Job Amenities & Earnings Inequality

Alex Bell, Harvard University*

October 31, 2019

Job Market Paper (click here for current draft & here web summary)

Abstract

Previous research on prices of job amenities has suffered from simultaneity bias due to unobserved worker ability, resulting in apparently “wrong-signed” compensating wage differentials. I propose a new estimator for amenity prices that uses only a single imprecise proxy for workers’ ability to identify the amenity price that holds ability fixed. My estimation strategy removes imprecision from the ability proxy by using predicted values from a regression of the ability proxy on wages and amenities. With price estimates for a set of observed job characteristics, I turn to investigating the role of job amenities in demographic income gaps. I find a large role for costly amenity substitution in explaining the gender pay gap. In contrast, substitution on the basis of observed amenities does not appear to play large roles in income inequalities by race or by parent background.

*Contact: alexanderbell@fas.harvard.edu. I am grateful to Raj Chetty, Roland Fryer, Nathan Hendren, Larry Katz, and Elie Tamer for research advice during graduate school. This work has also benefited from insightful conversations with Isaiah Andrews, Gary Chamberlain, David Cutler, Edward Glaeser, Claudia Goldin, Sandy Jencks, Corinne Low, Amanda Pallais, and Ariel Pakes. Finally, these ideas would not have gotten off the ground without extensive early-stage debates with Sophie Calder-Wang and Alex Olssen.
I Introduction

Workers differ both in terms of total compensation and how they split their compensation into wages and amenities. Determinants of total compensation may include productivity, discrimination by employers, search frictions, and rents. But income differences also arise when workers trade off between wages and other desirable attributes. This paper decomposes earned income gaps by gender, race, and parent background into differences in these two components. In doing so, I put forward a new methodology for estimating the trade-offs that workers face between income and observable non-monetary characteristics of their jobs.

The idea that job amenities command a wage premium, or compensating wage differential, dates back at least to Smith (1776). Two centuries, later Brown (1980) best summarized the existing empirical literature as holding “some clear support for the theory but an uncomfortable number of exceptions.” More recently, quasi-experimental evidence has revived interest in the theory (e.g., Lavetti (2018)). A persistent barrier to studying the full impact of compensating differentials on income inequality is the inherent difficulty of estimating prices for a wide set of job amenities in a sample representative of the labor market.

The problem stems from the researcher’s inability to isolate workers who had similar offer sets to each other but chose to split their compensation differently into wage and amenities. A classic illustration is that of physical safety on the job. Supposing workers tend to prefer jobs with higher levels of physical safety, one would expect workers in safer jobs to be paid less. Figure 1 shows the opposite is true in the NLSY. In general, safer jobs pay more, not less. The goal, however, is to look at how income and safety vary among workers who share the same level of productivity, and therefore could have obtained similar types of jobs. Panel B shows that although controlling for observable productivity-relevant characteristics weakens the relationship, in this case it does not even change it to the expected sign. Such a result might be expected if the productivity channels that determine a worker’s compensation are not fully observed. In the absence of a full set of controls, income and safety are both determined by differences in (unobserved) productivity and preferences. The regression of
wage on safety conflates compensating differentials with the returns to skill, much in the same way that a regression of quantities on prices conflates supply and demand. Controlling for a few observed ability measures has done as much to solve this simultaneity problem as controlling for a few shifters of supply would in a regression of price on quantity.\textsuperscript{1}

The core contribution of this paper is to demonstrate the data used to construct Figure 1 is more than sufficient to remove the simultaneity bias previously seen as inherent to observational analysis of compensating differentials. Similar to instrumental variables, my approach can be viewed as treating an observed measure of ability as a shifter of ability.\textsuperscript{2} It differs from IV estimation in that the variable that needs to be shifted – ability – is not observed.\textsuperscript{3}

The OLS approach to tracing out wage-amenity variation generated by preferences requires the researcher to control for the full set of variables that determine workers’ abilities. In contrast, my approach requires only one such variable. Both approaches, it should be noted, are sensitive to a normalization on the part of the researcher as to which variable (or variables) belong in the ability control. My approach allows me to explore the robustness of results when the ability channel is proxied by any one of various types of cognitive, social, and physical characteristics of workers.

Figure 2 illustrates the difference between OLS and my proposed strategy using hypothetical data constructed to mirror the important facets of real labor market data. Suppose the researcher observes only three coarse bins of workers’ ability, represented by colors. Clearly, the correspondence between observed and actual ability is quite noisy; some observably low-

\textsuperscript{1} Another analogy can be drawn to discrete choice analysis, in which choices are regressed on product characteristics (typically using logistic regression). That framework assumes that binary choices are informative of which options are better. My framework is similar, but uses an observed measure of worker productivity rather than choice. The typical estimating equation for compensating differentials in representative datasets is analogous to a regression of one product characteristic on other characteristics controlling for which product was chosen, which is generally not consistent as described by McFadden (1973).

\textsuperscript{2} What I refer to here as “ability” should also be thought to include other factors that lead workers to get more or less compensation that would not typically be called “ability,” such as connections, search frictions, or discrimination.

\textsuperscript{3} My approach shares this feature in common with Olley and Pakes (1996), who describe how observed investment can identify the unobserved productivity term in production functions.
skilled workers enjoy very high wage and amenity. Panel A depicts the standard regression
of wage on amenity with indicators for observed ability (as an average of regressions within
observed ability levels). The estimating equation is as follows, with coefficient of interest \( \hat{\beta} \):

\[
\hat{E}^*[\text{wage}|\text{amenity, obs. ability}] = \hat{\beta}\text{amenity} + I[\text{obs. ability}]
\]

Because we observe workers’ true abilities with noise, within every level of observed
ability there is still an upward-sloping relationship between wage and amenity. Hence, the
wage-safety relationship of Figure 1 remains upward-sloping, even with controls for observed
ability.

