Cross-Sectional Variation in Individuals' Earnings Instability

Scott Drewianka*

University of Wisconsin-Milwaukee

October 7, 2009

Abstract

There has been considerable recent interest in earnings instability—the variability of workers' earnings around their expected earnings paths. While previous work has measured trends in instability, often to illuminate trends in inequality, this paper investigates the variation across workers. Data from the Panel Study of Income Dynamics reveal considerable differences in earnings instability across demographic and occupational characteristics, generally in accordance with prior expectations. These results can also be used to develop a person-specific measure of instability for use in behavioral studies, and it is shown that the resulting metric correlates strongly with several decisions that are plausibly influenced by earnings risk.

Keywords: earnings instability, employment, wages.

JEL codes: D3, J0.

1 Introduction

Although much has been written about the role of risk in financial and capital markets, it is probably true that the most important risks facing the majority of people are those that arise in labor markets. It is well

*Dept. of Economics, Bolton Hall 860, 3210 N. Maryland Ave., Milwaukee, WI 53211; (414)229-2730; fax: (414)229-3860; *sdrewian@uwm.edu*. This project benefited from outstanding research assistance by Müge Karacal, and at an earlier stage by Menna Eltaki and Martin Priest. Helpful advice and comments were provided by Scott Adams, Florence Allen, Swarnjit Arora, Keith Bender, Bill Boal, Stacey Chen, Gary Fields, Nathan Grawe, Justin Gibson, Peter Gottschalk, John Heywood, Kristen Keith, Sunwoong Kim, Stephan Klasen (the editor), Thomas Kniesner, Peter Orazem, James Peoples, Richard Perlman, Justin Tobias, Don Williams, two anonymous referees, and participants in seminars at Iowa State University, the University of Wisconsin-Milwaukee, the Midwest Economics Association meetings, and the joint meetings of the Society of Labor Economists and European Association of Labour Economists. I am also grateful for funding from the UWM Graduate School Research Committee, the Center for Research on International Economics, and the National Institute for Child Health and Human Development (grant number T32HD07302). All remaining flaws are solely my responsibility.

known that earnings from labor constitute the largest portion of income for most individuals. Like all other sources of income, it carries some risk, as we see in the form of plant openings and closings, technological change, and unemployment. Moreover, it may be particularly difficult to insure against risks to labor income because a person can generally work for only a small number of employers at once and because many types of human capital are useful to only a limited number of employers.

Insofar as these risks are important, it is reasonable to expect them to influence decisions. This notion is perhaps most immediate in the context of occupational choice, where we expect that workers will not accept riskier jobs unless they are paid an expected compensating wage differential (Orazem and Mattila, 1991). Labor market risks may also affect decisions about specialization (Grossman and Shapiro, 1982), home ownership (Haurin and Gill, 1987; Diaz-Serrano, 2005a, 2005b), marriage and fertility (Oppenheimer, 1988; Weiss and Willis, 1997), or indeed any sort of lasting commitment (Shore and Sinai, forthcoming). Such risks have implications for public policy as well, as they influence, for example, the value of social insurance programs and the desirability of a progressive income tax code (Kniesner and Ziliak, 2002a, 2002b).

In light of these considerations, it would be desirable to measure labor market risks and their effects. However, as the next section explains, relatively little is currently known about how individuals differ in their exposure to those risks. The goal of this paper is thus to reduce that gap in our knowledge by identifying some characteristics of individuals who face relatively high or low levels of earnings risk. To be clear, the statistical relationships identified here are descriptive, but not necessarily causal. (For instance, we will find that self-employed workers face greater earnings risk than other workers. While it is certainly possible that they would face less risk if they were not self-employed, it is also possible that they would have faced elevated earnings risk anyway (e.g., due to volatile personalities) and use self-employment as a way to manage that risk.) Nevertheless, given how little we know about variation in earnings risk, there would seem to be considerable value in simply describing who faces it. In addition to the obvious distributional implications, the resulting estimates might also be used to create a person-specific index of risk that could then be used to identify how behavior varies with exposure to that risk.

Since the analysis will emphasize differences across individuals, the paper's title calls it a "cross-sectional" analysis. This does not mean that the longitudinal nature of the data is ignored, however. The first stage of the empirical analysis will identify earnings shocks using longitudinal earnings data and panel regression

methods, so those shocks are identified by fluctuations around individual-specific earnings paths. The second stage of the analysis then seeks to determine how the magnitude of those shocks varies with individuals' characteristics. As explained later in the paper, there are reasonable arguments both for and against using panel regression methods at that stage. We shall thus begin by presenting results from pooled regressions, but subsequent robustness checks will show that longitudinal regressions generally yield similar estimates.

Most previous work on variation in earnings instability has been more interested in differences over time, rather than across people, although a few studies have measured differences in earnings risk between large groups by stratifying the data according to factors like age and education. Insofar as comparisons are possible, the results presented here are generally consistent with the most common findings in the literature. For example, the estimates corroborate previous evidence from several countries that earnings instability has increased over time and that the youngest and oldest workers have the most volatile earnings. However, unlike most previous studies, our empirical strategy allows us to measure conditional effects. We are thus able to learn, e.g., that more educated workers have more stable earnings on average, but not when we control for their occupations and industries—suggesting that the effect mainly reflects an ability to avoid shocks, rather than an ability to adapt to them. Other key findings are that African-Americans and self-employed workers face unusually high earnings risk, and that earnings risk tends to be higher in occupation and industries with lower average earnings.

The discussion begins by reviewing the relevant literature, and then Sections 3 and 4 present the data and empirical methods that are used to measure covariates of labor market risks. Section 5 reveals the empirical results, checks for robustness, and provides some indication that our measure of risk has predictive power for some behaviors. The final section summarizes the results and suggests directions for additional work.

2 Background

Earnings volatility arises in many economic applications, extending back at least as far as Adam Smith (1776[1994], 119-133), but it is perhaps most closely associated with the Permanent Income Hypothesis (Friedman, 1957). In that well-known theory, a person's income at a particular time can be decomposed into "permanent" and "transitory" components, respectively representing his or her expected income path

and deviations from that path. The permanent component has a much larger effect on the person's ability to consume resources, but transitory components may also affect individuals' welfare if there are credit constraints or if the transitory shocks are substantially persistent over time.

This distinction between permanent and transitory earnings has recently received substantial attention in the literature on income inequality. Several papers have decomposed inequality into permanent and transitory components and compared their magnitudes. Generally speaking, studies from more developed countries like Canada (Kennedy, 1989), Germany (Burkhauser, Holtz-Eakin, and Rhody,1997), and the U.S. (Haider, 2001) tend to find that a majority of the cross-sectional variation in earnings reflects unequal lifetime incomes. Dahlberg and Gustavsson (2008) estimate that the components are of roughly equal magnitude in Sweden, although they caution that their methodology may understate the permanent component of inequality (e.g., because they stratify the data across counties and thus lose the between-county variations, which is likely to be predominantly permanent). In contrast, a larger fraction of the overall variation is attributed to transitory shocks in developing countries like Brazil (Santos and Souza, 2007) and Venezuela (Freije and Souza, 2002).

Other studies have attempted to determine the extent to which trends in income inequality are due to permanent or transitory differences in workers' incomes, with most results indicating that both components have grown considerably in Canada (Baker and Solon, 2003; Beach, Finnie, and Gray, 2003; Morissette and Ostrovsky, 2005), Sweden (Dahlberg and Gustavsson, 2008), the U.K. (Blundell and Preston, 1998; Dickens, 2000), and the U.S. (Gottschalk and Moffitt, 1994; Moffitt and Gottschalk, 1995; Haider, 2001; Hyslop, 2001, Moffitt and Gottschalk, 2002). The increase in transitory inequality has also been used to explain why consumption inequality has grown more slowly than income inequality and, relatedly, the expansion of consumer credit (Krueger and Perri, 2006).

Most of these papers emphasize changes in the variance of transitory earnings over time, rather than the cross-sectional variation that is the focus of the present paper. The prevailing emphasis is mainly driven by the goals of the earnings inequality literature, but it may also be a consequence of the most widely used empirical methodology, which involves estimating the parameters of earnings dynamics models by minimizing the distance between the autocovariance matrix predicted by a theoretical model and the observed population-wide autocovariance matrix. While there are strong reasons for using that approach in that context, the fact that it requires an empirical autocorrelation matrix derived from many individuals makes it inconvenient to

investigate cross-sectional variation in parameters.

Much of what we know about differences in earnings instability across groups comes from a well-known paper by Gottschalk and Moffitt (1994). Like other contributors to the literature, they are primarily interested in trends in earnings instability. However, part of their analysis involves stratifying the sample according to some characteristics, which allows them to assess how inequality and instability differ between groups (although it does not allow them to control for other covariates that may also differ between those groups). They find that the transitory variance is larger for U.S. men who are younger, less educated, earn less, do not belong to labor unions, and work in construction, services, or wholesale and retail trade. However, Gordon (1984) finds the opposite pattern for age and education (in an earlier period). In Canadian data, Baker and Solon (2003) and Morissette and Ostrovsky (2005) find a U-shaped pattern of earnings volatility over the life-cycle, with a minimum around age 45.

Other notable studies in the literature also provide some evidence about covariates of transitory earnings. While investigating the riskiness of human capital investments, Chen (2008) finds that a college education has little effect on the transitory volatility of worker's earnings, although there is substantial variation in permanent earnings among college graduates. Bostic (1997) asks whether differences in earnings volatility can explain racial differences in mortgage application rejection rates, but he finds little evidence that there are important interracial differences in earnings volatility. In contrast, several studies have shown that self-employment is associated with increased earnings volatility in the U.S. (Gordon, 1984; Rosen and Willen, 2002) and many European nations (Diaz-Serrano, 2005b).

3 Data

The data we will use were originally gathered in the 1980–1997 waves of the Panel Study of Income Dynamics (PSID), but they have been recoded for use in the Cross-National Equivalence File (CNEF). One advantage of using the CNEF data is that it has been more thoroughly prepared for immediate use. In addition, some variables have been recoded to make them comparable to variables in three other major international panel studies (Burkhauser, Butrica, Daly, and Lillard, 2000).

The analysis uses data on earnings, hours, and wages for non-disabled men aged 16-65 who report positive

earnings and hours worked and who did not report a labor force status of "student" or "retired." Since some characteristics vary across years, the data are aligned so that workers' characteristics at a point in time are matched with their earnings, hours, and wages for that year, which were gathered in the following year's survey. This means that the shocks we identify are *about to be experienced* (as opposed to recently experienced) by a person with a particular set of characteristics. This seems especially important for analyzing workers' occupations and industries, as an employment shock may plausibly cause a worker to change employers.

There is some danger in excluding men who worked zero hours, as we would underestimate labor market risks if we eliminated workers who were involuntarily unemployed for an entire year. That said, people who simply do not want to work should properly be excluded, and there are practical problems with observations with zero earnings (e.g., log earnings is not well-defined). Previous work has consistently excluded such observations, so we follow that convention. We will revisit this decision in Section 5.3, however.

One departure from previous work is that the analysis uses both the main PSID data and the associated Survey of Economic Opportunity (SEO), an oversample of persons who lived in low-income families in 1966.[1] Including the SEO raises the sample size among non-whites by over 70 percent, which provides sufficient observations to examine interracial differences. In contrast, most previous papers have excluded the SEO and focused on white men. One concern is that combining the samples may bias the results by overrepresenting persons from poor families. That said, Table 1 shows that the SEO is more representative of the broader population than one might imagine—the race-specific distributions of log earnings, log hours, log wages, age, and education, are generally quite similar in the two samples. Presumably this similarity, at least within racial groups, reflects churning of the income distribution, intergenerational regression to the mean, and entry of individuals who married into the sample between 1966 and 1981 (when our first observations were surveyed). Regardless, we will guard against unforeseen problems by including a dummy variable for the SEO observations in all regressions run below, and our first-stage regressions (see equation (1) below) will also allow separate year and education effects for the SEO sample. Very few of these parameters turn out to be statistically significant, however.

[PLACE TABLE 1 ABOUT HERE]

Table 2 lists summary statistics for the full sample, starting with our three dependent variables: logs of

earnings, hours, and wages. Dollar figures are given in 1990 dollars, as measured by the Consumer Price Index (CPI-U). The rest of the table shows how the observations are distributed across several demographic and occupational variables. Perhaps the most notable feature of the sample is the racial composition. Due in part to the inclusion of the SEO data, a disproportionate share of the sample is African-American. Latinos, and to a lesser extent Asians, comprise a smaller share of the sample than they do of the current population, mostly due to immigration patterns in the decades since the sample was drawn. The occupations and industries listed reflect either workers' current primary job or the most recent job on which we have information. The educational categories should be read as mutually exclusive—a person can belong to only one category at a time, although he may move to a new category after he obtains additional schooling.

[PLACE TABLE 2 ABOUT HERE]

4 Empirical Model

Earnings shocks are defined as deviations from individuals' expected earnings paths. Studies of transitory earnings thus begin by estimating a model of earnings dynamics, then interpret the residuals as transitory shocks. This paper will look for covariates of earnings risk by identifying factors that predict the magnitude of those residuals. Since workers' careers are far too short for a law of large numbers to apply to each person individually, the shocks one actually experiences can paint only a partial picture of the risks to which one was exposed (Friedman, 1953). Accordingly, we will make the identifying assumption that observably similar persons face similar risks. In practice this means that the absolute value of the residuals will eventually be regressed against individuals' characteristics, so that risk will be measured from both individuals who experience relatively large shocks and the presumably larger number who experience small shocks.

