

# CAREER CHOICE AND THE RISK PREMIUM IN THE LABOR MARKET

German Cubas and Pedro Silos\*

February, 2017

## Abstract

Is risk priced in the labor market? To answer this question we first document a strong, robust, and positive correlation between average earnings and the standard deviation of both temporary and permanent idiosyncratic shocks to earnings across 19 US industries. This result suggests that risk is compensated in the labor market. However, workers are heterogeneous in their industry-specific abilities and their industry choice is endogenous. Thus, the correlation may be entirely driven by selection. We construct an equilibrium model of career choice that allows us to decompose the estimated correlation between compensation for risk and compensation for unobserved abilities. In the model economy workers cannot perfectly insure against labor earnings shocks and their comparative advantage interacts with her risk aversion to determine an optimal career choice. We find that permanent shocks are highly compensated, whereas temporary shocks are not. Two additional results arise. First, workers accumulate different levels of wealth depending on the employment industry. Second, compensation for risk explains a sizable fraction of observed cross-industry differences in labor earnings.

*Key words:* Risk Premium, Labor Markets, Roy Model, Incomplete Markets

*JEL Classifications:* E21 · D91 · J31.

---

\*Affiliation: University of Houston, Temple University, respectively. Corresponding author: German Cubas, University of Houston, 3623 Cullen Boulevard Room 204, Houston, TX 77204, USA, gcubasnorando@uh.edu, 1-713-743-3187 (Phone), 1-713-743-3798 (Fax). We thank the Editor, Matthias Doepke, and three anonymous referees for their excellent comments and suggestions. We also thank D. Amengual, Y. Chang, M. Bils, F. Borraz, C. Carroll, M. Deveraux, J. Dubra, J. Hatchondo, G. Kambourov, D. Lagakos, M. Lopez-Daneri, J. Ponce, B. Ravikumar, V. Ríos-Rull, C. Robotti, R. Rogerson, Y. Shin, M. Tertilt, G. Ventura, and R. Warren for their comments and suggestions. Seminar participants at Fce-Udelar, Central Bank of Chile, Central Bank of Uruguay, Federal Reserve Bank of Atlanta, Federal Reserve Bank of Dallas, Federal Reserve Bank of St. Louis, Instituto Tecnológico Autónomo de México, National Bureau of Economic Research Summer Institute, Quantitative Society for Pensions and Saving workshop at Utah State University, Society for Economic Dynamics Meeting in Seoul, Tsinghua Workshop in Macroeconomics, Universidad de Los Andes, Universidad ORT, University of Georgia, University of Houston and University of Iowa, provided useful feedback. G. Canavire provided excellent research assistance at the beginning of this project. German Cubas acknowledges support from the Fondo Clemente Estable of ANII (project FCE-2-2011-1-6904). This research was also supported in part by the National Science Foundation through major research instrumentation grant number CNS-09-58854. Early versions of this paper circulated under the name "Comparative Advantage and the Risk Premia in the Labor Market".

# 1 Introduction

The risk-return trade-off in financial assets has been extensively studied in financial economics and macroeconomics. Much work has been devoted to constructing economic environments whose quantitative predictions match the data. Although the findings of a large empirical literature suggest the labor market is a major source of risk for most workers, to date, a quantitative analysis of the risk-return trade-off in the labor market has not been attempted. As a first attempt, this paper poses and answers the following questions: Are labor earnings positively correlated with their volatility? If so, what does the standard framework for studying imperfect risk-sharing imply for the risk-return trade-off workers face in the labor market? What additional ingredients does the standard framework need in order to match and understand the facts on labor earnings and their volatility? Surprisingly, these questions have never been answered. Answers to those questions would improve our understanding of central topics in labor economics and macroeconomics, ranging from the measurement of risk in the labor market to policy related issues such as redistribution and insurance policies and optimal taxation.

We begin by documenting two facts using the Survey of Income and Program Participation (SIPP). The first finding is a large dispersion in the variance (both transitory and permanent) of shocks to labor earnings across 19 US industries: Workers in the construction or transportation industries experience large permanent shocks to earnings, while those working in government and social services are insulated from earnings variability. Our second and new finding is a strong, positive, and robust correlation between mean earnings and earnings risk across industries. This correlation is obtained once we control for other industry characteristics that affect the average level of earnings (education, age, etc). The estimated coefficients imply a difference in average earnings, solely related to permanent risk, between the riskiest and safest industries of around 5%. When considering transitory shocks, the increase due to risk between earnings of workers in the safest and the riskiest industries is 3%.

These correlations must be approached with caution. The study of the risk-return trade-off in labor markets is more complex than in financial markets, and as a result that positive correlation cannot immediately be associated with the existence of a risk premium in labor markets. Since workers are heterogeneous in their abilities and industries value some abilities more than others, the estimated correlation may be driven entirely by selection. In other words, high average earnings in risky industries may reflect the relative scarcity of particular abilities. Workers lacking such abilities would be unable to command high earnings even if they were willing to expose themselves to risk.

To quantitatively evaluate the risk-return trade-off in labor markets, our starting point is the workhorse framework used to study risk-sharing in macroeconomics: an equilibrium consumption-savings model in which workers experience shocks (permanent and transitory) to their labor earnings. We introduce a career choice in which different careers represent industries. In our environment, risk-averse individuals choose an industry in which to supply labor services. Some industries are riskier than others and, everything else equal, they are less attractive. These industries are less attractive because shocks to labor earnings are not perfectly insurable. Since this environment says nothing about selection, it attributes any correlation between risk and the average level of earnings to compensation for risk. To account for selection, we assume that workers are ex ante heterogeneous: Each is endowed with an industry-specific ability, which drives a worker's comparative advantage, interacting with her risk aversion to determine an optimal career choice. To summarize, the model overlaps an Aiyagari (1994) economy, where markets are incomplete, with a Roy (1951) model, where workers self-select into different careers based on their comparative advantage.

We confront our model with US data from several sources. Earnings in the model are subject to permanent and transitory shocks. The stochastic processes driving these shocks are the same as those estimated using the SIPP. The parameters driving the industry-specific production functions are from National Income and Product Accounts data. In our benchmark case, we set the risk-aversion parameter to 3 and parameterize the distributions of abilities so that in equilibrium

the model delivers the mean and standard deviations of the cross-sectional distributions of earnings observed in the data for each of the industries in the economy. Therefore, by construction, we replicate the documented positive correlation between mean earnings and their volatility. Two important model predictions that are not matched in the calibration process arise and are noteworthy. First, the model predicts a distribution of workers into industries that resembles the distribution observed in US data. Second, the wealth-to-income ratios predicted by the model for each of the industries are positively correlated with their standard deviation of the permanent shock. This fact is established by Carroll and Samwick (1997) using data from the Panel Study of Income Dynamics (PSID).

Viewed through the lens of the model, the positive relationship between the permanent and transitory risk to earnings and the average level of earnings is a convolution of two forces: the compensation for risk and the compensation for industry-specific skills. Therefore, in order to separate the effect of these two forces into the observed differences in mean earnings, we proceed by performing a counterfactual exercise in which we eliminate individual differences in ability or comparative advantage. In other words, we consider individuals as *ex ante* homogeneous. In this counterfactual world, only the differences in the volatility of earnings across industries shape an individual's industry choice. This experiment yields two results. First, with the same value for the risk-aversion parameter, the model predicts the positive correlation between mean earnings and permanent risk (i.e. as in the data there is a risk premium).

Second, in this counterfactual exercise, the model predicts a no temporary risk premium. Thus, according to this result, the strong association between the standard deviation 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) arises entirely from selection and the permanent shock to earnings is the one that is compensated. We further investigate what drives the correlation between earnings and the permanent and transitory risk. Once the sorting of workers has taken place, mean earnings in a particular industry are the result of the product of the mean abilities of workers and the equilibrium wage rate. Thus,

our model allows us to quantify the relative roles of the amount of human capital (abilities) versus the unit price of human capital (the wage rate) in explaining the overall correlation between average earnings and our two measures of risk. We find that mean abilities are positively correlated with both permanent and transitory risk whereas the equilibrium wage rates are negatively correlated.

Finally, the model has implications for labor earnings inequality. Overall, inequality is driven almost entirely by shocks, with little effect caused by initial earnings differences implied by workers' comparative advantage. Regarding inter-industry earnings inequality, cross-industry differences in risk and abilities contribute almost equally to explaining the observed earnings inequality across industries.

**Related Literature** To our knowledge, the first attempt to empirically analyze the link between the variability of income and mean earnings was the seminal work of Kuznets and Friedman (1954) in their classic study of incomes of professionals. In the early 1980s, Abowd and Ashenfelter (1981), Feinberg (1981), and Leigh (1983) continued this work. This earlier literature focused on the cross-sectional dispersion of earnings, rather than the variation of earnings over time, and did account for earnings risk due to changes in employment status. Feinberg (1981) considers variation of income over time but uses a very small subset of the PSID to estimate the relationship between risk and earnings.<sup>1</sup> Moreover, all these authors interpret their empirical results as proof (or lack thereof) of the existence of a risk premium or compensating differential, with no mention of the possibility of selection influencing the empirical findings. Finally, they do not distinguish between transitory and permanent shocks to earnings.

One important contribution of our paper is that we measure idiosyncratic labor market risk by industry and analyze its macroeconomic implications. On the measurement side, we build on papers such as Storesletten, Telmer, and Yaron (2004b), Guvenen (2009), and Low, Meghir, and Pistaferri (2010), but we extend this literature by explicitly considering different industries. On the modeling side,

---

<sup>1</sup>He has annual earnings data for only 326 individuals whom he follows for 6 years.

our work belongs to the extensive quantitative macroeconomics literature with heterogeneous agents and incomplete markets initiated by Bewley (1977), Huggett (1993), and Aiyagari (1994). More recent contributions include those by Storesletten, Telmer, and Yaron (2004a) and Heathcote, Storesletten, and Violante (2008, 2009).

An important aspect of our work is that the estimated measures of earnings volatility are in part a result of a worker's industry choice. Workers self-select into industries and that determines the risk they will face in their careers. The problem of controlling for unobserved heterogeneity when estimating earnings risk has been tackled in a strand of the previous literature; examples are Heckman, Lochner, and Taber (1998), Chen (2008) and Cunha, Heckman, and Navarro (2005). However, none of those works deal with the sorting of workers across industries and with the question of whether risk is priced in the labor market.

Our work contributes to a growing literature that develops quantitative models of occupational choice and income dynamics. One important related paper is Kambourov and Manovskii (2009), who study the interplay between occupational mobility and wage inequality. Other important papers that study occupational mobility and occupation specific human capital are Neal (1999) and Gathmann and Schönberg (2010). We abstract from mobility; our focus is on industries, industry-specific risk and its interplay with workers' abilities. Mobility may be important as it could serve as a response to adverse income shocks, as studied in Neumuller (2015). The emphasis of this work is on the switching option as a way to obtain a negative relationship between earnings risk and mean earnings. Neumuller (2015) finds a negative relationship using biennial earnings data from the Panel Study of Income Dynamics (PSID). His empirical result contrasts with that reported by Dillon (2016), who uses the same data. In addition, we also abstract from the process of accumulating those abilities as we assume they are known by workers at the beginning of their working lives. It may be the case that there is a process of experimentation into different industry careers, as studied in Antonovics and Golan (2012).

More recently, in work contemporaneous to ours, Dillon (2016) finds a positive

relationship between the expected value and variance of lifetime earnings. Aside from the different methodology and dataset used by Dillon, our framework, (i) incorporates the industry choice of workers in a general equilibrium model and (ii) explicitly models the interplay between unobserved abilities and aversion to risk, which are absent in Dillon's work. As we describe in detail below, the general equilibrium aspect of our work is important as the sorting of workers into industries affects the price of industry-specific abilities. Nevertheless, she uses a richer econometric model and, more importantly, her results complement and confirm our main empirical finding.