In contrast, Panel B sketches two ways to implement my proposed estimation strategy
in a linear framework. Both approaches begin the same way. Rather than using observed
ability as a control, it is the outcome of a regression on both wage and amenity. I call this
Step 1 on the graph:

\[
\hat{E}^*[\text{obs. ability}|\text{wage, amenity}] = \hat{\delta}\text{wage} + \hat{\pi}\text{amenity}
\]

Although the regression of Step 1 is not structural, it does provide some intuition.\(^4\) If
wage and amenity are two forms of compensation that workers enjoy, we should expect both \( \hat{\delta} \) and \( \hat{\pi} \) to be positive. The relative magnitudes of the coefficients reveal the tendency
for higher-ability workers to get more wage and more amenity. The regression also has a
graphical interpretation in wage-amenity space. It points in the direction in which observed
ability is increasing. This direction is orthogonal to what we want to estimate; we seek to
estimate the wage-amenity trade-off when ability is held constant. The orthogonal direction
is labeled as Step 2 in the graph, and it can be obtained either from a ratio of coefficients

\(^4\)This equation differs from the “reverse” approach described in Goldberger (1984) because I do not regress
on the exogenous variable of race; all regressors here are structural outcomes. Whereas the typical estimation
strategy for job amenities is valid only under the assumptions of reverse regression (no search frictions), my
approach assumes unobserved ability and search frictions cause both wage and amenity. Online Appendix
D.A contrasts these assumptions.
or using predicted values from Step 1. In a linear model, the two approaches produce numerically equivalent estimates (much like instrumental variables and control functions).

The simplest approach is to read off the perpendicular line to the regression line of Step 1. If Step 1 told us the direction in which skill increases, the orthogonal direction must describe wage-amenity variation that holds skill constant, which is the relative price a worker faces. Because the regression line of Step 1 has a slope of \( \hat{\beta} \), the slope of the orthogonal line and thus the price estimate is \( -\frac{\hat{\beta}}{\hat{\pi}} \). I refer to this procedure as the ratio-of-coefficients approach; the instrumental variables estimator is also often thought of as a ratio of coefficients from reduced-form to first-stage regressions.

An alternative approach uses predicted values from the regression of Step 1. The strength of the predicted-values approach, as I elaborate on in Section III, is that it can easily accommodate a non-parametric estimator in place of the Step 1 linear regression. Referring to these predicted values as \( \hat{\text{ability}} \), Step 2 of the predicted-values approach is to regress wage on amenity with the inclusion of the constructed ability control:

\[
\hat{E}^*[\text{wage}|\text{amenity, ability}] = \hat{\beta}_{\text{amenity}} + \hat{\gamma}\hat{\text{ability}}
\]

This regression also need not be linear, although \( \hat{\beta} \) from the linear regression is numerically equivalent to the ratio of coefficients, \( -\frac{\hat{\pi}}{\hat{\beta}} \). In this case, because \( \hat{\text{ability}} \) is a linear combination of the outcome and a regressor, this approach yields a perfect fit; standard errors from the second-stage regression are not meaningful (I discuss identification-robust inference in Section IV.B). To build intuition for this predicted-values approach, recall that the basic structure of the problem is that workers’ wages and amenities are both caused by ability and preferences. The goal is to build a control variable for ability such that at a given level of the control, workers still obtain different combinations of wages and amenities (due to preferences). Thus, the ideal control variable for ability can be written as some combination of wages and amenities. A prediction exercise of observed ability guides the construction of that ideal control variable. By assumption, in a large sample we would obtain the same
predicted values if either observed or true ability were used as the dependent variable in Step 1 – both of which correspond to true ability – so \( \hat{\text{ability}} \) identifies true ability.

Whereas either estimation approach overcomes the problem of ability bias in observational data, a key limitation remains: in observational data, I still cannot price characteristics that I cannot observe. Real workers make choices taking into account many aspects of jobs. Although my estimator readily extends to pricing many amenities jointly, I do not claim to observe the full set. One operationalization of a compensating differential, to which I do not adhere in this paper, is the wage trade-off a worker faces when an amenity increases by a unit, and all other job attributes are held constant. Instead, I estimate the trade-offs workers face when they choose a typical job that has a unit more of the amenity. Both quantities may be interesting, and there is a guiding framework by which I think of the two as related. Empirically, I find that observed amenities tend to be correlated with each other such that observed amenities’ prices tend to be smaller when more amenities are priced together. If my observed amenities can be considered a random sample of all amenities, it seems likely that such a positive correlation would also exist between observed and unobserved amenities (Altonji et al., 2005; Oster, 2019). This general bundling of amenities is likely to lead my price estimates to be larger in absolute magnitude than price estimates that hold all other amenities fixed.

Although the problem of unobserved amenities calls for caution when interpreting price estimates, a positive correlation between observed and unobserved amenities if anything would strengthen my analysis of demographic pay gaps. The fact that I cannot price all the dimensions in which work differs suggests that I am likely to under-state the importance of amenities as a whole in pay gaps. But to the extent that many of the good amenities are bundled, an analysis that only takes into account the observed ones – at prices inclusive of how unobserved ones covary – may be a more reasonable approximation of the overall role of amenities. Again turning to the empirics, this paper’s findings on demographic pay gaps seem unlikely to be driven by problems of unobserved amenities. For instance, I find that
conditional on total wage and amenity compensation, women tend to earn better amenities and lower income than men. This finding implies amenities play a substantial role in the current gender pay gap, which seems difficult to explain by the problem of observing too few amenities. In stark contrast, I find more limited roles for the same set of amenities to explain gaps by race and by parent background. Unlike the gender gap, these income inequalities are shown to be primarily absolute gaps in total compensation rather than a function of how workers split their compensation into wages and amenities. Although the possibility still remains that additional racial and intergenerational income inequality is generated by amenities other than those I observe, my findings are consistent with a literature that has focused on the role of the absolute factors of skills and discrimination in causing these gaps. My findings on the gender gap are also consistent with a modern literature pointing to the importance of how work is remunerated.

