ABSTRACT

PUCKETT, XINGYI SU. A Simulation Approach for Evaluating Hedonic Wage Models' Ability of Recover Marginal Value of Risk Reduction. (Under the direction of Dr. Laura Taylor).

This research proposes a simulation-based approach to evaluate the impact of hedonic wage model specifications and researchers' econometric strategies on the accuracy of the VSL estimates. The analysis proceeds in three stages. In the first stage, a work-job sorting algorithm is developed to simulate a hedonic wage equilibrium which is in consistent with compensating wage differential theory. The simulated wage equilibrium is used to compute worker's "true" marginal-willingness-to-pay. In the second stage, the simulated data is used to estimate large set of reduced form hedonic wage models that vary by the modeling choices, including: i) which functional form is utilized in the model, ii) the inclusion of occupation and/or industry fixed effects, and iii) the cross-section or panel data estimators applied to the model. The hedonic wage models are also estimated in various omitted variable scenarios. The marginal value of risk estimated from the reduced form hedonic wage model are compared to the "true" marginal-willingness-to-pay to recognize that: i) how the different modeling choices influence the accuracy of the empirical estimates of marginal-willingness-to-pay for workplace fatal risk and ii) which combination of model specification and econometric strategies is most preferable. Finally, a set of model specifications suggested in the evaluation are used to estimate the VSLs to compare to the literature.

A Simulation Approach for Evaluating Hedonic Wage Models' Ability to Recover Marginal Value for Risk Reduction

by
Xingyi Su Puckett

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2013

APPROVED BY:

_____	_____
Laura Taylor Chair of Advisory Committee	Ivan Kandilov
_____	_____
Nicolai Kuminoff	Dan Phaneuf
_____	_____
Steven Sexton	Roger von Haefen

DEDICATION

To James and Noah

BIOGRAPHY

Xingyi S. Puckett was born in Shanghai, China on January 4th, 1984. After graduating from Shanghai Xiangming High School in 2002, she attended the Shanghai Shanda Univerisity. Xingyi received a Bachelor of Science degree in International Economics and Trade in 2006. The following year she attended graduate program in North Carolina State University. Xingyi received a Master of Science degree in Economics in 2008. After graduation, she pursued a doctorate in Economics at North Carolina State University.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest thanks to my advisor, Dr. Laura Taylor, for her tireless efforts in guiding me through the dissertation. I have been extremely fortunate to have had the opportunity to work with her. Her encouragement and advising style both on and off the research subject made the last five years truly memorable. I owe my special thanks to my committee member Dr. Nicolai Kuminoff, as his guidance and support make this thesis possible. His knowledge and insight continuously provided me with invaluable advice for the work in this thesis. I also thank my other committee members: Dr. Roger von Haefen, Dr. Dan Phaneuf, Dr. Ivan Kandilov and Dr. Steven Sexton. Through their research, teaching, and comments on my work, each committee member improved the quality of this dissertation and enhanced my appreciation of economics.

While at NC State, I also benefited from numerous discussions with Xiangping Liu, Nam Anh Tran, Jonathan Lee, Timothy Hamilton, Xiaomei Chen (Barbara), Tengying Weng.

I am also grateful to all my friends for their help throughout my ups and downs in the past five years. Finally and most importantly, I want to extend my thanks to my mother for her unconditional support, and to James, my husband, for everything he did for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Cost-benefit analyses of government programs affecting premature deaths frequently require placing a monetary value on mortality risk reductions. In regulatory analysis, the value of reduced mortality risk usually takes the form of a "value per statistical life" (VSL). The VSL is calculated by summing the average willingness-to-pay (WTP) for a small reduction in mortality risk over the population upon which the average is computed and who are affected by the risk reduction (Cameron, 2010).[1]To see how the VSL is computed, consider the following example. Suppose it is estimated that each member of a population of a hundred thousand is willing to pay $50 on average for 1/100,000 or a "micro-risk" decrease in his risk of dying during the next year, the corresponding VSL would be $ 50 × 100,000 or $5 million. Thus, the VSL provides policy makers with an *ex ante* estimate of the monetary benefit of a policy that reduces mortality risk.

The VSL estimate has been used as a primary input for cost-benefit analysis of U.S. Environmental Protection Agency (USEPA) regulations that have substantial economic impact. According to US Office of Management and Budget (OMB)'s *Circular A-4, Regulatory Analysis* (2003), EPA air pollution rules account for approximately 90 percent of total monetized benefits of significant federal health and safety regulations reviewed by OMB in 2004. Importantly, the monetized benefits for EPA's air pollution regulations are

[1] This VSL is not an identified life, but rather an aggregated marginal willingness to pay reductions. While it has commonly been referred as willingness to pay to save one anonymous life, this has led to unnecessary confusion and controversy in the media and policy dormain (Cameron, 2010; Scotton and Taylor, 2009).

attributable primarily to reductions in premature mortality and thus hinge critically on the VSL (OMB 2005).

In an attempt to find an objectively determined point estimate or range of values for the VSL that can be shown to be valid and reliable, a large empirical literature has developed evaluating individual behavior as related to risk-dollar tradeoffs (see McConnell, 2006; Kochi et al. 2006; Viscusi and Aldy, 2003; and Mrozek and Taylor, 2002 for recent reviews of the literature). Although several approaches exist to estimate the VSL, economists most frequently rely on hedonic wage models to examine equilibrium choices that workers make between wages and workplace risks. The hedonic wage method is based on the idea of compensating wages for risky working environments, originally formalized by Jones-Lee (1974) and Thaler and Rosen (1976) and recognized as far back as Adam Smith's *Wealth of Nations (1776).* The wage compensation concept suggests that if some jobs take place in an environment where a fatal injury is more likely to occur, we would expect to see a compensating wage differential to face this increased fatality risk. In other words, jobs that are more risky could be expected to offer higher wages to attract enough people to these jobs, and an estimate of the compensating wages required to attract a worker to a riskier job provide the basis for capturing the VSL as described above.

Empirically, in the hedonic wage model, wages are regressed on the workers' occupational risk levels, as well as other observed worker and job characteristics that are thought to influence wages. Typically, cross-section ordinary least squares (OLS) models of the following form are estimated:

$$\ln(wage_i)=\alpha+\beta * r_j+X_i\gamma+Z_j\theta+\varepsilon_i, \quad (1.1)$$

where log wage of the i^{th} worker is regressed on a measure of the risk of fatal injury r_j associated with the job j chosen by worker i, a vector of job attributes $\mathbf{Z_j}$, and a vector of workers' characteristics $\mathbf{X_i}$.[2] The estimated coefficient β represents the additional wage (in percentage terms) that workers require to accept an additional unit of risk i.e. the wage-risk premium. To compute the VSL, this monetary value for a marginal change in risk is then aggregated over the pool of workers at risk.

Hedonic wage studies are the foundation of VSL estimates that are used in the regulatory analysis. Currently, the USEPA's draft *Guidelines for Preparing Economic Analysis (2010)* (hereafter, the draft *Guidelines*) suggests a central VSL estimate of $7.9 million (in 2008 dollars), based on an earlier guideline document (USEPA, 2000). This value is derived from 26 VSL studies whose point estimates of the VSL range from $0.9 million to $20.6 million (in 2002 dollars). Within these 26 studies, 21 use hedonic wage models to estimate the VSL.

Notwithstanding the importance of the VSL in policy analysis, large variation in the underlying estimates upon which the EPA's central VSL estimate is based raises the concern about the validity of these estimates as a basis for policy analysis. Several meta-analyses of the VSL literature have explored the reasons for this variation and sought to provide "best estimates" for policy applications (Mrozek and Taylor, 2002; Viscusi and Aldy, 2003; Kochi et al,2006;). However, each of these studies also reports a wide range of the "best estimates"

[2] Both γ and θ are the vectors of coefficients for $\mathbf{X_i}$ and $\mathbf{Z_j}$.

for the VSL, ranging from about $2.0 million to $3.3 million in Mrozek and Taylor (2002) to $11.1 in Kochi et al. (2006) (in 2008 dollar). Black and Kniesner (2003), and Black et al. (2003) took a different approach and systematically examined the robustness of hedonic wage estimates using different risk measures and labor force data sets, aiming to identify the source of instability of VSL estimates. They concluded that other than the different methodologies and different samples which cause expected variations in the estimates, two major issues significantly contribute to the fragile estimates of the hedonic wage literature: poorly measured workplace risk rates and omitted variables.

Measurement error in workplace risk is mainly caused by limitations in the risk data available to researchers. For example, all past hedonic wage studies examined in meta-analyses use a risk measure based on a count of the number of fatal accidents for either a specific occupation or a specific industry across the US. This induces a correlation between the risk measure and wage differentials associated with different jobs or industries that are not due to risk differences. [3]

Over the past few years, significant advances have been made in the measurement of job risk, and much more pertinent measures of the risk are available. As described in Scotton (2007), the Census of Fatal Occupation Injuries (CFOI) provides detailed information on workplace deaths. These risk measures are associated with particular jobs than more broadly based indexes which enable the analyst to create risk rates that vary by occupation within

[3] The reasons for inter-industry wage differentials have been explored in Krueger and Summers (1988), Blanchflower et al (1990), Beckerman and Jenkinson (1990), Margolis and Salvanes (2001), Martins (2009) and Guertzgen (2010) among others. None of which suggests risk differentials are the source.

industry. CFOI data has been applied in recent cross-section studies (cite) and in panel models (Viscusi et al. 2007; Kniesnerl et. al, 2011 and Kochi, 2011). Despite the access to more refined data, variability across studies persists with current estimates varying from $2 million to $15 million (in 2008 dollar).

Omitted variable bias is the second underlying issue thought to be driving instability of VSL estimates, and is a particularly troublesome issue for hedonic wage models. Key worker characteristics such as heterogeneous risk preferences are not contained in commonly used labor force data. While unobserved workers characteristics can be addressed using panel data if these characteristics are time-invariant, more troublesome is unobserved job characteristics that may be correlated with both job risks and wages and, therefore, bias the coefficient estimates on risk in hedonic wage applications in an unknown direction (Hwang, 1998; Shogren and Stamland, 2002; DeLiere, Kohn and Timmins, 2010).

Specifically, bias in estimating the risk premium could be caused by the endogeneity of non-pecuniary characteristics if they are correlated with risk. For instance, one may find a positive correlation between fatal risk and physical exertion required for a job or risk and environmental factors such as noise, heat, or odor. In this case, an upward bias in estimating the risk premium could be caused by the endogeneity of these non-pecuniary characteristics. On the other hand, it is also possible that there is a negative correlation between fatal risk and job characteristics required for a job such as office work or documenting skills, in which case, the estimates may be downwardly biased. Ultimately, it is an empirical question and likely related to the jobs populated in any particular labor force data set.

Throughout the literature, most of the efforts are made to address these two underlying problems by varying the samples of workers, employing better measured job risk data, and adopting different methodologies (see Cropper et al, 2011, for detailed summary of recent literature). However, very few studies went beyond these two considerations and attempted to understand how researchers' econometric strategies might also influence VSL estimates. The importance of exploring this point is underscored by Mrozek and Taylor (2002) who found that 78% of 203 VSL estimates across 33 studies used a log-linear specification for their wage equations and a OLS estimator.

Black et al. (2003) systematically examine the robustness of the VSL estimate to the hedonic wage model specifications. They estimated hedonic wage equations using different combinations of worker and risk data sets. They also estimated different hedonic wage models with specifications that vary by the inclusion of dummy variables for the geographic location of the worker and for the occupation and/or industry fixed effects. In order to examine the influence of choices about the hedonic wage specification, they vary the models by using linear or logarithmic forms of the risk variable while keeping the wage rate in logarithmic form. Moreover, they estimate the wage-risk locus nonparametrically by employing the propensity score method. They found that the estimated coefficient on fatal risk varied widely. Changes in the inclusion of occupation and/or industry fixed effects or changes in the job risk specification can result in very large changes in the estimated price of risk. Their results also indicate that estimates from more flexible functional forms present similar sensitivity to the choice of risk data and other model specifications such as

occupation and/or industry fixed effects. These results lead to their conclusion that the instability of the VSL estimates is not the result of equation misspecification, but rather non-classical measurement error in the risk.

Black et al. (2003)'s was an important assessment of VSL literature prior to 2000, but it is based on the early fatal risk data which are severely flawed.[4] These flawed data make it difficult to decompose and identify the causes of the instability of the VSL. As noted in Black et al. (2003), it is unclear if the parametric representation of the wage equation is correct when the other problems with the estimation that appear in both parametric and nonparametric estimates. Therefore, it is still unknown whetehr the commonly used hedonic wage functional from is appropriate and that how the choice of functional forms affects the estimates of MWTP for risk.

In this research, I will take advantage of improved fatal risk data and enriched occupational attributes data to, revisiting the problem of hedonic wage modeling choice by addressing three questions: 1) whether the specifications commonly employed in past studies are appropriate, 2) whether using more flexible functional forms together with more advanced data sets and econometric methods can provide more accurate wage-risk estimates or narrow down the variance of the estimates, and 3) provide guidance on how researchers can systematically examine the accuracy of VSL estimates from more flexible hedonic wage functions. Furthermore, my work evaluates the approaches and data used in current literature (post-2000) in systematic way that has not yet been done.

[4] See Drudi (1995) for a comprehensive summary of the history of occupational risk data, and Scotton (2007) for discussion of measurement error of early risk data.

My work draws upon a simulation approach developed in the two hedonic property value literature by Cropper, Deck and McConnelll (1988) (henceforth CDM) and Kuminoff et al. (2010) (henceforth KPP) to evaluate of the relative performances of different empirical techniques in recovering the marginal value for an environmental amenity improvement. CDM has been an influential study that many other studies rely on to justify their empirical specifications. CDM conduct a simulation study of a housing market that aims to examine how the hedonic functional form influences the error associated with measuring marginal prices. In particular, they use Wheaton's (1974) linear programming algorithm to solve for an equilibrium vector of housing prices under the assumptions about the parametric form of utility, the distribution of preferences, and the supply of housing. The linear programming algorithm allows them to parameterize the utility function through which the "true" marginal-willingness-to-pay (MWTP) for housing characteristics can be derived. Then, they compare the "true" MWTP with empirical estimates from various hedonic price functions to test which functional form provides housing characteristics marginal prices that are closest to the 'true" MWTP. They found that the simpler linear specifications outperformed the more flexible functional forms in the face of omitted variables, while the more flexible specifications perform better when all the variables are observed.

More recently, KPP reviewed 61 residential property value studies and recognized that while linearity still remain as a consistent assumption in most of the empirical studies, methods of estimating hedonic pricing functions have been evolving in terms of the data quality and the advances in micro-econometrics techniques. They extended the approach in

CDM by applying newer econometrics techniques such as including spatial fixed effects, spatial error corrections and panel data methods in a controlled simulation to test the accuracy of common specifications and econometric approaches for hedonic housing price models to recover MWTP for house attributes. Their key finding is the flexible price function, such as the quadratic Box-Cox outperform the simpler linear, log-linear, and log-log specifications when spatial fixed effects are used to control for omitted variables. In panel data settings with time-varying omitted variables, KPP find the difference-in-difference estimator is the best suited to hedonic estimation in panel data. Further, KPP provide evidence that the neighborhood amenities have implicit prices. They find that time-varying implicit prices can be controlled by using a generalized difference-in-differences estimator that includes interaction between a time dummy and all of the characteristics of the hedonic price function.

Given the circumstances in the labor market is parallel with the housing market, I follow their simulation logic of KPP and CDM and extend it into hedonic wage modeling. My goal is to develop a simulated data base that allows me to test the accuracy of commonly used hedonic wage models in recovering the MWTP for reducing the risk of death at the workplace. From a simulated labor market equilibrium, two data bases can be constructed. The first is an equilibrium wage surface arising from each worker being matched to the wage/job bundle that maximizes his utility. Given a set of equilibrium wages for a vector of job and the characteristics of workers matched to these jobs, hedonic wage models may be estimated to recover the marginal implicit price of risk. These marginal prices may then be

compared to observations on the "true" marginal MWTP to reduce risk as computed from utility parameters for each worker that are jointly determined in the labor market simulation.

Following CDM and KPP, the aggregated difference (errors) between the "true" marginal-willingness-to-pay (MWTP) and the estimated wage-risk premium from the hedonic models can be used to suggest the direction and the magnitude of the bias from any particular empirical model. The observations on the direction and magnitude of the estimation biases are then employed as a tool to evaluate the accuracy of various hedonic wage models with different econometric specifications in recovering MWTP.

Following the logic of the simulation approach described above, this research proceeds in two stages. First, I compute the aggregated difference between the "true" MWTP and the estimated wage-risk premium for each hedonic wage model. To do so, I employ the salary-adjustment process developed in Crawford and Knoer (1988) and develop a structural model to construct a simulation framework that is consistent with compensating wage differential theory. The structural model I develop directly considers workers' heterogeneous preferences over the job characteristics including workplace fatal risk. In the structural model, workers' utility over wages and job attributes are defined and workers are matched to jobs through a sorting process. The model allows recovery of preference parameters over job attributes, including the risk of injury or death on the job. The simulation model is calibrated using observed labor market outcomes (i.e. observed wages). The calibration method allows me to capture workers' heterogeneous preferences, and in turn

each worker's utility function can be parameterized. Given the utility parameters, each worker's MWTP to reduce their workplace risk by one unit can be calculated directly.

In the second step of this research, I estimate various traditional hedonic wage models that regress equilibrium wages on job attributes and characteristics of workers holding the jobs. Specifically, I estimate 3,580 models that vary by the choices over functional form for the hedonic wage function, covariates included, and econometric strategies. Note, these models are estimated using the market equilibrium wages that arise from the simulation. The simulated wage data are linked to workers' actual characteristics (e.g. age, education attainment) through the job sorting process I develop for the simulation. Thus, each simulated equilibrium wage has associated job characteristics and a set of workers matched to that job (with its equilibrium wage) with associated demographic characteristics. Thus, I observe the demographic characteristics that are based on an actual sample of workers. The marginal price of risk obtained from these models is then compared to the "true" MWTP computed from each worker's utility parameters. The difference in these two values is referred to as the "bias" from the empirical hedonic wage model, and it is used as the basis to evaluate the relative performance of different modeling choices.

Since the labor market equilibrium is defined in the simulation model, the performance of the hedonic estimators can be tested within a scenario in which all covariates determining the equilibrium are observed. They can also be tested in scenarios where I omit variables, which mimics the likely setting of empirical hedonic wage studies. The performance of econometric strategies used to alleviate omitted variable bias can then be

evaluated. For example, studies relying on cross-section labor market data focus on using instrumental variables and fixed effects for the occupation and/or industry of worker to alleviate omitted variable bias.[5] Likewise, panel data studies control for workers' time invariant characteristics to alleviate the endogeneity bias associated with unobserved worker characteristics (Viscusi et al. 2007; Kniesnser et al. 2010; Kochi, 2011). To evaluate the relative performance of the standard econometric strategies in the literature, I propose to use both cross-section and panel data and apply several econometric approaches including Ordinary Least Square (OLS), first differences and fixed effects to systematically evaluate the relative performance of these techniques.

In the analysis, I will also vary the hedonic wage model's parametric specification. I propose to estimate the models that vary by three types of transformations of the dependent variable (wage rate): i) linear, ii) natural log, and iii) Box-Cox transformation. The models also vary in two types of transformations of the independent variable (fatal risk rate): i) linear; and ii) natural log. The combination of the transformations of the wage rate and fatal risk rates leads to four types of functional forms: i) linear; ii) log-linear; iii) double-log; and iv)Box-Cox. I also explore models which include or exclude quadratic risk term as covariates.

I also focus on various choices for the number of occupation industry dummy variables used in the hedonic wage model. The sensitivity of estimates of the price of risk to inclusion of dummy variables for occupation and industry has been noted in many studies,

[5] Arabsheibani &Martin (2001), Gunderson &Hyatt (2001) and Siebert &Wei (1994) have used IV approach in their cross-section models.

including Black et al. 2003, Hintermann et al.(2010), Leigh (1995), and Mrozek and Taylor (2002). There is a controversy over whether or not industry dummy variables should be included in hedonic wage estimation (Viscusi and Aldy, 2003). In early studies, the risk variables are generally constructed based on worker's industry affiliation, and that leads to a strong correlation between risk variables and industry dummy variables (see Black et al., 2003 for detailed discussion). Including industry dummy variables generally increases the variance of the risk coefficient and often leads to an insignificant risk coefficient. However, omitting industry dummy variables when strong correlation exists between risk and industry variables may bias the risk coefficient.

My evaluation yields several important results. First, in the presence of omitted variables, the simple linear functional form and the linear Box-Cox functional form perform best at recovering true MWTP, while semi-log and double-log functional forms consistently perform poorly. This is an important finding as the vast majority of past studies have used log-linear functional form in the estimation models. Specifically, more than 78 percent of the 200-plus specification included in Mrozek and Taylor (2002)'s meta-analysis used log-linear functional form. Second, adding a quadratic risk term to a function generally increases the vulnerability of the model to omitted variables and also increases sensitivity to the econometric strategies employed. The effect of including a quadratic risk is especially pronounced with the semi-log functional form, decreasing their performance substantially. This finding is important, since it is common in the past work to include risk as quadratic in

their wage hedonic wage function (Leigh and Folsum, 1984; Olson, 1981; Dorsey and Walzer, 1983; and Scotton and Taylor, 2003).

The third finding is that the impact of occupation and/or industry fixed effects on the estimates of the MWTP varies across the functional form. Inclusion of these fixed effects appears to improve the performance of linear Box-Cox functional form, but worsen the performance of the double-log functional forms. Moreover, in the cross-sectional estimator, the two types of semi-log (linear semi-log and semi-log with quadratic risk term) and two types of Box-Cox functional forms (linear Box-Cox and quadratic Box-Cox) are particularly sensitive to the different specification of occupation and/or industry fixed effects, especially when a quadratic risk term is included. Finally, in examining the performance of panel models, I find that the first differences model with wider time interval (i.e. differencing across two years and rather than one) outperforms other panel data models. It appears that regardless of the functional forms, applying the first difference with wider time interval effectively reduces the bias and significantly improves the accuracy of the estimates of MWTP.

The bias analysis allows me to identify the best performing hedonic wage specifications and use these specifications to estimate the VSL using two labor force data sets. The goal of this step is to take what is learned in the simulation and apply the best performing models to commonly used labor force data. I find that if the SIPP data are employed, the VSL estimate is $0.54 million (in 2008 $) with a standard error of $0.18 million. With a 95 percent confident interval of $0.18 to $ 1.08 million, this point estimate

overlaps with the panel model estimates in Kochi (2010) which also employs the SIPP data. The 95 percent confident interval of their estimates is between $0.43 to $4.03 million (in 2008 $). My estimate is closer to the lower-end value of their estimates.

With the PSID data, the estimation result is quite a bit larger than the SIPP estimate. It implies a VSL of $15.4 million with a standard error of $7.8 million (in 2008 $) with a 95% confident interval of $0.3 million to $31.4 million. The 95% confident interval of estimates overlaps the typical estimates from the literature which have ranged between $2 million to $11 million (Viscusi, 2004; Viscusi et el., 2007; Kniesner et al., 2012; and Kochi, 2011).

The sensitivity analysis of the VSL estimates indicates that the sensitivity of the estimates to the occupation and/or industry fixed effects depends on which data are used in estimation. While the estimates with the SIPP data present a clear pattern that industry fixed effects reduces the estimates of VSL substantially, the estimates with the PSID data are rather robust to the occupation and/or industry fixed effects. These empirical results for PSID and SIPP data are supportive to Black et al.'s (2003) conclusion that estimates for VSL are highly sensitive to the source of labor market data, even when conditioning on econometric methodology.

Overall, it appears that the wide variation of VSL still remains when the same model specifications are applied to different labor force data sets. This implies that the VSL estimates may largely depend on the source of the labor force data available to researchers. This finding is reinforced by results from sensitivity analysis. Incorporating the researcher's

judgments on choosing the labor force data to modeling choices and exploring the impact of the worker sample on the VSL estimates can be a topic for future research.

This research contributes to the hedonic wage literature in three ways. First, the simulation approach to solve for a hedonic wage equilibrium is novel to the hedonic wage literature. The sorting algorithm constructs a controlled simulation context where workers' job sorting behavior is modeled without losing the consistency with the underlying hedonic wage theory—something that has not been done in previous hedonic wage studies. Although the approach to evaluating the performance of hedonic wage models is directly analogous to that used in CDM and KPP, it contributes to our understanding the strengths and weaknesses of the empirical models specifications in recovering the marginal price of fatal risk. Also, this is the first study that uses O*NET data to incorporate a rich set of wage determinant factors in the hedonic wage models. Finally, in combination with the O*NET data and information on best performing models, I provide VSL estimates that are based on econometrics and specifications that have been carefully investigated.

The remainder of this thesis is as follows. Chapter 2 reviews the theory of compensating wage differentials (hedonic wage theory) as related to the simulation and the estimation models. The maintained and extended assumptions in this dissertation are also discussed in chapter 2. In Chapter 3, I describe the data elements needed for the analysis in the thesis. Chapter 4 reviews the literature that the simulation model draws upon, and describes the simulation approach I take to solve the job-worker matching problem numerically and compute the analytical "true" hedonic wage equilibrium. In Chapter 5, I

present a detailed review on the data and the approach that KPP used to evaluate the hedonic pricing model specifications and describe how I apply their evaluation method to the hedonic wage model and how hedonic wage modeling choices are examined. Chapter 5 also presents the evaluation results along with suggested “best practice” models minimize bias. Chapter 6 summarizes main findings of the dissertation, and its contribution to the hedonic wage literature. I also present the empirical estimates for hedonic wage models with preferred model specifications and actual labor force data. The empirical estimates are then compared with literature. Chapter 6 is closed with the discussion of future work.

CHAPTER 2

THEORY OF COMPENSATING WAGE DIFFERENTIALS

This chapter reviews the theory of compensating wage differentials (hedonic wage theory) as related to the simulation and the estimation models that will be carried out in chapter 4 and chapter 5. The theory of hedonic wages is largely based on the theory of hedonic pricing developed by Rosen (1974) and the concept of compensating wage differentials for risk as described in early work by Jones-Lee (1974). Rosen analyzes the price determining process for attributes of heterogeneous goods**,** while Jones-Lee analyzes the workers risk-wage taking behavior in a state dependent expected utility theory framework. The combination of the theories from Rosen and Jones-Lee is the foundation of the hedonic wage theory.

In a hedonic wage equilibrium, there is assumed to exist a hedonic wage schedule, *v*, which relates all relevant job characteristics to wages. The job characteristic of interest here is workplace fatal risk, *p*. The hedonic wage function is written as, $W(p, z)$, where *z* is the set of job characteristics other than risk. This hedonic wage schedule is a locus of equilibrium wage-risk combinations that emerges from *all* workers' utility maximizing choices and *all* firm's profit maximization process.

The setting for the hedonic model of the labor market is one where workers have preferences for job amenities and firms supply those amenities; and hire workers when making production decisions. Firms differ in the costs of providing the amenities and their costs of hiring. Workers differ in their preferences over the job amenities. These amenities

could be any non-wage attribute of the job, such as the level of comfort of the work place (perhaps in terms of temperature control), the existence and quality of on-site child care facilities, or the level of workplace safety (McConnell, 2006). The marketplace interaction of workers with different preferences for amenities and firms with different cost structures generates a locus of equilibrium wage-amenity combinations which are referred to as a hedonic wage function, $W(p, z)$.

In the following sections, I describe hedonic wage theory as described by Jones-Lee (1974) and Rosen (1974) in more detail. Then, based on McConnell's (2006) overview of hedonic wage models incorporating more recent findings from hedonic wage literature, I discuss assumptions underlying the hedonic wage model. One important assumption maintained in many past applications is homogeneity among workers and firms. This is the assumption I wish to extend in this dissertation.

2.1 Hedonic labor market

2.1.1 Preferences over Risks and Wages: the Worker

In the context of valuing workplace safety, the focus of the analysis relies on the combination of wage and workplace risk levels. Workers' risk-wage taking behavior can be described by using the state dependent expected utility theory developed by Jones-Lee (1974) and formalized as follows.

Workers face uncertain outcomes of death and survival at work, and the individual worker maximizes his expected utility between being in either the death state or the survival

state. Let *U(W)* represent the utility of a healthy worker with wage w in the survival state, where we assume the marginal income is positive:

$$U' \equiv \frac{\partial u(w)}{\partial w} \geq 0. \tag{2.1}$$

The utility in the death state is represented by $V(\tilde{W})$. The utility in death state is assumed to be a function of wealth (wage), since individuals may obtain a certain utility by bequesting their wealth to heirs. If this is not the case, then the utility in death state is zero. We also assume that:

$$v' \equiv \frac{\partial v(\tilde{W})}{\partial \tilde{W}} \geq 0. \tag{2.2}$$

and

$$v'' \equiv \frac{\partial v_{\prime}(\tilde{W})}{\partial \tilde{W}} \leq 0. \tag{2.3}$$

Further, we assume that workers obtain higher utility from survival than from death.

$$U(W) > V(\tilde{W}), \tag{2.4}$$

Workers choose a job with a probability of death, *p*, to maximize their expected utility. Suppose workers initial wealth is W (>0), and the on job level of fatality risk *p* is $0 < p < 1$. Consequently their expected utility is:

$$EU^0 = (1-p)U(w) + p\,V(w), \tag{2.5}$$

Now, suppose workers fatal risk level is increased to p' $(0 < p < p' < 1)$. To sustain the original level of utility, workers need to be compensated by additional wages. Let the compensation level that leaves workers in the same expected utility level be δ. This

compensation is referred to as the willingness to accept (WTA) compensation for an increment change of risk. It is also assumed that δ is a function of the risk preference parameter, α. By definition, the following equality holds:

$$EU^0 = (1- p') U(w+ \delta(\alpha)) + p' V(w+ \delta(\alpha)), \quad (2.6)$$

where $\delta(\alpha)$ is chosen to satisfy (2.6). Recall that EU^0=(1- p) U(w) +p V(w) and we can rearrange (2.6) and see that:

$$\delta = \delta(p, W, \alpha , EU_0). \quad (2.7)$$

Substituting this δ function back to (2.6), and total differentiating with respect to *p*, the marginal willingness to accept compensation for a marginal change in risk, $\partial\delta / \partial p$, is given by:

$$\frac{\partial \delta}{\partial p} = \frac{U(w+\delta)-V(w+\delta)}{(1-p)U'- p V'} > 0, \quad (2.8)$$

where $U' = \partial U(w+\delta)/\partial\delta$ and $V' = \partial V(w+\delta)/\partial\delta$. From (2.1), (2.3) and (2.4), the equation (2.8) is positive. The second derivative of (2.8) is also positive, indicating that the workers indifference curves are positively sloped and convex.

Figure 2.1 shows an example of two individuals' indifference curves. The indifference locus for worker 1 is δ_1 and δ_2 is the indifference locus for worker 2. Along each indifference locus, the utility level is constant. The worker's indifference function is a function of risk level, *p*, worker's risk preference parameter, α, and wage level, *W*. Since the indifference function is convex, the movement towards northwest direction increases their utility level. Therefore, worker 1 has a higher utility level in $\delta 1'$ than in $\delta 1$. In addition, worker 1 is more risk averse than worker 2, thus the δ_1 curve is steeper than $\delta 2$ curve. The

steeper δ curve indicates that the worker 1 requires more wage compensation to accept a marginal increase in risk than worker 2 when they face a same risk level change. In the same market setting, different optimal levels of risk-wage combination are generated by workers' different marginal rates of substitution between risk and compensation. In the market equilibrium, the slope of the indifference curve between *w* and *p* equals to the slope of the hedonic wage function.

2.1.2. Firm's Production and Risk

From the firm's perspective, job-related amenities such as safety are costly to supply. Isoprofit functions will reflect various combinations of wages and level of amenity that hold profits at a constant level. A firm's profit function can be complicated depending on which cost factors we include (McConnell, 2006). Here I assume a simple profit function that is:

$$\pi=\varphi k(L) - WL - c(p;\mu)L, \tag{2.9}$$

where φ is the price of output, $k(L)$ is the production function as defined over labor inputs, w is the wage rate, c is the cost of providing a certain level of safety per worker and μ is the efficiency parameter of providing a certain safety level s which can be considered as $s=1-p$.

For simplicity, we assume there is no factor other than labor to produce output. Firms maximize their profits subject to the hedonic wage schedule (thus $W=W(p;z)$) by choosing the number of workers hired and the level of safety provided. The profit maximizing conditions with respect to L and p are:

$$\frac{dk}{dL} = \frac{W+c}{\varphi}, \tag{2.10}$$

$$\frac{dw}{dp} = -\frac{dc}{dp}. \tag{2.11}$$

Equation (2.10) indicates that in order to maximize profits, the firm should choose labor inputs such that the marginal productivity of labor (left hand side) should equal to the wage rate plus the cost to provide a marginal reduction to the risk per worker (right hand side) assuming the price of output equals to one. Equation (2.11) shows that the firm's marginal implicit price of risk represented by the hedonic wage schedule (left hand side) should be equal to the marginal safety-cost saving from a marginal decrease of safety level (i.e. increase of risk level) (right hand side). Firms choose *L* and *p* such that (2.10) and (2.11) hold.

By rearranging firm's profit function (2.9), we obtain the firm's wage offer curve (OC):

$$OC=w(p, L, \mu, \pi)=(\varphi k (L) – c(p;\mu)L-\pi)/L. \tag{2.12}$$

This offer curve shows the firm's trade-off between providing safety and wage compensation for a given profit level. It is referred to as an isoprofit curve. The slope of the OC curve is given by *dOC/dp=-dc/dp*. Assuming the cost of providing additional unit of safety is increasing at a increasing rate as safety level increases. Equivalently, the cost of additional unit of safety level is decreasing at decreasing rate as risk increases. OC is a positively sloped concave function from below. At the maximum profit level, the conditions (2.10) and (2.11) hold.

Figure 1 illustrates the OC for two types of firms. To simplify the discussion all firms are assumed to offer identical job characteristics except fatality risk level. The only difference comes from the different efficiency of providing safety at work. Firm 1 has an

offer curve OC_1 and Firm 2 has an offer curve OC_2. Profits are higher on offer curves that are in the south-eastern direction. For example, firm 1's profit level is higher in $OC_{1'}$ than OC_1 since at every risk level, firm 1 pays less wages along OC1' than OC1. Firm 1 has a flatter offer curve than firm 2, indicating firm 1 has a higher efficiency of providing an additional unit of safety than firm 2. Firms maximize their profit where their offer curves are tangent to the hedonic wage schedule. Different shapes of the offer curves result in firms finding different optimal levels of providing safety and compensation given the market equilibrium wage/risk function. Firm 1 provides higher safety levels than firm 2, given the hedonic wage schedule.

In the basic hedonic wage framework presented above, the hedonic wage schedule is generated by a joint distribution of workers' risk preferences (α) and firms' efficiency in providing safety (μ). As illustrated in Figure 2.1, worker 1, who is more risk averse than worker 2, will be hired by firm 1 for whom providing safety is relatively inexpensive at a risk of p_1 and a wage of w_1. Worker 2, who is less risk averse than worker 1, will be hired by firm 2 for whom providing safety is relatively costly. The hedonic wage schedule is an envelope function which traces out the worker/firm pair's labor contract, where the workers' willingness to pay for a marginal decrease in risk level is equal to the firms' marginal cost of providing a marginal decrease in the risk level. If workers risk preferences are homogeneous and firms differed in the cost of providing safety, the hedonic wage schedule would trace out the common indifference curve of workers. If all firms' costs of providing safety are

homogeneous and workers' risk preferences differed, the hedonic wage schedule would trace out the common isoprofit function of firms.

There are two means by which the locus can shift. The first is through a change in the cost structure of supplying safety. If technical change makes supplying safety cheaper, firms' isoprofit functions will shift leftward, allowing workers to reach higher indifference curves. The location of the new envelop of isoprofit functions will be determined by the differential impact of technical change on different types of firms. The second force that can shift the locus is the change of the distribution of workers' risk preferences. Redistribution of workers' risk preference affects the labor supply to the different types of firms which results in the redistribution of isoprofit functions through changes in the wage level and the profit levels of firms.

2.2 Underlying assumptions

The theory outlined above relies on the following major assumptions:

a) workers are identical except their risk preference,
b) firms are identical except their efficiency in providing safety,
c) workers value even small changes in workplace safety,
d) workers are fully informed of the risk on every possible job,
e) safety cannot be provided costless,
f) labor market is competitive so that risk can be treated as a "transacted" attribute of a job.

Next, I will discuss assumptions that are maintained and extended in this research, and the extent to which they are relaxed, if they are relaxed.

2.2.1 Assumptions maintained in the research

The assumption that workers may not have perfect information about the occupational risk level of a particular job they are about to take (Smith, 1985, Scotton and Taylor, 2009, Slovic, 2000 and Lichtenstein and Slovic, 2006). The assumption (d) above has received considerable attention in the literature. The concern is that workers might have perceptions of fatal risks of jobs which may or may not be the same as actual risks. In addition to being unaware of actual fatalities, workers' subjective perception of risk may pose other problems such as cognitive biases. Perceived characteristics of a threat such as the amount of dread it evokes, the level of controllability one has in the face of it, and how observable or well-known it is to those affected and to science, are among the factors shown to bias an individual's risk perceptions (Scotton and Taylor, 2009).

A few studies have compiled workers' subjective preferences regarding risks and used this information in the econometric analysis (Viscusi, 1979, Viscusi and O'Conner, 1984, Gerking et al. 1988, and Liu and Hammitt, 1999). However, little is known about the correlation between the perceived risks and "actual" fatality risk levels as perceived risks might be more voluntary and individually controllable. The standard approach in the literature is to use "objective risk" measures that are based on the statistical rate of fatalities attributable to specific populations. I will follow this approach to apply the fatal risk in this research. In particular, I incorporate objective fatal risk measure as a factor affecting a

worker's job choice in a sorting process. In other words, it is assumed that workers are fully informed of the risk on every possible job when they sort into the jobs. Then, various hedonic wage estimation models will be carried out based on the equilibrium generated from the sorting process. Since the empirical model's specification is based on the structural model generated from a sorting process, the assumption (d) will be embeded in the analysis.

Another assumption that has been criticized is that workers are freely mobile and that the labor market and its participants are always in equilibrium (assumption f). Because transaction costs exist in the labor market, individuals wishing to readjust their wage-risk position will need to incur costs of job search and may forfeit benefits associated with job tenure if they change positions. If worker's mobility is restricted, then worker's observed choice may not satisfy the first order condition when attributes change (McConnell, 2006). However, since examining the direction and magnitude of the difference between the "true" hedonic wage locus and the empirical estimated hedonic wage function is the focus of this analysis, it is necessary to maintain the assumption of the existence of equilibrium and costless movement by workers among jobs.

Considering that assumption c is related to d, and assumption e is trivial, I maintain the assumptions c, d, e, f, stated earlier in this research. These four assumptions support that the existence of an equilibrium hedonic wage locus that reveal workers' MWTP for risk reduction.

2.2.2 Assumptions extended in this research

With the assumptions of homogeneity among workers in their productivity/ skills and firms in their job attributes (other than fatal risk), variation in risk preferences means that we can observe equilibria as pictured in Figure 2.1. It explains why individuals with similar productivity/ skills can be found at different wage-risk trade-offs.

However, it is also commonly observed that workers are offered higher pay in the safer jobs. This anomaly arises in the labor market because of differences in workers' productivity/ skills. For individuals who have exactly the same risk preferences but different productivities, they will be facing different offer curves and they will likely optimize at different wage-risk combinations. Figure 2.2 illustrates one example. High productive worker and low productive worker represent different groups of workers who are assumed to have identical preferences irrespective of productivity. Generally higher educated workers are more productive than lower educated workers because of their higher human capital. High productive worker can be though as the one with higher education attainment and low productive worker as the one with lower education attainment. Even though the marginal cost to provide safety is same for both groups at every risk level, firms would pay higher educated workers who have higher marginal productivity more than lower educated workers. Therefore, with same level of non-pecuniary job attributes, productivity level dictates that the utility level of each group attained in the labor market can be different. In this example, the workers with higher education attainment have higher utility. Thus higher educated workers have different hedonic wage schedules as illustrated in Figure 2.2.

Failure to control for either worker or job characteristics that are correlated with job risk will cause biased estimates. Here I use a simple example to explain the nature of the bias. Suppose the true model of wage determination is:

$$\ln(W_i) = \alpha_{oi} + \alpha_1 X_i + \beta p_i + b Z_i + u_i , \qquad (2.13)$$

where $\ln(W_i)$ is the natural logarithm of the ith worker's wage, X_i is the vector of covariate, p_i is the measure of risk, and (α_1, β) are coefficients to be estimated, and u_i is the error term of the regression, Z_i is a vector of job disamenities and b>0 by assumption. It is straightforward to show that:

$$p\lim\hat{\beta}^{OLS} = \beta + \frac{\operatorname{cov}(p_i, Z_i | X_i) b + \operatorname{cov}(p_i, u_i | X_i)}{\operatorname{var}(p_i | X_i)} . \qquad (2.14)$$

Unmeasured worker characteristics could bias the coefficient of risk in either direction. If worker productivity is measured imperfectly and more productive workers tend to accept safer jobs, then $\operatorname{cov}(p_i, u_i) < 0$ and $\operatorname{cov}(Z_i, u_i) < 0$ are expected. Thus the second term in the numerator tends to underestimate the price of risk (Hwang et al. 1992).

In contrast to general productivity, Garen (1988) and Shogren and Stamland (2002) emphasizes on risk-specific productivity. Risk specific productivity is a feature of workers which- for example as a result of cool-headedness or good physical agility-enables them to be more productive in hazardous professions than average workers. Therefore, workers who are highly skilled in handling risk select themselves into more risky occupations in which they have a comparative wage advantage. If risk-specific productivity is not observed by the researcher, the position correlation between risk-specific productivity and wages will lead to an upward bias of the coefficient of the fatal risk.