We highlight a source of wage inequality that is intimately related to individuals' choice of industry. Our framework allows us to decompose the observed overall inequality into shocks and unobserved abilities and we find that the former explains most of the observed inequality. This result contrasts with those obtained in Keane and Wolpin (1997) and Huggett, Ventura, and Yaron (2011) as in these papers the role of initial conditions is larger in explaining income inequality. These papers have richer set ups to model the process of human capital accumulation. In our case we take abilities at the beginning of the labor careers as given, and we focus on the multidimensional aspect of them. This way, our framework incorporates workers' comparative advantage and its effect on occupational choice, thus it is closely related to Roy's (1951) work. The empirical content of the original Roy model is studied in Heckman and Honore (1990) and Buera (2006). Roy's ideas are also adapted in modern dynamic discrete choice models to analyze the sources of income inequality—first, in an important paper, Keane and Wolpin (1997) and more recently in Hoffmann (2010). However, we consider our work as the first that integrates Roy's ideas into the analysis of career choice under uninsurable idiosyncratic labor earnings risk in a general equilibrium model. Along these lines we see our framework as a useful tool to be applied in future work incorporating workers' comparative advantage into the analysis of earnings dynamics and wage inequality.

## 2 Labor Market Risk and Mean Earnings

In this section, we briefly describe our dataset. Then we document that risk and return in earnings are positively correlated across industries. We do this in two steps. First, we estimate the labor earnings processes and the properties of the shocks faced by workers in different industries. Second, we characterize and estimate the empirical relation between mean earnings and the degree of uncertainty in earnings across industries.

Our definition of labor earnings is rather broad (but consistent with previous studies). In addition to the obvious variability in wage rates, we also consider changes in earnings due to changes in the number of hours worked or changes in employment status.<sup>2</sup> We focus on individuals who never change industries as this is most consistent with the quantitative framework we use.

### 2.1 The SIPP

To explore the relationship between the level of average earnings and the degree of unforeseen variability in those earnings, we turn to the data. Ideally, to obtain an accurate answer, researchers would hope for an extended high-frequency, large panel of individual labor earnings with characteristics describing both the employee and the employer. The richer that dataset, the easier it would be to separate risk from other features that could affect average earnings. For the United States, the SIPP is the best approximation of that ideal dataset. The SIPP is constructed by the U.S. Census Bureau and takes the form a series of continuous panels that follow a national sample of households. The first panel began in 1983 but the earlier panels had a short duration. 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 these panels are the ones on which we focus. More specifically, we use the 1996, 2001, and 2004 panels obtained from the Center for Economic and Policy Research, CEPR (2014).

The SIPP conducts interviews every four months, eliciting information at monthly

---

<sup>2</sup>We do not consider individuals who move in and out of the labor force, but we do consider employment to unemployment transitions and vice versa.



frequency on variables such as labor earnings, demographic characteristics, occupation, and so on. We use quarterly data constructed from the monthly data provided in SIPP. We are able to follow individuals for at most 16 quarters; this short duration prevents us from having entire life-cycle profiles of earnings. The appendix provides a step-by-step description of our choice of the sample of individuals on which we perform the analysis described in this section. In brief, we (i) focus on individuals between 22 and 66 years of age (ii) eliminate those who are self-employed, (iii) simultaneously report missing earnings but positive hours worked, (iv) those out of the labor force, and (v) focus on individuals with at least 10 consecutive quarters of responses. In addition, we define earnings to include unemployment insurance if an individual reports zero hours worked and is unemployed.

There are two main advantages to using SIPP data.<sup>3</sup> The first advantage is the number of respondents. The SIPP sample is considerably larger than that of the PSID, which surveys about 10,000 households, or the NLSYs, which interview between 9,000 and 13,000 individuals. This feature allows us to have 19 industries with a significant number of workers in each of them. The second advantage is the interview frequency. The SIPP provides a wealth of information at a monthly frequency; the PSID interviews annually (biennially since 1997) and the NLSY97 has also been biennial since 1994. Fortunately, for many individuals in the United States unemployment or declines in income are short-lived experiences (usually weeks or a few months). But given these are the risks on which this study focuses, that fact underscores the importance of the availability of information at higher frequencies.

## 2.2 Labor Income Shocks

The first step in our analysis computes earnings variability at the individual level with a regression approach used extensively in the literature, see for example,

---

<sup>3</sup>See the validation study by Abowd and Stinson (2011) who compare SIPP's good-quality earnings data with administrative data. In addition, Gottschalk and Huynh (2006) study the SIPP relative to other longitudinal panels such as the Panel Study of Income Dynamics (PSID) and the National Longitudinal Surveys of Youth (NLSY97 and NLSY79).

Carroll and Samwick (1997). We proceed by estimating a fixed effects 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

$$y_{ijt} = \alpha_{ij} + \beta_j \mathbf{X}_{ijt} + u_{ijt} \quad (1)$$

The vector  $\mathbf{X}$  includes 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 time dummies.

We estimate equation (1) for all individuals in a given industry. Repeating this procedure for all industries yields estimates  $\{\hat{\alpha}_{ij}, \hat{\beta}_j\}_{j=1}^{19}$ .

Depending on the persistence of the shocks and workers' risk aversion, small differences in variances across industries may imply large differences in welfare. To account for this difference in the nature of risk, we enrich our empirical analysis by allowing the error term to be decomposed into a permanent component and a transitory component. 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 such shocks successfully. Permanent shocks last longer and can be associated with, for instance, depreciation of job-specific human capital or permanent changes in how an industry operates. Smoothing out the latter type of shock through a buffer stock of savings is more difficult and permanent changes in consumption are often required. We follow Carroll and Samwick (1997) and Low, Meghir, and Pistaferri (2010), among others, by assuming that the residual is equal to the addition of a permanent and a transitory component. In addition, given we use quarterly data we enrich our analysis by allowing for the possibility of no occurrence of the shocks in every quarter. Moreover, we allow the probability of the occurrence of the shocks to be industry-specific.

Thus, we assume that

$$u_{ijt} = \eta_{ijt} + \omega_{ijt}, \quad (2)$$

where  $\eta_{ijt}$  is the transitory component and  $\omega_{ijt}$ , the permanent component which

is a random walk, that is,

$$\omega_{ijt} = \omega_{ij,t-1} + \epsilon_{ijt}. \quad (3)$$

As mentioned above, we further assume that

$$\eta_{ijt} = \begin{cases} 0 & \text{with probability } \phi_j \\ \tilde{\eta}_{ijt} & \text{with probability } 1 - \phi_j \end{cases} \quad (4)$$

with  $\tilde{\eta}_{ijt}$  distributed i.i.d.  $N(0, \sigma_{\tilde{\eta},j}^2)$ ; and

$$\epsilon_{ijt} = \begin{cases} 0 & \text{with probability } \lambda_j \\ \tilde{\epsilon}_{ijt} & \text{with probability } 1 - \lambda_j \end{cases} \quad (5)$$

with  $\tilde{\epsilon}_{ijt}$  distributed i.i.d.  $N(0, \sigma_{\tilde{\epsilon},j}^2)$ .

The estimation of equation (1), we obtain  $\{\{\hat{u}_{ijt}\}_{i=1}^{N_j}\}_{t=1}^T$ . Using those and for each industry  $j$ , we estimate the vector of parameters  $\{\sigma_{\tilde{\epsilon},j}^2, \sigma_{\tilde{\eta},j}^2, \lambda_j, \phi_j\}$  by the method of moments. We follow an identification procedure similar to the one proposed by Low, Meghir, and Pistaferri (2010) in which the moments to match are  $E[(\Delta u_{ijt})^2]$ ,  $E[(\Delta u_{ijt})^4]$ ,  $E[\Delta u_{ijt} \Delta u_{ijt-1}]$  and  $E[(\Delta u_{ijt})^2 (\Delta u_{ijt-1})^2]$ . The theoretical expressions for these moments are functions of the vector of parameters and they are given by the following equations.

$$E[(\Delta u_{ijt})^2] = 2(1 - \phi) \sigma_{\tilde{\eta},j}^2 + (1 - \lambda) \sigma_{\tilde{\epsilon},j}^2 \quad (6)$$

$$E[(\Delta u_{ijt})^4] = 6(1 - \phi)^2 \sigma_{\tilde{\eta},j}^4 + 12(1 - \phi)(1 - \lambda) \sigma_{\tilde{\eta},j}^2 \sigma_{\tilde{\epsilon},j}^2 + 6(1 - \phi) \sigma_{\tilde{\eta},j}^4 + 3(1 - \lambda) \sigma_{\tilde{\epsilon},j}^4 \quad (7)$$

$$E[\Delta u_{ijt} \Delta u_{ijt-1}] = -(1 - \phi) \sigma_{\tilde{\eta},j}^2 \quad (8)$$

and

$$E[(\Delta u_{ijt})^2 (\Delta u_{ijt-1})^2] = 3(1 - \phi)^2 \sigma_{\tilde{\eta},j}^4 + 4(1 - \phi)(1 - \lambda) \sigma_{\tilde{\eta},j}^2 \sigma_{\tilde{\epsilon},j}^2 + (1 - \lambda)^2 \sigma_{\tilde{\epsilon},j}^4 + (1 - \phi) \sigma_{\tilde{\eta},j}^4 \quad (9)$$

To estimate the variances of the two innovations, we proceed as follows. For a sample of workers in a given industry  $j$ , we estimate  $E(\widehat{\Delta u_{ijt} \Delta u_{ijt}})$ ,  $E(\widehat{\Delta u_{ijt} \Delta u_{ijt-1}})$ ,  $E([\widehat{\Delta u_{ijt} \Delta u_{ijt}}]^2)$  and  $E([\widehat{\Delta u_{ijt} \Delta u_{ijt-1}}]^2)$  using the sample analogs. Solving the system comprised of the previous four equations, we obtain  $\hat{\sigma}_{\varepsilon,j}^2$ ,  $\hat{\sigma}_{\eta,j}^2$ ,  $\hat{\lambda}_j$  and  $\hat{\phi}_j$ . As a result, the estimates of the variances of the permanent and transitory shocks are  $\hat{\sigma}_{\varepsilon_j}^2 = (1 - \hat{\lambda}_j)\hat{\sigma}_{\varepsilon_j}^2$  and  $\hat{\sigma}_{\eta_j}^2 = (1 - \hat{\phi}_j)\hat{\sigma}_{\eta_j}^2$ , respectively.

Table 1 presents the estimated variances for each of the 19 industries in our sample, standard errors obtained by block-bootstrap.<sup>4</sup> The weighted average of the estimated variances across industries is 0.0062 for the permanent shock and 0.0042 for the transitory shock. Although we use total earnings instead of earnings per hour and our data are at the quarterly frequency instead of the annual frequency, our estimates for the variance of the permanent shock are very much in line with the estimates obtained in the literature. The variance of the permanent shock in annual terms is 0.0231 in Dillon (2016) and Low, Meghir, and Pistaferri (2010) and 0.0217 in Carroll and Samwick (1997). As for the transitory shock, there is more variation in the estimates provided in the literature, the estimates ranges from 0.0074 (Low, Meghir, and Pistaferri (2010)) to 0.044 (Carroll and Samwick (1997)). Regarding the estimated probability of the occurrence of the shocks, the permanent shocks are more frequent than the transitory shocks: the estimated average probability for the permanent shock is 0.82 and for the transitory shock is 0.13.

However, both the magnitude of the shocks and the probability of their occurrence are important for the perspective of the worker, and those two moments differ by industry. In other words, this decomposition uncovers large differences across industries in the degree of permanent and transitory income volatility. This is clear in Figures 1 and 2 which show the boxplots of the variances for each industry.<sup>5</sup> In terms of the permanent shocks to earnings the riskiest industries are

---

<sup>4</sup>We eliminate Armed Forces and Mining as these industries present physical risk. Although that risk maybe important for the occupational choice of individuals, its modeling is beyond the scope of this paper.

<sup>5</sup>In addition, we perform a Kolmogorov-Smirnov test that leads us to reject the hypothesis that the estimated variances come from the same distributions at 5% level of significance for every pair of industries.

