the importance of idiosyncratic career preference is apparent in the data, we cannot separate large idiosyncratic career taste shocks from important idiosyn-
cratic career taste shocks. This informs the following transformations:

\[ \tilde{\gamma}_i = \alpha + \gamma_i - \alpha \gamma_i \quad (8) \]

\[ l^x(x^I_i, x^C_c) \equiv \frac{\alpha}{1 - \alpha} l^x(x^I_i, x^C_c); \tilde{l}_{i,c} \equiv \frac{\alpha}{1 - \alpha} \tilde{l}_{i,c} \quad (9) \]

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

\[ V(i, c) = y^I_i + y^C_c + y(x^I_i, x^C_c) + l^x(x^I_i, x^C_c) - \frac{1}{2} \tilde{\gamma}_i \sigma^2_c + y_{i,c} + \tilde{l}_{i,c} \quad (10) \]

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

\[ \varepsilon_{i,c} \equiv y_{i,c} + \tilde{l}_{i,c} \text{ and } v(x^I_i, x^C_c) \equiv y(x^I_i, x^C_c) + l^x(x^I_i, x^C_c) \quad (11) \]

then

\[ V(i, c) = y^I_i + y^C_c + v(x^I_i, x^C_c) - \frac{1}{2} \tilde{\gamma}_i \sigma^2_c + \varepsilon_{i,c}. \quad (12) \]

Equation (12) gives career choice a standard random utility, multinomial logit structure (McFadden, 1974). Individuals choose the career that gives them the highest utility, and this utility is affected by career attributes that affect everyone equally (here, \( y^C_c \)), career attributes that affect different individuals differently (here, observable and unobservable covariates \( v(x^I_i, x^C_c) \) and the utility cost of \( \sigma^2_c \) that depends on \( \tilde{\gamma}_i \)), and an error term (here, \( \varepsilon_{i,c} = y_{i,c} + \tilde{l}_{i,c} \)). What is unique here is that our economic model provides an economic interpretation to the utility level normalization (log of certainty equivalent income) and the coefficient on the \( \sigma^2_c \times \tilde{\gamma}_i \) term (\(-\frac{1}{2}\)). Because the economic model provides our normalization, coefficient estimates and the variance 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 or covariates, so that $y_{i,c} = \tilde{l}_{i,c} = 0$. In this setting, there is no person-specific heterogeneity conditional on $\gamma$ (save for heterogeneity in $y'$ which does not affect choice). All individuals with the same $\gamma$ are indifferent among any choices they choose with positive probability.

This implies a weakly (negatively) monotonic relationship between risk aversion and income risk choice. We should never see a more risk tolerant person choosing less income risk. Here, we consider a continuum of careers on some range of $\sigma^2$, which have full support in the sense that all careers are chosen by someone. Let $\gamma(\sigma^2)$ be the risk aversion of the person who chooses income risk $\sigma^2$.\footnote{Because of the full support assumption, each $\sigma^2$ is chosen by someone and therefore maps to a $\gamma$ though a measure zero set of $\sigma^2$ values may map to multiple $\gamma$ 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.}

At an interior optimum, the individual’s first order condition (holding career enjoyment fixed) requires that\footnote{Obtained by differentiating expected utility in equation (12) with respect to $\sigma^2$, setting equal to zero, and rearranging terms.}

$$\frac{dy^C}{d\sigma^2} = \frac{1}{2} \tilde{\gamma}$$

(13)

If equation (13) must hold for each $\{\sigma^2, \tilde{\gamma}\}$ pair and we know the risk aversion of the marginal individual for each $\sigma^2$, then we can trace out $y^C$ as

$$y^c = y^C_0 + \frac{1}{2} \int_0^{\sigma^2} \tilde{\gamma}(x) dx$$

(14)

Here, $y^C_0$ is the sum of log pay and enjoyment 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 observed by individuals with the same survey-based estimate of risk aversion. We model this heterogeneity in equation (12), where $X_{i,c}^\varepsilon \equiv \{y_{i,c}, \tilde{l}_{i,c}\}$ are idiosyncratic individual-career-specific productivity and taste shocks (for person $i$ in career $c$), respectively.

We require either that underlying idiosyncratic terms have an extreme value distribution or that the number of careers $N_C$ be large in this sense that we can use extreme value theory to describe the best career, given the independence and distributional assumptions from Section 2.1.3. As a result, $\lim_{N_C \to \infty} \max_{c \in \{1, ..., N_C\}} \{\varepsilon_{i,c}\}$ has an extreme value (Type I) distribution, with cdf

$$F(x \mid \mu, \beta) = e^{-e^{-(x - \mu)/\beta}}. \quad (15)$$

$\lim_{N_C \to \infty} \beta^2 \propto \text{var}(\varepsilon_{i,c})$; $\beta$ maps directly the standard deviation of $\varepsilon_{i,c}$, $\beta$ measures the degree to which individuals differ in their taste for or skill in various careers.

