Ikuho Kochi (USA, Mexico)

Endogeneity and estimates of the value of a statistical life

Abstract

This study examines the robustness of the recent panel hedonic wage studies that estimate wage-risk tradeoffs, or so called the value of a statistical life (VSL). Recent panel studies found that the VSL estimated from cross-sectional hedonic wage models are biased substantially due to unobserved worker heterogeneity. The VSL estimated from cross-sectional hedonic wage models has been extensively used in evaluating the benefit of reducing premature mortality levels, thus these results have significant policy implications. However, it is well known that VSL estimates from hedonic wage models show substantial variation depending on the sample of workers, type of risk measures, and the empirical specification of the models. Thus previous panel results may be specific to the data they used. This research employs panel data for a sample of workers that has not yet been used in the VSL estimation, but which is widely used in labor economics: the Survey of Income and Program Participation. Similar to past studies, substantial endogeneity bias in cross-sectional models is found. After controlling for unobserved worker heterogeneity, the author finds estimates of the VSL of $1.9 to 2.8 million; a 30 to 60 percent reduction from the VLS estimated through the cross-sectional hedonic model. These estimates are also a third of the VSL estimates from previous work which uses panel models, but within the range of plausible estimates suggested by several meta-analyses of the topic.

Keywords: compensating wage differential, value of statistical life, panel data analysis, individual heterogeneity, inter-industry wage differentials.
JEL Classification: C23, I10, J17, J28.

Introduction

Over the past 40 years, cross-sectional ordinary least squares (OLS) hedonic wage models have generally been employed to empirically measure the wage-premium associated with riskier jobs (see Mrozek and Taylor, 2002; Viscusi and Aldy, 2003; Cropper et al., 2011). Empirical estimates of the wage-risk premium have received substantial attention among economists because of their extensive use in policy analysis. For example, the United States Environmental Protection Agency (US EPA) and Office of Management and Budget (OMB) have used the wage-risk premium as a main reference to determine the value of statistical life (VSL) to evaluate the benefits of reducing mortality from air pollution control policy (OMB, 2003; USEPA, 1999; USEPA, 2005).

Although there have been significant advancements in hedonic wage modeling, important issues remain related to the specification of the empirical model. One of the main criticisms is the potential endogeneity bias in cross-sectional hedonic wage models due to omitted variables related to unobserved worker heterogeneity, such as risk preferences or the worker's skill in protecting themselves in a dangerous work environment (Cropper et al., 2011; Shogren and Stamland, 2002; Viscusi and Aldy, 2003). Standard econometric corrections would be to employ an instrumental variable (IV) approach or panel models if unobservables are time invariant. Earlier work in this area has used the IV approach and generally finds a substantial *downward* bias in the wage-risk premium in simple cross sectional models (Arabsheibani, 2001; Black et al., 2003; Garen, 1988; Gunderson and Hyatt, 2001; Siebert and Wei, 1994). However, the results of these studies are often sensitive to model specifications or sample selections, which raise the concern of the validity of results[1].

Two recent panel studies employing data from the United States (US) and the United Kingdom (UK) find opposite results from previous IV studies. Kniesner et al. (2011) find that unobserved worker heterogeneity substantially biases the wage-risk premium *upward* in cross-sectional hedonic wage models based on the US labor market data. Their VSL estimates based on the cross-sectional model are between $14 and $32.2 million[2]. Once they apply panel models and control for various endogeneity factors, they obtained VSL estimates of between $4 and $11 million, which is about 65% reduction from the VSL reported from their cross-sectional models. Hintermann et al. (2010) also found upward bias in cross-sectional OLS estimator using UK labor market data. However once they control for unobserved worker heterogeneity, they were unable to detect statistically significant wage-risk premium.

The implication of these panel studies is important for the US environmental policy analysis. Currently, US EPA heavily relies on the VSL estimated from cross-sectional hedonic wage modeling to analyze the benefit of reducing premature mortality from air pollution

[1] Arabsheibani and Marin (2001) argued that the instability of their wage-risk premium estimates are likely due to the difficulties associated with finding appropriate instruments and the resulting poor fit of the first stage risk equation.

[2] Based on Kniesner et al. (2011), Table 3, columns (3) and (4). All monetary values are adjusted to a 2005 dollar value using Consumer Price Index.

control policies (US EPA, 2005). These panel studies imply that US EPA may have overestimated the benefit of reduction of premature mortality substantially due to the incorrect hedonic model specification their benefits estimates are based upon.

As Mrozek and Taylor (2002) and Viscusi and Aldy (2003) illustrated, the estimated VSL from hedonic wage models are sensitive to the model specifications, sample selections, and data sources. In particular, cross-sectional hedonic wage models that used Panel Survey of Income Dynamics (PSID), one of the main labor market data in the US, tends to generate VSL estimates on the higher-end (e.g., Moore and Viscusi, 1990a; 1990b; Doman and Hagstrom, 1998). The VSL estimates from cross-sectional models in Kniesner et al. (2011) who also use the PSID are no exception. As such, the VSL based on panel models they estimate may also be unique to the PSID data. It is important to examine for the robustness of results to the different labor market data and provide information about the potential range of unbiased VSL estimates.

In this paper, we estimate the VSL using panel models with a US labor market data provided by the Survey of Income and Program Participation (SIPP) to examine the robustness of previous panel studies. The SIPP is a large scale national level longitudinal data set that is used extensively in labor economics, but to our knowledge, this is the first study to use the SIPP to estimate the VSL.

The SIPP has several preferable characters including rich information about individual income, asset, and labor force status. Hedonic wage models are estimated using 1996 SIPP panel and unobserved worker characteristics are controlled for using panel data models. In addition, instrumental variable (IV) panel methods are applied to more fully examine the potential for endogeneity in the panel models. Lastly, similar to Kniesner et al. (2011) a disaggregated risk measure is used.

Overall, the results indicate that failing to control for unobserved worker heterogeneity substantially influence risk estimators. Failing to account for the worker heterogeneity factors biases risk premium estimates upward as found in previous panel studies. Our panel-model estimates of the VSL are statistically significant, but the point estimates are far lower than Kniesner et al. (2011). Our comparable point VSL estimates from pooled cross-sectional models that do not control for endogeneity factors are between \$4 and \$6.7 million and point VSL estimates from the panel models are between \$1.9 million and \$2.8 million. Our results reinforce previous finding that VSL estimates depend largely on the data source. The unbiased point VSL estimates from Kniesner et al. (2011) (\$4-11 million) should be considered high-end estimates and our estimates (\$1.9-2.8 million) should be considered low-end estimates to be used in the policy analysis.

This paper is organized as follows. The next section discusses the issue of bias associated with omitted worker heterogeneity in hedonic wage modeling. Section 2 presents the data and econometric models and section 3 summarizes the results. Conclusions and policy implications are presented in the final section.

1. Unobserved worker characteristics and risk endogeneity in hedonic wage models

The standard hedonic wage model estimates the following equation by an ordinary least squares (OLS) model:

$$y_i = \beta r_i + X_i \gamma + \mu_i , \qquad (1)$$

where y_i is the wage level, r_i is the measure of fatal occupational risk, X_i is the vector of determinants of wages composed of both worker and job characteristics, and μ_i is the error term for individual i. The parameter for the risk variable, β, represents the additional wage workers require to accept an additional unit of risk, the wage-risk premium. This wage-risk premium is then aggregated over the pool of workers at risk to estimate the value that workers collectively place on reducing the risk that one among them dies, which is equivalent to the VSL. If the risk variable is endogeneous such that the $\text{cov}(r_i, \mu_i | X_i) \neq 0$, then the OLS risk estimator, $\hat{\beta}$, is biased and inconsistent, such as would be the case if the unobservable characteristics of workers, such as risk preferences or the worker's skill in protecting themselves in a dangerous work environment, are correlated with the average risk level associated with a job[1].