The outline for the rest of the paper is as follows. In the remainder of this section, I relate my work to previous research on compensating differentials, including previous attempts to diagnose and solve the problem of ability bias highlighted by Figure 1. In Section II, I provide my rendition of the usual structural model of compensating differentials; I discuss the identification assumption in the context of that structural model and in the reduced-form. Section III proves formally that the Rosen frontier is non-parametrically identified by observational data, and Section IV discusses estimation in the linear model, including potential sources of bias. Section V explores amenity price estimates in the NLSY linked with O*NET occupation data. Section VI relates these price estimates to differences in quantities of amenities by gender, race, and parent background. Finally, Section VII concludes.

I.A Relation to Literature Pricing Job Amenities

This paper builds on the conceptual framework of a labor market sorting equilibrium put forward by Thaler and Rosen (1976) and Rosen (1986). The only stylistic difference is that I more explicitly model unobserved skill heterogeneity. The potential bias from unobserved
skill heterogeneity in such a model has been shown and calibrated by Hwang et al. (1992), whose final paragraph nicely motivates this paper:

Where then should future research on the estimation of compensating wage differentials be directed? In our opinion, significant progress is most likely to come from new econometric methods that are better able to address the problem of unobserved productivity heterogeneity. We hope that this study will stimulate renewed efforts along this line.

An early wave of empirical literature on compensating differentials attempted to find evidence for the theory in nationally representative datasets, and the present paper uses many of the same datasets. Lucas (1977) offered “some of the first systematic empirical support for Adam Smith’s notion of equalizing wage differentials.” Lucas linked individual-level data from the Survey of Economic Opportunity with newly available occupational-level data from the Dictionary of Occupational Titles. Six categories of occupational characteristics were investigated, including sedentariness, repetitiveness, and physical conditions such as risk of injury. These job characteristics were priced in hedonic hourly wage equations, with productivity differences proxied by controls for race, gender, age, years of education, and union membership. Results were imprecise and mixed, but generally indicated that workers taking on repetitive work and work in poor physical conditions – both thought to be undesirable – were earning slightly higher wages as a result.

Brown (1980) suspected that previous studies testing relationships between wage and amenities controlling for observed productivity measures were biased by “the omission of important worker abilities.” To better control for unobserved productivity differences across workers, Brown introduced worker fixed effects into a panel derived from the NLSY also linked with DOT occupation data. Ultimately, Brown’s within-worker comparisons of wages and amenities proved quite similar to the cross-worker comparisons – that is, often apparently

5The SEO is now part of the PSID and DOT is now O*NET.
wrong-signed. My interpretation of this result is that even within-worker, there is a great deal of idiosyncrasy in total realized utility across job spells. Rather than differences in jobs across workers being caused by worker-level skill, this result suggests a great role for the type of idiosyncrasy in overall utilities emphasized by Mortensen (2003), which I model explicitly in the next section.

Sorkin (2018) also leverages worker moves, but across firms rather than across occupations, and finds some evidence in support of compensating differentials. Under strong assumptions about what information is conveyed when workers move across firms, Sorkin (2018) estimates that 15% of the overall variation in earnings across the economy is due to firm-level compensating differentials. He also concludes that half of the firm component of the variance of earnings is accounted for by these compensating differentials. His approach is not informative as to which job amenities different workers may value.

Most recent literature has used quasi-experimental methods to isolate variation in wage due to particular job characteristics. By combining longitudinal survey data with seasonal variation in safety conditions in the Bering Sea, Lavetti (2018) was able to illustrate that commercial fishermen in this area are paid higher wages to take on more fatality risk, and that this premium is falling as risk levels rise. Other quasi-experimental work to estimate value of statistical life have included the use of mandated speed limits by Ashenfelter and Greenstone (2004) and military re-enlistment by Greenstone et al. (2015).

The key advantage of quasi-experiments, relative to my work, is that they can recover price estimates that control for unobserved amenities. This suggests an opportunity to check my findings for bias due to omitted amenities. However, the vast majority of the quasi-experimental compensating differentials research seems to have only studied the price of amenities relating to safety and fatality risk, typically in various sub-populations that are unrepresentative of the labor market. The findings of this literature seem to be that safety

---

6Villanueva (2007) also follows a within-worker approach in Germany to examine price heavy workload, job insecurity, poor hours regulation, and skill mismatch. The author concludes the results “broadly support the existence of” compensating differentials in Germany.
is positively valued by the labor market; my findings certainly agree on safety. There is not enough variation in fatality risk in my occupation-level data to more precisely line-up with quasi-experimental estimates of value of statistical life (although my point estimate is in the same direction).

A related but different literature has focused on estimating workers’ valuations of amenities, rather than market prices. To clarify the distinction, a group of workers may value an amenity a great deal, but this fact alone does not imply they sacrifice wages to get it; that depends on the price. As discussed in the next section, a price is an equilibrium between workers’ heterogeneous valuations of an amenity and firms’ heterogeneous costs of providing it. The typical strategy to study workers’ valuations of amenities has been to survey particular groups of workers about real (or fictitious) offers, and observe which offers they do (or would) accept. For instance, a survey of postdoctoral biologists’ job offers collected by Stern (1999) found that scientists value jobs that provide freedom to publish, despite little relationship between wage and freedom in the cross section due to the problem of ability bias. More recent research has turned to a framework of eliciting stated hypothetical choices rather than observing revealed choice in the labor market. In a survey of undergraduates at a prestigious university, Wiswall and Zafar (2018) reported that female students tended to state a higher willingness to pay for job flexibility and stability, whereas male students reported preferences for earnings growth. Based on early-career surveys of the students, they concluded that “a substantial part of the early gender gap in earnings we observe is a compensating differential in which women are willing to give up higher earnings to obtain other job attributes.” Controlling for stated preferences reduced the gender earnings gap by one-quarter.7 Maestas et al. (2018) surveyed respondents to the American Working Conditions Survey about their preferences. Men and women reported large differences in willingness to pay for amenities, particularly for physical activity and paid time off. The

7Wiswall and Zafar (2018) collected preference data on six categories besides earnings: (1) earnings growth, (2) number of hours, (3) part-time option (“flexibility”), (4) likelihood of being fired (“stability”), (5) performance compensation, and (6) proportion men.
authors concluded that “conditional on their wage, women have better amenities” and when individuals' valuations of the amenities are factored in to the gender pay gap, the magnitude is reduced by 30%. Mas and Pallais (2017) also applied a stated-choice framework regarding job flexibility to real applicants at a call center and in a nationally representative internet survey. Although they find that “women do have a higher valuation for worker-friendly work arrangements and a stronger distaste for employer discretion than do men,” they conclude that differences along this particular dimension “are not large enough to explain a substantial part of gender wage gap.” There is a similarity between my estimating procedure and the estimation procedure used by studies that observe worker choice data; this similarity is discussed in Online Appendix J.