There are two technical advantages to an extreme value approach. First, $\lim_{N_C \to \infty} (\mu - \ln(ln(N_C)))$ is a constant; increasing the number of careers affects only the location parameter $\mu$, shifting the whole distribution up while leaving its shape (governed by parameter $\beta$) unchanged. As a result, we can normalize out $\mu$, so that we need

$\text{Note that the mean of this distribution is } \mu + \beta \gamma_{em} \text{ (where } \gamma_{em} \approx 0.577 \text{ refers to the Euler-Mascheroni constant) and the standard deviation is } \beta \pi / \sqrt{6}.}$

13
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, $\varepsilon_{i,c}$.

Let $r$ refer to the set of careers in a rectangle $\{ y^C + v(x_i^I, x_i^C), \sigma^2 \}$ for person $i$. 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. If $F(x | \mu, \beta)$ is the cdf for $\max_{c \in C} \varepsilon_{i,c}$, then

$$F_r(x | \mu_r, \beta_r) = (F(x | \mu, \beta))^{s_r} = e^{-e^{-e^{-\frac{(x - \mu_r - \beta \ln(s_r))}{\beta}}} = F(x | \mu + \beta \ln(s_r), \beta).$$

(16)

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 $\varepsilon_{i,c}$. We can then write the preferred career in each range $r$ as providing the following payoff:

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

(17)

Note that we replace $c$ subscripts on $X^C$ with $r$ to indicate that all careers in rectangle $r$ have approximately the same $\{ y^C + v(x_i^I, x_i^C), \sigma^2 \}$ values. In this case, as the number of careers in this range becomes large, the distribution converges to:

$$W(i, r) \equiv \lim_{C \to \infty} \max_{c \in r} V(i, c) = y_i^I + y_c^C + v(x_i^I, x_c^C) - \frac{1}{2} \tilde{\gamma}_i \sigma_c^2 + EV(\mu + \beta \ln s_r, \beta)$$

$$W(i, r) = \mu + \beta \gamma_{em} + y_i^I + \beta \ln(s_r e^{v(x_i^I, x_c^C) - \frac{1}{2} \tilde{\gamma}_C \sigma_c^2}) + EV(-\beta \gamma_{em}, \beta)$$

(18)

\footnote{Note that the extreme value distribution (EV) in equation (18) has mean zero. As $C$ increases, $\mu$ increases by $\ln(\ln(N_C))$. We envision a limiting setting in which for all $c$, $y^C$ falls at this same rate. Therefore, $\lim_{N_C \to \infty} \mu + y_c^C$ converges to a constant. As the number of careers increases, the average}
Given the extreme value distribution, the probability that an individual’s preferred career will lie in range \( r \) is

\[
prob(W(i, r) > W(i, q), \forall q \neq r) = \frac{s_r e^{(y^C_i^r + v(x^I_i^r, x^C_i^r) - \frac{1}{2} \gamma_i \sigma^2_i)} / \beta}{\sum_q s_q e^{(y^C_i^q + v(x^I_i^q, x^C_i^q) - \frac{1}{2} \gamma_i \sigma^2_i)} / \beta}.
\] (19)

The probability that a given range will have the highest value (equation 19) is nothing more than the pdf, the joint distribution of attributes \( X^C \) of careers chosen given \( X^I \). We re-write equation (19) by taking the size of each range to zero, so that the sums become integrals and \( s_r \) becomes \( f^C(X^C) \):

\[
f(X^C \mid X^I) = \frac{f(X^C | \tilde{\gamma} = 0, X^I) e^{-\frac{1}{2} \tilde{\gamma} \sigma^2}}{\int \int \int X^C \frac{f(X^C | \tilde{\gamma} = 0, X^I) e^{-\frac{1}{2} \tilde{\gamma} \sigma^2} dX^C}{f(X^C | \tilde{\gamma} = 0, X^I) e^{-\frac{1}{2} \tilde{\gamma} \sigma^2} dX^C}} (21)
\]

The model implies that a risk-neutral person (equation 22) 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 \( (y^C + v(x^I, x^C)) \). The importance of these career-specific attributes dominate the importance of career frequency when idiosyncratic career preference is small \((\beta \to 0)\). Without idiosyncratic preference, risk-neutral people merely choose the career with the highest \( y^C + v(x^I, x^C) \); the distribution of risk choices is extremely tight around the “best”

\[
W(X^I_i) \equiv E[W(i, r) \mid W(i, r) > W(i, q), \forall q \neq r] = \mu + \beta \gamma + y^I_i + \beta \ln(\sum_q s_q e^{(y^C_i^q + v(x^I_i^q, x^C_i^q) - \frac{1}{2} \gamma_i \sigma^2_i)} / \beta)
\] (20)