Transportation and Construction, whereas the safest are Government and Social Services. As for transitory shocks, the riskiest are Other Services and Medical Services, whereas the safest are Recreation and Entertainment and Business Services. As shown in Table 2, in some industries both permanent and transitory shocks have large variances. Other industries are relatively safe regarding both types of shocks. The riskiest industry with respect to both shocks is Transportation and the safest are Education, Social Services and Agriculture. There are industries that have low variance for the permanent shock but high transitory variance: Government, for instance. Finally, workers in Construction face low transitory risk but high permanent risk. The specification of a stochastic process for shocks with industry-specific probabilities of earnings adjustments offers new insights regarding the nature, frequency and magnitude of the shocks. For some industries the probability of occurrence is low but the variance of the shocks (conditional on occurring) is large, and for others, the frequency of the shocks is high but the actual magnitude of the shocks, when they occur, is smaller. For instance Business Services and Medical Services have almost the same variance of permanent shocks (0.0066 vs. 0.0064) but the frequency of non-adjustment  $\lambda$  is almost 37% for Business Services but only 7% for Medical Services. The same can be said about Construction and Personal services when evaluating the variances of temporary shocks. Construction has a much lower probability of adjustment than Personal Services but their overall variance is roughly the same.<sup>6</sup>

### 2.3 Correlation Between Earnings and Risk

Having estimated measures of risk for our group of industries, we are now ready to test the hypothesis that the level of risk and the average level of earnings across industries are positively correlated. Our claim, however, should be understood as 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

---

<sup>6</sup>The focus of our work is on the cross-industry differences in idiosyncratic risk. However, there exist differences in aggregate risk across industries. By aggregate risk at the industry level we refer to correlated shocks across all workers in an industry. In the Appendix we show that the bulk of the risk faced by workers is idiosyncratic and not aggregate.

else is constant across industries. Industries differ along many dimensions that may affect average earnings independently of their level of risk. This should raise suspicion that either the mix of workers or firms in a given industry is an important determinant of the industry's average level of earnings; this is addressed in the proposed quantitative model below. For now, we estimate the relationship between earnings and earnings risk in two different ways to account for this industry heterogeneity. The first approach involves computing industry averages (that is, averages across individuals who work in a given industry) of variables we deem relevant in determining average earnings. To estimate the relationship between average earnings and industry risk, we estimate the following regression equation:

$$y = \gamma + \theta Z + v, \quad (10)$$

where  $y$  is a vector whose  $j$ th element is the log of average earnings across individuals in industry  $j$ , and  $Z$  is a matrix of regressors. The  $j$ th row of  $Z$  has seven elements: average age, average age squared, the average level of education of all individuals working in industry  $j$ , the fraction of females in industry  $j$ , two dummies that represent two of the three panels of the SIPP used in the estimation, and the industry  $j$  standard deviations of permanent and transitory income shocks estimated above. Since the number of industries in each of the panels of our sample is 19,  $y$  is a column vector of dimension 57 and  $Z$  is of dimension  $57 \times 9$ . Finally,  $\gamma$  is a vector of intercepts and  $v$  is a vector of residuals.<sup>7</sup>

The second approach to estimating the relationship between industry risk and mean earnings involves estimating earnings net of an individual's main observed characteristics: age, age squared, level of education, and gender. We estimate the following pooled regression,

$$y_{ijt} = \gamma_0 + \gamma X_{ijt} + \xi_{ijt}. \quad (11)$$

The vector of coefficients  $\gamma$  represent the effect of the observed characteristics (age, age squared, level of education, and gender) on (log) earnings. These coeffi-

---

<sup>7</sup>For reasons commented above we exclude Armed Forces and Mining but our qualitative results do not change if we estimate this regression by adding these two industries.

cients are common across individuals and across time. We use these estimates to compute net (log) earnings at the individual level:  $\tilde{y}_{ijt} = y_{ijt} - \hat{\gamma}X_{ijt}$ . Averaging across time and across individuals in each industry gives the mean of net earnings by industry. Formally, for a given industry  $j$ , we compute

$$\tilde{y}_j = \log \left( \frac{1}{N_j} \sum_{i=1}^{N_j} e^{\tilde{y}_{ij}} \right), \quad (12)$$

where

$$e^{\tilde{y}_{ij}} = \frac{1}{T} \sum_{t=1}^T e^{\tilde{y}_{ijt}}. \quad (13)$$

We regress (log) average net earnings by industry on the standard deviations of earnings to estimate the relationship between earnings and volatility:

$$\tilde{y}_j = \alpha_0 + \alpha_1 \sigma_{\epsilon,j} + \alpha_2 \sigma_{\eta,j} + \alpha_3 I_{1996} + \alpha_4 I_{2001} + \xi_{\tilde{y},j}. \quad (14)$$

The number of left-hand-side observations is again 57 (19 for each of the three panels).  $\alpha_0$  is an intercept,  $\alpha_1$  is the coefficient that relates permanent risk to (net) earnings,  $\alpha_2$  is the coefficient that relates transitory risk to net earnings,  $\alpha_3$  and  $\alpha_4$  are the coefficients that represent the dummies for the 1996 and 2001 SIPP panels, respectively.

Table 3 presents the results. The second column of the table refers to the estimates of equation (10) and the third column refers to the estimates obtained using equation (14). In both cases all coefficients are positive and significant, and of similar magnitudes.

According to the estimates presented in column 3, the estimated value of the coefficient associated with the standard deviation of the permanent shock to earnings is 2.08. That means that moving from Government (the safest industry) to Construction (the riskiest industry) implies a 4.73% increase in net mean earnings. For the transitory shock, the value estimated for its associated coefficient is 1.15 and hence, switching from Recreation and Entertainment (the safest industry) to Medical Services (the riskiest industry) implies a 3.21% increase in mean net

earnings.

The data and our approach to linking labor earnings and their uncertainty yield estimates that appear consistent with a compensating differential for risk in the labor market. But caution is necessary in interpreting this result. 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 a factor individuals consider when making that choice. But their comparative advantage, in other words, their relatively higher productivity in a certain industry, which is consequence of a set of individual characteristics, plays a role as well. Some portion of that comparative advantage originates, for instance, from being a female or a college graduate, characteristics that we have accounted for to some degree. Much of the advantage, however, originates from unobserved characteristics, which are obviously difficult to control for.

How much of the estimated positive risk-earnings correlation is due to a compensating differential and how much is due to self-selection? In the next section, we use a quantitative framework in which comparative advantage and individuals' industry choice are explicitly modeled to determine the answer.

### **3 The Model**

In what follows we provide a description of our model. Our notation makes clear that we are restricting our analysis to a stationary environment. Our economy is populated by a mass of risk-averse individuals of total measure equal to 1. 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 industries. At birth, prior to



the industry choice, each individual draws a value for an industry-specific skill or ability from a given distribution specified below. These skills directly affect the productivity and, hence, earnings of an individual, thereby determining an individual's comparative advantage for, say, working in finance and not in agriculture. Since 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 industry-specific skills neither grow nor decrease over an individual's lifetime.

Once individuals are working within a particular industry (from which they cannot move), they are subject to idiosyncratic shocks to their labor income. The process driving these shocks differs from industry to industry, and workers in some industries experience a higher variability in earnings than workers in other industries. If workers are risk averse, riskier industries look less attractive.

When an individual is born in period 0 (i.e., when she enters the labor market), her problem is to choose one of the  $J$  mutually exclusive career alternatives to maximize the expected discounted value of her lifetime utility.

At time 0, individual  $i$  knows his set of abilities represented by the vector  $\Omega_{i,0}$ , which is defined as follows:

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

where  $\theta_{i,j}$  represents the individual-specific industry pre-labor-market skills (one per industry  $j$ ). Each value  $\theta_{i,j}$  is drawn from an industry-specific distribution  $N(\mu_{\theta_j}, \sigma_{\theta_j}^2)$ . We have one distribution per industry and we assume they are independent. Once the individual chooses an industry, only the  $\theta$  corresponding to the chosen industry affects her lifetime labor earnings. Each period, by inelastically supplying one unit of time to industry  $j$ , each individual receives labor earnings,  $w_j e^{\theta_{i,j}} e^{v_{i,j}}$ , composed of an industry-specific competitive wage rate ( $w_j$ ), individual-specific industry pre-labor-market skills ( $\theta_{i,j}$ ), and an individual-specific but time-varying labor productivity shock ( $v_{i,j}$ ).

For an individual of age  $s$  in industry  $j$ , the time-varying component of earn-

ings is the addition of two orthogonal random variables: a transitory,  $\eta_{s,j}$ , and a permanent component,  $\omega_{s,j}$ . Thus

$$v_{s,j} = \eta_{s,j} + \omega_{s,j}. \quad (15)$$

As in the previous Section,  $\omega_{s,j}$  follows a random walk:  $\omega_{s+1,j} = \omega_{s,j} + \epsilon_{s,j}$ . In addition, we assume that

$$\eta_{sj} = \begin{cases} 0 & \text{with probability } \phi_j \\ \tilde{\eta}_{sj} & \text{with probability } 1 - \phi_j \end{cases} \quad (16)$$

with  $\tilde{\eta}_{sj}$  distributed i.i.d.  $N(-\frac{1}{2}\frac{1}{\sigma_{\tilde{\eta},j}^2}, \sigma_{\tilde{\eta},j}^2)$ ; and

$$\epsilon_{sj} = \begin{cases} 0 & \text{with probability } \lambda_j \\ \tilde{\epsilon}_{sj} & \text{with probability } 1 - \lambda_j \end{cases} \quad (17)$$

with  $\tilde{\epsilon}_{sj}$  distributed i.i.d.  $N(-\frac{1}{2}\frac{1}{\sigma_{\tilde{\epsilon},j}^2}, \sigma_{\tilde{\epsilon},j}^2)$ .

By subscripting the variance by  $j$ , we clarify 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 has begun her working life. We assume individuals start their working life with zero wealth. 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.<sup>8</sup> The variables relevant to these decisions are the level of wealth ( $b$ ), the age of the individual ( $s$ ), and the following components of income: the time-varying component ( $\omega$  and  $\eta$ ) and the ability level for the chosen industry ( $\theta_j$ ). Thus, the vector of individual state variables can be

---

<sup>8</sup>Our model differs 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 the literature analyzing heterogeneous-agent economies. The only departure is that we allow for two different shocks with different statistical properties. This departure allows us to analyze the impact of transitory and permanent risk on industry choice.

denoted as  $x = (b, \omega, \eta, s, \theta_j)$ , where  $j$  is the chosen industry. Denote by  $\Psi_j(x)$  the distribution of industry  $j$  workers across assets, age, income, and abilities.<sup>9</sup> It is an aggregate state variable since it determines the wage rate in industry  $j$ .<sup>10</sup> Only the marginal distribution of age is identical across all industries. For convenience, let  $\mathcal{S} = \mathcal{S}_B \times \mathcal{S}_{E_\eta} \times \mathcal{S}_{E_\omega} \times \mathcal{S}_\theta \cup \{1, \dots, S\}$  denote the state space of the vector of state variables  $x$ .<sup>11</sup> It is convenient to write the problem recursively, and we denote the remaining lifetime utility for an individual working in industry  $j$  by  $V_j(x)$ . For ages  $\{1, \dots, S\}$ ,<sup>12</sup> it is defined by

$$V_j(x) = \max_{c, b'} \{u(c) + \beta EV_j(x')\}, \quad (18)$$

subject to

$$c + b' = w_j e^{\theta_j} e^\eta e^\omega + b(1 + r) \quad (19)$$

and

$$b' \geq \underline{b}, \quad b_0 = 0, \quad b_{S+1} \geq 0. \quad (20)$$

We follow relatively standard notation when we write  $x'$  to denote the values of  $x$  one period ahead. Equation (19) 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,  $w_j e^{\theta_j} e^\eta e^\omega$ . In addition to this budget constraint, individuals face a borrowing constraint that restricts the lower bound on asset holdings,  $\underline{b}$ . Also, individuals are born with zero wealth ( $b_0 = 0$ ) and face a non negativity constraint in their savings at the time of death ( $b_{S+1} \geq 0$ ).

As for the industry choice, at birth the individual chooses from a set of  $J$  industries the one that yields the highest utility:

---

<sup>9</sup>The distribution is subscripted by  $j$  because workers, facing different income shocks and self-selecting into industries based on different ability levels, will choose different levels of assets.

<sup>10</sup>Since we assume stationarity we omit including it as an argument in the value function.

<sup>11</sup>In general, the joint state space should have a subscript  $j$ . In our model, the borrowing constraint and longevity are identical across industries. Income innovations and abilities are all real numbers; hence, we can omit the subscript  $j$ .