Panel data estimation methods such as the fixed effects model and the first-differenced model are able to perfectly control for time-invariant unobserved worker characteristics. We also control time-variant unobserved variables through two-stage least squares (2SLS) regression models. We compare the estimated VSL from the cross-sectional, fixed effects and first-differenced models to identify the potential endogeneity bias in the cross-sectional estimators.

2. Data and estimating models

Data for individual hourly wage, job and socio-economic characteristics come from the Survey of Income and Program Participation (SIPP)[2]. The

[1] See Hwang et al. (1992) for an illustration of the unobserved worker heterogeneity problem in hedonic wage estimation.

[2] Detailed data description and data download is available at http://www.bls.census.gov/sipp/ (last retrieved on March 25, 2007).

SIPP is a national panel data administered by the US Census Bureau. The SIPP contains rich information about individual income, labor force status, and general demographic characteristics of US population. People are interviewed by phone or in person every four months. Each four month reference period is called a *wave*. Only one observation from each wave is used in our analysis[1]. The 1996 SIPP panel is used in this analysis, which lasted for four years and contained twelve waves. The total number of individuals included in 1996 panel is approximately 63,000.

Both the PSID and the 1996 SIPP are longitudinal labor market data set in the US but with different study design. The PSID is a long-term longitudinal dataset that follows the national representative sample in 1968. The 1996 SIPP follows sample for only four years, but has a larger sample size, with shorter intervals between interviews (four months) and more information about income and assets-holding data as compared to the PSID. The characteristics of the 1996 SIPP data are preferable for the objectives of this study. The short interval of interviews would help to ensure that sources of unobserved worker heterogeneity are time invariant. In addition, the 1996 SIPP contains more detailed information about financial assets, the perception about health insurance, and characteristics of employers such as firm size and employment benefit, which is an important determinant of wage but generally ignored in the literature (Bockstael and McConnell, 2006).

Potential wage determinants are: age, educational attainment, gender, race, marital status, number of children under 18 in the household, union status, residential location (urban vs. rural), whether or not the person works over-time, region of the worker, and the occupation and industry group of the firm for which workers work. We also collect the availability of employer provided health insurance and size of firms in the model.

Table 1 reports the definitions of variables extracted from the 1996 SIPP panel and the summary statistics for the sample of workers. The sample is hourly paid full time workers who hold only one job during a wave, are not self-employed, and work for wages[2]. Workers, who are earning less than minimum wage, or whose age is less than 18 or more than 65, are omitted from the analysis[3]. We use hourly wage information provided in the SIPP in our main analysis due to a concern about the accuracy of monthly wage information[4]. We will use monthly wage information in the sensitivity analysis to examine the effect of limiting the sample to hourly paid workers. There are a total of 141,299 observations for 23,860 hourly paid workers. We have an unbalanced panel, and the minimum, average and maximum number of observations per worker is 2, 5, 9 and 12, respectively[5]. The average hourly wage is $13.58 in 2005, which is lower than the average hourly earnings of $16 for the US labor market in 2005[6]. The average age of workers is 38 years old and 42% of the sample graduated from high school, 33% of the sample has attended college, and 8% of the sample has a bachelors or higher degree.

Table 1. Definition of variables and summary statistics

	Definition	1996 SIPP, all hourly paid worker-wave (N = 141,299)	
		Mean	(SD)
Wage	Hourly wage (2005, in dollars)	13.58	(5.89)
Risk	Fatal injury risk rate by occupation and industry per 10,000 workers	0.54	(0.94)
Age	Age in years	38.63	(11.20)
Ugdeg	1 if individual has a bachelor degree or more, 0 otherwise	0.08	
College	1 if individual attended only some college, 0 otherwise	0.33	

[1] We also only keep workers who had a job during entire fourth-month of the reference period. Some questions ask people to record information for every month during reference periods. In this case, only the fourth-month observation is used for the analysis. This is to be consistent with the dependent variable, hourly wage, which is sampled only for the fourth-month observation.

[2] Full time workers are defined as workers who work more than or equal to 35 hours per week. In addition, workers with top-coded wage or income values and workers with only one observation in entire panel are dropped from the analysis.

[3] The minimum wage level for service workers ($2.13 per hour) is used as a cutoff wage level.

[4] There are two main concerns with using the wage information for salaried workers in the SIPP. First, salaried worker only have information of monthly wages, which show more unexplained variation than hourly wage (i.e., large fluctuations of the wage level over time for a worker who shows no sign of job change, hours of work, location of work, or per-hour wage level). This unexplained variation in wages would inflate the variance as well as may bias the results. Second, monthly wage models need to control for the number of hours worked. However, information about hours worked is only available as "usually work hours" and not as "actual work hours". So we may introduce measurement error. In the sensitivity analysis, monthly wages are converted into hourly wages by using "usual work hours".

[5] The panel estimators will be biased if the decision of dropping out from the survey is correlated with the idiosyncratic errors (Wooldridge, 2001). Lamas et al. (1994) report that this attrition bias was not present in wage models using the SIPP 1990 panel. There is no study that examines the attrition bias in the SIPP 1996 panel, but we expect the bias is minimal, if exist, due to the similar sample design between the 1990 and 1996 panels.

[6] October 2005 Employment Situation Summary last retrieved on March 25, 2007 from http://www.bls.gov/news.release/empsit.nr0.htm.

Table 1 (cont.). Definition of variables and summary statistics

	Definition	1996 SIPP, all hourly paid worker-wave (N = 141,299)	
		Mean	(SD)
Hsgrad	1 if individual only graduated from high school, 0 otherwise	0.42	
Hispanic	1 if individual is of Hispanic origin, 0 otherwise	0.13	
Blacknh	1 if individual is black and non-Hispanic, 0 otherwise	0.13	
Othrace	1 if individual is non-white, non-black, and non-Hispanic, 0 otherwise	0.04	
Female	1 if individual is female, 0 otherwise	0.45	
Workov	1 if individual usually works more than 40 hours, 0 otherwise	0.19	
Union	1 if individual is a union member or covered by union, 0 otherwise	0.20	
Married	1 if individual is married, 0 otherwise	0.57	
Kids18	Number of kids under 18 years old	0.79	(1.10)
Hipart	1 if individual is provided part of health insurance by employer, 0 otherwise	0.47	
Hifull	1 if individual is provided full health insurance by employer, 0 otherwise	0.21	
Empall	1 if number of employees at all locations > 100, 0 otherwise	0.69	
Empsize	1 if number of employees at worker's location < 25, 0 otherwise	0.28	
Neast	1 if individual lives in the Northeastern region, 0 otherwise	0.17	
Midwest	1 if individual lives in the Midwestern region, 0 otherwise	0.27	
West	1 if individual lives in the West region, 0 otherwise	0.21	
South	1 if individual lives in the Southern region, 0 otherwise	0.35	
Urban	1 if individual lives in an urban area, 0 otherwise	0.78	
Agind	1 if individual works in the agricultural industry, 0 otherwise	0.02	
Constind	1 if individual works in the construction industry, 0 otherwise	0.07	
Tcuind	1 if individual works in the transportation, communication or utility industries, 0 otherwise	0.07	
Trdind	1 if individual works in the wholesale or retail trades industries, 0 otherwise	0.18	
Servind	1 if individual works in the service industry, 0 otherwise	0.31	
Manufind	1 if individual works in the manufacturing industry, 0 otherwise	0.27	
Pubind	1 if individual works in the public section, 0 otherwise	0.08	
Craftocc	1 if individual has a craftsman occupation, 0 otherwise	0.17	
Profocc	1 if individual has a professional occupation, 0 otherwise	0.11	
Techocc	1 if individual has a technical occupation, 0 otherwise	0.29	
Servocc	1 if individual has a service occupation, 0 otherwise	0.13	
Farmocc	1 if individual has a farming occupation, 0 otherwise	0.04	
Laborocc	1 if individual has a general labor occupation, 0 otherwise	0.26	

Note: Standard deviations (SD) for continuous variables are shown in parenthesis.