Finally, although not directly related to the content of this paper, my approach was influenced by a brief (and perhaps forgotten) literature on the econometrics of detecting discrimination in observational data. A debate centered on how to interpret the fact that at any level of wage, black workers have on average less education than white workers – a fact which is also true of the modern labor market – and whether this implied discrimination against whites. The conclusion of the debate is best summarized by Goldberger (1984). In organizing the debate, the paper provided a clear lesson on how to discuss the quantity estimated by a regression in relation to competing structural models of the labor market (e.g., those with and without search frictions).

II Economics of Compensating Differentials

My conceptual framework follows at a high level the model of Rosen (1986) in the Handbook of Labor Economics. Readers interested in a more mathematical treatment of the closed-form integral solution to a binary version of this model are directed to Online Appendix J.

---

8Maestas et al. (2018) collected preference data on schedule flexibility, telecommuting opportunities, physical job demands, pace of work, autonomy, paid time off, teamwork, opportunities for job training, and meaningfulness.

9This fact and similar analogues by gender appear to have been discovered close to 1980 (e.g., Hashimoto and Kochin (1980); Conway and Roberts (1983); Kamalich and Polachek (1982)).
To better emphasize the problem of ability bias, I introduce the standard continuous Rosen model first in terms of workers that are homogenous in their ability (but differ in preferences). I then introduce the second level of exogenous heterogeneity, ability. The exclusion restriction will be stated precisely in the context of the model’s primitives, and then in more general language.

### II.A Relative Taste Heterogeneity

To start, workers are equally productive and differ only in how their compensation is split into various forms. We start with just wage $w$ and a single amenity $z$, but the model generalizes to further observed forms of compensation.\(^{10}\) Worker utility is given by $u(w_i, z_i | \theta_i)$ and firm profits by $\pi(w_j, z_j | \Omega_j)$. The equilibrium price will be determined by the distributions of $\theta_i$ and $\Omega_j$.

Market clearing is defined as a matching of workers to firms and a wage and amenity paid out in each job. Efficiency dictates workers with the highest valuation of the amenity are matched to firms with the lowest cost of producing it: perfect assortative matching on $\theta_i$ and $\Omega_j$. Each worker-firm pair determines a level of income and amenity that is Pareto optimal, which means that the $(w, z)$ coordinate chosen will be at a tangency of the worker and firm indifference sets, specified by $u$ and $\pi$. Famously, all workers’ and firms level-sets of $u$ and $\pi$ are “kissing” in equilibrium. For the matching of workers and firms to be stable, the $(w, z)$ coordinate chosen by a particular pair will also be one that is less desirable to neighboring workers and firms with similar valuations, relative to what those neighbors have chosen.\(^{11}\) With a sufficiently continuous distribution of $\theta$ and $\Omega$, the relative location choice of each pair is unique and constitutes the unique equilibrium within the market.\(^{12}\)

---

\(^{10}\)Although wage refers to an hourly concept, the logic of the model would also generalize to different earnings measures, such as yearly earnings among a group of full-time salaried workers. The model is intended to explain the worker’s labor supply decision between different types of jobs, rather than of how many hours to supply.

\(^{11}\)Figure A.6 illustrates that if the frontier deviates from this condition, it is not a stable equilibrium.

\(^{12}\)While the location choice of all pairs relative to each other is unique, there is a level parameter that
From the perspective of a single participant who takes other participants’ choices as given, the convex combinations of \( w \) and \( z \) that have been agreed upon function as a budget set. The local slope of this set – how much wage must be given up to obtain a small amount more amenity – is the local price of the amenity in wage units. A single worker’s problem can be rewritten as maximizing \( u(w_i, z_i | \theta_i) \) subject to the non-linear budget constraint \( \psi(w_i, z_i) = \eta_i \), in which \( \psi \) is a continuous function that takes in as arguments levels of worker \( i \)'s wage \( w_i \) and amenity \( z_i \), and maps those arguments to a level of total compensation, denoted \( \eta_i \). Firms, too, maximize profits subject to the feasible set of options given by this \( \psi \) frontier. This is a partial-equilibrium interpretation of the frontier, but if many workers or firms changed their choices, the shape of the frontier could change in the new equilibrium. The first two panels of Figure 3 illustrate the relationship between the slope of the frontier and the distributions of preferences and firm costs that have contributed to its shape.

II.B Absolute Heterogeneity

Finally, we extend the model to incorporate the fact that some jobs afford better compensation than others. To incorporate this feature, I model the amount of total compensation the worker can obtain as heterogeneous and determined exogenously, perhaps by factors determined prior to labor market entrance such as intelligence, motivation, and the unobserved facets of job search. Firms differ in their exogenous staffing needs and pay out more total compensation, composed of wages and amenities, in accordance with what type of applicant they seek to attract.\textsuperscript{13} The multiple Rosen frontiers are shown in Panel C of Figure 3.

The critical modification is that total compensation, which we have denoted by \( \eta \), is no

\textsuperscript{13}A more complex model of the labor market would allow firms to have some elasticity of substitution between skilled and unskilled labor. However, the goal of the present paper is not to understand the factors determining why higher-skilled workers earn more than lower-skilled workers, but rather the income-amenity choices faced by workers of a similar skill level. Thus, I take the skill premium as determined outside of the model.
longer the same for all workers. Viewing $\eta$ as pre-determined, the equilibrium can still be found by sub-dividing the entire labor market into sub-markets indexed by $\eta$. Markets indexed by the highest $\eta$ tend to contain the workers with the most productive abilities, matched with firms that most value such human capital. Within each $\eta$-level sub-market, wage and amenity are determined in accordance with the Rosen framework in which all jobs entail an $\eta$ amount of total compensation, but in different forms. The amenity price is now local to the sub-market. For a sub-market of compensation level $\eta$, the price is governed by the frontier $\psi(w, z|\eta)$. The $\psi$ function should be strictly monotone in both arguments, such that the frontiers of different compensation levels do not intersect each other.

