Comparative Advantage and the Risk Premium in Labor Markets

German Cubas and Pedro Silos*

Preliminary Draft

July 7, 2012

Abstract

Using the Survey of Income and Program Participation (SIPP) we document a significant and positive association between earnings risk (both permanent and transitory) and average log-earnings across 21 industries. To understand the nature of that correlation we construct an environment where (i) risk-averse individuals choose an industry in which to supply labor (ii) labor earnings are subject to industry-specific risk estimated from SIPP, but opportunities to insure against that risk are limited, and (iii) individuals have different comparative advantages that affect their industry choice. We use this model to decompose how much of the empirical correlation between risk and mean earnings represents a compensating differential and how much of it is due to selection. We find the positive association between permanent risk and earnings to be mostly compensation for risk. But selection is responsible for the observed relationship between temporary risk and mean earnings.

1 Introduction

At any given point in time the distribution of earnings across individuals reflects, among other factors, compensation for undesirable job characteristics.Workers sort themselves into different careers which are often times attached to a sector or industry. If workers are risk-averse and sectors differ in the degree of earnings risk workers face, the distribution of labor earnings will reflect differences across industries. Imagine a risk-averse economist with a given set of skills entering the labor market for the first time. Suppose those skills can be applied equally well in two industries: finance and higher education. Of the two, finance features larger high-frequency variation in earnings. If this is the only difference between the two sectors, our risk-averse economist will demand a premium for working in the riskier industry. In general, the magnitude of this premium will depend on several factors such as the degree of risk-aversion of workers, the ability of individuals to insure against earnings risk, or the portability of skills should the worker decide to move to an alternative industry if shocks to earnings are persistent. The first goal of this paper is to estimate the relationship between temporary and permanent shocks to earnings, and the mean levels of earnings across industries.

The paper has two parts. The first part employs the Survey of Income and Program Participation (SIPP) to estimate quarterly labor earnings risk across 21 industries of the US economy. Our definition of earnings risk is broad, encompassing unemployment spells, unexpected drops in hours, and drops in wages. Both the definition of risk and the estimation methodology follow a large literature modeling earnings dynamics using panel data. We find substantial differences in the degree of labor earnings risk across industries. Workers in the financial or transportation industries experience large permanent shocks to earnings while those working in social services are insulated from earnings variability. Working for the government entails low permanent risk but high transitory risk. Moreover, the evidence favors a positive cor-

relation between mean earnings and earnings risk, once we control for other industry characteristics that affect the average level of earnings. The estimated coefficients imply that when considering the permanent shocks to earnings the difference in average earnings between the riskiest and safest industries is around 10%. When shocks are transitory, moving from the safest to the riskiest industry implies an increase in mean earnings of 8%.

It is tempting to interpret the estimated correlation as a differential that compensates risk-averse workers for bearing risk. But the correlation may well reflect that the distributions of workers' abilities increase average earnings in risky sectors. From reduced-form estimates it is not possible to disentangle the two; fixed-effects estimates from individuals' earnings regressions are a convolution of the two channels. To understand the nature of the estimated correlation, the second part presents an overlapping-generations model in which risk-averse individuals, besides a standard consumption versus savings choice, choose an industry in which to supply labor services. Some industries are riskier than others and, everything else equal, they are less attractive. Individuals are ex-ante heterogeneous since each of them has a sector-specific skill or comparative advantage. In the spirit of Roy's original model, an individual can be very productive in the Finance sector but not so much in Agriculture. Absent these fundamental differences in the distribution of abilities, when facing a high volatility of earnings in some industries, individuals seek safer alternatives, supplying more labor in low-risk sectors and hence depressing wages. In equilibrium, the nature of the earnings distributions across industries is shaped by the two different channels: On the one hand, the aversion of workers to supply labor to risky industries and on the other hand the distribution of abilities that determine their comparative advantage.

In our empirical application, we discipline the relative level of risk across industries by using the variances of persistent and transitory shocks estimated in the first part of the paper. In addition, our calibrated economy matches the labor's share

3

across different sectors of the US economy. We also parameterize the distribution of abilities so that the model delivers the mean and standard deviations of the cross-sectional distributions of earnings observed in the data. As a result of the sorting of workers, a natural distribution of mean earnings and industry-risk arises. Interestingly, the model predicts a distribution of workers into sectors that closely resembles the one observed in the US data.

Viewed through the lens of the model, the positive relationship between the variance of both the permanent and transitory shocks to earnings and the average level of earnings are a convolution of two forces: the compensation for risk and the compensation for sector specific skills. Therefore, in order to decompose the effect of these two forces into the observed differences in mean earnings we proceed to perform a counterfactual exercise in which we shut down all the action coming from the individuals pre-labor market skills or comparative advantages, in other words, we make individuals living in our model economy to be ex-ante homogeneous. In this counterfactual world only the differences in the volatility of earnings across sectors shape the individuals' sectoral choice.

With reasonable levels of risk aversion, the model over predicts the positive correlation between mean earnings and permanent risk, i.e. a risk premium that is higher than in the data. On the contrary, it predicts a temporary risk premium that is virtually zero. Therefore, the almost equally strong association between the variance of transitory shocks and mean earnings observed in the data which, in light of the reduced form model can be interpreted as a pure risk premium, is entirely coming from selection. A large fraction of individuals possesses skills which increase productivity in industries with relatively large transitory shocks. So even though they are risk-averse, based on their comparative advantage they work in high (temporary) risk industries. Finally, in this counterfactual economy, as expected the distribution of workers into sectors is at odds with the data.

The question at hand has ramifications across several areas in economics. Under-

4

standing sources of inequality should be central in the analysis of a wide range of policies considered in macroeconomics, public finance and labor economics. Whether a large fraction of inequality observed early in life arises solely from career choices will influence the benefits of policies targeting income redistribution. Providing a quantitative answer to that question is at the core of this paper.

Our work contributes to a growing literature that develop theoretical models of labor markets occupational choice and income dynamics that are taken to data. Regarding our empirical analysis, to our knowledge, the first attempt to empirically analyze the link between the variability of income and mean earnings goes back to the seminal work of Kuznets and Friedman (1939) in their classic study of income of professionals. Using data on just a few professions, they find that professions with high mean income where those with high variability of earnings over time. In the same line, Feinberg (1981) examines whether or not the labor market compensates workers for increased earnings-risk. Using only six years of PSID data the author performs a regression analysis and concludes that a significant part of average hourly earnings reflects compensation for earnings-risk. Moreover, the magnitude of the risk premium is significantly lower for women, laborers, and industry-changers than for others. The analysis was enriched both theoretically and empirically with the work of Leigh (1983) who modeled job choice across industries in a partial equilibrium framework under idiosyncratic earning risk. In his econometric analysis, the author focuses on two main categories of jobs across industries: white collar and blue collar. He found that white collar workers are compensated for taking jobs with high earning uncertainty but the blue collar workers are not.

Although answering a different research question, related to our focus on earnings variability across industries of the US economy, an influential work is Carroll and Samwick (1997) who analyze the relationship between income uncertainty and wealth. As in our case, they estimate permanent and transitory earning risk by industry, occupations and educational groups by proposing a novel econometric approach and

using PSID data. They provide evidence that wealth is higher for workers that face higher income uncertainty. More recently, a work contemporaneous to ours, Dillon (2012) estimates the parameters of a life cycle model of occupation, labor supply, and consumption choices using data from the Panel Study of Income Dynamics. Even though this model does not incorporate the occupational choice, it is rich enough to include multiple sources of occupation-specific earnings uncertainty. The main finding is that the relationship between the expected value and variance of lifetime earnings shows that compensation for earnings risk is a key explanation of variation in expected lifetime earnings across careers. We see this finding as complementary evidence of the empirical fact we document which motivates the development or our theoretical model.

Provided that an important contribution of our paper is to measure idiosyncratic labor market risk and its macroeconomic implications in a general equilibrium framework, our work is closely related to Storesletten, Telmer, and Yaron (2004a) Storesletten, Telmer, and Yaron (2004b), Heathcote, Storesletten, and Violante (2008), Heathcote, Storesletten, and Violante (2009) ; Low, Meghir, and Pistaferri (2010) and Guvenen (2009). We see our study as contribution to this strand of the literature since we not only estimate earning risk for the whole economy but also its differences across sectors and its interaction with worker allocations. Moreover, these sectoral differences could be of potential interest for future research regarding the estimation of the persistence of earning shocks for which the literature is far from a consensus (see Guvenen (2007)).