Compared to the current national trend of educational attainment in the US labor force, our sample under-represents the labor force with a bachelor's degree, and over-represents the labor force with less than high school diploma[1]. However, once we add monthly paid (or salaried) workers in the sensitivity analysis, the educational attainment level and wage level of sample workers become compatible with national average (see Table 1A in Appendix A).

Within the sample above, there are 11,164 workers (out of 23,860 total workers) who change their occupation or industry at the 3-digit classification level at some point in the four years. This is approximately 46% of all workers in the sample. The frequency of job change ranges from one to six times. About 90% of job-changers changed jobs only one or two times during the four year period.

2.1. Occupational fatal risk data. Occupational risk rates created by Scotton (2000) are used. Scotton (2000) creates 506 risk rates based on a 22 occupation × 23 industry matrix. To avoid measurement error due to yearly fluctuations of death incidences, Scotton computes a six year average risk rate between 1992 and 1997. The risk rate in each occupation-industry cell is calculated by the following formula:

$$r_{oi} = \frac{D_{oi}}{W_{oi}}, \qquad (2)$$

where r_{oi} is the fatal risk rate in occupation o and industry i, D_{oi} is the annual average number of death incidents in occupation o in industry i, and W_{oi} is the

[1] The current national trend is that approximately 30% of the workforce has graduated from high school, 30% have attended a college but have no degree, and 30% hold a bachelor's degree or more (The educational attainment level of labor force over time last retrieved on March 25, 2007 from http://www.bls.gov/cps/labor2005/chart2-1.pdf.)

annual average total number of workers in occupation o in industry i. The numerator in equation (2), D_{oi}, is obtained from the Census of Fatal Occupational Injuries (CFOI) files for the period of 1992-1997. W_{oi} is obtained by computing the average annual employment level in each industry and occupation pair from the Industry-Occupation Employment Matrix 1991-1996 administered by the BLS[1]. Scotton's occupation and industry classification is reproduced in Tables 2A-3A (see Appendix).

The mean fatal risk rate in the SIPP sample is 5.4×10^{-5} with standard deviation of 9.4×10^{-5} and the median risk rate is 2.0×10^{-5}. This is comparable to the mean risk rate of related studies which used the CFOI to create risk rates. Scotton and Taylor (2011) report a mean risk rate of their CPS sample (n = 43,261) of 4.8×10^{-5}, and Kniesner et al. (2011) report a mean 3-year average risk rate for their sample from the PSID (n = 7,931) of approximately 6.2×10^{-5}. About 0.6% of the panel data observations face zero risk (974 observations). There are 10,795 observations in the panel of 141,299 observations where the risk rate changes between waves due to a worker changing jobs[2]. This comprises approximately 8% of total observations.

We use both first-differenced (FD) and fixed effects (FE) models in our analysis to evaluate potential endogeneity bias in panel estimators and sensitivity of results to different panel models[3]. Assume the wages of the i_{th} worker in period t, $y_{i,t}$, are determined as follows:

$$y_{i,t} = \beta r_{i,t} + X_{i,t}\gamma + Z_i\delta + \mu_{i,t}, \quad (3)$$

where y, r and X are defined as in the equation (1), Z_i is a vector of unobserved time-invariant worker characteristics, $\mu_{i,t}$ is an error term, and $t = \{1, 2, ..., T\}$.

A first-differenced model with T periods implies the following estimating equation:

$$\Delta y_{i,t} = \beta\Delta r_{i,t} + \Delta X_{i,t}\gamma + \Delta\mu_{i,t}, \quad (4)$$

where,

$$\Delta y_{i,t} = y_{i,t} - y_{i,t-1},\ \Delta r_{i,t} = r_{i,t} - r_{i.t-1},$$
$$\Delta X_{i,t} = X_{i,t} - X_{i,t-1}, \text{ and } \Delta\mu_{i,t} = \mu_{i,t} - \mu_{i,t-1}$$

And a fixed effects model with T periods implies the following estimating equation:

$$\ddot{y}_{i,t} = \beta\ddot{r}_{i,t} + \ddot{X}_{i,t}\gamma + \ddot{\mu}_{i,t}, \quad (5)$$

where

$$\ddot{y}_{i,t} = y_{i,t} - \bar{y}_i,\ \ddot{r}_{i,t} = r_{i,t} - \bar{r}_i,\ \ddot{X}_{i,t} = X_{i,t} - \bar{X}_i, \text{ and}$$
$$\ddot{\mu}_{i,t} = \mu_{i,t} - \bar{\mu}_i,$$

and where

$$\bar{y}_i = \frac{\sum_{t=1}^{T} y_{i,t}}{T},\ \bar{r}_i = \frac{\sum_{t=1}^{T} r_{i,t}}{T},\ \bar{X}_i = \frac{\sum_{t=1}^{T} X_{i,t}}{T} \text{ and}$$
$$\bar{\mu}_i = \frac{\sum_{t=1}^{T} \mu_{i,t}}{T}.$$

Note that in both equations, the unobserved time-invariant heterogeneity Z is perfectly controlled.

Equations (1), (4) and (5) provide the basis for our empirical comparison. Equation (1) is estimated on the full pooled sample and standard errors are adjusted for correlation among observations of a same worker. Equations (4) and (5) will be applied to the same sample as equation (1) while perfectly controlling for unobserved time-invariant worker heterogeneity. Comparing the risk coefficient estimates between equation (1) and (4) or (5) will provide the magnitude of endogeneity bias due to time-invariant worker heterogeneity in equation (1).

It is known that measurement error bias may be exacerbated in panel models (Griliches and Hausman, 1986). Also there may be other sources of potential endogeneity in the panel models such as unobserved time-variant variables and simultaneity of the wage and risk variables. To control for these factors, two-stage least squares (2SLS) panel models will be employed. Lastly, the sensitivity of risk estimators to changes in sample composition will be examined.

3. Results

Table 2 reports the estimation results with pooled cross-sectional OLS (hereafter, simply referred as OLS), fixed effects (FE) and first-differenced (FD) models. The dependent variable is the log of gross hourly wages in 2005 (in dollars), and independent variables include all variables listed in Table 1. For succinctness, only risk coefficients are reported. Full

[1] Scotton's (2000) risk measure does not reflect the risk of self-employed workers, deaths related to suicides, deaths that occur 30 days after the injury, or deaths of military persons.