Note that 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-being.
choice. However, as idiosyncratic career preference becomes larger ($\beta \to \infty$), 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 $X^I$. We begin by integrating out unobservable components of careers.$^{16}$ Then, we integrate out unobservable individual attributes. To do so, we must make the strong assumption that any individual unobservables must not vary with career risk, so that

$$E \left[ e^{(y^C + v(x^I, x^C))}/\beta \mid x^IO, x^CO, x^CU \right]$$

should not vary with $\sigma^2$. In this case, equations (24) and (25) become:

$$f(\sigma^2 \mid \bar{\gamma}, x^IO, x^CO) = \frac{f(\sigma^2 \mid \bar{\gamma} = 0, x^IO, x^CO)e^{-\frac{1}{2} \bar{\gamma} \sigma^2}}{\int_{\sigma^2_q} f(\sigma^2 \mid \bar{\gamma} = 0, x^IO, x^CO)e^{-\frac{1}{2} \bar{\gamma} \sigma^2_q} d\sigma^2_q}$$

(27)

$$f(\sigma^2 \mid \bar{\gamma} = 0, x^IO, x^CO) = \frac{f^C(\sigma^2 \mid x^CO)E \left[ e^{(y^C + v(x^I, x^C))}/\beta \mid x^IO, x^CO \right]}{\int_{\sigma^2_q} f^C(\sigma^2 \mid x^CO)E \left[ e^{(y^C + v(x^I, x^C))}/\beta \mid x^IO, x^CO \right] d\sigma^2_q}$$

(28)

$^{16}$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, $y^C_{\epsilon_i,c}$) or unobservable career-specific attributes ($X^{CU}$). We integrate these out to obtain the marginal distribution of observable career choices. We require that $\bar{\gamma}$ does not affect the expected payoff of some risk levels more than others, so that

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

does not vary with $\bar{\gamma}$, in which case equations (21) and (22) (after applying Bayes Rule) become:

$$f(\sigma^2 \mid X^I, x^CO) = \frac{f(\sigma^2 \mid X^I, \bar{\gamma} = 0, x^CO)e^{-\frac{1}{2} \bar{\gamma} \sigma^2}}{\int f(\sigma^2 \mid X^I, \bar{\gamma} = 0, x^CO)e^{-\frac{1}{2} \bar{\gamma} \sigma^2_q} d\sigma^2_q}$$

(24)

$$f(\sigma^2 \mid \bar{\gamma} = 0, X^I, x^CO) = \frac{f^C(\sigma^2 \mid x^CO)E \left[ e^{(y^C + v(x^I, x^C))}/\beta \mid x^I, \sigma^2, x^CO \right]}{\int f^C(\sigma^2 \mid x^CO)E \left[ e^{(y^C + v(x^I, x^C))}/\beta \mid x^I, \sigma^2_q, x^CO \right] d\sigma^2_q}$$

(25)
The critical insight from equations (21), (24), or (27) is that the distribution of risk choices for risk-averse people \( f(\sigma^2|\gamma) \) is completely determined by the distribution for risk-neutral people \( \gamma = 0 \) and a single parameter \( \beta \). Each conditional distribution \( f(\sigma^2|\gamma) \) for a given \( \gamma \) is merely an exponential shift of another such conditional distribution for another \( \gamma \). The degree of that shift is governed by \( \beta \), which is proportional to the standard deviation of the idiosyncratic individual-specific-career taste shocks. As these shocks become larger, the shift becomes more modest and conditional distributions look more similar to one another (and more similar to the distribution of careers, \( f^C \)). As these shocks become smaller, the shift becomes more substantial and conditional distributions for high and low \( \gamma \) become more different (and each becomes more concentrated around the “best” choice for that \( \gamma \)). When idiosyncratic shocks are large, the distribution of risk choices by risk-neutral people will reflect primarily the distribution of career options (\( f^C(\sigma_q^2 | x^{CO}) \)); 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 overidentified when we observe the joint distribution of \( \sigma^2 \) and \( \gamma \). One conditional distribution completely determines the shape of all the others, and a single parameter determines the link between all conditional distributions.

### 2.4 Idiosyncratic Career Taste vs. Idiosyncratic Career Skill

Equation (27) provides a way to estimate \( \beta \), from the degree to which the conditional distribution of risk choices shifts with risk-aversion. \( \beta \) measures the standard deviation of \( y_{i,c} + \tilde{l}_{i,c} \). Without additional information, we cannot separate the relative importance of individual-specific shocks to skill in specific careers (\( y_{i,c}^{\varepsilon} \)) from individual-specific shocks to taste for (enjoyment of) those careers \( \tilde{l}_{i,c}^{\varepsilon} \).
However, we can separate these two affects using income data. Observed log pay (ignoring mean-zero income shock $\xi$) is:

$$ \log \text{ pay}_{i,c} \equiv y_i^I + y_c^C + y^x(x_i^I, x_c^C) + y_i^\varepsilon $$  \hspace{1cm} (29) 

Combining equations (12) and (29) yields:

$$ \log \text{ pay}_{i,c} = V_{i,c} - \hat{l}(x_i^I, x_c^C) + \frac{1}{2} \hat{\gamma}_i \sigma_c^2 - \tilde{l}_{i,c}^\varepsilon $$  \hspace{1cm} (30) 