Finally, our work is closely related to Roy's seminal work (Roy (1951)) which was the first to study the effect of self-selection into the observed differences in mean earnings of workers performing different occupations. Roy's ideas are also adapted in modern dynamic discrete choice models to analyze the sources of income inequality, firstly in an important paper, Keane and Wolpin (1997) and, recently in Hoffmann (2010). However, we see our work as being the first that integrates Roy's ideas into

the analysis of the risk that workers face by using a dynamic general equilibrium framework. We see this as our main technical innovation and it is crucial in helping us in our quantitative application since it allows us to obtain the true compensation for risk in the US labor market. Nevertheless, the framework proposed here could be used for future work interested in incorporating workers comparative advantages into the analysis of earnings dynamics or in the analysis of various sources of wage inequality.

2 Labor Market Risk and Mean Earnings: The Evidence

In this section, after briefly describing our data set, we document that risk and return in earnings are positively correlated across industries. We do this in two steps. First, we estimate the labor earning processes and properties of the shocks that workers of different sectors face in their work lives. Second, we characterize and estimate the empirical relation between mean earnings and the dispersion of earning shocks across sectors.

Our definition of labor earnings is rather broad (but consistent with previous studies). Besides the obvious variability in wage rates, we also consider changes in earnings due to changes in the amount of hours worked or changes in the employment status.[1] As we make clear below, those changes which may be predicted based on information about individuals are not included in our measure of risk. For instance, if on average women who are between 25 and 30 years old begin working part-time after having been full-time employees, this decrease the amount of hours worked, and the resulting earnings drop, is not considered risk by our methodology.

We focus on individuals who never change industries, as this is most consistent with the quantitative framework we use below. In reality, individuals can switch industries anytime they want, but given the small numbers of inter-industry switches

---

[1]We do not consider individuals that move in and out of the labor force, but we do consider employment to unemployment transitions and vice versa.

we observe in the SIPP, it seems that the cost of changing industries is high. Perhaps the low volume is caused by the close association between an industry and an occupation. After all, there are not many truck drivers in finance or medical doctors in agriculture, and there is a lot of human capital specific to an occupation that cannot be transferred easily to alternative occupations. On the other hand, not observing many industry changes could be due to the short length of the SIPP panel. Individuals' labor market histories usually span decades and industry switches could occur at lower frequencies than those represented by three- or four-year intervals observed in our data set. In any case, if industry switching is not used as an income-smoothing device in the face of the high-frequency shocks that are the focus of our research, its omission is probably inconsequential.

2.1 The SIPP

To explore whether the level of average earnings and the degree of unforeseen variability in those earnings are positively related, we turn to data. Ideally, to get an accurate answer to that question one would hope for a long high-frequency large panel of individual labor earnings with characteristics describing both the employee and the employer. The richer that data set, the easier would be to separate risk from other features that could affect average earnings. For the United States, the Survey of Income and Program Participation (SIPP) is the best approximation to that ideal data set. It is constructed by the U.S. Census Bureau and it takes the form a series of continuous panels which follow a national sample of households. The first panel began in 1983 but these earlier panels had a short duration. Starting in 1996 the Census Bureau began constructing longer panels with a larger number of households (more than 30,000 although the actual size varies) and those panels are the ones on which we focus on.

The SIPP conducts quarterly interviews which ask interviewees (individuals) to provide information at the monthly frequency on variables such as labor earnings,

demographic characteristics, occupation, etc.... These variables are collected for at most two jobs, but the survey also asks which of those is the primary job for the individual.[2]

Relative to other longitudinal panels such as the Panel Study of Income Dynamics (PSID) and the National Longitudinal Survey of Youth (NLSY97 and NLSY79), advantages of the SIPP are mainly two. The first is the number of respondents. It is considerably larger than the PSID, which surveys about 10,000 households, or the NLSYs, which interview between 9,000 and 13,000. The second advantage is the frequency of interviewing. The SIPP provides a wealth of information at the monthly frequency; the PSID interviews annually and the NLSY97 is now interviewing biennially. It is fortunate that for many individuals in the United States being unemployed or suffering a drop in income is a short-lived experience (usually weeks or months). But given those are the risks on which this study focuses, that fact underscores the importance of having information at higher frequencies. The main disadvantage of the SIPP is the short duration of its panels (at most four years) which prevents us from having entire life-cycle profiles of earnings.

2.2   Labor Income Shocks

The first step in our analysis computes earnings variability at the individual level using a regression approach used extensively in the literature. For example, in Carroll and Samwick (1997). We proceed by estimating a fixed effect model for each industry j in our sample. Given a panel of N individuals for whom we measure earnings (and other variables) over a period of time T, we assume that (log) earnings for individual i in industry j at time t, y_{ijt}, can be modeled as,

$$y_{ijt} = \alpha_{ij} + \beta_j X_{ijt} + u_{ijt} \tag{1}$$

---

[2]An Appendix to the paper contains a step-by-step description of how we choose the sample of individuals to perform the analysis described in this section.

The vector **X** comprises several variables that help predict changes in the level of log-earnings. Specifically, we include age, sex, ethnicity, years of schooling, an occupational dummy, and a seasonal dummy.[3] We first assume that the error term u_{ijt} is distributed i.i.d. $N(0, \sigma_{j,u}^2)$.

We estimate equation (1) by ordinary least squares for all individuals in a given industry. Repeating this procedure for all industries yields estimates $\{\hat{\alpha}_j, \hat{\beta}_j\}_{j=1}^{21}$ and $\widehat{\sigma_{j,u}^2}$.

We present the estimates of the variances of the innovations for each industry in Table 1.

The median of the estimated variances is 0.066 which corresponds those workers who work in the Education sector. The workers who face the least amount of uncertainty are, in order, those who work in Armed Forces, Agriculture and Forestry, Social Services, Mining and Utilities.[4] Workers in Finance, Medical Services, Other Services, Transportation and Hospitals are the industries with the highest levels of income uncertainty. Note that, according to this notion of risk, the Finance sector is more than twice as risky as Agriculture and Forestry. Finally, we test the hypothesis that all the estimated variances are equal and we reject it with a p-value of 5.95×10^{-10} .

2.3 Permanent and Transitory Shocks

We now enrich our empirical analysis by allowing the error term to be decomposed into a permanent component and a transitory component. The reason for distinguishing between the two types of shocks is that they affect the welfare of workers differently. Transitory shocks (e.g. the loss of an important customer for a consultant) are seldom a cause for concern; small levels of savings are usually enough for workers to weather that type of shock successfully. Permanent shocks are, by definition,

---

[3]An alternative interpretation of the seasonal dummy is a periodic change over time in the coefficient α_{ij}.

[4]Regarding Armed Forces, even though it is a low earnings risk sector, could be considered risky using alternative metrics (e.g. injuries and death while in service).

longer-lasting and can be associated with, for instance, depreciation of job-specific human capital or permanent changes in the way an industry operates. Smoothing out the latter type of shocks through a buffer stock of savings is more difficult and permanent changes in consumption are often times required. As the impact on the welfare of individuals is different for the two types of risk, one would expect the premium that workers demand for bearing them differs as well.

We follow Carroll and Samwick (1997) and Low, Meghir, and Pistaferri (2010), among others, by assuming that

$$u_{ijt} = \eta_{ijt} + \omega_{ijt}, \tag{2}$$

where η_{ijt},the transitory component, is distributed i.i.d. $N(0, \sigma_{j,\eta}^2)$, and ω_{ijt}, the permanent component, is a random walk, i.e.

$$\omega_{ijt} = \omega_{ij,t-1} + \epsilon_{ijt} \tag{3}$$

with i.i.d. innovations ϵ_{ijt} that are distributed $N(0, \sigma_{j,\epsilon}^2)$.

By estimating equation (1) we obtain $\{\{\hat{u}_{ijt}\}_{i=1}^{N_j}\}_{t=1}^{T}$.

We estimate the variances of the permanent and transitory components by following the identification procedure proposed in Low, Meghir, and Pistaferri (2010).