[2] As noted earlier, 11,164 workers changed jobs at a 3-digit level industry/occupation group. However, since risk rates are created for broader industry/occupation groups, changing jobs within a 3-digit level industry/occupation pair may not result in a different risk level, leading to a smaller number of observations who experience risk changes.

[3] Ziliak et al. (1999) noted that if there is no endogeneity in panel models and if the FE model is adjusted for non-stationarity, then the FD model and the FE model should have a same probability limit when more than two time period are contained in the data.

model results for models 1 through 6 are presented in Tables 1B-3B (see Appendix B).

Labor economists have long recognized that industry affiliation plays an important role in determining the wage level for reasons not related to occupational risks (see reviews by Leigh, 1995, and Dorman and Hagstrom, 1998). There is a controversy over whether or not industry dummy variables should be included in hedonic wage estimation (Viscusi and Aldy, 2003). Since the risk variables are generally constructed based on worker's industry affiliation, there is a strong correlation between risk variables and industry dummy variables. Including industry dummy variables generally increases the variance of the risk coefficient and often leads to an insignificant risk coefficient. However, omitting industry variables while strong correlation exists between risk and industry variables may bias the risk coefficient. The solution is to create risk variables in a way not highly correlated with industry variables, such as basing them on worker's occupation affiliation, or a combination of industry and occupation affiliations. For this reason, we use 22 occupation × 23 industry risk matrix for our analysis. We estimate both models with and without industry dummy variables to examine the bias associated with omitting industry dummy variables in panel models.

Models 1 through 3 in Table 2 exclude industry dummy variables, while they are included in Models 4 through 8. All models also include dummy variables that indicate the panel wave from which an observation is obtained to capture time trends in wages. Most explanatory variables are statistically significant, and results are generally consistent with findings in previous studies using similar risk measures such as Viscusi (2004). Age and education level are positively correlated with wages. Hispanics and African Americans earn less than whites, and females earn less than males. Workers who belong to a union receive higher wages than non-union workers and so do married workers compared to single workers.

Table 2. Regression analysis results with hourly paid workers

	OLS (clustered)	Fixed effects (FE)	First-differenced (FD)	IV-FE	IV-FD
Excluding industry dummy variables					
	Mode 1	Model 2	Model 3		
Risk coefficient	0.0361***	0.0141***	0.0112***		
(Standard error)	(0.0024)	(0.0015)	(0.0025)		
R^2 (overall)	0.42	0.17	0.21		
VSL (million $)	9.80	3.82	3.04		
Including industry dummy variables					
	Model 4	Model 5	Model 6	Model 7	Model 8
Risk coefficient	0.0150***	0.0086***	0.0071**	0.0104***	0.0085**
(Standard error)	(0.0027)	(0.0017)	(0.0028)	(0.0024)	(0.0039)
R^2 (overall)	0.44	0.20	0.25	0.07	0.01
Anderson LR statistics				P<0.01	P<0.01
Sargan statistics				P=0.67	P=0.91
Endogeneity test				P=0.26	P=0.60
VSL (million $)	4.07	2.33	1.92	2.82	2.30
Average hourly wage	13.58	13.58	13.58	13.58	13.58
N	141,299	141,299	99,611	141,299	99,611

Note: ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level.

The coefficient for the risk variable in Model 1 is positive and statistically significant at the 1% level. The implied VSL from this model is $9.80 million, and is in the range of previous VSL estimates (Kochi et al., 2006; Viscusi, 1992; Viscusi and Aldy, 2003)[1]. When we control for unobserved worker heterogeneity, the coefficient for the risk variable is dramatically reduced, yet still significant at the 1% level. The FE model (Model 2) and FD model (Model 3) in Table 2 indicate VSL estimates that are 61% and 68% less than the cross-sectional estimator, respectively. These results indicate that unobserved time-invariant worker characteristics significantly bias the OLS estimates upward resulting in overestimation of the wage-risk premium. This finding confirms previous panel studies that control for individual heterogeneity through panel models significantly reduces the risk coefficient.

Models 4 to 6 in Table 2 report the OLS, FE and FD models that incorporate seven industry-specific dummy variables. Many industry dummy variables are statistically significant at the 1% level across the models (see Appendix B, Table 2B). The risk variable is

[1] The VSL is estimated as follows: VSL = coefficient of risk variable ×Average hourly wage × 40 (hours) × 50 (weeks) × 10,000, where 10,000 is the unit measure for the fatal risk variable.

also significant at the 1% or 5% level indicating the multicolinearity between industry dummy variables and the risk variable is not an issue here. However, the risk coefficients are reduced by on average 44% as compared to the corresponding models without industry dummy variables indicating that omission of industry controls biases the risk estimator up, as also found in Leigh (1995), Dorman and Hagstrom (1998), and Mrozek and Taylor (2002). In models with industry dummy variables, the VSL estimated from the cross-sectional model is $4.07 million. The panel models generate substantial reduction of the VSL estimates, $1.92 and $2.33 million, indicating the upward endogeneity bias in the cross-sectional hedonic wage models.

To control for remaining potential endogeneity factors in our panel models, such as measurement error, time-variant unobserved variables or simultaneity between wages and the risk variable, we employ a 2SLS panel models. Instrumental variables are expected to influence the choice of risk level but not the wages received. Our instruments include the number of social security recipients in the household (*N_SS*), the monthly income from all financial investments (*inv_all*) and a dummy variable indicating that the employee's reason not having health insurance is because "don't believe in insurance" (*nohi_reason*)[1]. These variables are obtained from the 1996 SIPP.

We also developed an instrument from the risk data itself. This additional variable is the difference between the risk level of individual worker determined by their occupation and industry affiliation, and the average risk of the occupation in which the worker engages across all industries (*dif_rocc*). The variable *dif_rocc* is expected to have a strong correlation with the risk variable, but is only a valid instrument if the worker's deviation from the mean risk level within the same occupation is not correlated with the error term. It is difficult to say with certainty whether *dif_rocc* is a valid instrument. The conditions to be met are quite complicated in this context. The condition for the FD model is that the changes in the deviation from the mean occupational risk (across all industries) must be uncorrelated with changes in the error term from the regression estimating the changes in wage. There is not an intuitive story as to why this condition might hold. However, there is not a clear argument against its validity either. As discussed below, a number of validity test indicated *diff_rocc* is likely a valid instrument.

The second stage 2SLS panel model results relating to the risk variables are presented in Models 7 and 8 in Table 2. More detailed results of the first stage regressions are available in Appendix B, Table 3B. As indicated in Models 7 and 8, the second stage IV-FE model and IV-FD model show slightly higher coefficient estimates for the risk variable as compared to the FE and FD models (VSL of $2.3-$2.8 million). However, the Hausman test for endogeneity (shown in the *Endogeneity Test* row in Table 2) indicates that these coefficients are not significantly different from those in the FD and the FE models. These results suggest that there is no significant endogeneity bias in the FD and FE models resulting from contemporaneous correlation. The Sargan statistics, which evaluates the over-identification restriction, fails to reject the null hypothesis. Failing the null hypothesis of the over-identifying restriction indicates that the current set of instruments is valid, although this may be due to a low power of the test (Wooldridge, 2001). Nevertheless, the coefficient estimates in the IV-FE and IV-FD models are similar to each other, which indicate that the models may be well-specified. There is no significant change among non-risk variables when we estimate the IV-FE and IV-FD models as compared to the FE and FD models.