We can then take the expectation of pay conditional on career $c$ having the highest $V_{i,c}$ from the equation(20):

$$ E[\log \text{ pay}_{i,c} \mid V_{i,c} \geq V_{i,c'} \forall c'] = E[W(X^I) - \hat{l}(x_i^I, x_c^C) + \frac{1}{2} \hat{\gamma}_i \sigma_c^2 - E[\tilde{l}_{i,c}^\varepsilon \mid V_{i,c} > V_{i,c'} \forall c']] 
\hspace{2cm} = \mu + \beta \gamma_{em} + y_i^I + \beta \ln(\sum q s_q e^{(y_q^C + v(x^I_q, x^C_q) - \frac{1}{2} \hat{\gamma}_q \sigma_q^2) / \beta}) 
\hspace{2cm} - \hat{l}(x_i^I, x_c^C) + \frac{1}{2} \hat{\gamma}_i \sigma_c^2 - E[\tilde{l}_{i,c}^\varepsilon \mid V_{i,c} > V_{i,c'} \forall c'] $$  \hspace{1cm} (31) 

Next, we take the expectation of $V_{i,c}$ from equation (12) conditional on career $c$ having the highest $V_{i,c}$:

$$ E[V(i, c) \mid V_{i,c} \geq V_{i,c'} \forall c'] = E[W(X^I) = y_i^I + y_c^C + v(x_i^I, x_c^C) - \frac{1}{2} \hat{\gamma}_i \sigma_c^2 
\hspace{2cm} + E[y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon \mid V_{i,c} \geq V_{i,c'} \forall c'] $$  \hspace{1cm} (32) 

Plugging equation (20) into equation (32) and re-arranging terms yields:

$$ E[y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon \mid V_{i,c} \geq V_{i,c'} \forall c'] = \mu + \beta \gamma_{em} + \beta \ln(\sum q s_q e^{(y_q^C + v(x^I_q, x^C_q) - \frac{1}{2} \hat{\gamma}_q \sigma_q^2) / \beta}) - y_c^C - v(x_i^I, x_c^C) + \frac{1}{2} \hat{\gamma}_i \sigma_c^2. $$  \hspace{1cm} (33)
By assuming joint normality of $y_{i,c}^\varepsilon$ and $\tilde{l}_{i,c}^\varepsilon$, $E[y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon \mid V_{i,c} \ge V_{i,c}' \forall c']$ from equation (33) identifies $E[\tilde{l}_{i,c}^\varepsilon \mid V_{i,c} > V_{i,c}' \forall c']$ in equation (31):

$$E[\tilde{l}_{i,c}^\varepsilon \mid V_{i,c} \ge V_{i,c}' \forall c'] = \frac{\text{var}(\tilde{l}_{i,c}^\varepsilon)}{\text{var}(y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon)}(\mu + \beta \gamma_{em} + \beta \ln(\sum_q s_q e^{(y_q^\varepsilon + v(x_i^I,x_q^C) - \frac{1}{2}\gamma_i\sigma_q^2)/\beta}))$$

$$- y_{i,c}^C - v(x_{i,c}^I,x_{i,c}^C) + \frac{1}{2} \tilde{\gamma}_i \sigma_c^2). \quad (34)$$

Plugging equation (34) into equation (31) yields:

$$E[\log \text{pay}_{i,c} \mid V_{i,c} \ge V_{i,c}' \forall c'] = (\mu + \beta \gamma_{em})(1 - \frac{\text{var}(\tilde{l}_{i,c}^\varepsilon)}{\text{var}(y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon)})$$

$$+ y_{i,c}^I + \beta \ln(\sum_q s_q e^{(y_q^\varepsilon + v(x_i^I,x_q^C) - \frac{1}{2}\gamma_i\sigma_q^2)/\beta})(1 - \frac{\text{var}(\tilde{l}_{i,c}^\varepsilon)}{\text{var}(y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon)})$$

$$+ y_{i,c}^C \frac{\text{var}(\tilde{l}_{i,c}^\varepsilon)}{\text{var}(y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon)} - v(x_{i,c}^I,x_{i,c}^C)(1 - \frac{\text{var}(\tilde{l}_{i,c}^\varepsilon)}{\text{var}(y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon)})$$

$$+ \frac{1}{2} \tilde{\gamma}_i \sigma_c^2(1 - \frac{\text{var}(\tilde{l}_{i,c}^\varepsilon)}{\text{var}(y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon)}). \quad (35)$$

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 $y_{i,c}^I$ and $\gamma_i$) but not chosen career attributes. The third line depends on chosen career attributes (specifically $y_{i,c}^L$ and $x_{i,c}^{CO}$ but not $\sigma_c^2$) and individual observables ($x_{i,O}^I$). The final line depends on both an individual attribute ($\gamma_i$) and a career attribute ($\sigma_c^2$).