Taking first-differences in equation (1) and given the process specified in (2), we have that,

$$\Delta y_{ijt} = \Delta \beta_j X_{ijt} + \Delta \eta_{ijt} + \epsilon_{ijt}. \tag{4}$$

Now define,

$$g_{ijt} = \Delta(y_{ijt} - \beta_j X_{ijt}) = \Delta \eta_{ijt} + \epsilon_{ijt}. \tag{5}$$

To identify the parameters of interest, we compute,

$$E(g_{ijt}^2) = \sigma_{\epsilon_{ij}}^2 + 2\sigma_{\eta_{ij}}^2, \tag{6}$$

and

$$E(g_{ijt}g_{ijt-1}) = -\sigma_{\eta_{ij}}^2 \tag{7}$$

To estimate the variances of the two innovations, we proceed as follows. For an individual i in a given industry j, we estimate $\widehat{E(g_{ijt}^2)}$ and $\widehat{E(g_{ijt}g_{ijt-1})}$ by taking the sample moments. By solving the system we then obtain $\widehat{\sigma_{\epsilon_j}^2}$ and $\widehat{\sigma_{\eta_j}^2}$ by taking averages across individuals of the obtained estimated variances for each individual.

Table 2 shows the estimated variances.

The median of the estimated variances across industries are 0.0141 and 0.0037 for the permanent (column 2, Construction) and transitory shocks (column 4, Medical Services), respectively.

Regarding the variance of the permanent shock the group of relatively safe industries is composed by Armed Forces, Social Services, Utilities, Communication and Government. The most uncertain sectors are Finance, Transportation, Retail Trade, Education and Business Services.

The risky sectors according to the variability of the temporary component are Mining, Agriculture and Forestry, Finance, Government and Other Services. On the other hand, the sectors with the lowest variance of transitory income shocks are Recreation and Entertainment, Armed Forces, Business Services, Personal Services and Construction.

Without exception, the variance of the permanent component is higher than that of the transitory component by a factor of three, roughly.

2.4 Industry Risk and Risk Premia

Having estimated measures of risk for our group of industries, we are now ready to test the hypothesis that, across industries, the level of risk and the average level of earnings are positively correlated. Our claim, however, should be understood to be a *ceteris paribus* claim. That is, everything else constant, a higher level of risk should be associated with a higher level of earnings. Of course, not everything else is constant across industries. Industries differ along many dimensions that may affect average earnings independently of their level of risk. The facts that, for instance, better educated workers earn more or that larger firms pay higher wages, are well-established. This should lead one to suspect that the mix of workers or firms in a given industry are important determinants of its average level of earnings. From the econometric point of view, to account for this industry heterogeneity, we proceed in two ways. We first compute industry averages (that is, averages across individuals who work in a given industry) of variables we deem relevant in determining average earnings. More specifically, we establish the (conditional) sign of the relationship between average earnings and industry risk by estimating the following regression equation:

$$y = \gamma + \boldsymbol{\theta} \mathbf{Z} + \nu \tag{8}$$

where y is a vector whose j^{th} element is the average (log) earnings for individuals in industry j, and $\mathbf{Z}$ is a matrix of regressors. The jth row of $\mathbf{Z}$ has six elements. the average age, the average age squared, and the average level of education of all individuals working in industry j, the fraction of females in industry j, and the industry j variance of income shocks estimated above (see Table 1). Since the number of industries in our sample is 21, y is a column vector of dimension 21, and $\mathbf{Z}$ is of dimension 21×6.[5] Finally, γ is a vector of intercepts and ν a vector of residuals. We assume that

---

[5]We also estimate the equation 8 when the matrix $\mathbf{Z}$ includes the variances of the temporary and permanent component of income shocks. In that case, the matrix $\mathbf{Z}$ has 7 columns because we do not

the error term v_j is distributed i.i.d. $N(0, \sigma_v^2)$.

Column 2 of Table 3 shows the results of estimating equation 8. It presents the estimated values for the coefficients and their probabilities of being less than zero computed by bootstrap. All the coefficients are significant and have the expected signs. Workers' age and education levels are positively related to mean earnings and females labor earnings are on average lower than males. Our focus is on the sign and magnitude of the coefficient associated with risk, σ_ϵ^2. The table shows that this coefficient points to a strong and positive association between uncertainty and earnings. More importantly, the probability that this coefficient is less than zero is 0.0002. Note that the value of the coefficient associated with uncertainty implies that increasing the variance from 4.9% to 9% (we go from Agriculture and Forestry to Finance), increases the mean level of earnings by 28%. According to this econometric model, this result would consistent with the existence of a risk premium in the labor market.

Alternatively, in order to document the relationship between our measure of industry risk and mean earnings we also estimate individual's earnings net of its main observed characteristics: age, the square of age, education and gender. We proceed by estimating a pooled regression by OLS that allows us to obtain estimates for the net log earnings of each individual in our sample. Specifically, we estimate the following regression

$$y_{ijt} = \gamma_0 + \gamma X_{ijt} + \lambda_{ijt} \tag{9}$$

where the vector of coefficients γ represents the effect of the observed characteristics (age, education, square of age, and gender). These coefficients are common in the cross-section and across time. When then obtain, for each individual and at a point in time, log earnings net of observed characteristics by computing $\tilde{y}_{ijt} = y_{ijt} - \hat{\gamma} X_{ijt}$. By averaging across time and across individuals in each industry we obtain the mean

---

include the vector of overall variances.

of the net earnings by industry. Specifically, for a given industry j, we compute

$$\tilde{y}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \tilde{y}_{ij}, \tag{10}$$

where

$$\tilde{y}_{ij} = \frac{1}{T} \sum_{t=1}^{T} \tilde{y}_{ijt}. \tag{11}$$

We now use this estimated values of net earnings to estimate the following regression equation:

$$\tilde{y}_j = \alpha_0 + \alpha_1 \sigma_j^2 + \nu_{\tilde{y},j} \tag{12}$$

Again, since the number of industries in our sample is 21, each variable of the regression has 21 observations. α_0 is an intercept, α_1 is the coefficient that represents the effect of our measure of risk into the mean of net earnings and $\nu_{\tilde{y}}$ is the residual which is assumed to be i.i.d. $N(0, \sigma_{\nu_{\tilde{y}}}^2)$.

Column 3 of Table 3 shows the estimated values for the coefficients and the probabilities of that they are less than zero computed by bootstrap. Note that, somehow confirming the existence of a risk premium, the coefficient associated with risk, α_1, is positive which a probability of being less than zero equal to 0.0104. Note, however, that according to this approach, the value of the coefficient associated with uncertainty implies that increasing the variance from 4.9% to 9% (we go from Agriculture and Forestry to Finance), increases the mean level of earnings by 7%.

As mentioned above, the decisions of workers could greatly differ upon the nature of the shock, so it is important to consider the decomposition of the process into a temporary and a permanent component.

For these reasons, we estimate equation (8) using as regressors the variances of the two components, permanent and transitory, instead of just one variance that reflects

overall uncertainty. The second column of Table 4 presents the results.

All the coefficients are significant and have expected signs. Excepting the coefficients associated with the variances, their magnitudes are close to the ones found before. Turning now to the coefficients associated with uncertainty, we first observe that they are strongly positive and with probabilities of being less than zero of 18% and 16% for the permanent and transitory shocks, respectively. The estimated value for coefficient associated with the variance of the permanent shock to earnings is 9.3. Therefore, according to this result, going from the Social Service industry (the second safest) to Finance (the riskiest industry) implies an increase in mean earnings of 7%. Regarding the transitory shock, the value estimated for its associated coefficient is 20.3 and so, according to this result, moving from Recreation and Entertainment (the safest sector) to Mining (the riskiest sector) implies an increase of in mean earnings of 10%. As in the case with the total variance of earnings, we present the results by using our alternative specification to document the relationship between mean earnings and uncertainty. It is depicted in column 3 of Table 4. The estimation results point to a strong and positive relationship between mean earnings and the estimated variances, being the values of estimated coefficients for the permanent and transitory shocks to earnings are 6.9 and 16.6, respectively, with very low probabilities of being less than zero: 1.5% and 7.7%, respectively. According to these results, considering the permanent shock to earnings, moving from Social Services to Finance imply a an increases in mean earnings of 5%. If we look at the transitory shock to earnings, moving from Recreation and Entertainment to Mining implies a compensation in mean earnings of 8%.