Empirically, the unobserved heterogeneity issue has long been recognized as one of major reasons for inconsistent estimates (Black et al, 2003). However, it is very challenging to address this issue due to the data limitation. The difficulty is that not all productivity differences are measurable. Researchers have attempted to control for differences of workers by estimating the hedonic wage function for a relatively homogeneous subset of the labor market. In this research, I will follow the same logic to segment the market to alleviate the issue of omitting workers heterogeneity in terms of their productivity and skill. I will separate the labor market in two ways: 1) grouping workers into different skill levels based on their observed education attainment and working experience, and ii) grouping workers into different occupation-industry clusters based on available occupation/industry coding system. While the assumption of workers homogeneity in their productivity and skill is not entirely relaxed, it is extended into that the workers are assumed to share similar skill and productivity in the same groups.

In terms of job characteristics, bias in estimating the risk premium could be caused by the endogeneity of non-pecuniary characteristics if they are correlated with risk. For example, one may find a correlation between fatal risk and physical exertion required for a job or risk and environmental factors such as noise, heat, or odor. It can be illustrated with same example as in equation (2.13) and equation (2.14). As we expect that job disaminities covary positively $cov(p_i , Z_i | X_i) > 0$ and as $b>0$, the second term in the numerator causes us to overestimate the price of risk. In other word, if higher-risk jobs have undesirable

characteristics not measured by the researcher, the risk variable tends to capture these characteristics biasing the estimates upward.

In order to alleviate these heterogeneity issues, I further incorporate heterogeneity among workers and firms in the following ways. First, I have used data from O*NET, which is new source of occupational characteristics information. It allows me to relax the assumption of firms' homogenety in terms of their job attributes and define the job based on its associated job attributes (including workplace fatality risk). Second, a hypothetical labor market is constructed where workers heterogeneity is presented as their different preferences for various job attributes. Specifically, a structural model is developed in which worker's job choice is driven by the offered wage, job attributes including workplace fatal risk, and preferences for the various job attributes. By solving worker-job sorting process iteratively via a numerical approach, workers heterogeneous preferences will be captured. Given workers' heterogeneous preferences are the driving forces to the worker-job sorting process and determine workers' choices of jobs, the estimates of workers' preferences are expected to absorb unobserved factors affecting job choice decision.

Relaxing the assumption of homogeneity among workers and firms in the ways described above provides a context which is crucial to carry out the analysis in this research. Based on the simulated labor market which reveals the unobserved factors determining the "true" hedonic wage locus, the analysis will be focusing on exploring the effects of the unobserved heterogeneity among workers and firms on estimating hedonic wage models. The

"true" hedonic wage locus is used as the baseline value to compare to the estimation from both cross-sectional and panel models.

The data considerations for each of the above extensions, as well as a review of the relevant literature and the proposed models, are presented in detail in chapters 3, 4 and 5. In the next chapter, I will discuss the data in detail that is adopted in this research.

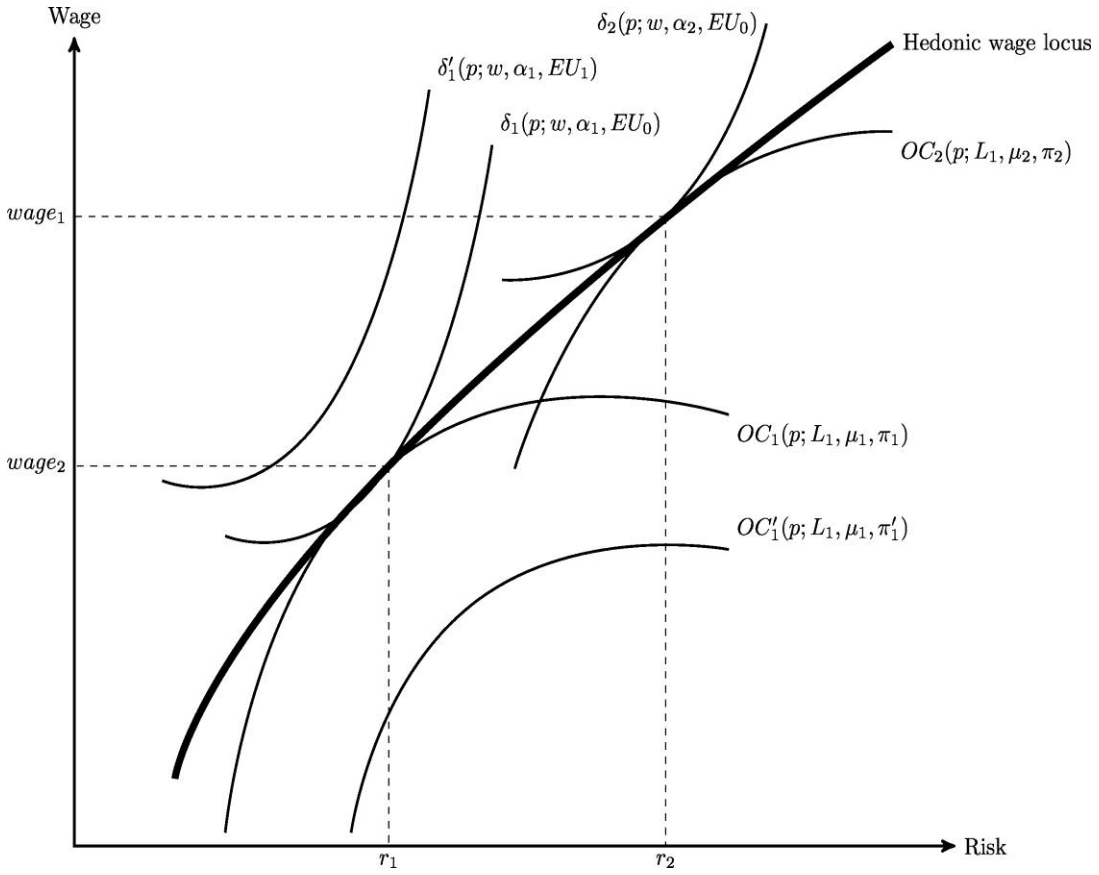


Figure 2.1: *Hedonic wage locus for heterogeneous workers and heterogeneous firms*

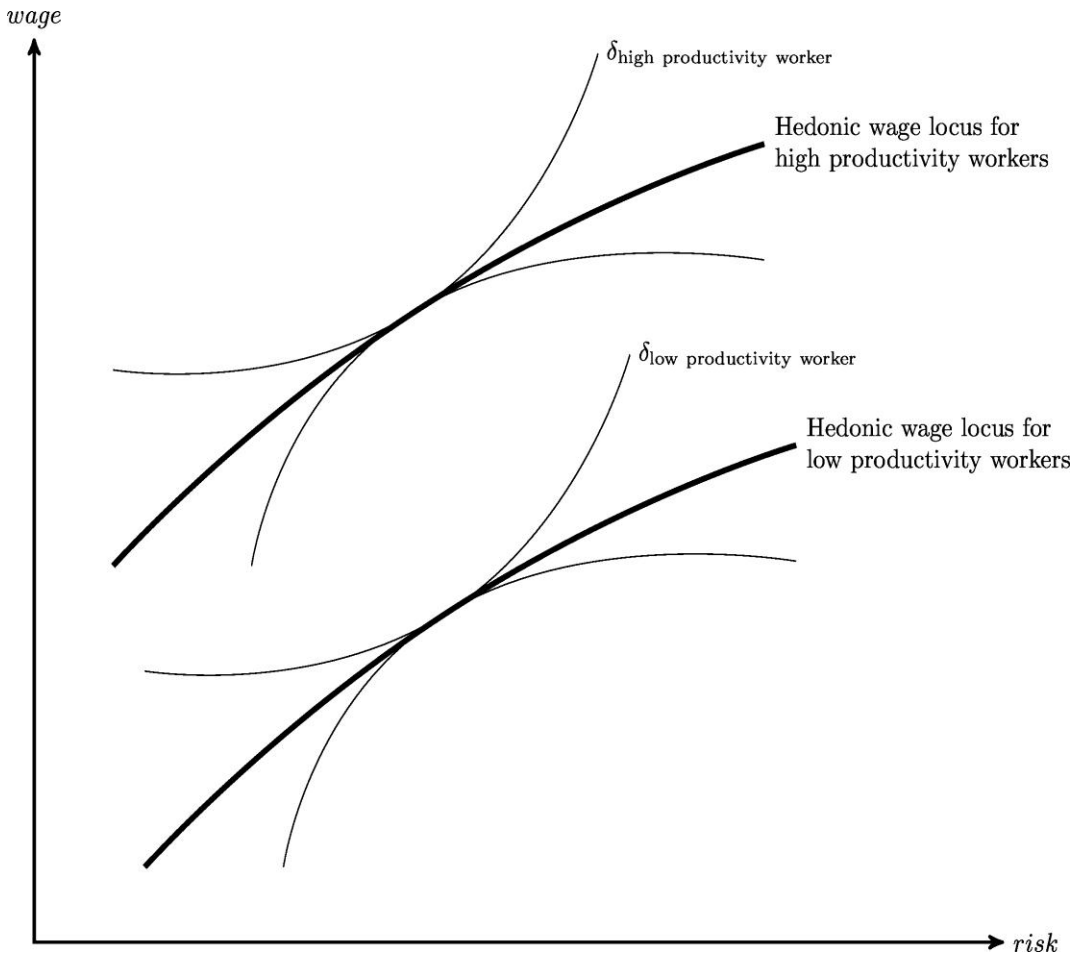


Figure 2.2: Hedonic wage locus for workers with different productivity

CHAPTER 3

DATA

In this chapter I describe the data elements needed for the analysis in the thesis. These are:

i) Panel data on U.S. workers' wages, job and demographic characteristics. I will use the Survey of Income and Program Participation (SIPP) that follows 1,892 workers from 2004 to 2007.

ii) Panel data on workplace fata risks. I will use confidential Census of Fatal Occupational Injuries (CFOI), which is a complete census of workplace deaths over the same time period.

iii) Panel data on workplace injuries from the BLS Injuries, Illnesses and Fatalities (IIF). It provides a count of injuries requiring days-away-from-work or job transfer / restricted work activity.

iv) Data on job characteristics other than injuries or deaths. This dataset is obtained from the O*NET, complied by BLS.

Each of the above is described in detail below.

3.1 Labor force data: the Survey of Income and Program Participation (SIPP)

Data for individual hourly wage, job and socio-economic characteristics are obtained from the 2004-2006 Survey of Income and Program Participation (SIPP).[6] The SIPP is national micro panel data collected by the U.S. Census Bureau, and it contains rich

[6] Detailed data description and data download is available at http://www.bls.census.gov/sipp/

information on workers' individual income, labor force status, and general demographic characteristics. People are interviewed by phone or in person every four months. Each four months reference period is called a *wave*. The 2004 SIPP data is used in this analysis, which is a panel that lasts for four years and contains twelve waves. Only one observation from each year is used because the primary fatal risk data employed in the analysis is reported annually.

The main advantage of using the SIPP data is its structure as panel data and the richness of the information it contains. Another two frequently used labor force data sets are the Current Population Survey (CPS) and Panel Study of Income Dynamic (PSID). The CPS is a cross-sectional data set which does not contain multiple observations of each worker over the years. The SIPP is a medium length time series data with a large sample size. The PSID is the long length time series data (since 1968), but with relatively small sample size (about 6000-8000 households). Compared to the CPS and PSID, the SIPP provides more information on each employee's current earned and critical non-labor income. According to the SIPP User Guide, the SIPP records up to 70 cash and in-kind sources of income data for the current year, while the CPS and the PSID contain limited information on workers' income. Of particular importance is that the data from the SIPP on *total earned income, total income, and total other unearned income* which are not available from the CPS and PSID, allow me to compute the worker's non-labor income, which is crucial to carry out the simulation model in chapter 4.

The major wage determinants commonly used to specify the hedonic wage estimators in the literature are available in the SIPP. These include age, educational attainment, gender, race, marital status, number of children under 18 in the household, union status, and the worker's occupation and industry group of the firm for which workers work. The worker's current wage is available on either an hourly or monthly basis. The hourly wage data is preferred since there are more missing observations in the monthly wage data. Also collected are residential location (urban vs. rural and region of residence), availability of employer provided health insurance, size of firms, and whether or not the person works over-time as potentially important wage determinants.

The sample is full time workers who hold only one job at a time. Workers whose earning is less than minimum wage, or whose age is less than 18 or more than 65 are omitted from the analysis. Given the framework for the simulation model, I select a balanced panel to keep track of workers' job choices over the years. This restriction results in a total of 5,676 observations for 1,892 individual workers that are used in the analysis.[7] Table 3.1 reports the summary statistics for the sample. The table includes variables describing worker and job attributes that are commonly used in hedonic wage studies. The average hourly wage is $19.04 in 2005 dollars, which is somewhat higher than the average hourly earnings in the U.S. labor market of $17.27 in 2005. The average age of workers is 44 years old, and 29.5% of the sample only have high school degree, 35.3% of the sample has attended college, and 28.8 % of the sample has a bachelors or higher degree. The national average trend over 2004 to 2006 is that approximately 30.2% of the workforce having graduated from high school,

[7] Applying the same cleaning method to the PSID will result in a sample size of around 800.

25.9% have attended a college but have no degree, and 24.2% hold a bachelor's degree or more.[8] Compared to the national average of educational attainment in the labor force during 2004 to 2006, the SIPP sample is somewhat more highly educated.

Forty-four percent of the sample is female and 66% of the sample is married. The majority of workers are white, 11.2% of workers are Hispanic and another 10.2% of workers are African American, which is nearly identical to the racial composition of the U.S. labor force during this period. [9] Seventy percent of workers are union members or covered by union benefits, which is slightly higher than the current average union membership rate of 12.5%.[10] The average number of children under 18 for the SIPP sample is 0.8 which approximately the same as national average of 0.9. The population of the workers is denser in middle Atlantic division, east north central division, south Atlantic division and pacific region. The rest of the workers are evenly distributed across New England, west north central, east south central, west south central, and mountain area. The share of workers who work in each major occupation and industry group is as follows: only 8.5% of workers work in service related occupations, 35.6% are in managerial, professional and related occupations, 26.5% are in sales and office occupations, 11.09% have a natural resource, construction or maintenance occupation, and 18.2 % are in production, transportation or material moving occupations.

---

[8] The education attainment data from labor force statistics is available at www.bls.gov/webapps/legacy/cpsatab4.htm

[9] The education attainment data from labor force statistics is available at www.bls.gov/webapps/legacy/cpsatab4.htm

[10] The education attainment data from labor force statistics is available at www.bls.gov/webapps/legacy/cpsatab4.htm

There are few workers engaging in the construction, transportation / communications / utility and public industry. These workers make-up only 8.9% of the total sample. In contrast, a high proportion (38.8%) of workers engage in the trade, service and manufacturing industries, in which 7.5% of workers engage in the service industry, 17.9% of workers engage in the wholesale or retail trade industry, and 40.3% of workers engage in the manufacturing industry.

For the period of 2004-2006, the SIPP uses 4-digit occupation and 4-digit industry codes from 2000 Census Occupation Classification System as described in the SIPP data dictionary to classify the occupation and industry, respectively.[11] In this analysis, the SIPP industry codes are matched to the NAICS industry code and SOC occupation code and converted to the 22 occupation and 23 industry groups constructed by Scotton (2000, 2010).

3.2 Fatal risk data: Census of Fatal Occupational Injuries (CFOI)

The number of occupational fatal incidences is obtained from the Census of Fatal Occupational Injuries (CFOI) file collected by the BLS for the period 2004 through 2007. This data is not publicly available. The non-public use CFOI file contains richer information on death circumstances of all workplace deaths in US, and it identifies both the industry and occupation of the worker at the time of death. This feature of the CFOI data enables one to create risk data varying by occupation within industry.

In this research, the aggregation of risk at the occupation and industry level and the construction of the risk rates replicates the approach in Scotton (2000, 2010). The occupation

[11] The SIPP data dictionary is available at www.census.gov/sipp/diction.html

and industry classification systems are used to delineate 23 occupations within 22 industries where 421 risk rates are measured.[12] The risk rate in each industry-occupation cell is calculated as follows:

$$p_{oi} = \frac{D_{oi}}{W_{oi}}, \tag{3.1}$$

where the numerator D_{oi} is the annually number of death incidents in occupation o and industry *i*, and the denominator W_{oi} is the annual average total number of workers in occupation *o* in industry *i*. The average number of workers will be obtained from the CPS database. [13]

The numerator in equation 3.1, *Doi*, is obtained from the CFOI files for the period 2004-2006, which contains 17,338 deaths. The CFOI contains various characteristics of deceased workers, such as gender, age, race, location of accident, size of firm, event of accident, occupation, industry, and time of accident. The list of all available variables in the CFOI files is reproduced in Appendix 3.A. The CFOI uses 6-digit Standard Occupational Classification (SOC) System 2002, and 6-digit North American Industry Classification System (NAICS) 2002 code to classify the occupation and industry, respectively. According to these coding systems, the workplace fatality data is reported from 473 occupation and 1,183 industry categories. Scotton (2000, and 2010) aggregated occupation and industry codes in the CFOI into a matrix of 23 occupation and 22 industry codes by using 2-digit

[12] An alternative way is to compute the risk rate by 10 industries within occupations. However, as shown in Scotton (forthcoming), the risk rate measures are very similar based on these two different calculation methods.

[13] The data source is at data.nber.org/morg/annual

NAICS industry and 2-digit SOC codes. The 23 occupation and 22 industry classification used in Scotton (2000, 2010) is reproduced in Appendix 3.B. I follow the same way to coarsely aggregate the occupation and industry groups for the CFOI and construct annual fatality risk measures for these 23 occupations by 22 industry cells.

After applying the method from Scotton (2001, 2010), all the occupations and industries reported in the CFOI are aggregated into 421 occupation industry groups. Merging to the SIPP data reduces the number of occupation industry groups to 211. The mean of average fatal rate in the SIPP sample is 0.33×10^{-4} with standard deviation of 0.55×10^{-4} and the median is 0.11×10^{-4}. The highest risk rate in the sample is 5.45×10^{-4}, which is faced by the construction/ extraction workers in mining industry.

Table 3.2 provides comparison of fatal risk rates' distribution before and after merging to the SIPP sample. Panel A compares the risk rates at few important percentiles. The difference of the risk rates at lower percentile (25%) and higher percentile (95% and up) implies that jobs that are not observed in the SIPP sample are more likely the ones with extremely low or extremely high (top five percent) fatal risk rates. This implication is reinforced by panel B which provides the comparison through the distribution of jobs at different risk rate levels. It appears that the distribution of jobs at the lower risk level ($0<$ risk rates ≤ 0.25) does not change much after the fatal risk rate measures are merged to the SIPP sample, but the amount of occupation-industry pairs having zero fatal risk rates is reduced in the sample. Also notice that jobs in the SIPP sample are more densely distributed

at the median level of risk rates (0.25< risk rates ≤ 1), but slightly less densely distributed at the high level of risk rates (risk rates > 1).

3.3 Non-fatal injury data: Injuries, Illnesses, and Fatalities (IIF) program

The risk rate of non-fatal injuries is obtained from the 2004-2006 Injuries, Illnesses, and Fatalities (IIF) program administered by the BLS. The IIF estimates the number and frequency (incidence rates) of workplace injuries and illnesses based on logs kept by employers during the (calendar) year. These records reflect not only the year's injury and illness experience, but also the employer's understanding of which cases are work-related under recordkeeping guidelines promulgated by the Occupational Safety and Health Administration (OSHA).

Nonfatal recordable workplace injuries and illnesses are those that result in any one or more of the following: 1) loss of consciousness; 2) days away from work; 3) restricted work activity or job transfer; 4) medical treatment beyond first aid. These four criteria also define SOII measures in the case types, which are 1) total recordable cases; 2) total cases; 3) days-away-from-work; 4) job transfer or restriction cases and 5) other recordable cases. [14]

According to the Guidelines, days-away-from-work cases are those resulting in days away from work (beyond the day of the injury or onset of illness).[15] Job transfer or restriction cases are those leading only to job transfer or restricted work activity. This occurs when, as the result of a work-related injury or illness, an employer or health-care professional keeps, or recommends keeping, an employee from doing the routine functions of his or her job or

[14] See http://www.bls.gov/iif/oshwc/osh/os/ostb2071.txt for reference in year 2006.
[15] The guidelines are available at http://www.osha.gov/recordkeeping/handbook/index.html#1904.7_1

from working the full workday that the employee would have been scheduled to work before the injury or illness occurred. [16] This may include instances where a) an employee is assigned to another job on a temporary basis; b) an employee works at a permanent job less than full time; c) an employee works at a permanently assigned job but is unable to perform all duties normally connected with it. For example, in the case of an employee who suffers a work-related injury resulting in 5 days away from work, and upon returning to work, the employee was unable to perform normal duties associated with the job for an additional 3 days. This case would be recorded as a days-away-from-work case with 5 days away from work and 3 days of restricted work activity. Other recordable cases are defined as cases which require medical treatment beyond first aid, which do not result in any days away from work, nor job transfer or restriction, beyond the day of the injury or onset of illness.[17] Total cases are the sum of the counted number of days-away-from-work cases and job transfer or restriction cases. And adding injury numbers or rates of total cases and other recordable cases gives the total recordable cases measure. In order to obtain a more complete report of probability of non-fatal injury risk, the measure of risk rate that applied in the study is *total cases* measure (i.e. the sum of the counted number of days-away-from-work injury cases and job transfer or restriction injury cases).[18]

[16] According to OSHA recordkeeping handbook, an employee's routine functions are those work activities the employee regularly performs at least once per week.
[17] For detailed definition of first aid, one can refer to OSHA recordkeeping handbook at www.osha.gov/recordkeeping/handbook/index.html
18 This measure has also been used in Viscusi (1992) and Viscusi and Aldy (2003).

The Injuries, Illness and Fatalities (IIF) data uses NAICS classification system. Data is provided on the basis of 2-digit to 6-digit level respectively. I convert the 4-digit industry code defined by 2000 Census industry classification system in SIPP to 3-digit 2000 NAICS code, so that I can take advantage of SIPP coding system to obtain most detailed non-fatal injury data. Appendix 3.D shows the industry group mapping. The first and second columns show the 3-digit NAICS industry group title and codes, respectively. The third and fourth columns show the 4-digit code for the industry groups which are defined by 2000 Census industry classification system and are included in the 3-digit NAICS industry groups.

Once the 3-digit level injury risk rates are merged to the period of the SIPP data, it shows that the injury risk is differentiated by 66 industries. In order to get the idea of correlation between the non-fatal injury risk and the fatal injury risk, I also report the summary statistics of non-fatal injury risk rates on a more aggregated level. Table 3.3 presents the injury risk rate measures aggregated into 19 industries to compare to the corresponding fatal risk rate measures. The highest non-fatal injury risk bearing industry is the couriers and messengers, where the risk rate is 694 per 10,000. The second highest non-fatal injury risk bearing industry is the transportation and warehouse where the injury risk rate is 394 per 10,000. The lowest risk bearing group is the information, finance and insurance industry where the risk rate is 33 per 10,000. In comparison, instead of the couriers and messengers industry, the highest fatal risk bearing industry is mining industry and the second highest risk bearing industry is transportation and warehouse. Industry of

management of companies and enterprises faces zero fatal risk. The second lowest fatal risk bearing industry is the information, finance and insurance industry.

Figure 3.1 presents the correlation between non-fatal injury risk and fatal risk in a more graphical way. Each marker indicates the value of fatal or non-fatal injury risk rates for a specific industry. The red dashed lines trace out the non-fatal injury risk levels over 19 different industries separately and blue straight line is fatal risk levels. They compare the trend of fatal and non-fatal injury risk rates over the industries. It appears that both non-fatal and fatal injury risks tend to be relatively higher in the industries such as construction, transportation and warehousing, and waste management and remediation services. And finance and insurance, professional, scientific, and technical services and education industries are the ones with relatively lower risks. The larger difference between two risk measures appears in the mining, manufacturing, and currier and messengers, warehouse and storage industries.

3.4 Occupation characteristics data: Occupational Information Network (O*NET)

To capture non-pecuniary aspects of jobs other than risks, I employ a relatively new database that contains rich information about occupational attributes. The Occupational Information Network (O*NET), produced by the U. S. Department of Labor's Employment and Training Administration, is a database system for collecting, organizing and describing data on job characteristics and skill requirements for workers within detailed occupations.

The first public release version of O*NET was in 1998 and I use O*NET 15.0, released in June 2010 which contains 572 O*NET job descriptors cross 854 occupations.

The job descriptors in O*NET cover a wide range of attributes, including types of job skills or attributes that are required to perform the tasks of an occupation. Most of O*NET job descriptors are measured using an "importance" and a "level" scale, although these two scales are highly correlated. Level values are scaled from 1 to 7 in which each scale value indicates a specific requirement regarding the task. For example, in the question of "how often are you required to wear protective and safety equipment ?", the 1 to 7 scale refers to "once per year or less", "more than once per year", "more than once per month", "more than once per week", " daily", "several times per day", "hourly or more often". The Importance value is scaled from 1 to 5 indicating "how important is this task to performance on this job?", where value 1 represents "that task is of no importance" and 5 represents "the task is extremely important". [19] As suggested in Hirsch and Schumacher (2010) and Ingram and Neumann (2006), the level scale provides more detailed information in terms of the intensity of one specific skill needed for a job, and hence I will retain only the descriptors measured in level form.

All O*NET job descriptors are initially categorized into the following four major groups: 1) worker characteristics such as workers' ability, interests, and preferred working style; 2) worker requirements such as basic skills, knowledge and education that are required to accomplish the task; 3) job experience and licensing requirement for the job; 4)

[19] See www.onetcenter.org/ombclearance.html for detailed questionnaires.

occupational requirements including working context and general working activity requirements which represent activities or behaviors that underlie the accomplishment of major work functions (Jeaneret, Borman, Kubisiak & Hanson, 1999). An example of the elements in these broad groups is shown in Appendix 3.E. I will focus on the *occupational requirements* group and use the data provided in this group as the job attributes data set in my model. Following DeLeire et. Al (2009), Ingram and Neumann (2006) and Abraham and Spletzer (2009), I group the descriptors into five "informative categories" seeking to extract: i) cognitive demands, ii) motor skill demands, iii) physical demands, iv) working conditions, and v) the demand for interaction with people from both public and coworkers while on the job. Detailed descriptions of grouping the attributes for each category are shown in Appendix 3.F.

As indicated in Ingram and Neumann (2006), within each of the five informative categories, there are a large set of descriptors and they are conceptually correlated to each other. I adopt the approach of factor analysis to collapse the O*NET occupation descriptors within each informative category into a small set of indexes. Factor analysis can be summarized as a two-step procedure to create uncorrelated latent variables for each of the five informative groups. The first step is to determine a small set of descriptors that are preferably uncorrelated with the large set of descriptors (most of which are correlated to each other). In the second step, indexes are created with the small set of descriptors that measure similar things conceptually in this context (see Ingram and Neumann, 2006 for more detailed description of factor analysis).

Factor analysis can be easily carried out using software with built-in packages such as STATA and SAS. A simple example of how the factor analysis might be done in STATA is presented in Appendix 3.G. Factor analysis loads the descriptors in a way that accounts for a maximum proportion of covariance of all the descriptors. The resulting "factor" represents a latent variable which is aggregated by a cluster of descriptors. The value of the factor indicates the relevance of the corresponding descriptor in defining that latent variable and the cluster is composed of the high relevant descriptors. One factor analysis may output multiple factors. It implies that multiple uncorrelated latent variables should fully reveal information that the loaded descriptors contain. Once the clusters of descriptors are selected, each uncorrelated latent variable will be generated by taking the average value of the associated high relevant cluster of descriptors. In processing O*NET data, I apply factor analysis into five informative categories. The resultant set of the uncorrelated latent aggregated variables represents each informative category. While the strength of the factor analysis is that it can generate a set of uncorrelated covariates, the drawback is that it tends to lose a clear economic interpretation of the resulting factors.

There are 15 latent aggregated variables that are constructed from the underlying O*NET variables to describe occupational characteristics. The 15 job attributes can be briefly described as following:

1) Cognitive demands involving information processing and analyzing data or information,

2) Cognitive demands involving creative thinking, strategies developing, planning, and decision making,

3) Cognitive demands involving monitoring process, inspecting equipment or material,

4) Physical demands that require physical activities,

5) Physical demands that require climbing ladder or poles, or kneeling and crawling,

6) Physical demands that require repetitive physical motion, operating machine or handling tools,

7) Low levels of hazardous work environments,

8) Working environment that is exposed to the weather or hot, cold temperature,

9) High levels of hazardous working environments,

10) Skills require handling technical or electronic or mechanical device or equipment,

11) Interacting with computers,

12) Interacting with people indirectly,

13) Interacting with people directly, or being involved in conflicting situations,

14) Job requires group or team-work and taking responsibility for the outcome of the group performances,

15) Working directly with the public.

After creating the above set of job attributes, I match them to the labor market and risk data. To match O*NET job attributes to the SIPP data, a cross-walk matching the two different occupation coding schemes has to be created. The SIPP code identifies workers in

330 occupations based on three-digit 2000 Census Occupation Classification system, while O*NET provides a more detailed eight-digit Standard Occupational Classification 2000 (SOC 2000) breakdown with 854 occupational categories. I follow a crosswalk recoding O*NET data into 6-digit occupation groups to merge to the SIPP.[20] For those occupations that are not one-to-one matches from O*NET to SIPP, I calculate O*NET descriptor values aggregately based on the weighted average across the components of O*NET occupations, using total employment reported in the Occupational Employment Statistics (OES) as weights. For a small number of occupations that are defined at a level of detail that OES employment figures are not available, they are assigned equal weights.[21]

Table 3.4 reports the summary statistics for the 15 job attributes across occupations held by workers in the SIPP sample. The statistics are reported by gender and consistent with intuition. It indicates that men are more likely to choose in jobs with higher requirements for physical practice or demand of motor skill associated with machines or tools. Women tend to sort into the jobs requiring more interaction with people.

Table 3.5 reports the correlations coefficient between the job characteristics and workplace fatal and non-fatal injury risks. In general, the correlation between fatal risk, injury risk and other job attributes are consistent with intuition. For example, injury risk is

[20] The cross-walk of occupational coding system is Census 2000 occupational classification, with 2000 Standard Occupational Classification (SOC) equivalents, which can be obtained from www.bls.gov.soc.socguide.htm.

[21] There is one-to-one match from O*NET to the SIPP for 485 occupations. 262 of the remaining involve a mapping of two or more O*NET occupations to a SIPP category and they can be weighted by OES employment data. The last 107 occupational code in O*NET data is defined in such a detail way that OES employment figures are not available. In these cases, component O*NET occupations are assigned equal weights.

positively related to fatal risk. Hazardous working conditions are positively correlated to fatal risk, where the high level of hazard working condition (exposure 2) has the highest correlation. As expected, the job requirements involving technical or electronic or mechanical device or equipment handling and repetitive physical motions (cognitive skill 3, physical demand 2, 3 and motor skill 1) also show positive correlations to the fatal risk.

In sum, each of the 221 possible jobs defined in my analysis is differentiated by 17 attributes of which fatal risk is measured using CFOI, non-fatal injury risk is from the IIF and 15 job attributes are created using O*NET data.

Table 3.1

Statistic summary of SIPP sample from 2004 wave (2004-2006): Worker characteristics and job characteristics for labor force sample (N=5676)

Variable	Definition	Mean	Std.	[min max]
hr_wage	hourly income	19.08	9.56	[3.5 73.19]
age	age in years	43.52	9.56	[19 64]
blacknh	1 if individual is black and non-hispanic	0.1	0.3	[0 1]
college	1 if individual attended college but with no degree	0.35	0.47	[0 1]
female	1 if individual is female	0.44	0.49	[0 1]
hispanic	1 if individual has a hispanic origin	0.11	0.31	[0 1]
hsgrad	1 if individual graduated from high school	0.28	0.45	[0 1]
kid18	1 if individual has any kids under 18 years old	0.43	0.49	[0 1]
married	1 if individual is married	0.69	0.46	[0 1]
othrace	1 if individual is non-white, non-black, non-hispanic	0.04	0.21	[0 1]
ugdeg	1 if individual have bachelor degree or more	0.28	0.45	[0 1]
	Job Characteristics			
Empall	1 if number of employee at all locations >100	0.91	0..27	[0 1]
empsize	1 if number of employee at worker's location <25	0.25	0.43	[0 1]
hifull	1 if individual is provided part o f health insurance by employer	0.23	0.42	[0 1]
hipart	1 if individual is provided full health insurance by employer	0.74	0.43	[0 1]
salaried	1 if individual gets monthly payment	0.51	0.49	[0 1]
workot	1 if individual usually works more than 40 hours	0.28	0.45	[0 1]
union	1 if individual is a union member or covered by union	0.17	0.37	[0 1]
	Worker's location			
metro	1 if individual lives in metro area	0.735	0.53	[0 1]
Region1/9	Dummy variables indicating in which of the following nine census regions the worker lives:			
	individual lives in New England Division	0.05	0.23	[0 1]
	individual lives in Middle Atlantic Division	0.14	0.34	[0 1]

Table 3.1 Continued

individual lives in East North Central Division	0.17	0.37	[0	1]
individual lives in West North Central Division	0.08	0.27	[0	1]
individual lives in South Atlantic Division	0.18	0.38	[0	1]
individual lives in East South Central	0.07	0.25	[0	1]
individual lives in West South Central	0.08	0.28	[0	1]
individual lives in Mountain	0.07	0.26	[0	1]
Individual lives in Pacific	0.13	0.33	[0	1]

Table 3.2

Distribution of fatal risk rates created from CFOI before / after merging to SIPP sample [a]

Risk rate Measure	22 occ x23 ind.	
	Before merge to the SIPP	After merge to the SIPP
	N [b] =419	N=211
Panel A. distribution of risk rates		
25%	0 [c]	0.03
50%	0.09	0.11
75%	0.38	0.25
90%	1.01	1.01
95%	1.92	1.27
99%	4.65	2.82
Max	17.05	5.45
Panel B. Number of jobs		
risk rate=0	31.02%	21.11%
risk rate>0 & ≤ 0.25	35.32%	34.78%
risk rate>0.25 & ≤ 0.50	10.73%	17.08%
risk rate>0.50 & ≤ 1.0	11.21%	16.14%
risk rate>1.0	11.45%	10.86%

a. Rates are reported as deaths per 10,000 workers.
b. N is the number of the jobs presented in the data.
c. The risk rate is reported at various percentiles

Table 3.3

Summary of non-fatal injury risk and fatal risk across 19 industry groups in the SIPP sample

			Non-fatal injury rate (per 10,000)			fatal risk rate (per 10, 000)		
Ind.Code	Industry group	% of workers [a]	Mean (std.)	Min	max	Mean (std.)	Min	max
2100	mining	0.39%	235.45 (57.62)	[89.99	290]	3.49 (1.41)	[0.97	5.45]
2200	utility	1.20%	236.47 (12.55)	[220	250]	0.49 (0.53)	[0	1.98]
2300	construction	7.12%	333.41 (9.41)	[320	340]	1.09 (0.58)	[0.06	2.69]
3123	manufacturing	20.97%	344.14 (129.76)	[100	580]	0.26 (0.24)	[0	1.03]
4200	wholesale trade	6.34%	286.13 (81.06)	[210	390]	0.45 (0.48)	[0	1.53]
4445	retail trade	9.27%	303.27 (107.12)	[89.99	430]	0.23 (0.16)	[0	1.19]
4800	transportation and warehousing	2.98%	396.44 (173.94)	[170	770]	1.73 (1.23)	[0.0	2.93]
4923	currier and messengers, warehouse and storage	0.85%	694.16 (131.43)	[540	880]	0.36 (0.47)	[0	1.97]
5100	information	2.41%	124.26 (17.52)	[30	138.09]	0.24 (0.46)	[0	4.33]
5200	finance and insurance	6.43%	33.34 (4.72)	[30	40]	0.05 (0.06)	[0	0.98]
5300	real estate and rental and leasing	1.60%	193.4 (12.66)	[180	210]	0.27 (0.34)	[0	1.65]
5400	professional, scientific, and technical services	5.16%	53.34 (4.72)	[50	60]	0.15 (0.3)	[0	1.96]

Table 3.3 Continued

5500	management of companies and enterprises	0.05%	130 (20)	[110 150]	0		
5600	administrative and support and waste management and remediation services	2.26%	239.84 (116.73)	[170 530]	0.56 (0.64)	[0	2.26]
6100	educational services	11.79%	96.66 (4.71)	[89.99 100]	0.06 (0.1)	[0	1.04]
6200	health care and social assistance	13.71%	276.91 (146.93)	[100 580]	0.06 (0.05)	[0	0.76]
7100	arts, entertainment, and recreation	1.27%	278.88 (39.27)	[200 330]	0.32 (0.29)	[0	1.52]
7200	accomodation and food services	3.07%	195.68 (84.94)	[130 330]	0.17 (0.16)	[0	1.64]
8100	other services	3.15%	139.83 (31.71)	[100 190]	0.35 (0.32)	[0	1.59]

a. Both of the non-fatal injury risk rates and fatal risk rates are aggregated into 19 industry groups by using 2-digit NAICS code. Thus, the percentage of the workers in each group is the same for non-fatal injury risk and fatal risk measures.

Table 3.4

Summary of occupational attributes of the SIPP sample by gender

Characteristics		Description	Male (N=3177)	Female (N=2499)
			Mean (std.)	Mean (std.)
Exposure	1	low level of hazardous work environment	2.63 (0.90)	1.97 (0.60)
	2	providing working environment exposed to the weather or hot, cold temperature	2.13 (0.75)	1.47 (0.38)
	3	high level of hazardous working environment, specifying technical or electronic or mechanical device or equipment	2.26 (0.30)	2.45 (0.45)
Cognitive demand	1	cognitive demands involving information processing and analyzing data or information	3.45 (0.44)	3.56 (0.48)
	2	cognitive demands involving creative thinking, strategies developing, planning, and decision making	3.27 (0.45)	3.26 (0.49)
	3	cognitive demands involving monitoring process, inspecting equipment or material	3.35 (0.37)	3.07 (0.42)
Physical demand	1	physical demands requiring physical activities	2.85 (0.80)	2.60 (0.78)
	2	requires climbing ladders, poles, or kneeling and crawling	2.25 (0.41)	1.96 (0.25)
	3	requires repetitive physical motion, operating machine or handling tools	3.09 (0.75)	2.70 (0.61)
Motor skill	1	specifying technical or electronic or mechanical device or equipment	2.18 (0.68)	1.60 (0.42)
	2	interacting with computer	3.24 (1.01)	3.67 (0.93)
People skill	1	requires interacting with people indirectly	3.41 (0.70)	3.67 (0.62)
	2	interacting with people directly, or involving in the conflicting situations	2.92 (0.38)	3.14 (0.44)
	3	requires group or team working and taking responsibility for the outcome of the performances	3.57 (0.44)	3.48 (0.43)
	4	Working directly with public. Each of the jobs is differentiated by the job attributes and the injury risk	3.06 (0.63)	3.26 (0.55)

Table 3.5

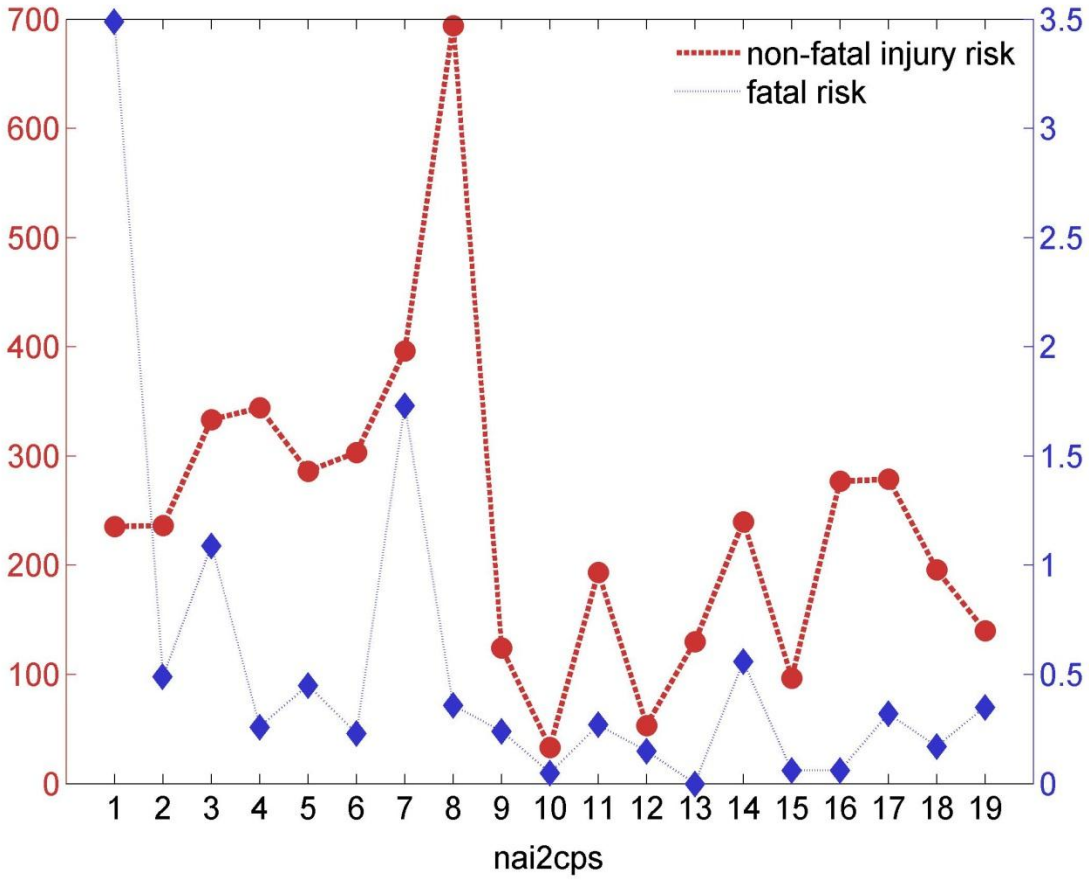
Pairwise correlations of job characteristics and injury risks [a]

		Fatal	Injury	Exposure			Cognitive skill			Physical demand			Motor skill		People skill			
				1	2	3	1	2	3	1	2	3	1	2	1	2	3	4
Fatal risk		1																
Injury risk		0.22	1															
	1	0.49	0.34	1														
	2	0.69	0.27	0.72	1													
Exposure	3	-0.2	-0	0.02	-0.2	1												
	1	-0.3	-0.2	-0.4	-0.3	0.1	1											
	2	-0.2	-0.2	-0.3	-0.2	0.2	0.72	1										
Cognitive skill	3	0.33	0.24	0.62	0.4	0.2	0.16	0.25	1									
Physical demand	1	0.24	0.27	0.69	0.4	0.3	-0.6	-0.4	0.32	1								
	2	0.62	0.3	0.77	0.8	-0.1	-0.4	-0.3	0.48	0.57	1							
	3	0.42	0.32	0.77	0.5	0	-0.6	-0.6	0.42	0.04	0.7	1						
	1	0.43	0.27	0.81	0.6	-0.1	-0.1	-0	0.67	0.42	0.7	0.64	1					
Motor skill	2	-0.4	-0.3	-0.6	-0.5	-0.1	0.77	0.49	-0.2	-0.7	-0.6	-0.7	-0.3	1				
	1	-0.4	-0.3	-0.7	-0.4	0.1	0.71	0.7	-0.2	-0.7	-0.6	-0.9	-0.5	0.73	1			
	2	-0.1	-0.1	-0.2	-0.1	0.6	0.27	0.38	-0	0.04	-0.2	-0.3	-0.3	0.11	0.43	1		
	3	0.11	0.05	0.14	0.2	0.3	0.18	0.51	0.36	0.19	0.1	-0.1	0.08	-0.1	0.28	0.5	1	
People skill	4	-0.1	-0.2	-0.5	-0.2	0.3	0.34	0.5	-0.2	-0.2	-0.3	-0.6	-0.4	0.32	0.67	0.7	0.3	1

a. The entry of the cell is the correlation coefficient.