**II.C Equilibrium Exclusion Restriction**

Decompose worker $i$’s compensation $\eta_i$ into an observable part $h_i$ and an unobservable part $\varepsilon_i$. The variable $\varepsilon$ can contain unmeasured skills, and also random search and matching frictions which result in observably identical workers obtaining better or worse jobs. The exclusion restriction is that among workers of a total compensation level $\eta$, $h$ must be independent of worker and firm tastes, $\theta$ and $\Omega$. Formally:

$$\left( h \perp \theta \right) | \eta \text{ and } \left( h \perp \Omega \right) | \eta$$

(1)

There is some redundancy here in that the condition on the relationship with $\theta$ is really the same as the one on the relationship with $\Omega$. Under the sorting framework, at job $k$, $\theta_k = \Omega_k$ so $\left( h \perp \theta \right) | \eta$ is interchangeable with $\left( h \perp \Omega \right) | \eta$. However, I state both conditions to emphasize that we can evaluate the exclusion restriction from either the firm’s perspective or the worker’s perspective – and an argument in violation of either is sufficient to rule out

---

14We can think of $\eta$ as differing not only due to productivity, but also to some degree due to luck in the Mortensen (2003) sense.

15To the best of my knowledge, all prior work on estimating market prices of amenities appears to have made the assumption that wage is stochastic but amenity is not. But to the extent that search frictions exist in the labor market, it seems implausible that they would affect wage but not amenities. My model treats total compensation, and thus both wage and amenity, as stochastic.
a candidate \( h \) variable.\(^1\)

Finally, the parameters \( \theta \) or \( \Omega \) in Condition 1 can be replaced with the joint distribution of \( w \) and \( z \) in favor of a more reduced-form statement of the exclusion restriction.

\[
\left( (w, z) \perp h \right) | \eta
\]

(2)

Identification relies on a variable that is informative about how high of a compensation frontier the worker lies on \((\eta)\), but adds no further information about the location of the worker along that frontier. Put another way, \( h \) would be irrelevant to \( w \) and \( z \) if we could observe \( \eta \).

**II.D No “bad controls”**

From a reduced-form perspective, it is straightforward to show that the exclusion restriction is a weakening of the typical OLS assumption for studying compensating differentials in observational data. Any variable appropriate to include in the standard approach is also admissible here as an \( h \) variable.

The typical amenity pricing approach seeks to estimate the joint distribution of \( w \) and \( z \) conditional on \( \eta \), \((w, z) | \eta \). For instance, this may be linearly approximated by a regression of \( w \) on \( z \) with controls. A control is a bad control if it can cause bias, for instance if adding it to an already-identified model would cause identification of the structural parameters to be lost.\(^1\) Defining \((w, z) | \eta \) as the identified distribution of interest, then, a **good control** would have the property that including it in the regression does not change the relationship of interest:

\(^{16}\)From the worker side, we say that within the \( \eta \)-level workers, those that have high \( h \) can’t have different valuations of the amenity \((\theta)\) from those that have low \( h \). From the firm side, we say that among firms that pay \( \eta \)-level compensation, those that employ high-\( h \) workers can’t have different valuations of the amenity \((\Omega)\) from those that employ low-\( h \) workers.

\(^{17}\)Angrist and Pischke (2009) write: “Some variables are bad controls and should not be included in a regression model even when their inclusion might be expected to change the short regression coefficients. Bad controls are variables that are themselves outcome variables in the notional experiment at hand. That is, bad controls might just as well be dependent variables too. Good controls are variables that we can think of as having been fixed at the time the regressor of interest was determined.”
(w, z)|η, h = (w, z)|η

Rewrite the “good control” condition in terms of PDF’s:

\[ f_{wz|η,h} = f_{wz|η} \]

Re-writing the left-hand side:

\[ \frac{f_{wzh|η}}{f_{h|η}} = f_{wz|η} \]

Re-arranging,

\[ f_{wzh|η} = f_{wz|η}f_{h|η} \]

By definition of independence, we arrive back at Condition 2:

\[ ((w, z) \perp h)|η \quad (3) \]

In the typical analysis of wage on amenity with ability controls, Condition 2 is necessary to include h as a control variable. The only difference is that with the standard regression approach, we would also require the rest of the full set of ability controls. Therefore, the proposed estimator entails a weakening of the assumptions required by the OLS approach. Any single variable that allowed to be included as a control under OLS should also be a valid h variable.

II.E AFQT as an h variable

I argue that score on the Armed Forces Qualification Test (AFQT) is correlated with the level of the wage-amenity frontier, but is independent of workers’ locations along their particular frontiers. The AFQT score is the primary summary score from the Armed Services Vocational Aptitude Battery, which was administered to respondents of the NLSY79 in 1980.
As the purpose of the test is to measure basic skills that are applicable to a wide range of vocations, the score by construction should be independent of how employers located at various positions on a wage-amenity frontier view a worker. Although AFQT is in no sense “innate,” any effort by a worker to increase his or her AFQT score should result only in a movement to a higher compensation frontier, rather than a movement along a given frontier. Perhaps the most pressing concern is that the decision to invest in AFQT-relevant skills could for some reason be correlated with a preference for wage or for certain types of amenities. For instance, such an argument might be that workers who have a stronger affinity for certain amenities invest more into training for general-purpose skills. Still, such an argument is hard to rationalize with utility-maximizing individuals.

In a more reduced-form sense, we can plausibly think of AFQT as a noisy proxy for productivity, in which the noise is uncorrelated with preferences. Heterogeneity in income versus amenities among \( \eta \)-level workers and firms would be determined not by AFQT but rather by workers’ and firms’ heterogeneous valuations of the amenity. The key is that once we’ve conditioned on \( \eta \), AFQT should be irrelevant to both the firms’ and workers’ problems.