Equation (35) suggests a simple regression that can be used to recover the relative importance of taste shocks ($\tilde{l}_{i,c}^\varepsilon$) compared with all shocks ($y_{i,c}^\varepsilon + \tilde{l}_{i,c}^\varepsilon$). In a regression to predict pay with a constant (to capture the first line), individual-specific controls (including risk aversion, to capture the second line), career-attribute controls (particularly a measure of income risk, to capture the third line), and the interaction of individual- and career-specific controls (besides risk aversion and income...
risk) then the fourth line shows that the coefficient interation between risk aversion ($\gamma_i$) and income risk ($\gamma_i$) identifies $\frac{1}{2} \times \left(1 - \frac{\text{var}(\tilde{t}_{i,c})}{\text{var}(y_{i,c} + \tilde{t}_{i,c})}\right)$.

If people dislike risk, they must be compensated in some way for taking on 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.

2.5 Dynamics

We enrich this model by making it dynamic. This involves allowing more realistic income dynamics as well as optimizing inter-temporal consumption decisions. We assume that an individual chooses a “career”, namely a starting salary (which includes a risk premium) and a corresponding level of income risk. We consider optimal career-switching later, but for now we assume that income volatility and risk premia never change.

We assume that income evolves in response to permanent and transitory shocks (following Hall and Mishkin, 1982; Carroll and Samwick, 1997, and many others). Excess log income $y_{i,t}$ is modeled as the sum of permanent income and transitory income.

$$y_{i,t} = y_{i,0} + \sum_{k=1}^{t} \omega_{i,k} + \varepsilon_{i,t} \quad (36)$$

Excess log income refers to the unpredictable component of income (the same
across all careers) that can be predicted from immutable demographic characteristics. We can calibrate this to average age-based data for male household heads from the PSID. Risk premia and ability are captured in initial excess log income, \( y_{i,0} \). Permanent income is i.i.d. in (log) changes; transitory income is i.i.d. in (log) levels. Permanent income changes (\( \omega_{i,t} \)) have mean zero and variance \( \sigma_{\omega,i,t}^2 \), also referred to here as the permanent variance or volatility; transitory income (\( \varepsilon_{i,t} \)) has mean zero and variance \( \sigma_{\varepsilon,i,t}^2 \), also referred to here as the transitory variance or volatility; This is merely the stylized version of the income process estimated by Jensen and Shore (2009a,b) and used in Section 3.

We replace the one-period utility function (equation 1) with its multi-period analog:

\[
Eu_{i,t} = \sum_{s=t}^{T} \beta^{T-s} E \left[ \frac{C_{i,s}^{1-\gamma}}{1 - \gamma} \right]
\]  

(37)

We assume the agent chooses an optimal consumption path, subject to the intertemporal budget constraint linking wealth \( W_{i,t} \), income \( Y_{i,t} \), and consumption \( C_{i,t} \):

\[
W_{i,t+1} = (W_{i,t} + Y_{i,t} - C_{i,t}) \times (1 + r_t); \quad W_{i,t} \geq 0.
\]  

(38)

In this exercise, we consider a forty-year time horizon, a typical work life for an individual who begins work in their early twenties. In this setting, it is straightforward to solve the optimal consumption/saving decision through backward induction. This yields an initial value function, \( V(\gamma, y_0, \sigma_\omega, \sigma_\varepsilon) \). It is straightforward to compare careers numerically, either to trace out risk-return indifference curves or to choose between discrete career choices.
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 male household heads aged 22 through 60\(^\text{17}\) gives us 52,181 observations on 3,041 individuals with 17 years of recorded data per individual on average. As mentioned previously, we focus on excess log income as our outcome measure, which is the residual from a regression to predict the natural log of labor income. Summary statistics about the demographics of this group are shown in Table 1.

3.1 Income Volatility

Jensen and Shore calculate the volatility of excess log income as our outcome measure, which is the residual from a regression to predict the natural log of labor income. This regression is weighted by PSID-provided sample weights, normal-

ized 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.\^18

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.

Our income dynamics 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 \(e_{i,t}\).

\[
y_{i,t} = \sum_{k=1}^{t-3} \omega_{i,k} + \sum_{k=t-2}^{t} \phi_{\omega_{i,t-k}} \cdot \omega_{i,k} + \sum_{k=t-2}^{t} \phi_{\varepsilon_{i,t-k}} \cdot \varepsilon_{i,k} + e_{i,t} \quad (39)
\]

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.

In this framework model, permanent shocks come into effect over three pe-

\^18 Working with excess log real income is also standard in this literature (Meghir and Pistaferri, 2004; Carroll and Samwick, 1997).
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 \(\phi\) 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 \((\sum_k \phi_{\varepsilon,k} = 1)\).