<sup>12</sup>For ages  $S \geq S + 1$  the value function is zero.

<sup>13</sup>As is clear from equations (17) and (16),  $\omega$  and  $\eta$  follow a mixture distribution. The expectation operator in (18) is taken with respect to that mixture distribution.

$$j^* = \operatorname{argmax} \{W_1, \dots, W_J\}, \quad (21)$$

where  $W_{j^*}$  for an individual  $i$  is defined as the value of choosing occupation  $j^*$  at age 0:

$$W_{j^*} = \mathbb{E}_0 \{V_{j^*}(x) | \Omega_{i,0}\}. \quad (22)$$

When choosing an industry,  $\Omega_{i,0}$  (the vector of abilities drawn at birth) is in an individual's information set, thus appearing to the right of the conditioning sign. The individual also knows the statistical properties of shocks experienced by workers in each industry. As a result, 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 equation (21) induces an endogenous distribution of workers across industries. Let  $\pi_j$  denote the mass of workers in industry  $j$  with  $\sum_{j=1}^J \pi_j = 1$ .

**Firms** Our model economy can be pictured as a small, open economy consisting of a set of islands in which each island represents an industry. Each industry produces a consumption good according to the following industry-level technology:

$$Y_j = N_j^{\alpha_j}, \quad (23)$$

where  $Y_j$  is the output of industry  $j$ ,  $N_j$  represents the labor input of that industry measured in efficiency units,<sup>14</sup> and  $\alpha$  is the share of labor in output (with  $\alpha < 1$ ). Firms are owned by foreigners who operate them, pay wages, and enjoy profits. We do not consider any kind of inter-industry trade in goods, so readers may assume that goods produced across islands are identical.<sup>15</sup>

---

<sup>14</sup>The measure of efficiency takes into account both the time-varying productivity component and the industry-specific abilities.

<sup>15</sup>Alternatively, one can picture  $J$  different goods and assume that an individual working in industry  $j$  obtains utility from consuming only the good produced in that industry, and not those from other industries.

**Equilibrium** We can now define a stationary competitive equilibrium that consists of (i) a set of industry wages  $\{w_j\}_{j=1}^J$ , (ii) industry populations (or masses)  $\{\pi_j\}_{j=1}^J$ , (iii) industry-specific distributions  $\{\Psi_j(x)\}_{j=1}^J$ , (iv) industry-level efficiency-weighted employment levels  $\{N_j\}_{j=1}^J$ , (v) an optimal industry choice  $j^*$ , and (vi) industry-specific decision rules  $\{b'_j(x), c_j(x)\}_{j=1}^J$  and associated value functions  $\{V_j(x)\}_{j=1}^J$  that 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 (18) yielding value functions  $\{V_j(x)\}_{j=1}^J$ . In addition,  $j^*$  is the industry that maximizes lifetime utility at time zero.
2. Given our stationarity assumption the set of industry-specific populations  $\{\pi_j\}_{j=1}^J$  are constant and the distributions of abilities across industries are consistent with the optimal industry choice (21). For any given industry  $j$ , its population satisfies  $\pi_j = \text{Prob}(W_j > W_{-j})$ , where we define the vector  $W_{-j}$  as equal to  $\{W_1, \dots, W_{j-1}, W_{j+1}, \dots, W_J\}$ . The cumulative distribution of  $\theta_j$  in a given industry  $j$  is defined by

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

where  $\Theta_j$  is the support of  $\theta_j$  and  $\Theta_{-j}$  is the support of  $\theta_{-j}$ , and  $\chi_{\{\theta_j: W_j > W_{-j}\}}$  is an indicator function that takes the value 1 when an individual with ability  $\theta_j$  chooses industry  $j$ . Finally,  $F(\theta_j)$  is the cumulative distribution function of  $\theta_j$  before agents sort.

3. Wages in industry  $j$  are equal to the marginal product of a marginal unit of average efficiency in that industry, denoted as

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

4. Labor markets clear at the industry level,

$$N_j = \pi_j \int_S e^{\theta_j} e^{\eta} e^{\omega} d\Psi_j(x).$$

5. In a given industry  $j$ ,  $\Psi_j(x)$  is the stationary distribution associated with workers' abilities the transition function implied by the optimal decision rule  $b'_j(x)$  and the law of motion for the exogenous shocks  $\eta$  and  $\omega$ .
6. For each industry  $j$ , the following resource constraint is satisfied:

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

Some of these equilibrium conditions do not require further explanation. However, conditions (2) and (6) need some clarification. Condition (2) summarizes the post-sorting distribution of workers across industries. This distribution is characterized by a cumulative distribution function  $G_j(\tilde{\theta})$  representing the fraction of workers working in industry  $j$  with an ability level for that industry  $\theta_j$  lower than  $\tilde{\theta}$ . The numerator in the definition of  $G_j$  gives the conditional probability that a worker with a  $\theta_j$  lower than  $\tilde{\theta}$ , and a vector  $\theta_{-j}$  of ability levels in all other industries, chooses industry  $j$ . The denominator gives the unconditional probability of a worker choosing industry  $j$ .

Condition (6) gives the industry-level resource constraint. Recall that industries are small open economies, able to borrow or lend from abroad at an exogenous and fixed net rate  $r$ . As a result, at the industry level total labor earnings must equal consumption plus the net asset position of that industry.<sup>16</sup>

## 4 Quantitative Analysis

This section presents the quantitative analysis. For this purpose, we use the theoretical model developed in the previous section computed and calibrated to mimic the US economy.<sup>17</sup>

It is informative to relate workers' earnings in the model to the fixed effects regression estimated from the SIPP. Earnings for an individual  $i$  working in industry

---

<sup>16</sup>Multiplying both sides of (19) by  $\pi_j$ , and integrating the left- and right-hand sides of that same equation with respect to  $\Psi_j(x)$ , gives equilibrium condition (6).

<sup>17</sup>The online appendix includes details about the procedure used to compute the model.

$j$  at time  $t$  are denoted by

$$y_{ijt} = w_j e^{\theta_{ij}} e^{\eta_{ijt}} e^{\omega_{ijt}} \quad (24)$$

Taking logs in the definition of earnings yields

$$\log(y_{ijt}) = \log(w_j) + \theta_{ij} + \eta_{ijt} + \omega_{ijt}. \quad (25)$$

Note that earnings are written in a form identical to that of estimated equation (1) (except for the exogenous observables  $\mathbf{X}$ ). The model counterpart to the estimated fixed effect  $\alpha_{ij}$  is given by  $\log(w_j) + \theta_{ij}$ . That object is a convolution of two endogenous and unobserved components: the price of human capital in industry  $j$ ,  $w_j$ , and individual  $i$ 's industry-specific ability  $\theta_{ij}$ . Note that the estimated reduced-form regression does not allow differentiating between the two, and hence it is not possible to know how much of the difference in average fixed effects across industries is driven by comparative advantage. The structural model makes this decomposition transparent.

## 4.1 Parameter Values

The model period is set equal to one quarter and the total lifetime for an individual is 120 periods. These two values correspond to a 30-year employment history. We exogenously set the annual interest rate to be 3.5%. We set  $\underline{b} \geq 0$  and choose  $\beta$  equal to 0.974 in order to match an aggregate (annual) wealth-to-income ratio of 3. We restrict preferences to be of the constant relative risk aversion class with the coefficient of risk-aversion equal to 3.<sup>18</sup> In addition, we need to assign values for the parameters that govern returns to scale at the industry level,  $\alpha_j$ 's. These parameters represent labor's share of total revenue in each of the industries. Following Hopenhayn and Rogerson (1993), who use the same decreasing returns to scale technology, we use National Income and Product Accounts (NIPA) data to determine their values. Specifically, we use data on compensation of full-time equivalent employees and gross domestic product (GDP) at the industry level from the NIPA to set the labor share of each industry. Table 4 reports the calibrated labor

---

<sup>18</sup>In the appendix we present the results of a sensitivity analysis in which we consider different value for this parameter.

share for each industry.<sup>19</sup>

One of the driving forces of a non degenerate wage distribution across industries is an industry-specific level of risk. We use the estimates for the variances of the two components of income we estimate from the SIPP in Section 2 as a measure of this risk. Hence, we set  $J$ , the total number of industries, to 19 and feed the model with the estimated values of the variances of both the permanent and transitory shocks reported in the second and fourth columns of Table 1.

Finally, we need to parameterize the distributions of pre-labor-market skills or abilities; that is we need to find values for the 38 parameters  $\{\mu_{j,\theta}, \sigma_{j,\theta}^2\}_{j=1}^{j=19}$ . To do so, we pick values for these parameters so the model approximates the cross-sectional means and standard deviations of the distribution of net earnings for each of the 19 industries (columns 2 and 3 of Table 5). The use of net earnings is justified because 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 (11)). The resulting parameter values are shown in Table 6. Two assumptions identify the means and variances of the distributions of abilities. The first is the assumption of independence in abilities across industries.<sup>20</sup> The second assumption is that only pecuniary considerations affect the career choice.

## 4.2 Selection versus Risk

Solving the model for the set of parameter values just described delivers an equilibrium distribution of earnings within and across industries. In equilibrium, individuals' earnings depend on the industry wage, ability level, and realization of the earnings shocks. Since we assume that shocks to labor earnings have zero

---

<sup>19</sup>There are differences between the SIPP and NIPA regarding industry classification. For this reason, we reclassify a few industries to make the classification comparable. This is the case for Finance in the SIPP, which corresponds to Finance and Insurance/Real Estate and Rental and Leasing in NIPA. Business services (SIPP) corresponds to Professional, Scientific, and Technical Services plus Management of Companies and Enterprises plus Administrative and Waste Management Services (NIPA). Recreation and Entertainment in SIPP corresponds to Arts, Entertainment, and Recreation plus Accommodation and Food Services in NIPA. Medical in SIPP corresponds to Ambulatory Health Care Services in NIPA. Social Services in SIPP corresponds to nursing and residential care facilities plus social assistance in NIPA. In addition, personal services is not an industry in NIPA so it is assigned the labor share of business services.

<sup>20</sup>In the appendix we analyze the effects of relaxing this assumption.



mean, in equilibrium average earnings within an industry depend exclusively on the average ability level and the wage rate for that industry. However, insofar as the variances of the shocks affect the sorting of risk-averse workers, they influence the equilibrium distribution of earnings across industries as well.

Our goal is to link our model results with the findings of the econometric analysis in Section 2. In our model, there are 19 industries as in the SIPP, but since the model is stationary, it is unclear how to map an empirical regression with 57 left-hand-side observations (19 per SIPP panel) into one with only 19. We choose to use the weighted averages across panels of average (net) earnings and standard deviations as our left-hand-side and right-hand-side variables, respectively. In other words, we regress average net earnings by industry on the two measures of volatility after taking time averages across the three panels. The coefficients of this regression are presented in the second column of Table 7. The model counterpart regresses average earnings by industry obtained by using our calibrated model on the standard deviations of earnings shocks obtained in our econometric analysis. The result of this benchmark case is presented in the third column of Table 7.

Since the parameters of the distributions of the pre-labor-market skills are chosen such that the model approximates the mean of net earnings of each for the 19 industries, the model delivers the positive relationship between mean earnings and the standard deviations of the transitory and permanent shocks to earnings. This is clear when comparing columns 2 and 3 in Table 7. By construction, the coefficients are fairly close (6.23 versus 6.54) in the case of the permanent shock and almost equal (5.86 versus 5.84) for the case of the transitory shock.

As noted previously, the documented strong and positive relationship between mean earnings and their standard deviations suggests the existence of a risk premium in the US labor market. However, the presence of individual specific pre-labor-market skills affects the sorting of individuals into the industries of the economy, and hence changes the interpretation of the regression coefficient. In other words, we may be assigning all of the observed differences in mean earnings to compensation for risk while in fact they are compensation for the skills of individuals. Fortunately, our theory is rich enough to allow for a decomposition of the

correlation between compensation for risk and for unobserved abilities. For this purpose, we perform a counterfactual experiment in which we set all individuals' abilities to 1. Hence, there are no ex-ante differences across individuals, and the sorting across industries is driven entirely by differences in risk.