The data and approach we use to link labor earnings and their uncertainty, yield estimates which appear to be consistent with a compensating differential for risk in the labor market. But one ought to be cautious. The distribution of average earnings across individuals in an industry is an endogenous outcome resulting from individuals' decisions of where to supply their labor services. The level of earnings risk is

certainly something individuals consider when making that choice. But their comparative advantage, in other words, their relatively higher productivity in a certain sector, consequence of a set of individual characteristics, plays a role as well. Some of that comparative advantage originates from being, for instance, a female or a college-grad, characteristics which we have accounted for to some degree. Much of the advantage, however, originates from unobserved characteristics which are, obviously, difficult to control for. To help us decompose how much of the estimated earnings risk *premium* is due to a compensating differential and how much to self-selection, the next section describes a quantitative framework in which comparative advantage and individuals' industry choice are explicitly taken into account.

3   The Model

Our artificial economy is populated by a mass of risk-averse individuals of total measure equal to one. Time is discrete and individuals live for S periods which correspond to their working lives. In other words, they are born into a labor market and never retire. Each individual is endowed with one unit of time each period that is supplied inelastically in the labor market. When an individual reaches time $S + 1$ and dies, another age 0 individual replaces her, so the total population is constant. At the beginning of their lifetimes, individuals choose to work in one of J mutually exclusive job opportunities indexed by j, which we interpret as sectors or industries. At birth, prior to the industry choice, each individual draws a value for sector-specific skill or ability from a given distribution specified below. These skills enter directly the productivity and hence earnings of an individual and therefore determine an individual's comparative advantage for, say, working in Finance and not in Agriculture. As these skills are random draws, we are silent about their origin but they could loosely be interpreted as innate abilities or human capital acquired before entering the labor market. Finally, the values for the sector-specific skills do neither grow nor decrease

over an individual's lifetime.

In addition, once working for an industry (from which they cannot move), individuals are subject to idiosyncratic shocks to their labor income. The process driving those shocks differs from industry to industry. In particular, workers in some industries experience a higher variability of earnings relative to workers in other industries. If workers are risk-averse, riskier industries look less attractive than safer industries.

When an individual is born in period 0 (i.e. when he enters the labor market), his problem is to choose one of the J mutually exclusive career alternatives in order to maximize the expected discounted value of her life-time utility:

$$\mathbb{E}_0 \left\{ \left[\sum_{s=1}^{S} \beta^{s-1} \sum_{j} \mathbf{1}_j u(c_{s,j}), \right] \middle| \Omega_{i,0} \right\},$$

where $\mathbf{1}_j$ is an indicator function with value $\mathbf{1}_j = 1$ if the individual chooses to work in industry j and 0 otherwise. The function $u(c_{s,j})$ denotes the individual's per-period utility derived from choosing $j \in J$ and consuming $c_{s,j}$; we assume $u_c > 0$ and $u_{cc} < 0$. The only source of uncertainty are shocks to labor earnings and we describe those in detail below. For now it suffices to say that expectations in (3) are taken with respect to the distribution of those shocks. The vector $\Omega_{i,0}$ represents the information set of an individual i at time 0 and it is formally the vector

$$\Omega_{i,0} = \{\theta_{i,1}, \ldots, \theta_{i,J}\}$$

where the logarithm of each value $\theta_{i,j}$ is drawn from an industry-specific distribution $N(\mu_{\theta_j}, \sigma^2_{\theta_j})$. Each period, by inelastically supplying one unit of time to sector j, each individual receives labor earnings, $w_j \theta_{i,j} e^{v_{i,j}}$, comprised of a sector specific competitive wage rate (w_j), individual-specific sectoral pre-labor-market skills ($\theta_{i,j}$) and, an individual-specific but time-varying labor productivity shock ($v_{i,j}$). Once the individual makes her sectoral choice, only the θ corresponding to the chosen industry affects her lifetime labor earnings.

18

For an individual of age s, the time-varying component of earnings is the addition of two orthogonal stochastic components,

$$v_{s,j} = \eta_{s,j} + \omega_{s,j},\tag{13}$$

where η_j is an i.i.d. transitory shock to log earnings distributed as $N(-\frac{1}{2}\frac{1}{\sigma_{j,\eta}^2}, \sigma_{j,\eta}^2)$ and $\omega_{s+1,j}$ is the permanent component that follows a random walk: $\omega_{s+1,j} = \omega_{s,j} + \epsilon_{s,j}$ with ϵ_j being $N(-\frac{1}{2}\frac{1}{\sigma_{j,\epsilon}^2}, \sigma_{j,\epsilon}^2)$ i.i.d innovations.[6] By subscripting the variance by j, we make clear that the nature of the shock process is industry-specific. Despite the inability of consumers to change industry in midlife, we allow them to partially insure against labor income shocks by saving in a one period risk-free non-contingent bond with an exogenous interest rate equal to r.

Individual's Decision Problem Suppose an individual has chosen an industry in which to supply labor and begun his working life. Every period, optimization for this individual entails choosing how much to consume and the amount of savings or quantity of one-period bonds to purchase.[7] The variables relevant to these decisions are the level of wealth, the age of the individual, and the following components of income: the time-varying component, the ability level for the chosen industry. Let the vector of individual state variables be denoted as $x = (b, \omega, \eta, s, \theta_j)$, where j is the chosen industry. Denote by $\Psi_j(x)$ the industry j workers' distribution across assets, age, income, and abilities.[8] It is an aggregate state variable since it determines the wage rate in industry j. Only the marginal distribution of age is identical across all

---

[6]In the quantitative application we approximate the random walk by a highly persistent process. It is close to a unit root but stationary nevertheless.

[7]Our model is set apart from others in the literature in the optimal choice of an industry and its general-equilibrium implications. Once the individual has chosen an industry, the optimization problem of the consumer is essentially identical to many examples in a literature analyzing heterogeneous agents economies. The only departure is that we allow two different shocks with different statistical properties. This departure allows us to analyze the impact of transitory and permanent risk on industry choice.

[8]The distribution is subscripted by j because workers, facing different income shocks and self-selecting into industries based on different abilities levels, will choose different levels of assets.

industries. For convenience denote by $\mathcal{S} = \mathcal{S}_B \times \mathcal{S}_{E_\eta} \times \mathcal{S}_{E_\omega} \times \mathcal{S}_\theta \cup \{1, \ldots, S\}$ the state space of the vector of state variables x.[9] It is convenient to write the problem recursively, and we denote the remaining lifetime utility for an age-$\bar{s}$ individual working in industry j by the $V_j(x|s = \bar{s})$. It is defined by,

$$V_j(x|s = \bar{s}, \Psi_j) = \max_{c,b'} \left\{ u(c) + \beta E V_j(x'|s = \bar{s} + 1, \Psi_j') \right\} \quad \text{if} \quad 1 \le s \le S \quad \text{and} \quad 0 \quad \text{otherwise},$$

subject to,

$$c + b' = w_j(\Psi_j)\theta_j e^\eta e^\omega + b(1 + r) \tag{14}$$

and,

$$b \ge \underline{b}, \;\; b_0 = 0, \;\; b_{S+1} \ge 0 \tag{15}$$

We follow relatively standard notation when we denote by x' the values of x one period ahead. Equation (14) is a standard flow budget constraint that equates consumption plus savings to total earnings from capital holdings $b(1 + r)$, and earnings from supplied labor $w_j(\Psi_j)\theta_j e^\nu$. In addition to this budget constraint, individuals face a borrowing constraint that restricts the lower bound on asset holdings. Also, individuals are born with zero wealth ($b_0 = 0$) and they face a non-negativity constraint in their savings at the time of death ($b_{S+1} \ge 0$).

At birth, the individual chooses from a set of J industries the one that yields the highest utility.

$$j^* = argmax \left\{ W_1, \ldots, W_J \right\} \tag{16}$$

where W_{j^*} for an individual i is defined as

---

[9]In general, the joint state space should have a subscript j. In our particular model, the borrowing constraint and longevity are identical across industries. Income innovations and abilities are all real numbers. Hence we can omit the subscript j.

$$W_{j^*} = \mathbb{E}_0 \left\{ V_{j^*}(x|s=1)|\Omega_{i,0} \right\}. \tag{17}$$

When choosing an industry, $\Omega_{i,0}$ - the vector of abilities drawn at birth - is in a person's information set, thus appearing to the right of the conditioning sign. The individual knows as well the statistical properties of shocks that he will experience in each industry. As a result, and although not explicitly written, it should be understood that the expectation is taken with respect to a different distribution if the worker computes W_j for $j \neq j^*$. The choice in (16) induces an endogenous distribution of workers across industries. Let μ_j denote the mass of workers in industry j; $\sum_{k=1}^{J} \mu_k = 1$.