Figure 3.1: *Plot of fatal and non-fatal risk rates across 19 major industries* [a]

a. Both of the non-fatal injury risk and fatal risk rates are measured per 10,000 workers. The values of non-fatal injury risk and fatal risk, however, are of different magnitudes. Thus, in order to present the comparison in a clear way, two scale charts are used. The y-axis in red shows the sale values for non-fatal injury risk measure. The y-axis in blue shows the scale values for fatal risk rates measure.
b. The numbers on the x-axis represent 19 major industry categories that are observed in the SIPP sample, which are:
   1= mining
   2= utility
   3= construction
   4= manufacturing
   5= wholesale trade
   6= retail trade
   7= transportation and warehousing
   8= currier and messengers, warehouse and storage
   9= information
   10= finance and insurance
   11= real estate and rental and leasing
   12= professional, scientific, and technical services
   13= management of companies and enterprises
   14= administrative and support and waste management and remediation services
   15= educational services
   16= health care and social assistance
   17= arts, entertainment, and recreation
   18= accommodation and food services
   19= other services


non-fatal injury risk
fatal risk
700
600
500
400
300
200
100
0
3.5
3
2.5
2
1.5
1
0.5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
nai2cps

CHAPTER 4

SIMULATING HEDONIC LABOR MARKET

4.1 Introduction

In this chapter, I propose a simulation approach to solve the job-worker matching problem numerically to characterize a hedonic wage equilibrium, with which I can explore the relative importance of econometric challenges in estimating the VSL. The simulation exercise focuses on conducting a sorting algorithm that is driven by heterogeneous preferences that workers have over non-pecuniary job attributes. In the simulation exercise, I estimate how a range of factors which are not commonly observed by the researcher affects the choices of workers and employers, as well as how these choices interact to determine the equilibrium allocation of workers across jobs. The primary goal of this chapter is to build a framework where worker heterogeneity is directly modeled rather than treated as unobservable as is often done in the literature. In such context, the "true" wage schedule can be traced out through each worker's wage-risk combination achieved at the market equilibrium.

The remainder of the chapter is organized as follows. Section 4.2 focuses on discussing the simulation strategies and the sorting model which this research is built upon. Section 4.3 introduces the sorting algorithm employed in this research. The implication of the sorting algorithm is presented in section 4.4, and section 4.5 presents the computational details and results.

4.2 Related work

To complete my analysis, I draw on four distinct literatures. The first is concerned with the generalization of the Roy model. This model provides the intuition to construct a worker-job sorting process which is not only based on the wage, but on non-pecuniary job attributes. Second, the literature of simulation approaches informs the logic that I follow to solve the hedonic equilibrium numerically. Finally, I draw on the literature of two-sided matching model in the labor market setting which provides the basis upon which the sorting algorithm is developed.

4.2.1 Roy Model sorting

In addition to the theory of compensating wage differentials, the analysis in this chapter is closely related to the idea of occupational (Roy) sorting based on idiosyncratic returns in DeLeire, Kahn and Timmins (2010). In particular, they extended the standard Roy model to a more general version that allows for a common non-pecuniary component of utility associated with each sector that workers sort into.

In its initial form, the Roy model has been used to characterize the optimal choice of occupation or economic sector given wage offers from two or more sectors. The idea behind the Roy model is that occupational choice can be seen as a setting of self-selection, where workers choose the sector which provides them with the higher wage. However, the standard Roy model is restrictive in the way that non-pecuniary aspects do not matter in the job choice decision.

DeLeire, Kahn and Timmins (2010) used minimum order statistics to prove and correct for the bias in estimating hedonic wage model without accounting for workers' non-pecuniary return from job selection. Intuitively, the minimum order statistic identifies the difference in the non-pecuniary component of utility follows directly from the observation that no individual will choose a less-preferred choice (on the basis of non-pecuniary considerations) unless the offered exceeds this threshold. Thus, the minimum wage observed in the less-preferred sector should be exactly the minimum wage observed in the more-preferred sector plus the difference in non-pecuniary components. After identifying the non-pecuniary component of utility for multiple occupations, they estimated the effect of non-pecuniary occupation characteristics including fatality risk by regression. Their results indicated that non-pecuniary return matters in job sorting decision and the traditional hedonic wage model without controlling for Roy sorting bias yield significantly downwardly biased estimates of the VSL.

While the minimum order statistic model is not applied in this analysis, the idea of the generalization of the Roy model provides the intuition of considering non-pecuniary aspects of the choice of jobs in the sorting algorithm developed in this chapter. In order to build up a context where the wage-risk locus can be mapped out at the market equilibrium, I incorporate a non-pecuniary component in the utility specification and non-pecuniary job attributes are considered as factors during the job choice decision.

4.2.2 The Simulation approach

Based on the idea of generalizing the Roy model described above, the underlying mechanism generating the hedonic labor market equilibrium is an assortative process where firms supply jobs with different levels of the attributes and different wages, and workers with different tastes for the job attributes will sort themselves accordingly. Such a link between jobs and workers can be used to make inferences regarding workers' underlying preferences in equilibrium.

The logic of constructing a simulated hedonic market builds upon the idea first developed in Cropper, Deck and McConnelll (1988) (hereafter CDM) and expanded in Kuminoff et al. (2010). In both studies, the hedonic equilibrium was computed in a housing market setting. The theoretical framework can be described as follows:

In the hedonic property value model, *J* houses are defined over a vector of characteristics *Z*. Household *i*'s utility depends on the characteristics of its house Z_j and on its consumption of a composite numeraire good *C*, given its idiosyncratic income y_i and preferencesα_i. Households bid against each other for the housing stock, with houses sold to the highest bidders. Each household is assumed to choose a specific house and quantity of *C* that maximize its utility subject to a budget constraint:

$$max_{j,c}\ \mathrm{U}\ (Z, \mathrm{C}; \alpha_i) \text{ subject to } y_i = \mathrm{C} + p_j, \tag{4.1}$$

where p_j represents expenditures on house *j*.

A hedonic equilibrium is achieved when every household occupies its utility-maximizing location, given prices and characteristics. It implies that no household is willing to pay more for house *j* than the household buying the house (i.e. houses are sold to the highest bidder). Each household occupies exactly one home, for which it has the maximum bid. Moreover, the household could not receive higher utility by purchasing any other house. I describe the assignment problem with same notation as in CDM and Kuminoff et al (2010). If b_{ij} denotes household *i*'s bid for the j^{th} house and A_{ij} is an indicator where A_{ij}=1 if and only if household *i* occupies that home, then equilibrium can be presented as:

$$b_{ij} = \max \{b_{ij}\} \text{ iff } A_{ij} = 1, \tag{4.2}$$

$$\sum_i A_{ij} = \sum_j A_{ij} = 1. \tag{4.3}$$

Given a specification for utility and the data of housing characteristics, the bid function can be used to compute a vector of prices and an assignment of people to homes that jointly satisfy the conditions for a hedonic equilibrium. In both CDM and Kuminoff et al (2010), the process of solving for a locational equilibrium begins by assigning each household *i* a reference level $\tilde{u}$. The utility function is inverted to solve for C =U^{-1} (x, $\tilde{u}$; α). Inserting U^{-1}(.) into the budget constraint and rearranging terms provides an expression for the household's maximum willingness to pay for each home:

$$b = y - U^{-1}(x, \tilde{u}; \alpha). \tag{4.4}$$

In CDM, they used translog and Diewert functional form for utility. The general form of utility in CDM is:

$$U_{ij} = g(C) + \sum_j a_{ij} (M_i)\, g(Z_j) + 0.5 \sum_j \sum_k \beta_{jk}\, g(Z_j)\, g(Z_k). \tag{4.5}$$

Household's preference parameters a_{ij} depend on a vector of measured buyer characteristics, M_i, which includes race, family size, whether the household has children, and the education and occupation of the household head. Preferences also reflect unmeasured, individual-specific taste factor over housing attributes α_{ij}, thus $a_{ij}= \alpha_{ij} + \delta_i' M_i$. In Kuminoff et al. (2010), they explored the Cobb-Douglas, translog, and Diewert specifications for utility function. Moreover, instead of constructing the preference parameters as a function of buyer's characteristics, household's relative taste factors are directly introduced in the model as: U_{ij} = ln(c) + $\sum_j \alpha_{ij}$ g(Z_j) + 0.5$\sum_j \sum_k \beta_{ik}$ g(Z_j) g(Z_k).

The sorting algorithms that are used to solve hedonic equilibrium numerically are different in CDM and Kuminoff et al (2010). The algorithm in CDM is based on a utility adjustment mechanism by Wheaton (1974) where utility is endogenously determined by the equilibrium house allocation process. Utility adjustment follows changes of the parameters in the utility function. In Kuminoff et al (2010), the algorithm is constructed by using a series of hypothetical second-price auctions for individual houses. I will briefly discuss these two algorithms below.

The equilibrium prices in CDM (1988) were computed by iteratively solving an assignment process developed by Wheaton (1974). In the process, the number of households is given exogenously and housing stock is at fixed amount in the market. The given number of households must be allocated to the houses, and the allocation of houses must not exceed the amount of houses available in the market. The iterative mechanism is driven by a "shadow price". When shadow price is positive, the marginal household in the market is

bidding more than the house's actual price. When shadow price is negative, it is bidding lower than the actual price. As Wheaton noted, a nonzero shadow price is an indication of disequilibrium. Household's bid is correspond to a given level of utility. Households will decrease their bid if they can (because that raises their utility), or raise their bid if they have to (making themselves worse off). Competition in the housing market drives households to adjust its utility until its bid are just such as to ensure the shadow prices to be zero.

The iterative bidding algorithm (IBA) in Kuminoff et al. (2010) was developed in Kuminoff and Jarrah (2010). In their algorithm and also in the CDM, household's utility is assumed to be monotonic in the numeraire. Every household submits a bid and the houses are assigned to the highest bidders. The household who won the bidding will pay the second highest bid plus ε, which is less than or equal to its bid. Therefore, the new assignment must either increase household's reference utility or leave it unchanged. After solving for house price and updating utility, the auction is moved to the next home. The algorithm continues running second-price auctions until the occupant of every home is paying an ε above the second highest bid for that home. The IBA iterates over a series of second-price auctions for individual houses until subsequent bidding has no effect on prices or the assignment of people to houses.

Comparing to Wheaton's linear programming algorithm in CDM, iterative bidding algorithm in Kuminoff and Jarrah (2008) does not constrain households to be uniquely assigned to houses on intermediate iterations of the algorithm. In other words, same household could win the different auctions for different houses. The difference of the prices

that the biding winner is asked to pay for multiple assigned houses motivates the household to choose between houses. In this way, the algorithm will keep evolving. The IBA continues to iterate until the vector of housing prices converges, signaling the market has cleared. Releasing the constraint of unique assignment on intermediate iterations provides the advantage of the IBA that it avoids the need to store large assignment matrices in the computer's memory, enabling them to increase the size of their simulated market. In both of these simulation approaches, the allocation of houses to the individuals and the prices of the houses in equilibrium are determined by individuals' heterogeneous preferences. And these preferences are recovered parametrically

To make the simulation realistic, both CDM and Kuminoff et al. (2010) employed real housing data to construct the simulated housing market. Specifically, CDM drawn the data from houses sold in Baltimore City or Baltimore County in 1977-1978. The attributes of houses come from Multiple Listing data, and the attributes of neighborhoods from the 1980 census of housing and population. Kuminoff et al (2010) employed housing stock data on approximately 104,000 single-family housing sales in Wake County, North Carolina between 1992 and 2000. The values for the housing attributes are based on the 1990 Census of Population and Housing. The data contain the sale price of each house, its structural characteristics, and the characteristics of its neighborhood. The housing attributes data are used to distinguish difference houses and the sale price of each house is used to calibrate the sorting process so that the preferences parameter of the utility function can be recovered numerically and hedonic equilibrium can be computed.

The advantage of simulation approach is that the MWTP of each household for house characteristics can be computed through identified utility. Since these "true" MWTP measures are known in the simulated context, various empirical hedonic models are estimated and judged on the basis of how closely it estimates "true" marginal price within the same simulation context. Both CDM and Kuminoff et al. (2010) focus on evaluating the performance of various hedonic modeling specifications in dealing with omitted variable problem.

4.2.3 Labor market two-sided matching model

In the same way that housing markets are modeled by a sorting process, the equilibrium relationship between worker's wage, job attributes and worker's characteristics in hedonic wage models are generated by an assortative matching process. Employers can be viewed as offering a package of job characteristics (including the quality of the work environment) to the workers with different characteristics and preferences. The equilibrium arises from the matching of workers to firms with varying levels of the job characteristic.

Given the parallel circumstances in the labor market, I extend the logic of simulation in CDM and Kuminoff (2009) into a hedonic wage framework and develop a sorting algorithm based on a two-sided matching model. The two-sided matching literature is built upon the work of one-to-one matching model of Koopmans and Beckmann (1957), Gale (1960), and Shapley and Shubik (1971). In these models, agents on one side of the market are matched to objects (or agents) on the other side, and each "match" generates a pair-specific surplus. An efficient matching is the one that maximizes the sum of all involved parties'

payoffs. The assignment model has been applied to a broad range of settings in which individuals in one group are matched with individuals, agents or firms in a separate second group; referred to as one-to-many matching and many-to-many matching. Examples include models of college attendance in Fox (2006) and hospital-intern markets in Roth (1984, 1991).

The sorting algorithm in this research builds upon the matching model in Crawford and Knoer (1981). It is conducted in a labor market setting, in which heterogeneous firms form matches with heterogeneous workers. In their framework, one firm can be matched to multiple workers, but each worker can be matched to at most one firm and preferences are separable across pairs. In other words, the payoff from a particular firm-worker pair is independent of the other matches the firm forms. Instead of relying on the linear programming duality theory commonly used in the assignment model, they used the Gale-Shapley deferred acceptance algorithm where firms bid for workers in simultaneous ascending auctions (Gale and Shapley 1962; Roth and Sotomayor 1990). The algorithm begins with each firm proposing employment to its most preferred set of workers at the one arbitrarily low wage. When a set of workers reject the offer, the firm makes offers to other workers to fill its remaining openings. The algorithm stops when no rejections are issued, singaling that the market is in equilibrium.[22]

[22] In a subsequent paper, Kelso and Crawford (1982) extended the model to the setting which allows the employers' hiring capacity to be decided endogenously instead of being fixed as in Crawford and Knoer (1981). Given this research relies on the data from observed matches to model workers sorting behavior, the hiring capacity of each firm is treated as known and exogenous in the model. Therefore, the market setting in Crawford and Knoer (1981) is more suitable to be applied in this research.

4.3 The sorting algorithm

When considering the actual job market, one may think of job-worker matching as a one-to-one matching in which a complete set of job characteristics identify a unique job to which one particular worker is assigned. However, empirically there is a lack of detailed information on employers' and job characteristics, and so one can only construct "job types" that group jobs broadly. Therefore, analytically, the worker-job matching is done as a many-to-one match type, which is the same setting as in Crawford and Knore (1981). In equilibrium, while job assignment to each worker is unique, multiple workers can sort into the same kind of job.

4.3.1 Job-sorting algorithm framework

Consider a labor market where each job is characterized by a vector of attributes and workers differ in their "tastes", (i.e. their preferences over non-wage job characteristics specific to each particular job). Workers sort into a fixed number of jobs. During the sorting process, workers maximize the expected value of the utility and firms maximize their profits. The set of wages is chosen so that the supply equals demand and thus at all pairs of wage-risk combinations. When all agents are matched, the market is in equilibrium.

The set up can be formalized as a labor market with *m* workers and *n* firms, indexed $i = 1 \ldots m$ and $j = 1 \ldots n$ respectively. From a worker's perspective, a job can be viewed as a differentiated product, that is, a good with a bundle of characteristics including fatal risk level (the probability of death) and the job attributes other than fatal risk. The supply of jobs

is assumed exogenous to the model. Each worker i, is assumed to choose a specific job that maximizes his expected utility subject to a budget constraint:

$$\max EU = \max (1-P_j)*U(C_i, Z_j ; \alpha_i) + P_j*U(\overline{y_i}+B_i^j)$$

$$\text{st. } \overline{y_i} + W_i^j - C_i = 0, \tag{4.6}$$

where C_i represents the composite commodity, P_j is the fatal risk rate for job j, Z_j is the vector of job attributes (not including fatal risk) for job j, $\overline{y_i}$ is non-labor (exogenous) income, W_i^j is the wage from job j, α_i is a vector of the worker's preferences parameters toward job characteristics and B_i^j is the bequest benefit in the state of death. After substituting the budget constraint, the expected utility function can be presented as:

$$\max EU = \max (1-P_j)*U(\overline{y_i} + W_i^j, Z_j ; \alpha_i) + P_j*U(\overline{y_i}+B_i^j). \tag{4.7}$$

A firm's profit function can be complicated depending on which cost factors we include. Here I assume a simple profit function based on that in McConnell (2006), which is:

$$\pi = V(S^j) - \sum_{i \epsilon S^j} W_i^j - C(p;\mu)\, S^j, \tag{4.8}$$

where S^j is the hiring capacity of firm j, V(.) is the production function as defined over labor inputs, C is the cost function of providing a certain level of safety and μ is the efficiency parameter of providing a certain safety level. For simplicity, I assume there is no factor to produce output other than labor and the output price is normalized to one. Initially, assume that productivity is homogeneous among workers within each firm. This assumption will be discussed and relaxed in the following section. Hiring capacity is fixed and the cost of providing safety for each worker is exogenous. Firms maximize their profits by offering

wages and providing a certain level of safety to the workers. Given these assumptions, the firms' maximizing problem is simplified to be one of minimizing the hiring costs in terms of wages.

The process of solving for a hedonic equilibrium in the labor market begins by assigning each worker *i* a reference level of utility $\tilde{U}_i$. As Wheaton (1974) and Kuminoff et al. (2010) chose reference utility randomly, I assume a baseline utility, $\tilde{U}_i$, is a function of only the worker's non-labor income, $\overline{y_i}$. This baseline utility is defined as a utility level that worker gets when being unemployed. $\tilde{W}^j_{i,t=0}$ is defined as the wage level which makes the worker indifferent between being in job j and being unemployed:

$$\tilde{U}_i = \mathrm{U}(\overline{y_i}) = (1-P_j)\mathrm{U}(\overline{y_i} + \tilde{W}^j_{i,t=0}, Z_j ; \alpha_i), \qquad (4.9)$$

where the subscript t=0 indicates the initial round of the sorting algorithm. Note that the assumption of no ex-ante utility in the state of death remains in this model. Solving equation (4.9) for $\tilde{W}^j_{i,t=0}$, implies $\tilde{W}^j_{i,t=0}=U^{-1}(\tilde{U}_i, Z_j, P_j, ; \alpha_i)-\overline{y_i}$, which is the initial minimum wages required by each worker. This minimum wage initializes the sorting process as follows. In order to attract worker *i* to take job *j*, the potential wage offer from employer j can be derived by increasing $\tilde{W}^j_{i,t=0}$by a small amount ε, i.e. $W^j_{i,t=1} = \tilde{W}^j_{i,t=0} + \varepsilon$. Therefore, once baseline utility is defined, the information on each job's attributes, fatal risk level and the data on each worker's preference parameters α_i can be used to solve for potential wage offer that each employer needs to provide if they want to hire one particular worker. Each firm initially

makes offers to its favorite set of workers. Firm j, for example, makes offers to S^j set of workers where $V(S^j) - \sum_{i \in S^j} W^j_{i,t=1} - c(p;\mu)\, S^j \geq 0$.

The sorting algorithm is then carried out as follows.[23]

Round0: $\tilde{U}_i$ is determined by each workers' non-labor income, which explicitly gives the reference utility level when he is not consuming any job attributes and not providing any labor services (i.e. from being unemployed).

Round1: In the first round of the sorting algorithm, given a job's attributes and skill requirements, each firm derives the potential wage which is enough to attract each worker to take their jobs, across all the workers. After evaluating the hiring cost, each firm makes offers, up to the hiring capacity of the firm, to the workers who require the lowest wage, second lowest wage, third lowest wages, and so forth, in order.

Round2: Workers update their utility level by using wage offers and corresponding job attributes from received job offers. Recall that $W^j_{i,t=0}$ is formulated as $W^j_{i,t=0} = U^{-1}(\tilde{U}_i, Z_j, P_j, ; \alpha_i) - \overline{y_i} + \varepsilon$ for all i and j. It is likely that some workers will receive more than one offer. For example, the worker who requires a lower wage for job j, may also be ranked as the lowest potential wage for job k. In this case, this worker will receive two offers and one rejection will be issued. Workers who receive

[23] For the ease of exposition, I will refer to "jobs" and "firms" interchangeably.

more than one offer reject all but the one which gives highest utility level. If there are rejections, the sorting algorithm goes to the next round.

Round3: Offers not being rejected in previous rounds remain in force. If worker i rejects an offer from firm j and accept the offer from firm k in round t-1, then the potential wage offer from firm j in round t increases to $W_{i,t}^{j}=U^{-1}(Z_k, P_k, W_{i,t-1}^{k}, Z_j, P_j; \alpha_i) - \overline{y_i} + \varsigma$ which induces worker i to switch the jobs. If wage in round t is not rejected, then $W_{i,t}^{j}= W_{i,t-1}^{j}$. Firms including the rejected ones in last round re-evaluate the potential wage cost taking into account their current potential wages ($W_{i,t}^{j}= W_{i,t-1}^{j}$ or $W_{i,t}^{j}=U^{-1}(Z_k, P_k, W_{i,t-1}^{k}, Z_j, P_j; \alpha_i) - \overline{y_i} + \varsigma$) and make offers to their favorite workers up to their capacity (not necessarily the same workers they offered to in the previous rounds).

In real competitive labor market, jobs are normally subject to negotiation between employers and prospective employees. To mimic this fact, this step allows the possibility that rejected employers will keep offering to the same worker after re-evaluating the potential offers.

Round4. The process stops when no rejection is issued in the current round. Workers then accept the offers that remain in force from the jobs they have not rejected.

Since the sorting algorithm is based on the two-sided matching model in Crawford and Knoer (1981), strategic behavior by workers and firms is not included in the mechanism. In particular, the rules for making, rejecting, and accepting wage offers determine that the

side "proposing" captures all the rents, i.e. the firms have full bargain power. This is in contrast to the standard approaches in industrial organization literature, which emphasize the link between worker's wage and the firm's profitability. They typically confront the bargain or rent sharing problems in imperfectly competitive market. For instance, bargaining models find that workers will receive wages in excess of their best alternative, and that this difference will depend positively on their firms' rents (eg., Blanchflower et al, 1990; Beckerman and Jenkinson, 1990; Margolis and Salvanes, 2001; Martins, 2009; and Guertzgen, 2010). Search models incorporate searching costs as trading frictions that would generate rents to firm-worker pairs even if all firms and all workers were identical (e.g., Diamond 1982; Pissarides 1990; Stole and Zwiebel, 1996; Cahuc, Marque and Wasmer, 2008).

To be able to incorporate rent sharing effect on determining wages, one needs to have data containing information about the worker's wage over time and simultaneously about the profitability of the worker's firm. These appropriate matched employer-employee data, however, are not available in US labor market.[24] While two-sided matching model on which this sorting algorithm is based, avoids the bargain problem, ignoring the existence of rent sharing leaves out one dimension of wage disparities which might persist between workers with identical observed individual characteristics, employed in different sectors. As it will be discussed in the following section, the sorting models are carried out and calibrated in each

[24] Currently, they only exist in some European countries. The best known matched employer-employee panel data sets are those of Denmark, Finland, Portugal, and Sweden. Each data set covers all firms in each country and all employees of each firm in each year over a relatively long period.

of 9 major industries to alleviate the problem of ignoring sectorial wage premia over different industries.

In the sorting algorithm described above, job-worker sorting outcomes change in two ways. First, conditional on worker i being one of the workers that employer j wants to hire across t rounds, and the employer being rejected by worker i across t rounds, then $W_{i,t}^{j}=U^{-1}(Z_{k'}, P_{k}, W_{i,t-1}^{k}, Z_{j'}, P_{j}; \alpha_{i})$ -$\overline{y_{i}}$+ ς where k ≠ j, keeps evolving and worker i is induced to change his offer-taking decision. Second, the rejected employers keep increasing the potential wage offer to their highest ranked workers up to the point where they change their set of favorite workers to whom they make offers by comparing the potential wages for all workers. This implicitly evolving process is consistent with Jones-Lee hedonic framework and drives the model to converge to equilibrium. The convergence will be proved formally in the following section.

4.3.2 Proof of the convergence

Under perfect information, the notion of equilibrium in a labor market which is implied by the convergence of the algorithm, requires that no firm and worker should be able to negotiate an agreement that is mutually beneficial when compared to (any of) those in force with their current partners. The equilibrium is formalized as following.

Definition1: An equilibrium, Φ is an assignment of *m* workers to *n* firms together with a wage schedule which is represented as (Φ; $W_{1}^{\varphi(1)}$,$W_{2}^{\varphi(2)}$, …$W_{m}^{\varphi(m)}$), where individual rational allocation φ(i) is denoted as the firm that the i^{th} worker is assigned to. S_{φ}^{j} is the set of worker that hired by firm j at equilibrium (so that S^{j} is the hiring capacity of firm j

which is exogeneous to the model). $\tilde{U}_i$ is the baseline utility of worker i when being unemployed.

$$U_i^{\varphi(i)}(C_i, Z_{\varphi(i)}, W_i^{\varphi(i)}; \alpha_i) \geq \tilde{U}_i \,, \tag{4.10}$$

$$V(S_\varphi^j) - \sum_{i \in S_\varphi^j} W_i^j - c(p;\mu)\, S_\varphi^j \geq 0 \,, \tag{4.11}$$

Definition2. A core allocations is denoted by an individually rational allocation (φ; $W_1^{\varphi(1)}$, $W_2^{\varphi(2)}$, …$W_m^{\varphi(m)}$) such that there is no firm and worker coliation (j, $\tilde{S^j}$) and $\tilde{W_i^j}$ that satisfy

$$U_i^j\left(C_i, Z_j, \tilde{W_i^j}; \alpha_i\right) \geq U_i^{\varphi(i)}(C_i, Z_{\varphi(i)}, W_i^{\varphi(i)}; \alpha_i) \,, \tag{4.12}$$

$$V(\tilde{S^j}) - \sum_{i \in \tilde{S^j}} \tilde{W_i^j} - c(p;\mu)\, \tilde{S^j} \geq V(S_\varphi^j) - \sum_{i \in S_\varphi^j} W_i^j - c(p;\mu)\, S_\varphi^j \,, \tag{4.13}$$

where the notation of $\tilde{S^j}$ and $\tilde{W_i^j}$ is to distinguish from the current set of workers S^j hired by firm j and current wage W_i^j for worker i.

Proposition 1. If worker *i* has at least one offer at time t, he or she always has at least one offer at any time t'> t.

Proof. From Round2 and Round3, workers who have one offer, remain their permitted wage in any period when other workers issue rejection. That is, their competitive advantage will increase, to the point that they will receive at least more than one offer. Same argument is valid for those workers who have more than one offer in previous rounds. They will remain their permitted wage for the offers they did not reject. Consequently, those workers have competitive advantages of those offers which were not rejected. While their number of offers will decrease with the process evolving, they will always have at least one offer at any round t'>t.

Proposition 2. After a finite number of rounds, every worker has exactly one offer and the process stops.

Proof. Suppose worker *i* has no offers at some point in the process. As long as worker *i* has no offers, the wages at which firms are permitted to make offers to him remain constant by R3. On the other hand, as long as worker i remains without an offer, at least one other worker, say, worker *k*, must be issuing rejections. Since rejections by worker *k* cause his permitted salaries to rise for all bidders but one, and at least one other worker's permitted wages must remain constant in any period in which he issues a rejection, worker *k* cannot issue rejections indefinitely without reducing the number of his offers to one, at which point worker k can no longer issue rejections. As this argument is equally valid for any worker who has more than one offer, worker *i*'s competitive disadvantage must fall, in finite time, to the point where he or she gets at least one offer. When every worker who was initially without an offer gets one, the process stops by R4.

Proposition 3. The process converges to an individually rational allocation.

Proof. Since for each worker $\tilde{U}_{i,t=0} = U(\overline{y_i})$ and $\tilde{W}^j_{i,t=0} = U^{-1}(\tilde{U}_i, Z_j, P_j, ; \alpha_i) - \overline{y_i}$ for all (i, j) and the W^j_{it} never fall by R1 and R3. $U_i^{\varphi(i)}(t^*) \geq \tilde{U}_{i,t=0}$ for all (i, j), where t^* is the time at which the process stops.

To prove that $V(S^j) - \sum_{i \epsilon S^j} W_i^j(t^*) - c(p;\mu)\, S^j \geq 0$ if firm j hires set of workers, S^j, when the process converges, suppose to the contrary that $V(S^j) - \sum_{i \epsilon S^j} W_i^j(t^*) - c(p;\mu)\, S^j < 0$. By R3, firm j would not have made offers at wages W_i^j to the set of wokers S^j unless $V(\tilde{S^j}) - \sum_{i \epsilon \tilde{S^j}} W_i^j(t*) - c(p;\mu)\, \tilde{S^j} < 0$ were true for all other sets of $\tilde{S^j}$ as well. Given the assumption that $V(S^j) - \sum_{i \epsilon S^j} W_{i,t=1}^j - c(p;\mu)\, S^j \geq 0$, and from R1 and R3, $V(\tilde{S}^j) - \sum_{i \epsilon \tilde{S}^j} W_i^j(t*) - c(p;\mu)\tilde{S}^j < 0$ implies that firm j was rejected at some salary at least once at some time strictly before t^* by every worker. But this is impossible, since before t^* at least one worker was always without an offer by Proposition 1 and R4; by R2, that worker could not have rejected frim j before t^*. This contradiction completes the proof.

Proposition 3. The salary-adjustment process converges in finite time to a core allocation in the discrete market for which it is defined.

Proof. By definition the process converges to an equilibrium denoted by $(\Phi; W_1^{\varphi(1)}, W_2^{\varphi(2)}, \ldots W_m^{\varphi(m)})$, and S_φ^j is the set of workers assigned to firm j by Φ. Suppose the $(\Phi; W_1^{\varphi(1)}, W_2^{\varphi(2)}, \ldots W_m^{\varphi(m)})$ is not a core allocation. Since by Proposition 3 $(\Phi; W_1^{\varphi(1)}, W_2^{\varphi(2)}, \ldots W_m^{\varphi(m)})$ is individually rational, there must be a coliation (j, $\tilde{S^j}$) and $\widetilde{W_i^j}$ that:

$$U_i^j\left(C_i, Z_j, \widetilde{W_i^j}; \alpha_i\right) > U_i^{\varphi(i)}(C_i, Z_{\varphi(i)}, W_i^{\varphi(i)}(t^*); \quad \text{for all } i \epsilon \tilde{S^j}, \tag{4.14}$$

$$V(\tilde{S^j}) - \sum_{i \epsilon \tilde{S^j}} \widetilde{W_i^j} - c(p;\mu)\, \tilde{S^j} \geq V(S_\varphi^j) - \sum_{i \epsilon S_\varphi^j} W_i^j(t^*) - c(p;\mu)\, S_\varphi^j, \tag{4.15}$$

By R4, workers never lose offers except by rejecting them for ones providing higher utility level. Thus, by equation 4.14, worker i in set $\widetilde{S^j}$ must never have received (and therefore never have rejected) an offer from firm j at $\widetilde{W_i^j}$. In other word, the wage that firm j offers to the set of workers, $\widetilde{S^j}$, should be $\sum_{i \in S_\varphi^j} W_i^j(t^*) \leq \sum_{i \in \widetilde{S^j}} \widetilde{W_i^j}$. Given the assumption that workers' productivities are homogeneous to each firm, then $V(\widetilde{S^j}) - \sum_{i \in \widetilde{S^j}} \widetilde{W_i^j} - c(p;\mu)\,\widetilde{S^j} \leq V(S_\varphi^j) - \sum_{i \in S_\varphi^j} W_i^j(t^*) - c(p;\mu)\,S_\varphi^j$.

The contradiction establishes the proposition.

The sorting algorithm has been demonstrated to solve for a vector of hedonic prices and a matching of workers to jobs that jointly define a hedonic equilibrium. Nevertheless, the convergence of the algorithm and the existence of the hedonic wage equilibrium rely on the assumptions that firms/employers' hiring capacity and the cost of providing safety facility are exogenous to the model, and the productivity of workers is assumed homogeneous to each firm. These assumptions will be discussed more in detail in next section.

4.4. Empirical implication of the algorithm

Employers are differentiated by their workplace fatal risk, non-fatal injury risk, and job attributes related to cognitive skill requirements, physical demands, working conditions, motor skills demands and the degree to which interaction with people is necessary. Workers are heterogeneous in terms of their preferences over these job attributes, as presented by the specific form for the utility function that is chosen.

The hedonic wage equilibrium is computed for Cobb-Douglas utility function, i.e. U_i^j =$(W_i^j + \overline{y_i})Z_j^{\alpha_i}$, which has been employed in Hwang (1998) and Han (2009), given that it is general enough to allow for the heterogeneous valuation of job characteristics. It is also consistent with the intuition of generalization of Roy model which incorporates non-pecuniary job return as utility component (DeLeire, Kahn and Timmins, 2010). Based on this standard multiplicative utility function worker *i*'s expected logarithm utility from consuming job j's attributes and wages, can be represented as:

$$E(\ln U_i^j) = (1-P_j)[\ln(W_i^j + \overline{y_i}) + \alpha_i \ln(Z_j)]. \quad (4.16)$$

where P_j is the fatal risk rate for job j, Z_j is the vector of job attributes (not including fatal risk) for job j, $\overline{y_i}$ is non-labor (exogenous) income, W_i^j is the wage from job j, α_i is a vector of parameters representing the worker's preferences toward job characteristics. It is assumed that worker has no bequest benefit in the state of death. Worker heterogeneity is captured by differences in the value of α. Differences in the value of α meter the differences in attitude among individuals regarding non-pecuniary job attributes. Workers' preference parameters, α, are allowed to be dispersed in both positive and negative domain, as certain job attributes which would appeal to some individuals and be disamenities to others.

4.4.1 Reproducing the market

Given the definition for the two sides of the market and the specification of worker's preference function, the simulation framework can be used to solve for a vector of hedonic wages that reflects wage patterns in an actual market. The algorithm has been demonstrated to solve for a vector of hedonic wages and a matching of workers to jobs that jointly define

equilibrium. Nevertheless, as noted above, the convergence of the algorithm and the existence of the hedonic wage equilibrium heavily rely on the two major assumptions: i) that firm's preferences are separable across different workers, and ii) workers' productivity or skills is homogeneous to the firm.

The assumption of separable preferences is directly taken from Crawford and Knoer (1981). It implies that the payoff from a particular employer-worker pair is independent of the other matches the employer forms. In other words, there is no substitutability or complementarity among worker types in production. Based on this condition, the employer's preferred set of worker from any available set of workers can be deduced using the rankings of individual workers and the number of job vacancies the employer wishes to fill. This assumption is essential to establish the existence of a stable match. It guarantees that no firm ever regrets having to offer employment at subsequent steps of the algorithm to workers who have not rejected its earlier offers. It drives the change of worker's competitive advantage which ensures that each worker will have one offer in a finite time (Roth and Sotomayor, 1990).

The assumption of homogeneity of workers' productivity, however, is the major difference between this model and Crawford and Knoer's framework. In their setting, each worker is assigned a monetary value (production) and they are ranked by their productivity values net of salaries. In this study, due to the data constrain, the information of workers' productivity or skill in particular jobs are not available. While productivity or production information is available on the employers' database, they are aggregated on the firm level

which cannot be matched back to individual worker's level in the labor force database. While the assumption of homogeneous productivity is imposed to alleviate the data issue, it needs to be applied with caution.

As noted in Freeman (2003), employers are actually choosing from a set of workers of different characteristics and cannot be indifferent to the productive characteristics of employees. Simply imposing the assumption of workers' homogeneous productivity eliminates the mechanism for ensuring a distinguishing feature of hedonic labor market. Moreover, since jobs vary with specific skill requirements on an industry level or occupation level, it is problematic to assume that all the workers are necessarily viable candidates for every job. For example, a miner should probably be treated as a prefect substitute for a teacher and an IT technician is generally not qualified for a management position. Therefore, simply running the sorting model based on the assumption of homogeneous productivity will gives us a misleading outcome.

To address this issue, I segment the labor market into different job clusters. In empirical practice, labor market have been segmented on the basis of geography, education, skill requirements or into different occupational groups which are believed a priori grounds to share similar skill and productivity requirement (Freeman, 2003). There are studies that divide labor market into various submarkets by workers' risk aversion level and firms' safety equipment efficiency level (Kniesner and Leeth, 1988, Kahn and Lang, 1988, and Mongemory, 2002). I propose to segment the market in two ways. One is to segment the

labor market into occupation-industry clusters and then apply the sorting algorithm within each job cluster instead of the whole market. This method is based on the idea that the division of the occupational labor market is formed by differentiation and specialization of the jobs (Smith, 1983; Kalleberg, 2003). As described in McConnel (2006), the worker's productivity in the given job matters (e.g. ability education and / or skill). While some workers will not be eligible for certain jobs for lack of appropriate skills or training, others may be eligible but less qualified and will be hired only at different wage risk combinations. For instance, surgeons face higher wages and lower risks than manufacturing worker, and differences in skills prevent arbitrage. Based on this concept, I hypothesize that workers in the similar occupations present similar skills or productivity due to different skills and investment in training and qualifications.

In the SIPP labor force data, the occupation code is from 2000 Standard Occupational Classification (SOC) System in BLS and the industry code is from North American Industry Classification System (NAICS). The SOC and NAICS classification system allow me to aggregate the occupations and industries into 6 major occupational classifications and 9 major industrial classifications as defined in Scotton (2000, 2010). I apply the sorting algorithm within each of these 40 occupation-industry clusters (excluding agricultural industry and military occupation). The industry group and occupation group classifications are presented in the Appendix 4.A and the Appendix 4.B.

Another way to increase the realism of the homogeneity assumption is to create job clusters that further group observed workers into different skill levels within each

occupation-industry cluster. I create a "skill level" as a composite, S_i, based on three observable characteristics affecting skill as developed by Portela (2001):

$$S_i = \text{mschool} \times \left(0.5 + \frac{e^{\frac{school_i - mschool}{sschool}}}{1 + e^{\frac{school_i - mschool}{sschool}}}\right) \times \left(0.5 + \frac{e^{\frac{exper_i - mexper|_{school_i}}{sexper|_{school_i}}}}{1 + e^{\frac{exper_i - mexper|_{school_i}}{sexper|_{school_i}}}}\right), \quad (4.17)$$

where $school_i$ is for the schooling level of worker i, evaluated as years of schooling, *mschool* represents the average level of schooling in the population, and *sschool* is the standard deviation of schooling. The first argument of the equation, $\text{mschool} \times \left(0.5 + \frac{e^{\frac{school_i - mschool}{sschool}}}{1 + e^{\frac{school_i - mschool}{sschool}}}\right)$ represents the base skill of individual i and it is obtained by standardizing the relative position of the worker in the schooling distribution. The correction factor, $0.5 + \frac{e^{\frac{school_i - mschool}{sschool}}}{1 + e^{\frac{school_i - mschool}{sschool}}}$ takes account of the actual position of the individual in the schooling distribution. It takes values between 0.5 and 1.5. Values above 1 is reached when the schooling level of the worker is above the average in the economy, while levels of schooling below the average are associated with a correction factor of less than 1.