For risk-averse workers, 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 cannot be an equilibrium allocation because industry-level technologies display decreasing returns to scale. As a result, the more workers within an industry, the lower the wages and viceversa. The resulting equilibrium features relatively safe industries with low wages and a large mass of workers. Riskier industries display the opposite characteristics.

What does the model predict regarding the correlation between mean earnings and their volatility in this counterfactual environment? Table 7 shows the correlation between the mean earnings predicted by the model and the standard deviation of the permanent and transitory shocks (column 4). Two important results emerge: First, in the counterfactual exercise the coefficient associated with the permanent shock is 4.1. This coefficient points to a strong and positive association between this counterfactual mean earnings and the standard deviation of the permanent shock to earnings. Moreover, comparing the coefficients associated with the permanent shock in columns 3 and 4 (4.13 in the counterfactual case versus 6.54 in the benchmark) it is clear that the compensation for unobserved abilities appears to be compensation for risk but it is not. Nevertheless, the underlying compensation for permanent risk in our environment is still strong as it is implied by the correlation between earnings and permanent risk. However, the coefficient corresponding to the transitory shock reduces to only  $-0.02$ , whereas in the benchmark case this coefficient is 5.84. In light of the result of this counterfactual experiment, what appears to be compensation for risk associated with transitory shocks (as suggested by the estimated correlation of the econometric model) is mostly compensation for the unobserved abilities of individuals. One interpretation of this result is that by saving in one-period bonds, workers can smooth transitory shocks to labor earn-

ings quite well. Hence, they do not need to be compensated for bearing that type of risk in the labor market.<sup>21</sup>

As mentioned when calibrating the model we pick the parameters of the ability distributions such that the model delivers the mean and variance of earnings of the cross section of earnings in each of the industries. Thus, by construction, our benchmark economy gets close to generating the correlation between earnings and standard deviations of the permanent and transitory shocks (see columns 2 and 3 of Table 7). For this reason, this result cannot be interpreted as an independent test for the model. Nevertheless, the model does well in predicting the sorting of workers into the different industries we observe in the data: the correlation between the model and the data is 0.3. More importantly, in the counterfactual exercise in which we equalize individual's abilities, this correlation is -0.3. This result can be viewed as validating the mechanism proposed in the model.

In our benchmark economy, we further investigate what drives the correlation between mean earnings and permanent and transitory risk. As mentioned above, after the sorting of individuals into industries, mean earnings in a particular industry is the result of the product of the mean abilities of workers and the equilibrium wage rate. We wish to decompose the product of those two components for each industry and to do so, we proceed to estimate two additional regressions. First, we regress (log) average ability levels across industries on permanent and temporary standard deviations. Second, we regress (log) wage rates across industries on the two standard deviations of shocks. Both experiments help illustrate the relative roles of the amount of human capital (abilities) versus the unit price of human capital (the wage rate) in explaining the overall correlation between average earnings and our two measures of risk.

Table 8 summarizes the result of the separation of total earnings between an ability and a wage component. The second and third columns show the coefficients of regressing either variable in the first row, on the standard deviations of the permanent and the transitory shocks. The coefficients associated with mean

---

<sup>21</sup>One could argue that in the real world some workers do not participate in the financial market and therefore do not have access to the one-period risk-free bond included in our model. In that case, transitory risk would recover some of its importance.

earnings are those corresponding to our benchmark economy (shown in Table 7). The second and third row display the coefficients associated with abilities (both are positive: 11.99 and 5.16) and those associated with wages rates (-8.49 and 0.37). Our interpretation is that a high ability acts as an insurance device for workers, allowing them to work in relatively riskier industries. Regarding the wage rates, there are two forces that determine them in equilibrium. First, risk averse workers will try to avoid risky industries, so these industries will have fewer of them and given our assumption of decreasing returns at the industry level, wages will be relatively higher. However, as just mentioned, more able workers self-select into riskier industries and that lowers wages in those industries. This second force dominates in the case of the permanent shock and that is responsible for the negative coefficient we find between the wage rates and our measures of volatility.

### **4.3 Implications for Wealth and Income Inequality**

The model delivers predictions regarding the amount of wealth workers accumulate to smooth consumption when they face shocks to labor earnings. Everything else constant, the riskier the industry in which an individual works, the more she saves. Nonetheless, average ability levels also affect precautionary savings: two individuals facing the same variance of shocks but having different ability levels display different savings behavior. Since the abilities distributions help the model match average earnings in the data by industry, it is not necessarily true that the relationship between risk and savings is monotonic. However, across all 19 industries the relationship between risk and wealth-to-income ratios is positive on average. The model predicts that the mean wealth-to-income ratio of individuals working in Construction (the riskiest industry) is higher than that of those working in Government. Figure 3 shows the relationship between average wealth-to-income ratios for each industry in two model economies. The benchmark economy is represented by blue circles and a counterfactual economy in which all industries feature the same income risk is represented by red circles. Each circle represents a pairing of the industry wealth-to-income ratio and the standard deviation of the permanent shock. As is clear from the figure, the benchmark economy delivers a

strong and positive correlation between wealth and permanent risk. Overall, our analysis of risk and precautionary savings appears to confirm the findings of Carroll and Samwick (1997): Individuals who work in risky industries save more than those working in safer industries. However, note that although the red circles are much more concentrated than the blue circles they do not concentrate in one point showing that some portion of wealth inequality can be attributed to differential abilities across industries. This is the result of the sorting of workers with different abilities which makes them differ in their saving behavior.

We now analyze the model's implications for labor income inequality. The model is rich enough to allow for the separation of cross-industry inequality into an ability component versus a wage component. Table 9 shows the decomposition. It displays the standard deviation of: i) the log of earnings, ii) the log of average abilities, and iii) the log of wage rates across industries. Earnings across industries are roughly half as smooth (a standard deviation of 0.11) as either abilities (0.26) or wages (0.24).

In addition, we also study income inequality. Table 10 reports Gini coefficients and standard deviations for the distribution of earnings in alternative environments. We report both overall inequality and inter-industry inequality. We examine four alternative environments. "Benchmark" refers to the parameterization described previously: Agents are ex ante and ex post heterogeneous, and the degree of earnings uncertainty and aggregate technology vary across industries. In the second case, "Equal Risk", the difference with respect to the benchmark is that it equates uncertainty in earnings risk across industries (we set it to the cross-industry average). The last two environments eliminate any ex-ante heterogeneity across individuals. The "No *Ex-Ante* Heterogeneity" economy corresponds to the counterfactual described in the previous section: The abilities of all individuals in all industries are set to 1 and industries vary by uncertainty in earnings and in their labor shares. The last environment also eliminates any differences in risk across industries.

The first two columns of Table 10 show that overall inequality is driven almost entirely by shocks with little driven by initial earnings differences. The Gini coef-

efficient in the economy with no ex ante heterogeneity and equal variances across industries is 0.33.<sup>22</sup> The Gini rises in the benchmark to 0.38, an increase of 15%. Equating risk across industries does not appear to affect inequality that much. The last two columns of 10 show the same statistics for inter-industry earnings inequality. It is clear from the table that cross-industry differences in risk and abilities contribute almost equally to explaining the observed earnings inequality across industries.

## 5 Concluding Remarks

This paper documents a strong and positive correlation between the mean and the standard deviation of both transitory and permanent shocks to labor earnings across industries of the US economy. To properly determine the compensation for risk in the labor market, we develop an equilibrium model of career choice with incomplete markets in which risk-averse and heterogeneous individuals choose an industry in which to supply their labor services. Some industries are riskier than others and workers are endowed with an industry-specific ability. When mapped to the data, the model assigns almost no compensation for transitory shocks. Conversely, the higher earnings observed in industries with higher permanent risk reflect a pure compensation for risk. Furthermore, according to our model workers accumulate different levels of wealth depending on the industry of employment. More importantly, compensation for risk explains a sizable fraction of observed cross-industry earnings inequality.

An important aspect of the framework proposed in this study is the interplay between risk and comparative advantages in shaping individuals' career choices. According to our results wealth and income inequality are partly the outcome of workers' career choices. For this reason, the proposed mechanisms should be incorporated in the analysis of a wide range of policies studied in macroeconomics, public finance, and labor economics. For instance, on the effect of policies targeted to modify initial conditions and those directed at shocks over an individual's work-

---

<sup>22</sup>Measuring inequality using the Gini coefficient or using the standard deviation of log earnings yields the same inference. We nonetheless report both measures in the table.

ing lifetime.

Finally, our framework does not address several forces that shape both individuals' skills and their career choices, our results may help to motivate work to shed light on these forces. A short (but by no means exhaustive list) would include (i) the consideration of industry-specific human capital accumulation to better understand how individuals build their comparative advantages, (ii) the inclusion of a richer asset market structure as well as the analysis of workers participation in such markets to better understand the possibilities of self-insurance against industry-specific earnings shocks, and (iii) the career choice of financially constrained individuals.

## References

- Abowd, J., and M. Stinson (2011): "Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Bureau Survey and SSA Administrative Data," Technical Papers 01/2002 11-20, Center for Economic Studies, U.S. Census Bureau.
- Abowd, J. M., and O. C. Ashenfelter (1981): "Anticipated Unemployment, Temporary Layoffs, and Compensating Wage Differentials," in *Studies in Labor Markets*, Sherwin Rosen, ed., pp. 141–186. National Bureau of Economic Research.
- Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, 109(3), 659–84.
- Antonovics, K., and L. Golan (2012): "Experimentation and Job Choice," *Journal of Labor Economics*, 30(2), 333 – 366.
- Bewley, T. (1977): "The Permanent Income Hypothesis: A Theoretical Formulation," *Journal of Economic Theory*, 16(2), 252–292.
- Buera, F. J. (2006): "Non-parametric Identification and Testable Implications of the Roy Model," Discussion paper, Northwestern University.
- Carroll, C. D., and A. A. Samwick (1997): "The Nature of Precautionary Wealth," *Journal of Monetary Economics*, 40(1), 41–71.
- CEPR (2014): "Center for Economic and Policy Research. 2014. SIPP Uniform Extracts, Version 2.1.7. Washington, DC.," .
- Chen, S. H. (2008): "Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity," *The Review of Economics and Statistics*, 90(2), 275–289.
- Cunha, F., J. Heckman, and S. Navarro (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57(2), 191–261.
- Dillon, E. W. (2016): "Risk and Return Tradeoffs in Lifetime Earnings," Discussion paper, Arizona State University, Department of Economics.

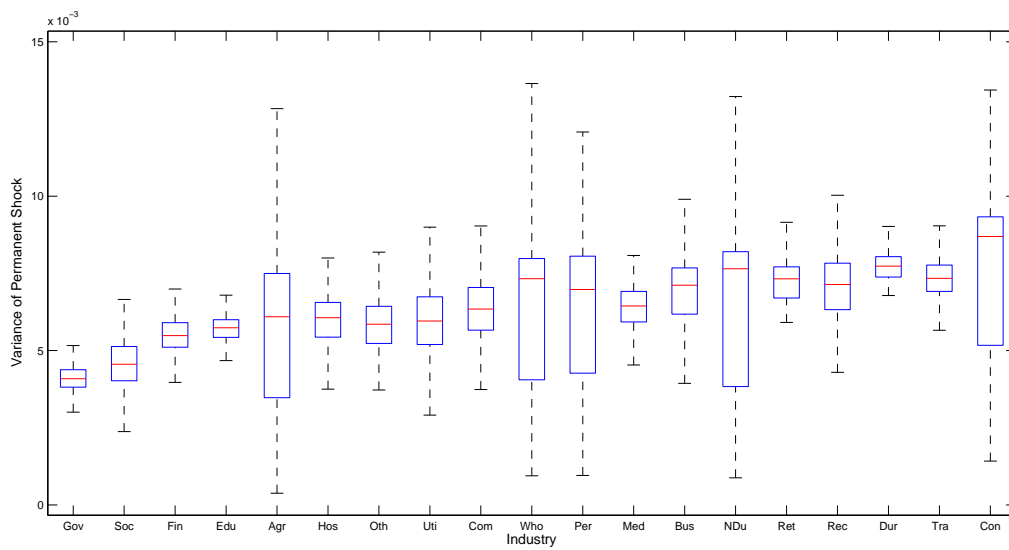


- Feinberg, R. M. (1981): "Earnings-Risk as a Compensating Differential," *Southern Economic Journal*, 48(1), 156–163.
- Gathmann, C., and U. Schönberg (2010): "How General Is Human Capital? A Task-Based Approach," *Journal of Labor Economics*, 28(1), 1–49.
- Gottschalk, P. T., and M. Huynh (2006): "Are Earnings Inequality and Mobility Overstated? The Impact of Non-Classical Measurement Error," IZA Discussion Paper No. 2327, Institute for the Study of Labor (IZA).
- Guvenen, F. (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.
- Heckman, J., L. Lochner, and C. Taber (1998): "Explaining Rising Wage Inequality: Explanations With A Dynamic General Equilibrium Model of Labor Earnings With Heterogeneous Agents," *Review of Economic Dynamics*, 1(1), 1–58.
- Heckman, J. J., and B. E. Honore (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), 1121–49.
- Hoffmann, F. (2010): "An Empirical Model of Life-Cycle Earnings and Mobility Dynamics," Discussion paper, University of British Columbia, Department of Economics.
- Hopenhayn, H., and R. Rogerson (1993): "Job Turnover and Policy Evaluation: A General Equilibrium Analysis," *Journal of Political Economy*, 101(5), 915–38.
- Huggett, M. (1993): "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," *Journal of Economic Dynamics and Control*, 17(5-6), 953–969.