4. Sensitivity analysis

Overall, the results thus far suggest that the FE and FD models are reasonably well specified and further corrections for endogeneity do not alter our conclusions qualitatively. Next, additional sets of models are presented which explore the sensitivity of our results to two important features of our labor force data.

First we include salaried workers in the sample to examine if our results apply to a more general population. Table 3 replicates Table 2 but with both salaried and hourly paid workers[2]. Although the FD models no longer show significant risk coefficients, the general conclusions do not change when comparing OLS and FE models: omitting unobserved worker characteristics and industry dummy variables substantially biases the risk coefficient upward. The estimated wage-risk premia across models with all workers are very similar to those based on the sample of hourly paid workers.

[1] The wage level should be determined according to the worker's productivity. The incomes that are earned through non-wage sources, the number of social security recipients in household, or their lack of belief in health insurance would not likely affect the worker's productivity. On the other hand, the level of total wealth, number of dependents or belief in health insurance may be related to the worker's risk taking behavior.

[2] Models in Table 3 include a dummy variable that indicates if a worker is hourly paid or not.

Table 3. Regression analysis results with all wage-type workers[a]

	OLS (clustered)	Fixed effects (FE)	First-differenced (FD)
Without industry dummy variables			
	Model 1	Model 2	Model 3
Risk coefficient	0.0333***	0.0112***	0.0063
(Standard error)	(0.0022)	(0.0018)	(0.0042)
R^2 (overall)	0.42	0.18	0.12
VSL (million $)	11.38	3.83	
With industry dummy variables			
	Model 4	Model 5	Model 6
Risk coefficient	0.0198***	0.0073***	0.0037
(Standard error)	(0.0025)	(0.0021)	(0.0046)
R^2 (overall)	0.43	0.20	0.14
VSL (million $)	6.77	2.49	
Average hourly wage	17.10	17.10	17.10
N	260,439	260,439	198,029

Note: [a]IV-FD and IV-FE models are omitted from the table because Sargan statistics indicate that a set of instruments is not valid with this sample. ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level.

Second, we restricted sample to male workers who changed jobs within four-year panel period to control potential remaining confounding factors. Female and male workers may have different preference towards risk, which may confound our results. Also panel estimates of the wage-risk premium are based on workers who changed their job-related risk levels due to job changes over time while the cross-sectional model incorporates all workers including those who did not change jobs or risks.

Table 4. Regression results: male job-changer sample

	Hourly paid workers			All wage-type workers		
	OLS (clustered)	Fixed effects (FE)	First-differenced (FD)	OLS (clustered)	Fixed effects (FE)	First-differenced (FD)
Excluding industry dummy variables						
	Mode 1	Model 2	Model 3	Model 7	Model 8	Model 9
Risk coefficient	0.0309***	0.0158***	0.0125***	0.0235***	0.0106***	0.0056
(Standard error)	(0.0033)	(0.0018)	(0.0029)	(0.0031)	(0.0021)	(0.0045)
R^2 (overall)	0.40	0.09	0.07	0.42	0.13	0.17
VSL (million $)	8.6	4.4	3.4	8.4	3.8	
Including industry dummy variables						
	Model 4	Model 5	Model 6	Model 10	Model 11	Model 12
Risk coefficient	0.0152***	0.0096***	0.0085**	0.0118***	0.0070***	0.0058
(Standard error)	(0.0042)	(0.0022)	(0.0033)	(0.0039)	(0.0025)	(0.0051)
R^2 (overall)	0.42	0.10	0.11	0.42	0.14	0.17
VSL (million $)	4.2	2.6	2.3	4.2	2.5	
Average hourly wage	13.95	13.95	13.95	17.94	17.94	17.94
N	36,670	36,670	24,491	66,828	66,828	48,213

Note: ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level.

The difference between the cross-sectional and panel estimators may be due to the systematic difference between job changers and non-job changers. Thus, excluding female and non-job changer sample will show the sensitivity of our results to such potential confounding factors. Table 4 shows the estimated risk coefficient for male job-changer sample for each OLS, FE and FD model. Overall, there are only marginal changes in the results compare to our base models. The estimated VSL only slightly changed from the base model, and overall conclusions are unchanged: omitting unobserved worker characteristics and industry dummy variables substantially biases the risk coefficient upward. After controlling inter-industry differentials and unobserved worker heterogeneity, the estimated VSL ranges between $2.3 and 2.6 million.

Conclusion

This study aims to provide a robustness analysis to correct for endogeneity bias in cross-sectional hedonic wage models frequently used to estimate the VSL. Using the SIPP data and disaggregated workplace fatal risk data, the estimated VSL after controlling for endogeneity bias is $1.92-2.82 millions, a 30% to 60% reduction from the cross-sectional model[1]. The 2SLS panel models support the results of the FE and FD models and our findings are robust to changes in the sample composition.

Our study support the recent panel studies that found upward endogeneity bias in cross-sectional hedonic wage models. However, our VSL estimates after controlling endogeneity bias are much smaller than the estimates from Kniesner et al. (2011) and larger than Hintermann et al. (2010). The difference between our estimates and Kniesner et al. (2011) is substantial, even though we both use the sample from the US labor market. One of the main differences in sampling design between the PSID and the SIPP is the intervals between interviews. The PSID interviews sample every two years, while the SIPP interviews the sample every four months. Different intervals between each observation may contribute to the different levels of VSL. We re-estimate the models with one observation per year and per two years using the male job changer sample. We did not find any evidence that different time intervals between observations affect the risk-wage estimators.

Although the difference between VSL estimates of our study and Kniesner et al. (2011) is substantial, it is not surprising. Previous cross-sectional hedonic wage models showed substantial variation in estimators depending on the model specification and labor market data source, and it is expected to observe such variation in panel hedonic wage models as well. We should consider estimates of Kniesner et al. (2011) as high-end, and our estimates as low-end VSL estimates. Combining the results from Kniesner et al. (2011) and ours, the plausible range of VSL for the US policy analysis is between $2 and $11 millions.

The impact of inter-industry wage differentials would be heavily influenced by how a researcher creates occupational risk measures. This study suggests that even with a risk variable that varies by both, occupation and industry, the correlation between the risk variable and broad-level industry dummy variables is strong. Inter-industry wage differentials are a well-established phenomenon, and this study further underscores that the hedonic wage model should control for industry characteristics in the model to avoid a potential upward bias in wage-risk estimates.

In future analyses, the robustness of this study's results should be tested using different labor market data such as the Current Population Survey. The estimated VSL from various studies that correct endogeneity bias in hedonic wage models using different labor market sample should be combined to provide plausible range of the VSL estimates. Nevertheless, in the future hedonic wage analysis it is critical to control for unobserved individual heterogeneity and inter-industry wage differentials.

Acknowledgements

This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS. I thank Scott Richardson at the BLS for allowing access to a special research file. I also thank Matthew Gunter and Gregory Fayard at the BLS and Patricia A. Champ for facilitating my access to the data. The author would like to thank Laura O. Taylor for comments on earlier work, and Carol R. Scotton for allowing me to use her data.