Firms   One can picture our model economy as a small open economy containing a set of islands with each of the islands representing an industry. In each industry, a consumption good is produced according to the following industry-level technology:

$$Y_j = N_j^{\alpha_j}, \tag{18}$$

where Y_j is the output of sector j, N_j represents the labor input of that sector measured in efficiency units,[10] and α is the share of labor in output (with $\alpha < 1$). Firms are owned by foreigners who operate it, pay wages, and enjoy profits. We do not consider any kind of inter-industry trade in goods, so the reader can assume that goods produced across islands are identical. [11].

Equilibrium   We can now define a stationary competitive equilibrium which consists of a set of industry wages $\{w_j\}_{j=1}^{J}$, industry populations (or masses) $\{\mu_j\}_{j=1}^{J}$,

---

[10]The measure of efficiency takes into account both the time-varying productivity component and the industry-specific abilities.

[11]Alternatively, one can picture J different goods and assume that an individual working in industry j obtains utility from consuming the good produced in that industry only, and not those from other islands

industry-specific distributions $\{\Psi_j(x)\}_{j=1}^{J}$, industry-level efficiency-weighted employment levels $\{N_j\}_{j=1}^{J}$, and industry-specific decision rules $\{b_j'(x), c_j(x)\}_{j=1}^{J}$ and associated value functions $\{V_j(x)\}_{j=1}^{J}$, which satisfy the following conditions:

1. Given wages, the industry-specific decision rules $\{b_j'(x), c_j(x)\}_{j=1}^{J}$ solve the optimization problem (3) yielding value functions $\{V_j(x)\}_{j=1}^{J}$.

2. The set of industry-specific populations $\{\mu_j\}_{j=1}^{J}$ and the distributions of abilities across industries are consistent with the optimal industry choice (16). For any given industry j, its population satisfies $\mu_j = Prob(W_j > W_{-j})$ where we define the vector W_{-j} to be equal to $\{W_1, \ldots, W_{j-1}, W_{j+1}, \ldots, W_J\}$. The cumulative distribution of θ_j in a given industry j is defined by,

$$G_j(\theta_{0,j}) = \frac{\int_{\Theta_{-j}} \int_{\{\theta_j \in \Theta_j : \theta_j < \theta_{0,j}\}} \chi_{\{\theta_j : W_j > W_{-j} | \theta_{-j}\}} dF(\theta_j) dF(\theta_{-j})}{\int_{\Theta_{-j}} \int_{\Theta_j} \chi_{\{\theta_j : W_j > W_{-j} | \theta_{-j}\}} dF(\theta_j) dF(\theta_{-j})} = \int_{\mathcal{S}} \chi_{\{\theta_j \leq \theta_{0,j}\}} d\Psi_j(x)$$

where Θ_j is the support of θ_j and Θ_{-j} is the support of θ_{-j} and $\chi_{\{\theta_j : W_j > W_{-j}\}}$ is and indicator function that takes the value 1 when an individual with ability θ_j chooses industry j. Finally, $F(\theta_j)$ is the c.d.f of θ_j before sorting of agents.

3. Wages in industry j are equal to the marginal product of a marginal unit of average efficiency in that industry:

$$w_j = \alpha_j N_j^{\alpha_j - 1},$$

where the industry-level measures of employment are defined as
$N_j = \mu_j \int_{\mathcal{S}} \theta_j e^{\eta} e^{\omega} d\Psi_j(x)$.

4. For an individual in an industry j, the decision rules $b_j'(x)$ and $c_j(x)$ solve the individuals' dynamic problem (3), and $V_j(x)$ is the associated value function.

22

5. In a given industry j, $\Psi_j(x)$ is the stationary distribution associated with the transition function implied by the optimal decision rule $b'_j(x)$ and the law of motion for the exogenous shocks.

6. At the industry level, the following resource constraint is satisfied:

$$w_j N_j = \int_{\mathcal{S}} \{c_j(x) + b'_j(x) - b_j(x)(1+r)\} d\Psi_j(x)$$

4 Quantitative Analysis

This section presents the quantitative analysis. For this purpose, we use the theoretical model developed in the previous section which is computed and calibrated to mimic the US economy. Besides the standard complexities associated with computing standard life cycle economies there is another layer of difficulty in this particular model that has to do with the existence of the pre-labor market skills or abilities distributions. The main reason has to do with computing values functions for each possible combination of the abilities draws for each simulated individual that lives in our model economy. Even though the presence of these variables dramatically enrich our analysis it restrict the number of industries for which it is computationally feasible to solve the model. We therefore, we restrict our quantitative analysis to 4 industries of the US economy: Agriculture, Manufacturing, Services and Public Sector, which are the result of aggregating the 21 industries detailed above. In Table 5 we present mean of net earnings the variance of the permanent and transitory shocks for these 4 aggregate sectors. Note also that even though we have aggregated the 21 industries, the strong and positive relation between mean earnings and the variances of the permanent and transitory shock is preserved as it is reflected in the correlation coefficient which is 0.88 and 0.69 for the permanent and transitory shock, respectively. In addition, if we regress the mean net earnings on a constant and both variances we get that the coefficient associated with the variance of the permanent shock is 8.5 and

with the variance of the transitory shock is 8.4. This implies that, considering the permanent shock to earnings, moving from the Public Sector (the safest) to Manufacturing (the riskiest) implies an increase in net earnings of 3.5%. If we consider the transitory shock to earnings, moving from the Public Sector (which is again the safest) to Agriculture (the riskiest) implies an increase in mean earnings of 2%. We now turn to parameterize the model economy.

4.1 Parameter Values

We start by setting the model period equal to a quarter, and the total lifetime for an individual to be 120 periods. These two values correspond to 30-year employment histories. We exogenously set the interest rate to be 5% and β to be 0.96. We restrict preferences to be of the constant relative risk aversion class with coefficient or risk aversion equal to 2. In our benchmark case we start by setting $\underline{b} \geq 0$ In addition, by using data of the Compensation of Employees and GDP at the industry level from the National Income and Product Accounts for the period 1990 through 2009 we set the labor share os Agriculture equal to 0.30, the Manufacturing labor Share equal to 0.63, Services equal to 0.51 and the Public Sector equal to 0.85.

One of the driving forces of a non-degenerate wage distribution across industries is an industry-specific level of risk. As a measure of this risk, we use the estimates for the variances of the two components of income we estimate from SIPP in Section 2 and we aggregate to the 4 industries we we focus in this section. Hence, we set J, the total number of industries, to be 4 and we feed the model with the estimated values of the variances of both the permanent and transitory shocks depicted in the fourth and fifth column of Table 5.

Finally, it still remains to parameterize the distributions of pre-labor market skills or abilities, i.e. to find values for 8 parameters: $\{\mu_{j,\theta}, \sigma^2_{j,\theta}\}^{j=J}_{j=1}$. For this purpose we pick values for these parameters so that the model delivers the mean and standard deviation of the net earnings for each of the 4 industries (column 2 and 3 of Table

24

5). The use of net earnings or, in other words, earnings net of observables such us education or age, is justified by the fact that we in our model economy all individuals are equal in terms of sex and education , and there is no age-specific productivity (i.e. all the observables we control for in equation 9). The resulting parameter values are shown in Table 6.

4.2   Results

The experiment consists of solving the model for the set of parameter values just described. Since the parameter of the distributions of the pre-market labor skills are picked so that the model exactly replicates the mean of net earnings of each of the 4 industries, it will also replicates the relationship between the mean earnings and the variances of the transitory and permanent shocks to earnings. Therefore, this empirical relationship cannot be used as an independent test for the model. Interestingly, the model has testable implications with regard to the sorting of workers into the 4 sectors of the economy. Specifically, the model predicts the mass of individual that work in each of the sectors in equilibrium. Table 7 shows the model predictions and it data counterpart. The correlation between the share of individuals in each sector in the model and in the data is 0.98 being particularly good the model predictions regarding the mass of workers that work in Manufacturing and the Public Sector.