- Huggett, M., G. Ventura, and A. Yaron (2011): "Sources of Lifetime Inequality," *American Economic Review*, 101(7), 2923–54.
- Kambourov, G., and I. Manovskii (2009): "Occupational Mobility and Wage Inequality," *Review of Economic Studies*, 76(2), 731–759.
- Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.
- Kuznets, S., and M. Friedman (1954): "Incomes from Independent Professional Practice, 1929-1936," in *Incomes from Independent Professional Practice, 1929-1936*, NBER Books. National Bureau of Economic Research.
- 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.
- Neal, D. (1999): "The Complexity of Job Mobility among Young Men," *Journal of Labor Economics*, 17(2), 237–61.
- Neumuller, S. (2015): "Inter-industry wage differentials revisited: Wage volatility and the option value of mobility," *Journal of Monetary Economics*, 76(C), 38–54.
- Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.
- 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.

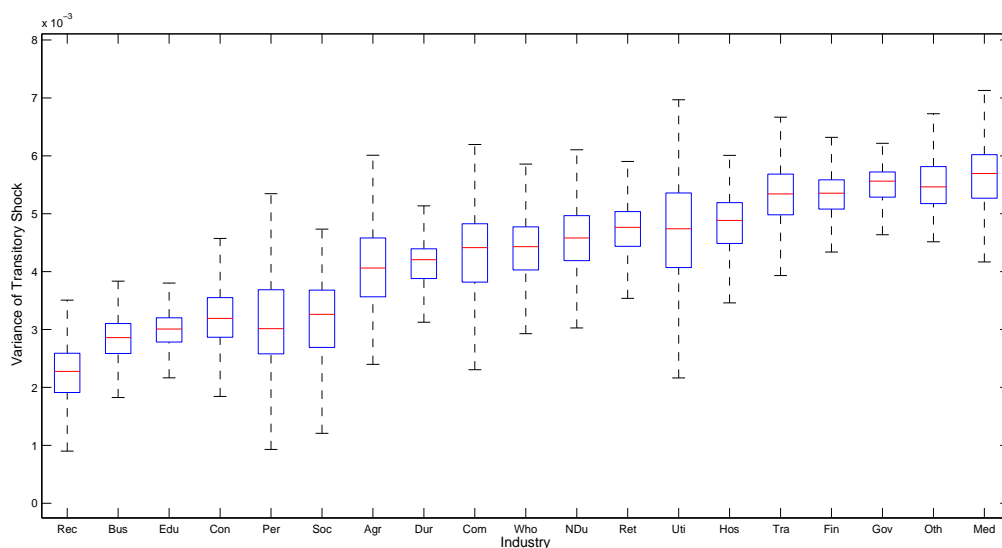
# Figures

Figure 1: Boxplot for the Variances of the Permanent Shock to Earnings<sup>a</sup>



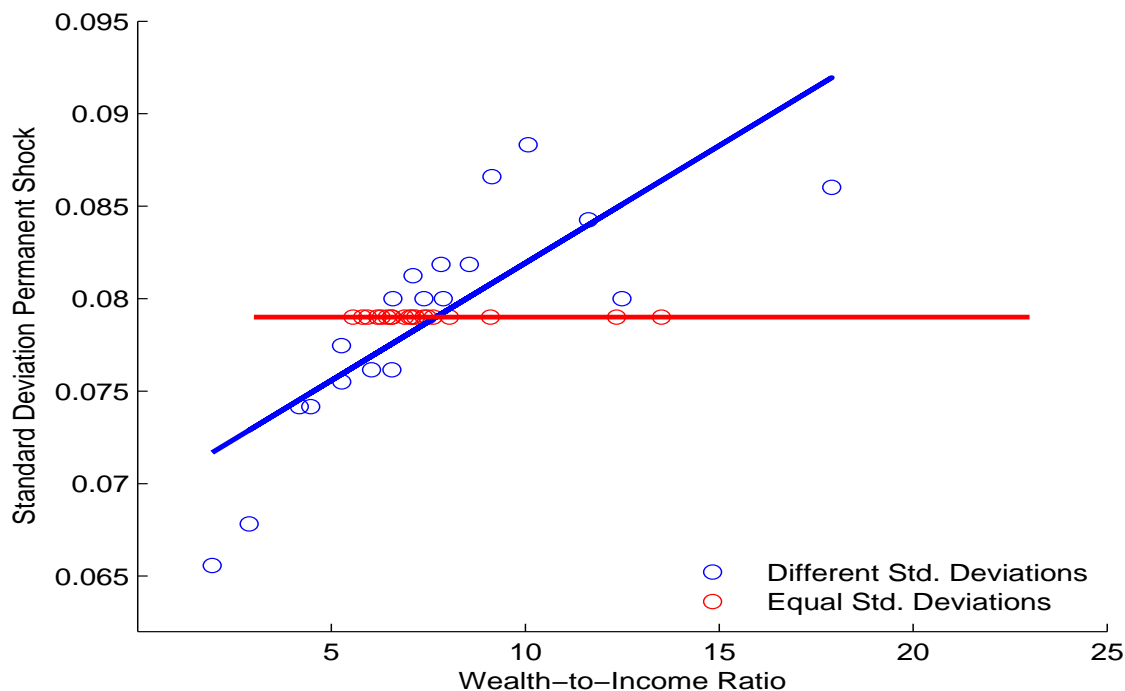
<sup>a</sup> Soc is social services, Agr is agriculture and forestry, Bus is business services, Edu is education, NDU is nondurable goods manufacturing, Rec is recreation and entertainment, Fin is finance, Ret is retail trade, Gov is government, Per is personal services, Com is communication, Uti is utilities, Con is construction, Who is wholesale trade, Tra is transportation, Oth is other services, Med is medical services, Dur is durable goods manufacturing and Hos is hospitals.

Figure 2: Boxplot for the Variances of the Transitory Shock to Earnings <sup>a</sup>



<sup>a</sup> Soc is social services, Agr is agriculture and forestry, Bus is business services, Edu is education, NDu is nondurable goods manufacturing, Rec is recreation and entertainment, Fin is finance, Ret is retail trade, Gov is government, Per is personal services, Com is communication, Uti is utilities, Con is construction, Who is wholesale trade, Tra is transportation, Oth is other services, Med is medical services, Dur is durable goods manufacturing and Hos is hospitals.

Figure 3: Wealth-to-Income Ratios



Note: Each circle (red or blue) represents an industry. The blue line represents the fitted values of regressing the wealth-to-income ratio on the variance of permanent shocks. The red line represents the fitted values of regressing the wealth-to-income ratio on the variances of permanent shocks in the counterfactual experiment in which the variances of permanent shocks to earnings are equal across industries.

# Tables

Table 1: Variance By Industry - Permanent and Transitory

	$\sigma_{\epsilon,j}^2$	Ranking	$\sigma_{\eta,j}^2$	Ranking	$\lambda_j$	$\phi_j$	Number Individuals
1 Agriculture and Forestry	0.0057 (0.0024)	5	0.0040 (0.0009)	7	0.42 (0.23)	0.81 (0.06)	137
2 Construction	0.0078 (0.0023)	19	0.0031 (0.0007)	4	0.16 (0.19)	0.93 (0.02)	756
3 Durable Goods Manufacturing	0.0074 (0.0015)	17	0.0040 (0.0006)	8	0.10 (0.17)	0.91 (0.01)	2175
4 Non Durable Goods Manufacturing	0.0067 (0.0023)	14	0.0045 (0.0008)	11	0.21 (0.20)	0.90 (0.02)	1218
5 Transportation	0.0075 (0.0011)	18	0.0052 (0.0009)	15	0.17 (0.14)	0.86 (0.02)	758
6 Communication	0.0064 (0.0012)	9	0.0041 (0.0011)	9	0.10 (0.19)	0.89 (0.02)	385
7 Utilities	0.0060 (0.0016)	8	0.0047 (0.0011)	13	0.15 (0.17)	0.83 (0.04)	263
8 Wholesale Trade	0.0064 (0.0023)	10	0.0043 (0.0008)	10	0.16 (0.20)	0.90 (0.02)	687
9 Retail Trade	0.0067 (0.0017)	15	0.0046 (0.0007)	12	0.13 (0.19)	0.89 (0.02)	1472
10 Finance	0.0055 (0.0012)	3	0.0052 (0.0009)	16	0.04 (0.12)	0.88 (0.02)	1089
11 Business Services	0.0066 (0.0017)	13	0.0027 (0.0006)	2	0.37 (0.22)	0.94 (0.02)	932
12 Personal Services	0.0064 (0.0023)	11	0.0031 (0.0008)	5	0.34 (0.22)	0.84 (0.04)	232
13 Recreation and Entertainment	0.0071 (0.0011)	16	0.0022 (0.0006)	1	0.11 (0.19)	0.90 (0.02)	327
14 Hospitals	0.0058 (0.0014)	6	0.0047 (0.0007)	14	0.31 (0.18)	0.80 (0.03)	967
15 Medical Services	0.0064 (0.0012)	12	0.0056 (0.0008)	19	0.07 (0.13)	0.85 (0.03)	758
16 Education	0.0055 (0.0012)	4	0.0029 (0.0005)	3	0.11 (0.18)	0.93 (0.01)	1914
17 Social Services	0.0046 (0.0010)	2	0.0032 (0.0006)	6	0.16 (0.22)	0.84 (0.03)	450
18 Other Services	0.0058 (0.0017)	7	0.0054 (0.0009)	18	0.20 (0.13)	0.83 (0.03)	482
19 Government	0.0043 (0.0010)	1	0.0053 (0.0008)	17	0.08 (0.14)	0.79 (0.03)	1590

Note: The permanent variance is  $\tilde{\sigma}_{\epsilon,j}^2(1 - \lambda_j)$  and the transitory variance is  $\tilde{\sigma}_{\eta,j}^2(1 - \phi_j)$ . Bootstrap standard errors are shown in parentheses. The column Ranking just ranks the industries according to their estimate of the variance.

Table 2: Permanent and Transitory Shocks Across Industries

		Transitory shock		
		Low	Medium	High
Permanent shock	Low	17,16,1	14	10,19
	Medium	11,12	6,8	7,15,18
	High	2,13	3,4,9	5

Note: This table shows the classification of industries across six dimensions according to the value of the variance of both the transitory (vertical) and permanent (horizontal) shocks into terciles. For instance, an industry is classified into the low permanent shock and high temporary shock is an industry in which its estimated value for the permanent shock and transitory shock are in the bottom and top tercile, respectively. The industries are represented by their numbers as defined in Table 1

Table 3: Regression Results - Permanent and Transitory

Dependent Variable	Earnings Coefficient	Net Earnings Coefficient
Intercept	2.513 (0.033)	7.008 (0.000)
Female	-0.487 (0.000)	
Age	0.263 (0.000)	
Age. Sq.	-0.003 (0.000)	
Educ.	0.241 (0.000)	
$\sigma_\epsilon$	2.270 (0.008)	2.078 (0.008)
$\sigma_\eta$	0.869 (0.042)	1.149 (0.018)
I <sub>96</sub>	-0.176 (0.003)	0.043 (0.273)
I <sub>01</sub>	-0.180 (0.010)	-0.204 (0.003)

The second column shows the estimation results of regressing log earnings by industry on 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 and on our estimates for the standard deviations of the permanent shock and transitory shocks to labor earnings ( $\sigma_\epsilon$  and  $\sigma_\eta$ , respectively). For positive (negative) coefficients the values in parentheses show the probability that the coefficient is less (bigger) than zero computed by bootstrap.