4.1 Estimating Cross-Sectional Variability in Earnings

We begin by specifying a model of earnings. Let Y_{igt} be the log annual earnings of person i from group g in year t, and let A_{it} be i's age. Suppose that i's expected age-earnings profile is the sum of two polynomials

in age, Q and R, and let ϵ_{igt} be deviations from that path:

$$Y_{igt} = Q(A_{it}, \psi_{gt}) + R(A_{it}, \gamma_i) + \epsilon_{igt}. \tag{1}$$

The polynomial Q is a "baseline" age-earnings profile for group g, where groups are defined by race, birth cohorts, and level of education. It represents differences in earnings that an econometrician could predict from agents' ages and groups. As in previous work, Q is specified as a quartic function of age, and its coefficients $\psi_{gt} \equiv (\psi_{gt0}, \ldots, \psi_{gt4})$ are allowed to vary over time to reflect changes in the labor market.[2]

One might be tempted to interpret Q as an individual's expected earnings, in which case one would say that earnings shocks are the difference between the person's actual earnings and Q. A reasonable objection to that plan is that agents undoubtedly have more information about their own expected earnings profile than the econometrician does. For example, although an econometrician might only know that a particular 27 year old white male is a college graduate, that person himself would also know what he studied, what his grades were, what his family background is, and all sorts of other idiosyncratic information that would help him form a better forecast of his own future earnings than the econometrician's. Put another way, such private information would create (from the econometrician's perspective) unobserved heterogeneity in individuals' idiosyncratic expected age-earnings profiles.

The polynomial R is intended to address this objection It represents the difference between the individual's forecast of his own earnings and the econometrician's forecast—in other words, person i's expected age-earnings profile relative to that of his group. The sum $(Q + R)$ is thus the age-earnings profile the individual actually expects. Although heterogeneity in earnings that is associated with $(Q + R)$ represents earnings *inequality* across individuals, it does not represent earnings *instability* because the workers can predict that component of their earnings. Instead, the earnings shocks—the deviations ϵ—are the only actual realizations of the earnings risk we seek to understand.

We will take R to be a random polynomial in age, with coefficients $\gamma_i \equiv (\gamma_{i0}, \ldots, \gamma_{iK})$:

$$R(A_{it}, \gamma_i) = \sum_{k=1}^{K} \gamma_{ik} A_{it}^k \tag{2}$$

$$(\gamma_{i0}, \ldots, \gamma_{iK}) \sim N(0, \Sigma_\gamma) \tag{3}$$

We place no restrictions on Σ_γ (other than properties of all covariance matricies, like symmetry and positive definiteness), so the data are free to inform us about the correlation between the coefficients. In the sections that follow, we will experiment with different orders (K) for this polynomial, and will will ultimately decide to work with a quadratic form ($K = 2$) because including additional terms does not increase the log-likelihood (although it does dramatically increase computation time).[3] As we shall see in Section 5, in practice the order of R does not make a great deal of difference, as the residuals are very similar whether K is 0,1, or 2.

4.2 Comparison with Previous Models of Earnings

Before turning to the residuals, it may be helpful to compare this model to those used previously in the literature. A representative model is Haider's (2001):

$$Y_{it} = Q(A_{it}, \psi_t) + p_t(\alpha_{i0} + \alpha_{i1} A_{it}) + \epsilon_{it}, \tag{4}$$

where the notation above has been preserved to the extent possible. The function Q is exactly as in (1), except that its parameters do not vary across groups. Parameters α_{i0} and α_{i1} allow heterogeneity across individuals in the baseline level of earnings and earnings growth, much like γ_{i0} and γ_{i1} in (1). Thus, the main difference is the loading factor p_t that allows the cross-sectional variation in earnings to vary over time.

If the goal is to measure the trend in permanent inequality, including a population-wide parameter like p_t simplifies the interpretation of the results. However, if the focus is on the cross-sectional variation in the distribution of ϵ_{it}, it is not critical to partition fluctuations into parts associated with aggregate trends (p_t) and individual heterogeneity (α_0, α_1). Rather, the goal at this point is just to identify individuals' expected earnings path in order to get the best possible estimates of the residuals. As long as p_t can be reasonably approximated by a low-order polynomial (a simple linear specification fits Haider's estimates of p_t for 1973–1991 with $R^2 = 0.935$), the specification in (1) is more flexible than that in (4) because it does not force the trend in inequality to affect all persons equally. Thus, the specification used here is similar to Haider's, but somewhat better suited to our needs.

Haider's model can also be used to illustrate another attractive feature of model (1): its tractability. Although Haider's is just one of the many models that appear in the literature on earnings dynamics,[4] it

is estimated by the same technique used in nearly every paper in that literature. The estimation strategy begins by deriving the aggregate autocovariance matrix that is predicted by the model of earnings dynamics. For example, Haider's model (4) implies that

$$\mathrm{Var}(Y_{it}|A_{it}) = p_t^2(\sigma_0^2 + 2\sigma_{01}A_{it} + \sigma_1^2 A_{it}^2) + \sigma_{\epsilon t}^2 \tag{5}$$

where $\mathrm{Var}(\alpha_{i0}) \equiv \sigma_0^2$, $\mathrm{Var}(\alpha_{i1}) \equiv \sigma_1^2$, $\mathrm{Cov}(\alpha_{i0}, \alpha_{i1}) \equiv \sigma_{01}$, and $\mathrm{Var}(\epsilon_{it}) \equiv \sigma_{\epsilon t}^2$. There is a similar prediction for each element of the earnings covariance matrix. One then estimates the parameters $(\sigma_0^2, \sigma_1^2, \sigma_{01}, p_t^2, \sigma_{\epsilon t}^2)$ by minimizing the distance between the predicted- and observed autocovariance matricies.

While this technique is ideal for estimating (among other parameters) the variance of aggregate earnings risk in each period ($\sigma_{\epsilon t}^2$), it is not very useful for estimating covariates of earnings risk because it only estimates parameters of the aggregate distribution—i.e., it does not generate a residual for each individual observation. In principle, one could partition the sample according to some characteristics (e.g., level of education) and compute a separate variance for each subsample, but this would be tedious and still would not provide much opportunity to explore conditional earnings risk (e.g., we could not see how earnings risk varied with education conditional on workers' race or self-employment status). This is why Section 2 claimed that this methodology is inconvenient for investigating heterogeneity in $\sigma_{\epsilon t}^2$ across workers.

In contrast, our model of earnings dynamics (1) can be used to compute a residual for each observation. It is a standard mixed model, and thus (unlike Haider's model or most used in the literature) it is relatively straightforward to compute a best linear unbiased predictor of the person-specific random effects $\widehat{\gamma_i}$ and residuals $\widehat{\epsilon_{igt}}$ (see, e.g., Wooldridge, 2002). The next section explains how we will use those residuals to investigate the covariates of earnings risk.

4.3 Estimating Earnings Risk from Earnings Residuals

Since the estimated earnings residuals $\widehat{\epsilon_{igt}}$ are our estimates of earnings shocks, our goal is to identify covariates of their magnitude. Our main strategy is inspired by Glejser's (1969) familiar method of adjusting

for heteroskedasticity. That is, we shall regress the absolute values of $\widehat{\epsilon_{igt}}$ on characteristics Z_{igt}:

$$E[\,|\widehat{\epsilon_{igt}}|\,|\,Z_{igt}] = \delta_0 + \delta_t + Z_{igt}\delta. \tag{6}$$

The coefficients δ are interpreted as effects on the standard deviation of earnings shocks, and the predicted value is a measure of the individual's labor market risk. In effect, this approach addresses the concern that many individuals may experience no unusual outcomes even though they are at risk by effectively averaging the size of the unpredictable components of earnings across all people with those characteristics.

One possible objection to this procedure arises from the fact that the first-stage regression (1) did not control for some variables in Z—in particular, workers' occupation, industry, and self-employment status. There is a good reason for their exclusion: since that regression includes the person-specific polynomial R, the estimated effects of employment characteristics would be largely identified from cases in which workers change jobs. Whether such a change reflects a layoff or a firing or a new job offer, it would seem to represent exactly the sort of shock we hope to measure, so it does not seem appropriate to include those factors in regression (1). However, excluding those employment factors raises the possibility that $E\,[\widehat{\epsilon}|Z] \neq 0$, which may (or may not) alter our interpretation of $|\widehat{\epsilon}|$. At any rate, it may be reassuring to know the results below are robust to this consideration. When regression (6) is run with $|\widehat{\epsilon_{igt}} - E\,[\widehat{\epsilon_{igt}}|Z_{igt}]|$as the dependent variable, most parameter estimates change by a only negligible amount, and no substantive conclusions change.

Quantile regression (Koenker and Bassett, 1978) provides another reasonable alternative to (6), especially if we are concerned that the residuals may be oddly distributed or that the results may be unduly influenced by extreme outliers. That said, given our goal of measuring risk, it is not clear that we really want to minimize the role of the largest outliers, which presumably correspond to the most consequential shocks. To investigate, the results from regression (6) were compared to those from a quantile regression designed to predict the 90-10 interdecile range. The two methods produced remarkably similar measures of risk: the correlation between the predicted absolute deviations and the predicted interdecile range was 0.98 for both earnings and hours, and 0.99 for wages. Thus, in this context the difference between the two methods appears inconsequential. Since regression (6) is much less computationally intensive, Section 5 will present results from that approach.

At any rate, it should be acknowledged once again that none of these approaches allow us to make strong claims about causality. For instance, if we find that artists have a relatively large δ, we cannot immediately distinguish between the hypothesis that art markets are unusually volatile and the hypothesis that artists are unusually capricious. To be sure, many potentially interesting questions would require some understanding of causality, including several of the topics this paper discusses for future research. Nevertheless, given that relatively little is currently known about the distribution of these risks across individuals and that an empirical strategy to determine causality would likely be as controversial as it is informative, at this point it seems best to de-emphasize questions of causality and to focus on simply providing descriptive evidence about which workers face the most risk.

4.4 Measurement Error

One potential problem with this empirical strategy is the effect of measurement error. If there is no measurement error, regression (1) provides the best linear unbiased predictor of the residual, but if the dependent variable is measured with error the estimated residual will also include the portion of the measurement error that cannot be predicted from the person-specific age-earnings profile. Insofar as the remaining measurement error is correlated with characteristics Z, the estimated cross-sectional differences in volatility ($E[\,|\widehat{\epsilon}|\,|\,Z]$) would be biased. However, the remaining measurement error would be less consequential if it were uncorrelated with Z, as only the intercept would be estimated with bias. For many purposes such a bias would be irrelevant, essentially just a rescaling of our earnings instability index.

Thanks to a series of validation studies, a fair amount is known about measurement error in the PSID (Duncan *et al.*, 1985; Duncan and Hill, 1985; Rodgers, Brown, and Duncan, 1993; Bound *et al.*, 1994; Bound, Brown, and Mathiowetz, 2001). The average earnings as reported by workers is very close to the average earnings as reported by their employers, and the mean absolute difference between the two is not especially large, around ten percent of the mean wage. In addition, reporting errors are positively correlated over time for individuals, so it seems likely that much of it will be captured in the individual-specific portion (R) of regression (1). For annual hours worked, the mean absolute reporting error is a bit larger, but that is due in part to the fact that workers systematically overreport their hours worked, which suggests that much of the measurement error in hours would also be captured in the intercept of regression (1). Moreover, for

both earnings and hours the magnitude of the measurement error is generally not strongly correlated with variables like age, education, race, and gender—the only major difference is that the earnings (but not hours) of blacks are reported with larger absolute errors than are those of other groups.

All of this would tend to suggest that measurement error is unlikely to bias our results strongly. However, Pischke (1995) has argued that much of the measurement error stems from underreporting of transitory shocks. If this is correct, and if the tendency to underreport transitory shocks varied across characteristics Z, the method used here would mismeasure differences in earnings volatility across groups. On the other hand, if the underreporting is fairly uniform across the population (as suggested by the fact that there are few strong covariates of absolute measurement error), its primary effect would be a downward bias for the intercept of regression (6). As noted above, this would effectively rescale our volatility index, but the index would still appropriately reflect cross-sectional differences in volatility.

Thus, while one could never entirely dismiss the possibility, the existing evidence on the extent, nature, and covariates of measurement error in the PSID suggests that it is unlikely to bias our results seriously.

5 Results

In presenting the results, let us begin with a brief look at the results of regression (1). For each dependent variable, Table 3 reports some key parameter estimates and the distribution of residuals from three different orders specified for the random polynomial R. For example, the first column of the table presents the results when we specify R as a zero-order random polynomial—in other words, a simple random effects model. Such a model allows for heterogeneity in individuals' earnings, but does not allow that heterogeneity to change as the individuals age. Previous work has often considered such a model too restrictive, but we present it for the sake of comparison. The second column then presents the results when we specify R as a first-order polynomial in age (so $R(A, \gamma) = \gamma_0 + \gamma_1 A$, where $(\gamma_0, \gamma_1) \sim N(0, \Sigma_\gamma)$ for some arbitrary covariance matrix $\Sigma\gamma$), which allows for heterogeneity in both levels of earnings and earnings growth. The results in the first portion of the table describe Σ_γ. The results in the lower portion of the table describe the residuals from these regressions, which are our estimates of the individuals' earnings shocks and what will be analyzed below when we seek to identify covariate of earnings risk.

[PLACE TABLE 3 ABOUT HERE]

Two main conclusions emerge from Table 3. First, the results appear robust to alternate specifications of R. Adding higher-order terms improves the goodness of fit, at least up to a quadratic form. That said, it only changes the standard deviation of $\widehat{\epsilon}$ by a small amount, and estimates of ϵ are highly correlated across specifications, especially beyond the first-order.[5] Since it provides the best fit, the analysis below will be based on a quadratic R, but it should be clear that the results are not sensitive to this choice.