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 \(\sigma_{\omega,i,t}^2 \equiv E[\omega_{i,t}^2]\); the transitory shocks \(\varepsilon_{i,t}\) have mean zero and variance \(\sigma_{\varepsilon,i,t}^2 \equiv E[\varepsilon_{i,t}^2]\):

\[
\begin{pmatrix} \omega_{i,t} \\ \varepsilon_{i,t} \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\omega,i,t}^2 & 0 \\ 0 & \sigma_{\varepsilon,i,t}^2 \end{bmatrix} \right)
\]

We refer to \(\sigma_{i,t}^2 \equiv (\sigma_{\varepsilon,i,t}^2, \sigma_{\omega,i,t}^2)\) jointly as the volatility parameters, Finally, we have “noise variance” which refers to the variance of measurement error \(\gamma^2 \equiv E[e_{i,t}^2]\) 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 individual-specific estimates of volatility and their evolution as well as the cross-sectional distribution of volatility and its evolution. In a standard Dirichlet Process (DP) prior model, the pair \(\sigma_{i,t}^2 \equiv (\sigma_{\varepsilon,i,t}^2, \sigma_{\omega,i,t}^2)\) can have one of \(L\) possible values, \(\{\sigma_{i}^2\}_{i=1}^L\) (where these values and their number are estimated, but there is a prior on the number of such values). In a hierarchical DP (HDP) model (recently developed by Teh, Jordan, Beal, and Blei, 2007), the usual DP model is extended so by adding a second tuning parameter, \(\Theta_i\), which implicitly places a prior on the total number of unique parameter values for any given individual, \(L_i\); we set \(\Theta_i = 1\). Jensen and Shore (2009a,b) extend this
Diagram describes evolution of volatility parameters.

The decision tree for evolving volatility is shown in Figure 3. Volatility can change from one year to the next, or remain unchanged. If it changes, it can change to a value that this individual had in another year or a new value the individual has never had before. If it changes to a new value to that individual, it can change to a value that another individual in the population has had at some point, or to a new value never seen before in the population. See Jensen and Shore (2009a,b) for the probabilities of these. This paper uses 1997 estimates of permanent volatility obtained in Jensen and Shore (2009a,b).
3.2 Risk-Aversion

In 1996, the PSID included a series of survey questions which aimed to elicit estimates of risk tolerance, $1/\gamma$. Respondents were asked whether they would be willing to take a series of lotteries which varied in compensation for risk. Based on which gambles the respondents were and were not willing to take, risk tolerance was identified as within 4 ranges.

Table 2 presents the joint distribution of volatility and risk aversion. In that table, $\sigma^2$ values are divided into 10 bins. These are formed using $\sigma$ intervals of 0.01, grouping such intervals together until each grouping contained at least 0.5% of the $\sigma$ values. The distribution of $\sigma^2$ values is also shown in the upper-left panel of Figure 4; this distribution is broken into the coarser bins in the lower-left panel. The upper-right panel shows the proportion of the data in each of these 10 bins; the upper-right and lower-left panels are identical except that their axes are scaled differently. Note that individuals with high risk aversion were slightly less likely to have the highest volatility values. However, the conditional distribution of income volatility is similar for those identified as risk tolerant or risk intolerant.

4 Estimation

If we could observe the joint distribution of data $\{\sigma^2, \tilde{\gamma}, x^{IO}, x^{CO}\}$, then estimation of equations (27) and (28) by maximum likelihood is straightforward. We need only choose a parametric (or nonparametric) structure for $f(\sigma^2 \mid \tilde{\gamma} = 0, x^{IO}, x^{CO})$, and estimate its parameters along with $\beta$ by maximum likelihood. Table 2 shows the non-parametric approach we pursue, splitting $\sigma^2$ into 10 ranges. We assign each range the $\sigma^2$ value of its mid-point. Covariates aside, we need only estimate nine probabilities, the probability that a risk-neutral person will land in each of the
The upper-left panel presents the distribution of 1997 σ estimates from Jensen and Shore, a histogram of the standard deviation of permanent income changes. The lower-left panel shows how we collect these values into 10 bins. The upper-right panel is identical to the lower-left in showing the fraction of individuals in each σ bin; it differs only in the x-axis, which shows bin instead of σ values. The lower-right panel presents estimates of $f(\sigma | \tilde{\gamma} = 0)$. These are normalized by dividing by the value in the upper-right panel 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).
Table 2: Conditional distribution of income volatility, $\sigma$

<table>
<thead>
<tr>
<th>$\sigma &gt; \sigma \leq$</th>
<th>probability in a given $\sigma$ range given $\tilde{\gamma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[0, 2)</td>
</tr>
<tr>
<td>0.13 0.14</td>
<td>0.8%</td>
</tr>
<tr>
<td>0.14 0.15</td>
<td>40.4%</td>
</tr>
<tr>
<td>0.15 0.16</td>
<td>50.3%</td>
</tr>
<tr>
<td>0.16 0.17</td>
<td>0.5%</td>
</tr>
<tr>
<td>0.17 0.18</td>
<td>1.8%</td>
</tr>
<tr>
<td>0.18 0.19</td>
<td>0.8%</td>
</tr>
<tr>
<td>0.19 0.21</td>
<td>1.6%</td>
</tr>
<tr>
<td>0.21 0.29</td>
<td>1.3%</td>
</tr>
<tr>
<td>0.29 0.99</td>
<td>1.6%</td>
</tr>
<tr>
<td>0.99 1</td>
<td>1.0%</td>
</tr>
<tr>
<td># of Obs.</td>
<td>384</td>
</tr>
<tr>
<td>% of Obs.</td>
<td>22.0%</td>
</tr>
</tbody>
</table>

Table shows the distribution of $\sigma^2$ estimates. $\sigma^2$ estimates are from Jensen and Shore and are top-coded at 1. $\gamma$ ranges are from the coarsely-binned responses to the 1996 risk-tolerance supplement to the PSID.