The second argument of the equation is to correct incorporable experience. *experi* is the experience of the individual, $mexper|_{school_i}$ is the average experience within schooling level *schooli*, $sexper|_{school_i}$ is the standard deviation. When the experience is below average, the coefficient to multiply for the base skill will be less than 1, representing a penalty on the ability of the worker. When the experience is greater than the average, the

coefficient will be higher than 1, and it will increase the initial measure of skill associated with the schooling.

The experience correction factor is intended to reflect the fact that individuals with identical schooling levels can have different skills, and that part of this difference can be captured by their experience. Therefore, for individuals with the same schooling level, the skill is higher the greater the positive deviance of experience from the mean. However, a higher dispersion of the experience for a group of workers with a certain schooling level, implies that the relevance of this difference will be reduced, which is consistent with common situation. The higher the deviance is from the mean experience, the lower the impact of the skill is. Therefore, the working experience is considered as a complement to schooling in the model. The impact of experience on the ability of a worker depends on his/her schooling. The same length of experience in the labor market has a different impact on the productivity skill, depending on the individual's schooling.[25]

By using information on schooling and working experience from the SIPP data, Figure 4.1 depicts the function $y= \frac{e^{\frac{x-mschool}{sschool}}}{1+e^{\frac{x-mschool}{sschool}}}$, where x is $school_i$.[26] It is consistent

[25] The approach of dealing with worker's heterogeneous skill is consistent with a Leontief-type production over the heterogeneous components of skill. In principle, this approach could also be extended to housing market applications to distinguish between buyer types such as renters versus outright buyers versus people taking out loans to buy.

[26] They are collected from the survey questions: *education attainment* and *how long have you been in the same line of work.* The SIPP data dictionary is available at www.census.gov/sipp/diction.html

with the intuition that for individuals who have low education, their skill level increases rapidly with schooling. But for those who have high education, the marginal increase in the schooling has a smaller impact on the increase in working skill level. Experience shows a similar relationship.

Applying the algorithm in the job clusters is expected to provide a more intuitively reasonable context for the assumption of worker homogeneity in productivity and reduce the restrictiveness of the assumption. Aggregating the jobs based on the major occupation and industry categories within classification system and the skill levels presented by observed workers in the labor force provides 132 job clusters in total.[27] The skill is ranked into four levels with percentile 25%, 50%, 75% and 99%. Table 4.1 summarizes the skill level for each of 6 major occupations and 9 major industries. Panel A shows that the workers in the education services, health and social assistant industry appear higher skilled, while workers in art, entertainment, recreational, mining and utility industries are lower skilled. Panel B presents that management, professional and related occupations are mostly highly skilled and service occupations appear to be lower skilled occupations.

4.4.2 Panel feature of the data

In order to generate a hypothetical market where panel study approaches can be applied, I simulate the job over time so that the panel feature can be captured. I take advantage of the panel feature of SIPP data in my sorting algorithm. In 2004, the baseline utility level for each worker is computed by using their non-labor income as described in the

[27] 84 job clusters defined by occupation-industry cell and skill levels do not contain any observations.

last section. After applying the sorting algorithm with the observed labor market data in 2004, the resulting utility level that each worker gets from his/her job is used as the baseline utility level for the sorting algorithm in 2005. In 2005, workers are allowed to resort based on observed jobs available in the real data and resulting utility level from 2004 simulation results. Note that any new wage/job combination does not decrease the utility as compared to the baseline utility in 2004. The same practice will be carried out for 2006. In other words, worker' preference parameter, α, and the utility produced from simulation in the previous year are used as the starting values of α and baseline utility for the simulation in 2005 and 2006. The jobs that workers sort into each year are based on the real labor force data in the SIPP. In the real data, same workers are observed in different jobs over the years. These different sets of jobs induce workers to resort in the simulation for different years.

The advantage of this approach is that it is possible to keep track of the workers job changing behavior in the simulated context. Also, instead of choosing between being unemployed or employed as modeled in 2004, it is more realistic to consider the workers' job sorting behavior from an existing job. The sorting result in 2006 incorporates information about workers, risk rates and job attributes across year 2004 to 2006, but is not a fully dynamic model. Nonetheless, the process better reflects the dynamic feature of the labor market than sorting based on 2006 data alone.

4.5 *Computational detail and Results*

4.5.1 Computational detail

The simulation results provide an assignment of workers to jobs and a vector of wages that defines a hedonic labor market equilibrium. Recall that worker's wage is determined as $W_{i,t}^{j}=U^{-1}(Z_k, P_k, W_{i,t-1}^{k}, Z_j, P_j; \alpha_i)$ -$\overline{y_i}$+ ς and the initial minimum wage is $\widetilde{W}_{i,t=0}^{j}=U^{-1}(\widetilde{U}_i, Z_j, P_j; \alpha_i)$-$\overline{y_i}$.[28] Given that the utility function, vector of Z, fatal risk P, non-labor income $\overline{y_i}$ is known and the $\widetilde{U}_i$ is calculated as a function of non-labor income, wages can be calculated during the sorting process and they are calibrated to approximately reproduce the pattern of wages observed in SIPP labor force data.

As mentioned before, the bidding rules in the sorting algorithm do not fully account for the wage determinant factors which are not observed from data. However, calibrating worker's preferences allows to adding a dimension of wage differences between workers with different observed and unobserved characteristics. In the model, the wage disparities between workers depends on worker's baseline utility, heterogeneous preferences toward various job attributes and the compensation to the job attributes that firm offers. While the model still ignores the potential patterns of correlation between wage difference and firm and worker's strategic behavior, the computed hedonic wage equilibrium does present both layer of wage disparities which can be explained by observed data and the layer of wage disparities which are unexplainable by the observed data. Therefore, the calibration helps to reduce the

[28] ς >0 and is a marginal increment to ensure the new wage offer increases worker i's utility.

unrealism of the simulated data set and increase the accuracy of empirical model evaluation that will be carried out in the following chapter.

This calibration was done by using the Nelder-Mead algorithm to solve for the 16x1 vector of α that minimizes the distance between the predicted and observed distributions of worker' wage. Each parameter α is drawn independently from 16 normal distributions with different mean and variance.[29] This is different from the approach in Kuminoff et al (2009) where the preference parameters are initially drawn from the same gamma distribution with same shape and scalar parameters. While drawing from different distribution may greatly adding computational time, assuming different normal distribution of each preference parameter allows the influences from various job characteristics to be more flexible. Using normal distribution helps to capture the aspect that the preferences are subjective. Some job attributes would appeal to some individuals, but not to others. The normal distribution allows for parameter to be dispersed within the positive and negative domain.

The parameters of the initial distributions for drawing the vector of α in the model are determined by using a random utility model (RUM). The idea of using a RUM model to determine the starting values for the preference parameters in the simulation is also similar to Klaiber and Smith (2013). As a discrete choice model, RUM has typically used to value product attributes. It views the individual as choosing the product that gives the highest utility out of all the products in a choice set with utility being a function of product attributes. In labor market setting, RUM has been used to examine worker sorting across occupations in response to the risk of death on the job (DeLeire and Levey, 2007). The resultant estimates

[29] Actually, there are 32 parameters which include mean and std. for each of 16 distributions.

reflect the weights on different job characteristics in determining occupation choices. In my application, I group sample of workers into 7 major industry categories (excluding agriculture and military force), and estimate RUM model within each of the industry categories to obtain the vectors of weights for job characteristics separately. And the estimated weights are used as the starting value for drawing the vector of α in the algorithm.

Specifically, we assume a random utility model in which individuals choose from a variety of occupations. The utility an individual derives from a particular occupation depends on individuals characteristics, the wage he or she can receive on the job, and the characteristics of the job:

$$U_{ij}=U(X_i, W_{ij}, Z_j), \tag{4.18}$$

where *i* indexes individuals and *j* indexes occupations. The wage an individual receives in occupation *j* is a function of the same (or a subset of) individual (X_i) and job (Z_i) characteristics as in equation 4.19:

$$W_{ij}=f(X_i, Z_j), \tag{4.19}$$

Substituting the wage equation into equation 4.18, assuming a linear function form, and adding an independently and identically distributed with type I extreme value distribution disturbance term yields:

$$U_{ij}=\beta X_i + \alpha Z_j + \varepsilon_{ij}. \tag{4.20}$$

An individual will choose among J occupations the one that yields the highest utility, such as:

$$U_{ij}^* > U_{ik}^*, \quad k\neq j. \tag{4.21}$$

Define U_{ij}=1 if individual *i* chooses occupation j and U_{ij}=0 otherwise. Given the assumption on the distribution of the error term, we can estimate the parameters of the random utility model by McFadden's conditional logit (for a description, see Maddala 1983):

$$\text{Prob}\,(U_{ij}=1)=\frac{\exp\{\beta X_i+\alpha Z_j\}}{\sum_{j=1}^{J}\exp\{\beta X_i+\alpha Z_j\}}\,. \qquad (4.22)$$

In the model, X_i is a vector of variables describing worker i including:

hr_wage	hourly income
age	age in years
blacknh	1 if individual is black and non-Hispanic
college	1 if individual attended college but with no degree
female	1 if individual is female
hispanic	1 if individual has a Hispanic origin
hsgrad	1 if individual graduated from high school
kid18	1 if individual has any kids under 18 years old
married	1 if individual is married
othrace	1 if individual is non-white, non-black, non-Hispanic
ugdeg	1 if individual have bachelor degree or more

RUM is carried out with worker's characteristics, X_i, a full set of O*NET occupational characteristics plus fatal and non-fatal risk. Note that β cannot be estimated because βX_i will drop out of equation 4.22.The vector of parameters α reflects the weights on different job characteristics Z in determining occupation choices.

Table 4.2 reports the statistics of the estimates from the RUM model. The table contains parameter estimates from logit models estimated separately for seven industry categories defined by 2 digit NAICS coding system. The coefficient on fatal death risk is negative for all the industries. The coefficient on nonfatal injury risk is positive for mining,

construction and manufacturing industries and other services, but they are negative for all other industries.[30] While this coefficient might have been expected to be negative in all industries, the sign of the estimates for mining, construction and manufacturing industries and other services, could be a result of a "job requirements" situation where workers choose to be in risky occupations because risky occupations are also ones requiring, for example physical strength or outdoor working environment which would appeal to those workers.

In addition to Table 4.2, Figure 4.2 uses histograms more visually comparing the distributions of parameters across seven different industries. Recall that the 15 job attributes can be categorized into five groups: 1) cognitive skills; 2) working condition; 3) physical demand; 4) motor skill; 5) interpersonal skill. In figure 4.2, each histogram presents the distributions of parameters one group of job attributes across industries. Five different histograms are created to present the comparison separately.

The general patterns of estimated coefficients across the industries are consistent with the intuition hypothesis of worker sorting on the basis of preferences. According to the occupation and industry coding in the data, in mining, utility, construction industries, most of workers are in production, transportation, and material moving or maintenance occupations. They tend to choose job requires hand handling, tools using and the jobs with group working environment. In information, scientific technical service or financial industries, most of the

[30] Other services (except Public Adminstration) sector comprises establishments engaged in providing services not specifically provided for elsewhere in the classification system. Establishments in this sector are primarily engaged in activities, such equipment and machinery repairing, providing drycleaning and laundry services, personal care services, death care services, pet care service, photofinishing services. More about the Other Services are available on www.bls.gov/iag/tgs/iag81.htm#about

workers are in office related occupations. Those workers appear to be in the jobs which involve monitoring process, inspecting, and interacting with computer. In education, health care or social service industries, workers are professional and office related occupations. The coefficients suggest that they prefer the jobs associated with cognitive skill such as data analyzing, monitoring process and establishing interpersonal relationship, performing work directly with public. In retail industries, most of the workers are in sales occupations. The coefficients indicate that they are more appealed to the job required direct interacting with the public.

The approach of calibration allows adding a dimension of wage differences between workers with different observed and unobserved characteristics. In particular, distinguishing workers' preferences across different occupation-industry categories helps to account for patterns of wage disparities over different major industries and occupations which have been recognized in the literature. However, the calibration on workers' preference still couldn't fully capture the wage determinant factors which are not observed from data. The sources underlies the wage disparities among workers are attributed to worker's baseline utility, heterogeneous preferences toward various job attributes and the compensation to the job attributes that firm offers. The potential correlation between workers' wage differential and firms' profitability is still missing in the model. This problem may affect the evaluation which will be carried out in chapter 5, where the impact of occupation/ industry fixed effect control might not be fully detected.

4.5.2 Results of simulation

To evaluate the performance of the algorithm, I compare the simulated wages to the actual ones observed from the SIPP data for 2004, 2005 and 2006. Figure 4.3 and Figure 4.4 contrast the distributions of simulated wages and actual wages. Figure 4.3 presents the results from the model aggregating jobs into 40 clusters, while Figure 4.4 shows the results from the model aggregating jobs into 130 job clusters.[31] In both figures, the actual wages are ranked from low to high. The red solid line represents the distribution of actual wages. The blue dashed line represents the equilibrium wage for those workers from the matching in the simulation. Panel A, Panel C and Panel E in both of the figures compare the cumulative distribution function of simulated equilibrium wages and actual wages for 2004, 2005, and 2006. Differences between actual wages and simulated wages reflect the magnitudes of the positive and negative prediction errors. These prediction errors arise from the ways in which the sorting process is abostract from the reality, including: i) using Cobb-Douglas utility function form, ii) lack of data on employer's cost associated with safety providing and worker's productivity that may affect the wages, and iii) the work-job matching lacks the dimension of employer's profitability and worker's bargain power that may also influence the wages. However, Panel A, C, E still illustrate that the simulation model clearly reproduces the general wage trends in 2004, 2005 and 2006.

This match can be further illustrated by the corresponding comparison of probability density functions of natural log of weekly wages presented in Panel B, D and F in Figure 4.3

[31] 40 job clusters are produced by aggregating jobs by the pairs of occupation and industry. 130 job clusters are produced by aggregating jobs by occupation and industry categories within different workers' skill levels.

and Figure 4.4. These panels present the comparison of percentage of the total workers distributed across different wages. In both of Figure 4.3 and Figure 4.4, Panel B, D and F show that comparing to the distribution in the real data, there are fewer workers at the lower end of wage from the simulation. It could be caused by using worker's non-labor income as reference utility to start the algorithm. A high reference utility makes it impossible for the algorithm to converge to a very low simulated income.

In Figure 4.3, most of the differences between simulated data and the real data are presented in Panel B and F which are the simulated results from 2004 and 2006. In Figure 3, Panel B, more workers are distributed at median higher wage in the simulation. In Figure 4.3, Panel F, fewer workers are distributed at the median wage level in the simulation compared to the ones in real data.

In Figure 4.4, most of the differences between simulated data and the real data are presented in Panel B and Panel D, where more workers are distributed at median higher wage level in the simulation compared to the ones in the real data. While some predicted wages still differ considerably from the actual values as showed in the figures and the workers' job assignments are not the same as the observed jobs that workers fill in, the simulation results produced by both job clustering approaches reflect the general wage trend in the actual data and a reasonable approximation to the actual labor market.

Table 4.3 provides the statistics to compare the wages generated from simulation models with the wage sets observed in the real data. Panel A and panel B compare the statistics from models using different job clustering approaches separately. Notice that the

size of samples is reduced when the jobs are clustered by broad occupation-industry categories and the skill levels. The reason is that further segmenting the market by workers' skill level produces more job clusters with fewer workers in each cluster. When the number of workers in the job cluster is less than 5, the sorting algorithm will fail to carry out due to the insufficient rounds of offer and counteroffer between workers and the employers. Therefore, those job clusters are dropped in the analysis, which causes a loss of the 40 samples of workers. It also results in different summary statistics of actual wages in the models using different job clustering methods.

Based on the simulated results from clustering jobs by occupation and industry, Panel A indicates that the simulated wages tend to be higher than the real wages at the 25, 50 and 75 percentiles for 2004, and the simulated wages are lower than the real wages at the 50 percentile for 2006. In Panel B, it shows that when clustering jobs by occupation, industry and skill levels, the simulated wages are higher than the real ones at the 25 and 50 percentile for both year of 2004 and 2005.

In sum, the simulation results demonstrate that the proposed sorting algorithm is able to reproduce the wage trend in the real labor market. Simulating a hedonic labor market equilibrium based on the actual market trends helps to add realism to the subsequent analysis of functional form. Based on this study, the simulated hedonic labor market can be utilized to

construct hypothetical scenarios that allow us to address economic issues that cannot currently be investigated econometrically.

Table 4.1

Distribution of skilled workers within industry and occupation [a]

Panel A. Percentage of skilled workers within industry					
Ind. categories	Skill Level				Total #of workers [c]
	0-25% [b]	25%-50%	50%-75%	75%-100%	
Mining and utility	29.35	20.04	28.74	21.86	494
Manufacturing	24.79	23.38	29.33	22.61	1,190
WRT [1]	27.56	31.82	23.84	16.77	1,103
FIRE [2]	23.89	23.30	30.29	22.52	1,017
Educ./Health [3]	16.59	19.63	24.46	39.32	1,447
AERF [4]	42.68	27.24	18.70	11.38	246
Other	22.91	35.20	21.79	20.11	179
Panel B. Percentage of skilled workers within occupation					
Occ. categories	Skill Level				Total #of workers
	0-25%	25%-50%	50%-75%	75%-100%	
Management, Professional	8.87	16.60	32.80	41.72	2,018
Service	50.92	32.44	13.96	2.67	487
Sales and related	25.71	29.03	26.44	18.82	1,509
NRCM [5]	29.14	24.36	27.07	19.43	628
PTM [6]	36.27	28.43	19.54	15.76	1,034

a. The value in the cell is the percentage of the total workers within each industry or occupation.
b. The skill levels are divided by four inter-quantile ranges of skill index.
c. Total number is the total over 2004-2006
1. WRT = wholesale trade, retail, transportation industries.
2. FIRE = finance and insurance, real estate professional, scientific, and technical services industries
3. Educ./Health = education service, health care and social assistance industries
4. AERF = arts, entertainment and recreation, accommodation and food services industries.
5. NRCM = natural resources, construction, and maintenance occupations.
6. PTM = production, transportation and material moving occupations.

Table 4.2

Descriptive statistics of job characteristics from Random Utility model [a]

	Ind2[1]	ind3[2]	Ind4[3]	Ind5[4]	Ind6[5]	Ind7[6]	Ind8[7]
	Coeff. (S.E.)	Coeff. (S.E.)	Coeff. (S.E.)	Coeff. (S.E.)	Coeff. (S.E.)	Coeff. (S.E.)	Coeff. (S.E.)
injury	0.83 (-0.66)	0.048 (0.03)	-0.02 (0.03)	-0.3 (0.10)	-0.21 (0.07)	-0.68 (0.41)	0.19 (0.10)
worcon_f1	-0.02 (-0.08)	-0.08 (0.01)	-0.42 (0.25)	0.18 (0.30)	0.38 (0.20)	0.64 (1.46)	-0.11 (0.30)
worcon_f2	0.51 (1.15)	-0.31 (0.08)	0.46 (0.25)	-0.26 (0.25)	-0.37 (0.33)	-0.10 (0.69)	-0.62 (0.25)
worcon_f3	-0.42 (0.05)	-0.24 (0.03)	-0.24 (0.38)	-0.86 (0.37)	0.73 (0.16)	1.34 (0.34)	-1.30 (0.37)
scmplx_f1	0.03 (-0.92)	-0.50 (-0.06)	0.67 (0.40)	0.22 (0.31)	-1.08 (0.33)	1.11 (12.22)	1.33 (0.31)
scmplx_f2	0.15 (1.07)	0.02 (0.01)	-0.31 (0.35)	-0.47 (0.22)	0.82 (0.28)	0.26 (0.93)	-0.10 (0.22)
scmplx_f3	0.41 (0.74)	0.63 (-0.87)	-0.71 (0.31)	0.29 (0.26)	-0.41 (0.30)	-1.18 (0.34)	-1.47 (0.26)
phydds_f1	0.02 (-0.42)	-0.11 (0.05)	0.51 (0.17)	-0.04 (0.22)	1.58 (0.19)	0.25 (0.53)	0.63 (0.22)
phydds_f2	0.62 (0.44)	0.25 (0.18)	0.16 (0.38)	-0.24 (0.61)	-0.16 (0.38)	0.22 (0.87)	2.76 (0.61)
phydds_f3	-1.24 (-0.07)	-0.31 (-0.02)	0.318 (0.27)	0.37 (0.29)	-1.06 (0.28)	-0.92 (1.09)	-0.43 (0.29)
mskill_f1	0.61 (0.02)	0.21 (-0.13)	-0.12 (0.32)	-0.40 (0.34)	-0.34 (0.24)	0.11 (2.9)	0.26 (0.34)
mskill_f2	-0.57 (-1.14)	0.10 (0.01)	-0.24 (0.18)	0.31 (0.21)	0.81 (0.15)	-0.32 (0.98)	-0.35 (0.21)
intpeople_f1	-0.17 (-0.50)	-0.42 (-0.39)	0.49 (0.32)	0.81 (0.32)	1.24 (0.25)	-0.61 (1.21)	-0.07 (0.32)
intpeople_f2	1.45 (-0.5)	0.37 (0.29)	-0.8 (0.30)	0.13 (0.26)	0.24 (0.21)	0.10 (0.87)	0.23 (0.26)
intpeople_f3	0.39 (1.05)	0.30 (0.01)	0.14 (0.22)	-0.30 (0.21)	-0.66 (0.25)	0.34 (2.06)	0.52 (0.21)
intpeople_f4	-0.82 (-0.34)	-0.22 (-0.35)	0.97 (0.23)	0.49 (0.19)	-0.75 (0.19)	-0.39 (2.05)	0.44 (0.19)

Table 4.2 Continued

a. The estimated coefficients are reported for the non-fatal risk job attributes which are actually applied in the simulation.

1. Ind2= mining and utility
2. Ind3= manufacturing
3. Ind4= Wholesale trade, Retail, transportation, warehouse and couriers
4. Ind5= Information, Finance and Insurance, Real Estate Professional, Scientific, and Technical Services
5. Ind6= Education service, health Care and Social Assistance.
6. Ind7= Arts, Entertainment, and Recreation, Accommodation and Food Services
7. Ind8= Other Services (except Public Administration and private households)

Table 4.3

Statistical summary of true and simulated wages of sample workers for year 2004-2006 [a]

Clustering job by 6 occupation within 9 industries *(N=1,892)*						
	2004		2005		2006	
Stat. Summary	True wage	Simulated wage	True wage	Simulated wage	True wage	Simulated wage
Mean	862.74	830.87	798.48	751.48	820.64	764.03
Std. dev.	516.57	464.81	409.14	452.92	424.82	475.97
25th	517.05	527.57	479.23	443.04	491.19	430.83
50th	772.17	778.04	707.31	666.22	724.49	666.24
75th	1,077.59	1,123.34	1,020.81	1,018.31	1,047.16	1,065.16
95th	1,774.82	1,575.24	1,652.74	1,552.66	1,705.52	1,608.39
Clustering job by 6 occupation within 9 industries and 3 skill levels *(N=1,851)*						
	2004		2005		2006	
Stat. Summary	True wage	Simulated wage	True wage	Simulated wage	True wage	Simulated wage
Mean	826.97	839.57	761.59	758.67	853.39	838.0
Std. dev.	520.53	397.03	400.95	399.21	430.84	447.67
25th	516.21	533.95	466.35	460.41	522.62	499.94
50th	762.17	778.22	654.95	661.55	769.64	734.52
75th	1,071.32	1,146.26	952.36	972.05	1,062.01	1,088.79
95th	1,744.65	1,531.89	1,607.50	1,490.39	1,723.46	1,698.73

a. The value of wages is measured as weekly income (in 2005 $value)

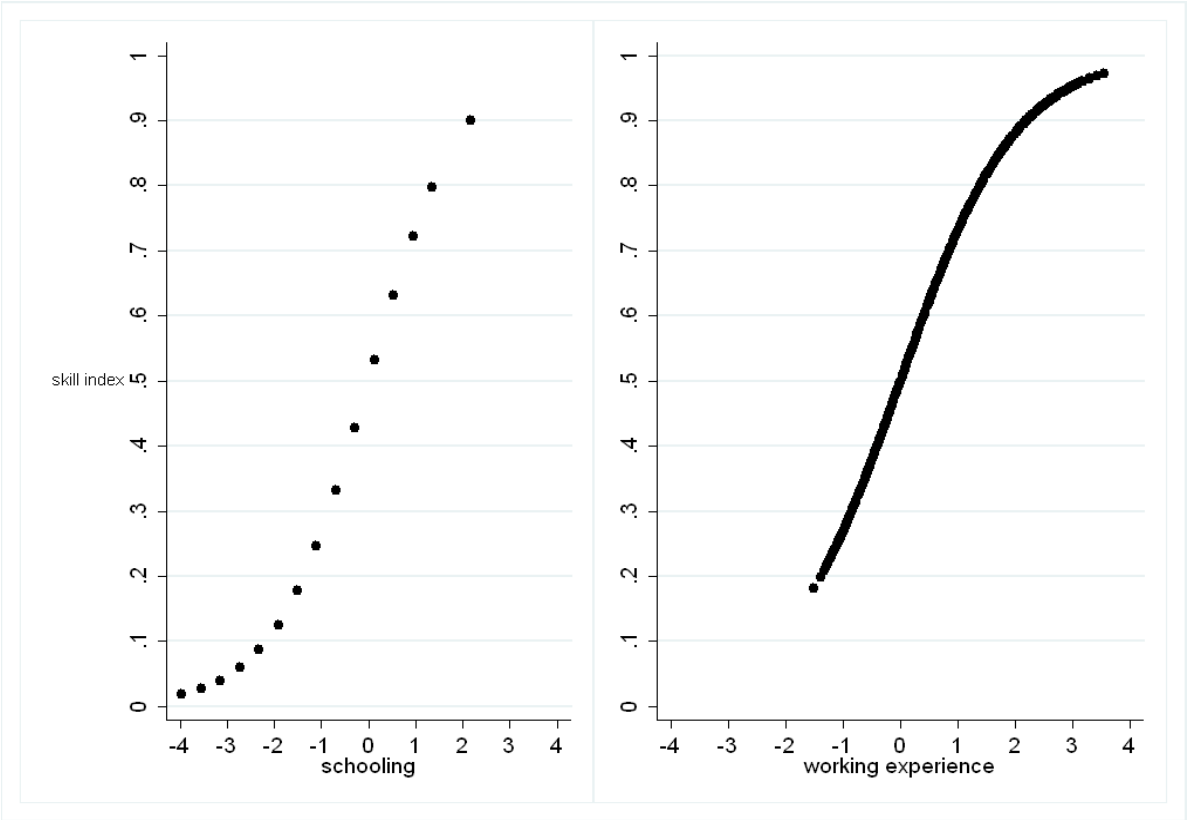


Figure 4.1 *Logistic cumulative ditribution of skill index over schooling and working experience* [a]

a. The value of schooling and working experience on x-axis is normalized

Figure 4.2. *Histogram plot of preference parameters estimated from Random utility model* [a]

a. The random utility model is adopted to estimate the starting value of preference parameters for sorting algorithm. Therefore the random utility model is only applied to 2004 data. The estimated coefficients are reported for non-fatal risk job attributes which are the ones actually applied in the simulation.
b. On the x-axis, the type of job attributes are indicated as :
   injury= Non-fatal injury risk rates;
   exp.1= low level of hazardous work environment;
   exp.2= working environment exposed to the weather or temperature;
   exp.3= high level of hazardous working environment;
   cognitive1= cognitive demands involving information processing and analyzing data;
   cognitive2= involving creative thinking, strategies developing, planning, and decision making;
   cognitive3= cognitive demands involving monitoring process, inspecting equipment or material;
   physical1= physical demands requiring physical related activities;
   physical2= requires climbing ladders, poles, or kneeling and crawling;
   physical3= requires repetitive physical motion, operating machine or handling tools;
   motor1= specifying technical or electronic or mechanical device or equipment;
   motor2= interacting with computer;
   people1= requires interacting with people indirectly;
   people2=interacting with people directly, or involving in the conflicting situations;
   people3= job requires group or team working and taking responsibility for the outcome of the group performances;
   people4= working directly with the public
c. The 7 major industry categories (excluding agriculture and army force) are indicated in 7 different colors, which are:
   ind1= mining and utilities;
   ind2= manufacturing;
   ind3= Wholesale trade, Retail, transportation, warehouse and couriers;
   ind4= Information, Finance and Insurance, Real Estate Professional, Scientific, and Technical Services;
   ind5= Education service, health Care and Social Assistance;
   ind6= Arts, Entertainment, and Recreation, Accommodation and Food Services;
   ind7= Other Services (except Public Administration and private households)

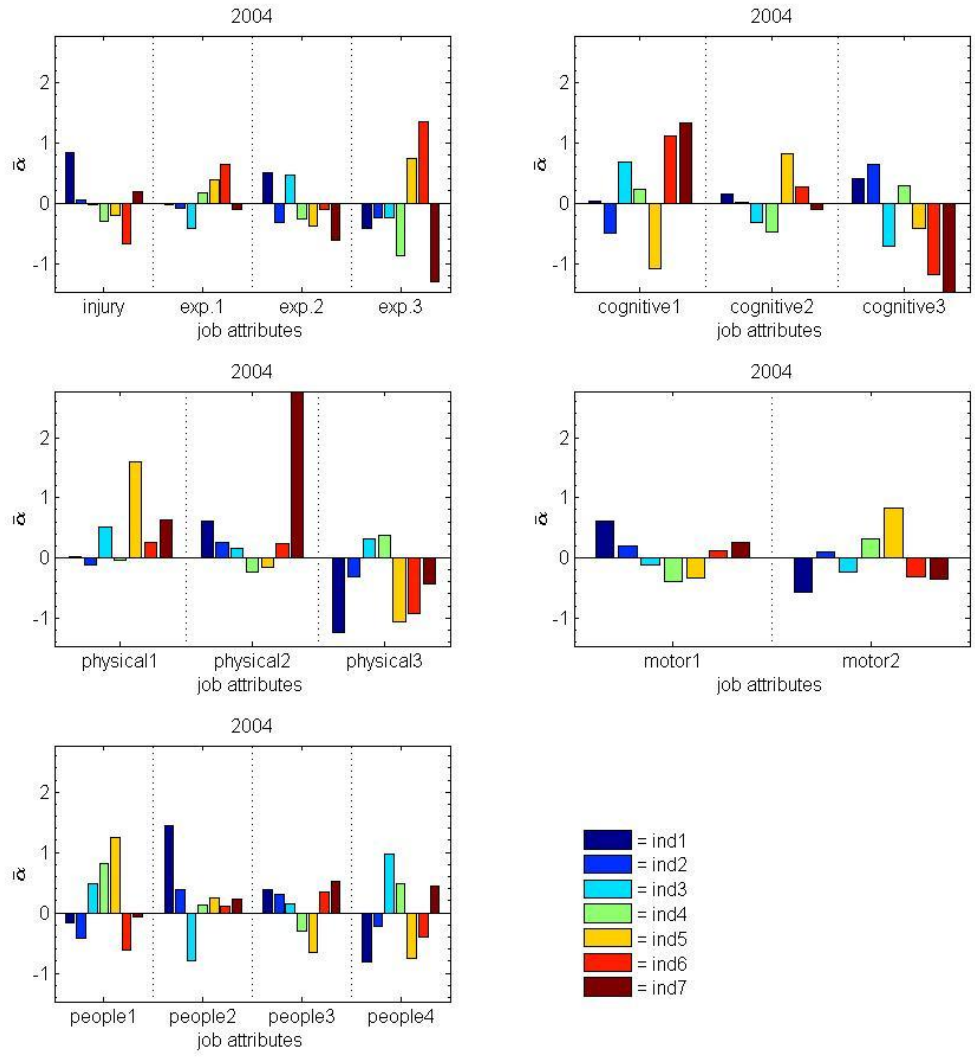
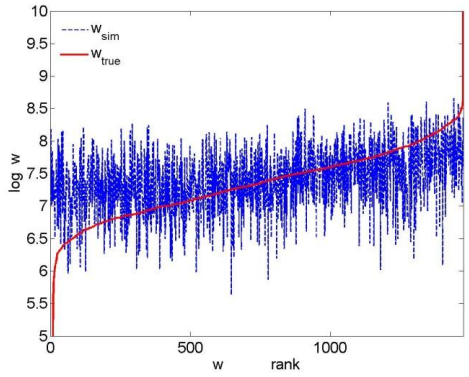
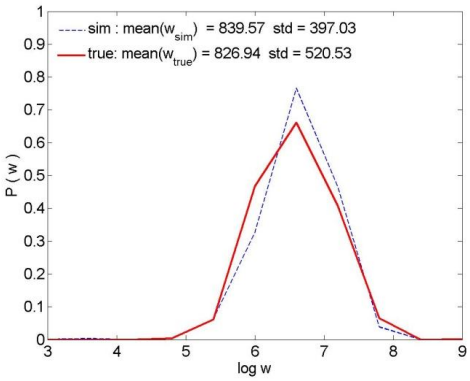

2004
injury
exp.1
exp.2
exp.3
job attributes
2004
cognitive1
cognitive2
cognitive3
job attributes
2004
physical1
physical2
physical3
job attributes
2004
motor1
motor2
job attributes
2004
people1
people2
people3
people4
job attributes
= ind1
= ind2
= ind3
= ind4
= ind5
= ind6
= ind7

Figure 4.3. *The difference between actual wage and simulated wage distribtuion from year 2004 to year 2006(segmenting the market by grouping occupation and industry)* [a]

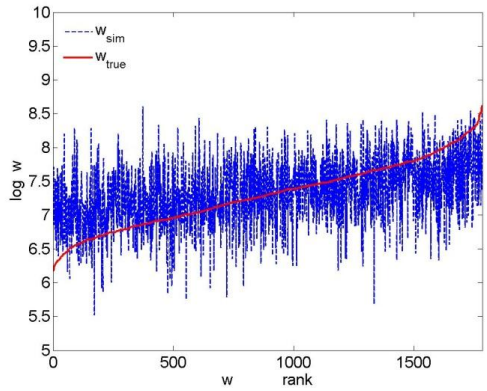
a. The legend shows the value of mean and standard deviation of the weekly wages for the both observed and simulated wages (in 2005 $value).



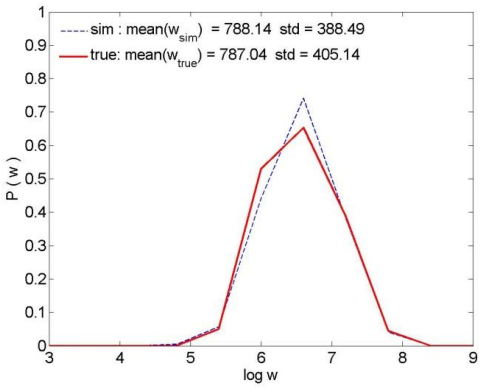
A. Derivative from CDF (year 2004) [a]



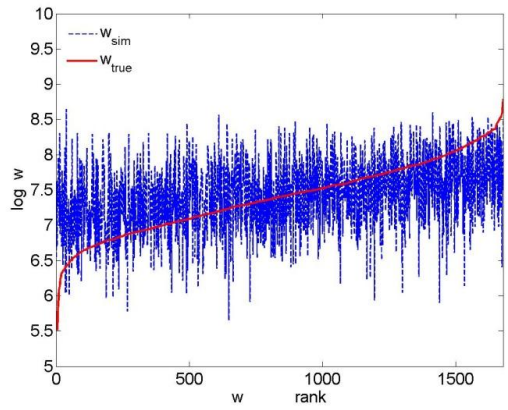
B. Derivative from PDF (year 2004)



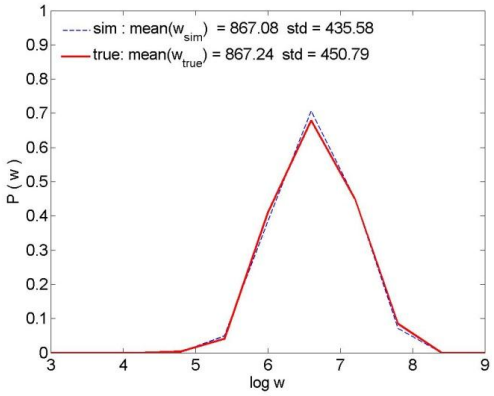
C. Derivative from CDF (year 2005)



D. Derivative from PDF (year 2005)



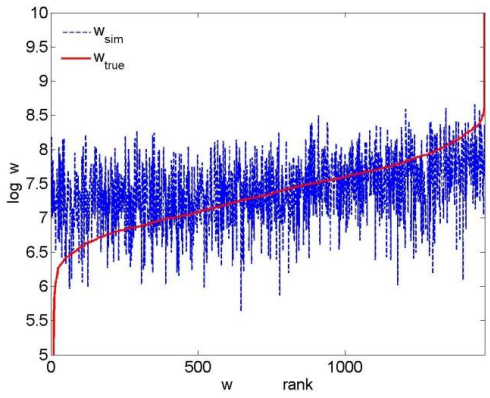
E. Derivative from CDF (year 2006)



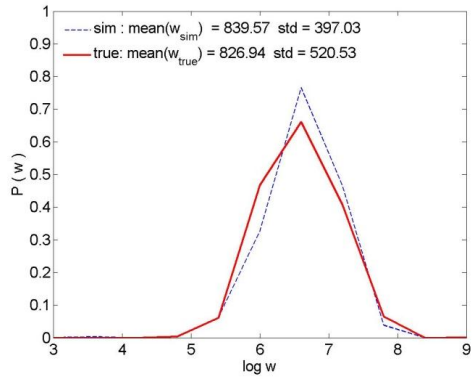
F. Derivative from PDF (year 2006)

Figure 4.4 *The difference between actual wage and simulated wage distribtuion from year 2004 to year 2006(segmenting the market by grouping skill, occupation and industry)* [a]

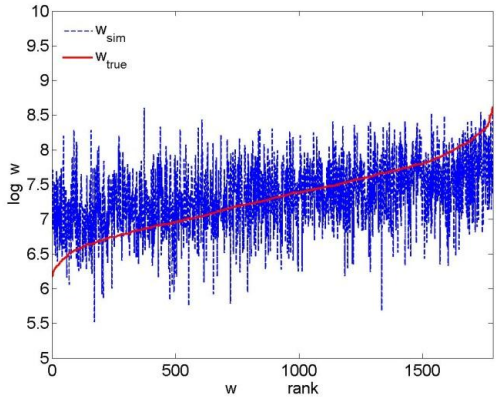
a. The legend shows the value of mean and standard deviation of the weekly wages for the both observed and simulated wages (in 2005 $value).



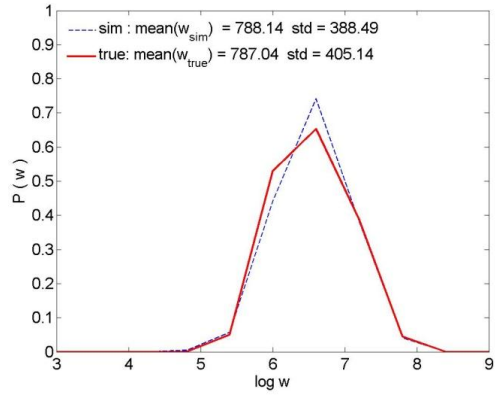
A. Derivative from CDF (year 2004) [a]



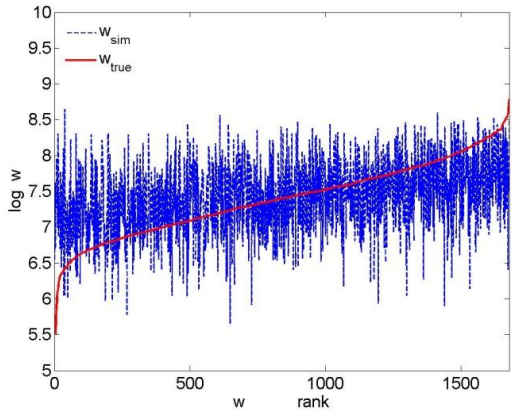
B. Derivative from PDF (year 2004)



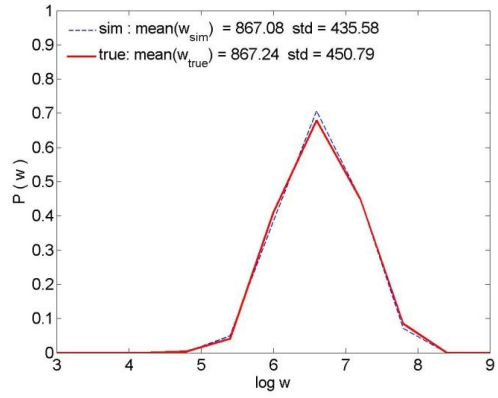
C. Derivative from CDF (year 2005)



D. Derivative from PDF (year 2005)



E. Derivative from CDF (year 2006)



F. Derivative from PDF (year 2006)

CHAPTER 5

EVALUATING HEDONIC WAGE MODEL SPECIFICATIONS

5.1 Introduction

In the previous chapter, I used a simulation model and the SIPP labor force data to simulate a hedonic equilibrium in which workers maximize utility as they reached in job-worker match. The results give me utility parameters, with which I can compute each worker's true MWTP to reduce risk by one unit given worker's non-labor income, non-pecuniary job attributes and heterogeneous preferences over job attributes.