As it was highlighted above, the documented strong and positive relationship between mean earnings and their variances, can be interpreted as evidence in favor of a existence of a pure risk premium in the US labor market. However, the presence of individual specific pre-labor market skills or individual comparative advantages affect the sorting of individuals into the sectors of the economy. This this is unobserved for the econometrician and so it can greatly change the interpretation of the results since we could be mistakenly assigning all of the observed differences in mean earnings to compensation for risk while in fact they are compensation for the skills of the individuals. Fortunately, our theory is rich enough so that we can proceed to perform

a counterfactual experiment in which we shut down all the differences across individuals and across sectors in the pre-labor market skills. This counterfactual exercise allows us to decompose the observed relationship between mean earnings and the variance of shocks into pure compensation for risk and compensation for unobserved skills. Specifically, we solve the model again but instead of picking the parameters of the skills distributions to match the moments of the net earnings distributions, we just endow all individuals with the same skill level for each of the 4 industries that compose our economy.

By doing this our model economy dramatically changes its properties since the individual sectoral choice is mainly affected by the cross-sectoral differences in the volatility of earnings. The individuals are now ex-ante homogenous but still subject to idiosyncratic shocks to their labor income; hence they are *ex-post* heterogeneous. It is still the case that workers in some industries experience a higher variability of earnings relative to workers in other industries. If workers are risk-averse, riskier industries look less attractive than safer industries. Everything else equal, all workers would concentrate in the safest industry, with all but one industry having no workers and hence no output. In our environment, this can not be an equilibrium because we assume technology, at the industry level, displays decreasing returns to scale. As a result, the more workers populate an industry, the lower the wages, and vice-versa. The resulting equilibrium features relatively safe industries with low wages and a large mass of workers. Riskier industries display the opposite characteristics.

In Table 8 we present the mean net earnings and share of workers predicted by the model in this counterfactual exercise. Note that the share of workers in each industry is at odds with the data (correlation of -0.22). Regarding the mean of net earnings the model prediction is highly correlated with data and also highly correlated with the variance of the permanent shock (correlation coefficient of 0.97). However, it is less correlated with the variance of the transitory shock when compared with the benchmark model (correlation coefficient of 0.39 versus 0.58). Moreover, when we regress

26

the predicted mean of net earnings with respect to the variance of the permanent and transitory shocks we find that the coefficient associated with the permanent shock is 15.1 which points to a strong and positive association between mean earnings and the permanent shock to earnings. Moreover, according to this result imply that the increase in mean net earnings from moving from the Public Sector to Manufacturing would be around 6.2% while in the data it is actually 3.5%.

However, the results regarding the transitory shock are totally different since coefficient corresponding to the transitory shock reduces to only 0.9. Recall that in the case of the benchmark model, this coefficient was 8.4 which means that what appears to be compensation for the risk associated with transitory shocks is actually compensation for unobserved comparative advantages of individuals that have endogenously sorted into the different sectors of the economy.

5 Concluding Remarks

Using data from the Survey of Income and Program Participation (SIPP) one finds that the level of risk and the mean level of (log) earnings are positively related. Irrespective of whether risk is of a transitory or permanent nature, the positive association is clear. The question we ask in this paper is whether that estimated correlation is a pure compensating differential or masks other factors, for instance self-selection due to comparative advantage. A standard Roy model places a lot of weight when choosing an industry in a worker's comparative advantage. However, if the latter is driven by unobserved characteristics, it is not possible to tell how much of the observed equilibrium relation between risk and earnings is due to comparative advantage moving into particular industries and how much to a compensating differential (individuals moving away from risky industries). The model we construct assigns a zero compensating differential component to the relationship between the variance of transitory shocks and (log) earnings. On the other hand, the higher earnings observed in industries

with higher permanent risk reflect a pure compensating differential.

References

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

D'Erasmo, P. N., and H. J. Moscoso-Boedo (2011): "Intangibles and Endogenous Firm Volatility over the Business Cycle," Virginia Economics Online Papers 400, University of Virginia, Department of Economics.

Dickens, W. T., and L. F. Katz (1987): "Interindustry Wage Differences and Industry Characteristics," NBER Working Papers 2014, National Bureau of Economic Research, Inc.

Dillon, E. W. (2012): "Risk and Return Tradeoffs in Lifetime Earnings," Discussion paper, University of Michigan, Department of Economics.

Feinberg, R. M. (1981): "Earnings-Risk as a Compensating Differential," *Southern Economic Journal*, 48(1), 156–163.

Fields, J., and E. N. Wolff (1995): "Interindustry wage differentials and the gender wage gap," *Industrial and Labor Relations Review*, 49(1), 105–120.

Guvenen, F. (2007): "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *American Economic Review*, 97(3), 687–712.

——— (2009): "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, 12(1), 58–79.

Heathcote, J., K. Storesletten, and G. L. Violante (2008): "Insurance and opportunities: A welfare analysis of labor market risk," *Journal of Monetary Economics*, 55(3), 501–525.

——— (2009): "Quantitative Macroeconomics with Heterogeneous Households," *Annual Review of Economics*, 1(1), 319–354.

Hoffmann, F. (2010): "An Empirical Model of Life-Cycle Earnings and Mobility Dynamics," Discussion paper, University of British Columbia, Department of Economics.

Katz, L. F., and L. H. Summers (1989): "Can Interindustry Wage Differentials Justify Strategic Trade Policy?," in *Trade Policies for International Competitiveness*, NBER Chapters, pp. 85–124. National Bureau of Economic Research, Inc.

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

Kopecky, K., and R. Suen (2010): "Finite State Markov-chain Approximations to Highly Persistent Processes," *Review of Economic Dynamics*, 13(3), 701–714.

Krueger, A. B., and L. H. Summers (1988): "Efficiency Wages and the Inter-industry Wage Structure," *Econometrica*, 56(2), 259–93.

Kuznets, S., and M. Friedman (1939): "Incomes from Independent Professional Practice, 1929-1936," in *Incomes from Independent Professional Practice, 1929-1936*, NBER Chapters. National Bureau of Economic Research, Inc.

Leigh, J. P. (1983): "Job Choice across Industries when Earnings Are Uncertain," *Quarterly Review of Economics and Business*, 23(3), 54–69.

Low, H., C. Meghir, and L. Pistaferri (2010): "Wage Risk and Employment Risk over the Life Cycle," *American Economic Review*, 100(4), 1432–67.

McGoldrick, K. (1995): "Do Women Receive Compensating Wages for Earnings Uncertainty?," *Southern Economic Journal*, 62(1), 210–222.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

Siow, A. (1984): "Occupational Choice under Uncertainty," *Econometrica*, 52(3), 631–45.

Storesletten, K., C. I. Telmer, and A. Yaron (2004a): "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, 51(3), 609–633.

Storesletten, K., C. I. Telmer, and A. Yaron (2004b): "Cyclical Dynamics in Idiosyncratic Labor Market Risk," *Journal of Political Economy*, 112(3), 695–717.

Thaler, R. H. (1989): "Interindustry Wage Differentials," *Journal of Economic Perspectives*, 3(2), 181–93.

Appendix A: Data

We use three Surveys from the Survey of Income and Program Participation (SIPP): 1996, 2001 and 2004. Cleaning of the data until we reach the final sample is identical across the three surveys. We use the public data files from SIPP available at http://www.ceprDATA.org, maintained by the CEPR (Center for Economic and Policy Research) in Washington, D.C. For each Survey, we perform the following steps:

1. Eliminate individuals who simultaneously report missing earnings but positive hours worked.

2. Eliminate individuals who report working in two different industries or those who do not report their industry (self-employed).

3. Eliminate individuals who report being out of the labor force.

4. Eliminate secondary jobs (i.e. we focus on the primary job of the individual).

5. Restrict analysis to individuals older than 22 but younger than 66.

6. Restrict analysis to individuals who are married.

7. Eliminate individuals who do not report complete samples.

We redefine earnings to be unemployment insurance if an individual reports zero hours worked and reports being unemployed. For those individuals who are not unemployed we also eliminated those with very low earnings (less than 600 1996 dollars per month). [To be completed: Report number of observations lost in each step for all three surveys.]

Appendix B: Model Computation

This appendix describes the solution algorithm employed to solve the model described in Section 3.

1. The first step is to discretize the distributions for the selection parameters (i.e. the industry-specific skills or productivities). Recall from the model description that we assume normality for the selection parameters: $\theta_j \sim N(\mu_{\theta_j}, \sigma_{\theta_j}^2)$. We construct an equi-spaced grid of length N_R for the support of each distribution $G_R^j = \left\{\hat{\theta}_j^1, \ldots, \hat{\theta}_j^{N_R}\right\}$, assuming $\hat{\theta}_j^1 = \mu_{\theta_j} - w_R \sigma_{\theta_j}$ and $\hat{\theta}_j^{N_R} = \mu_{\theta_j} + w_R \sigma_{\theta_j}$ and setting $w_R = 4$ and $N_R = 10$.