References

1. Arabsheibani, G.R. and A. Marin (2001). "Self-selectivity Bias with a Continuous Variable: Potential Pitfall in a Common Procedure", *Applied Economics*, 33 (15), pp. 1903-1910.
2. Black, Dan A., Jose Galdo, and Liquin Lin (2003). "How Robust Are Hedonic Wage Estimates of the Price of Risk?" in Value of Statistical Life Analysis and Environmental Policy: A White Paper for Presentation to Science Advisory Board – Environmental Economics Advisory Committee, United States Environmental Protection Agency (Ed.): Environmental Protection Agency, Office of Policy, Economics and Innovation, National Center for Environmental Economics.
3. Cropper, Maureen L., James K. Hammitt, and Lisa A. Robinson (2011). "Valuing Mortality Risk Reductions: Progress and Challenges", in NBER Working Paper Series, Cambridge: National Bureau of Economic Research.
4. Dorman, Peter and Paul Hagstrom (1998). "Wage Compensation for Dangerous Work Revisited", *Industrial and Labor Relations Review*, 52 (1), pp. 116-135.
5. Garen, John (1988). "Compensating Wage Differentials and the Endogeneity of Job Riskiness", *The Review of Economics and Statistics*, 70 (1), pp. 9-16.
6. Griliches, Zvi and Jerry A. Hausman (1986). "Errors in Variables in Panel Data", *Journal of Econometrics*, 31 (1), pp. 93-118.
7. Gunderson, Morley and Douglas Hyatt (2001). "Workplace Risks and Wages: Canadian Evidence from Alternative Models", *Canadian Journal of Economics*, 34 (2), pp. 377-395.

---

[1] One of the closest VSL estimates to our estimates is found in Mrozek and Taylor (2002). The authors find that the "best practice" VSL estimated from the meta-analysis model that controls for inter-industry wage differentials, but not unobserved worker heterogeneity, is around $2.4 million. US EPA (2005) and OMB (2003) use this estimate as a lower bound VSL estimate for their policy analysis.

8. Hintermann, Beat, Anna Alberini, and Anil Markandya (2010). "Estimating the Value of Safety with Labour Market Data: Are the Results Trustworthy?" *Applied Economics*, 42 (9), pp. 1085-1100.
9. Hwang, Hae-Shin, W. Robert Reed, and Carlton Hubbard (1992). "Compensating Wage Differentials and Unobserved Productivity", *Journal of Political Economy*, 100 (4), pp. 835-858.
10. Kniesner, Thomas J., W. Kip Viscusi, Christopher Woock, and James P. Ziliak (2011). "The Value of a Statistical Life: Evidence from Panel Data", working paper, dated on January 28, 2011.
11. Kochi, Ikuho, Bryan Hubbell, and Randall Kramer (2006). "An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis", *Environmental and Resource Economics*, 34 (3), pp. 385-406.
12. Lamas, Enrique, Jan Tin, and Judith Eargle (1994). "The Effect of Attrition on Income and Poverty Estimates from the Survey of Income and Program Participation (SIPP)", in SIPP Working Paper (SIPP WP190), Washington D.C.: Census Bureau.
13. Leigh, J. Paul (1995). "Compensating Wages, Value of A Statistical Life, and Inter-Industry Differentials", *Journal of Environmental Economics and Management*, 28 (1), pp. 83-97.
14. Nancy E. Bockstael and Kenneth E. McConnell (2006). "Chapter 7: Hedonic Wage Analysis", in Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models (The Economics of Non-Market Goods and Resources), Springer.
15. Moore, Michael J. and W. Kip Viscusi (1990a). "Discounting Environmental Health Risks: New Evidence and Policy Implications", *Journal of Environmental Economics and Management*, 18 (2.2), pp. S51-S62.
16. _______ (1990b). "Models for Estimating Discount Rates for Long-Term Health Risks Using Labor Market Data", *Journal of Risk and Uncertainty*, 3 (4), pp. 381-401.
17. Mrozek, Janusz R. and Laura O. Taylor (2002). "What Determines the Value of Life? A Meta-Analysis", *Journal of Policy Analysis and Management*, 21 (2), pp. 253-270.
18. Office of Management and Budget (OMB) (2003). "*Informing Regulatory Decisions: 2003 Report to Congress on the Costs and Benefits of Federal Regulations and Unfunded Mandates on State, Local, and Tribal Entities*", OMB.
19. Scotton, Carol R. (2000). "The Impact of Workplace-Risk Measurement on Estimates of the Value of a Statistical Life", Georgia State University.
20. Scotton, Carol R. and Laura O. Taylor (2011). "Valuing Risk Reductions: Incorporating Risk Heterogeneity into a Revealed Preference Framework", *Resource and Energy Economics*, 33 (2), pp. 381-397.
21. Shogren, Jason F. and Tommy Stamland (2002). "Skill and the Value of Life", *Journal of Political Economy*, 110 (5), pp. 1168-1173.
22. Siebert, W.S. and X. Wei (1994). "Compensating Wage Differentials for Workplace Accidents: Evidence for Union and Nonunion Workers in the US", *Journal of Risk and Uncertainty*, 9 (1), pp. 61-76.
23. US Environmental Protection Agency (USEPA) (1999). *The Benefits and Costs of the Clean Air Act 1990 to 2010, EPA Report to Congress*, USEPA.
24. ______ (2005). "*Regulatory Impact Analysis for the Final Clean Air Interstate Rule*", USEPA.
25. Viscusi, W. Kip (1992). *Fatal Tradeoffs: Public and Private Responsibilities for Risk*, New York: Oxford University Press.
26. Viscusi, W. Kip and Joseph E. Aldy (2003). "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World", *Journal of Risk and Uncertainty*, 27 (1), pp. 5-76.
27. Viscusi, W. Kip (2004). "The Value of Life: Estimates with Risks by Occupation and Industry", *Economic Inquiry*, 42 (1), pp. 29-48.
28. Wooldridge, Jeffrey M. (2001). *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.
29. Ziliak, James P., Beth A. Wilson, and Joe A. Stone (1999). "Spatial Dynamics and Heterogeneity in the Cyclicality of Real Wages", *The Review of Economics and Statistics*, 81 (2), pp. 227-236.

Appendix A

Table 1A. Summary statistics for hourly and salaried workers

	Definition	1996SIPP, all worker-wave (N = 260,439)	
		Mean	(SD)
Wage	Hourly wage (2005, in dollars)	17.10	(10.24)
Risk	Fatal injury risk rate by occupation and industry per 10,000 workers	0.46	(0.86)
Age	Age in years	39.66	(10.82)
Ugdeg	1 if individual has a bachelor degree or more, 0 otherwise	0.27	
College	1 if individual attended only some college, 0 otherwise	0.30	
Hsgrad	1 if individual only graduated from high school, 0 otherwise	0.31	
Hispanic	1 if individual is of Hispanic origin, 0 otherwise	0.10	
Blacknh	1 if individual is black and non-Hispanic, 0 otherwise	0.11	
Othrace	1 if individual is non-white, non-black, and non-Hispanic, 0 otherwise	0.04	
Female	1 if individual is female, 0 otherwise	0.44	