Second, the distribution of $\widehat{\epsilon}$ is nearly symmetrical. In each case, the median value of $\widehat{\epsilon}$ is a small positive number, which implies that workers are slightly more likely to experience a very large negative shock than a very large positive one. Nevertheless, these distributions are otherwise quite symmetrical, which helps to explain why the results of regression (6) are robust to the alternative approaches discussed in Section (4.1).

5.1 Covariates of Earnings Risk

Tables 4–7 present the evidence on covariates of earnings risk. Table 4 presents the standard deviations of the dependent variables and their residuals ($\widehat{\epsilon}$) across years and across individuals' demographic characteristics. For this table only, each reported standard deviation is measured from a stratified sample. For instance, the second row of results in the left-hand column indicates that among respondents who are not high school graduates, the standard deviation of actual log earnings is 1.04, and the standard deviation of residual log earnings (after the first stage regressions) is 0.50. The next line then presents the same measures for the population of respondents who have exactly 12 years of schooling, and so on. It should be noted that this is the only point where the data have been partitioned; every subsequent table uses data from regressions that use all of the available data.

[PLACE TABLE 4 ABOUT HERE]

Note also that each observation in the full data set appears once within each of the reported categories. For example, an observation from a 27 year old man with 12 years of schooling is used to compute both (a) the standard deviation of log earnings for men aged 26-30, and (b) the standard deviation of log earnings among high school-educated men. Consequently, the figures reported in Table 4 should be viewed as unconditional results. That is, when the table tells us that the standard deviation of earnings for men aged 16-20 is larger

than that of men aged 21-25, we cannot immediately conclude that the relationship is truly attributable to differences in age. An alternate explanation might be that the difference is really a college attendance effect, since men over age 21 are more likely to have attended college and the standard deviation of earnings is smaller among men who attended college.

To provide a better sense of conditional effects, Tables 5–7 list the results of regressions of absolute residuals ($|\widehat{\epsilon}|$) against all of those factors simultaneously.[6] It would be too strong to call these estimates of causal effects—for example, we might incorrectly conclude that some profession is especially risky if it simply tends to attract workers whose unobserved personality traits would tend to make their earnings unstable regardless of which occupation they chose. That said, the regressions effectively determine how closely earnings risk is associated with each observed characteristic conditional on all of the others, so they provide some insight into which factors are at least plausibly most important.

[PLACE TABLE 5 ABOUT HERE]

[PLACE TABLE 6 ABOUT HERE]

[PLACE TABLE 7 ABOUT HERE]

It is perhaps most interesting to discuss the effects of each factor across these tables, so the following discussion is organized by categories of explanatory variables.

Education. One of the more striking patterns in Table 4 is the inverse relationship between education and instability, especially hours instability. The variance of earnings around the expected earnings path of a typical worker with less than a high school education is on the order of 50 percent higher than that of a typical college graduate, and the difference is even greater for hours variability. This difference calls to mind any number of interesting hypotheses—including Schultz's (1975) contention that education increases individuals' stability by making them more adaptable to change ("the ability to deal with disequilibria"). Alternatively, it may reflect differences in exposure to layoffs or differences in quit rates.

The role of education appears a bit different in the regression results, however. Workers who have not graduated from high school continue to exhibit unusually unstable earnings and hours (and perhaps wages), and education still predicts some stability in hours, although that effect now seems to plateau once individuals enroll in college. Yet all else equal, education beyond high school is associated with increased wage volatility,

and those with some college (but not college graduates) have the least earnings volatility. While this finding may seem surprising in light of the evidence in Table 4, it is consistent with recent estimates by Chen (2008).

In order to identify the source of the discrepancy between the patterns of unconditional (from Table 4) and conditional (Tables 5–7) differences across educational groups, the regressions were run again with subsets of the explanatory variables excluded. The pattern of unconditional differences are robust to the inclusion of most of the factors considered. For earnings volatility, the new pattern emerges only when we add controls for age and occupations—more educated workers are more likely to work in more stable occupations, and during this particular era the workers most likely to be highly educated are those aged 30–50, the age range with the greatest earnings stability. Similarly, the difference in hours volatility arises because more educated workers are more likely to work in more stable industries and occupations, and the difference in wage volatility is entirely due to occupation effects. It thus appears that much of the greater stability enjoyed by more educated workers stems from their career choices—suggesting that it does not so much reflect an ability to deal with disequilibria as it does an ability to avoid disequilibria in the first place.

Age. Table 4 indicates that earnings volatility follows a U-shaped pattern over the life-cycle: sharply declining in the early years of one's career, remaining relatively constant for roughly 25 years, and then rising as retirement approaches. This is consistent with recent evidence from Canada (Baker and Solon, 2003; Morissette and Ostrovsky, 2005). The volatilities of hours and wages follow a similar trend, although it is much sharper for hours. Contrary to our experience with education, the effects are very similar in the regression format, the only change being that workers aged 16–20 no longer seem to have unusually unstable earnings. This small difference emerges only once we control for occupations and industries, which is not very surprising when one considers the type of jobs that teenagers typically hold.

Race. The results in Tables 4–7 consistently indicate that African-American and Latino men experience much more labor market risk than members of other racial groups. The variance of earnings shocks is estimated to be about 50 percent higher for African-Americans than for Whites, and over twice as high as that of Asian-Americans, and the black-white difference is even greater in hours volatility. The differences are not quite as large for Latinos, but they are still considerable. There is also some indication that African-Americans have more unstable wages, although the interracial differences are much smaller in that regard. All of these differences are reduced in the regression format, but in each case the effect remains relatively

large and statistically significant.

As noted earlier, the PSID validation studies leave open the possibility that the results for African-Americans are biased upwards by that group's larger measurement error. That said, it is perhaps more likely that these gaps understate the true difference in risk for African-Americans because they face a greater risk of several events that would cause them to be dropped from the sample, including mortality, incarceration, and year-long unemployment.

The results for Asian-Americans are also interesting. Members of that group experience the smallest fluctuations in all three variables, and especially so in earnings. The typical Asian-American's variance of transitory earnings is about 1/3 less than that of the typical White, for example. Nevertheless, the regressions indicate that Asian-Americans do not have unusually low instability compared to others who have the same traits, and they may even have greater hours instability (the estimate is significant at the 10 percent level). The initial pattern was robust to the inclusion of all of our explanatory variables except the occupation and industry dummies, so it seems that Asian-Americans' greater stability is a function of their career choices, and if anything they may have a bit more instability than members of other groups who hold similar jobs.

Trends over Time. As noted earlier, the previous literature on earnings instability has focused almost exclusively on changes over time. The results in Tables 4–7 are generally consistent with previous results showing that instability has increased in several countries (e.g., Blundell and Preston, 1998; Haider, 2001; Baker and Solon, 2003; Dahlberg and Gustavsson, 2008), so there is little reason to belabor the discussion. Perhaps the strongest (and least surprising) result is that risks rose relative to the trend during the recession years of the early 1980s and early 1990s. There is also some evidence of a trend toward greater instability of both earnings and wages, especially toward the end of the period. The average stability of hours worked (in Table 4) seems to be increasing slightly, but the trend vanishes in the regression format, especially when we control for workers' ages. (This was determined by running a series of regressions that excluded each category of explanatory variable, as explained above in the discussion of educational effects.) Thus, the small increase in hours stability seems to reflect changes in the age structure of the population. One possible reason is that during this period the large Baby Boom cohort (born 1946-1964) was entering the portion of the life-cycle in which workers' hours are typically most stable (ages 36-55, according to Table 6).

Self-Employment. Self-employment is perhaps the strongest predictor of unstable earnings. The average

transitory variance of earnings for fully-self employed workers (who do not also hold another job) is more than double that of other workers. While this may seem large, Diaz-Serrano (2005b) has previously shown that self-employment is a particularly strong indicator for earnings risk in several EU nations. Unlike many of the characteristics discussed previously, this difference is entirely due to greater wage instability. In fact, the average self-employed worker actually has more stable hours than the average non-self-employed worker, and self-employed workers who also hold another job ("partially self-employed") have the most stable hours of all.

One might imagine that the difference in hours stability reflects little more than the fact that self-employed workers are unlikely to lay themselves off. However, the regression results presented in Table 6 show that self-employed workers actually have *less* stable hours than *similar* workers. As it turns out, the main reason that the average self-employed worker has more stable hours than other workers is that they are more likely to be white and middle-aged—as we have seen, two traits associated with more stable hours.

Note that it is not clear that the differences in wage- and hours instability between self-employed and other similar workers were caused by self-employment itself. Even if we assume that some portion of the gap in earnings risk between self-employed and other workers is causal, that difference would presumably influence the selection of workers into self-employment. For example, any risks that are inherent in self-employment are plausibly more palatable to workers who would face greater risk in any event, or alternatively who have more stability in other aspects of their lives. Another possibility is that self-employment attracts workers who have a greater tolerance for risk (Brown *et al.*, 2006, Brown *et al.*, 2007), perhaps because their next best alternative may also be a job with elevated earnings risk. Of course, similar caveats apply to nearly all of the factors examined here, especially those that reflect workers' career choices (education, occupation, and industry). As emphasized in the paper's introduction, this is ultimately why the estimates are more appropriately interpreted as descriptive evidence rather than causal effects.

Insofar as being self-employed increases workers' exposure to earnings risk, it would be interesting to know whether there is a compensating differential for accepting this greater risk. In the raw data, the average self-employed worker earns more than 30 percent more than the average non-self -employed worker (log earnings of 10.29, versus 9.98 for others)—in part by working 20 percent more hours, but also because their hourly wages are 12 percent higher. It is beyond the scope or ability of the current paper to determine how much of

this difference is a compensating differential and how much is due to selection bias, but it seems like a very interesting question for future research.

Industries. There is great variability in instability across industries—to cite the extremes, the average worker in the agriculture and forestry sectors has a transitory earnings variance over six times greater than the average worker in the chemicals industry, and about double that of the average worker in the sample. Differences in other characteristics are responsible for some of these gaps, but the regression results verify that most of the qualitative conclusions hold even conditional on those differences. One exception is the financial sector, where the average worker has relatively stable earnings and hours, but the regressions indicate that similar workers in other industries experience similar stability.

The volatilities of earnings, hours, and wages are positively correlated across industries. Along with agriculture, the construction, retail, and service sectors emerge as the most volatile. Several of these industries are strongly cyclical, so in that respect the results correspond to our intuition. At the other end of the spectrum, instability is lowest in less cyclical industries like education, insurance, intermediate products (chemicals, wood/paper, earth/clay/stone, and energy and water), and perhaps health services. A few other industries have relatively low wage volatility, but more typical hours volatility: public administration, the postal system, wholesale trade, clothing and textiles, and engineering. The finding that several of the industries with less volatile earnings are dominated by the public sector (including education) is probably not surprising, and it matches previous evidence from every EU country studied by Diaz-Serrano (2005b).

It is also not very surprising that workers with unknown industry affiliation experience less stability. This might reflect unemployment as of the survey date, but remember that workers do not appear in the sample unless they have positive hours and earnings during the year. Such workers might also be more itinerant; workers have been assigned to their most recent industry whenever that is possible, but that we are less likely to have a prior industry for workers who do not remain in the sample for long. Otherwise, when a worker's industry is listed as unknown, that may simply mean that his or her industry defies clear categorization—which might be an indication of instability in and of itself.

Occupations. There is also a lot of variability across occupations. As with the industry effects, there is a strong positive correlation between occupations' hours instability and wage instability, and much of it compounds the differences across industries. Perhaps the most striking example is that of farm managers.

Even in the regression format—which also assigns a good deal of additional volatility to the agriculture industry— farm managers have the highest wage volatility of all occupations and the second-highest earnings volatility (exceeded only by workers whose occupations are unknown), although their hours are not abnormally volatile.[7] Similarly, bricklayers and carpenters have relatively unstable earnings even beyond the fact that they work in the construction industry.

It is also easy to see why the apparent effects of higher education are diminished when we control for workers' occupations. Human-capital intensive professions (often quantitatively oriented ones) emerge as the most stable occupations, including mathematicians, engineers, accountants, business managers, and lawyers. The most unstable occupations tend to have lower educational requirements—for example, cooks and waiters, laborers, and janitors.

Finally, on a related note, high-paying jobs tend to be more stable. If we run a regression like (6), except with log earnings as the dependent variable, the estimated occupation coefficients would represent premiums paid to workers over observably similar workers in other jobs. Those earnings premiums are substantially negatively correlated with the estimated occupational differences in earnings volatility reported in Table 5 (from the "actual" regression 6, with $|\widehat{\varepsilon}|$ as the dependent variable): $\rho = -0.83$ if the effects are weighted by the number of workers in each occupation.[8] One possible explanation is that the negative correlation reflects returns to tenure, since workers in more stable occupations presumably have greater average tenure. We can test that hypothesis by adding a quartic in worker's tenure to the log earnings regression, and this exercise does reveal a significant tenure premium: an extra year of job tenure is associated with an estimated 3.9 percent increase in wages at the mean tenure (5.36 years) and a 6.0 percent increase at the median tenure (3 years). Nevertheless, the correlation between occupational earnings premiums conditional on tenure and the occupational effects from Table 5 remains strongly negative, $\rho = -0.79$. Another possibility is that the correlation is driven by differences in occupational unemployment rates (i.e., higher rates reduce average earnings and increase volatility). Indeed, the correlation is even higher (-0.87, and -0.86 controlling for tenure) if the exercise is repeated using hours worked in place of earnings. Yet the correlation is also substantially negative (-0.56, and -0.52 controlling for tenure) for hourly wages, indicating that the pattern is not entirely due to variation in unemployment. The theory of compensating differentials would have suggested a positive correlation, so one suspects that the large negative correlation reflects workers who command an earnings

(or wage) premium use some portion of it to "buy" greater stability. Proving that suspicion would obviously require controlling for selection bias—a task that is beyond the scope of a paper that primarily aims to provide descriptive evidence, but an interesting question for future work.