In this chapter, I compute the true MWTP for marginal change in fatal risk for each worker in the SIPP based on their simulated job assignment, the associated job attributes (including fatal risk) and the worker's preference parameters recovered in the simulation. Based on the simulation results, the computed MWTP is then compared to an estimate for the implicit price of risk that is obtained from a traditional hedonic wage model. By using the simulated data, 3,580 hedonic wage models are estimated, that vary by hedonic wage functional forms, estimators, the inclusion of non-pecuniary job attributes and the presence of different occupation and/or industry fixed effects. Modeling choices are then evaluated by the difference between true MWTP and the implicit price of risk estimated from each regression model.

The remainder of the chapter is organized as follows. Section 5.2 reviews a similar model undertaken by Kuminoff et al. (2010) in a housing market context. Then, section 5.3 describes the empirical application in detail, followed by the evaluation results in section 5.4.

Section 5.5 presents the data used in estimating the VSL to compare to the literature and section 5.6 summarizes the estimation results.

5.2 Related work

Ideally, the ability of an empirical modeling would be examined by comparing resultant compensating wage differentials with the true hedonic wage locus. The difficulty is that the true hedonic wage locus is unknown. Cropper, Deck amd McConnel (1988) (hereafter CDM) were the first to meet this challenge by developing a theoretically consistent framework for computing a hedonic equilibrium in a housing market setting. As described in chapter 4, they use a simulation approach to compute the hedonic housing market equilibrium. They define a household utility by using the simulated house allocation, house attributes, the price supporting the allocation and the characteristics of households. And then they compute each household's preference and the MWTP for house characteristics. Given the MWTP computed in the equilibrium, they examine the accuracy of the marginal price of risk from hedonic models on the basis of how close it is to the MWTP. They vary the hedonic models by different functional forms such as: linear, semi-log, double-log, quadratic, linear Box-Cox (LBC), and quadratic Box-Cox (QBC). The key findings in CDM were that flexible functional forms (i.e. LBC and QBC) performed best when all variables were included in the model, but simpler functional forms (i.e. linear, log-linear, and log-log) performed best in the presence of omitted variables.

Recently, Kuminoff et al (2010) (hereafter, KPP) revisit the work of CDM, incorporating hedonic econometric strategies that are commonly used in more recent studies. Instead of simply testing the accuracy of the functional forms, they examined the performance of the models with different combination of functional forms and covariate choices. Specifically, KPP focused on recent choices for covariates that researchers employ in an attempt to mitigate omitted variable bias. In this research, I follow CDM and KPP's simulation approach closely, given the circumstances in the labor market is parallel with the housing market. The simulation methods that are used by CDM and KPP to construct the hedonic housing market equilibrium were presented in chapter 4. In this section, I will describe work by KPP in detail, since I directly apply their evaluation method in the analysis. [32] Below I describe the data that used in the KPP's analysis and the empirical hedonic models KPP evaluated against the simulated equilibrium marginal values.

KPP employ housing stock data from approximately 104,000 single-family housing sales data in Wake County, North Carolina between 1992 and 2000. The data contain the sale price of each home, its structural characteristics, the characteristics of its neighborhood (median commute time, median household income, and the % under 18 in each census block), and urban amenities that vary within and between neighborhoods (proximity to open space and proximity to shopping centers). The distribution of household income was defined using data from the *2000 Census of Population and Housing*, which reports the number of households in each of 16 income bins.

[32] Since KPP follow CDM, but introduce more modern data and empirical methods, I only review KPP in detail.

KPP employ the Wake County data to solve for 100 different hedonic equilibria for each of three utility functions (Cobb-Douglas, Translog, and Diewert utility functions). On each replication, a sample of households is randomly drawn from the Census income distribution for Wake County under the assumption that households are uniformly distributed within each of 16 income bin. Then a random sample of homes is drawn from Wake County database of 104,000 housing transactions. Given a random sample of homes and a random sample of households, they solve for the assignment of households to homes and the vector of equilibrium prices which jointly clear the market for housing. The resultant price vector and the associated housing characteristics comprise the simulated data they use to evaluate alternative cross-sectional model specifications for the hedonic price function.

In addition to evaluating a cross-section estimator, they further expand the scope of the simulation to investigate the performances of panel data estimators. Although KPP do not have true panel data available, they generate panel data through shocking a subset of neighborhoods with new amenity levels and solve for the new equilibrium prices and housing assignments after the shock. They base the shock on the actual change of commute times in Wake County that occurred following the construction of a new beltline highway in the late 1990s. Then they determine the percentage reduction in commute time associated with the section of tracts in the county. The data on equilibrium prices for the same homes before and after the market shock are used as panel data in the estimators.

Once the hedonic equilibrium is simulated, KPP estimate traditional hedonic price models using information on the home that household is assigned in and associated house

attributes. They vary the hedonic price models by different functional forms with different estimators. They apply six functional forms (linear, quadratic, semi-log, double-log, LBC, and QBC) to the OLS estimator in cross-section data and the pooled OLS estimator in panel data. They explore the linear, semi-log, and double-log price functional forms with two panel data estimators (first difference (FD) and difference in difference (DID)).

In addition to the variation of the functional forms, KPP also vary the hedonic models with different covariates in an attempt to mimic the omitted variable issues in the recent studies and to explore the effect of econometrics strategies that are employed by researchers to mitigate the omitted variable bias. KPP concentrate on two spatially delineated variables: *commute time* and *nearest park* to evaluate the performance of hedonic models with spatial fixed effects. These two spatially delineated variables are chosen to represent varying degrees of spatial heterogeneity. They consider two types of omitted variable scenarios. One is where they observe all of the structural characteristics of the home but omit all three spatially delineated attributes from every regression.[33] Another scenario is where KPP randomly choose three characteristics from the full set of housing characteristics, neighborhood characteristics, and urban amenities to omit. This set up is to address the concern that the pattern of results may only reflect the particular variables they choose to omit from the regression.

The variation of their models results in a large set of hedonic estimates along with the modeling choices that are applied to the estimation model. It builds up a dataset with which

[33] Three spatially delineated attributes are: (1) median income in the census block, (2) share of the population under 18 in the census block, and (3) distance from the home to the nearest shopping center.

the effects of different model specifications can be decomposed and identified. This approach is valuable to my analysis. As will be discussed below, the choice of covariates in hedonic wage model, the econometrics strategies and the inclusion of occupation and industry fixed effects are of particular interest in my analysis.

In summary, KPP run 54,000 regressions to evaluate 540 different hedonic models using cross-sectional and panel data generated from the equilibria.[34] The models varied according to the shape of the utility function (*Cobb-Douglas*, *Translog*, *Diewert*); the form of the price function (*linear*, *quadratic*, *semi-log*, *double-log*, *LBC*, and *QBC*); the spatial controls for omitted variables (*no controls*, *fixed effects for census tracts*, *spatial error model*, *spatial lag model*); estimator type (*cross-sectional OLS*, *pooled cross-section*, *differences-in-differences*, *first differences estimator*); the variables omitted from the regression (*none, three neighborhood amenities, three randomly chosen variables*); and the number of homes in the simulated market (*200, 2000*). Between two and eleven variables entered the price function in each regression, they generate more than half a billion estimates for the MWTP of individual attributes across the 54,000 regressions.

The estimates for the MWTP from 54,000 regressions are then used to systematically summarize the performances of different modeling choices. For each regression model, the

[34] They carry out 100 Monte Carlo replications to solve for 100 different hedonic equilibria for each of the three utility functions. With these 100 simulated results, they replicate each of the 540 regression models 100 times and provide the overall summary of models' specifications.

difference between every household' MWTP for housing attributes and the estimated marginal price of attributes are calculated. The difference provides information on the direction and magnitude of the bias that is caused by omitted variables and functional form misspecification. Based on the calculated differences, KPP summarize the evaluation results in two ways. One is to select subset of the results and provide model-specific summary on the difference between the MWTP and the marginal price of attributes. They highlight the results on using spatial fixed effects to control for omitted variables in cross-section data and the performance of panel data estimators. They also use a meta-analysis to synthesize the full set of results. Specifically, they regress the absolute value for normalized bias on a set of indicator variables for modeling choices. The modeling choices include functional form, spatial fixed effect control, the type of estimator and market size. The estimated coefficients of indicator variables are interpreted as the effect of modeling choices on the magnitude of bias.

Their key finding is that when spatial fixed effects are used to control for omitted variables, the flexible specifications for the price function, such as the QBC model, outperform the simpler linear, log-linear and log-log specifications by producing lowest difference between the MWTP and the marginal price of attributes. They also provide evidence that alternative panel models such as difference-in difference and first-difference can effectively eliminate the biases caused by temporal variant of functional form misspecification. Moreover, they find that difference-in-differences and first-differenced

specifications for the price function can perform well in the presence of time-varying omitted variables as long as the spatial fixed effects are included in the model.

5.3 Evaluation of hedonic wage models

In this research, I directly follow the approach in KPP and extend it to examine the accuracy of hedonic wage modeling specifications.[35] The benchmark data to which hedonic wage models are compared is the simulated wage equilibrium computed numerically in the previous chapter. This is a pseudo dataset which is defined by observed worker's characteristics and the simulated jobs to which they sort with associated job characteristics and simulated wages. Specifically, in this pseudo data each worker's demographic characteristics remain the same as the ones observed from the SIPP labor force data. However, each worker's wage, fatal risk and other job attributes that are associated with the jobs are actually determined by the worker-job matching algorithm in the simulation. As such, the worker-job match is different from what is observed in the labor force data. In addition to worker characteristics, simulated wage and job risks, this pseudo dataset also includes worker's heterogeneous preferences toward the different job characteristics. The preference parameters are computed as part of the simulation exercise.

5.3.1 Empirical estimators employed

Below I will describe the estimators that are used in the analysis. I begin with an econometrically familiar representation of the hedonic wage equation. For worker i (i =

[35] In order to distinguish the models in chapter 4 and chapter, I will refer to hedonic wage equilibrium from chapter 4 as "simulated wage equilibrium" and the reduced form models in chapter 5 as "hedonic wage models".

1,…,N) in industry j (j = 1,…,J) and occupation k (k = 1,…,K) at time t (t=1,… T), the hedonic tradeoff between the wage and risk of fatality is:

$$f(W_{ijkt}) = \alpha_{oi} + \alpha_1 X_{it} + \alpha_2 Z_{ikt} + \alpha_3 D_{jkt} + \beta\, f(p_{jk}) + u_{ijkt}, \qquad (5.1)$$

where p_{jk} is the industry and occupation specific fatality rate; X_i is a vector of demographic characteristics such as worker's education, age, race, marital status, union status, and dummy variables for region of residence. D_{jk} is a vector of dummy variables for industry-based fixed effects and/or occupation-based fixed effect; Z_{ik} is a vector containing occupational attributes included in the model. In this equation, $f(W_{ijk})$ is a transformation of the wage rate which may be linear, natural log, or Box-Cox transformation. In Box-Cox model, the transformation of dependent variable has form: $W_{ijkt}(\lambda) = \frac{W^{\lambda}-1}{\lambda}$ for $\lambda \neq 0$. $f(p_{ijk})$ is a transformation the fatal risk rate which may be linear, and natural log. The combination of transformation of W_{ijk} and p_{ijk} results in four basic functional forms: linear, semi-log, double-log, and LBC. In addition to these four functional forms, I also explore three functional forms with quadratic fatal risk term: linear with quadratic fatal risk term included, semi-log functional form with quadratic fatal risk term included and QBC. These three functional forms including quadratic risk term have been often used in the hedonic wage literature such as Moore and Viscusi (1988), Scotton and Taylor (2009), and Viscusi and Hersch (2000).

I evaluate five approaches to estimations. The first two estimators are simple one-year cross-sectional OLS and pooled OLS.[36] For the cross-section OLS estimator, the 2006 sample of the simulated data is used. For the pooled OLS estimator, the 2004, 2005 and 2006 simulated data are pooled. The pooled OLS models include occupation and/or industry dummy variables but ignore the temporal dimension of the data.

Given the availability of panel data, I also explore the basic first-differenced (FD) and fixed effects (FE) panel estimators. These models have been applied in the recent panel studies to eliminate potential endogeneity bias (Viscusi et al. 2007; Kniesner et al. 2011, Kochi, 2011). In this analysis, the basic FD and FE estimators are applied to the same sample as in the pooled OLS. It is known that omitted variables and measurement errors can lead to a problem of serially correlated errors in first difference (FD) models (Griliches and Hausman, 1986). Viscusi et al. (2007) suggests using a wider time interval when exploring FD model to mitigate the effect of autocorrelation in errors. Given the data they used is across 1993 to 2001, they differ the variables in 8 years intervals (i.e. ln W_{i2001} − ln W_{i1993}). In this analysis, the data generated by the simulation is across three years, 2004-2006. Therefore, in addition to the FD estimator which differences the variables between periods t-1 and t, i.e. Δ=1, I also estimate FD models with two-year difference, i.e. Δ = 2.

All seven functional forms are tested in both simple one-year cross-section OLS and pooled OLS estimators. The OLS estimator use the function described in equation 5.1 with transformation to linear, semi-log, double-log, LBC, QBC, linear with quadratic fatal risk term and semi-log with fatal risk term. In the FD and FE estimation models, five functional forms are tested: linear, semi-log,

[36] Most of early hedonic wage studies rely on cross-sectional model. More recent studies include Arabshebani (2001), Black et al. (2003), and Gunderson and Hyatt (2001). Pooled cross-sectional models have also been included in the studies to compare with the simple cross-sectional models (Viscusi, 2007; Kniesner et al., 2010; and Kochi, 2010).

double-log, linear with quadratic risk term and semi-log with quadratic risk term. An example of a FD estimator using the natural log of wage as dependent variable is:

$$\Delta \ln W_{ijkt} = \alpha_1 \Delta X_{it} + \alpha_2 \Delta Z_{jkt} + \alpha_3 \Delta D_{ijkt} + \beta \Delta P_{ijkt} + \Upsilon \Delta \delta_t + \Delta u_{ijkt}, \qquad (5.2)$$

where Δ refers to the first-difference operator. For instance, $\Delta X_{it} = X_{it} - X_{it-1}$, and δ_t is the time dummy.

An example of linear wage specification in a fixed effects model with *t* periods implies the following estimating equation:

$$(W_{ijkt} - \overline{W_{ijk}}) = \alpha_1 (X_i - \bar{X}_i) + \alpha_2 (Z_{ijk} - \overline{Z_{ijk}}) + \alpha_3 (D_{ijkt} - \overline{D_{ijk}}) + \beta (p_{jk} - \overline{p_{jk}}) + \Upsilon (\delta_t - \bar{\delta}_t) + (u_{ijk} - \overline{u_{ijk}}) \qquad (5.3)$$

where $\overline{W_{ijk}} = \frac{1}{T}\sum_{t=1}^{T} W_{ijkt}$, $\overline{p_{ijk}} = \frac{1}{T}\sum_{t=1}^{T} p_{ijkt}$, $\overline{X_{ijk}} = \frac{1}{T}\sum_{t=1}^{T} X_{ijkt}$, $\overline{D_{ijk}} = \frac{1}{T}\sum_{t=1}^{T} D_{ijkt}$, $\overline{Z_{ijk}} = \frac{1}{T}\sum_{t=1}^{T} Z_{ijkt}$, and $\overline{u_{ijk}} = \frac{1}{T}\sum_{t=1}^{T} u_{ijkt}$.

Similar to KPP, the performance of various hedonic wage estimators are examined in two scenarios in which, i) all job attributes are observed, and ii) there are omitted variables in the model. As described in the previous chapter, the sorting model assumed 23 worker and job characteristics influenced the wage equilibrium. These variables are presented in Table 5.1, and are grouped by major type of characteristics (i.e. worker characteristics, general job attributes such as whether working overtime is required, and occupation-specific characteristics developed from the O*NET database.) The full-information scenario is defined as the one including all the attribute variables presented in table 5.1. Recall that 15 variables describing O*NET job attributes are grouped into five categories: i) cognitive demands, ii) motor skill demands, iii) physical skill demands, iv) hazard exposure, and v)

people skill demand. These 15 O*NET job attributes are considered when the sorting model in chapter 4 is carried out computing worker's true MWTP.

The omitted variable scenario is constructed where the *groups* of omitted variables are randomly chosen instead of one particular variable being omitted. This set up is to mimic the situation that researchers are more likely to lack a set of information relative to one particular type of job attribute. By randomly omitting groups of the O*NET job attributes, I construct thirty-one omitted variable scenarios with different combinations of the O*NET job characteristics that are omitted.

There is a controversy over whether or not industry dummy variables should be included in hedonic wage estimation (Viscusi and Aldy, 2003). In the early studies, the risk variables are generally constructed based on worker's industry affiliation, and that leads to a strong correlation between risk variables and industry dummy variables (see Black et al., 2003 for detailed discussion). Including industry dummy variables generally increases the variance of the risk coefficient and often leads to an insignificant risk coefficient. However, omitting industry variables where strong correlation exists between risk and industry variables may bias the risk coefficient. Since the CFOI data provide the risk rates measured by both occupation and industry, I explore the effect of occupation/ industry dummy variables in the omitted variable scenario where the O*NET job attributes are omitted.

As described in chapter 3, the fatal risk rates in this research are measured on a 2-digit combination of occupation and industry affiliations, and the O*NET occupational attributes are measured on a 6-digit occupational level based on the coding system of Standard Occupational Classification 2000. I include 1-digit level occupation dummies and 1-digit level industry dummies in the models. These industry and occupation dummies are used to evaluate the performances of hedonic

wage models in the presence of industry and occupational fixed effects. Specifically, the models will vary by: i) only including occupational fixed effects, ii) only including industry fixed effects, and iii) including both occupational and industry fixed effects.

In sum, the estimation models to be evaluated vary by: i) whether or not the covariates are fully observed or some are omitted *(the omitted variables are randomly chosen from O*NET job characteristics groups)*; ii) functional forms of the wage equation *(linear, linear with quadratic fatal risk, log-linear, log-linear with quadratic fatal risk, double-log, LBC,* and *QBC);* iii) the fixed effect controls for omitted job characteristics *(no controls, fixed effects for occupational categories and/or industry categories)*; iv) the approaches to cross-section OLS estimation *(OLS on a simple cross-section or pooled OLS)*; and v) the approaches to panel data estimation *(FE, FD (*Δ*=1) and FD (*Δ*=2));*

5.3.2 Evaluation method

To quantify the performance of an individual model, I directly follow the approach in KPP computing the difference between every worker's MWTP for risk and the corresponding partial derivative of the hedonic wage function.[37] A worker's MWTP is computed by employing Von Neuman-Morgenstern expected utility function which identifies the worker's maximization problem as:

$$E(U)=\max (1-p)\, U_0\, (\bar{y} + W(p)) + p\, U_1\, (\bar{y} + \tilde{W}), \qquad (5.4)$$

And the individual worker's MWTP can be derived as:

[37] This is also the method employed by Cropper, Deck and McConnell (1988).

$$\text{MWTP} \equiv \frac{dw}{dp} = \frac{U_0(\bar{y} + \text{W(p)}) - U_1(\bar{y} + \tilde{W})}{(1-p)U_0'(\bar{y} + \text{W(p)}) + pU_1'(\bar{y} + \tilde{W})}. \quad (5.5)$$

The Workers' utility functional form is defined as Cobb-Douglas, $U_i^j = (W_i^j + \overline{y_i}) Z_j^{\alpha_i}$. Based on this standard multiplicative utility function, worker *i*'s expected logarithm utility from consuming job j's attributes and wages is: $U_{0i}^j = ln(W_i^j(\text{p}) + \overline{y_i}) + \alpha_i lnZ_j$, which is also the same utility function used in the sorting algorithm. As discussed in chapter 4, workers are assumed to have no bequest values and do not have ex-ante utility for the wealth in the state of death, i.e. $U_{i1}(\bar{y} + \tilde{W}) = 0$. Given the utility functional specification, each argument in equation 5.5 can be specified. The non-labor income, $\bar{y}$, of each worker comes from observed data from the SIPP. Workers' preference parameters, fatal risk rates, job characteristics and wages are from the simulated labor market equilibrium.

In empirical work, the hedonic wage model is commonly defined as: f(W_i) = α_{oi} + $\alpha_1 X_i$ + $\alpha_2 Z_j$ + $\alpha_3 D_j$ + β f(p_j) + u_i (all variables were described in equation 5.1). The difference between worker's MWTP computed in equation 5.5 and the corresponding partial derivative of the estimated hedonic model (equation 5.1) is defined as the bias (error), e_p of worker's estimated valuation of fatal risk:

$$e_{pi} = \frac{\partial \text{W}(p_i)}{\partial p_i} - MWTP_i . \quad (5.6)$$

For each of hedonic wage model estimated, e_{pi} is computed for every worker and varies across workers in the following way.

First, the marginal price of risk, $\frac{\partial W(p_i)}{\partial p_i}$, may vary across workers depending on the functional forms and specifications of the hedonic wage model. In linear models, there is only one marginal price of risk, given the $\frac{\partial W(p_i)}{\partial p_i}$ is simply the coefficient of fatal risk rates. In quadratic models, the marginal price of risk varies by the fatal risk rates, given $\frac{\partial W(p_i)}{\partial p_i}$ depends on not only the coefficients of linear and quadratic fatal risk rate term, but also the fatal risk rate, p_j. For instance, if the workers are in 211 different cells of occupation/industry fatal risk rates matrix, there will be 211 $\frac{\partial W(p_i)}{\partial p_i}$, since each worker's marginal price of risk is calculated with the fatal risk rate associated with his job. In other estimation models with semi-log, double-log, and Box-Cox functional forms, the calculation of $\frac{\partial W(p_i)}{\partial p_i}$ also depends on the worker's wage which is different across workers. Thus, the marginal price of risk in these models varies across workers. And the MWTP varies by workers since each has unique preference (α_i) and initial non-labor income ($\bar{y}$).

After computing the bias, the mean and the standard deviation of e_{pi} associated with each hedonic wage model can be obtained. The mean $\bar{e}_p$ is defined as $N^{-1}\sum_i e_{pi}$ and $\bar{s}_p$, the standard deviation of e_{pi} is given by: $\sqrt{N^{-1}\sum_i(e_{pi}-\bar{e}_p)^2}$. Finally, I normalize the mean and the standard deviation of the bias by the average MWTP for each hedonic wage model. Specifically, I define β_p as the mean bias $\bar{e}_p$, normalized by the average MWTP as :

$$\beta_p = \frac{\bar{e}_p}{N^{-1}\sum_i MWTP_i}, \tag{5.7}$$

The standard deviation of the error, e_p, in equation 5.6 is given by:

$$S_p = \frac{\bar{s}_p}{N^{-1}\sum_i MWTP_i}, \tag{5.8}$$

The summary of bias indicates the performance of each estimation model. The variation of the hedonic wage modeling choices results in 3,580 regressions to evaluate. These 3,580 regressions produce 6,766,200 e_{pi} and 3,580 $\bar{e}_p$ and β_p with $\bar{S}_p$ and S_p.

I present the results in the following ways. First, I pool all the evaluation results and provide a general summary. The general summary focuses on reporting the relative performance of different functional forms in full-observed scenario and omitted variable scenario. Then, I highlight the results from the models including occupation/industry fixed effects and the results indicating the performance of panel data estimators. Finally, I conduct a meta-analysis of models that regresses the values for | β | on indicators for various modeling choices. In section 5.5, I use what we learn in model performance and estimate hedonic wage models using "best practices" model specifications as defined by best performing models and report VSL estimates and compare them to the literature.

5.4 Testing Results

5.4.1. General summary

Table 5.2 reports the summary statistics of biases from 3,580 regression models using seven functional forms. These functional forms are: basic linear, semi-log, double-log, LBC and QBC, linear functional form with quadratic fatal risk term included, and the semi-log functional form with quadratic fatal risk term included. Panel A presents the summary of bias from the models in fully observed scenario. The estimation models vary by the estimators: *i)*

cross-section OLS, ii) pooled OLS, iii) FE, iiv) FD (Δ=1) and v) FD (Δ=2). There are 29 regression models.[38] The numbers in the table represent the mean and standard deviations of the biases. The mean and the standard deviation of bias are further averaged over regression models varying by the estimators.

The estimated biases vary across the functional forms. The absolute mean biases range from as low as 11% in the LBC model to as high as 154% in the semi-log model. Comparing the simpler functions (linear, semi-log, and double-log) with more flexible functions (LBC and QBC), one may notice that the more flexible functions tend to produce smaller bias. Moreover, adding a quadratic risk variable into the regression model has a very different effect on the performance across different models. For the LBC models, inclusion of a quadratic risk variable appears to approximately double the absolute value of bias as compared to models without quadratic risk term. Including quadratic risk in the semi-log form, however, produces substantially lower absolute mean bias, from 154% to 70%.

Among all the functional forms, the LBC and QBC functions have the lowest standard deviation, 0.16 and 0.17, respectively. The semi-log and the semi-log with quadratic risk term models have highest standard deviation 2.54 and 1.81, respectively. The effect of including a quadratic risk term on the standard deviation also seems to be inconsistent across the functional forms. For the linear functional form, it appears that including a quadratic risk term tends to triple the standard deviation of the biases. For the semi-log functional form, adding a quadratic risk term decreases the standard deviation. And for the Box-Cox functional form, there is very little the impact of adding a quadratic risk term to the model.

[38] In panel data estimators, the LBC and QBC functional forms are not applies.

The large spread of the biases exhibited in the models such as semi-log models, implies that these estimation models might predict marginal prices poorly at extreme levels of fatal risks.

Based on the criterion of mean and standard deviation of the bias, the results in Panel A indicate that across all five estimators in the fully observed scenario, the LBC and QBC functional forms outperform others as they have relatively low mean bias and small standard deviation of the bias. These findings are consistent with CDM and KPP.

Now consider the omitted variable scenario. This scenario is constructed in an attempt to depict a more realistic case where the researchers are unable to observe some features of jobs that matter to workers. Panel B presents the results from the single cross-section models where the O*NET job attributes are all omitted. The estimation models vary by different presence of occupational and/or industry fixed effects: i) *no fixed effect controls; ii) fixed effects for occupations only; iii) fixed effects controls for industries only;* and *iv) control for both occupation and industry.* There are 28 regression models. The mean and standard deviation of the bias are average over these 28 regression models.

Comparing row (5) and row (1), we can see that the value of biases increases for six out seven functional forms. The increase of the absolute bias indicates, not surprisingly, that omitting job attributes hurts the performance of single cross-section models, regardless of the inclusion of occupation/ industry fixed effects. Among all the functional forms, the linear function form produces the least average bias, 7%. The double-log and semi-log functional forms produce largest absolute bias, 200% and 197%, respectively. While the Box-Cox functional forms produce relatively lower average biases, their biases are substantially higher

than the one from linear models. The impact of including quadratic risk term on the average bias is consistent with that in fully observed scenario. Inclusion of quadratic risk variable increases the absolute bias for linear and Box-Cox models, but decreases the absolute bias for semi-log models. However, the difference of the biases between the models with and without quadratic risk term is bigger in panel B than in panel A. It implies that the estimation models tend to be more sensitive to the functional specifications in the omitted variable scenarios than in the fully-observed scenario.

Contrary to the fully observed scenario, the semi-log functional form produces least standard deviation among all the functional forms. The QBC functional form which has least standard deviation in panel A has relatively large standard deviation in the cross-sectional models with job attributes variables omitted. Furthermore, it appears that adding quadratic risk term increases the standard deviation of bias for linear, semi-log and Box-Cox functional forms. Notice that the standard deviation for semi-log with quadratic risk term is about ten times as large as the one for semi-log without quadratic risk term. Large absolute bias and small standard deviation of bias from semi-log functional form reflects that this functional form consistently performs poorly with cross-sectional models, regardless of the occupation/industry fixed effects. This finding does raise concern that these commonly used functional forms are probably not appropriate to be used for cross-section estimations with the existence of omitted variable problem.

Panel C presents the results from the panel data models where the O*NET job attributes are all omitted. The models vary by pooled cross-sectional OLS and panel data estimators*: i) pooled OLS, ii) FE, iii) FD (Δ=1)* and *iv) FD (Δ=2)*.[39] The estimation model also vary by different presence of occupational and/or industry fixed effect controls: i) *no fixed effect controls; ii) fixed effects for occupations only; iii) fixed effects controls for industries only;* and *iv) control for both occupation and industry.* There are 88 regression models, and the mean and standard deviation of the bias are averaged over these 88 regression models.

The results in Panel C indicate a pattern that is similar to the one presented in Panel B. Among all seven functional forms, the double-log and semi-log function produces highest absolute bias value, 156% and 105%, respectively. The linear functional form produces lowest absolute averaged biases of 11%. Adding a quadratic risk term reduces the average bias for the semi-log functional form, but increases the value of the bias for the linear functions and Box-Cox functions. Adding a quadratic risk to the semi-log functional form also appears to have a large impact on the average bias. In the single cross-section models, adding a quadratic risk term to semi-log functional form reduces the average bias from 197% to 93%. In panel data models, adding quadratic risk term to semi-log functional form reduces the average bias from 105% to 36%. With both single cross-section and panel data estimators, the inclusion of the quadratic risk term in the semi-log functional form reduces

[39] In cross-section estimator, the simulated data from 2004 is applied. In panel data estimators, the simulated data from all three years, 2004-2006 are applied.

about a third of the average bias. However, the contrast of the average biases value from these two models shows that semi-log models perform better with panel data overall.

The pattern of dispersion for the bias presented in Panel C across the functional forms is very similar to panel B. The LBC and QBC have the lowest standard deviation, 0.31 and 0.50, respectively. Models with quadratic risk tend to have larger standard deviation of biases. However, in comparison to row (6), the average standard deviation of the bias is significantly larger in panel C for five out of seven functional forms. In contrast to panel B, the average standard deviation of the bias from the semi-log without quadratic risk is almost as large as the one produced by the semi-log with quadratic risk function. The increase of spread of the biases in the panel models, especially with the semi-log functional form, indicates that the marginal price of risk from panel data models could be more sensitive to the choice of sample of workers.

Panel D reports the results from all five estimators: *i) cross-section; ii) pooled OLS, iii) FE, iv) FD (*Δ*=1) and v) FD (*Δ*=2)*. The models also vary by the presence of occupation industry fixed effect control. The job attributes that are constructed from O*NET are still omitted all at once to mimic the omitted variable scenario commonly existing in the literature. There are 116 regressions models. The mean and the standard deviation of bias are averaged over these 116 regression models.

The biases reported in panel D are averaged over the models presented in panel B and panel C. Therefore, the general pattern of functional forms' performance is the product of the pattern summarized in panel B and panel C. With all cross-section and panel data estimators

being considered, the linear function and LBC function outperform other functional forms with lowest absolute averaged biases, 10% and 11%, and the smallest standard deviation, 0.44 and 0.49, respectively. Among all seven functional forms, the double-log and semi-log functions perform worst.

Table 5.2, Panel E reports the bias from the estimations in the omitted variable scenario where the omitted job attributes groups are randomly chosen. In addition to varying functional form, presence of fixed effect, and different econometrics methods (i.e. cross-section or panel data approaches), the models also vary by which O*NET variable groups are omitted. Recall that I construct thirty-one omitted variable scenarios with different combinations of the O*NET job characteristics that are omitted. There are 3,480 regression models in omitted variable scenario. The mean and standard deviation of bias are averaged over these 3,480 models. For all the functions, the pattern and the magnitude of the bias are comparable to the ones presented in panel D where the O*NET job attributes are all omitted. It implies that the performance of the functional forms in panel D is typical even when some, but not all job attributes are available to the researchers.

In order to present the general pattern of bias distribution across the seven different functional forms, we use a Box-plot to graphically display the results summarized in Table 5.2 Panel B, D and E. The plot is presented in Figure 5.1. The "box" itself represents the middle 50-percent of the data. The upper and lower boundary of the box locates the 75^{th} and 25^{th} percentile of the data, respectively, and mid line indicates the median of the data. The "whiskers" indicate variability outside the upper and lower quartiles, and are the value of the

observation closest to being 1.5 times the inter-quartile range without being larger. The box-plot visually summarizes the bias distribution across functional forms in terms of the range, median, normality and skewness.

The size of the boxes is different across the functional forms. Clearly, the boxes for linear functional form and Box-Cox functional form are significantly smaller than the other functional forms. This size indicates a distribution of the bias within two inter-quartiles. Contrary to the linear model and Box-Cox models, the double-log, linear with quadratic risk and semi-log with quadratic risk models produce bigger boxes which signify a larger spread in the middle two inter-quartiles and instability of the associated estimates of MWTP for risk.

The location of box in between the whiskers and the position of the median line in the box provide insight on the normality of the bias's distribution. For most of the linear and LBC models, the median line is around the middle of the box, and the box is approximately centered in the middle of the whiskers. This position indicates that the biases produced by these models are more symmetrically distributed. Notice that using a linear functional form in the cross-section models with all the O*NET job attributes being omitted produces a median level which is shifted significantly to the low end of the box (i.e. skewed to the left). This implies that most of these models produce large downward biases, but there are a few exceptionally large upward biases or small downward biases. Those exceptional values lead to a significantly shifted median line. However, the box width for these models is very small with bias values at 25th percentile being between -0.5 to 0. This implies the majority of the biases are still close zero. Overall, the plots indicate clearly that linear functional form and

LBC functional form perform best with majority of the biases being densely distributed around zero.

The length of the whisker and the position of the median level of bias from double-log, linear with quadratic risk, and semi-log with quadratic risk functional forms reinforce the sensitivity of the corresponding estimates to the omitted variables. In particular, the length of the whisker based on the models varying by the econometric strategies and the groups of omitted variables implies double-log, linear with quadratic risk, and semi-log with quadratic risk functional forms are vulnerable to different modeling choices.

5.4.2. Effect of occupational and industry fixed effect controls

Adding occupation and/or industry dummies to regression models has been a conventional way in the hedonic wage studies to alleviate omitted variables bias (Krueger and Summer, 1988; Leigh, 1995; Dorman and Hagstrom, 1998; Black et al. 2003; Viscusi, 2004; Hintermann et al., 2010; and Kochi, 2010). Occupation dummies are often included to proxy job characteristics other than fatal risk and industry dummies are used to purge out the inter-industry wage differentials.

Table 5.3 presents the results from 840 single cross-sectional models *i) with no industry or occupational dummies, ii) with only industry dummies, iii) with only occupational dummies,* and *iv) with both occupational and industry dummies*. The models also vary by different combinations of the O*NET job characteristics groups that are omitted.

The results from single cross-section models in full-observed scenario (no fixed effects included) are presented for comparison in Panel A. The general pattern of bias across

the functional forms is consistent with the one presented in Table 5.2, Panel A. The two Box-Cox functional forms produce relatively lower absolute bias. The double-log functional form does worst with large mean and high standard deviation of the bias. The impact of adding a quadratic risk term to the cross-section models, however, is different from that presented in Table 5.2, Panel A where both the cross-section and panel data models are considered. It appears that including quadratic risk term in the semi-log and Box-Cox functional forms lower the absolute mean bias substantially. In particular, the inclusion reduces the bias from 96% to only 3%.

Moreover, when only single cross-sectional models are considered, the effect of quadratic risk term on the standard deviation of bias is more consistent across the functional forms. For linear, semi-log and Box-Cox functional forms, adding a quadratic risk term increases the standard deviation of bias by three to seven times. Especially for the semi-log functional form, the inclusion of quadratic risk term increases the standard deviation of bias from 0.12 to 0.94. Comparing the semi-log functional form with quadratic risk term with the semi-log functional form without quadratic risk term, one notices the significant difference on the mean and the standard deviation of the biases. This contrast implies in the fully observed scenario, applying semi-log functional forms to cross-section estimator may need caution. While the semi-log with quadratic risk term has low mean bias, the large standard deviation implies a large spread of the bias.

The comparison between row (1) and row (3) shows that the correlation between observed and the omitted job attributes has different impact on different functional forms.

The general pattern is that the omitted job attributes reduce the estimates of MWTP in the linear model and LBC models. Specifically, the positive bias of the MWTP estimate from linear model is reduced from 0.25 to 0.11, and the positive bias of the MWTP estimate from the LBC model is reduced from 0.15 to 0.01. The omitted job attributes, however, tend to produce higher estimates for MWTP in the models with quadratic risk term. For the linear functional form with quadratic risk term, the negative bias is reduced from -0.75 to -0.45. For semi-log with quadratic risk, the positive bias increases from 0.03 to 0.31. And for the QBC, the negative bias is changed from -0.07 to 0.12. Notice that the direction of the effect on four out of seven functions contradicts the conventional wisdom. For linear, linear with quadratic risk, double-log and LBC functions, the omitted variable bias moves estimates closer to the true MWTP for the risk. For example, the linear model overestimates the MWTP by 25% when all variables are observed. Omitting the job attributes reduces the bias to 11%. This counter-intuitive result is similar to what KPP report in their comparison of marginal prices for housing attributes relative to simulated MWTP. As pointed out in their study, this counterintuitive result may be caused by the unknown true shape of the equilibrium functions. In the simulation, the "true" shape of equilibrium wage function relies heavily on the shape of the utility function. Failing to specify it correctly can lead to badly biased estimates for MWTP even when omitted variables do not present a confounding influence.

In Table 5.3, Panel C, the mean and standard deviation of the bias is averaged over the models that vary by three choices of fixed effects: *i), fixed effects for occupations only; ii) fixed effects controls for industries only;* and *iii) control for both occupation and industry.*

One may notice that adding fixed effects to the model further decreases the estimates of the MWTP for all the functions (comparing row (3) and row (5)). The impact of fixed effects is not consistent across the functional forms, however. Adding fixed effect to the models improves the performance of linear function (the absolute value of average bias is reduced from 11% to 1%), but worsen the performance of the double-log resulting in the highest absolute value of average bias of 229%. Interestingly, for the semi-log model without quadratic risk, the comparison of average biases from row (3) and row (5) indicates that fixed effects seems to have very little impact on the bias.

The comparison between row (4) and row (6) reflects how much the occupation and industry fixed effects affect the spread of the biases. While the general pattern in row (6) is comparable to row (4), the sensitivity of different functions to occupation and/or industry fixed effects does appear to vary across the functions. For example, the semi-log functions has smallest change in standard deviation (from 0.08 to 0.15), while the double-log function, is highly sensitive to different fixed effects choices with largest change of standard deviation (from 0.26 to 1.26). Moreover, the models with a quadratic risk variable included tend to be more sensitive to the occupation and industry fixed effects. For instance, controlling for fixed effects increases the QBC function's standard deviation from 0.39 to 0.93, while it only changes LBC function's standard deviation from 0.26 to 0.45.

In order to examine the impact of fixed effects on the bias more closely, Panel D describes results from the single cross-sectional models using different combinations of occupation and industry fixed effects. The difference between row (3) and (7) indicates that

controlling for industry fixed effects in the single cross-section models tend to increase the estimate of the MWTP of risk for six out of seven functional forms. For example, controlling for industry fixed effect in the LBC model increases the upward bias of the estimates from 8% to 26%. This finding is reinforced by the comparison between row (9) and (11). It is, however, not consistent with findings in the literature that the coefficient on risk often tends to be significantly lower when industry dummies are includes (Leigh, 1995; Dorman and Hagstrom, 1998; Mrozek and Taylor, 2002; Kochi, 2010 and Scotton, 2013). This result could be caused by not fully accounting for inter-industrial wage differentials in the data for evaluation. The wage disparities among measurably similar workers that are employed in different jobs come from workers' preferences to various job attributes and the firms' compensation for the job attributes they offer. The job attributes, however, are measured only by occupations. Moreover, the wage differentials do not account for the sectorial wage premia across different industries as documented (Helwege, 1992; Zanchi, 1992; Gannon et al., 2005; Hartog et al., 2000, among others). Therefore, it was not possible to build into the simulation the necessary structure to create inter-industry wage differentials in the data. And that prevents effective evaluation of the impact of industry fixed effects on the estimates of MWTP.

The impact of including only occupation fixed effects is understood by comparing row (3) and (9) and the comparison of row (7) and (11). It appears that including occupation fixed effects reduces estimates of MWTP substantially. This result is not surprising. Recall that the omitted variables are the O*NET job attributes which are measured by occupation

only and the simulated "true" wage equilibrium is constructed in the way that workers care about many job attributes including fatal risk. The reduction of the estimates from the model with occupational fixed effects implies that the omission of these O*NET job attributes possibly introduces an upward bias into the models.

The box-plot in Figure 5.2 illustrates the general pattern of the bias distribution across functional forms with result to occupational and/or industry fixed effects. Linear functional form and LBC functional forms outperform all other functional forms with smaller boxes and narrower whiskers. In particular, when the occupational fixed effects are included, the biases from most of linear models are densely distributed around median value which is close to zero. While the LBC functional form does not perform as well as the linear functional form, it does better than the other functions. When including occupational fixed effects (with or without industry fixed effects), double-log and semi-log with quadratic risk functional forms perform poorly with the double-log performing worst. In particular, the long whisker reflects that the double-log functional form tends to produce biases that are spread out across a wider range. The shifted boxes indicate the distribution is skewed to the right which implies the double-log function tend to produce exceptionally large downward biases.

As shown in table 5.3, applying occupation and industry fixed effects to linear cross-section model does move the average bias very close to zero. However, most of the single cross-sectional models are still suffering from large bias values, regardless of the fixed effect methods. It implies that with presence of potential omitted variable biases caused by

unobserved worker's characteristics or preferences, simply adding fixed effect are not sufficient enough to improve models' performances.