Table 1A (cont.). Summary statistics for hourly and salaried workers

	Definition	1996SIPP, all worker-wave (N = 260,439)	
		Mean	(SD)
Workov	1 if individual usually works more than 40 hours, 0 otherwise	0.30	
Union	1 if individual is a union member or covered by union, 0 otherwise	0.17	
Married	1 if individual is married, 0 otherwise	0.62	
Kids18	Number of kids under 18 years old	0.78	(1.08)
Salary	1 if individual is not hourly paid worker, 0 otherwise	0.44	
Hipart	1 if individual is provided part of health insurance by employer, 0 otherwise	0.49	
Hifull	1 if individual is provided full health insurance by employer, 0 otherwise	0.23	
Empall	1 if number of employees at all locations > 100, 0 otherwise	0.70	
Empsize	1 if number of employees at worker's location < 25, 0 otherwise	0.28	
Neast	1 if individual lives in the Northeastern region, 0 otherwise	0.18	
Midwest	1 if individual lives in the Midwestern region, 0 otherwise	0.25	
West	1 if individual lives in the West region, 0 otherwise	0.22	
South	1 if individual lives in the Southern region, 0 otherwise	0.35	
Urban	1 if individual lives in an urban area, 0 otherwise	0.81	
Agind	1 if individual works in the agricultural industry, 0 otherwise	0.02	
Constind	1 if individual works in the construction industry, 0 otherwise	0.05	
Tcuind	1 if individual works in the transportation, communication or utility industries, 0 otherwise	0.07	
Trdind	1 if individual works in the wholesale or retail trades industries, 0 otherwise	0.17	
Servind	1 if individual works in the service industry, 0 otherwise	0.38	
Manufind	1 if individual works in the manufacturing industry, 0 otherwise	0.21	
Pubind	1 if individual works in the public section, 0 otherwise	0.10	
Craftocc	1 if individual has a craftsman occupation, 0 otherwise	0.12	
Profocc	1 if individual has a professional occupation, 0 otherwise	0.30	
Techocc	1 if individual has a technical occupation, 0 otherwise	0.29	
Servocc	1 if individual has a service occupation, 0 otherwise	0.09	
Farmocc	1 if individual has a farming occupation, 0 otherwise	0.04	
Laborocc	1 if individual has a general labor occupation, 0 otherwise	0.16	

Note: Standard deviations (SD) for continuous variables are shown in parenthesis.

Table 2A. Occupation group mapping[a]

Occ code	22 occupation groups	Census occupation classification codes
70120	Executive & administrative positions	004-022
70300	Management related occupations	023-037
70400	Engineers	044-059
71290	Professional occupations (except engineers)	043, 063-199
71590	Technicians (includes air craft pilots)	203-235
71900	Marketing and sales occupations	243-285
72300	Secretaries & typists	313-315
72400	Financial records keepers	337-344
72600	Administrative support occupations (except finance & secretaries)	303-309, 316-336, 345-389
73100	Cleaning & building service and maintenance	448-455
73200	Service workers (except cleaning & building service and maintenance)	403-447, 456-469
73350	Mechanics (all types)	505-549
73400	Blue-collar worker supervisors	503, 553-558, 613, 628, 803, 843, 864
73490	Construction tradesmen	563-599
73510	Extractive occupations	614-617
73540	Precision workers	634-699, 796-799
73630	Machine operators	703-779
73700	Fabricators & hand workers	783-795
73820	Truck drivers	804
73900	Motor vehicle & material moving equip operators	806-834, 844-859
74000	General laborers	865-889
74390	Farming, forestry & fishing occupations	473-499

Source: [a]Scotton (2000, p. 200) with some corrections.

Table 3A. Industry group mapping[a]

23 industry groups	23 inds code	Industry (2-digit SIC code)	SIC	SIPP code
Agriculture, forestry, and fishing	9010	Agricultural production crops	01	010-032
	9010	Agricultural production livestock and animal specialties	02	
	9010	Agricultural services	07	
	9010	Forestry	08	
	9010	Fishing, hunting, and trapping	09	
Mining, extraction and quarrying	9020	Metal mining	10	040-050
	9020	Coal mining	12	
	9020	Oil and gas extraction	13	
	9020	Mining and quarrying of nonmetallic minerals, except fuels	14	
Construction	9030	Building construction general contractors and operative builders	15	060
	9030	Heavy construction other than building construction contractors	16	
	9030	Construction special trade contractors	17	
Food and tobacco products	9420	Food and kindred products	20	100-130
	9420	Tobacco products	21	
Textile mill and apparel products	9423	Textile mill products	22	132-152
	9423	Apparel and other finished products from fabrics & similar materials	23	
Lumber/wood/stone/glass products	9432	Lumber and wood products, except furniture	24	230-262
	9432	Furniture and fixtures	25	
	9432	Stone, clay, glass and concrete products	32	
Paper and printing products	9427	Paper and allied products	26	160-172
	9427	Printing, publishing, and allied industries	27	
Chemicals/petro/plastics/leather goods	9431	Chemicals and allied products	28	180-222
	9431	Petroleum refining and related industries	29	
	9431	Rubber and miscellaneous plastics products	30	
	9431	Leather and leather products	31	
Metals, machinery, and misc. Manufacturing industries	9435	Primary metal industries	33	270-350
	9435	Fabricated metal products, except machinery & transportation equipment	34	
	9435	Industrial and commercial machinery and computer equipment	35	
	9435	Electronic & other electrical equipment, components, except computer equipment	36	
	9435	Measuring, analyzing, and controlling instruments; photographic, medical and optical goods; watches and clocks	38	
	9435	Miscellaneous manufacturing industries	39	
Motor vehicle and equipment manufacturing	9437	Transportation equipment	39	351-370
Railroad and water transportation	9500	Railroad transportation	40	400, 420
	9500	Water transportation	44	
Personal transportation services (ground)	9541	Local/suburban transit & interurban highway passenger	41	401, 402, 432
	9541	Transportation services	47	
Trucking, warehousing and air transportation	9545	Motor freight transportation and warehousing	42	410-411, 421
	9545	Transportation by air	45	
Communications, utilities and sanitary services	9549	Communications	48	422, 440-442, 450-472
	9549	Electric, gas, and sanitary services	49	
	9549	Pipelines, except natural gas	46	
Wholesale trade	9651	Wholesale trade-durable goods	50	500-574
	9651	Wholesale trade-non-durable goods	51	
Retail trade	9652	Building materials, hardware, garden supply and mobile home dealers	52	580-694
	9652	General merchandise stores	53	
	9652	Food stores	54	
	9652	Automotive dealers and gasoline service stations	55	
	9652	Apparel and accessory stores	56	
	9652	Eating and drinking places	58	
	9652	Miscellaneous retail (liquor and drug stores)	59	

Table 3A (cont.). Industry group mapping[a]

23 industry groups	23 inds code	Industry (2-digit SIC code)	SIC	SIPP code
Finance, insurance and real estate	9760	Depository institutions	60	700-714
	9760	Non-depository credit institutions	61	
	9760	Insurance carriers	63	
	9760	Insurance agents, brokers and service	64	
	9760	Real estate	65	
	9760	Holding and other investment offices	67	
Personal services	9872	Personal services	72	761, 771-795
	9872	Private households	88	
Business, auto and repair services	9876	Business services	73	721-760, 801, 882-893
	9876	Automotive repair, services and parking	75	
	9876	Miscellaneous repair services	76	
	9876	Engineering, accounting, research, management and related services	87	
Entertainment services	9879	Motion pictures	78	800, 802, 810
	9879	Amusement and recreation services	79	
Health services	9880	Health services	80	812-840
Social, legal, educational and other services	9885	Hotels, rooming houses, camps, and other lodging places	70	762-770, 841-881
	9885	Legal services	81	
	9885	Educational services	82	
	9885	Social services	83	
	9885	Museums, art galleries, and botanical and zoological gardens	84	
	9885	Membership organizations	86	
Public administration & USPS	9990	United states postal service	43	412, 900-932
		All other public administration	91-99	

Source: [a]Scotton (2000, pp. 194-198) with some modifications.