5.2 Robustness Check: Longitudinal Estimates

One potential objection to the analysis above is that it uses a pooled regression that does not take advantage of the longitudinal nature of the data. For example, one could imagine running equation (6) with person-level heterogeneity in baseline levels of earnings risk:

$$E[\,|\widehat{\epsilon_{igt}}|\,|\,Z_{igt}] = \delta_i + \delta_t + Z_{igt}\delta. \tag{7}$$

Depending on what one assumes, this model could be implemented either as a random effects regression or a fixed effects regression. We would estimate a random effects regression if we were willing to assume that the heterogeneity were normally distributed ($\delta_i \sim N(0, \sigma_\delta^2)$) and that the person-specific effects δ_i are uncorrelated with observable characteristics and circumstances Z. The latter assumption would be especially objectionable if one wanted to measure causal effects of Z on $E[\,|\widehat{\epsilon_{igt}}|]$, as it ignores our original concern that persons who are inherently prone to earnings risk may be systematically more likely to enter some circumstances than others. Still, random effects estimates would provide descriptive evidence about which agents face the most earnings risk, and they may be more precise estimates of that relationship than the pooled least squares estimates presented earlier.

The other common alternative is to treat δ_i as a fixed effect. The effects of circumstances Z on earnings risk would then be measured entirely from within-person changes in shock magnitude—for instance, the effect of self-employment on earnings risk would be measured as the increase in the magnitude of earnings residual when a person is self-employed compared to when he is not. Although it might be tempting, it would also be premature to interpret that estimate as a causal effect because it still could be either (a) that the change in circumstances were due to changes in the person's underlying earnings risk (e.g., if people were more willing to become self-employed when they felt there was more stable demand for their services), or (b) that some third factor affected both the person's circumstances and his earnings risk (e.g., if some workers become

self-employed after suffering health shocks that cause their hours worked to fluctuate more than usual). That said, the fixed effects approach would at least distinguish between permanent differences in individuals' earnings risk and differences that arise with (if not necessarily because of) changes in circumstances.

The fixed effects approach also has some other shortcomings that may be more damning. For one thing, it is impossible to investigate the association between earnings risk and permanent personal characteristics in a fixed effects model. This seems unfortunate given that some interesting covariates of earnings risk do not change over individuals' lives—for example, race. Another difficulty is that when panels are short, fixed effects estimates may falsely attribute large shocks to permanent interpersonal differences. This is a real concern in our sample, where the average subject is observed for about 8 years. To see this, suppose that a person is in the sample for five years, experienced only small earnings shocks for four of those years, but a large shock in the other year that is due to a change in occupation. A fixed effects regression would then attribute one-fifth of the large shock to the individual-specific fixed effect, which could then lead to substantial underestimation of the occupation effect. More generally, fixed effects models are usually considered preferable when the sample contains nearly the complete universe of possible cases (such as if one had a data from most or all of the 50 U.S. states), but that is not a good description of our data.[9]

Table 8 presents the results of both specifications. To facilitate comparison, the first column repeats estimates from pooled regressions (Table 5), and then the second and third columns respectively report the results from random- and fixed effects versions of regression (7). The final column reports estimates from a more elaborate random effects regression that allows heterogeneous growth in earnings risk over the life-cycle:

$$E[\,|\widehat{\epsilon_{igt}}|\,|\,Z_{igt}] = \delta_i + \delta_i^* A_{igt} + \delta_t + Z_{igt}\delta,$$

where (δ_i, δ_i^*) is joint-normally distributed with arbitrary covariance. Only the most interesting estimates are reported. In particular, estimated year and race effects are not shown because they are nearly identical across methods (except that race effects cannot be estimated at all in a fixed effects regression), and occupations and industries are shown only when estimates are statistically significant in at least one specification.

[PLACE TABLE 8 ABOUT HERE]

Several interesting conclusions emerge from Table 8. First, the two sets of random effects estimates are

very similar to one another, and they are generally quite similar to the estimates from the pooled regressions reported in Table 5. If anything, the relationship between education and earnings risk looks a bit stronger in the random effects estimates. Earnings risk is still rises a bit with college graduation, but to a lesser extent than before, and it now seems to drop a bit more with graduate studies. The new results also show a somewhat clearer U-shaped pattern of earnings risk over the life-cycle, but again the difference is not dramatic. In most cases, the estimated self-employment, industry, and occupation effects are similar to those reported in Table 5, the largest difference being that some of the larger positive estimates are now smaller in magnitude (although still typically positive and statistically significant). The reduction in magnitude indicates that workers in those occupations (farm managers, cooks and waiters, insurance representatives, janitors, and workers in other or unknown occupations, as well as fully- or partially self-employed workers) experienced relatively large earnings shocks even in years in which they worked in other jobs. On the other hand, workers in the mechanical engineering and insurance industries now appear to be modestly less shielded from earnings risk than indicated by the pooled regression estimates, apparently because the types of workers working in those industries experienced relatively small shocks even when they worked in other jobs.

The fixed effects estimates are also qualitatively similar to the pooled estimates, especially for the self-employment, industry, and occupation dummies. There are some more notable differences in the estimated education effects, however—earnings risk now appears to be monotonically decreasing in education. (Note that this finding is consistent with the unconditional correlation we saw in Table 4.) This difference is informative in light of the differences between the estimation strategies. The fixed effects estimates measure changes in earnings risk as individuals acquire more education, so the result means that a given worker faces less risk as he earns additional degrees. This is probably not too surprising considering that individuals are probably more likely to hold relatively itinerant jobs while they are in school. In contrast, pooled and random effects estimates obtained some identification from differences across individuals. Thus, the difference between the fixed and random (and pooled) results suggests that some persons who pursued higher education may have faced relatively higher levels of baseline earnings risk than those who did not, but that risk then decreased as they obtained more schooling.

The other notable difference between the fixed and random effects results lies in the estimated age effects. The fixed effects estimates show much higher levels of earnings risk at the beginning and end of workers'

careers, and the bottom of the U-shaped age-risk profile is earlier and sharper than in the pooled or random effects results. The difference between the random and fixed effects results indicates that person-specific effects are correlated with the ages at which the individuals are observed in the sample. Considering that all individuals pass through each age during their careers, this may seem like a peculiar correlation, but it may be possible if earnings risk differs across birth cohorts (especially since we only observe workers' earnings for part of their careers). For instance, the two sets of estimates could be reconciled if individuals observed at ages greater than 40 (and possibly those observed at ages less than 20) typically faced less earnings risk than individuals who were observed mainly at ages between 20 and 40. Since the earnings data are from 1980 to 1997, that would suggest that the greatest baseline earnings risk would confront persons born in or near the Baby Boom era (1946-1964). Other explanations may also be possible, however.

While these differences are interesting and perhaps instructive, we should not let them obscure the fact that most estimates are quantitatively robust across all four specifications, and most of the remaining results are qualitatively robust. There is little here to dispute the conclusions from Section 5.1, and indeed most of the evidence confirms them.

5.3 Predictive Power

Having established that there are substantial differences in the magnitudes of residuals across a number of characteristics, two questions remain. First, do those differences actually correspond to what we think they do, labor market risks? Second, is there any indication that our measure of those risks has predictive power for cross-sectional differences in behavior?

The results in Table 9 provide some preliminary evidence on these questions. The top panel provides information on the cross-sectional differences in our measures of risk, while the lower panel shows the estimated slope coefficients ($\partial \Pr(Y)/\partial x$) from a series of probits that uses our measure of risk as an independent variable. Each probit also includes as explanatory variables all of the characteristics Z that were used to estimate the measure of risk, except for the industry and occupation dummy variables. The risk effects are thus measured only from the variation in the workers' most recent jobs. Since workers who could not be associated with any occupation or industry exhibited greater volatility than other workers, they have been excluded from the sample in order to avoid any biases that might arise from our ignorance or from the unusual

nature of those individuals.

[PLACE TABLE 9 ABOUT HERE]

The first two probits address the first question from above. The dependent variable is a dummy variable that equals 1 if the person reported zero annual hours or earnings (even though the person claimed to be part of the labor force at the previous interview date), and 0 if the person has positive annual hours and earnings. The probit thus asks whether $E[\,|\widehat{\epsilon}|\,|\,Z]$ helps to predict which workers do not work for an entire year, as it should if those workers really have suffered an adverse shock and if the measure actually does correspond to exposure to those risks. The measure of risk is (out of necessity) predicted out-of-sample for persons who did not report positive hours or earnings, so if anything the method is biased against finding that it influences the probability of belonging to that group.

Nevertheless, the probit results show that estimated risk is strongly associated with a higher probability of not working: at the mean, a one standard deviation increase estimated earnings risk raises by 45 percent the predicted probability of reporting zero hours or earnings, and a one standard deviation increase in estimated hours risk raises that risk by 60 percent. It is also reassuring that the latter figure is somewhat higher—after all, these people are not claiming to have experienced a fluctuation in wages, but rather long-term unemployment. If the probit is run on just the measure of hours risk and no other variables, the estimated increase is even larger (95 percent), and the estimated effect is even more statistically significant (Z-statistic=20.57). To gauge the association another way, if we obtain predicted values from a probit of the dummy variable on Z, including the occupation and industry dummies but excluding the measure of risk, that predicted risk of non-work conditional on Z has a high correlation with the estimate of hours risk, $\rho = 0.52$. Given that these out-of-sample predictions are so successful, it seems safe to conclude that the predicted values generated above measure exactly what we think they do. It should also ease any concerns that one might have had about excluding observations with zero hours or earnings from the sample we used to construct our measure of risk.

The remaining probits ask whether our estimate of earnings risk helps to predict three behaviors that are of broad interest: the probability that married people divorce within the next year, that a newly formed couple decides to get legally married rather than cohabit, and that the respondent is in jail at the time of the

next interview. It is important to realize that none of these estimates are intended to be definitive—on the contrary, we not addressing important issues of causality, and at any rate these topics warrant much greater attention than can be provided here. The point of these estimates is simply to establish the relevance of our measure of risk by showing that it is sufficiently correlated with these behaviors (and presumably others) to merit further attention.

Previous studies have found that divorce is related to earnings shocks (Weiss and Willis, 1997), so it is not surprising that the probits indicate a relatively strong relationship between earnings risk and divorce. What is most interesting is that the effect of earnings risk has a similar magnitude to that of earnings itself: a one standard deviation decrease in a man's actual log earnings is associated with a 21 percent increase in the risk of divorce, whereas a one standard deviation increase in his earnings risk predicts a 17 percent increase in the risk of divorce. However, as the fourth probit shows, there is a greater discrepancy if we distinguish between the man's expected log earnings (i.e., his permanent income predicted by regression (1)) and his current deviation from it (the residual). In that case, the estimated effect of a standard deviation increase in risk falls to a 13 percent increase in the risk of divorce, but a standard deviation decrease in permanent earnings predicts a 34 percent increase in the risk of divorce—a larger effect than for actual earnings. Transitory earnings shocks do not appear to have an independent effect, however. This is quite sensible—temporary shocks should be irrelevant except insofar as they are likely to be repeated in the future.

The fifth probit indicates that choices between marriage and cohabitation are more closely related to earnings risk than to actual earnings.[10] The latter has little predictive power, but a one standard deviation increase in estimated earnings risk raises the probability that a new match will be a cohabitation by 16 percent. This finding is consistent with the hypothesis that some couples cohabit to hedge against uncertainty.

The final probit shows that persons with greater earnings risk are more likely to go to jail. Since these probits control for the demographic factors used to estimate earnings risk, this result cannot be attributed to life-cycle, racial, or educational effects. A standard deviation increase in earnings risk predicts a 19 percent increase in the probability of going to jail, versus a 34 percent increase from a comparable decrease in actual earnings. Although the direction of causality is far from clear, as with all of these issues, there appears to be a meaningful relationship to explore.[11]

6 Conclusion

At the very least, the results discussed above demonstrate that our measure of earnings volatility varies substantially and robustly across factors in ways that correspond to both our intuition and to the limited evidence that was previously available. Insofar as our ultimate goal is to measure the effects of labor market risk on behavior, it seems especially useful that there are large differences across occupations and industries, as it is not obvious why within-person changes in those job characteristics would influence a person's decisions about, e.g., savings or marriage apart from their effects on the person's expected lifetime earnings and labor market risks.

The paper has also identified a number of promising topics for additional research. In closing, it seems worthwhile to reflect briefly on some other possible extensions. First, it would be interesting to know how well observed factors can account for international differences in the level of earnings instability. For example, Burkhauser, Holtz-Eakin, and Rhody (1997) estimate that the standard deviation of earnings shocks is about 28 percent higher among U.S. men than among German men, but we also know that a higher percentage of men in the U.S. are younger.[12] One might thus wonder whether the difference in earnings risks would persist if the age distributions were identical. A back of the envelope calculation using the 1993 age distributions suggests and the raw standard deviations of earnings shocks reported in Table 4 suggests that U.S. earnings instability would be about 1 percent lower if the U.S. population had the German age distribution. This is small compared to the estimated difference between the countries, but it is possible that the gap might be closed further (or for that matter, expanded) if we also considered differences in other factors like education and industrial composition. That said, while such an exercise is potentially quite interesting, caution is warranted any time earnings instability is compared across data sets because differences in measurement or reporting error would bias estimated differences in earnings risk. Thus, if the question is really whether workers face greater earnings risk in one country or another, the answer may have to wait until there is an international longitudinal data set that collects earnings data in the same way across countries. On the other hand, as in the calculation above, international difference in observed factors may provide a useful benchmark for counterfactual analyses.