Finally, several foundational papers in labor economics have used AFQT score from the NLSY as a noisy measure of worker productivity. Employer learning and statistical discrimination have been explored by Farber and Gibbons (1996) and Altonji and Pierret (2001), who used AFQT as a measure of worker skills that are unobserved to the employer. In a different vein, Neal and Johnson (1996) used AFQT as a “measure of basic skills that helps predict actual job performance,” noting in particular its usefulness of measuring skills present prior to labor market entrance. The importance of AFQT in explaining the wage gap has often been cited as evidence of the racial “skill gap.” A central assumption in prior literature has been that AFQT measures general-purpose skills, not preferences. For instance, although many interpretations of the findings of Neal and Johnson (1996) have emerged that do not involve a racial “skill gap,” no arguments have been made that controlling for AFQT is revealing a “preference gap” for job amenities.
III Non-Parametric Identification

In this section, I prove that the Rosen frontier can be non-parametrically identified under the equilibrium exclusion restriction of Condition 2, in which ability proxy $h$ is assumed irrelevant to wage-amenity sorting conditional on true ability $\eta$. In the next section, I apply the non-parametric result to a data-generating process assumed to be linear, in which I discuss issues of potential bias.

III.A Object of Interest

Denote wage by $w$, amenity by $z$, and total compensation by $\eta$. We want to learn what wage and amenity values are associated with each level of total compensation. For a given level of compensation $\eta_j$, I will denote combinations of wage and amenity associated with that $\eta_j$ by the Rosen frontier notation: $\psi(w, z|\eta_j)$.

III.B Assumptions

I make the usual standard assumptions: an IID sample of $w$, $z$, and $h$; and, all variables are assumed continuous. The substantive assumptions are:

1. Conditional independence of $h$: $((w, z) \perp h) | \eta$

2. Monotonic relevance of $h$: $E[h|\eta]$ is everywhere strictly monotone in $\eta$

3. $\eta$ is degenerate given values of $w$ and $z$; i.e., there exists a non-stochastic function $g$ such that $\eta \equiv g(w, z)$

Assumption 1 is what we have referred to as the “identifying assumption.” However, Assumption 3 is also a critical component of the basic structure of the research question that enables identification. The global strict monotonicity of Assumption 2 can be relaxed; Online Appendix F discusses identification under a more local form of monotonicity.
III.C Proposition

Under these assumptions, $\psi(w, z|\eta)$ is identified for all levels of $\eta$.\textsuperscript{18}

III.D Estimator sketch

Define the random variable $\hat{h}$ as an approximation to $E[h|w, z]$. For simplicity, $\hat{h}$ may simply be a conditional mean of $h$ in given bins of $w$ and $z$.\textsuperscript{19} Denote with subscripts $\hat{h}_j$ as quantile $j$ of the constructed variable $\hat{h}$ and $\eta_k$ as quantile $k$ of the unobserved variable $\eta$. For all observed $w$ and $z$, the distribution $\psi(w, z|\hat{h}_j)$ approximates the distribution $\psi(w, z|\eta_k)$ for some unknown quantiles of the variables, denoted by $j$ and $k$.

III.E Proof of Identification

We will see that with a large amount of data, conditioning the distribution on $\hat{h}$ becomes the same as conditioning on $\eta$.

The first step is to note that the constructed variable $\hat{h} = \hat{E}[h|w, z]$ can be equated, in probability limit, with $E[h|w, z, \eta]$. This is because $\eta$ is assumed degenerate given $w$ and $z$, so adding it to the conditioning is redundant (Assumption 3).

In the second step, we rely on conditional independence of Assumption 1 to equate $E[h|w, z, \eta]$ from above with $E[h|\eta]$. This is intuitive, since the wage-amenity split is by assumption not related to $h$.\textsuperscript{20}

At this point, we have seen that under the conditional independence assumption, $\hat{h}$ is as good as $E[h|\eta]$. Our last assumption is that $E[h|\eta]$ is strictly monotone in $\eta$ (Assumption 2). If this is the case, then each quantile of the control function corresponds precisely to a

---

\textsuperscript{18}To clarify, the units of $\eta$ are to be viewed as a nuisance parameter. We seek to learn the values of $w$ and $z$ associated with the same $\eta$, but will remain agnostic about the cardinal units of that $\eta$.

\textsuperscript{19}Alternatively, optimal kernel estimators might be considered for efficiency, which is not the present focus.

\textsuperscript{20}A full derivation of this step is provided in Online Appendix C.
different quantile of $\eta$. Conditioning on an arbitrarily small range of the control function makes the correspondence to the level of $\eta$ arbitrarily close.

Quantiles of $\hat{h}$ can be mapped to quantiles of $\eta$ as follows. Consider a worker $i$ with observed $w_i$ and $z_i$, and an unknown level of compensation, $\eta_i$. The researcher is interested in learning what other observed values of $w$ and $z$ are associated with this particular amount of total compensation; this is the distribution we denote as $\psi(w, z|\eta_i)$. To find the quantile $k$ of $\hat{h}$ such that $\psi(w, z|\hat{h}_k)$ approximates $\psi(w, z|\eta_i)$, all that one needs to do is find the distribution of $\psi(w, z|\hat{h}_k)$ that contains the point $(w_i, z_i)$. Since $\hat{h}$ will be defined on any point for which $w$ and $z$ are observed, we know that for all observed $w$ and $z$, $\psi(w, z|\hat{h})$ identifies $\psi(w, z|\eta)$.

IV Estimation in the Linear Model

For the purpose of discussing potential biases, it is now useful to narrow our vision to a model in which price can be summarized by a single parameter. The non-parametric estimator entails nothing more than plotting conditional means of the $h$ variable by bins of wage and amenities. The linear proposed estimator is best understood as a special case of that non-parametric estimator, in which conditional expectations are allowed to be replaced with their linear counterparts and can thus be estimated by linear regression. Relative to the non-parametric framework, we will introduce two additional assumptions of linearity: linearity of the Rosen frontier and a linear relevance of $h$ to $\eta$.