Table 4: Labor Shares

Industry no.	Industry	Labor Share
1	Agriculture and Forestry	0.29
2	Construction	0.66
3	Durable goods manufacturing	0.69
4	Nondurable goods manufacturing	0.48
5	Transportation	0.65
6	Communication	0.45
7	Utilities	0.26
8	Wholesale trade	0.53
9	Retail trade	0.58
10	Finance	0.24
11	Business services	0.70
12	Personal services	0.70
13	Recreation and entertainment	0.61
14	Hospitals	0.89
15	Medical services	0.75
16	Education	0.88
17	Social services	0.81
18	Other services	0.65
19	Government	0.84

Note: The table presents the labor shares calculated by dividing the compensation of full-time equivalent employees by the GDP of each industry in the period 1990-2009 using the NIPA by industry provided by the Bureau of Economic Analysis.



Table 5: Mean and Dispersion of Earnings by Industry

Industry no.	Industry	Mean Earnings	Std. Dev. Earnings
1	Agriculture and Forestry	7.06	0.33
2	Construction	7.26	0.34
3	Durable Goods Manufacturing	7.24	0.32
4	Non Durable Goods Manufacturing	7.20	0.34
5	Transportation	7.31	0.33
6	Communication	7.35	0.33
7	Utilities	7.36	0.30
8	Wholesale Trade	7.21	0.35
9	Retail Trade	7.07	0.37
10	Finance	7.26	0.35
11	Business Services	7.23	0.37
12	Personal Services	7.07	0.37
13	Recreation and Entertainment	7.07	0.40
14	Hospitals	7.21	0.35
15	Medical Services	7.12	0.34
16	Education	7.08	0.31
17	Social Services	7.01	0.34
18	Other Services	7.14	0.39
19	Government	7.26	0.33

Note: This table shows the earnings statistics of the 19 industries included in our sample. It contains the log of mean earnings and the standard deviation of (log) earnings.

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

Industry no.	Industry	$\mu_{j,\theta}$	$\sigma_{j,\theta}$
1	Agriculture and Forestry	7.37	0.65
2	Construction	7.41	0.72
3	Durable Goods Manufacturing	8.98	0.63
4	Non Durable Goods Manufacturing	7.54	0.73
5	Transportation	7.24	0.76
6	Communication	7.23	0.84
7	Utilities	7.40	0.87
8	Wholesale Trade	7.73	0.75
9	Retail Trade	7.48	0.63
10	Finance	7.22	0.81
11	Business Services	7.20	0.75
12	Personal Services	8.46	0.55
13	Recreation and Entertainment	8.06	0.59
14	Hospitals	7.48	0.76
15	Medical Services	7.50	0.66
16	Education	7.14	0.70
17	Social Services	7.29	0.66
18	Other Services	7.82	0.72
19	Government	7.29	0.86

Note: This table shows the calibrated values for the mean ( $\mu_{j,\theta}$ ) and standard deviations ( $\sigma_{j,\theta}$ ) of the distribution of pre-labor-market skills for the 19 industries considered.

Table 7: Model Predictions: Mean and Volatility of Earnings

Variable	Data Coefficient	Benchmark Coefficient	Counterfactual Coefficient
Permanent $\sigma_\epsilon$	6.23	6.54	4.13
Transitory $\sigma_\eta$	5.86	5.84	-0.02

Note: The second column shows the estimation results of regressing log mean earnings by industry (weighted mean by panel of SIPP) on the standard deviations of permanent and transitory shocks in the benchmark model. The third (fourth) column presents the estimation results of the same regression but using the mean earnings predicted by the model both in the benchmark case (counterfactual case) described in the text.

Table 8: Earnings Decomposition: Abilities vs. Wage Rates

Dependent Variable	Mean Earnings Coefficient	Mean Abilities Coefficient	Wage Rate Coefficient
Permanent $\sigma_\epsilon$	6.54	11.99	-8.49
Transitory $\sigma_\eta$	5.84	5.16	0.37

Note: The table shows the coefficient estimated when regressing mean earnings (second column), mean abilities (third column) and wage rates (fourth column) on the standard deviations of the permanent and transitory shocks. The second column is the benchmark case and the dependent variable of the other two regressions comes from the decomposition of earnings using the model's output as described in the text.

Table 9: Sources of Inter-Industry Inequality

	Standard Deviation
Log Mean Earnings	0.11
Log Mean Abilities	0.26
Log Wage Rates	0.24

Note: The table shows the standard deviation of (log) mean industry earnings, abilities and wage rates calculated by using the model's output as described in the text.

Table 10: Income Inequality

Environment	Overall		Inter-Industry	
	Gini	Std. Dev.	Gini	Std. Dev.
Benchmark	0.382	0.710	0.059	0.106
Equal Risk	0.379	0.701	0.058	0.105
No ex ante Heterogeneity	0.329	0.611	0.022	0.040
No ex ante Heterogeneity and Equal Risk	0.328	0.610	0.000	0.000

Note: The table shows the Gini coefficients and standard deviation for the whole distribution of earnings (second and third columns) and for mean industry earnings (fourth and fifth columns) for different environments (rows) described in the text.

# Online Appendix (Not for Publication)

## A Data and Sample Selection

The main data source is the Survey of Income and Program Participation (SIPP). More specifically, we use the 1996, 2001, and 2004 panels. The actual files used are those maintained by the think-tank Centre for Economic and Policy Research (CEPR), in Washington DC.<sup>23</sup> Data files are divided by topic. We merge selected variables from the following data extracts:<sup>24</sup>

- A (“ID and Weights”): month, year, and age.
- B (“Demographics”): education, marital status, ethnicity, gender.
- D (“Employment and Wages”): earnings, employment status, hours, industry, employer size, occupation.
- G (“Income Transfers”): unemployment insurance (state and federal).

Our sample selection procedure can be summarized as follows:

- Restrict individuals to be between the ages of 22 to 66.
- Eliminate those not in the labor force. To assign individuals to either employment, unemployment, or out of the labor force we use the employment status recode (ESR). That variable can take eight possible values. We assign an individual to the unemployment pool if in any given month his/her reply is: a) “with job one or more weeks, spent one or more weeks looking or on layoff”, b) “no job during month, spent entire month looking or on layoff”, or c) “no job during month, spent one or more weeks looking or on layoff”. The out-of-the-labor-force individuals are those who reply “no job during month, no time spent looking or on layoff”. We consider a person employed if they are in the labor force and not unemployed.

---

<sup>23</sup>All the data files are publicly available at <http://ceprdata.org/sipp-uniform-data-extracts/>.

<sup>24</sup>The variables id number, month of reference and wave are common to all files.

- Eliminate those with earnings equal to zero. Our definition of earnings includes only labor earnings for those who are working. Unemployed individuals' earnings are reported receipts of unemployment insurance.
- Eliminate those workers who are employed but who report working zero hours.
- The CEPR constructs a "Major Industry Group" from individuals' responses regarding their employer's industry. There are a total of 23 industries in their classification. The number of industries we employ is 19, which results from a) merging Forestry, Hunting, and Fishing with Agriculture, b) eliminating Armed Forces and Mining for reasons discussed in the text and c) eliminating the Household Services category. We eliminate workers who do not report the industry of employment.<sup>25</sup>
- Restrict analysis to individuals with at least 10 consecutive quarters of responses.
- Eliminate the top and bottom 5% of the distribution of earnings levels. We also restrict earnings (per hour) increases to be smaller than 150% and to not drop by more than 40% for employed individuals.

## **B Industry-Level Risk: Aggregate vs. Idiosyncratic**

Our paper deals with industry-level idiosyncratic risk. It is nonetheless useful to assess the relative importance of that risk vis-a-vis aggregate risk at the industry level (i.e. risk that affects all workers working in a given industry). One way to quantify aggregate industry risk is to examine the variance of the time-dummies in our first-stage individual-level regression. Since the regression is estimated separately for each industry, it yields a set of time dummies for each industry and each panel. We compute the variance of the de-trended time dummies prior to comparing their volatility to the variance in idiosyncratic risk. The reason for the de-trending is that time dummies capture the secular growth in real income and

<sup>25</sup>The reason for eliminating "Household Services" is the very small sample.

Table B.1: Industry-Specific Aggregate Risk vs. Industry-Specific Idiosyncratic Risk

	Aggregate	Idiosyncratic	Ratio
Agriculture and Forestry	0.017	0.108	0.16
Mining	0.028	0.123	0.23
Construction	0.015	0.115	0.13
Durable Goods Manufacturing	0.011	0.114	0.10
Non Durable Goods Manufacturing	0.012	0.115	0.11
Transportation	0.015	0.120	0.12
Communication	0.014	0.120	0.12
Utilities	0.016	0.114	0.14
Wholesale Trade	0.011	0.116	0.10
Retail Trade	0.011	0.117	0.10
Finance	0.012	0.113	0.11
Business Services	0.011	0.113	0.10
Personal Services	0.017	0.118	0.14
Recreation and Entertainment	0.016	0.116	0.14
Hospitals	0.013	0.126	0.11
Medical Services	0.010	0.121	0.08
Education	0.010	0.101	0.10
Social Services	0.015	0.108	0.13
Other Services	0.009	0.108	0.09
Government	0.010	0.107	0.10
Armed Forces	0.021	0.105	0.20

Note: The column labeled *Aggregate* shows the standard deviations of the (de-trended) industry-specific time dummies as explained in the text. The column labeled *Idiosyncratic* displays the standard deviation of residuals by industry, which is of the same order of magnitude as the sum of standard deviations of temporary and permanent shocks shown in Table 1.

that secular growth can't be considered risk. Calculating that standard deviation by industry (and taking averages across the three panels) yields the first column of Table B.1. In most industries, idiosyncratic risk is considerably larger than aggregate risk. The ratio of the latter to the former fluctuates between 8% and 25%, with an average of roughly 11%.

## C Additional Results

In this section we report results for a set of alternative regressions. The purpose is to assess the sensitivity of the relationship between earnings and volatility to alternative measures of earnings.

Tables C.2, C.3, and C.4 report the coefficients from regressing earnings on a set of observables, indicator variables for the different SIPP panels, and the two measures of volatility. In each table the measure of earnings is different. In the first case the left-hand side variable is still total earnings but we impose no bounds on earnings changes (we still eliminate the top and bottom 5% of the distribution of earnings levels). In the second case, Table C.3 we use earnings per hour to bound growth rates between consecutive periods, being the limits (bounds) 60% for decreases and 250% for increases. Finally, Table C.4 reports coefficients when the left-hand side variable in the regression is earnings per-hour. Overall, most of the conclusions drawn in the main body of the text are robust to these alternative measures of earnings. Coefficients are positive and quite tightly estimated. Irrespective of the regression being estimated for earnings or net earnings, the coefficients are large and positive.<sup>26</sup>

---

<sup>26</sup>Not to overwhelm the reader with too many tables, we do not report results using total variance as the regressor measuring volatility. These results are available upon request.

Table C.2: Regression Results - Permanent and Transitory - Unbounded Earnings Changes

Dependent Variable	Earnings Coefficient	Net Earnings Coefficient
Intercept	2.857 (0.033)	6.989 (0.000)
Female	-0.476 (0.000)	
Age	0.237 (0.002)	
Age. Sq.	-0.003 (0.002)	
Educ.	0.248 (0.000)	
$\sigma_\epsilon$	1.949 (0.013)	1.721 (0.023)
$\sigma_\eta$	1.377 (0.015)	1.254 (0.030)
I <sub>96</sub>	-0.193 (0.003)	0.111 (0.058)
I <sub>01</sub>	-0.210 (0.005)	-0.133 (0.060)

The second column shows the estimation results of regressing log earnings by industry on 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 and on our estimates for the standard deviations of the permanent shock and transitory shocks to labor earnings ( $\sigma_\epsilon$  and  $\sigma_\eta$ , respectively). For positive (negative) coefficients the values in parentheses show the probability that the coefficient is less (bigger) than zero computed by bootstrap.