2. Guess a distribution of masses $\left\{\mu_j\right\}_{j=1}^J$ and efficiency levels $\left\{\theta_j^*\right\}_{j=1}^J$ for each of the industries. This yields aggregate employment levels (in efficiency units) for each of the four industries $\left\{N_j\right\}_{j=1}^J$. From our technology assumption, the wage rate in each industry is equal to the marginal product of a unit of average efficiency.

3. Given a set of wages $\left\{w_j\right\}_{j=1}^J$, we compute the individual's life-cycle problem for each industry and for each value of the industry-specific ability. To solve for the value and policy functions we discretize the space of bond holdings. Current, not future, bond-holdings are required to lie on a grid $G_B = \{b_1, \ldots, b_{N_B}\}$, with $N_B = 100$, and we use linear interpolation to approximate future value functions. We discretize the values of the persistent and temporary shocks, ω and η.[12]. We use $N_P = 5$ points to approximate the persistent component and $N_T = 2$ to approximate the i.i.d component.[13] The construction of the grid and the computation of the transition matrix for the persistent component follow the procedure outlined in Kopecky and Suen (2010).

---

[12]For computational reasons, we approximate the random walk with a very persistent process; an autocorrelation of 0.999.

[13]Two grid points for the iid component matches the mean and standard deviation, the only two moments that are relevant if the process follows an iid normal distribution.

4. The previous step yields a set of N_R expected value functions for each industry j conditional on a given level of ability,

$$\left\{ \left\{ \mathbb{V}_j^k = \int V_j(x|\theta_j = \hat{\theta}_j^k) d\Psi_j(x) \right\}_{k=1}^{N_R} \right\}_{j=1}^J.$$

To increase the degree of accuracy when updating the aggregate efficiency units of labor, we construct a second grid of length $N_{\tilde{R}} > N_R$: $G_{\tilde{R}j} = \left\{ \tilde{\theta}_j^1, \ldots, \tilde{\theta}_j^{N_{\tilde{R}}} \right\}$. The two endpoints of the set $G_{\tilde{R}j}$ equal the two endpoints of G_R^j. For any point in $G_{\tilde{R}j}$ not in G_R^j, we linearly interpolate the value functions of its two nearest neighbors in G_R^j. Denote the value function given an ability $\tilde{\theta}_j^k$ by $\tilde{\mathbb{V}}_j^k$. Finally, the results shown in the main body of the paper assume $N_{\tilde{R}} = 65$.

5. Completing the previous step yields, four each industry, a set of three vectors: a grid $G_{\tilde{R}j} = \left\{ \tilde{\theta}_j^1, \ldots, \tilde{\theta}_j^{N_{\tilde{R}}} \right\}$, a vector of associated probabilities for each element in $G_{\tilde{R}j}$, $\left\{ \tilde{p}_j^1, \ldots, \tilde{p}_j^{N_{\tilde{R}}} \right\}$,[14] and a vector of associated value functions $\left\{ \left\{ \tilde{\mathbb{V}}_j^k \right\}_{k=1}^{N_{\tilde{R}}} \right\}_{j=1}^J$. Denote by $K^* = (N_{\tilde{R}})^J$ the set of all possible combinations of the J ability parameters. In other words there are K^* possible values for the vector $\left\{ \tilde{\theta}_1^{i_1}, \ldots, \tilde{\theta}_J^{i_J} \right\}_{i_1,\ldots,i_J=1}^{N_{\tilde{R}}}$. Let $T : \{1, \ldots, N_{\tilde{R}}\}^J \rightarrow \{1, \ldots, K^*\}$ be a mapping that yields a value in the set $\{1, \ldots, K^*\}$ given a J-tuple $\{i_1, \ldots, i_J\}$, where each element $i_1, \ldots, i_J$ belongs to the set $\{1, \ldots, N_{\tilde{R}}\}$. The number $p_{T(i_1,\ldots,i_J)} = p_1^{i_1} \times \ldots \times p_J^{i_J}$ is the probability attached to the event an individual draws the vector $\theta_1^{i_1}, \ldots, \theta_J^{i_J}$. There are K^* such probabilities and $\sum_{k=1}^{K^*} p_k = 1$. For each J-tuple $\{i_1, \ldots, i_J\}$ there is also a set of value functions $\left\{ \tilde{\mathbb{V}}_1^{i_1}, \ldots, \tilde{\mathbb{V}}_J^{i_J} \right\}$, and an associated index $j^* = argmax \left\{ \tilde{\mathbb{V}}_1^{i_1}, \ldots, \tilde{\mathbb{V}}_J^{i_J} \right\}$ that represents the optimal indus-

---

[14]Since we discretize the state-space for the ability distribution, a given probability is computed as

$$\tilde{p}_j^i = \frac{\phi(\tilde{\theta}_j^i, \mu_{\theta_j}, \sigma_{\theta_j}^2)}{\sum_{k=1}^{N_{\tilde{R}}} \phi(\tilde{\theta}_j^k, \mu_{\theta_j}, \sigma_{\theta_j}^2)},$$

where $\phi(\theta, \mu, \sigma^2)$ is the density of a normally-distributed random variable with mean μ and variance σ^2 evaluated at θ.

try choice for that particular vector of industry-specific skills.

6. Once we have computed the optimal industry j^* for each combination of skill-specific vectors, we are ready to update the guesses for the industry populations and the average efficiencies in each industry. The new mass for industry j, μ_j is computed as:

$$\mu_j = \sum_{k=1}^{K^*} \chi_{\{j^*=j\}} p_k,$$

where χ_A is an indicator variable that takes the value 1 if the event A occurs and zero otherwise. In other words, we sum over all probabilities associated with the event "industry j is the optimal choice". To update the average ability value we proceed analogously by computing,

$$\theta_j^* = \sum_{k=1}^{K^*} \theta_j^{\{i_j \in \{1,...,N_{\tilde{R}}\}:k=T(i_1,...,i_J)\}} \chi_{\{j^*=j\}} (p_k) / \mu_j.$$

The previous expression computes the average ability re-weighting the probabilities to constrain them to sum to 1 (hence the presence of μ_j dividing each of the probabilities). These two quantities can then be used to update wages, and hence compute new value functions, repeating the above steps until the maximum of the absolute values between the guessed efficiencies and the newly computed efficiencies,and the absolute value of the guessed masses and the newly computed value is less than 10^{-4},

$$\max \left\{ \left\{ |\mu_j^{(i)} - \mu_j^{(i-1)}|, |\theta^*{}_j^{(i)} - \theta^*{}_j^{(i-1)}| \right\}_{j=1}^{J} \right\} < 10^{-4},$$

where $i - 1$ and i are two consecutive iterations.

Appendix C: Hours and Occupations

To show that the results are robust to using earnings per hour (as opposed to total earnings), we perform the same empirical analysis as that of the main body of the text. We use only the 1996 sample as the quality of hours seems to be better than in the other two surveys. Except transforming earnings into per-hour units, the steps are the same as those described above. Table 9 shows the equivalent to Table 3 after having re-estimated equations (1) and (8) with earnings per-hour on the left-hand side.

Tables

Table 1: Variance of Earnings by Industry

	σ_j^2	Ranking
1 Agriculture and Forestry	0.0446	2
	(0.0041)	
2 Mining	0.0545	4
	(0.0063)	
3 Construction	0.0654	9
	(0.0030)	
4 Durable Goods Manufacturing	0.0663	10
	(0.0019)	
5 Non Durable Goods Manufacturing	0.0663	12
	(0.0026)	
6 Transportation	0.0718	18
	(0.0028)	
7 Communication	0.0596	6
	(0.0033)	
8 Utilities	0.0587	5
	(0.0045)	
9 Wholesale Trade	0.0693	14
	(0.0029)	
10 Retail Trade	0.0703	16
	(0.0019)	
11 Finance	0.0901	21
	(0.0039)	
13 Business Services	0.0701	15
	(0.0032)	
14 Personal Services	0.0684	13
	(0.0048)	
15 Recreation and Entertainment	0.0615	8
	(0.0046)	
16 Hospitals	0.0706	17
	(0.0029)	
17 Medical Services	0.0744	20
	(0.0031)	
18 Education	0.0663	11
	(0.0023)	
19 Social Services	0.0531	3
	(0.0028)	
20 Other Services	0.0724	19
	(0.0040)	
21 Government	0.0597	7
	(0.0020)	
22 Armed Forces	0.0395	1
	(0.0070)	

Note: σ_j^2 is the estimate of the variance for the shocks to labor earnings (in logs) for industry j. Bootstrap standard errors are shown in parenthesis. The column Ranking just ranks the industries according to their estimate of the variance.