Likewise, it would be quite interesting to see whether the covariates of earnings volatility differ across

countries. This may be a more feasible exercise than the one proposed in the last paragraph because measurement error would only pose a problem if its variance were correlated with those covariates. While this paper has found large differences in earnings instability across industries within the United States, it is far from clear that the same would be true in a country with more extensive unionization and greater regulation of the labor market. If not, then the difference may constitute and important benefit or cost of those institutions. On the other hand, if the covariates were similar across countries with different institutions, that would suggest that cross-industry difference in volatility are related to differences in competition or technology. Moreover, if the data allowed us to measure trends in industries' effects on the volatility of hours and wages, it may also be possible to learn something about changes over time in industries' technology or competitiveness, whether that is related to increased international competition, greater availability of information, faster technological change, or changes in the relative importance of general or job-specific human capital.

It would be useful to know whether the same covariates predict after-tax earnings instability. Like most of the literature, this paper has focused on pre-tax earnings, but Kniesner and Ziliak (2002a, 2002b) have shown that earnings stabilization is an important benefit of income taxation. By comparing the covariates of pre- and post-tax earnings instability, one could identify which taxpayers benefit the most from this stabilization. Perhaps more importantly, if some other people have highly unstable after-tax earnings, it may be possible to design a welfare-improving revision to the tax code.

Another possible extension is to compare covariates of earnings risk across data sets within a given country. In principle, one might try to conduct the same analysis on different data in order to test the robustness of the empirical results, but that may not be very practical in light of differences between samples. The only other long-term longitudinal study of earnings in the U.S. is the National Longitudinal Survey of Youth (NLSY), which unlike the PSID covers only individuals born in a single cohort. That said, one might compare the results in this paper to those of a similar analysis on the NLSY as a way of distinguishing between age and cohort effects.

Finally, insofar as we are interested in the causes and effects of earnings risk, it may be helpful to supplement this analysis with a study of heterogeneity in risk tolerance. Several recent papers have used questions from the PSID and other data sets to compute measures of individuals' willingness to bear risk (Barsky *et al.*, 1997; Brown and Taylor, 2006; Brown *et al.*, 2007; Kimball, Sahm, and Shapiro, forthcoming).

If these measures are considered reasonably reliable, one could imagine asking how they vary with individuals' actual exposure to earnings risk. Since preferences are likely among the most important factors that have not been observed in this paper, including them in an empirical analysis would be a big step toward identifying causal effects on earnings risk. Moreover, the ability to control for risk tolerance directly would overcome one of the most important challenges facing those who wish to measure causal effects of risk on behavior.

To be sure, we will only be able to answer such questions if we have some confidence in our ability to measure both earnings volatility and its variation across people. Volatility is inherently difficult to measure because doing so requires us to imagine other outcomes that could have occurred under different realizations of uncertainty. It is thus hoped that the methods developed here will be a first step toward overcoming those challenges and answering the intriguing questions that rest upon them.

Notes

[1]The SEO pre-dates the main sample, which began in 1968. It was first drawn by the Bureau of the Census, and a subsample was followed after the PSID began. See Hill (1992).

[2]Note that this specification implicitly assumes that individuals can predict macroeconomic shocks. Little is lost by excluding aggregate risks from our risk index because they do not vary across workers and there is little hope of insuring against them.

[3]In principle, equation (1) could be estimated directly (e.g., with Stata's *xtmixed* command), but computational considerations lead us to estimate it in two steps, first estimating Q from Y and then estimating R from the residuals. This is standard in the literature.

[4]See Guvenen (2007) for a recent critical survey of earnings dynamics models. He advocates models that allow heterogeneity across persons in rates of earnings growth, as in our regression (1) and Haider's (4).

[5]The correlation is even higher between $\widehat{\epsilon}$ estimated from second- and third-order specifications of R, greater than 0.999 for all three dependent variables.

[6]The listed standard errors adjust for heteroskedasticity and residuals correlated at the individual level.

[7]Unlike self-employment, working as a farm manager does not convey a large earnings premium. The mean log earnings for that group is 9.61, versus 10.02 for all other occupations.

[8]It would be interesting to know whether such a pattern can explain why earnings risk is so much higher among households in the bottom third of the Canadian income distribution (Morissette and Ostrovsky, 2005).

[9]The relative merits of fixed and random effects models have been widely discussed. See, e.g., Wooldridge (2002, pp. 247-291).

[10]This probit uses only data from after 1982, since earlier waves of the PSID did not distinguish between marriage and cohabitation.

[11]This may seem to conflict with Dahlberg and Gustavsson's (2005) finding that Swedish regional crime rates vary with changes in permanent inequality, but not with changes in transitory inequality. However, the effects of within-region fluctuations may differ considerably from those of (semi-permanent) cross-sectional differences in transitory inequality.

[12]This is easy to see from the population age pyramids available from the U.S. Census Bureau's International Data Base (*http://www.census.gov/ipc/www/idb/*).

References

[1] Baker, Michael, and Gary Solon, "Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records," *Journal of Labor Economics,* 21(2), 289-322, April 2003.

[2] Barsky, Robert B., F. Thomas Juster, Miles S. Kimball, and Matthew D. Shapiro, "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Survey," *Quarterly Journal of Economics*, 112(2), 537-579, May 1997.

[3] Beach, Charles M., Ross Finnie, and David Gray, "Earnings Variability and Earnings Instability of Women and Men in Canada: How Do the 1990s Compare to the 1980s?" *Canadian Public Policy*, 29 (Supplement), S41-S63, January 2003.

[4] Blundell, Richard, and Ian Preston, "Consumption Inequality and Income Uncertainty," *Quarterly Journal of Economics*, 113(2), 603-640, May 1998.

[5] Bostic, Raphael W., "Racial Differences in Short-Run Earnings Stability and Implications for Credit Markets," *Federal Reserve Board Finance and Economics Discussion Series*, 1997-34, June 1997.

[6] Bound, John, Charles Brown, Greg J. Duncan, and Willard L. Rodgers, "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 12(3), 345–368, July 1994.

[7] Bound, John, Charles Brown, and Nancy Mathiowetz, "Measurement Error in Survey Data," in James J. Heckman and Edward Leamer, eds., *Handbook of Econometrics, Volume 5*, 3705-3843, Elsevier Science, Amsterdam, 2001.

[8] Brown, Sarah, Michael Dietrich, Aurora Ortiz, and Karl Taylor, "Self-Employment and Risk Preference," *Sheffield Economic Research Paper Series*, number 2007008, 2007.

[9] Brown, Sarah, Lisa Farrell, Mark N. Harris, and John G. Sessions, "Risk Preference and Employment Contract Type," *Journal of the Royal Statistical Society, Series A*, 169(4), 1-15, October 2006.

[10] Brown, Sarah, and Karl Taylor, "Education, Risk Preference and Wages," *Sheffield Economic Research Paper Series*, number 2006002, 2006.

[11] Burkhauser, Richard V., Barbara A. Butrica, Mary C. Daly, and Dean R. Lillard, "The Cross-National Equivalent File: A product of cross-national research," working paper, Cornell University, 2000.

[12] Burkhauser, Richard V., Douglas Holtz-Eakin, and Stephen E. Rhody, "Labor Earnings Mobility and Inequality in the United States and Germany During the Growth Years of the 1980s," *International Economic Review*, 38(4), 775-794, November 1997.

[13] Chen, Stacey H., "Estimating the Variance of Wages in the Presence of Selection and Unobservable Heterogeneity," *Review of Economics and Statistics*, 90(2), 275-289, May 2008.

[14] Dahlberg, Matz, and Magnus Gustavsson, "Inequality and Crime: Separating the Effects of Permanent and Transitory Income," *Oxford Bulletin of Economics and Statistics*, 70(2), 129-153, April 2008.

[15] Diaz-Serrano, Luis, "Labor Income Uncertainty, Skewness and Homeownership: A Panel Data Study for Germany and Spain," *Journal of Urban Economics*, 58(1), 156-176, July 2005a.

[16] ———, "Income Volatility and Residential Mortgage Delinquency across the EU," *Journal of Housing Economics*, 14(3), 153-177, September 2005b.

[17] Dickens, Richard, "The Evolution of Individual Male Earnings in Great Britain: 1975-95," *Economic Journal*, 110 (460), 27-49, January 2000.

[18] Duncan, Greg J., and Daniel H. Hill, "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data," *Journal of Labor Economics*, 3(4), 508-532, October 1985.

[19] Duncan, Greg J., and Nancy A. Mathiowetz, with Charles F. Cannell, Daniel H. Hill, and Michael Ponza, *A Validation Study of Economic Survey Data*, Survey Research Center, Ann Arbor, MI, 1985.

[20] Freije, Samuel, and André Portella Souza, "Earnings Dynamics and Inequality in Venezuela: 1995-1997," Vanderbilt University, Department of Economics, working paper number 02-W11, May 2002.

[21] Friedman, Milton, "Choice, Chance, and the Personal Distribution of Income," *Journal of Political Economy*, 61(4), 277-290, August 1953.

[22] ———, *A Theory of the Consumption Function*, Princeton University Press, Princeton, NJ, 1957.

[23] Glejser, H., "A New Test for Heteroskedasticity," *Journal of the American Statistical Association*, 64(325), 316-325, March 1969.

[24] Gordon, Roger H., *Differences in Earnings and Ability*, Garland Publishing, Inc., New York, 1984.

[25] Gottschalk, Peter, and Robert Moffitt, "The Growth of Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity*, 1994(2), 217-254, 1994.

[26] Grossman, Gene M., and Carl Shapiro, "A Theory of Factor Mobility," *Journal of Political Economy*, 90(5), 1054-1069, October 1982.

[27] Guvenen, Fatih, "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?" *American Economic Review*, 97(3), 687-712, June 2007.

[28] Haider, Steven J., "Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991," *Journal of Labor Economics*, 19(4), 799-836, October 2001.

[29] Haurin, Donald R., and H. Leroy Gill, "Effects of Income Variability on the Demand for Owner-Occupied Housing," *Journal of Urban Economics*, 22(2), 136-150, September 1987.

[30] Hill, Martha S., *The Panel Study of Income Dynamics: A User's Guide, Volume 2*, Sage Publications, Newbury Park, CA, 1992.

[31] Hyslop, Dean R., "Rising U.S. Earnings Inequality and Family Labor Supply: The Covariance Structure of Intrafamily Earnings," *American Economic Review*, 91(4), 755-777, September 2001.

[32] Kennedy, Bruce, "Mobility and Instability in Canadian Earnings," *Canadian Journal of Economics*, 22(2), 383-394, May 1989.

[33] Kimball, Miles S., Claudia R. Sahm, and Matthew D. Shapiro, "Imputing Risk Tolerance from Survey Responses," *Journal of the American Statistical Association: Applications*, forthcoming.

[34] Kniesner, Thomas J., and James P. Ziliak, "Tax Reform and Automatic Stabilization," *American Economic Review*, 92(3), 590-612, June 2002a.

[35] ———, "Explicit Versus Implicit Income Insurance," *Journal of Risk and Uncertainty*, 25(1), 5-20, July 2002b.

[36] Koenker, Roger, and Gilbert Bassett, "Regression Quantiles," *Econometrica*, 46(1), 33-50, January 1978.

[37] Krueger, Dirk, and Fabrizio Perri, "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *Review of Economic Studies*, 73(1), 163-193, January 2006.

[38] Moffitt, Robert A., and Peter Gottschalk, "Trends in the Covariance Structure of Earnings in the U.S.: 1969–1987," working paper, Johns Hopkins University and Boston College, 1995.

[39] ———, "Trends in the Transitory Variance of Earnings in the United States," *Economic Journal*, 112(478), C68-C73, March 2002.

[40] Morissette, René, and Yuri Ostrovsky, "The Instability of Family Earnings and Family Incomes in Canada, 1986-1991 and 1996-2001," *Canadian Public Policy–Analyse de Politiques*, 31(3), 273-302, September 2005.

[41] Oppenheimer, Valerie, "A Theory of Marriage Timing," *American Journal of Sociology*, 94(3), 563-591, November 1988.

[42] Orazem, Peter F., and J. Peter Mattila, "Human Capital, Uncertain Wage Distributions, and Occupational and Educational Choices," *International Economic Review*, 32(1), 103-122, February 1991.

[43] Pischke, Jörn Steffen, "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study," *Journal of Business & Economic Statistics*, 13(3), 305-314, July 1995.

[44] Rodgers, Willard L., Charles Brown, and Greg J. Duncan, "Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages," *Journal of the American Statistical Association*, 88(424), 1208-1218, December 1993.

[45] Rosen, Harvey, and Paul Willen, "Risk, Return, and Self-Employment," working paper, Princeton University and University of Chicago, 2002.

[46] Santos, Antonio Loureiro, and André Portella Souza, "Earnings Inequality in Brazil: Is it Permanent or Transitory?" *Brazilian Review of Econometrics*, 27(2), 259-283, November 2007.

[47] Schultz, Theodore W., "The Value of the Ability to Deal with Disequilibria," *Journal of Economic Literature*, 13(3), 827-846, September 1975.

[48] Shore, Stephen H., and Todd Sinai, "Commitment, Risk, and Consumption: Do Birds of a Feather Have Bigger Nests?" *Review of Economics and Statistics*, forthcoming.

[49] Smith, Adam, *An Inquiry into the Nature and Causes of the Wealth of Nations*, 1994 Modern Library Edition, Edwin Cannan, ed., The Modern Library, New York, 1776.