## C.1 Robustness to Excluding Imputed Earnings

The SIPP imputes earnings for those individuals with missing observations. The fraction of imputed earnings is about 20% in the three panels we use. Using imputed values to measure longitudinal changes in earnings may be misleading and



Table C.3: Regression Results - Permanent and Transitory - Wide Bounds on Earnings Changes

Dependent Variable	Earnings Coefficient	Net Earnings Coefficient
Intercept	2.926 (0.035)	7.030 (0.000)
Female	-0.479 (0.000)	
Age	0.238 (0.000)	
Age. Sq.	-0.003 (0.003)	
Educ.	0.247 (0.000)	
$\sigma_\epsilon$	1.949 (0.015)	1.900 (0.017)
$\sigma_\eta$	0.766 (0.060)	0.805 (0.065)
I <sub>96</sub>	-0.170 (0.010)	0.119 (0.047)
I <sub>01</sub>	-0.169 (0.023)	-0.150 (0.032)

The second column shows the estimation results of regressing log earnings by industry on 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 and on our estimates for the standard deviations of the permanent shock and transitory shocks to labor earnings ( $\sigma_\epsilon$  and  $\sigma_\eta$ , respectively). For positive (negative) coefficients the values in parentheses show the probability that the coefficient is less (bigger) than zero computed by bootstrap.

Table C.4: Regression Results - Permanent and Transitory - Earnings Per Hour

Dependent Variable	Earnings Coefficient	Net Earnings Coefficient
Intercept	2.022 (0.022)	2.377 (0.000)
Female	-0.131 (0.000)	
Age	0.059 (0.103)	
Age. Sq.	-0.001 (0.118)	
Educ.	0.068 (0.000)	
$\sigma_\epsilon$	1.230 (0.007)	1.241 (0.002)
$\sigma_\eta$	0.747 (0.003)	0.826 (0.003)
I <sub>96</sub>	-0.027 (0.247)	0.272 (0.000)
I <sub>01</sub>	-0.010 (0.427)	0.246 (0.000)

The second column shows the estimation results of regressing log earnings by industry on 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 and on our estimates for the standard deviations of the permanent shock and transitory shocks to labor earnings ( $\sigma_\epsilon$  and  $\sigma_\eta$ , respectively). For positive (negative) coefficients the values in parentheses show the probability that the coefficient is less (bigger) than zero computed by bootstrap.

distort inference. In this section we examine how eliminating imputed values changes the estimated relationship between earnings and volatility. Table C.5 reports the results of regressing average earnings on earnings volatility using the two alternative approaches presented in the main text. The only difference in the sample is that we do not use observations flagged as being imputed values and not actual responses. Although there are numerical differences in the coefficient values, it does not appear as if eliminating those earnings observations based on imputation alters the results by much.

Table C.5: Regression Results - Permanent and Transitory - No Imputed Earnings

Dependent Variable	Earnings Coefficient	Net Earnings Coefficient
Intercept	2.891 (0.025)	7.077 (0.000)
Female	-0.494 (0.000)	
Age	0.245 (0.002)	
Age. Sq.	-0.003 (0.002)	
Educ.	0.237 (0.000)	
$\sigma_\epsilon$	1.674 (0.028)	1.178 (0.077)
$\sigma_\eta$	0.742 (0.093)	0.735 (0.107)
I <sub>96</sub>	-0.127 (0.033)	0.103 (0.050)
I <sub>01</sub>	-0.125 (0.047)	-0.092 (0.107)

The second column shows the estimation results of regressing log earnings by industry on 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 and on our estimates for the standard deviations of the permanent shock and transitory shocks to labor earnings ( $\sigma_\epsilon$  and  $\sigma_\eta$ , respectively). For positive (negative) coefficients the values in parentheses show the probability that the coefficient is less (bigger) than zero computed by bootstrap.

## D Model Computation

1. To facilitate its solution, the model is transformed by dividing the objective function and the budget constraint by the permanent component of labor earnings. Defining  $\tilde{x}$  to be  $\frac{x}{e^{\omega_t}}$ , that is, a variable divided by the permanent component of earnings, the objective function for the worker becomes,

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

where  $\gamma_s$  is growth rate of permanent income between periods  $s$  and  $s - 1$ . The budget constraint, omitting the subscript  $j$ , for the individual at time  $t$  becomes,

$$\tilde{c} + \tilde{b}' = \frac{\tilde{b}}{e^\epsilon} (1 + r) + w e^\theta e^{\eta j} \quad (26)$$

2. The first step to solving the transformed model is to discretize the distributions for the selection parameters (i.e. the industry-specific skills or productivities) and the two shocks (transitory and permanent). 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  $G_R^j = \{\hat{\theta}_j^1, \dots, \hat{\theta}_j^{N_R}\}$ , 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  $w_R = 2.5$  and  $N_R = 6$ . To discretize the two shocks, we proceed as follows. We discuss the discretization of the permanent shock only; the method used for the temporary shock is identical. We discretize the support of the distribution using a grid of length  $N_P = 5$ ,  $G_P^j = \{\tilde{\epsilon}_j^1, \dots, \tilde{\epsilon}_j^{N_P}\}$ . We set  $\tilde{\epsilon}_j^3$  to zero. We approximate the probability mass of each point in  $G_P^j$  using a Gaussian pdf, which gives  $\tilde{P}r_P^j = \{p(\tilde{\epsilon}_j^1), \dots, p(\tilde{\epsilon}_j^{N_P})\}$ . However, that probability mass vector does not take into account that shocks only occur with a probability  $(1 - \lambda_j)$ . To account for that leptokurtosis, we multiply  $\tilde{P}r_P$  times  $(1 - \lambda_j)$ , so that  $\sum_{i=1}^{N_P} p(\tilde{\epsilon}_j^i) = (1 - \lambda_j)$  and add the remaining mass, equal to  $\lambda_j$ , to the value zero. As a result, the final probability mass vector is

$$Pr_P^j = \left\{ p(\tilde{\epsilon}_j^1), \dots, p(\tilde{\epsilon}_j^3) + \lambda_j, \dots, p(\tilde{\epsilon}_j^{N_P}) \right\}.$$

3. Guess a distribution of masses  $\{\pi_j\}_{j=1}^J$  and efficiency levels  $\{\theta_j^*\}_{j=1}^J$  for each of the industries. This yields aggregate employment levels (in efficiency units) for each of the 19 industries  $\{N_j\}_{j=1}^J$ . From our technology assumption, the wage rate in each industry is equal to the marginal product of a unit of labor efficiency.
4. Given a set of wages  $\{w_j\}_{j=1}^J$ , 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  $G_B = \{b_1, \dots, b_{N_B}\}$ , with  $N_B = 100$ , and we interpolate using a cubic spline to approximate future value functions. We discretize the values of the persistent and temporary shocks,  $\omega$  and  $\eta$ . 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).
5. The previous step yields a set of  $N_R$  expected value functions for each industry  $j$  conditional on a given level of ability,

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

6. We simulate a set of  $N_I = 2,500,000$  life cycle histories. A draw from the abilities distribution determines the optimal choice of industry. To compute the maximum value function across industries we use linear interpolation (recall that we only compute a value function for a set of  $N_R$  point in each industry-specific ability distribution).<sup>27</sup>
7. The simulation yields an updated set of masses  $\{\tilde{\pi}_j\}_{j=1}^J$  and efficiency levels  $\{\tilde{\theta}_j^*\}_{j=1}^J$ . An equilibrium is found when  $\max \{|\pi_j - \tilde{\pi}_j|\}_{j=1}^J$  and  $\max \{|\theta_j^* - \tilde{\theta}_j^*|\}_{j=1}^J$  are both less than some tolerance level.

---

<sup>27</sup>Results are presented for a smaller sample of 500,000 life-cycle histories.

# E Sensitivity Analysis

## E.1 Risk Aversion

In our calibration we assume a value of 3 for risk aversion coefficient of our CRRA utility function. In order to address the importance of this assumption for our results we conduct a sensitivity analysis in which we consider alternative levels of risk aversion. We do not re-calibrate the cross-industry distributions of abilities; rather, we re-solve the model and examine the relationship between mean earnings and risk when agents' degree of risk aversion is changed.

The main inference obtained from the baseline case - that temporary shocks are not priced in the labor market and the positive correlation with earnings is an artifact of selection of heterogeneous workers - remains for different levels of risk aversion. Given that we are not recalibrating the abilities distributions when we modify the degree of risk aversion, the relationship between average earnings and our two measures of volatility changes as well. Table E.6, however, shows the result of regressing (log) average earnings across industries on the two measures of volatility for the model with selection and the model with no ex-ante heterogeneity.

Each row of the table corresponds to a specification with a different degree of risk aversion. We consider values for the coefficient of relative risk aversion,  $\gamma$ , from 1 to 5<sup>28</sup>. As expected, everything else equal, a lower risk aversion coefficients imply a lower regression coefficient of average earnings on the standard deviation of the permanent shock. However, irrespective of the degree of risk aversion, the model with homogeneous workers delivers a coefficient on the permanent shock larger than that of the model with heterogeneous workers. In other words, the compensation for risk is large for all levels of the risk aversion parameter. Moreover, the coefficient on the transitory shock becomes very small in all five cases when the selection channel (heterogeneity) is shut-off; for all the levels of the risk aversion parameter, the compensation for transitory risk is virtually zero.

---

<sup>28</sup>The baseline calibration corresponds to  $\gamma = 3$ .

Table E.6: Predictions with Different Degrees of Risk Aversion

	Baseline		Ex-ante Homogeneous	
	$\beta_{\sigma_e}$	$\beta_{\sigma_n}$	$\beta_{\sigma_e}$	$\beta_{\sigma_n}$
$\gamma = 1$	6.42	3.39	1.02	-0.01
$\gamma = 2$	6.34	4.89	2.78	-0.01
$\gamma = 3$	6.54	5.84	4.13	-0.02
$\gamma = 4$	6.93	5.59	5.22	-0.03
$\gamma = 5$	7.25	5.35	6.11	-0.04

## E.2 Correlated Abilities

We perform an experiment in which we allow individuals' abilities to be positively correlated. Specifically, we relax the assumption of uncorrelated abilities by conducting the following experiment. We use the standard deviations for the abilities distributions that we employed in our benchmark analysis. To determine an appropriate covariance structure of abilities draws across industries, given the large number of industries we consider we opt for a parsimonious specification in which a common covariance term drives all pair-wise correlations in abilities. In other words, covariances are all equal to a parameter  $\kappa = 0.08$  (note that correlations coefficients differ because standard deviations differ). Some of our industries are sufficiently different in the type of human capital they require (e.g. agriculture and social services), while other pairs of industries may be similar (e.g. social services and government). Although arguably arbitrarily we have targeted average correlation across pairs of industries equal to 0.34. That number implies a positive but not too strong correlation of abilities.

Table E.7 displays the regression coefficients of the model with correlated abilities. Alongside, it also shows the coefficients obtained from the SIPP data as well as the ones obtained from the model with no ex-ante heterogeneity (where obviously the positive correlation across abilities is irrelevant). The third column shows that the coefficients for the model with selection are larger (particularly the one as-



sociated with the standard deviation of the permanent shock). Even though that affects the magnitude of the compensation for risk, it does not appear to invalidate our main result as is clear from column 4.

Table E.7: Predictions with Correlated Abilities: Mean and Volatility of Earnings

Variable	Data Coefficient	Model - Correlated Abilities Coefficient	Counterfactual Coefficient
Permanent $\sigma_\epsilon$	6.23	10.25	4.13
Transitory $\sigma_\eta$	5.86	5.88	-0.02

Note: The second column shows the estimation results of regressing log mean earnings by industry (weighted mean by panel of SIPP) on the standard deviations of permanent and transitory shocks. The third (fourth) column presents the estimation results of the same regression but using the mean earnings predicted by the model both in the correlated-abilities case (no ex-ante heterogeneity) described in the text.