10 volatility bins.

The complication is that we do not observe $\tilde{\gamma}$ exactly; we see only into which of four coarse bins $\tilde{\gamma}$ falls. Furthermore, there is measurement error in $\tilde{\gamma}$, so that the true value for $\tilde{\gamma}$ 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 $\tilde{\gamma}$ 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 $\tilde{\gamma}$:

$$\ln(1/\tilde{\gamma}) = \ln(1/\gamma) + e$$ (41)

$$\begin{bmatrix} \ln(\gamma) \\ e \end{bmatrix} \sim N \left( \begin{bmatrix} -1.05 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.76 & 0 \\ 0 & 1.69 \end{bmatrix} \right)$$ (42)

We observe true log risk tolerance ($\ln(1/\tilde{\gamma})$) plus noise ($e$), placed into bins, so that
a given observation lies in a given bin if \( \ln(1/\gamma) > \text{bin} \) and \( \ln(1/\gamma) < \text{bin} \). Again, Table 2 shows these ranges and the fraction of observed data that falls into each.\(^{19}\)

We can then identify the relationship between our data \( f(\sigma^2 | 1/\ln(\gamma) \text{ bin}) \) and the object we wish to estimate \( f(\sigma^2 | \gamma) \) from equation (28):

\[
f \left( \sigma^2 | 1/\ln(\gamma) \text{ bin} \right) = \int_{\gamma} f \left( \sigma^2 | \gamma \right) f_{\ln(1/\gamma)} \left( \ln \left( 1/\gamma \right) | \text{bin} \right) d\gamma \quad (43)
\]

\[
f \left( \ln \left( 1/\gamma \right) | 1/\ln(\gamma) \text{ bin} \right) = f_{\ln(1/\gamma)} \left( \ln \left( 1/1/\gamma \right) \right) \frac{\text{pr} \left( 1/\ln(\gamma) \text{ bin} | \gamma \right)}{\text{pr} \left( 1/\ln(\gamma) \text{ bin} \right)} \quad (44)
\]

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

Armed with this distribution of \( \gamma \), we can search for maximum likelihood estimates of \( f(\sigma^2 | \gamma = 0) \) and \( \beta \) iteratively. First, we guess values of \( f(\sigma^2 | \gamma = 0) \) and \( \beta \). Next, we calculate \( f(\sigma^2 | \gamma) \) for each value of \( \sigma^2 \) and \( \gamma \) on our grid. Next we calculate \( f(\sigma^2 | \gamma) \) for each of the 10 grid values of \( \sigma^2 \) and each of the four coarse bins for \( \gamma \) by integrating over each value of \( \gamma \) 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 2. We can then compute the likelihood of observing the data in Table 2. We search over \( f(\sigma^2 | \gamma = 0) \) and \( \beta \) to find the one that maximizes the likelihood function.

\(^{19}\)We approximate this distribution with a 38 element grid, assigning a probability that \( \gamma \) 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\}.
5 Results

Equations (27) and (28) show the key model parameters to estimate. We want to estimate \( f(\sigma^2 \mid \tilde{\gamma} = 0, x^{IO}, x^{CO}) \) (the distribution of \( \sigma^2 \) chosen by risk-neutral people in equation 28) and \( \beta \) (\( \propto \text{var}(\varepsilon_{i,c}) \) from equation 27, the shift in the distribution of \( \sigma^2 \) as risk-aversion increases).

The lower-left panel of Figure 4 shows the distribution of \( f(\sigma^2 \mid \tilde{\gamma} = 0, x^{IO}, x^{CO}) = f(\sigma^2 \mid \tilde{\gamma} = 0) \) when the model is estimated without additional covariates. We obtain a 95% confidence interval around these estimates by finding the highest and lowest values of \( f(\sigma^2 \mid \tilde{\gamma} = 0, x^{IO}, x^{CO}) \) (separately for each \( \sigma^2 \) bin such that the restricted model fails to reject the likelihood ratio test that these restricted \( f(\sigma^2 \mid \tilde{\gamma} = 0, x^{IO}, x^{CO}) \) values are correct.

The \( \beta \) point estimate is 1.16, so that the standard deviation of idiosyncratic career values is \( 116\% \times \pi/\sqrt{6} \) of income (in log points). While we cannot reject the hypothesis of an infinite \( \beta \), we can place a lower bound on \( \beta > 0.38 \). While these estimates are large, we view the lower-bound on \( \beta \) estimates as entirely plausible; it implies a dispersion of idiosyncratic taste or skill of 48% of income.