Table 2: Variance of Earnings by Industry: Transitory and Permanent

	$\sigma^2_{\epsilon,j}$	Ranking	$\sigma^2_{\eta,j}$	Ranking
1 Agriculture and Forestry	0.0143	14	0.0050	20
	(0.0013)		(0.0007)	
2 Mining	0.0139	10	0.0066	21
	(0.0015)		(0.0022)	
3 Construction	0.0141	11	0.0032	5
	(0.0007)		(0.0004)	
4 Durable Goods Manufacturing	0.0151	16	0.0036	10
	(0.0005)		(0.0002)	
5 Non Durable Goods Manufacturing	0.0137	8	0.0037	12
	(0.0005)		(0.0002)	
6 Transportation	0.0156	20	0.0036	9
	(0.0007)		(0.0003)	
7 Communication	0.0121	4	0.0039	13
	(0.0009)		(0.0004)	
8 Utilities	0.0118	3	0.0036	7
	(0.0010)		(0.0004)	
9 Wholesale Trade	0.0142	13	0.0036	8
	(0.0007)		(0.0003)	
10 Retail Trade	0.0155	19	0.0041	16
	(0.0005)		(0.0002)	
11 Finance	0.0177	21	0.0047	19
	(0.0008)		(0.0003)	
13 Business Services	0.0151	17	0.0029	3
	(0.0006)		(0.0003)	
14 Personal Services	0.0129	6	0.0031	4
	(0.0009)		(0.0005)	
15 Recreation and Entertainment	0.0130	7	0.0019	1
	(0.0010)		(0.0004)	
16 Hospitals	0.0142	12	0.0040	14
	(0.0006)		(0.0003)	
17 Medical Services	0.0150	15	0.0037	11
	(0.0007)		(0.0003)	
18 Education	0.0154	18	0.0033	6
	(0.0006)		(0.0002)	
19 Social Services	0.0105	2	0.0041	15
	(0.0007)		(0.0004)	
20 Other Services	0.0138	9	0.0042	17
	(0.0008)		(0.0005)	
21 Government	0.0123	5	0.0043	18
	(0.0004)		(0.0002)	
22 Armed Forces	0.0080	1	0.0025	2
	(0.0013)		(0.0008)	

Note: $\sigma^2_{\epsilon,j}$ and $\sigma^2_{\eta,j}$ are the estimates of the variance of permanent and transitory shocks to labor earnings (in logs), respectively, for industry j. Bootstrap standard errors are shown in parenthesis. The columns called Ranking just rank the industries according to their estimate of the two types of variances.

Table 3: Regression Results - Total Risk

Dependent Variable	Earnings Coefficient	Net Earnings Coefficient
constant	−15.31	6.44
	(0.0231)	(0.0000)
female	−0.55	
	(0.0000)	
age	1.11	
	(0.0003)	
age^2	−0.01	
	(0.0000)	
education	0.315	
	(0.0000)	
σ^2	6.06	1.46
	(0.0002)	(0.0104)

The second column shows the estimation results of regressing log earnings by industry the variables listed in its first column. The third column presents the estimation results of regressing the net earnings obtained in a previous step on a constant term (*constant*) and on our estimates for the total variance of the earnings shock(σ^2). For positive (negative) coefficients the values in parenthesis shows the probability that the coefficient is less (bigger) than zero computed by Bootstrap.

Table 4: Regression Results - Permanent and Transitory

Dependent Variable	Earnings Coefficient	Net Earnings Coefficient
constant	−12.21	6.37
	(0.0646)	(0.0000)
female	−0.49	
	(0.0000)	
age	0.95	
	(0.0011)	
age²	−0.01	
	(0.0011)	
education	0.32	
	(0.0000)	
σ_ϵ^2	9.30	6.87
	(0.18)	(0.0152)
σ_η^2	20.33	16.59
	(0.16)	(0.0771)

The second column shows the estimation results of regressing log earnings by industry the variables listed in its first column. The third column presents the estimation results of regressing the net earnings obtained in a previous step on a constant term (*constant*) and on our estimates for the variance of the permanent shock and transitory shocks to labor earnings (σ_ϵ^2 and σ_η^2, respectively). For positive (negative) coefficients the values in parenthesis shows the probability that the coefficient is less (bigger) than zero computed by Bootstrap.

Table 5: Earnings and Variance of Earnings - 4 Industries

	Mean Earnings	Std. Dev. Mean Earnings	σ_ϵ^2	σ_η^2
Agriculture	6.55	0.3687	0.0141	0.0058
Manufacturing	6.54	0.3869	0.0143	0.0035
Services	6.53	0.3287	0.0141	0.0036
Public Sector	6.50	0.4095	0.0101	0.0034
Correlation with Mean Earnings			0.88	0.69

This table shows the earnings statistics that correspond to the aggregation of the 21 industries into 4 main industries. It contains the mean earnings in logs (second column), the standard deviation of mean earnings, the variance of the permanent shock (third column) and the variance of the transitory shock (fourth column).

Table 6: Parameters - Distribution of Pre-Labor Market Skills

	$\mu_{j,\theta}$	$\sigma_{j,\theta}^2$
Agriculture	6.46	0.31
Manufacturing	6.35	0.26
Services	7.40	0.24
Public Sector	5.43	0.36

This table shows the calibrated values for the mean ($\mu_{j,\theta}$) and variance ($\sigma_{j,\theta}^2$) of the distribution of pre-labor market skills for the 4 industries considered.

Table 7: Share of Workers by Industry

	Model	Data
Agriculture	0.13	0.02
Manufacturing	0.27	0.24
Services	0.49	0.65
Public Sector	0.11	0.10

This table shows the model predictions and their data counterpart of the share of workers in each of the 4 industries considered.

Table 8: Model Predictions - Counterfactual Exercise

	Mean Net Earnings	Share of Workers
Agriculture	6.54	0.14
Manufacturing	6.56	0.17
Services	6.53	0.18
Public Sector	6.48	0.35
Correlation with Data	0.90	−0.22

This table shows the model predictions for the mean of log earnings and the share of workers in each of the 4 industries considered.

Table 9: Regression Results (Earnings Per Hour) - Total Risk

Dependent Variable	Earnings Coefficient	Fixed Effect Coefficient
constant	−9.36	1.42
	(0.0017)	(0.0000)
female	−0.32	
	(0.0011)	
age	0.57	
	(0.0002)	
age²	−0.01	
	(0.0004)	
education	0.17	
	(0.0000)	
σ^2	5.42	1.67
	(0.0055)	(0.0663)

The second column shows the estimation results of regressing log earnings per hour by industry the variables listed in its first column. The third column presents the estimation results of regressing the net earnings per hour obtained in a previous step on a constant term (*constant*) and on our estimates for the total variance of the earnings shock(σ^2). For positive (negative) coefficients the values in parenthesis shows the probability that the coefficient is less (bigger) than zero computed by Bootstrap.

Table 10: Regression Results (Earnings Per Hour) - Permanent and Transitory

Dependent Variable	Earnings Coefficient	Net Earnings Coefficient
constant	-8.12	1.3928
	(0.0055)	(0.0000)
female	-0.34	
	(0.0041)	
age	0.49	
	(0.0012)	
age²	-0.01	
	(0.0020)	
education	0.15	
	(0.0008)	
σ_ϵ^2	6.42	3.09
	(0.0509)	(0.0425)
σ_η^2	17.30	0.26
	(0.1338)	(0.5387)

he second column shows the estimation results of regressing log earnings per hour by industry the variables listed in its first column. The third column presents the estimation results of regressing the net earnings per hour obtained in a previous step on a constant term (*constant*) and on our estimates for the variance of the permanent shock and transitory shocks to labor earnings (σ_ϵ^2 and σ_η^2, respectively). For positive (negative) coefficients the values in parenthesis shows the probability that the coefficient is less (bigger) than zero computed by Bootstrap.