In the non-parametric framework, we sought to identify the distribution of $w$ and $z$ given unobserved $\eta$. Here, we will take the object of interest to be the coefficient on $z$ from a hypothetical regression of $w$ on $z$ controlling for $\eta$. That is, we will assume $E[w|z, \eta]$ is linear in $z$, with homogeneous parameter $\beta$, such that each Rosen frontier can be written of the linear form:

$$w - \beta z = \eta$$

(4)
I have omitted a scale coefficient on $\eta$ as it makes no difference since $\eta$ is unobserved and we have normalized the price of wage to 1. The content of Equation 4 is simply that we need to assume that conditional on $\eta$, the values of $w$ and $z$ form a line if we are to summarize this relationship with a single parameter. The second assumption of linearity is more nuanced.

In the non-parametric framework, we assumed $E[h|\eta]$ was strictly monotone in $\eta$. Now, we will strengthen monotonicity to linearity. In other words, $E[h|\eta]$ can be written $\gamma \eta$ for some unknown $\gamma$.

With these additional assumptions, we can modify the proof of identification in Section III to replace all conditional expectations with linear regressions. The Table below provides a side-by-side comparison of the non-parametric and linear estimators. We already saw graphical intuition for the linear approach using predicted values in Figure 2; the non-parametric approach simply uses predicted values that are not required to be linear functions.

<table>
<thead>
<tr>
<th>Non-Parametric</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $\hat{h} = E[h</td>
<td>w, z]$</td>
</tr>
<tr>
<td>2. Plot $w$ on $z$ at various quantiles of $\hat{h}$</td>
<td>2. Regress $w$ on $z$ controlling for $\hat{h}$</td>
</tr>
</tbody>
</table>

**IV.A Nature of bias when $(w, z) \not\perp h|\eta$**

In the linearized model, the bias in $\hat{\beta}$ takes the sign of $cov^*(w, h|\eta)$. Online Appendix I contains a formal proof, but Figure 4 conveys the key intuition in the linear model. No bias occurs when the $h$ variable is evenly distributed along each Rosen frontier. To the extent that high-$h$ workers bunch on one end of the frontier, they will pull the estimate with them.

For instance, perhaps we suspect that a candidate $h$ variable is positively correlated with workers’ preferences for amenities (relative to wage). The excess bunching of high-$h$ workers at high amenity levels would cause us to estimate that this is a more costly amenity than it actually is. This argument matches what is depicted in the figure, as would a story in which $h$ was a skill that was particularly valued in high-amenity jobs (relative to high-wage jobs).
IV.B Finite-Sample Inference

In general, the linear proposed estimator is vulnerable to a finite-sample problem similar to that of weak instrumental variables whenever the relationship between $h$ and $w$ is difficult to detect empirically. The problem is the result of reporting the price of the amenity per unit of wage, which puts the wage coefficient into a denominator.

Recall the ratio of coefficients estimator for the price is $-\hat{\pi}/\hat{\delta}$ from a regression of $h$ on $w$ and $z$:

$$\hat{E}^*[h|w, z] = \hat{\delta}w + \hat{\pi}z$$

Whenever $\hat{\delta}$ may dip close to 0, the ratio of coefficients explodes, causing price estimates to skew large. In light of the potential for bias in finite samples, I report Anderson-Rubin confidence intervals when appropriate, following the suggestion and approach of Andrews et al. (2019).

IV.C Nature of Bias from Measurement Error

This estimator is affected by measurement error in a very different way from the typical OLS approach. Like the finite sample bias, the bias from measurement error is best understood from the ratio-of-coefficients approach, which runs the regression $\hat{E}^*[h|w, z] = \hat{\delta}w + \hat{\pi}z$ and takes the ratio $-\hat{\pi}/\hat{\delta}$.

Clearly, classical measurement error in $h$ causes no particular bias on $\hat{\pi}$ or $\hat{\delta}$. This is a key advantage of my estimator; noise in the dependent variable of a regression only causes imprecision but not bias. (Although, as previously discussed, this can itself cause finite-sample bias.)

Measurement error in a regressor, on the other hand, causes the coefficient on that regressor to be attenuated. If $w$ is measured with noise, then $\hat{\delta}$ is too small, so the estimated price of the amenity is too large in magnitude. If $z$ is measured with noise, then $\hat{\pi}$ is
too small, and the price of \( z \) is too small. Because all components of compensations are regresors, measurement error in any piece of compensation will simply make it appear less important or valuable. Of course, because measurement error in wage and amenity have the potential to balance each other out, there is also the case in which both variables are measured with a balanced amount of noise such that the resulting price estimate is unbiased. This section contains an intuitive treatment of measurement error in the proposed estimator, and a fuller treatment is given in the reduced-form statistical model of Online Appendix G.

V Price Estimates

V.A Data

The primary data source for the analysis is NLSY respondents employed in 2012 linked at the occupation-level with O*NET.

The NLSY is a longitudinal survey dataset constructed by the Bureau of Labor Statistics. The first wave, administered in 1979, sampled 12,686 men and women between the ages of 14 and 22. Subsequent survey waves were administered annually until 1994, and once every two years since. The NLSY contains three sub-samples. The primary sample (6,111 people) is “designed to represent the non-institutionalized civilian segment of young people living in the United States in 1979 and born January 1, 1957, through December 31, 1964.” Two additional sub-samples are contained; the first (5,295) is a sub-sample of racial minorities and economically disadvantaged people, and the second (1,280) is designed to represent those serving in the military as of September 30, 1978. To maximize power, I include all three samples.\(^{21}\)

I focus on data in 2012, as the occupation data has changed across years. By 2012, only 7,300 of the initial NLSY sample was retained. Of the 5,386 not interviewed, the biggest

\(^{21}\)In the final analysis sample, only 6 workers list Armed Forces (Census code 984) as their occupation in 2012.
driver has been administrative decisions to reduce the sizes of the non-representative sub-
samples. 84% of the military sub-sample was dropped after the 1984 survey, and in 1990
31% of the disadvantaged sub-sample was dropped (poor whites). After the 2,722 individuals
dropped for one of these reasons, the most common reasons for attrition were refusal (903),
can’t locate (466), deceased (689), other (481), and difficult cases (125).