[50] Weiss, Yoram , and Robert J. Willis, "Match Quality, New Information, and Marital Dissolution," *Journal of Labor Economics*, 15(1, part 2), S293-S329, January 1997.

[51] Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA, 2002.

Table 1: Distributions of Key Variables, by Race and Sample

	Log Earnings				Log Hours				Log Wages				Age				Years of Education			
	White		Black		White		Black		White		Black		White		Black		White		Black	
Percentile	Main	SEO	Main	SEO	Main	SEO	Main	SEO	Main	SEO	Main	SEO	Main	SEO	Main	SEO	Main	SEO	Main	SEO
0	4.61	4.61	4.61	4.62	1.10	2.83	0.69	1.39	-8.25	-7.07	-7.66	-7.41	16	16	16	16	1	2	1	3
5	8.77	8.54	7.43	7.70	6.95	6.79	5.99	6.26	1.48	1.41	1.02	0.99	21	21	20	21	10	9	7	8
10	9.30	9.10	8.27	8.46	7.31	7.19	6.68	6.95	1.77	1.68	1.38	1.38	23	23	22	23	11	10	9	10
15	9.55	9.39	8.74	8.91	7.46	7.39	7.00	7.24	1.95	1.85	1.54	1.57	25	25	24	24	12	11	10	10
20	9.72	9.58	9.02	9.17	7.54	7.50	7.27	7.40	2.08	1.99	1.68	1.71	27	26	25	26	12	12	11	11
25	9.86	9.74	9.20	9.35	7.57	7.55	7.41	7.50	2.19	2.11	1.78	1.82	28	28	26	27	12	12	11	11
30	9.96	9.86	9.36	9.48	7.59	7.58	7.50	7.55	2.29	2.20	1.88	1.91	29	29	27	28	12	12	11	12
35	10.06	9.95	9.48	9.59	7.61	7.60	7.55	7.58	2.37	2.29	1.96	2.01	31	30	28	29	12	12	12	12
40	10.15	10.05	9.58	9.69	7.62	7.61	7.57	7.59	2.45	2.37	2.04	2.08	32	32	30	30	12	12	12	12
45	10.23	10.14	9.68	9.78	7.64	7.63	7.59	7.60	2.53	2.45	2.10	2.15	33	33	31	31	12	12	12	12
50	10.31	10.22	9.76	9.86	7.67	7.65	7.60	7.62	2.61	2.53	2.18	2.22	35	34	32	33	13	12	12	12
55	10.38	10.28	9.85	9.96	7.70	7.69	7.61	7.63	2.68	2.60	2.25	2.30	36	36	33	34	13	13	12	12
60	10.45	10.36	9.94	10.03	7.73	7.72	7.63	7.64	2.76	2.67	2.33	2.37	38	37	35	35	14	13	12	12
65	10.53	10.44	10.04	10.10	7.76	7.76	7.64	7.66	2.83	2.74	2.41	2.45	39	39	36	36	14	14	12	12
70	10.61	10.51	10.13	10.18	7.80	7.80	7.65	7.70	2.91	2.82	2.50	2.54	41	41	38	38	16	14	12	13
75	10.69	10.61	10.22	10.27	7.82	7.84	7.70	7.74	2.99	2.90	2.60	2.62	43	43	40	39	16	15	13	13
80	10.78	10.70	10.32	10.36	7.86	7.88	7.75	7.78	3.08	2.99	2.69	2.70	46	46	43	41	16	16	14	14
85	10.90	10.81	10.43	10.46	7.92	7.95	7.81	7.83	3.19	3.09	2.80	2.82	48	49	46	43	16	16	14	14
90	11.06	10.96	10.57	10.58	8.00	8.03	7.87	7.92	3.34	3.23	2.93	2.93	52	53	50	47	17	16	15	16
95	11.30	11.25	10.75	10.77	8.09	8.13	8.03	8.05	3.58	3.48	3.12	3.10	57	58	55	53	17	17	16	16
100	14.05	13.30	12.16	12.16	8.87	8.78	8.67	8.67	6.74	5.24	5.37	7.62	65	65	65	65	17	17	17	17
# obs.	38,045	5,731	10,730	8,516	38,045	5,731	10,730	8,516	38,045	5,731	10,730	8,516	38,045	5,731	10,730	8,516	38,045	5,731	10,730	8,516

Table 2: Sample Characteristics

	Mean or %		%		%
Dependent Variables		Education		Industries (2-digit)	
ln(earnings)	10.01	Less than 12 years	18.1	Retail	10.8
ln(hours)	7.57	12 years	37.9	Other services	6.7
ln(wage)	2.43	13-15 years	20.8	Construction related	6.7
		16 years	14.3	Public administration	6.7
Age		17+ years	9.0	Mechanical engineering	6.7
Age 16-20	4.3			Other transportation	4.8
Age 21-25	13.6	Self-employment		Wood/paper/print	4.6
Age 26-30	19.0	Not self-employed	89.0	Wholesale	4.5
Age 31-35	18.9	Partially self-employed	1.2	Education/sport	4.2
Age 36-40	15.2	Fully self-employed	9.8	Health service	3.1
Age 41-45	10.7			Agriculture/forestry	3.0
Age 46-50	7.2	Occupations		Energy/water	2.9
Age 51-55	5.4	Priv. business leader	11.6	Construction	2.9
Age 56-60	4.0	Transportation operator	6.8	Iron/steel	2.7
Age 61-65	1.7	Machine fitter	6.7	Electrical engineering	2.7
		Labor/craftsman	5.0	Postal system	2.5
Birth Year		Bricklayer/carpenter	4.3	Food industry	2.1
Before 1935	8.7	Conveyor operator	4.2	Clothing/text	2.0
1935-44	12.0	Vendor	3.3	Legal services	1.7
1945-50	16.3	Architect/engineer	2.7	Service industry	1.5
1951-55	19.1	Inspector	2.6	Chemicals	1.5
1956-60	20.3	Electrical engineer	2.5	Financial institutions	1.3
1961-65	14.2	Pipe fitter	2.5	Volunteer/church	1.3
1966-70	6.9	Security service	2.4	Insurance	1.2
After 1970	2.6	Janitor	2.2	Mining	1.1
		Educator	2.1	Synthetics	0.9
Race		Soldier	1.9	Earth/clay/stone	0.8
White	67.4	Office worker	1.9	Restaurants	0.7
Black	29.7	Farm manager	1.9	Train system	0.5
Native American	1.0	Mailman	1.7	Fisheries	0.1
Asian/Pacific Islander	0.4	Mathematician	1.6	Private household	0.0
Latino	0.7	Cook/waiter	1.5	Other/unknown	7.9
Other	0.9	Other (56 categories)	30.4		
Number of observations:	64,915	Unique individuals:	7,389		
Main sample	50,365	Main sample	5,547		
SEO	14,550	SEO	1,842		

Table 3: Results from the First-Stage Regressions

Dependent Variable:	Earnings			Hours			Wages		
Order of R (random polynomial in age)	**0**	**1**	**2**	**0**	**1**	**2**	**0**	**1**	**2**
Distribution of parameters in R									
std dev (γ_0)	0.59	1.28	2.55	0.31	0.81	2.09	0.42	0.81	1.54
std dev (γ_1) x 10		0.37	1.44		0.20	1.09		0.26	0.95
std dev (γ_2) x 100			0.17			0.13			0.12
corr(γ_0,γ_1)		-0.90	-0.95		-0.95	-0.98		-0.88	-0.95
corr(γ_0,γ_2)			0.86			0.95			0.83
corr(γ_1,γ_2)			-0.97			-0.99			-0.96
Distribution of residuals (ε):									
std dev (ε)	0.51	0.48	0.47	0.38	0.37	0.36	0.44	0.42	0.41
Centiles (ε)									
10	-0.44	-0.39	-0.38	-0.27	-0.25	-0.24	-0.38	-0.35	-0.34
20	-0.21	-0.19	-0.18	-0.13	-0.12	-0.11	-0.21	-0.19	-0.19
30	-0.11	-0.10	-0.09	-0.06	-0.06	-0.06	-0.12	-0.11	-0.10
40	-0.03	-0.03	-0.03	-0.02	-0.02	-0.02	-0.05	-0.05	-0.05
50	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.01
60	0.10	0.09	0.08	0.07	0.07	0.06	0.07	0.06	0.06
70	0.18	0.16	0.15	0.12	0.11	0.11	0.14	0.12	0.12
80	0.28	0.25	0.24	0.19	0.18	0.18	0.22	0.20	0.19
90	0.45	0.41	0.40	0.32	0.30	0.29	0.36	0.33	0.32
Correlations of ε across specifications	1.000	0.972	0.954	1.000	0.988	0.969	1.000	0.976	0.958
		1.000	0.992		1.000	0.988		1.000	0.990
			1.000			1.000			1.000
Log L	-57215	-56005	-55376	-35695	-34966	-34282	-46298	-45294	-44747

Note: For each dependent variable, the table reports results from three different specifications of regression (1) that differ only in the order of the random polynomial in age, R(A,γ). The first specification includes R as a zero-order polynomial--i.e., it uses only person-specific random effects. The second specification adds a second random component that allows for constant unobserved heterogeneity in earnings growth rates. The final specification adds a third random component that also permits heterogeneity in earnings acceleration.

Table 4: Standard Deviations of Actual and Residual Log Earnings, Hours, and Wages, by Demographic Groups, Years, and Employment Characteristics

	Earnings ($1990)		Hours Worked		Hourly Wage			Earnings ($1990)		Hours Worked		Hourly Wage	
	Y	ε_{Yi}	H	ε_{Hi}	W	ε_{Wi}		Y	ε_{Yi}	H	ε_{Hi}	W	ε_{Wi}
All observations	0.92	0.44	0.52	0.33	0.76	0.32	Largest Industries						
							Agriculture/Forestry	1.09	0.62	0.53	0.35	1.06	0.46
By education							Other/Unknown	1.15	0.59	0.94	0.57	0.75	0.38
Less than HS	1.04	0.50	0.72	0.44	0.74	0.34	Legal Services	1.02	0.52	0.40	0.29	0.95	0.38
HS only	0.83	0.44	0.50	0.35	0.68	0.31	Service Industry	0.93	0.50	0.53	0.35	0.80	0.37
Some college	0.75	0.40	0.41	0.29	0.66	0.32	Construction	0.91	0.50	0.50	0.36	0.75	0.37
College grad	0.82	0.41	0.38	0.26	0.78	0.32	Other Services	0.95	0.48	0.48	0.35	0.85	0.36
Graduate studies	0.75	0.37	0.33	0.21	0.70	0.31	Construction Related	0.83	0.44	0.46	0.34	0.68	0.34
By age							Retail	0.87	0.44	0.48	0.33	0.73	0.34
Age 16-20	1.03	0.50	0.89	0.48	0.59	0.34	Health Service	1.00	0.43	0.42	0.30	0.89	0.31
Age 21-25	0.88	0.49	0.63	0.41	0.63	0.33	Other Transportation	0.79	0.43	0.46	0.32	0.70	0.32
Age 26-30	0.85	0.45	0.50	0.36	0.68	0.32	Wholesale	0.83	0.43	0.39	0.29	0.74	0.31
Age 31-35	0.82	0.42	0.44	0.31	0.70	0.30	Iron/steel	0.69	0.41	0.44	0.35	0.54	0.29
Age 36-40	0.84	0.41	0.41	0.29	0.76	0.32	Clothing/Text	0.72	0.39	0.40	0.31	0.59	0.27
Age 41-45	0.80	0.41	0.38	0.27	0.75	0.31	Energy/Water	0.72	0.39	0.39	0.30	0.62	0.26
Age 46-50	0.83	0.41	0.41	0.29	0.78	0.32	Wood/Paper/Print	0.77	0.39	0.39	0.29	0.66	0.29
Age 51-55	0.82	0.40	0.41	0.30	0.75	0.32	Public Administration	0.80	0.38	0.38	0.28	0.75	0.27
Age 56-60	0.87	0.42	0.45	0.30	0.80	0.32	Food Industry	0.77	0.36	0.42	0.31	0.63	0.29
Age 61-65	0.95	0.44	0.54	0.31	0.80	0.34	Postal System	0.63	0.36	0.37	0.26	0.55	0.25
By race							Mechanical Engineer	0.64	0.36	0.34	0.27	0.58	0.26
Black	0.98	0.50	0.64	0.43	0.73	0.34	Electrical Engineer	0.73	0.34	0.32	0.24	0.66	0.26
Latino	0.97	0.49	0.47	0.33	0.78	0.33	Financial Institutions	0.74	0.34	0.33	0.25	0.64	0.30
White	0.84	0.40	0.44	0.29	0.73	0.31	Education/Sport	0.68	0.32	0.37	0.24	0.60	0.26
Native American	0.90	0.40	0.58	0.36	0.67	0.30	Chemicals	0.60	0.24	0.27	0.19	0.54	0.23
Other	0.92	0.38	0.50	0.30	0.76	0.36	Largest Occupations						
Asian/Pac. Islander	0.82	0.33	0.53	0.28	0.70	0.30	Farm manager	1.14	0.69	0.52	0.34	1.15	0.51
By year							Other/Unknown	1.16	0.59	0.93	0.57	0.76	0.38
1980	0.86	0.39	0.52	0.32	0.63	0.29	Cook/Waiter	0.97	0.57	0.62	0.47	0.77	0.39
1981	0.89	0.39	0.53	0.31	0.67	0.29	Bricklayer/Carpenter	0.85	0.49	0.52	0.39	0.69	0.35
1982	0.97	0.44	0.66	0.42	0.69	0.30	Janitor	0.86	0.47	0.53	0.36	0.65	0.32
1983	0.95	0.44	0.60	0.39	0.68	0.30	Labor/craftsman	0.83	0.46	0.52	0.39	0.66	0.31
1984	0.92	0.42	0.55	0.36	0.70	0.31	Conveyor Operator	0.82	0.46	0.49	0.38	0.68	0.32
1985	0.93	0.41	0.54	0.34	0.71	0.31	Pipe Fitter	0.75	0.45	0.45	0.35	0.63	0.29
1986	0.94	0.41	0.53	0.34	0.72	0.30	Transportation Operator	0.70	0.43	0.44	0.33	0.61	0.32
1987	0.90	0.39	0.49	0.31	0.70	0.30	Priv. Bus. Leader	0.80	0.42	0.33	0.24	0.80	0.34
1988	0.91	0.42	0.52	0.33	0.71	0.31	Related Medical Job	0.76	0.40	0.44	0.34	0.62	0.29
1989	0.91	0.42	0.49	0.33	0.72	0.31	Mailman	0.66	0.39	0.39	0.30	0.57	0.27
1990	0.88	0.40	0.47	0.30	0.70	0.30	Machine Fitter	0.67	0.39	0.39	0.30	0.59	0.29
1991	0.93	0.42	0.50	0.32	0.76	0.32	Office Worker	0.67	0.38	0.38	0.30	0.57	0.29
1992	0.98	0.53	0.56	0.41	0.88	0.42	Security Service	0.69	0.38	0.38	0.29	0.61	0.27
1993	0.93	0.52	0.43	0.30	0.91	0.35	Vendor	0.79	0.38	0.36	0.26	0.68	0.32
1994	0.89	0.45	0.43	0.29	0.83	0.33	Electrical Engineer	0.61	0.34	0.31	0.24	0.55	0.27
1995	0.90	0.46	0.40	0.28	0.87	0.33	Mathematician	0.50	0.32	0.25	0.19	0.48	0.24
1996	0.98	0.46	0.45	0.28	0.92	0.35	Educator	0.57	0.32	0.36	0.23	0.54	0.25
By self-employment							Inspector	0.57	0.31	0.30	0.23	0.50	0.23
Fully self-employed	1.06	0.59	0.47	0.31	1.05	0.48	Soldier	0.54	0.30	0.32	0.25	0.54	0.30
Partially self-emp.	0.84	0.42	0.41	0.28	0.80	0.37	Business Manager	0.63	0.29	0.28	0.20	0.59	0.26
Not self-employed	0.90	0.41	0.52	0.34	0.71	0.30	Architect/Engineer	0.46	0.28	0.24	0.18	0.43	0.25