Appendix B. Full model results with hourly paid worker samples

Table 1B. Cross-section, fixed effects and first-differenced regressions models without industry dummy variables[a]

	Model 1		Model 2		Model 3	
	OLS (clustered)	Standard error	Fixed effects	Standard error	First-differenced	Standard error
Risk	0.0361***	0.0024	0.0141***	0.0015	0.0112***	0.0025
Age	0.0313***	0.0012	0.0568***	0.0021	0.0157***	0.0029
Age2	-0.0003***	0.00001	-0.0007***	0.00001	-0.0001***	0.00003
Ugdeg	0.2240***	0.0102	0.0873***	0.0211	0.0526*	0.0303
College	0.1505***	0.0066	0.0070	0.0162	0.0166	0.0226
Hsgrad	0.0839***	0.0061	-0.0026	0.0142	0.0155	0.0180
Hispanic	-0.0977***	0.0067				
Blacknh	-0.0523***	0.0062				
Othrace	-0.0811***	0.0115				
Female	-0.1254***	0.0049				
Workov	0.0562***	0.0041	0.0143***	0.0016	0.0060***	0.0012
Union	0.2212***	0.0052	0.0466***	0.0024	0.0119***	0.0019
Kids18	0.0072***	0.0019	0.0013	0.0014	0.0012	0.0018
Married	0.0787***	0.0044	0.0168***	0.0032	0.0052	0.0040
Hipart	0.1447***	0.0041	0.0334***	0.0017	0.0098***	0.0015
Hifull	0.1529***	0.0048	0.0367***	0.0020	0.0095***	0.0016
Empall	0.0435***	0.0042	0.0147***	0.0017	0.0046***	0.0016
Empsize	-0.0512***	0.0043	-0.0110***	0.0017	-0.0010	0.0016
Neast	-0.0227***	0.0072	-0.0050	0.0237	0.0315	0.0444
Midwest	-0.0461***	0.0063	-0.0299	0.0186	0.0450	0.0362
South	-0.0953***	0.0061	-0.0565***	0.0175	-0.0312	0.0370
Urban	0.0719***	0.0048	0.0117***	0.0039	0.0072	0.0052
Craftocc	0.3130***	0.0155	0.0661***	0.0094	0.0255	0.0157
Profocc	0.3617***	0.0167	0.0691***	0.0098	0.0328**	0.0159

Table 1B (cont.). Cross-section, fixed effects and first-differenced regressions models without industry dummy variables[a]

	Model 1		Model 2		Model 3	
	OLS (clustered)	Standard error	Fixed effects	Standard error	First-differenced	Standard error
Techocc	0.2059***	0.0156	0.0382***	0.0094	0.0082	0.0156
Servocc	0.0277**	0.0160	-0.0408***	0.0096	-0.0469***	0.0159
laborocc	0.1350***	0.0151	0.0266***	0.0091	-0.0062	0.0152
Constant	1.3343***	0.0268	1.3595***	0.0669	0.0110	0.0005
N (# group)	141,299	(23,860)	141,299	(23,860)	99,611	(23,860)
R^2 (overall)	0.42		0.17		0.21	
VSL (million $)	9.80		3.82		3.04	

Note: ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level. [a]Wave variables are omitted for succinctness.

Table 2B. Cross-section, fixed effects and first-differenced regression models with industry dummy variables[a]

	Model 4		Model 5		Model 6	
	OLS (clustered)	Standard error	Fixed Effects	Standard error	First-differenced	Standard error
Risk	0.0150***	0.0027	0.0086***	0.0017	0.0071**	0.0028
Age	0.0297***	0.0011	0.0559***	0.0021	0.0156***	0.0029
Age2	-0.0003***	0.00001	-0.0007***	0.00001	-0.0001***	0.00003
Ugdeg	0.2100***	0.0100	0.0862***	0.0210	0.0523*	0.0303
College	0.1431***	0.0065	0.0071	0.0162	0.0168	0.0226
Hsgrad	0.0824***	0.0060	-0.0025	0.0141	0.0156	0.0179
Hispanic	-0.0994***	0.0065				
Blacknh	-0.0591***	0.0062				
Othrace	-0.0821***	0.0114				
Female	-0.1273***	0.0048				
Workov	0.0640***	0.0040	0.0143***	0.0016	0.0061***	0.0012
Union	0.1968***	0.0052	0.0446***	0.0024	0.0117***	0.0019
Kids18	0.0069***	0.0019	0.0014	0.0014	0.0012	0.0018
Married	0.0737***	0.0043	0.0170***	0.0032	0.0054	0.0040
Hipart	0.1361***	0.0040	0.0318***	0.0017	0.0097***	0.0015
Hifull	0.1436***	0.0047	0.0348***	0.0020	0.0095***	0.0016
Empall	0.0507***	0.0042	0.0152***	0.0017	0.0044***	0.0016
Empsize	-0.0405***	0.0041	-0.0090***	0.0017	-0.0009	0.0016
Neast	-0.0217***	0.0070	0.0007	0.0236	0.0312	0.0443
Midwest	-0.0460***	0.0061	-0.0300	0.0186	0.0439	0.0362
South	-0.0987***	0.0059	-0.0499***	0.0175	-0.0296	0.0370
Urban	0.0721***	0.0047	0.0121***	0.0039	0.0072	0.0052
Agind	-0.1245***	0.0189	-0.0519***	0.0127	-0.0386	0.0265
Constind	0.0105	0.0125	0.0060	0.0095	-0.0603***	0.0197
Tcuind	-0.0352***	0.0116	-0.0241**	0.0095	-0.0688***	0.0196
Trdind	-0.2183***	0.0097	-0.0938***	0.0082	-0.1305***	0.0171
Servind	-0.0966***	0.0093	-0.0564***	0.0078	-0.1051***	0.0164
Manufind	-0.0980***	0.0099	-0.0173**	0.0083	-0.0753***	0.0171
Craftocc	0.2616***	0.0183	0.0481***	0.0104	0.0305*	0.0168
Profocc	0.3155***	0.0190	0.0601***	0.0107	0.0391**	0.0169
Techocc	0.1823***	0.0182	0.0328***	0.0104	0.0166	0.0166
Servocc	0.0019	0.0186	-0.0434***	0.0106	-0.0354**	0.0170
laborocc	0.1147***	0.0181	0.0130	0.0102	0.0028	0.0164
Constant	1.5246***	0.0298	1.4280***	0.0673	0.0110***	0.0005
N (# group)	141,299	(23,860)	141,299	(23,860)	99,611	(23,860)
R^2 (overall)	0.44		0.20		0.25	
VSL (million $)	4.07		2.33		1.92	

Note: ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level. [a]Wave variables are omitted for succinctness.

Table 3B. Selected variables from the first stage regression results for Models 7 and 8 in Table 2.

	First-stage results for Model 7	Standard error	First-stage results for Model 8	Standard error
Dif_rocc	0.6426***	0.0017	0.6571***	0.0019
Inv_all	0.00001	0.00001	5.47×10^{-6}	7.25×10^{-6}
N_SS	0.0026	0.0024	-0.0010	0.0016
Nohi_reason	0.0316*	0.0171	0.0070	0.0083
R^2 (overall)	0.77		0.67	

Note: ***significant at the 1% level, **significant at the 5% level, *significant at the 10% level.