Equation (28) shows that the distribution of \( \sigma^2 \) choices by risk-neutral people may reflect the distribution of career options \( f^C \) or the relative value of those options \( (E[(y^C + v(x^I, x^C))^{1/\beta} \mid \sigma^2]) \). There is no way to differentiate these options without a model of wage adjustment. At one extreme, we can assume that the demand for 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 the risk-return menu \( (y^C + v(x^I, x^C)) \), assuming no heterogeneity conditional on \( \sigma^2 \) as \( (\frac{f^C(\sigma^2=0)}{f(\sigma^2)})^{1/\beta} \). Estimates of this risk-return menu are shown in Figure 5.

At the other extreme, we can assume that demand for each career is completely
elastic, so that the value of each career is the same in expectation so that $f(\sigma^2 | \tilde{\gamma}) = 0 = f^C$. These are shown in the lower-right panel of Figure 4.

Next, we use income data to decompose $\varepsilon_{i,c}$ into idiosyncratic career skill ($y_{i,c}$) and taste ($\tilde{l}_{i,c}$) using the result from equation (35). 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 3 shows the results from regressions to predict pay with $\gamma$, $\sigma^2$, their interaction, and covariates. Here, $\gamma$ refers to $E[\gamma | \tilde{\gamma} \text{bin}]$ based on the distribution of $\tilde{\gamma}$ and measurement error proposed by Kimball, Sahm, and Shapiro (2009). Note that the coefficient on $\frac{1}{2} \times \sigma^2 \times \gamma$ is approximately zero, and we can reject the hypothesis that
Table 3: Impact of income risk and risk-aversion on income

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Log Average Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
<td>0.190 0.078 0.068 -0.039 2.252</td>
</tr>
<tr>
<td></td>
<td>(0.114) (0.099) (0.261) (0.226) (0.968)*</td>
</tr>
<tr>
<td>$\sigma^2 \times \sigma^2$</td>
<td>-0.486</td>
</tr>
<tr>
<td></td>
<td>(0.256)</td>
</tr>
<tr>
<td>$\ln(\sigma^2)$</td>
<td>-0.449</td>
</tr>
<tr>
<td></td>
<td>(0.117)**</td>
</tr>
<tr>
<td>$\gamma = \text{lowest}$</td>
<td>omitted category</td>
</tr>
<tr>
<td></td>
<td>0.014 -0.026 -0.028</td>
</tr>
<tr>
<td></td>
<td>(0.063) (0.055) (0.055)</td>
</tr>
<tr>
<td>$\gamma = 2^{\text{nd}} \text{lowest}$</td>
<td>0.086 0.064 0.066</td>
</tr>
<tr>
<td></td>
<td>(0.064) (0.056) (0.056)</td>
</tr>
<tr>
<td>$\gamma = 2^{\text{nd}} \text{highest}$</td>
<td>0.007 -0.010 -0.006</td>
</tr>
<tr>
<td></td>
<td>(0.052) (0.047) (0.047)</td>
</tr>
<tr>
<td>$\gamma = \text{highest}$</td>
<td>0.007 -0.010 -0.006</td>
</tr>
<tr>
<td></td>
<td>(0.052) (0.047) (0.047)</td>
</tr>
<tr>
<td>$\frac{1}{2} \times \sigma^2 \times \gamma$</td>
<td>0.060 0.057 -0.028</td>
</tr>
<tr>
<td></td>
<td>(0.121) (0.105) (0.128)</td>
</tr>
<tr>
<td>age</td>
<td>no yes no yes yes</td>
</tr>
<tr>
<td>family size</td>
<td>no yes no yes yes</td>
</tr>
<tr>
<td>race</td>
<td>no yes no yes yes</td>
</tr>
<tr>
<td>education</td>
<td>no yes no yes yes</td>
</tr>
<tr>
<td>observations</td>
<td>1,876 1,876 1,876 1,876 1,876</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.00 0.26 0.27 0.27 0.27</td>
</tr>
</tbody>
</table>

Standard errors in parenthesis: *significant at 5% level; **significant at 1% level. 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.
it is greater than about 25%. This implies that \((1 - \frac{\text{var}(\tilde{l}_{i,c})}{\text{var}(\tilde{y}_{i,c}) + \tilde{l}_{i,c}}) < 0.25\) from equation (35). Nearly all idiosyncratic variation in career taste or skill comes from taste and not skill. People vary widely in how much they think they would like various jobs, not in how good they would be at those jobs. Note the similarity between this regression and the risk-augmented Mincer equations from Hartog (2009).

6 Conclusion

Recent research has documented an increase in income volatility – the variance of income changes – over the last 40 years in the U.S. To the degree that income volatility is chosen (given risk preferences), this implies a change in the menu of risk-return options. We back out the change in the risk-return menu implied by changes in the distribution of income volatility (documented in Jensen and Shore (2009a)) and by implication the joint distribution of risk-free earning potential (ability) and risk aversion. Absent sorting of risk-tolerant people into high volatility jobs, the welfare cost of increased risk would be large; once we account for this sorting, the welfare cost of increasing volatility is small.

Our structural model provides an interpretation of the relative importance of risk and individual-career-specific skill in career choice.
References