5.4.3 Effect of panel models

In this section, I examine the performance of panel data strategies for estimating the MWTP. Table 5.4 reports the bias in MWTP from panel data estimations. The results from the fully observed scenario are displayed in Panel A. These results are based on the models without any omitted job attributes and including no occupation and/or industry fixed effects. Compared to the average bias in single year cross-section models (Table 5.3, Panel A), the pooled OLS estimator produces substantially larger estimates of MWTP than single one year cross-sectional estimator. This result is supportive of previous studies that suggest when panel data is used as a cross-section data, the estimates of the MWTP tend to be large (Viscusi and Aldy 2003). Moreover, the increase in bias is especially sizable for semi-log and double-log functions. The comparison of the average bias from single cross-section models and the pooled OLS models reflects the inability of pooled estimation to account for changes in the shape of the wage function.

The comparison of row (1), (3), (5) and (7) indicates that the FD estimator with a wider time interval (Δ=2) results in the smallest absolute value of the bias and pooled OLS performs worst. While the FE estimator produces a larger absolute value of the bias as compared to FD (Δ=2), it performs significantly better than FD (Δ=1) regardless of functional form. The large difference of the average bias values from the FD (Δ=1) estimator and the FD (Δ=2) estimator could be caused by serial correlation of the model errors. Recall

that in generating the panel data, the worker's utility and preferences from the simulation based on 2004 wages and non-wage income are used as the starting value for the 2005 simulation, and the same practice is applied to 2006 simulation. During this process, any new wage/job combination must not decrease utility as compared to the baseline utility in the previous year. The difference in simulated wages and utility over the years is thus driven by the change of worker's preferences which are the products of unobservables in the regression models. As presented in chapter 4, the pattern of change in wages from 2004 to 2005 leads to the same pattern of changes of the wages from 2005 to 2006. For example, if a worker's wage increases from 2004 to 2005, there is also an increase in his wage from 2005 to 2006. Therefore, it is possible that there is a correlation between the change of the unobservables (e.g. unobserved preferences), $\Delta u_{i2005-2004}$ and $\Delta u_{i2006-2005}$, and this autocorrelation could bias the FD (Δ=1) models.

Now consider the omitted variable scenario. Panel B presents the mean and standard deviation of the bias from four estimators. In these estimation models, occupation and industry fixed effects are not included. The presented results are averaged over models with different choices of omitted variable groups.

The effect of the omitted variables varies across the panel data methods. Compared to the full-observed scenario, it appears that the pooled OLS model produces higher estimate of the MWTP for five out seven functions. This effect is contrary to the ones for the FE and FD models, in which the omitted variables reduce the estimates of MWTP substantially, regardless of the functional forms. While this decrease hurts the performance of the FE and

FD (Δ=1) models, it is sufficiently large enough to offset upward bias produced by FD (Δ=2) and moves the estimates closer to the true MWTP (see row (15), panel B).

The relative performance of FE, FD (Δ=1) and FD (Δ=2) are similar to those in the fully observed scenario with FD (Δ=2) performing best and the pooled OLS performing worst. The FE performs better than the pooled OLS, but worse than FD (Δ=2) and the omitted variables clearly hurt the performance of this model. Except for the two semi-log functions, the results from FD (Δ=1) are comparable to the FE estimator. As in fully observed scenario, the FD (Δ=2) estimator perform well, regardless of the functional forms. With the presence of omitted variables, the lowest average bias produced by the FD (Δ=2) estimator is 2%. And the highest average bias produced by the FD (Δ=2) estimator is 6%.

I present the general pattern of bias distribution across the functional forms with the effect of the panel data methods in box-plots in Figure 5.3. The plot is based on the results reported in Table 5.4, Panel B. The distribution of the biases reinforces the findings described above. In particular, the biases produced by FD (Δ=2) estimator are densely distributed around zero, regardless of the functional forms. The wide spread of the boxes from the FE, FD and the pooled OLS estimators implies the estimates from semi-log model are very sensitive to the choice of the estimators.

For all of the panel data estimators, adding quadratic risk to the function appears to exaggerate the biases substantially. The pattern is especially clear for pooled, FE and FD (Δ=1). For instance, the position of the boxes shows that for the FE and FD (Δ=1) estimator, the majority of the linear with quadratic risk models produce more downwardly biased

estimates than the ones without quadratic risk. And for the pooled OLS, the majority of the linear with quadratic risk models produce more upwardly biased estimates than the ones without quadratic risk. The same pattern exists in the semi-log and Box-Cox models. Moreover, adding a quadratic risk term to the model increases the instability of the estimates to the modeling choices. This finding is illustrated by the comparison of the length of whiskers, the wider spread of the boxes across the estimators, and the size of the boxes.

The key results in the previous sections can be summarized as follows. First, in the fully observed scenario, with all the estimators in consideration, the LBC and QBC functional forms outperform others. The results are consistent with the findings in CDM and KPP that the more flexible functional forms (LBC and QBC), outperform all other functional forms when there are no omitted variables. Second, in the models where variables are omitted, simple linear OLS and LBC functional forms tend to perform best. The double-log and semi-log functional forms consistently do poorly in the presence of omitted variables. Third, adding a quadratic risk term to a function generally increases the vulnerability of the model to omitted variables and also increases sensitivity to the econometric strategies employed. The effect of including a quadratic risk is especially pronounced with the semi-log functional form, decreasing their performance substantially. This finding is troubling given how commonly these functional forms are used in the hedonic wage literature (Mrozek and Taylor, 2002; Kniesner and Leeth, 2009).

The fourth finding is that the impact of occupation and/or industry fixed effects on the estimates of the MWTP varies across the functional form. Inclusion of these fixed effects

appears to improve the performance of LBC functional form, but worsen the performance of the double-log functional form. Moreover, in the cross-sectional estimator, the semi-log and Box-Cox functional forms are particularly sensitive to the different specification of occupation and/or industry fixed effects, especially when a quadratic risk term is included. Finally, in examining the performance of panel models, I find that the FD model with wider time interval (Δ=2) outperforms other panel data models. It appears that regardless of the functional forms, applying FD (Δ=2) estimator effectively reduces the bias and significantly improve the accuracy of the estimates of MWTP.

While the comparisons presented in the previous sections highlight key results, the scope for these comparisons is limited. The comparison is based on a selected subset of the results. As pointed out in Learner (1983), a selective approach to summarizing a specification search can invalidate the statistical properties of the research design. To address these concerns, I further adapt the meta-analytic approach to summarize the results.

Meta-analysis has traditionally provided a simple and informative way to summarize results across a large number of distinct studies that investigate a common problem. Banzhaf and Smith (2007) first extend meta-analysis into the context of a single study to summarize how subjective modeling decisions influence economic predictions. Their internal approach to meta-analysis estimates a large number of candidate specifications for the same underlying model where each specification is defined by a set of indicator variables describing its econometric features. The advantage of this approach is that regressing economic outcomes

on indicators for the features of the econometric specification can help to summarize the ways in which the researcher's modeling decision influences their findings.

5.4.4. Internal Meta-Analysis results

In this section, I follow Banzhaf and Smith (2007) and KPP to adapt the meta-analytic approach to synthesize the full set of results, while simultaneously allowing for broader comparisons across specifications. This meta-analytic approach involves regressing model outputs on indicator variables for model inputs. The estimated coefficients summarize the ways in which subjective modeling decisions contribute to the performance of the model. Since the goal of meta-analysis in my application is to examine which model specification leads to smallest bias, the focus is on the magnitude of the bias instead of the direction of the bias. Therefore, in my application the value for the normalized welfare bias, $|\beta|$, is regressed on indicator variables for: (1) functional form choices, (2) the presence of industry and occupational dummy variables in the model, (3) the type of estimator (OLS, FD and FE), and (4) the omitted job attributes cases.

Table 5.5 presents the meta-analysis results. The regression uses 3,370 observations on the absolute bias in measuring MWTP in the presence of omitted variables.[40] Since all of the regressors are indicator variables, interpretation of results depends on the reference set of modeling choices defined by the excluded indicators. The reference model is a single cross-section OLS model using a linear functional form without any industry or occupation fixed effects.

[40] I dropped 210 observations with biases greater than 500%. While dropping these observations does not change the overall pattern of results, it leads to a substantial improvement in model fit.

Panel A of Table 5.5 indicates that the linear functional form performs best, ceteris paribus. For instance, the coefficient for a linear model with quadratic risk specification is 0.91, which indicates that adding quadratic risk variable to the linear model produces estimates for the MWTP with 91% more bias than the baseline linear model, all else equal. The double-log and the semi-log with quadratic risk perform the worst. With all else constant, the double-log function produces estimates for MWTP with about 118 % more bias than the baseline linear model. And the semi-log with quadratic risk produces estimates for the MWTP with about 101 % more bias than the baseline linear model. While the LBC functional form appears to produces estimates for the MWTP with less bias than the baseline linear model, the coefficient is not statistically significant at any significance level. Although the coefficient of QBC functional form indicates that it performs better than linear functional form with quadratic risk term, semi-log and semi-log with quadratic risk term, it produces about 55% more bias than the baseline linear model.

The set of occupational indicator variables in Table 5.5, Panel B describes the effect of inclusion of the job attributes on the performance of the estimation models. The coefficient for each of five groups of non-pecuniary job attributes shows that they are all preferable to be included in the model. In particular, the inclusion of *exposure, motor skill variable,* and *people skills* variables has significant impact on improving the model's performance. The models with these variables tend to produce the MWTP for fatal risk with 7% to 17% less bias than the models without controlling for these non-pecuniary jobs attributes.

Another set of indicator variables in Table 5.5, Panel C describes the relative performance of different cross-section and panel data estimators. Compared to the baseline cross-section OLS estimator, the FD (Δ=2) performs best, producing estimates for the MWTP with 119% less bias than the baseline single cross-section model. The FE reduces bias by about 21% less bias than the baseline model. In comparison, the FD (Δ=1) makes the estimates worse. The coefficients indicate that the FD (Δ=1) produces estimates for MWTP with 37% more bias than the baseline model, with all else equal. While it appears that the pooled OLS produces less bias than the baseline model, the improvement of the pooled OLS is not statistically significant.

Finally, the coefficients for variables in Table 5.5, Panel D indicate that controlling for both occupational and industry fixed effect to the model are preferable strategies for addressing correlated omitted variables. They produce 6% less bias than the baseline model without fixed effects control. The impact of only controlling for industry fixed effects or occupation fixed effects on the bias, however, is not statistically significant.

In sum, the results from meta-analysis indicate that linear functional form with both linear wage and linear fatal risk is the most preferable functional form. The presence of both industry and occupational dummy variables in the model has significant impact to reduce the bias, and FD (Δ=2) is the most preferable type of estimator. The inclusion of non-pecuniary job attributes does improve the accuracy of the estimates. Among five groups of the job attributes considered in the analysis, job conditions related to exposure, motor skills and

demand of interacting with people skill have significant effects on reducing the bias of the estimates.

Based on the results from meta-analysis reported in table 5.5, I will estimate the VSL using observed labor force data to compare to the previous studies. In comparing the literature, I attempt to find: i) where my estimates are situated in the literature, once I consider full set of modeling choices suggested in table 5.5, ii) how sensitive the VSL estimates is to the occupation and/or industry fixed effects when the improved risk data, more preferred estimators and functional form are applied to the estimation model, and iii) how the inclusion of the O*NET affects the VSL estimates, since the O*NET job attributes have not been incorporated in the previous hedonic wage studies.

To answer the first question, I will estimate the VSL by using a linear FD (Δ=2) model with both occupation and industry fixed effect controls. The model will also include *exposure, demand of motor skill,* and *people skill* O*NET job attributes in the covariates. In an attempt to answer the second and third questions, I will provide a sensitivity analysis of the estimation results by estimating additional set of models that have different occupation and/or industry fixed effects. And these models will vary by inclusion of O*NET job attributes suggested in table 5.5.

5.5 Conclusion

This chapter presents the results from an evaluation analysis. The process begins by estimating 3,580 hedonic wage models that vary by: i) the functional form of the hedonic wage model; ii) the number and the presence of occupation and/or industry fixed effects in

the regression; iii) the approach used for cross-section OLS estimation; iv) the approaches used for panel data estimation; and vi) the omitted job characteristics. The marginal value of risk reductions obtained from each hedonic wage regression is compared to the "true" MWTP obtained for each worker in my simulated wage equilibrium. The difference in these two values is referred to as the "bias" from the empirical model. The performance of each hedonic wage model is indicated by the bias summarized over the observations of bias in each model.

In the evaluation, both the general comparison pattern and the internal meta-analysis results indicate that with all the econometric strategies considered, it is the linear functional form that outperforms all other functional forms in all scenarios in which one or more covariates are omitted from the regression. Given that semi-log functional form is the most commonly used in the literature, it is surprising that the semi-log functions perform the second-worst of all forms of the hedonic wage function in the evaluation. The semi-log functions are especially likely to produce the marginal price of risk biased, on average, by an order of magnitude. The large standard deviation of the bias from the semi-log hedonic wage model further implies that it is highly sensitive to other modeling choices such as the econometric strategies or the inclusion of the non-wage job characteristics in the covariates.

The compensating wage differential theory provides no guidance on appropriate functional forms. A semi-log specification implies that the compensating wage differentials will increase proportionately with wage levels. Since the "true" MWTP is computed in the simulation, I can graph the "true" MWTP to against wage to explore the correlation between

compensating wage differentials and the wage. Figure 5.4 Panel A, B, and C present scatter plots of worker's MWTP against wage based on the 2004, 2005 and 2006 simulation results respectively. In the plots, workers are ranked from lowest to highest by hourly wage. As displayed in panel A, B and C, most of MWTP computed in the simulation are randomly located between 0 and 1. In all three plots, there appears to be little correlation between MWTP and wage, which implies that worker's MWTP for risk is independent of wage level and marginal price of risk is also independent of the level of wage. Thus, these plots are consistent with the findings in the evaluation analysis that linear functional form is a more reasonable approximation to the hedonic wage locus.

A natural concern, however, is that the pattern of evaluation results may simply reflect the shape of the utility function chosen for the analysis: Cobb-Douglas utility. Future work could consider using different parametric forms of utility function in the sorting algorithm to provide robustness checks of the evaluation results.

Table 5.1

Variables included in a fully specified regression model for evaluation

Variable name	Definition
Worker Characteristics	
hr_wage	hourly income
fatal risk	Workplace fatal risk rate per 10,000 workers
non-fatal injury risk	Non-fatal injury risk rate per 10,000 workers
age	age in years
blacknh	1 if individual is black and non-hispanic
college	1 if individual attended college but with no degree
female	1 if individual is female
hispanic	1 if individual has a hispanic origin
hsgrad	1 if individual graduated from high school
kid18	1 if individual has any kids under 18 years old
married	1 if individual is married
othrace	1 if individual is non-white, non-black, non-hispanic
ugdeg	1 if individual have bachelor degree or more
union	1 if individual is convered by a union contract
region1/9	individual lives in New England Division
	individual lives in Middle Atlantic Division
	individual lives in East North Central Division
	individual lives in West North Central Division
	individual lives in South Atlantic Division
	individual lives in East South Central
	individual lives in West South Central
	individual lives in Mountain
	Individual lives in Pacific
Job Characteristics from labor force data	
salaried	1 if individual gets monthly payment
workot	1 if individual usually works more than 40 hours
union	1 if individual is a union member or covered by union
*O*NET occupational characteristics*	

Table 5.1 Continued

exposure	e1	low level of hazardous work environment
	e2	providing working environment exposed to the weather or hot, cold temperature
	e3	high level of hazardous working environment, specifying technical or electronic or mechanical device or equipment
cognitive	c1	cognitive demands involving information processing and analyzing data or information
	c2	cognitive demands involving creative thinking, strategies developing, planning, and decision making
	c3	cognitive demands involving monitoring process, inspecting equipment or material
physical skill	p1	physical demands requiring physical activities
	p2	requires climbing ladders, poles, or kneeling and crawling
	p3	requires repetitive physical motion, operating machine or handling tools
motor skill	m1	specifying technical or electronic or mechanical device or equipment
	m2	interacting with computer
people skill	p1	requires interacting with people indirectly
	p2	interacting with people directly, or involving in the conflicting situations
	3	requires group or team working and taking responsibility for the outcome of the performances
	4	Working directly with public. Each of the jobs is differentiated by the job attributes and the injury risk
Major industry dummy variables		
ind1		Mining, utilities and construction
ind2		Food processing, textiles and equip manufacturing
ind 3		Retail and transportation
ind4		Fire, telecoms and publishing
ind5		education , health and social services
ind6		Entertainment and hotels
ind7		Personal, professional, civic and religious services (except Public Administration and private housholds)
Major occupation dummy variables		

Table 5.1 Continued

Occ1	Management, Business and Financial, Professional and Related
Occ2	Services
Occ3	Sales and Related, Office and Administrative Support
Occ4	Farming, Fishing, Forestry, Construction and Extraction, Installation, maintenance, and Repair
Occ5	Production, Transportation and Material Moving

Table 5.2

General results from models in fully observed scenario and omitted variable scenario

		Linear	Linear (r2)	Semi-log	Semi-log (r2)	Double log	Linear Box-Cox	Quad. Box-Cox
Panel A: All models (fully observed scenario) (# obs. of bias e_p: 54,868) (# of models: 29)								
(1)	β	0.25	-0.55	-1.54	-0.71	-1.29	0.11	-0.20
(2)	Std.	0.45	1.02	2.54	1.81	1.11	0.17	0.16
(3)	max	|-2.41|	|-1.80|	|-5.92|	|-3.64|	|-2.96|	|-0.01|	|0.08|
(4)	min	|0.93	|0.69|	|0.74|	|0.72|	|-0.61|	|0.24|	|0.32|
Panel B: Cross-section model (omitting all O*NET variables) *(# obs. of bias e_p: 51,100) (# Of models: 28)*								
(5)	β	-0.07	-1.43	-1.97	-0.93	-2.00	-0.21	-0.76
(6)	Std.	0.23	0.73	0.11	1.01	1.12	0.32	0.74
(7)	max	|-0.30|	|-2.54|	|-1.25|	|-2.32|	|-3.96|	|-0.65|	|-1.77|
(8)	min	|0.30|	|-0.34|	|0.18|	|0.46|	|-0.45|	|0.32|	|0.28|
Panel C: Panel data models (omitting all O*NET variables) *(# obs. of bias e_p: 163,976) (# of models: 88)*								
(9)	β	-0.11	-0.77	-1.05	-0.36	-1.56	0.41	-0.52
(10)	Std.	0.53	1.20	2.40	3.09	1.92	0.31	0.50
(11)	max	|-0.95|	|-2.47|	|-7.26|	|-10.27|	|-7.11|	|0.85|	|1.24|
(12)	min	|0.79|	|1.15|	|2.10|	|4.81|	|1.10|	|-0.20|	|-0.37|
Panel D: All models (omitting all O*NET variables) *(# obs. of bias e_p: 215,076) (# of models: 116)*								
(13)	β	-0.10	-0.91	-1.04	-0.48	-1.67	0.10	-0.11
(14)	Std.	0.49	1.15	2.14	2.80	1.76	0.44	0.90
(15)	max	|-0.95|	|-2.54|	|-7.26|	|-10.27|	|-7.11|	|0.85|	|-1.77|
(16)	min	|0.79|	|1.15|	|2.10|	|4.80|	|1.10|	|-0.65|	|1.24|
Panel E: All models (omitting O*NET variables randomly) *(# obs. of biase_p: 6,452,280) (# of models: 3,480)*								
(17)	β	-0.08	-0.82	-1.34	-1.12	-2.00	0.01	-0.15
(18)	Std.	0.51	1.08	2.27	3.32	1.94	0.48	0.85
(19)	max	|-1.29|	|-4.29|	|-7.85|	|-12.70|	|-7.86|	|-1.43|	|-3.19|
(20)	min	|1.21|	|1.78|	|3.15|	|11.00|	|1.50|	|1.31|	|1.79|

Table 5.3

Results from single cross-section models with different fixed effect

		Linear	Linear (r2)	Semi-log	Semi-log (r2)	Double log	Linear Box-Cox	Quad. Box-Cox
		Panel A: Fully observed scenario (including no fixed effects)						
(1)	β	0.25	-0.75	-0.96	0.03	-1.23	0.15	-0.07
(2)	Std.	0.11	0.61	0.12	0.94	0.86	0.28	0.66
		Panel B: Omitted variable scenario (including no fixed effects)						
(3)	β	0.11	-0.45	-0.96	0.31	-1.00	0.08	0.12
(4)	Std.	0.31	0.52	0.08	0.49	0.45	0.26	0.39
		Panel C: Omitted variable scenario (with all fixed effects)						
(5)	β	-0.01	-1.41	-0.97	-0.83	-2.29	-0.15	-0.69
(6)	Std.	0.37	0.97	0.15	1.22	1.26	0.45	0.93
		Panel D: Omitted variable scenario (different fixed effect choices)						
(7)	β (Ind.) [a]	0.17	-0.52	-0.95	0.50	-0.92	0.26	0.28
(8)	Std.	0.36	0.63	0.10	0.63	0.50	0.32	0.50
(9)	β (Occ.) [b]	-0.12	-1.78	-0.98	-1.67	-3.17	-0.47	-1.30
(10)	Std.	0.33	0.76	0.10	0.77	0.89	0.32	0.62
(11)	β(Occ_ind.) [c]	-0.08	-1.93	-0.97	-1.34	-2.79	-0.25	-1.07
(12)	Std.	0.36	0.83	0.11	0.88	0.92	0.35	0.70

a. The average bias is reported for the models controlling for only industry fixed effects.
b. The average bias is reported for the models controlling for only occupation fixed effects.
c. The average bias is reported for the models controlling for both occupation and industry fixed effects.

Table 5.4

Comparison of results from panel models

		Linear	Linear (r2)	Semi-log	Semi-log (r2)	Double log	Linear Box-Cox	Quad. Box-Cox
		Panel A: fully observed scenario [a]						
(1)	pooled OLS	0.37	0.8	2.86	5.65	3.6	0.62	1.6
(2)	Std	0.44	0.89	1.23	1.38	1.05	0.47	0.7
(3)	FE	-0.49	-1.08	-0.54	-0.6	-0.82		
(4)	Std	0.09	0.23	0.05	0.12	0.08		
(5)	FD ((Δ=1)	-0.69	-1.31	-3.87	-4.23	-1.06		
(6)	Std	0.01	0.06	1.25	2.5	0.04		
(7)	FD (Δ=2)	0.17	0.12	0.2	0.16	0.13		
(8)	std	0.14	0.15	0.10	0.11	0.10		
		Panel B: omitted variable scenario (without fixed effect)						
(9)	pooled ols	0.28	0.95	2.37	6.9	2.28	0.51	1.09
(10)	std.	0.41	0.98	0.88	2.54	1.33	0.43	0.73
(11)	FE	-0.69	-2.03	-1.13	-1.78	-1.34		
(12)	std.	0.2	0.44	0.24	0.44	0.29		
(13)	Fd (Δ=1)	-0.89	-2.3	-4.72	-0.26	-1.32		
(14)	std.	0.2	0.48	1.69	3.51	0.26		
(15)	Fd (Δ=2)	-0.02	-0.03	-0.03	-0.06	0.02		
(16)	std.	0.21	0.21	0.2	0.2	0.18		

a. In fully observed scenario, models do not vary by choice of omitted variables or fixed effects. Thus, the normalized standard deviation of bias only reflects the dispersion of bias across the workers.

Table 5.5

Meta-Analysis of hedonic results [a]

Modeling choice	Coef. (Std. Err.)	[95% Conf.	Interval]
Panel A: Functional form indicator (*excluded indicator: linear*)			
Linear (r2)	0.91*** (0.07)	[0.84	0.97]
Semi-log	0.63*** (0.12)	[0.54	0.72]
Semi-log (r2)	1.01*** (0.04)	[0.92	1.09]
Double-log	1.18*** (0.05)	[1.07	1.30]
LBC	-0.06 (0.05)	[-0.16	0.04]
QBC	0.55*** (0.07)	[0.41	0.69]
Panel B: Occupational attributes indicator			
Exposure	-0.07** (0.03)	[-0.13	-0.016]
cognitive	-0.01 (0.03)	[-0.07	0.04]
Physical demand	-0.04 (0.03)	[-0.09	0.01]
Motor skill	-0.17*** (0.03)	[-0.23	0.11]
People skill	-0.12*** (0.03)	[-0.17	-0.06]
Panel C: Estimation indicators *(excluded indicator: cross-section OLS)*			
Pooled OLS	-0.17*** (0.32)	[-0.27	-0.08]
Fixe effect	-0.20*** (0.04)	[-0.28	-0.12]
First difference Δ=1	0.37*** (0.07)	[0.25	0.50]
First difference Δ=2	-1.19*** (0.04)	[-1.27	-1.11]

Table 5. 5. Continued

Panel D: Occupation/ Industry Fixed effect indicator *(excluded indicator: no fixed effect)*		
Occ_ind [b]	-0.06***	[-0.14 0.15]
	(0.01)	
Occ_dummy_only[c]	0.02	[-0.06 0.10]
	(0.64)	
Ind_dummy_ondly [d]	-0.07	[-0.15 0.01]
	(0.04)	
N	3,370	
R^2	47.24%	

a. Legend: *** significant at the 1% level, **significant at the 5% level, *significant at the 10% level.
b. Occ_ind: including both seven occupation and nine industry fixed effects.
c. Occ_dummy_only: including only seven occupation fixed effects.

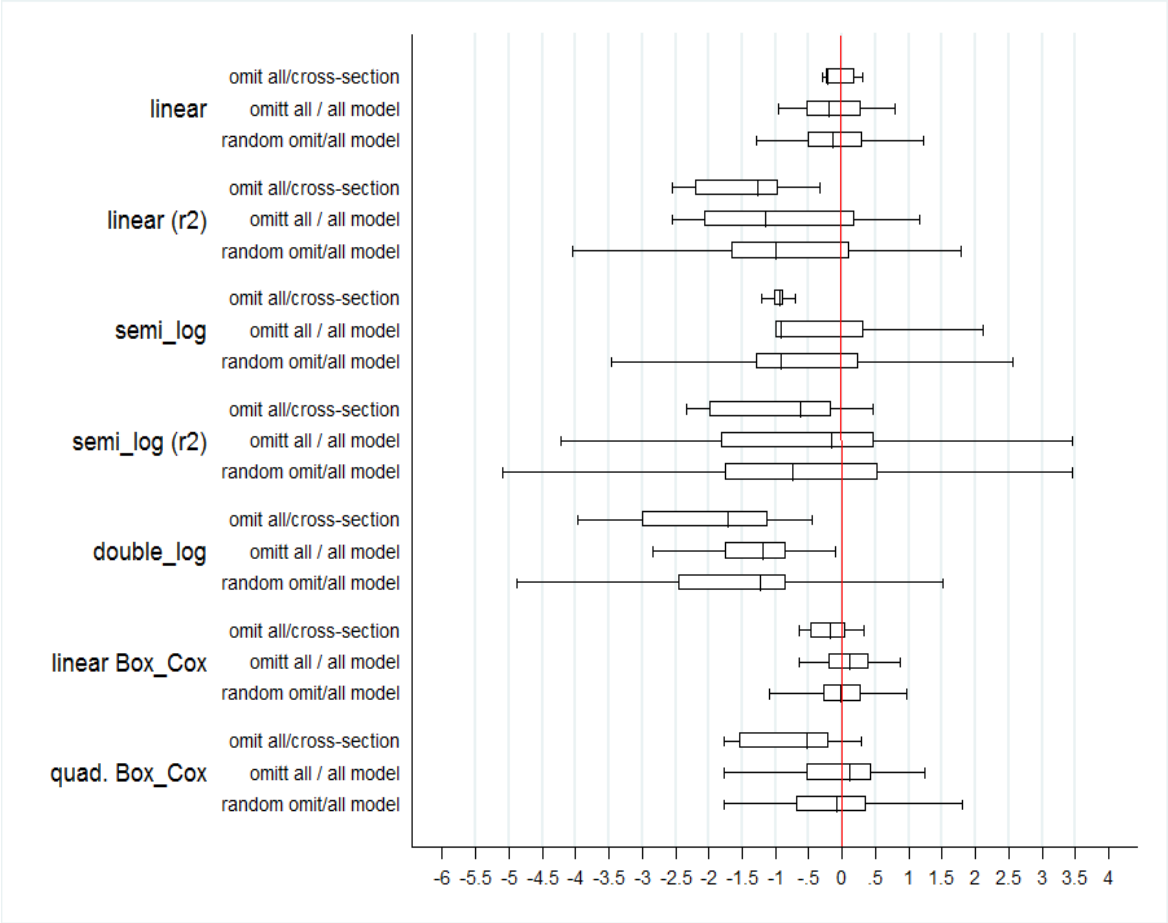


Figure 5.1 *General comparison of bias distribution*

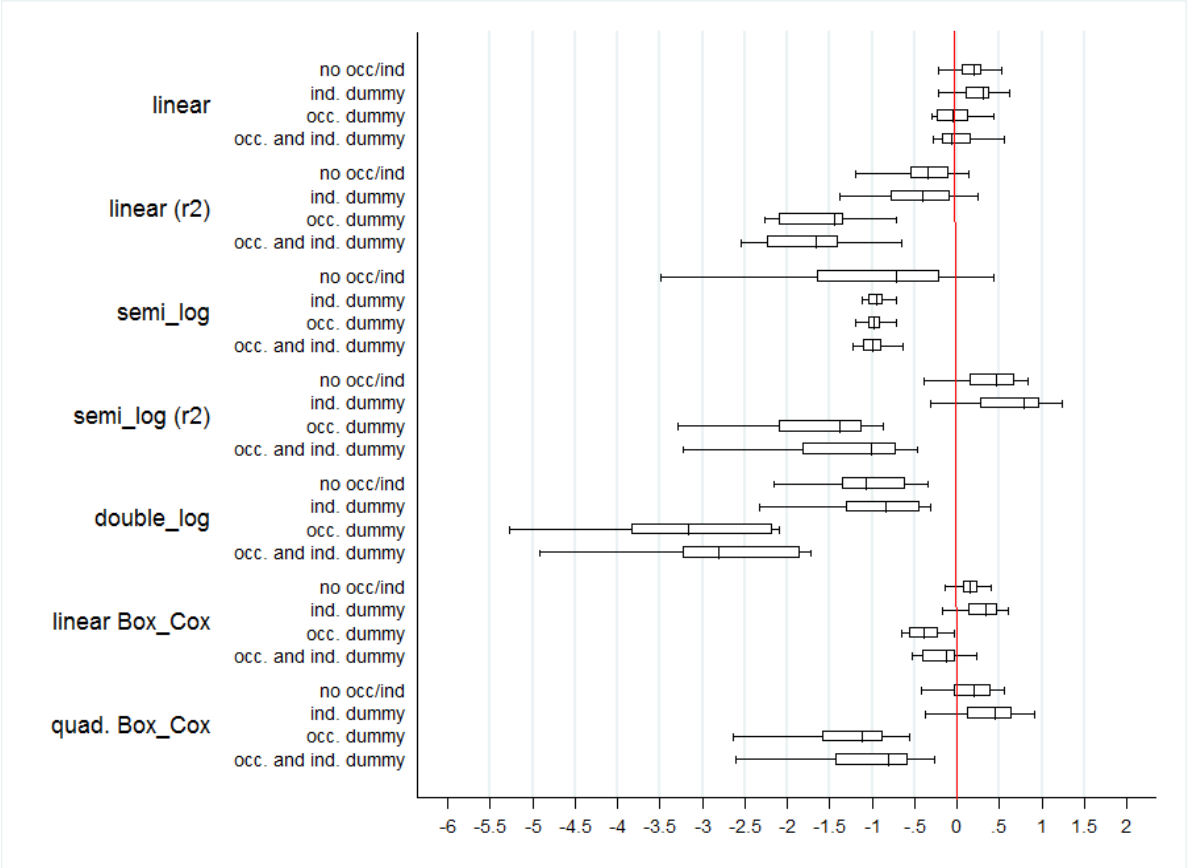


Figure 5.2 *Distribution of biases produced by single cross-section models*

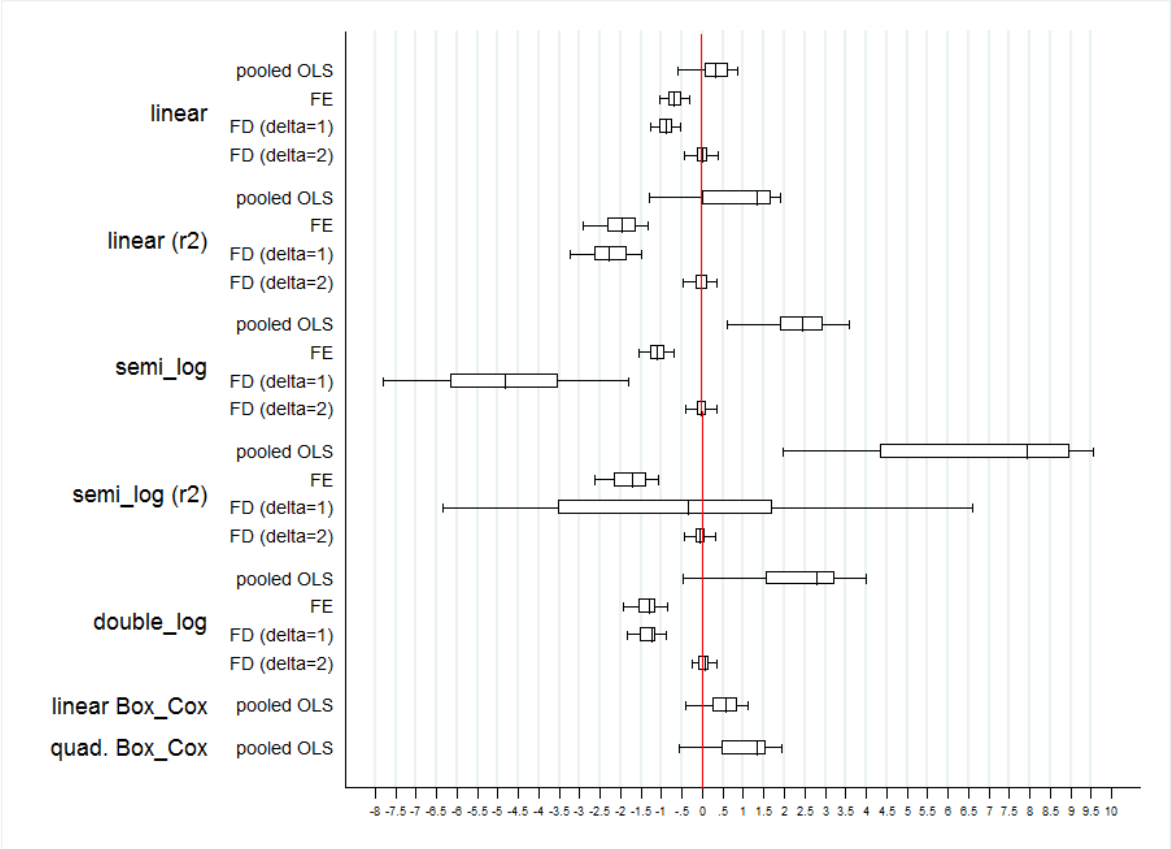
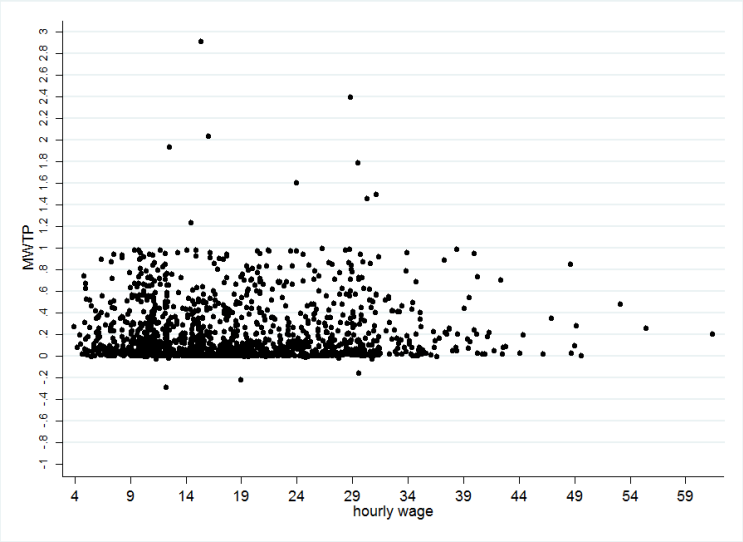
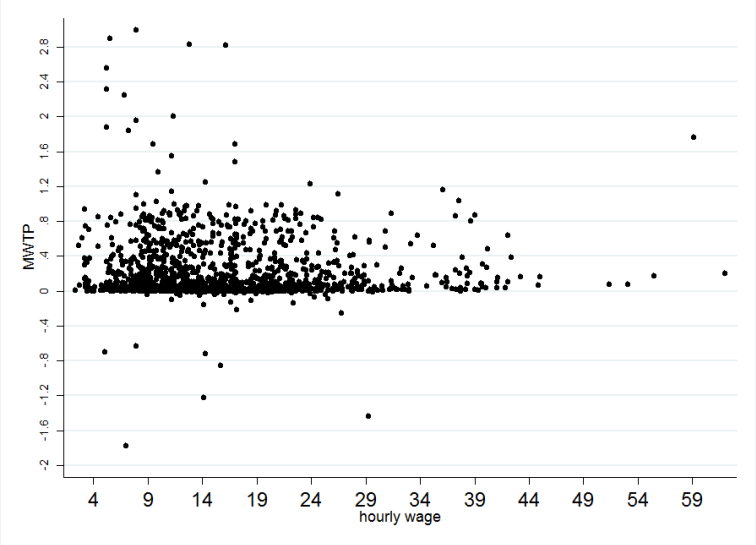


Figure 5.3 *Distribution of biases produced by Pooled OLS and panel data models with no fixed effect*

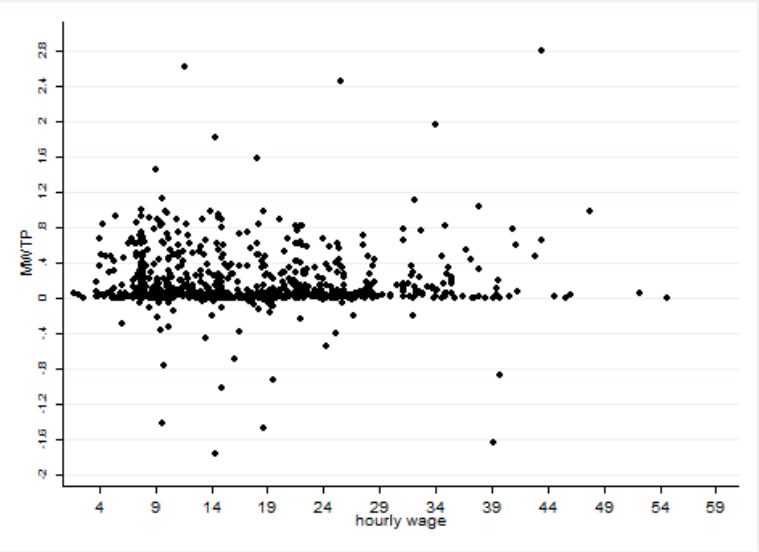
Figure 5.4 *Scatter plot of worker's MWTP against income from 2004-2006 simulation result*



A. Scatter plot based on 2004 simulation results



B. Scatter plot based on 2005 simulation results



C. Scatter plot based on 2006 simulation results

CHAPTER 6

CONCLUSION

Hedonic wage studies have produced a wide range of VSL estimates. Some of this variation can be expected on theoretical grounds, while some is due to the variations in model specifications. Researchers are confronted by two problems in hedonic wage analysis: the measurement error in fatal risk data and the omitted variable issue. In the early studies, these two issues are inter-related, as most of the early studies use risk measures at aggregated industry level which induces a correlation between the risk measure and wage differentials associated with different jobs or industries that are not due to risk difference. Differences in the treatment of these problems have led to vastly different hedonic wage results and implied VSL estimates. However, the inter-relationship between measurement error in risk data and omitted variables in earlier studies makes it impossible to understand and identify the effect of different correction methods applied in model specification and econometric strategies.

The previous chapters in this dissertation sought to use a simulation approach to systematically evaluate hedonic wage modeling choices and identify the influence of functional form and the econometric choices on the accuracy of the MWTP estimates. First, I proposed a sorting model to solve a job-worker matching problem numerically and provide an analytical "true" hedonic wage equilibrium. The sorting model manifests compensating wage differentials theory and the generalized Roy Sorting model' logic by translating information on workers and their choices over jobs into parameters that represent their heterogeneous preferences for occupational attributes. Given the worker's parameterized

utility function, the resultant wage and job attributes that workers sort into as the market equilibrium occurs, the sorting model provides complete information to compute worker's true MWTP for fatal risk. In the second stage of the analysis, a large set of reduced-form hedonic wage models are estimated with the simulated data. These reduced-form models vary by the model specifications such as the functional form and econometric strategies. The estimated marginal price of risk from the reduced form models are compared to the true MWTP for risk to explore the models' vulnerability to the omitted variable issue. The simulation model is used to answer the following questions: i) whether the commonly employed specifications in the past studies are appropriate, 2) how the different combination of modeling choices influence the accuracy of the estimates of MWTP for risk in the presence of omitted variable issues. The simulation model extends hedonic wage literature on the model selection approaches. The key findings from the analysis provide insights on the ability of commonly used hedonic wage models to recover the MWTP for workplace fatal risk. Importantly, the results from simulation model helps to better understand the sources of variation presented in the VSL estimates.