Note: Where no natural order exists, categories are listed in descending order of residual earnings variance.

Table 5: Regression of Absolute Residual Earnings on Characteristics

Variable	Coef.	SE	t
Constant	0.221	0.014	16.01
Education			
Less than HS	0.041	0.007	6.05
Some college	-0.011	0.006	1.96
College grad	0.010	0.007	1.44
Graduate studies	0.009	0.009	1.01
Age			
Age 16-20	0.000	0.010	0.04
Age 21-25	0.024	0.006	4.26
Age 31-35	-0.028	0.005	5.94
Age 36-40	-0.040	0.006	6.94
Age 41-45	-0.049	0.007	7.50
Age 46-50	-0.048	0.007	6.38
Age 51-55	-0.048	0.008	5.71
Age 56-60	-0.036	0.010	3.68
Age 61-65	-0.013	0.013	0.99
Race			
Black	0.064	0.006	11.17
Native American	-0.023	0.017	1.32
Asian	0.015	0.026	0.57
Latino	0.057	0.027	2.07
Other race	0.002	0.020	0.12
Year			
Year 1980	-0.031	0.007	4.63
Year 1981	-0.026	0.006	4.08
Year 1982	0.004	0.007	0.55
Year 1983	0.008	0.007	1.12
Year 1984	-0.006	0.007	0.95
Year 1985	-0.004	0.006	0.66
Year 1986	0.001	0.006	0.12
Year 1987	-0.015	0.006	2.47
Year 1989	-0.005	0.006	0.88
Year 1990	-0.007	0.006	1.16
Year 1991	0.006	0.006	1.02
Year 1992	0.057	0.008	7.39
Year 1993	0.064	0.008	7.72
Year 1994	0.023	0.007	3.19
Year 1995	0.026	0.007	3.63
Year 1996	0.028	0.008	3.54
Self-employment			
Partially self-emp.	0.057	0.013	4.46
Fully self-emp.	0.152	0.009	17.80
Industries (relative to retail trade)			
Other Services	0.036	0.010	3.44
Legal Services	0.032	0.021	1.51
Agriculture/Forestry	0.020	0.023	0.87
Construction	0.018	0.014	1.29
Service Industry	0.009	0.020	0.45
Other/Unknown	0.005	0.022	0.23
Other Transportation	0.003	0.011	0.30
Restaurants	-0.005	0.025	0.18
Public Admininistration	-0.006	0.015	0.38
Mining	-0.008	0.019	0.41
Iron/Steel	-0.011	0.013	0.80
Industries (continued)			
Construction Related	-0.012	0.011	1.07
Health Service	-0.012	0.016	0.74
Wholesale	-0.013	0.011	1.21
Financial Institutions	-0.017	0.015	1.16
Electrical Engineer	-0.018	0.012	1.49
Postal System	-0.019	0.015	1.29
Volunteer/Church	-0.019	0.017	1.10
Mechanical Engineer	-0.023	0.010	2.36
Food Industry	-0.024	0.013	1.83
Clothing/Textiles	-0.026	0.016	1.63
Wood/Paper/Print	-0.028	0.011	2.62
Earth/Clay/Stone	-0.028	0.021	1.36
Synthetics	-0.034	0.019	1.75
Energy/Water	-0.036	0.014	2.66
Education/Sport	-0.051	0.013	3.81
Insurance	-0.057	0.018	3.25
Chemicals	-0.072	0.011	6.52
Occupations (relative to priv. bus. leaders)			
Other/Unknown	0.148	0.022	6.78
Farm manager	0.122	0.032	3.77
Cook/Waiter	0.110	0.021	5.26
Insurance Represenative	0.053	0.023	2.32
Pipe Fitter	0.049	0.014	3.45
Bricklayer/Carpenter	0.047	0.013	3.70
Laborer/Craftsman	0.041	0.011	3.75
Conveyor Operator	0.035	0.012	2.93
Janitor	0.030	0.015	2.01
Painter	0.026	0.020	1.35
Transportation Operator	0.019	0.010	1.84
Vendor	0.013	0.011	1.21
Tool/Die Maker	0.012	0.019	0.62
Related Medical Job	0.003	0.020	0.16
Mailman	0.001	0.016	0.09
Educator	0.000	0.015	0.01
Office Worker	-0.001	0.013	0.06
Machine Fitter	-0.003	0.010	0.33
Restaurant/Store Manager	-0.006	0.022	0.27
Tailor	-0.007	0.026	0.25
Food Producer	-0.007	0.016	0.43
Engineering Tech. Expert	-0.012	0.015	0.76
Inspector	-0.013	0.010	1.34
Electrical Engineer	-0.016	0.015	1.07
Doctor/Dentist/Veternarian	-0.017	0.028	0.62
Security Service	-0.019	0.017	1.16
Scientist	-0.021	0.015	1.39
Business Manager	-0.025	0.011	2.26
Soldier	-0.033	0.016	2.01
Architect/Engineer	-0.042	0.010	4.30
Mathematician	-0.043	0.015	2.83
Lawyer	-0.049	0.028	1.77
Accountant	-0.051	0.017	2.96
No. obs.	64,906		
R sq.	0.076		

Note: Regression also included dummies for the SEO sample, number of observations for individuals, and an additional 42 occupations and 3 industries with fewer than 400 observations.

Table 6: Regression of Absolute Residual Hours on Characteristics

Variable	Coef.	SE	t
Constant	0.159	0.010	16.15
Education			
Less than HS	0.045	0.006	7.89
Some college	-0.012	0.004	2.89
College grad	-0.008	0.005	1.42
Graduate studies	-0.012	0.006	2.20
Age			
Age 16-20	0.029	0.010	2.90
Age 21-25	0.022	0.005	4.50
Age 31-35	-0.021	0.004	5.81
Age 36-40	-0.030	0.004	6.76
Age 41-45	-0.035	0.005	7.29
Age 46-50	-0.035	0.006	6.22
Age 51-55	-0.033	0.007	5.05
Age 56-60	-0.026	0.007	3.48
Age 61-65	-0.008	0.010	0.82
Race			
Black	0.059	0.005	12.23
Native American	0.005	0.016	0.29
Asian	0.036	0.022	1.68
Latino	0.033	0.016	2.05
Other race	0.016	0.013	1.21
Year			
Year 1980	-0.023	0.006	4.05
Year 1981	-0.022	0.005	4.01
Year 1982	0.034	0.007	5.15
Year 1983	0.026	0.006	4.28
Year 1984	0.004	0.006	0.67
Year 1985	0.002	0.005	0.39
Year 1986	0.007	0.005	1.36
Year 1987	-0.009	0.005	1.82
Year 1989	0.000	0.005	0.02
Year 1990	-0.006	0.005	1.24
Year 1991	0.002	0.005	0.30
Year 1992	0.046	0.006	7.21
Year 1993	0.015	0.005	2.75
Year 1994	0.001	0.005	0.24
Year 1995	0.000	0.005	0.07
Year 1996	-0.006	0.006	1.03
Self-employment			
Partially self-emp.	0.016	0.008	1.89
Fully self-emp.	0.028	0.005	5.90
Industries (relative to retail trade)			
Other Services	0.032	0.007	4.30
Agriculture/Forestry	0.027	0.019	1.45
Other/Unknown	0.021	0.021	0.99
Legal Services	0.019	0.014	1.41
Construction	0.018	0.010	1.74
Restaurants	0.014	0.021	0.64
Service Industry	0.013	0.016	0.84
Volunteer/Church	0.009	0.016	0.57
Other Transportation	0.008	0.008	0.94
Mining	0.007	0.016	0.43
Iron/steel	0.006	0.011	0.52
Industries (continued)			
Financial Institutions	-0.001	0.012	0.08
Construction Related	-0.001	0.008	0.15
Food Industry	-0.005	0.011	0.44
Wholesale	-0.005	0.007	0.77
Public Administration	-0.006	0.010	0.57
Postal System	-0.006	0.011	0.59
Mechanical Engineer	-0.009	0.007	1.30
Clothing/Text	-0.014	0.013	1.09
Electrical Engineer	-0.014	0.008	1.67
Health Service	-0.015	0.010	1.47
Synthetics	-0.016	0.016	0.98
Energy/Water	-0.022	0.010	2.26
Education/Sport	-0.024	0.011	2.14
Wood/Paper/Print	-0.025	0.008	3.02
Insurance	-0.026	0.011	2.43
Chemicals	-0.044	0.009	5.11
Earth/Clay/Stone	-0.045	0.012	3.77
Occupations (relative to priv. bus. leaders)			
Other/Unknown	0.149	0.019	7.74
Cook/Waiter	0.108	0.017	6.31
Labor/craftsman	0.042	0.008	5.15
Bricklayer/Carpenter	0.041	0.010	4.22
Related Medical Job	0.041	0.016	2.62
Educator	0.041	0.012	3.49
Conveyor Operator	0.039	0.009	4.32
Painter	0.037	0.017	2.16
Pipe Fitter	0.037	0.011	3.44
Transportation Operator	0.024	0.008	3.12
Janitor	0.022	0.011	1.96
Insurance Represenative	0.018	0.012	1.46
Security Service	0.015	0.012	1.29
Restaurant/Store Manager	0.014	0.016	0.87
Tool/Die Maker	0.013	0.013	1.01
Vendor	0.013	0.007	1.78
Scientist	0.012	0.013	0.93
Food Producer	0.012	0.017	0.74
Doctor/Dentist/Vet	0.009	0.015	0.60
Farm manager	0.007	0.021	0.35
Machine Fitter	0.006	0.007	0.94
Office Worker	0.006	0.009	0.66
Mailman	0.006	0.012	0.48
Engineering Tech. Expert	0.001	0.010	0.12
Tailor	-0.001	0.019	0.06
Soldier	-0.003	0.012	0.24
Inspector	-0.005	0.007	0.67
Electrical Engineer	-0.008	0.008	0.93
Business Manager	-0.016	0.007	2.18
Architect/Engineer	-0.023	0.006	3.75
Mathematician	-0.023	0.008	2.76
Accountant	-0.042	0.010	4.16
Lawyer	-0.043	0.011	3.77
No. obs.	65,560		
R sq.	0.086		

Note: Regression also included dummies for the SEO sample, number of observations for individuals, and an additional 42 occupations and 3 industries with fewer than 400 observations.