6.1 An alternative model selection approach

In hedonic wage studies, maximum likelihood (ML) estimation technique has been generally utilized to search for a proper specification of hedonic wage functions (Blomquist et al, 1988; Atrostic, 1982; Atkinson and Halvorsen, 1990; and Cheshire and Sheppard, 1995). Previous studies have used the Box-Cox ML estimation to choose the optimal transformation for the dependent and independent variables, to maximize the probability that

the normality assumption holds and that consistent estimates of the hedonic parameters are produced. (Michaelides, 2008; Lavetti, 2012)

However, in these studies the Box-Cox MLE approach is applied to only address the issues associated with the cross-sectional OLS estimators. Recent empirical hedonic wage studies have increasingly used panel data estimators along with different occupation and industry fixed effects. Large variation of the hedonic wage estimates arise from the models with different econometric strategies. For instance, Kniesner et al (2010) find that controlling for unobserved worker heterogeneity reduces estimates by about sixty percent from estimates obtained using a single cross-section of data. The estimates of the marginal price of risk also present high sensitivity to the inclusion of industry and occupation fixed effect (Cropper et al., 2011). These findings imply that unobserved characteristics may disproportionately affect the performance of parametric specifications for the hedonic wage function when different econometric methods are applied. Therefore, the investigation of the hedonic wage functional forms should not be separated from the context where specific econometric methods are applied. The proposed simulation approach provides alternative means to investigate the model specification more completely. In addition to search for a proper functional form, the simulation model also provides insights on: i) what combination of model specification and econometric methods provide the most accurate estimates for MWTP when some variables are omitted; and ii) which estimators are more robust to misspecification of the estimation function.

6.2 Results from simulation model

The application of the simulation model reveals several important points about the hedonic wage model specifications. The general pattern of performances of seven hedonic wage functional forms (*linear, semi-log, double-log, linear Box-Cox, linear with quadratic fatal risk term included, semi-log functional form with quadratic fatal risk term included, and quadratic Box-Cox),* can be summarized as follows. When all the job attributes are observed, the *linear Box-Cox* and *quadratic Box-Cox* functional perform best with lowest average bias and the smallest standard deviation of bias. This result is consistent with the findings in Copper, Deck and McConnell (1988) and Kuminoff, Pameter and Pope (2010) that more flexible functional forms (linear Box-Cox and quadratic Box-Cox), outperform all other functional forms when there are no omitted variables. In the presence of omitted job attributes, simple linear and linear Box-Cox functional forms tend to perform best, while double-log and semi-log functional forms consistently do poorly in the models with different econometric strategies. The analysis also provides evidence that adding a quadratic risk term to a function generally increases sensitivity to the employed econometric strategies. The effect of including a quadratic risk is especially pronounced with the semi-log functional form, decreasing its performance substantially.

Moreover, the evaluation results also indicate that the impact of econometric strategies varies across different functional forms. Inclusion of occupation and/or industry fixed effects appears to improve the performance of LBC functional form, but worsen the performance of the double-log functional form. The semi-log and Box-Cox functional forms

are particularly sensitive to the different specification of occupation and/or industry fixed effects, especially when a quadratic risk term is included.

Among five estimators that are commonly used in the hedonic wage studies, first-difference with a two-year time difference (FD(Δ=2)) is a more preferable estimator in recovering the true MWTP. FD (Δ=2) model performs especially well when linear functional form is applied.[41]

The pooled OLS estimator tends to overestimate MWTP. The effect of pooled OLS is emphasized when semi-log and double-log functional forms are applied to the models. FE and FD (Δ=1) models tend to underestimate the MWTP. The performance of FE and FD (Δ=1) models are comparable when linear, linear functional form with quadratic risk terms included and the double-log functional forms are applied. However, when the semi-log functional forms (with or without quadratic risk term) are applied, FD (Δ=1) models perform significantly worse than FE models by producing large downward biased and unstable estimates.

In order to synthesize the full set of evaluation results and provide broader comparison of the modeling choices, I undertake an "internal" meta-analysis which suggests a preferable model producing the estimated MWTP for risk with least bias. That is a linear FD (Δ=2) model with both occupation and industry fixed effect controls. The internal meta-analysis also suggests that including *exposure, demand of motor skill,* and *people skill*

[41] The five specifications considered are: single cross-sectional OLS, pooled OLS, Fixed effect, First difference (Δ=1), and First difference (Δ=2).

O*NET job attributes in the covariates tends to reduce the bias of estimated marginal price of risk.

6.3 Comparison to the literature

The hedonic wage literature includes numerous studies that have been summarized in a series of meta-analysis. Each of these meta-analyses constructs a "best estimate" of the VSL. Miller (2000), Mrozek and Taylor (2002) and Viscusi and Aldy (2003) use meta-regression to see how VSL estimates from individual studies vary with study and respondent characteristics. Kochi et al., uses a Bayesian approach to adjust estimates for within and between study variability. About thirty to forty U.S. hedonic wage studies are included in each meta-analysis. The resulting estimates range from $2.0 million to $11.1 million (in 2008 $). However, almost all studies included in these meta-analysis were completed before the improved CFOI data and do not reflect recent advances in econometric modeling.

In recent studies, the VSL estimates from Kniesner et al. (2011) which are based on the cross-sectional model are between $14 and $32.2 million. Once they apply panel models and control for various endogeneity factors, they obtained VSL estimates which are between $4 to $11 million. In Kochi (2011), their 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.

In attempt to see where my estimates can be situated in the literature once the simulation informed approach is applied, I estimate the VSL in this section by using a linear FD (Δ=2) model with both occupation and industry fixed effect controls. The model will also

include *exposure*, *demand of motor skill*, and *people skill* O*NET job attributes in the covariates. In addition to estimate the model with full set of specifications suggested by the evaluation analysis, I will provide a sensitivity analysis of the estimation results by estimating additional set of models that include different occupation and/or industry fixed effects, and the O*NET job attributes suggested in table 5.5. The sensitivity analysis aims to reveal: i) how sensitive the VSL estimates is to the occupation and/or industry fixed effects when the improved fatal risk data is utilized, more robust estimators and functional form are applied to the estimation model, and ii) how the inclusion of the O*NET affects the VSL estimates as the O*NET job attributes are usually not accounted for in the previous hedonic wage studies.

In order to provide a complete comparison of the estimates based on the model specification suggested by simulation results, to the estimates in the literature, I estimate the VSL by using both the SIPP and Panel Study of Income Dynamic (PSID). These two panel data have been commonly used in recent studies to provide point estimates of VSL (Viscusi and Aldy, 2003; Viscusi et al., 2007; Kniesner et al., 2010; and Kochi, 2010).

6.3.1 VSL estimates with SIPP data

The SIPP data used for VSL estimate is the same data set that the simulation in chapter 4 is based on. It is a balanced panel from 2004-2006 with a total of 5,676 observations for 1,896 hourly paid workers.[42] In the estimation models, the dependent variable is gross hourly wages. The independent variables include the worker's

[42] The sample of workers is limited to the ones who have constant observations over 2004, 2005 and 2006, so that worker's job choices over the years can be modeled in the simulation.

characteristics, such as age, educational attainment, gender, race, marital status, number of children under 18 in the household, union status, and the worker's occupation/ industry category of the firms for which the worker works. In addition, the independent variables also include job attributes that are available in the SIPP labor force data such as the size of the firm, and the worker's insurance status. These independent variables are included in the hedonic wage models in recent panel studies using the SIPP data (Kochi, 2011). Finally, this estimation model incorporates the O*NET job attributes which improves the model's performances suggested by the meta-analysis. The full set of variables used in estimation is presented in appendix 6.A.

6.3.1.1 Estimation results

Table 6.1 reports the estimation results using the preferred type of modeling specification (linear FD (Δ=2) model with inclusion of both occupation and industry fixed effects). For succinctness, only risk coefficients and the corresponding estimates of the VSL are reported. The full set of estimation results are reported in appendix 6.A, column (1). The coefficient for the risk variable is positive and statistically significant at the 5% level. The point estimate of VSL implied by the coefficient for the annual fatality rate is $0.54 million (in 2008 $) with a 95% confident interval of $0.18 to $ 1.08 million (in 2008 $). This point estimate is substantially smaller than the panel model estimates in Kochi (2010) which are between $1.9 to $2.8 million. However, the 95% confidence intervals of her estimate is between $0.43 to $4.03 million (in 2008 $), which overlaps with the 95% confident interval of my estimates.

6.3.1.2 Sensitivity analysis

Model 1 through 4 in Table 6.2 includes O*NET job attributes, while they are excluded in model 5 through 8. All models use the balanced SIPP panel which has 3,760 observations for 1,880 hourly paid workers. Model 1 through 4 varies by the presence of occupation/ industry fixed effects, and the same pattern applies to model 5 through model 8. All the estimation models still use a linear FD (Δ=2) estimator. The full set of estimation results from model 1 through model 8 is reported in Appendix 5.A, column (2) to (4). The estimates from five out of eight models are positive and significant at 5% level.

The results indicate that the estimated wage-risk premia across models with O*NET job attributes are larger than the ones omitting O*NET variables. The point VSL estimates from the models including the O*NET job attributes are between $0.54 and $1.83 million (in 2008 $), and the point VSL estimates from the model excluding the O*NET job attributes are between $0.24 and $0.99 million (in 2008 $). Comparing these estimates to those of Kochi (2011) which are between $1.9 and $2.8 million (in 2008 $), my estimates are closer to the lower-end value of hers.

In addition to the comparison of the point estimates, table 6.2 also reflects the impact of occupation and industry fixed effects. The occupation fixed effects appears to increase the estimate. The industry fixed effects seem to have strong impact on the estimates of the VSL. The models with industry dummy variables, however, produce smaller estimates than the ones without industry dummies, regardless of the inclusion of the O*NET variables. In particular, the coefficient on risk in model 3 and 7 becomes insignificant when the model

only controls for industry dummies. This finding implies that inter-industrial wage differentials maybe an important influence in the 2004 SIPP panel dataset, which could lead to high sensitivity of the VSL estimates to the industry fixed effects.

6.3.2 VSL estimate with the PSID data

The Panel Study of Income Dynamic (PSID) is a longitudinal study of a representative sample of U.S. individuals (men, women, and children) and the family units in which they reside. The PSID has followed the same families and their descendants since 1968, constructing a panel of 36 interviews as of 2009. Although the PSID began with just under 5,000 households, over time attrition resulted in only 3,000 households. The PSID contains individual-level data on wages, industry and occupation of employment, and socio-economic characteristics. Beginning in 1997, people are interviewed every other year. Each two-year reference period is a "wave".

The data from the 2003, 2005, and 2007 waves will be used in this analysis. There are a total of 24,113 observations in these three waves. In order to be comparable to the other recent studies employing this data, I will follow the selection criteria in Kniesner et al. (2006) to choose the sample for the analysis. The sample will consist of heads of household ages 18–65 that are in the random Survey Research Center (SRC) portion of the PSID. This sample excludes the oversampling of the poor and Latino sub-samples that are present in other PSID data products. The heads of the households in the sample are those who worked for hourly or salary pay at some point in the previous calendar year, and who are not permanently disabled or institutionalized, and who are not in agriculture or the armed forces.

There are a total of 6,021 observations for 2,460 workers in the selected sample. The detailed description of PSID sample is presented in Appendix 6.B.

6.3.2.1 Estimation results

In the estimation models, the dependent variable is gross hourly wages, and the independent variables include the ones commonly used in the hedonic wage literature such as age, educational attainment, gender, race, marital status, number of children under 18 in the household, union status, the worker's occupation, and the industry category of the firms for which the household head works. In addition, the independent variables also include the O*NET job attributes which are estimated to improve the model's performances. The full set of variables is reported in Appendix 6.C column (1).

Table 6.3 reports the estimation results from preferred model specification: a linear FD (Δ=2) model with both occupation and industry fixed effect included. For succinctness, only risk coefficients and the estimates of VSL are reported in table 6.3. The full set of estimation results is reported in appendix 6.C column (1). In table 6.3, the estimation results indicate that the coefficient on fatal risk is only significant at 10% level and implies a VSL of $15.4 million with a 95% confident interval of $0.3 million to $31.4 million. The 95% confident is quiet wide and thus include the typical point estimates from the literature which have ranged between $4 million to $11 million (Viscusi, 2004; Viscusi et el., 2007; Kniesner et al., 2012).

6.3.2.2 Sensitivity analysis

Table 6.4 replicates table 6.2 but with PSID data. Model 1 through 4 in table 6.4 includes the O*NET job attributes, while they are excluded in model 5 through 8. Model 1 through 4 varies by the presence of occupation/ industry fixed effects, and the same variation applies to model 5 through model 8. All the estimation models still use linear FD (Δ=2) estimator. The full set of the estimation results are reported in Appendix 6.C, column (2) to (4).

It appears that the estimated wage-risk premia across models with O*NET job attributes are larger than the ones omitting O*NET variables. The point estimates of the VSL from the models including the O*NET job attributes are between $14.8 and $16.3 million, while those excluding the O*NET job attributes are between $12.2 and $13.8 million. These estimates are closer to the upper end value of typical estimates from the PSID panel models ($4 million - $11 million), but are quite a bit larger than SIPP data estimates.

Contrary to the estimates with the SIPP data, the inclusion of occupation and/or industry fixed effects does not appear to have much impact on the coefficient estimates for fatal risk. Looking across columns, there is not consistent pattern or particularly large impact of occupation and/ or industry fixed effects on the coefficient estimates. The estimates with PSID data seems to be rather robust in these regards.

In sum, when applying the modeling specification suggested by the simulation results, the point estimates of the VSL are between $0.1 and $1.83 million with the SIPP data and the estimates are between $12.1 to $16.9 million with the PSID data. Comparing to the

typical estimates in the literature, the estimates with the SIPP data is similar to those of Kochi (2010). The estimates with the PSID data are significantly larger than those with SIPP data and are closer to the upper range of estiamtes in Viscusi et el. (2007) and Kniesner et al., (2012). The sensitivity analysis of the VSL estimates with data sets indicates that including O*NET job attributes tends to increase the estimated work-risk premia. However, the impact of occupation and/or industry fixed effects on the estimates depends on which data are used in estimation.

6.4 Future work

The idea of constructing a simulated hedonic market to evaluate the reduced form hedonic functions dates back CDM's (1988) study and has been applied to examine the hedonic price models in the housing market setting. However, the proposed simulation model in the labor market setting is still at a very early stage of development. This section considers ways to improve future application to the labor market by further exploiting the job-worker sorting algorithm and expanding the evaluation analysis.

6.4.1 Different utility function

Recall that the hedonic wage equilibrium is computed for Cobb-Douglas utility function, given that it is general enough to allow for the heterogeneous valuation of job characteristics. Cobb-Douglas utility function is also consistent with the intuition of generalization of Roy model which incorporates non-pecuniary job returns as utility component (DeLeire and Timmins, 2008). A natural concern, however, is that the pattern of evaluation results may simply reflect the shape of the utility function. Future work could

consider using different parametric forms of utility function in addition to Cobb-Douglas in the sorting algorithm. The simulation model could be repeated for different utility functions to provide robustness check of the evaluation results.

Furthermore, the fatal risk level is introduced as the probability in constructing the expected utility. An extension of the model can be adding subjective risk aversion as a determinant factor in the job-worker sorting process. In this way, the simulation model will allow not only worker's heterogeneous preference toward non-fatal injury job attributes, but also heterogeneous fatal risk aversion level.

6.4.2 Incorporating labor force data

The estimation results from section 6.3 indicate that even with the most robust model specification and econometric strategies, there is large difference between the VSL estimates from SIPP and PSID data. And it appears that the SIPP estimates are far more sensitive to the occupation and industry fixed effect then the PSID estimates. These findings suggest that VSL estimates may largely depend on the source of the labor force data available to researchers. The difference between SIPP estimates and the PSID estimates could be caused by the dimensions of wage differentials presented in the data set. For instance, the high sensitivity of the SIPP estimate to the inclusion of industry fixed effects implies a strong existence of inter-industry wage differentials, while PSID data may not present this dimension of wage differentials as strong.

In the future work, the simulation model could be applied to the alternative sample of workers in addition to the SIPP sample, so that the evaluation analysis can be expanded to

incorporate the choice of worker sample. For instance, the proposed simulation framework could be applied to different labor force data such as CPS or PSID, or to different groups of workers segmented by gender, or age or working classes. Expanding the simulation to account for different labor force data or sample workers not only provides insights on the accuracy of empirical model specification within specific data context, but also helps to identify the set of model specifications that are robust to the data choices.

Table 6.1

Estimates of wage-fatal risk tradeoff from SIPP

Estimator	First difference (Δ=2)
Functional form (fixed effect)	Linear (Occupation & industry fixed effect)
coefficient of annual fatality risk (std)	0.027** (0.009)
implied VSL ($ Millions) [95% confident Interval]	0.54 [0.18 0.88]
R2 (overall)	82%
Number of Observation	1,880

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

Table 6.2

Additional estimates of wage-fatal risk tradeoff from SIPP for sensitivity analysis

Linear, First-difference (Δ=2), with O*NET job attributes				
	occ. and ind. dummy variables	occ. dummy variables	ind. dummy variables	no occ. & ind. dummy variables
	Model 1	Model 2	Model 3	Model 4
Risk coefficient (std)	0.027** (0.009)	0.091** (0.035)	-0.032 (0.22)	0.043** (0.019)
VSL ($ Millions)	0.54	1.83		0.84
R2 (overall)	82%	78%	80%	75%
N	1,880	1,880	1,880	1,880
Linear, First-difference (Δ=2), without O*NET job attributes				
	Model 5	Model 6	Model 7	Model 8
Risk coefficient (std)	0.005 (0.0021)	0.049** (0.020)	-0.06 (0.31)	0.012** (0.0047)
VSL ($ Millions)		0.99		0.24
R2 (overall)	72%	63%	71%	58%
N	1,880	1,880	1,880	1,880

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

Table 6.3

Estimates of wage-fatal risk tradeoff from PSID

Estimator	First difference (Δ=2)
Functional form (fixed effect)	Linear (Occupation & industry fixed effect)
coefficient of annual fatality risk (std)	0.78* (0.39)
implied VSL ($ Millions) [95% Confident Interval]	15.4 [0.3 31.4]
Number of Observation	2,356

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

Table 6.4

Additional estimates of wage-fatal risk tradeoff from PSID for sensitivity analysis

Linear, First-difference (Δ=2), with O*NET job attributes				
	occ. and ind. dummy variables	occ. dummy variables	ind. dummy variables	no occ. & ind. dummy variables
	Model 1	Model 2	Model 3	Model 4
Risk coefficient (std)	0.78* (0.39)	0.74** (0.38)	0.77* (0.44)	0.81* (0.46)
VSL ($ Millions)	15.4	14.8	15.4	16.3
R2 (overall)	25%	24%	25%	20%
N	2,356	2,356	2,356	2,356
Linear, First-difference (Δ=2), without O*NET job attributes				
	Model 5	Model 6	Model 7	Model 8
Risk coefficient (std)	0.69* (0.30)	0.62* (0.30)	0.61* (0.28)	0.66** (0.29)
VSL ($ Millions)	13.8	12.5	12.2	13.2
R2 (overall)	18%	16%	18%	14%
N	2,356	2,356	2,356	2,356

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

REFERENCES

Abraham, K. G., and J. R. Spletzer. (2009). "New Evidence on the Returns to Job Skills." *American Economic Review: Papers and Proceedings* 99(2) p. 52-57.

Arabsheibani, G. R., and A. Martin. (2001). "Self-Selection Bias with a Continuous Variable: Potential Pitfall in a Common Procedure." *Applied Economics,* Vol. 33, No. 15 (December), p. 1903-1910.

Bayer, P., and C. Timmins. (2007). "Estimating Equilibrium Models Of Sorting Across Locations." *The Economic Journal* 117 (518) p. 353-374

Biddle, J.E., and G.A. Zarkin. (1988). "WorkerPreference and Market Compensation for Job Risk." *The Review of Economics and Statistics* 70 (4) p. 660–667

Black, D.A., J. Galdo, and L. Liqun. (2003). "How Robust are Hedonic Wage Estimates of the Price of Risk?" *Final Report to the USEPA, June 2003,* R 82943001

Black, D.A., and T.J. Kniesner. (2003)."On the Measurement of Job Risk in Hedonic Wage Models." *Journal of Risk and Uncertainty 27(3*) p. 205-220

Blomquist, G. C., and C. Glenn. (2004). "Self-Protection and Averting Behavior, Values of Statistical Lifes, and Benefit Cost Analysis of Environmental Policy." *Review of Economics of the Household 2* p. 89-110.

Bowland, B. J., and J.C. Beghin. (2001). "Robust Estimates of Values of a Statistical Life for Developing Economies." *Journal of Policy Modeling 23* p. 385-396.

Brown, J. E., and D. E. Ethridge (1995). "Functional Form Model Specification: An Application to Hedonic Pricing."*Agricultural and Resource Economics Review* 24 (2) p. 166-173

Costrell, R. M., and G. C. Loury. (2004). "Distribution of Ability and Earnings in a Hierarchical Job Assignment Model." *Journal of Political Economy* 112 (6) p.1322-1362

Crawford, V. P., and E.M. Knoer (1988). "Job matching with heterogeneous firms and workers". *Econometrica* 49 (2) p. 437–450

Cropper, M. L., L. B. Deck and K. E. McConnell (1988). "On the Choice of Functional Form for Hedonic Price Functions." *The Review of Economics and Statistics* 70 (4) p. 668-675

DeLeire T,, and C. Timmins (2008). "Roy model sorting and nonrandom selection in the valuation of a statistical life." Natl. Bur. Econ. Res.Work. Pap. 14364, Cambridge, M

Ekeland,I., J. J. Heckman., and L. Nesheim. (2004). "Identification and Estimation of Hedonic Models." *Journal of Political Economy* 112(S1), S60-S109

Garen, J. (1988). "Compensating Wage Differentials and the Endogeneity of job Riskness." *The Review of Economics and Statistics* 70(1) p. 9-16.

Glomb, T. M., J. D. Kammeyer-Mueller, and M. Rotundo (2004). "Emotional labor demands and compensating wage differentials." *Journal of Applied Psychology* 89 (4) p. 700-714

Griliches, Z., and J. A. Hausman. (1986). "Errors in Variables in Panel Data." *Journal of Econometrics 34(8) p. 1396-1400.*

Gunderson, M., and D. Hyatt. (2001). "Workplace Risks and Wages: Canadian Evidence from Alternative models." *Canadian Journal of Economics,* 34(2) p. 377-395.

Hadden, W. C., N. Kravets, and C. Muntaner, (2004). "Descriptive Dimensions of US Occupations with Data from the O*NET." *Social Science Research* 33 p. 64–78

Hausman J.A. and W.E. Taylor (1981), "Panel Data and Unobservable Individual Effects," *Econometrica* 49(6) p. 1377-1398.

Heckman, J. J. and B. E. Honore (1990). "The Empirical Content of the Roy Model". *Econometrica* 58 (5) p. 1121-1149

Heckman, J. J., and G. Sedlacek (1985). "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market." *The Journal of Political Economy* 93 p. 1077-1125

Heckman, J. J., R Matzkin and L. P. Nesheim (2004). "Simulation and Estimation of Hedonic Models." *In Frontiers of Applied General Equilibrium Modeling: Essays in Honor of Herbert Scarf*

Hirsch, B.T., (2005). "Why do Part-time Workers Earn Less? The Role of Worker and Job Skills." *Industrial and Labor Relations Review* 22(3) p. 689-722.

Hirsch, B.T., and E. Schumacher, (2010). "Underpaid or Overpaid? Wage Analysis for Nurses Using Job and Worker Attributes." *Southern Economic Journal,* revise and resubmit.

Hinntermann, B., A. Alberini, and A. Markandya, (2010). "Estimating the Value of Safety with Labour Market Data: Are the Results Trustworthy?" *Applied Economics* 42(7-9): 1085-100.

Hwang, H., D.T. Mortensen, and W.R. Reed (1998). "Hedonic wages and labor market search." *Journal of Labor Economics* 16 (4) p. 815–847

Ingram, B. B. and G. R. Neumann (2006). "The Returns to Skill". *Education* 13 (1) p. 35-59

Jeanneret, P.R., W.C. Borman., U.C. Kubisiak., and M.A. Hanson (1999). "An Occupational Information System for the 21st Century: The Developemnt of the O*NET." *Washington. DC: APA Books*

Jones-Lee, M. W. (1974). "The Value of Changes in the Probability of Death or Injury." *Journal of Political Economy, 82*(4), 835-849.

Kaplow, L. (2003). "The Value of a Statistical Life and the Coefficient of Relative Risk Aversion." *Journal of Risk and Uncertainty* 31(1) p. 23-34

Kelso A. S. JR. and V P Crawford. (1982). "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica* 50 (6) p. 1483-1504

Klaiber, H. Allen and V. Kerry Smith. 2013. "Quasi Experiments, Hedonic Models, and Estimating Tradeoffs for Local Amenities," *Land Economics forthcoming* (also NBER working paper #4957)

Kniesner, T. J., W. K. Viscusi, C. Woock, and J. P. Ziliak, (2006). "Pinning Down the Value of Statistical Life." Center for Policy Research Working Papers No. 85.

Kniesner, T. J., W. K. Viscusi, and J. P. Ziliak (2010). "Policy Relevant Heterogeneity in the Value of Statistical Life: New evidence From Panel Data Quantile Regressions." *Journal of Risk and Uncertainty* 40 p. 15-31

Kniesner, T. J. and W K. Viscusi, C. Woock., and J. P. Ziliak (2010). "The Value of a Statistical Life : Evidence from Panel Data." Working Paper, Syracus University, Center for Policy Research, No. 122

Kuminoff, N. V., C. Parmete, and J. Pope. (2010). "Which Hedonic Models Can We Trust to Recover the Marginal Willingness to Pay for Environmental Amenities?" *Journal of Environmental Economic and Management* (forthcoming)

Kuminoff, N. V. and A. S. Jarrah. (2010). "A New Approach to Computing Hedonic Equilibria and Investigating the Properties of Locational Sorting Models." *Journal of Urban Economics, 2010,* vol. 67, issue 3, pages 322-335

Kochi, I., B Hubbell and R. Kramer. (2006). "An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis". *Environmental Resource Economics* 34 (3) p. 385-406

Lee, F. (2002). "Statistics and Data Analysis Principal Component Analysis vs Exploratory Factor Analysis." *Analysis* 2 p. 1-11

McConnell, K . E. (2006). Chapter 7: Hedonic wage analysis. In N. E. Bockstael & K. E. McConnell (Eds.), *Environmental and Resource Valuation with Revealed Preferences; A Theoretical Guide to Empirical Models (The Economics of Non-Market Goods and Resources)* Springer.

Mrozek, J. R., and L. O. Taylor, (2002). "What Determines the Value of Life? A Meta-analysis."Journal of Policy Analysis and Management 21(2) p. 253-270.

Miguel Portela, (2000). "Measuring skill: a multi-dimensional index" *Economics letters* 72, 27-32

Perri, G., and C. Sparber (2009). "Task Specialization, Immigration and Wages*." American Economic Journal: Applied Economics 1(3) p. 135-169.*

Scotton, C. R. (2000). "The Impact of Workplace-Risk Measurement on Estimates of the Value of a Statistical Life*."* Georgia State University, Atlanta.

Scotton, C. R. (2013). "New Risk rates, Inter-Industry Differentials and the Magnitude of VSL estimates*." Journal of Benefit-Cos Analysis 4(1): 39-80.*

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) p. 89-110.

Smith, V. K., M. F. Evans, H. Kim, and D.H. Jr. Taylor. (2004). "Do the 'Near' Elderly Value Mortality Risks Differently?" *Review of Economics and Statistics 86(1) p.* 423-429

Smith, V. K., S. K. Pattanayak and G. L. Van Houtven (2003). "VSL reconsidered : what do labor supply estimates reveal about risk preferences ?" *Economic Letters* 80(2) p. 147-153

Spengler, H. and S. Schaffner. (2006). "Using Job Changes to Evaluate the Bias of theValue of a Statistical Life." Unpublished paper

Thaler, R. and S. Rosen. (1975). "The Value of Saving a Life: Evidence from the Labor Market." In N.E. Terleckyj (ed.), *Household Production and Consumption.* NewYork: Columbia University Press, pp. 265–300.

U.S. Environmental Protection Agency. (1999). “The Benefits and Costs of the Clean Air Act, 1970 to1990”. USEPA.

U.S. Environmental Protection Agency. (1999). “The Benefits and Costs of the Clean Air Act 1990 to 2010: EPA Report to Congress.” Washington D.C.: US EPA

U.S. Environmental Protection Agency. (2004). “Value of Statistical Life Analysis and Environmental Policy: A White Paper.” Washington D.C.: USEPA.

U.S. Environmental Protection Agency. (2010). “Value of Statistical Life Analysis and Environmental Policy: A White Paper.” Washington D.C.: USEPA.

U.S. Environmental Protection Agency. (2005).“Regulatory Impact Analysis for the Final Clean Air Interstate Rule.” Washington D.C.: US EPA

Viscusi, W. K., & Aldy, J. E. (2003). “The value of a statistical life: A critical review of market estimates throughout the world.” *Journal of Risk and Uncertainty* 27(1) p. 5-76

APPENDICES

Appendix 3.A

CFOI research file element

Table 3.A.1

CFOI research file element

Element	Description
REC	Record ID
REF	Reference year
YEA	Year of injury
MON	Month of injury
DAY	Day of week
TII	Time of incident
NAT	Nature of injury
PAR	Part of body
EVE	Event or exposure
SOU	Source of injury
SEC	Secondary source of injury
ACT	Worker activity
LOC	Location
REG	Region
OCC	Occupation
USO	Usual lifetime occupation
IND	Industry
USI	Usual lifetime occupation
OWN	Ownership
EST	Establishment size class
EMP	Employee status
TIE	Time with employer
GEN	Gender
AGE	Age group
RAC	Race
HIS	Hispanic origin
FOR	Foreign born
SUR	Days survived
NARR	Narrative

Appendix 3.B

Occupation group mapping

Table 3.B.1

Occupation group mapping

occ code	22 occupation group	6-digit SOC code	Census Occupation Classification
11	Management Occupations	11-1011—11-9199	001-043
13	Business and Financial Operations Occupations	13-1011—13-2099	050-095
15	Computer and Mathematical Occupations	15-1011—15-2099	100-124
17	Architecture and Engineering Occupations	17-1011—17-3031	130-156
19	Life, Physical, and Social Science Occupations	19-1011—19-4099	160-196
21	Community and Social Services Occupations	21-1011—21-2099	200-206
23	Legal Occupations	23-1011—23-2099	210-215
25	Education, Training, and Library Occupations	25-1011—25-9031	220-255
27	Arts, Design, Entertainment, Sports, and Media Occupations	27-1011—27-4099	260-296
29	Healthcare Practitioners and Technical Occupations	29-1011—29-9099	300-354
31	Healthcare Support Occupations	31-1011—31-9099	360-365
33	Protective Service Occupations	33-1011—33-9099	370-395
35	Food Preparation and Serving Related Occupations	35-1011—35-9099	400-416
37	Building and Grounds Cleaning and Maintenance Occupations	37-1011—37-3019	420-425
39	Personal Care and Service Occupations	39-1011—39-9099	430-465
41	Sales and Related Occupations	41-1011—41-9099	470-496
43	Office and Administrative Support Occupations	43-1011—43-9199	501-593
45	Farming, Fishing, and Forestry Occupations	45-1011—45-4029	601-613

Table 3.B.1 Continued

47	Construction and Extraction Occupations	47-1011—47-5099	620-694
49	Installation, Maintenance, and Repair Occupations	49-1011—49-9099	700-762
51	Production Occupations	51-1011—51-9199	770-896
53	Transportation and Material Moving Occupations	53-1011—53-7199	900-975

Appendix 3.C

Industry group mapping

Table 3.C.1

Industry group mapping

ind code	23 industry group	6-digit NAICS code	Census Industry Classification
1100	Agriculture, Forestry, Fishing and Hunting	11-1110—11-5310	0-300
2100	Mining	21-1111—21-3115	370-500
2200	Utilities	22-1111—22-1330	570-700
2300	Construction	23-1110—23-5990	770
3123	Manufacturing	31-1111—31-9999	1070-3990
4200	Wholesale Trade	42-1110—42-2990	4070-4599
4445	Retail Trade	44-1110—45-4390	4670-5790
4800	Transportation	48-1111—48-8999	6070-6290
4910	Postal	49-1110	6370
4923	Couriers and Warehousing, except postal	49-2110—49-3190	6380-6390
5100	Information	51-1110—51-4210	6470-6780
5200	Finance and Insurance	52-1110—52-5990	6870-6990
5300	Real Estate and Rental and Leasing	53-1110—53-3110	7070-7190
5400	Professional, Scientific, and Technical Services	54-1110—54-1990	7270-7490
5500	Management of Companies and Enterprises	55-1111—55-1114	7570

Table 3. C.1 Continued

5600	Administrative and Support and Waste Management and Remediation Services	56-111---56-2998	7580-7790
6100	Educational Services	61-1110—61-1710	7860-7890
6200	Health Care and Social Assistance	62-1111—62-4410	7970-8470
7100	Arts, Entertainment, and Recreation	71-1110—71-3990	8560-8590
7200	Accommodation and Food Services	72-1110—72-3410	8660-8690
8100	Other Services (except Public Administration and private households)	81-1111—81-3990	8770-9190
8140	Private households	81-4110	9290
9200	Public Administration except nat defense	92-1110—92-8120	9300-9590

Appendix 3.D

Table 3.D.1

3-digit industry group mapping

3-digit NAICS code	Title	Subsets of 2000 Census Industry group
111000	Crop Production	170
112000	Animal Production	180
113000	Forestry and Logging	190, 270
114000	Fishing, Hunting and Trapping	280
115000	Support Activities for Agriculture and Forestry	290
211000	Oil and Gas Extraction	370
212000	Mining (except Oil and Gas)	380, 390, 470
213000	Support Activities for Mining	490
221000	Utilities	570, 580, 590, 670, 680
230000	Construction	770
311000	Food Manufacturing	1070, 1080
312000	Beverage and Tobacco Manufacturing	1370, 13390
313000		1670
313315	Textile Mills and Apparel Manufacturing	1470, 1480, 1490, 1680, 1690
314000	Textile Product Mills	1570, 1590
316000	Leather and Allied Product Manufacturing	1770, 1790
321000	Wood Product Manufacturing	3770, 3780, 3790, 3870
322000	Paper Manufacturing	1870, 1880, 1890
323000	Printing and Related Support Activities	1990

Table 3.D.1 Continued

324000	Petroleum and Coal Products Manufacturing	2070, 2090 ,
325000	Chemical Manufacturing	2170, 2180, 2190, 2270, 2280, 2290
326000	Plastics and Rubber Products Manufacturing	2370, 2380, 2390
327000	Nonmetallic Mineral Product Manufacturing	2470, 2480, 2490, 2570, 2590
331000	Primary Metal Manufacturing	2670, 2680, 2690, 2770
332000	Fabricated Metal Product Manufacturing	2780, 2790, 2870, 2880, 2890, 2970, 2980
333000	Machinery Manufacturing	3070, 3080, 3090, 3170, 3180, 3190, 3290
334000	Computer and Electronic Product Manufacturing	3360, 3370, 3380, 3390
335000	Electrical Equipment, Appliance, and Component Manufacturing	3470
336000	Transportation Equipment Manufacturing	3470
337000	Furniture Related Product Manufacturing	3890
339000	Miscellaneous Manufacturing	3890
423000	Durable Goods Merchant Wholesalers	3890
424000	Nondurable Goods Merchant Wholesalers	3890
425000	Wholesale Electronic Markets and Agents and Brokers	4585
441000	Motor Vehicle and Parts Dealers	4670, 4680, 4690
442000	Furniture and Home Furnishing Stores	4770
443451	Electronics and Appliance Stores, Sporting Goods, Hobby, Book, and Music Stores	5270-5370
444000	Building Material and Garden Equipment and Supplies Dealers	4870, 4880, 4890

Table 3.D.1 Continued

445000	Food and Beverage Stores	4970, 4980, 4990
446000	Health and Personal Care Stores	5070, 5080
447000	Gasoline Stations	5090,
448000	Clothing and Clothing Accessories Stores	5170, 5180, 5190
452000	General Merchandise Stores	5380, 5390
453000	Miscellaneous Store Retailers	5470, 5480, 5490, 5570, 5580
454000	Nonstore Retailers	5590, 5591, 5592, 5670, 5680, 5690
481000	Air Transportation	6070
482000	Rail Transportation	6080
483000	Water Transportation	6090
484000	Truck Transportation	6170
485000	Transit and Ground Passenger Transportation	6180, 6190
486000	Pipeline Transportation	6270
487000	Scenic and Sightseeing Transportation	6280
488000	Support Activities for Transportation	6290
492000	Couriers and Messengers	6380
493000	Warehousing and Storage	6390
511000	Publishing Industries (except Internet)	6470, 6480, 6490
512000	Motion Picture and Sound Recording Industries	6570, 6590
515517	Broadcasting and Telecommunications	6670, 6680, 6690
518000	Internet Service Providers, Web Search Portals, and Data Processing Services	6692, 6695
519000	Other Information Services	6770, 6780

Table 3.D.1 Continued

520000	Finance and Insurance	6870, 6880, 6890, 6970, 6990, 7460, 7470
530000	Real Estate and Rental and Leasing	7070-7190,7270, 7280, 7290, 7370, 7380, 7390, 7460,
541000	Professional, Scientific, and Technical Services	7470, 7480, 7490,7570,7580, 7590, 7670, 7680, 7690, 7770,
551000	Management of Companies and Enterprises	7780,
561000	Administrative and Support Services	7790,,
562000	Waste Management and Remediation Services	7790
611000	Educational Services	7860-7890,7970, 7980, 7990, 8070, 8080, 8090, 8170,
621000	Ambulatory Health Care Services	8180
622000	Hospitals	8190
623000	Nursing and Residential Care Facilities	8270, 8290
624000	Social Assistance	8370, 8380, 8390, 8470
711000	Performing Arts, Spectator Sports, and Related Industries	8560
712000	Museums, Historical Sites, and Similar Institutions	8570
713000	Amusement, Gambling, and Recreation Industries	8580, 8590
721000	Accommodation	8660, 8670
722000	Food Services and Drinking Places	8680
811000	Repair and Maintenance	8780, 8790, 8870, 8880, 8890
812000	Personal and Laundry Services	8970, 8980, 8990, 9070, 9080, 9090
813000	Religious, Grantmaking, Civic, Professional, and Similar Organizations	9160, 9170, 9180, 9190

Appendix 3.E

Summary of annual non-fatal injury risk across 66 industry groups in the SIPP sample

Table 3.E.1

Summary of annual non-fatal injury risk across 66 industry groups in the SIPP sample

3-digit NAICS code	Title	Mean [a]	Std. Dev.	% of workers
211000	Oil and Gas Extraction	110	17.32	0.05
212000	Mining (except Oil and Gas)	271.54	17.25	0.23
213000	Support Activities for Mining	220	0	0.11
221000	Utilities	236.47	12.55	1.2
230000	Construction	333.42	9.41	7.12
311000	Food Manufacturing	499.58	21.61	1.25
312000	Beverage and Tobacco Manufacturing	557.50	19.09	0.14
313315	Textile Mills and Apparel Manufacturing	190.41	9.28	0.49
314000	Textile Product Mills	276.67	13.032	0.21
321000	Wood Product Manufacturing	510	29.77	0.79
322000	Paper Manufacturing	264	19.32	0.7
323000	Printing and Related Support Activities	236.67	4.77	0.74
324000	Petroleum and Coal Products Manufacturing	146.67	26.05	0.21
325000	Chemical Manufacturing	183.24	12.39	1.2
326000	Plastics and Rubber Products Manufacturing	424.03	34.47	1.18
327000	Nonmetallic Mineral Product Manufacturing	460	35.99	0.79
331000	Primary Metal Manufacturing	486.67	23.77	1.06
332000	Fabricated Metal Product Manufacturing	383.31	4.72	2.13
333000	Machinery Manufacturing	303.43	4.77	1.8

Table 3.E.1 Continued

334000	Computer and Electronic Product Manufacturing	106.55	9.43	1.99
335000	Electrical Equipment, Appliance, and Component Manufacturing	280	8.24	0.95
336000	Transportation Equipment Manufacturing	450	14.17	4.02
337000	Furniture Related Product Manufacturing	430	36.06	0.05
339000	Miscellaneous Manufacturing	233.66	17.01	1.25
423000	Durable Goods Merchant Wholesalers	216.76	4.69	3.65
424000	Nondurable Goods Merchant Wholesalers	380	8.27	2.7
441000	Motor Vehicle and Parts Dealers	206.51	12.58	1.87
442000	Furniture and Home Furnishing Stores	293.33	19.52	0.26
444000	Building Material and Garden Equipment and Supplies Dealers	416.67	18.95	1.8
445000	Food and Beverage Stores	323.33	4.74	1.69
446000	Health and Personal Care Stores	110	8.32	0.48
447000	Gasoline Stations	156.67	9.66	0.37
448000	Clothing and Clothing Accessories Stores	100	8.45	0.26
452000	General Merchandise Stores	403.81	19.10	1.99
453000	Miscellaneous Store Retailers	187.20	19.04	0.44
454000	Nonstore Retailers	270	8.94	0.11
481000	Air Transportation	763.33	9.61	0.48
482000	Rail Transportation	187.20	12.75	0.44
484000	Truck Transportation	383.33	9.51	1.06
485000	Transit and Ground Passenger Transportation	336.67	26.65	0.58
488000	Support Activities for Transportation	316.67	41.14	0.42
492000	Couriers and Messengers	820	66.56	0.48
493000	Warehousing and Storage	558.09	16.62	0.37

Table 3.E.1 Continued

511000	Publishing Industries (except Internet)	110	0	0.79
515517	Broadcasting and Telecommunications	134.71	3.28	1.55
518000	Internet Service Providers, Web Search Portals, and Data Processing Services	40	10	0.05
519000	Other Information Services	100	0	0.02
520000	Finance and Insurance	33.34	4.72	6.43
530000	Real Estate and Rental and Leasing	193.41	12.67	1.6
541000	Professional, Scientific, and Technical Services	53.34	4.73	5.16
551000	Management of Companies and Enterprises	130	20	0.05
561000	Administrative and Support Services	182.94	12.55	1.8
562000	Waste Management and Remediation Services	463.08	59.58	0.46
611000	Educational Services	96.67	4.72	11.79
621000	Ambulatory Health Care Services	110.08	8.20	4.47
622000	Hospitals	329.97	8.17	6.17
623000	Nursing and Residential Care Facilities	563.27	17.15	1.78
624000	Social Assistance	206.85	12.569	1.29
711000	Performing Arts, Spectator Sports, and Related Industries	250	26.83	0.11
712000	Museums, Historical Sites, and Similar Institutions	230	22.36	0.26
713000	Amusement, Gambling, and Recreation Industries	296.67	28.96	0.9
721000	Accommodation	316.67	9.51	1
722000	Food Services and Drinking Places	136.75	4.70	2.06
811000	Repair and Maintenance	176.73	12.48	0.97
812000	Personal and Laundry Services	156.67	4.78	0.63

Table 3.E.1 Continued

813000	Religious, Grantmaking, Civic, Professional, and Similar Organizations	109.89	8.23	1.55

a. Rates are reported as deaths per 10,000 workers.

Appendix 3.F

*Example of O*NET descriptors*

Table 3.F.1

*Example of O*NET descriptors*

The tables below list an example for each major categories of O*NET descriptors. To illustrate the data structure and provide the general idea of data structure and contents, each table lists out only few descriptors instead of full content of the data.

Element ID	Element name	Description
1	Worker Characteristics	Worker characteristics
1.A	Abilities	Enduring attributes of the individual that influence performance
1.A1.a	Verbal ability	Abilities that influence the acquisition and applica of verbal information in problem solving
1.A.1.a1	Oral comprehension	The ability to listen to and understand information ideas presented through spoken words and sentences.
1.A.1.a.2	Written comprehension	The ability to read and understand information and ideas presented in writing.

Element ID	Element name	Description
2	Worker Requirements	Worker Requirements
2.A	Basic skills	Developed capacities that facilitate learning or the more rapid acquisition of knowledge
2.A.1.a	Reading Comprehension	Understanding written sentences and paragraphs work related documents.
2.A.1.c	Mathematics	Communicating effectively in writing as appropr for the needs of the audience.

Table 3. F.1 Continued

Element ID	Element name	Description
3	Experience Requirements	Experience Requirements
3.A	Experience and Training	If someone were being hired to perform this j how much of the following would be required
3.A.1	Related Work Experience	Amount of related work experience required t hired for the job?
3.A.2	On-Site or In-Plant Training	Amount of on-site or in-plant training (e.g., organized class room instruction) required to perform the job?
3.A.3	On-the-Job Training	Amount of on the job training required to per the job?
Element ID	Element name	Description
4	Occupational Requirements	Occupational Requirements
4.A	Generalized Work Activities	General types of job behaviors occurring on multiple jobs
4.A.1.a.1	Getting Information	Observing, receiving, and otherwise obtainir information from all relevant sources.
4.A.1.a.2	Monitor Processes, Materials, or Surroundings	Monitoring and reviewing information from materials, events, or the environment, to dete assess problems.
4.A.1.b.1	Identifying Objects, Actions, and Events	Identifying information by categorizing, estimating, recognizing differences or simila and detecting changes in circumstances or ev

Appendix 3.G

Major groups of occupational requirements

Following Deleire et.al (2009), Ingram and Neumann (2006) and Abraham and Spletzer (2009), I construct five informative categories within occupational requirements. These five informative categories seek to extract 1) cognitive demands, 2) motor skill demand, 3) physical demands, 4) working condition, and 5) the requirement of interaction with people both in public or with coworkers from the job. There is big amount of descriptors under each of these five categories, usually between 50 and 80. Below I provide broad descriptions for each category. The informative categories are in bold and O*NET descriptors are listed below them.

Cognitive Skills
Reading comprehension
Writing letters and memos
Mathematics skills and knowledge
Complex problem solving skills
Critical thinking skills
Judgment and decision making skills
Thinking creatively
Documenting/Recording information skills
Processing information
Importance of continuous, repetitious physical activities or mental activities (such as checking entries in a ledger, key entry)
Clerical knowledge

Motor skills
Visual requirement
Finger dexterity
Manual dexterity
Motor coordination
Machine handling skills

Physical strength demand

Handling and moving objects
Performing general physical activities
Time spent bending or twisting body
Time spent climbing ladders, scaffolds, poles,etc.
Time spent keeping or regaining balance
Time spent kneeling crouching, stooping, or crawling
Time spent standing
Time spent using hands to handle, control, or feel objects, tools, or controls
Time spend walking or running
Keeping a pace set by machinery or equipment
Time spend making repetitive motions

Working conditions
Indoor environment control
Outdoor, exposure to extreme weather, high places
Exposure to vehicle and instrument
Exposure to sound, noise level is distracting or uncomfortable
Exposure radiation
Exposure to hazardous conditions and equipments
Exposure to disease and infection
Requiring wearing protective or safety equipment

Interaction with people
Assisting and caring for others
Communicating with people outside the organization
Performing for or working directly with the public
Interactions that require you to deal with external customers or the public in general
Customer and personal service
Management of material resources (such supplier)
Communicating with people outside the organization
Instructing
Coaching and developing others
Training and teaching others
Education and Training

Appendix 3.H

Example of factor analysis

In order to provide a straight forward explaination of the mechanism of the factor analysis that can be easily understood, I post an STATA example with the desciption of mechanism of the factor analysis. This example is taken directly from a STATA suppllementary document which can be accessed from:
http://dss.princeton.edu/training/Factor.pdf
The copied STATA screens present the steps to carry out the factor analysis and they are followed by the lists of explainations of each output from the STATA commands.

```
. factor  ideol equality owner respon competition, pcf
(obs=1125)

Factor analysis/correlation                       Number of obs    =      1125
    Method: principal-component factors           Retained factors =         2
    Rotation: (unrotated)                         Number of params =         9

    Factor  |  Eigenvalue   Difference    Proportion   Cumulative
------------+------------------------------------------------------------
    Factor1 |      1.54524      0.21290      0.3090       0.3090
    Factor2 |      1.33235      0.49085      0.2665       0.5755
    Factor3 |      0.84149      0.12808      0.1683       0.7438
    Factor4 |      0.71341      0.14590      0.1427       0.8865
    Factor5 |      0.56751            .      0.1135       1.0000
------------------------------------------------------------------------
    LR test: independent vs. saturated:  chi2(10) =  398.10 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

    Variable |   Factor1   Factor2 |  Uniqueness
-------------+---------------------+--------------
       ideol |      0.4719   0.4019 |       0.6157
    equality |      0.4066   0.6424 |       0.4220
       owner |      0.6179  -0.5762 |       0.2861
      respon |      0.5807   0.4130 |       0.4922
 competition |      0.6619  -0.5056 |       0.3063
```

1.The sum of all eigenvalues=total number of varialbes
2. Proportion indicate the relative weight of each factor in the total variance. For example, 1.54525/5=0.3090. The first factor explains 30.9% of the total variance.
3. Cumulative shows the amount of variance explained by n+(n-1) factors. For example, factor 1 and factor 2 account for 57.55% of the total variance.
4. Uniqueness is the variance that is 'unique' to the variable and not shared with other variables. For example, 61.57% of the variance in 'ideol' is not share with other variables in the overall factor model. On the contrary 'owner' has low variance not accounted by other variables (28.61%). Notice that the greater 'uniqueness' the lower the relevance of the variable in the factor model.

5. Factor loadings are the weights and correlations between each variable and the factor. The higher the load is, the more relevant in defining the factor's dimensionality. A negative value indicates an inverse impact on the factor. Here, two factors are retained because both have eigenvalues over 1. It seems that 'owner' and 'competition' define factor1, and 'equality', 'respon' and 'ideol' define factor2. Factor 1 and Factor 2 can be considered as two latent variables.

```
. rotate

Factor analysis/correlation                       Number of obs    =      1125
    Method: principal-component factors           Retained factors =         2
    Rotation: orthogonal varimax (Kaiser off)     Number of params =         9

    Factor     Variance   Difference    Proportion   Cumulative
    Factor1     1.45169      0.02579        0.2903       0.2903
    Factor2     1.42590            .        0.2852       0.5755

    LR test: independent vs. saturated:  chi2(10) =  398.10 Prob>chi2 = 0.0000

Rotated factor loadings (pattern matrix) and unique variances

    Variable    Factor1    Factor2   Uniqueness
       ideol     0.0869     0.6138       0.6157
    equality    -0.1214     0.7505       0.4220
       owner     0.8446    -0.0218       0.2861
       respon    0.1610     0.6941       0.4922
 competition     0.8307     0.0603       0.3063

Factor rotation matrix

                Factor1   Factor2
    Factor1      0.7487    0.6629
    Factor2     -0.6629    0.7487
```

6. By default the rotation is varimax which produces orthogonal factors. This means that factors are not correlated to each other. This setting is recommended when you want to identify variables to create indexes or new variables without inter-correlated components.

7. The pattern matrix here offers a clearer picture of the relevance of each variable in the factor. The latent variable factor1 is mostly defined by 'owner' and 'competition' and latent variable factor2 is defined by 'equality', 'respon' and 'ideol'.

8. Latent variables can be created out of each cluster of variables. For example, 'owner' and 'competition' define one factor. You could aggregate these two to create a new variable to measure 'market oriented attitudes'. On the other hand you could aggregate 'ideol', 'equality' and 'respon' to create an index to measure 'egalitarian attitudes'. Since all variables are in the same valence (liberal for small values, capitalist for larger values), we can create the two new variables as:

```
gen market = (owner + competition)/2
gen egalitatiran = (ideol + equality +respon)/3
```

Appendix 4.A

9 Major industry categories from North American Industry Classification System

Table 4.A.1

9 Major industry categories from North American Industry Classification System

Major industry group	NAIC code
Agriculture, Forestry, Fishing and Hunting	1100
Mining and utilities	2100-2300
Manufacturing	3000-3999
Wholesale trade, Retail trade, transportation, warehouse and couriers	4000-4999
Information, Finance and Insurance, Real Estate and Rental and Leasing, Professional, Scientific, and Technical Services	5000-5999
Education service, health Care and Social Assistance	6000-6999
Arts, Entertainment, and Recreation, Accommodation and Food Services	7000-7999
Other Services (except Public Administration and private housholds)	8000-8999
Public Administration	9200

Appendix 4.B

6 Major occupation categories from Standard Occupational Classification System

Table 4.B.1

6 Major occupation categories from Standard Occupational Classification System

Major occupation group	SOC Code
Management, Business and Financial, Professional and Related	11-29
Service	31-39
Sales and Related, Office and Administrative Support	41-43
Farming, Fishing, Forestry, Construction and Extraction, Installation, maintenance, and Repair	45-49
Production, Transportation and Material Moving	51-53

Appendix 6.A

Regression results with the SIPP data

Table 6.A.1

Regression results with the SIPP data

This appendix contains full sets of regression results in table 5.6 and table 5.8

	Column 1	Column 2	Column 3	Column 4
Panel A: Linear, First-difference (Δ=2), with O*NET job attributes				
hourly_income	occ. and ind.	occ.	ind.	no occ. & ind.
variables [a]	Model 1 (std)	Model 2 (std)	Model 3 (std)	Model 4 (std)
fatal risk	0.027 (0.009)	0.091 (0.035)	-0.032 (0.22)	0.043 (0.019)
age	-0.015 (0.0029)	-0.008 (0.0014)	-0.009 (0.004)	-0.011 (0.004)
age2	-0.03 (0.014)	-0.04 (0.0014)	-0.003 (0.0015)	-0.003 (0.0015)
ugdeg	0.097 (0.087)	0.133 (0.086)	-0.091 (0.087)	0.128 (0.086)
college	-0.042 (0.064)	-0.062 (0.053)	-0.034 (0.633)	-0.051 (0.630)
hsgrad	0.021 (0.019)	0.019 (0.017)	-0.690 (0.38)	-0.691 (0.38)
hispanic				
black				
othrace				
female				
workov	0.168 (0.081)	0.162 (0.081)	0.173 (0.082)	0.169 (0.085)
union	0.069 (0.028)	0.070 (0.029)	0.070 (0.288)	0.072 (0.289)
kids 18	0.0014 (0.011)	0.0035 (0.011)	0.0046 (0.011)	0.0041 (0.016)
hipart	0.192 (0.053)	0.191 (0.154)	0.211 (0.153)	0.209 (0.155)

Table 6. A.1 Continued

hifull	0.021 (0.016)	0.021 (0.016)	0.023 (0.016)	0.023 (0.016)
empall	0.024 (0.251)	0.016 (0.260)	0.015 (0.257)	0.028 (0.261)
empsize	0.011 (0.017)	0.024 (0.017)	0.024 (0.017)	0.023 (0.016)
injury risk	0.010 (0.011)	0.009 (0.012)	0.012 (0.011)	0.011 (0.012)
region 1	-1.636 (1.336)	-1.60 (1.32)	-1.67 (1.32)	-1.62 (1.32)
region 2	-0.521 (0.885)	-0.445 (0.86)	-0.480 (0.90)	-0.389 (0.89)
region 3	-1.098 (0.956)	-1.070 (0.951)	-1.117 (0.950)	-1.075 (0.948)
region 4	-0.932 (0.900)	-0.842 (0.906)	-0.901 (0.903)	-0.796 (0.911)
region 5	-0.990 (0.760)	-0.906 (0.745)	-0.950 (0.760)	-0.853 (0.748)
region 6	0.803 (0.762)	-0.691 (0.751)	-0.887 (0.764)	-0.774 (0.762)
region 7	-0.008 (1.178)	0.026 (1.182)	-0.042 (1.168)	0.004 (1.172)
region 8	-0.571 (1.053)	-0.437 (1.032)	-0.572 (1.056)	-0.428 (1.037)
worcon_f1	0.595 (0.273)	0.677 (0.274)	0.391 (0.259)	0.466 (0.259)
worcon_f2	-0.225 (0.134)	-0.295 (0.234)	-0.202 (0.233)	-0.255 (0.232)
worcon_f3	-0.361 (0.317)	-0.435 (0.321)	-0.155 (0.309)	-0.209 (0.313)
mskill_f1	-0.129 (0.024)	-0.156 (0.213)	-0.093 (0.207)	-0.084 (0.207)
mskill_f2	-0.002 (0.130)	-0.016 (0.138)	0.050 (0.138)	0.024 (0.139)
intpeople_f1	0.081 (0.024)	0.170 (0.259)	0.290 (0.251)	0.415 (0.246)
intpeople_f2	-0.138 (0.298)	-0.110 (0.290)	-0.314 (0.285)	-0.304 (0.287)

Table 6.A.1 Continued

intpeople_f3	0.330 (0.022)	0.331 (0.027)	0.333 (0.021)	0.325 (0.226)
intpeople_f4	0.219 (0.021)	0.160 (0.021)	0.232 (0.021)	0.161 (0.022)
ind1	-0.109 (0.079)		0.114 (0.079)	
ind2	0.323 (0.062)		0.302 (0.613)	
ind3	-0.34 (0.064)		-0.36 (0.063)	
ind4	0.26 (0.061)		0.27 (0.061)	
ind5	0.341 (0.063)		0.376 (0.062)	
ind6	-0.300 (1.07)		-0.249 (0.845)	
Occ1	0.867 (0.153)	0.764 (0.153)		
Occ2	0.024 (0.156)	-0.109 (0.154)		
Occ3	0.464 (0.152)	0.298 (0.155)		
Occ4	-0.04 (0.141)	-0.181 (0.143)		
constant	0.504 (0.105)	0.506 (0.105)	0.503 (0.107)	0.508 (0.106)
N (#group)	1,880	1,880	1,880	1,880
R2(overall)	82%	78%	80%	75%
VSL (million $)	0.54	1.83		0.84
Panel B: Linear, First-difference (Δ=2), without O*NET job attributes				
	Model 5 (std)	Model 6 (std)	Model 7 (std)	Model 8 (std)
fatal risk	0.005 (0.0021)	0.049 (0.020)	-0.06 (0.31)	0.012 (0.0047)

Table 6.A.1 Continued

age	-0.004 (0.047)	-0.005 (0.047)	-0.005 (0.048)	-0.005 (0.048)
age2	-0.03 (0.014)	-0.03 (0.014)	-0.039 (0.015)	-0.03 (0.014)
ugdeg	-0.169 (0.088)	-0.052 (0.087)	0.029 (0.088)	0.097 (0.087)
college	-0.042 (0.064)	-0.066 (0.064)	-0.037 (0.064)	-0.042 (0.063)
hsgrad	-0.688 (0.387)	-0.691 (0.387)	-0.677 (0.384)	-0.681 (0.384)
hispanic				
black				
othrace				
female				
workov	0.172 (0.082)	0.168 (0.082)	0.182 (0.082)	0.180 (0.083)
union	0.681 (0.029)	0.725 (0.029)	0.679 (0.029)	0.709 (0.029)
kids 18	0.029 (0.011)	0.025 (0.011)	0.032 (0.011)	0.026 (0.012)
hipart	0.194 (0.153)	0.198 (0.154)	0.199 (0.153)	0.194 (0.155)
hifull	0.021 (0.016)	0.020 (0.015)	0.023 (0.016)	0.021 (0.016)
empall	0.032 (0.25)	-0.05 (0.26)	0.004 (0.25)	-0.04 (0.26)
empsize	0.237 (0.176)	0.227 (0.176)	0.231 (0.176)	0.221 (0.177)
injury risk	0.017 (0.013)	0.025 (0.011)	0.018 (0.009)	0.025 (0.011)
Region 1	-1.590 (1.328)	-1.570 (1.310)	-1.557 (1.330)	-1.533 (1.32)
Region 2	-0.461 (0.901)	-0.390 (0.880)	-0.420 (0.918)	-0.35 (0.899)
Region 3	-1.147 (0.959)	-1.113 (0.956)	-1.079 (0.966)	-1.022 (0.964)

Table 6.A.1 Continued

Region 4	-0.943 (0.910)	-0.851 (0.919)	-0.859 (0.915)	-0.753 (0.926)
Region 5	-1.060 (0.763)	-0.986 (0.749)	-0.972 (0.770)	-0.876 (0.756)
Region 6	-0.815 (0.763)	-0.712 (0.760)	-0.819 (0.774)	-0.711 (0.769)
Region 7	-0.024 (1.174)	-0.014 (1.184)	-0.028 (1.173)	-0.023 (1.184)
Region 8	-0.564 (1.057)	-0.434 (1.037)	-0.495 (1.063)	-0.345 (1.044)
ind1	1.13 (0.80)		1.182 (0.804)	
ind2	0.300 (0.620)		0.248 (0.619)	
ind3	-0.431 (0.637)		-0.484 (0.633)	
ind4	0.163 (0.617)		0.167 (0.615)	
ind5	0.273 (0.637)		0.274 (0.628)	
ind6	-0.333 (0.856)		-0.447 (0.847)	
Occ1	0.772 (0.399)	0.799 (0.384)		
Occ2	0.368 (0.409)	0.311 (0.290)		
Occ3	0.368 (0.522)	0.509 (0.510)		
Occ4	0.154 (0.449)	0.157 (0.426)		
constant	0.497 (0.105)	0.501 (0.103)	0.503(0.107)	0.508 (0.106)
N (#group)	1,880	1,880	1,880	1,880
R2(overall)	72%	63%	71%	58%
VSL (million $)		0.99		0.24

Appendix 6.B

Panel Study of Dynamic income (PSID) labor force data

This appendix describes the PSID sample in estimating the VSL. The worker's demographic characteristics, work-place fatal and non-injury risk and the O*NET job attributes that associated with worker's job are discussed below.

Table 6.B.1 summarizes and compares the PSID and the SIPP sample. The variables include wage, worker's demographic characteristics and the job attributes other than occupational fatal risk. The demographic characteristics age, educational attainment, race, U.S. citizenship, gender, usual work hours, union status, marital status, location of residence. In the PSID, the average hourly wage of workers is 12.52. If the weekly wage in the SIPP sample is converted to hourly wage by using the average of usual working hours reported in the data, the average hourly wage will be $16.25 which is considerably higher than the average hourly income of the PSID sample. Average age of the SIPP sample is 40 years old, which is very similar to the PSID sample. Compared to the SIPP sample, the educational attainment level is lower among the PSID sample. Only 6% of the sample graduated from a four year college, while there is about 28% of the sample in the SIPP graduated from a four year college. In the SIPP sample, there is 11.7% the sample is of Hispanic origin. The portion of Hispanic workers in the PSID is very small, 1%. There is about 45% of the SIPP sample is female, only 21% of workers in the PSID sample is female. In general, comparing to the SIPP, the PSID represents more male workers with lower education attainment.

For the survey year of 2003, 2005, 2007, PSID uses 4-digit occupation code and 4-digit industry code from 2000 Census Occupation Classification System to classify the occupation and industry. Based on the same cross-walk and the same approach in Scotton (2000, 2010), I coarsely aggregate the occupation and industry groups based on 3-digit level and merge the fatal risk rates created from the CFOI to the corresponding 23 occupation 22 industry groups.

The risk is created annually from the 2003-2007 to match the period of workforce data. After being merged to the PSID data, the sample represents 242 of the 419 possible jobs defined by the risk rate square matrix. The highest risk level is 4.64×10^{-4}, which is faced by the construction and extracting workers in mining industry. The lowest risk level is 0.007×10^{-4}, which is faced by computer and mathematical workers in professional, scientific, and technical services industry. The mean of fatal rate in the PSID sample is 0.41×10^{-4} with standard deviation of 0.44×10^{-4}. The median is 0.21×10^{-4}. About 3.1% of total observations, 212 observations in the PSID, face zero risk.

Table 6.B.2 provides comparison of fatal risk rates' distribution before and after merging to the PSID sample. Panel A compares the risk rates at a few important percentiles. Notice the risk rates of PSID sample are higher at most of the percentiles (25%, 50% and 75%). At the higher percentile, however, the risk rates are smaller. This difference implies that jobs that are not observed in the PSID sample are more likely the ones with low risk or extremely high (top five percent) fatal risk rates. Panel B provides the comparison through the distribution of jobs at different risk rate levels. It appears that there are substantially less

occupation-industry pairs having zero fatal risk rates in the sample. Jobs in the PSID sample are more densely distributed at the median low level of risk rates (0< risk rates ≤ 0.5). It indicates that the PSID sample represents more of workers in the median level risky jobs and less of the workers in the extreme low or high jobs.

The non-fatal injury risk data is from the 2003, 2005, 2007 IIF. Since the industry coding system in the PSID is the same as the one in the SIPP, I apply the same approach to convert 4-digit industry codes in PSID to 3-digit 2000 NAICS categories (see appendix 3.5). Once the 3-digit level injury risk rates are merged to the period of the PSID data, the injury risk of the sample in the PSID differentiates by 70 industries. The highest risk bearing industry is the couriers and messengers, where the risk is 0.074. The lowest risk bearing group is the information, finance and insurance industry where the risk is 0.004.

In an attempt to understand the correlation between the non-fata, injury-risk and the fatal injury risk, I also report summary statistics for non-fatal, injury-risk rates on a more aggregated level. Table 6.B.3 presents the injury risk rate measures aggregated into 19 industries that are observed in the sample and the comparison to the fatal risk rate measures aggregated on the same level. The highest non-fatal, injury-risk industry is couriers and messengers, where the risk rate is 743 per 10,000. The second highest non-fatal injury risk bearing industry is the transportation and warehouse industry where the injury risk rate is 433 per 10,000. The lowest risk bearing group is the information, finance and insurance industry where the risk rate is 40 per 10,000. In comparison to the fatal risk data, it seems that

construction and transportation industries tend to have both high fatal and non-fatal injury risks, while other industries do not appear to have this consistency.

In addition to commonly included wage determinants, the occupational attributes created from O*NET is merged to the PSID to enrich the choices of wage determinants variables in order to increase the accuracy of the empirical estimates. Occupational attributes data is created and merged to PSID in the same way that is applied to match with the SIPP.

The occupational classification system of PSID is the same as the one in SIPP. I convert the O*NET data in 176 occupation groups in the same way as in the PSID. Table 6.B.4 reports the summary statistics for the 15 job attributes across occupations held by workers in the SIPP sample. The statistics are reported by gender and are consistent with the intuition, indicating that men are in jobs with higher requirements for physical practice or demand of motor skill associated with machines or tools, where women tend to sort into the jobs requiring more interaction with people.

Table 6.B.5 reports the correlations coefficients between the job characteristics and workplace fatal and non-fatal injury risks. In general, the correlation between fatal risk, injury risk and other job attributes are consistent with the intuition. Based on the PSID sample, injury risk is positively related to the fatal risk. Hazard working conditions are positively correlated to the fatal risk, where the high level of hazard working condition (exposure 2) has the highest correlation. As it is expected, the job requirements involving technical or electronic or mechanical device or equipment handling and repetitive physical motions (cognitive skill 3, physical demand 2, 3 and motor skill 1) also show positive

correlations to the fatal risk. Moreover, the direction of correlation between the other job attributes and non-fatal injury risk are consistent with the one between them and fatal risk.

Table 6.B.1

Worker characteristics and job characteristics for the PSID & SIPP sample

Variable	Definition	PSID (N=6,021)	SIPP (N=5676)
hr_wage	hourly income	12.52 (10.01) [2 100]	19.08 (9.56) [3.5 73.19]
age	age in years	40.19 (11.73) [18 65]	43.52 (9.56) [19 64]
blacknh	1 if individual is black and non-hispanic	0.07	0.1
college	1 if individual attended college but with no degree	0.47	0.35
female	1 if individual is female	0.21	0.44
hispanic	1 if individual has a hispanic origin	0.01	0.11
hsgrad	1 if individual graduated from high school	0.89	0.28
kid18	1 if individual has any kids under 18 years old	0.46	0.43
married	1 if individual is married	0.58	0.69
othrace	1 if individual is non-white, non-black, non-hispanic	0.03	0.04
ugdeg	1 if individual have bachelor degree or more	0.06	0.28
	Job Characteristics		
salaried	1 if individual gets monthly payment	0.38	0.51
workot	1 if individual usually works more than 40 hours	0.22	0.28
union	1 if individual is a union member or covered by union	0.14	0.17
	Worker's location		
Region1/9			
	individual lives in New England Division	0.09	0.05
	individual lives in Middle Atlantic Division	0.05	0.14
	individual lives in East North Central Division	0.10	0.17
	individual lives in West North Central Division	0.05	0.08
	individual lives in South Atlantic Division	0.11	0.18
	individual lives in East South Central	0.08	0.07

Table 6.B.1 Continued

individual lives in West South Central	0.16	0.08
individual lives in Mountain	0.14	0.07
Individual lives in Pacific	0.22	0.13

Table 6.B.2

Distribution of fatal risk rates created from CFOI before / after merging to the PSID [a]

Risk rate Measure	22 occ x23 ind.	
	Before merge to the PSID N [b] =419	After merge to the PSID N=242
Panel A. distribution of risk rates		
25%	0 [c]	0.07
50%	0.09	0.21
75%	0.38	0.58
90%	1.01	1.26
95%	1.92	1.52
99%	4.65	3..07
Max	17.05	4.63
Panel B. Number of jobs		
risk rate=0	31.02%	21.60%
risk rate>0 & ≤ 0.25	35.32%	41.20%
risk rate>0.25 & ≤ 0.50	10.73%	12.09%
risk rate>0.50 & ≤ 1.0	11.21%	11.55%
risk rate>1.0	11.45%	13.56%

a. Rates are reported as deaths per 10,000 workers.
b. *N* is the number of the jobs presented in the data.
c. The cell value is the risk rate per 10,000 workers.

Table 6.B.3

Summary of non-fatal injury risk and fatal risk across 19 industry groups in the PSID sample

			Non-fatal injury rate (per 10,000)			fatal risk rate (per 10, 000)		
Ind.Code	Industry group	% of workers [a]	Mean (std.)	Min	max	Mean (std.)	Min	max
2100	mining	0.60%	191.11 (71.02)	[80	310]	2.51 (1.54)	[0	4.63]
2200	utility	1.50%	221.93 (12.35)	[210	240]	0.44 (0.40)	[0	1.22]
2300	construction	13.69%	322.98 (34.50)	[280	360]	1.07 (0.47)	[0.06	3.44]
3123	manufacturing	19.43%	347.29 (129.99)	[100	710]	0.27 (0.23)	[0.17	1.14]
4200	wholesale trade	5.89%	274.31 (66.13)	[220	380]	0.55 (0.49)	[0	1.77]
4445	retail trade	11.33%	273.67 (105.53)	[100	430]	0.23 (0.15)	[0	0.89]
4800	transportation and warehousing	3.64%	433.30 (166.23)	[100	800]	1.70 (1.30)	[0.0	3.07]
4923	currier and messengers, warehouse and storage	1.22%	743.15 (96.52)	[540	880]	0.63 (0.42)	[0	1.29]
5100	information	3.24%	207.86 (154.61)	[50	480]	0.18 (0.43)	[0	4.33]
5200	finance and insurance	5.02%	40 (4.72)	[30	40]	0.05 (0.03)	[0	0.11]
5300	real estate and rental and leasing	1.72%	193.4 (23.62)	[160	210]	0.31 (0.40)	[0	1.83]
5400	professional, scientific, and technical services	6.39%	56.28 (4.83)	[50	60]	0.11 (0.20)	[0	1.83]

Table 6.B.3 Continued

5500	management of companies and enterprises	0.13%	128.75 (18.85)	[90	160]	0		
5600	administrative and support and waste management and remediation services	2.71%	251.48 (128.52)	[170	570]	0.89 (0.94)	[0	3.95]
6100	educational services	2.51%	107.2 (9.63)	[100	120]	0.10 (0.17)	[0	1.04]
6200	health care and social assistance	9.64%	307.37 (172.58)	[100	630]	0.08 (0.13)	[0	1.41]
7100	arts, entertainment, and recreation	1.42%	276.47 (35.81)	[200	310]	0.41 (0.51)	[0	2.61]
7200	accomodation and food services	5.64%	166.88 (67.34)	[130	360]	0.22 (0.47)	[0	4.24]
8100	other services	3.86%	162.51 (33.30)	[110	210]	0.43 (0.30)	[0	1.65]

a. Both of the non-fatal injury risk rates and fatal risk rates are aggregated into 19 industry groups by using 2-digit NAICS code. Thus, the percentage of the workers in each group is the same for non-fatal injury risk and fatal risk measures.

Table 6.B.4

Summary of occupational attributes of the PSID sample by gender

Characteristics		Description	Male (N=4,712) Mean (std.)	Female (N=1,246) Mean (std.)
Exposure	1	low level of hazardous work environment	2.62 (0.90)	2.10 (0.60)
	2	providing working environment exposed to the weather or hot, cold temperature	2.20 (0.76)	1.54 (0.47)
	3	high level of hazardous working environment, specifying technical or electronic or mechanical device or equipment	2.26 (0.31)	2.51 (0.49)
Cognitive demand	1	cognitive demands involving information processing and analyzing data or information	3.49 (0.44)	3.40 (0.52)
	2	cognitive demands involving creative thinking, strategies developing, planning, and decision making	3.34 (0.47)	3.12 (0.48)
	3	cognitive demands involving monitoring process, inspecting equipment or material	3.36 (0.41)	3.08 (0.44)
Physical demand	1	physical demands requiring physical activities	2.80 (0.81)	2.60 (0.78)
	2	requires climbing ladders, poles, or kneeling and crawling	2.26 (0.43)	2.04 (0.28)
	3	requires repetitive physical motion, operating machine or handling tools	3.00 (0.76)	2.90 (0.54)
Motor skill	1	specifying technical or electronic or mechanical device or equipment	2.16 (0.68)	1.65 (0.42)
	2	interacting with computer	3.26 (1.01)	3.38 (0.99)
People skill	1	requires interacting with people indirectly	3.49 (0.68)	3.46 (0.61)
	2	interacting with people directly, or involving in the conflicting situations	2.97 (0.36)	3.20 (0.41)
	3	requires group or team working and taking responsibility for the outcome of the performances	3.64 (0.45)	3.47 (0.44)
	4	Working directly with public. Each of the jobs is differentiated by the job attributes and the injury risk	3.17 (0.59)	3.33 (0.54)

Table 6.B.5

Pairwise correlations of job characteristics and injury risks [a]

		Fatal	Injury	Exposure			Cognitive skill			Physical demand			Motor skill		People skill			
				1	2	3	1	2	3	1	2	3	1	2	1	2	3	4
Fatal risk		1																
Injury risk		0.11	1.00															
Exposure	1	0.29	0.19	1.00														
	2	0.55	0.09	0.47	1.00													
	3	-0.14	0.10	0.40	-0.27	1.00												
Cognitive skill	1	-0.19	-0.09	-0.15	-0.22	0.18	1.00											
	2	-0.12	-0.06	-0.01	0.00	0.20	0.78	1.00										
	3	0.20	0.10	0.62	0.26	0.39	0.43	0.53	1.00									
Physical demand	1	0.06	0.18	0.63	0.20	0.39	-0.58	-0.36	0.20	1.00								
	2	0.39	0.23	0.66	0.57	0.23	-0.20	-0.10	0.46	0.44	1							
	3	0.27	0.14	0.62	0.25	0.21	-0.46	-0.47	0.24	0.67	0.44	1						
Motor skill	1	0.35	0.10	0.72	0.49	0.12	0.01	0.14	0.62	0.31	0.54	0.54	1					
	2	-0.32	-0.22	-0.53	-0.33	-0.20	0.68	0.42	-0.16	-0.73	-0.55	-0.57	-0.2	1				

Table 6.B.5 Continued

	1	-0.29	-0.19	-0.51	-0.17	-0.06	0.70	0.69	-0.04	-0.70	-0.44	-0.72	-0.29	0.75	1			
	2	-0.09	0.07	0.20	-0.14	0.76	0.27	0.35	0.35	0.23	0.16	-0.02	0.004	-0.07	0.17	1		
	3	-0.04	0.04	0.43	0.19	0.41	0.19	0.47	0.53	0.33	0.25	0.04	0.29	-0.14	0.15	0.55	1	
People skill	4	-0.10	-0.13	-0.33	-0.06	0.14	0.27	0.45	-0.04	-0.13	-0.34	-0.03	-0.23	0.32	0.54	0.46	0.19	1

a. The entry of the cell is the correlation coefficient.

Appendix 6.C

Regression results with the PSID data

Table 6.C.1

This appendix contains the full sets of results in table 6.3 and table 6.4.

	Column 1	Column 2	Column 3	Column 4
Panel A: Linear, First-difference (Δ=2), with O*NET job attributes				
Variables [a]	occ. and ind.	occ.	ind.	no occ. & ind.
hourly_income	Model 1 (std)	Model 2 (std)	Model 3 (std)	Model 4 (std)
fatal risk	0.78* (0.39)	0.74** (0.38)	0.77* (0.44)	0.81* (0.46)
female				
hispanic				
black				
othrace				
age	0.87 (0.27)	0.862 (0.280)	0.898 (0.279)	0.881 (0.284)
age2	-0.006 (0.003)	-0.006 (0.003)	-0.006 (0.003)	-0.006 (0.003)
kid18	-0.454 (0.481)	-0.460 (0.483)	-0.456 (0.487)	-0.460 (0.482)
ugdeg	9.571 (1.801)	9.93 (1.760)	9.725 (1.734)	9.927 (1.684)
college	-1.33 (0.64)	-1.355 (0.636)	-1.347 (0.646)	-1.392 (0.640)
hsgrad	-6.67 (3.688)	-5.837 (3.784)	-6.199 (3.443)	-5.128 (3.798)
salary	-12.169 (1.283)	-11.886 (1.279)	-12.171 (1.283)	-11.814 (1.289)
workot	0.629 (0.295)	0.666 (0.296)	0.611 (0.294)	0.652 (0.296)
union	1.181 (0.601)	1.234 (0.607)	1.145 (0.600)	1.228 (0.611)
married	-0.569 (0.704)	-0.595 (0.710)	-0.575 (0.709)	-0.561 (0.720)

Table 6.C.1 Continued

northeast	3.218 (3.242)	3.31 (3.31)	3.275 (3.26)	3.499 (3.417)
northcentral	3.527 (3.273)	3.528 (3.350)	3.584 (3.290)	3.713 (3.449)
south	3.744 (3.409)	3.906 (3.469)	3.671 (3.435)	3.933 (3.567)
injury risk	-0.001 (0.0009)	-0.001 (0.0009)	-0.001 (0.001)	-0.001 (0.001)
worcon_f1	1.217 (0.699)	1.215 (0.704	1.138 (0.705)	1.172 (0.708)
worcon_f2	-1.217 (0.699)	-1.107 (0.493)	-0.853 (0.477)	-0.797 (0.482)
worcon_f3	-1.065 (0.494)	0.735 (0.731)	0.284 (0.730)	0.365 (0.721)
mskill_f1	-0.850 (0.737)	-0.896 (0.704)	-0.671 (0.660)	-0.624 (0.661)
mskill_f2	-0.057 (0.387)	-0.125 (0.393)	0.093 (0.368)	0.061 (0.369)
intpeople_f1	0.021 (0.668)	0.297 (0.686)	-0.033 (0.641)	0.315 (0.653)
intpeople_f2	-0.689 (0.983)	-0.323 (0.989)	-0.973 (0.995)	-0.877 (0.995)
intpeople_f3	0.84 (0.39)	0.695 (0.608)	0.777 (0.608)	0.618 (0.623)
intpeople_f4	0.73 (0.66)	0.284 (0.658)	1.016 (0.663)	0.654 (0.661)
ind1 [b]	0.807 (1.086)		1.013 (1.067)	
ind2	0.778 (0.965)		0.333 (0.021)	
ind3	0.363 (0.086)		0.756 (0.952)	
ind4	1.36 (1.05)		1.207 (1.026)	
ind5	1.889 (1.208)		1.447 (1.197)	

Table 6.C.1 Continued

ind6	-3.621 (1.277)		-4.621 (1.282)	
ind7	9.571 (1.802)			
Occ1 [c]	-0.563 (1.089)	-0.754 (1.070)		
Occ2	-2.893 (1.227)	-3.955 (1.199)		
Occ3	-0.329 (1.132)	-0.702 (1.106)		
Occ4	-0.619 (0.760)	-0.702 (1.106)		
constant	0.228 (0.279)	0.304 (0.277)	0.29 (0.27)	0.282 (0.279)
N (#group)	2,356	2,356	2,356	2,356
R2(overall)	25%	24%	25%	20%
VSL (million $)	15.4	14.8	15.4	16.3
Panel B: Linear, First-difference (Δ=2), without O*NET job attributes				
	Model 5 (std)	Model 6 (std)	Model 7 (std)	Model 8 (std)
fatal risk	0.69* (0.30)	0.62* (0.30)	0.61* (0.28)	0.66** (0.29)
female				
hispanic				
black				
othrace				
age	0.901 (0.279)	0.889 (0.284)	0.922 (0.288)	0.912 (0.286)
age2	-0.006 (0.003)	-0.006 (0.003)	-0.006 (0.003)	-0.006 (0.003)
kid18	-0.479 (0.482)	-0.502 (0.480)	-0.462 (0.479)	-0.483 (0.475)
ugdeg	10.644 (1.648)	11.130 (1.6188)	10.52 (1.589)	10.962 (1.533)

Table 6.C.1 Continued

college	-1.368 (0.650)	-1.402 (0.642)	-1.366 (0.656)	-1.400 (0.648)
hsgrad	-6.551 (3.312)	-5.577 (3.710)	-6.179 (3.217)	-5.192 (3.704)
salary	-12.134 (1.282)	-11.84 (1.279)	-12.109 (1.277)	-11.750 (1.271)
workot	0.691 (0.294)	0.736 (0.295)	0.665 (0.293)	0.711 (0.294)
union	1.058 (0.621)	1.144 (0.631)	0.996 (0.614)	1.094 (0.627)
married	-0.573 (0.701)	-0.614 (0.709)	-0.586 (0.705)	-0.569 (0.717)
northeast	3.037 (3.130)	3.22 (3.26)	3.059 (3.115)	3.384 (3.347)
northcentral	0.237 (0.176)	3.393 (3.299)	3.359 (3.142)	3.561 (3.178)
south	3.365 (3.159)	3.919 (3.416)	3.529 (3.289)	3.902 (3.492)
injury risk	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
ind1	0.919 (1.098)		1.085 (1.086)	
ind2	0.781 (0.970)		0.704 (0.954)	
ind3	0.314 (0.858)		0.364 (0.854)	
ind4	1.254 (1.054)		1.120 (1.034)	
ind5	2.058 (1.182)			
ind6	-3.723 (1.369)			
Occ1	-0.266 (0.900)	-0.371 (0.874)		
Occ2	-1.841 (0.976)	-2.599 (0.934)		

Table 6.C.1 Continued

Occ3	-0.001 (0.824)	-0.709 (0.629)		
constant	0.211 (0.281)	0.190 (0.280)	0.173 (0.280)	0.262 (0.282)
N (#group)	2,356	2,356	2,356	2,356
R2(overall)	18%	16%	18%	14%
VSL (million $)	13.8	12.5	12.2	13.2

a. The variables for worker's characteristics, workplace fatal risk and non-fatal injury risk, and job attributes are defined in Table 5.C.1.
b. Seven major occupation categories are (agriculture and military industry are excluded in the sample):
   Ind1: mining, utility and construction
   Ind2: food processing, texitiles and equip manufacturing
   Ind3: retail and transportation
   Ind4: fire, telecomds and publishing
   Ind5: education, health care
   Ind6: entertainment and hotels
c. Five major occupation categories are:
   Occ1: management, business and financial
   Occ2: service
   Occ3: sales and office
   Occ4: natural resources, construction
   Occ5: production, transportation