Table 7: Regression of Absolute Residual Wages on Characteristics

Variable	Coef.	SE	t
Constant	0.190	0.009	21.40
Education			
Less than HS	0.018	0.004	4.62
Some college	0.005	0.004	1.25
College grad	0.013	0.005	2.57
Graduate studies	0.016	0.006	2.46
Age			
Age 16-20	-0.001	0.006	0.21
Age 21-25	0.008	0.004	2.19
Age 31-35	-0.021	0.003	6.88
Age 36-40	-0.024	0.004	6.43
Age 41-45	-0.029	0.004	6.84
Age 46-50	-0.027	0.005	5.46
Age 51-55	-0.027	0.006	4.81
Age 56-60	-0.021	0.007	3.19
Age 61-65	-0.005	0.009	0.61
Race			
Black	0.028	0.003	8.47
Native American	-0.017	0.012	1.37
Asian	0.002	0.028	0.09
Latino	0.014	0.017	0.80
Other race	0.034	0.021	1.64
Year			
Year 1980	-0.022	0.005	4.65
Year 1981	-0.023	0.005	4.98
Year 1982	-0.019	0.005	3.91
Year 1983	-0.014	0.005	3.08
Year 1984	-0.010	0.005	2.10
Year 1985	-0.008	0.005	1.68
Year 1986	-0.010	0.004	2.18
Year 1987	-0.007	0.004	1.59
Year 1989	-0.007	0.004	1.66
Year 1990	-0.009	0.004	2.09
Year 1991	0.005	0.005	0.95
Year 1992	0.053	0.006	8.85
Year 1993	0.041	0.005	7.71
Year 1994	0.007	0.005	1.34
Year 1995	0.010	0.005	1.91
Year 1996	0.022	0.006	4.04
Self-employment			
Partially self-emp.	0.077	0.011	6.84
Fully self-emp.	0.146	0.006	22.98
Industries (relative to retail trade)			
Mining	0.032	0.015	2.12
Legal Services	0.021	0.012	1.77
Other Services	0.017	0.007	2.45
Construction	0.010	0.010	0.98
Restaurants	0.008	0.016	0.53
Agriculture/Forestry	0.007	0.015	0.47
Service Industry	0.005	0.012	0.39
Other Transportation	0.000	0.007	0.04
Construction Related	-0.001	0.007	0.12
Other/Unknown	-0.003	0.011	0.23
Financial Institutions	-0.003	0.013	0.28
Industries (continued)			
Iron/steel	-0.009	0.008	1.05
Food Industry	-0.011	0.009	1.27
Wholesale	-0.016	0.007	2.41
Electrical Engineer	-0.016	0.008	2.16
Synthetics	-0.019	0.012	1.54
Mechanical Engineer	-0.021	0.006	3.53
Wood/Paper/Print	-0.021	0.007	3.13
Insurance	-0.022	0.012	1.86
Volunteer/Church	-0.022	0.011	1.95
Earth/Clay/Stone	-0.022	0.015	1.47
Postal System	-0.022	0.008	2.69
Clothing/Text	-0.023	0.008	2.70
Energy/Water	-0.028	0.008	3.39
Health Service	-0.030	0.011	2.72
Education/Sport	-0.034	0.010	3.32
Public Administration	-0.035	0.008	4.32
Chemicals	-0.037	0.009	4.16
Occupations (relative to priv. bus. leaders)			
Farm manager	0.093	0.019	4.96
Cook/Waiter	0.066	0.012	5.60
Other/Unknown	0.066	0.011	6.03
Soldier	0.046	0.010	4.47
Doctor/Dentist/Vet	0.025	0.024	1.05
Related Medical Job	0.024	0.013	1.90
Insurance Rep.	0.021	0.013	1.70
Janitor	0.020	0.010	2.11
Vendor	0.018	0.007	2.45
Bricklayer/Carpenter	0.017	0.009	1.94
Conveyor Operator	0.016	0.007	2.29
Transportation Operator	0.014	0.007	2.14
Pipe Fitter	0.012	0.008	1.50
Labor/craftsman	0.011	0.007	1.64
Security Service	0.010	0.010	1.04
Food Producer	0.008	0.014	0.59
Educator	0.006	0.011	0.50
Painter	0.003	0.014	0.21
Office Worker	-0.001	0.009	0.10
Tailor	-0.002	0.015	0.11
Tool/Die Maker	-0.004	0.011	0.35
Scientist	-0.006	0.012	0.49
Machine Fitter	-0.007	0.006	1.19
Restaurant/Store Manager	-0.008	0.017	0.47
Business Manager	-0.008	0.009	0.93
Mailman	-0.010	0.009	1.10
Electrical Engineer	-0.011	0.008	1.36
Inspector	-0.019	0.007	2.84
Engineering Tech. Expert	-0.022	0.009	2.37
Architect/Engineer	-0.031	0.008	4.01
Accountant	-0.031	0.013	2.46
Mathematician	-0.038	0.009	4.26
Lawyer	-0.040	0.016	2.44
No. obs.	64,345		
R sq.	0.074		

Note: Regression also included dummies for the SEO sample, number of observations for individuals, and an additional 42 occupations and 3 industries with fewer than 400 observations.

Table 8: Comparison of Estimated Effects on Absolute Earnings Residuals from Four Regression Specifications

	Pooled			Random Effects			Fixed Effects			Random w/ trends		
Variable	**Est**	**SE**	**P**	**Est**	**SE**	**P**	**Est**	**SE**	**P**	**Est**	**SE**	**P**
Education												
Less than HS	0.041	0.007	0.00	0.039	0.006	0.00	0.025	0.014	0.08	0.043	0.006	0.00
Some college	-0.011	0.006	0.05	-0.027	0.006	0.00	-0.045	0.011	0.00	-0.028	0.006	0.00
College grad	0.010	0.007	0.15	-0.009	0.007	0.18	-0.045	0.017	0.01	-0.013	0.007	0.08
Graduate studies	0.009	0.009	0.31	-0.019	0.010	0.05	-0.076	0.020	0.00	-0.026	0.010	0.01
Age												
Age 16-20	0.000	0.010	0.97	0.017	0.008	0.03	0.037	0.013	0.01	0.011	0.008	0.18
Age 21-25	0.024	0.006	0.00	0.026	0.005	0.00	0.025	0.007	0.00	0.023	0.005	0.00
Age 31-35	-0.028	0.005	0.00	-0.028	0.004	0.00	-0.021	0.006	0.00	-0.025	0.004	0.00
Age 36-40	-0.040	0.006	0.00	-0.035	0.005	0.00	-0.017	0.010	0.08	-0.032	0.005	0.00
Age 41-45	-0.049	0.007	0.00	-0.040	0.006	0.00	-0.011	0.013	0.40	-0.038	0.006	0.00
Age 46-50	-0.048	0.007	0.00	-0.037	0.007	0.00	0.005	0.018	0.80	-0.035	0.007	0.00
Age 51-55	-0.048	0.008	0.00	-0.037	0.008	0.00	0.020	0.022	0.35	-0.031	0.009	0.00
Age 56-60	-0.036	0.010	0.00	-0.008	0.009	0.38	0.070	0.026	0.01	0.005	0.011	0.63
Age 61-65	-0.013	0.013	0.32	0.021	0.012	0.07	0.115	0.031	0.00	0.043	0.015	0.00
Self-employment												
Partially self-emp.	0.057	0.013	0.00	0.022	0.011	0.06	0.010	0.012	0.38	0.024	0.011	0.04
Fully self-emp.	0.152	0.009	0.00	0.119	0.006	0.00	0.099	0.011	0.00	0.115	0.006	0.00
Industries												
Other Services	0.036	0.010	0.00	0.027	0.008	0.00	0.021	0.011	0.05	0.026	0.008	0.00
Legal Services	0.032	0.021	0.13	0.036	0.015	0.01	0.041	0.023	0.07	0.040	0.015	0.01
Volunteer/Church	-0.019	0.017	0.27	-0.049	0.017	0.00	-0.063	0.026	0.02	-0.042	0.017	0.02
Mechanical Engineer	-0.023	0.010	0.02	-0.007	0.009	0.40	0.013	0.012	0.25	-0.007	0.009	0.45
Wood/Paper/Print	-0.028	0.011	0.01	-0.015	0.010	0.12	-0.002	0.013	0.89	-0.016	0.010	0.10
Energy/Water	-0.036	0.014	0.01	-0.020	0.012	0.09	-0.010	0.017	0.56	-0.019	0.012	0.11
Education/Sport	-0.051	0.013	0.00	-0.028	0.013	0.03	-0.021	0.016	0.20	-0.026	0.013	0.05
Insurance	-0.057	0.018	0.00	-0.014	0.019	0.48	0.019	0.025	0.46	-0.012	0.020	0.53
Chemicals	-0.072	0.011	0.00	-0.056	0.015	0.00	-0.043	0.016	0.01	-0.057	0.015	0.00
Occupations												
Other/Unknown Occ.	0.148	0.022	0.00	0.096	0.012	0.00	0.064	0.019	0.00	0.094	0.012	0.00
Farm manager	0.122	0.032	0.00	0.077	0.020	0.00	0.036	0.037	0.34	0.081	0.021	0.00
Cook/Waiter	0.110	0.021	0.00	0.046	0.014	0.00	0.012	0.024	0.63	0.050	0.014	0.00
Insurance Rep.	0.053	0.023	0.02	0.022	0.017	0.21	0.007	0.019	0.72	0.022	0.018	0.21
Pipe Fitter	0.049	0.014	0.00	0.041	0.011	0.00	0.035	0.015	0.02	0.040	0.011	0.00
Bricklayer/Carpenter	0.047	0.013	0.00	0.024	0.010	0.01	0.012	0.013	0.35	0.029	0.010	0.00
Labor/craftsman	0.041	0.011	0.00	0.017	0.008	0.05	0.006	0.011	0.57	0.020	0.008	0.02
Conveyor Operator	0.035	0.012	0.00	0.017	0.009	0.06	0.009	0.012	0.47	0.020	0.009	0.03
Janitor	0.030	0.015	0.04	-0.004	0.012	0.76	-0.021	0.016	0.21	0.005	0.012	0.70
Transp. Operator	0.019	0.010	0.07	0.007	0.008	0.39	-0.001	0.011	0.92	0.010	0.008	0.26
Mailman	0.001	0.016	0.93	-0.019	0.014	0.16	-0.036	0.018	0.05	-0.018	0.014	0.18
Machine Fitter	-0.003	0.010	0.74	-0.020	0.008	0.01	-0.026	0.009	0.01	-0.017	0.008	0.04
Security Service	-0.019	0.017	0.25	-0.025	0.013	0.06	-0.019	0.022	0.39	-0.032	0.013	0.02
Business Manager	-0.025	0.011	0.02	-0.024	0.013	0.06	-0.024	0.013	0.07	-0.022	0.013	0.09
Soldier	-0.033	0.016	0.05	-0.043	0.016	0.01	-0.036	0.021	0.09	-0.046	0.016	0.00
Architect/Engineer	-0.042	0.010	0.00	-0.038	0.011	0.00	-0.030	0.012	0.01	-0.041	0.011	0.00
Mathematician	-0.043	0.015	0.01	-0.046	0.014	0.00	-0.033	0.015	0.02	-0.048	0.014	0.00
Lawyer	-0.049	0.028	0.08	-0.055	0.029	0.06	-0.123	0.045	0.01	-0.062	0.029	0.03
Accountant	-0.051	0.017	0.00	-0.039	0.017	0.02	-0.033	0.014	0.02	-0.044	0.017	0.01

Note: The first specification repeats Table 5 (pooled regression), the second and third add person-specific random and fixed effects (respectively), and the fourth includes both random effects and uncorrelated random growth. Only key parameters are shown; all regressions include the same dependent variables as in Tables 5 to 7. Estimated year and race effects are not shown because estimates do not vary meaningfully across methods. (Race effects cannot be estimated in the fixed effects model because race is a fixed characteristic for each individual.) Industries and occupations are reported only if estimates are statistically significant in at least one specification.

Table 9: Features of Predicted Absolute Residuals and Their Predictive Power

A. Summary Statistics of Estimated Residuals and Predicted Absolute Residuals

Raw Residuals			All Observations with E>0, H>0			Observations with E>0, H>0, Occupation and Industry Known		Including Observations with E=0 or H=0 (Out of Sample Predictions)	
Variable	Mean	SD	Variable	Mean	SD	Mean	SD	Mean	SD
$\lvert\varepsilon_E\rvert$	0.269	0.342	$E[\lvert\varepsilon_E\rvert]$	0.269	0.094	0.254	0.084	0.281	0.101
$\lvert\varepsilon_H\rvert$	0.194	0.273	$E[\lvert\varepsilon_H\rvert]$	0.194	0.081	0.176	0.059	0.205	0.090
$\lvert\varepsilon_W\rvert$	0.217	0.235	$E[\lvert\varepsilon_W\rvert]$	0.217	0.064	0.212	0.064	0.223	0.065
N obs	64,906		N obs	64,915		59,002		77,530	

B. Estimated Slope Coefficients from Probits of Several Dependent Variables on Predicted Absolute Residuals

Dependent Variable (Y):	No work (earnings=0 or hours = 0)		No work (earnings=0 or hours = 0)		Divorce (versus stayed married)		Divorce (versus stayed married)		Started a cohabitation (versus got married)		In jail at time of next interview	
Variable	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
$E[\lvert\varepsilon_E\rvert \mid Z]$	0.058	0.009			0.051	0.017	0.038	0.017	0.645	0.264	0.039	0.015
$E[\lvert\varepsilon_H\rvert \mid Z]$			0.087	0.011								
(actual) ln E					-0.007	0.001			0.004	0.015	-0.008	0.001
E(ln E | Z)							-0.011	0.001				
ε_E							-0.001	0.001				
Mean Pr(Y)	0.013		0.013		0.025		0.025		0.340		0.018	
Δ ln Pr(Y) per std dev Δ $E[\lvert\varepsilon\rvert]$	0.45		0.60		0.17		0.13		0.16		0.19	
Δ ln Pr(Y) per std dev Δ (ln E)					-0.21		-0.34		0.01		-0.36	
pseudo R^2	0.111		0.113		0.055		0.058		0.080		0.185	
N obs	63,404		63,404		45,894		45,890		2,208		55,480	

Note: All regressions in Panel B also include all factors used to predict the absolute residuals, except for workers' occupations and industries, and exclude cases in which workers cannot be identified with an occupation or industry. Standard errors adjust for correlations at the individual level.