My final analysis sample consists of 4,868 NLSY respondents who reported being em-
ployed in 2012, when they were aged 48-55. This number excludes 152 individuals
who I drop due to missing job amenity fields. To be employed, the respondent must have
worked at least ten hours per week and at least 9 weeks since the date of the last interview,
or is self-employed. Although I conduct the main analysis on this broad sample, I also test
robustness to selecting on the full-time employed population. In the analysis of demographic
pay gaps, I will divide the NLSY by gender and by race. Gender and race both correspond
to the respondent’s data in 1979, at the start of the survey, as determined by an interviewer.
Sex and gender were not coded separately at this time, so I use them interchangeably. The
racial/ethnic categories are as follows: “hispanic,” “black,” or “non-black/non-hispanic.” For
the analysis by race, I will compare the black group to the non-black/non-hispanic group,
which I will at times refer to by the approximate term “white” for brevity.

For additional results on income inequality by parent background, a PSID sample was
constructed to mirror the sample definitions of the NLSY. The PSID sample contains 929
workers in 2013 linked to their parents. For each child, I define a measure of parental income
to be the father’s income averaged over all years for which the father was between 30 and
50. Because of the requirement that I observe income for both workers and their fathers,
this sample of workers is substantially younger than the NLSY one, between ages 30 and 40.

I make two key deviations from the logic of the simple conceptual framework in which
workers choose jobs indexed by a wage and amenity level. First, while the basic framework
considers the trade-offs workers face between wage and amenities, in practice most workers
in the sample do not report being paid an hourly wage (Figure A.2). Although it is possible
to impute an hourly wage based on workers’ estimates of hours worked, to more accurately mirror the level at which most workers are making decisions, I use yearly labor income rather than hourly wages. Labor income in the NLSY is defined as total pre-tax income from wages and salary in the past calendar year, but if a worker holds more than one job then it is aggregated over all jobs. The amenities assigned, however, will correspond to the worker’s “primary” job. Assuming that any additional jobs held by the worker have similar amenities to the primary job, then this data definition can be viewed as the relevant yearly income and typical amenity levels consumed by the worker.

Second, whereas the canonical framework allows workers to choose particular jobs, in my data many amenities are only observed at the occupation level via the O*NET occupation survey. O*NET, which stands for Occupational Information Network, is a database of occupational characteristics maintained by the US Department of Labor. The data are derived primarily by occupational surveys and supplemented in some cases by occupational experts. I draw primarily on the O*NET Work Context table, which includes attributes that workers might plausibly have varying preferences over. Examples include how regular the schedule is, how often minor cuts and burns arise on the job, whether the work is indoors or outdoors, or how much freedom there is to make decisions. As the scales of most of these measures are difficult to compare, I translate income and all amenity measures to standard deviation units across workers in the sample. Table A.1 describes the process by which several amenity measures were constructed from O*NET.

The relative scarcity of job-level rather than occupational-level amenities has been discussed in prior work (e.g., Goldin (2014); Denning et al. (2019)). Due to this limitation, I aggregate the entire analysis to the occupation level, rather than job level, and weight by the number of workers in each occupation.$^{22}$ I also show robustness of the final results to estimation at the individual level with no aggregation. The occupation-level nature of the

$^{22}$Table A.2 contrasts the findings of the typical OLS approach on data at the individual versus occupation level; even with OLS, aggregation of the entire analysis to the occupation-level does seem to remove some, but not all, of the wrong-signed bias.
analysis entails a loss of power. After mapping the O*NET SOC occupation codes to the same Census format as is used in NLSY in 2012, I am left with only 397 occupations that are represented in the NLSY in 2012 (451 across all years). This means that 397 is the effective sample size, or number of clusters, for most of the analysis.

V.B Functional Form & Relation to Prior Approaches

As a first look at empirics, Figure 5 compares the OLS approach with the proposed approach, overlaying linear estimates on-top of their non-parametric counterparts. Similar to Figure 1, the OLS approach in Panel A entails regressing income on safety, controlling for the ability proxy of AFQT by way of terciles. Within terciles of AFQT, the income-safety relationship is upward sloping. In contrast, the proposed approach in Panel B uses precisely the same data but plots mean AFQT by bins of income and safety. Despite identical underlying data, the finding is the opposite: within each tercile of mean AFQT, which should be thought of as the empirical stand-in for the offer curve, the income-safety relationship is now downward sloping, implying that a trade-off clearly exists in the labor market between income and safety.

Just how steep is the trade off, and is it well characterized by a single parameter? Our answers to these questions are limited by the precision of the data. In Panel B, for instance, although a downward trend is obvious, we can also visually ascertain that some bins fall off the trend. This is likely due to sampling variation. In Panel C, a formal statistical test is applied to do inference on the graph. Because every bin represents a sample mean of AFQT, we can apply a standard two-sample \( t \)-test for difference in means to help determine which regions of the graph correspond to different offer curves. I have arbitrarily picked the center point as the reference point, by which all other means are tested. The test constructs a graphical confidence interval for the price, allowing the researcher to evaluate potential non-linearities around the center point. Although the graphical test can reject

\(^{23}\)This test was suggested by Isaiah Andrews.
many upward-sloping patterns that the wage-amenity frontier passing through the center point might have taken on, there is still a great deal of uncertainty about the exact shape of the frontier. Due to the low level of precision that can be obtained from this relatively small sample, for most of the remainder of the paper, I simply estimate linear average prices rather than attempting to gain non-parametric precision.\textsuperscript{24}

In a linear model, Figure 6 compares several of the typical strategies with the proposed estimator. In this case, I define the Mincerian estimator to be the regression of wage on the amenity with controls for education, AFQT, age, and year. I study the price of amenities one-by-one in relation to income. For instance, examining the amenity of safety, the raw, Mincerian, and individual fixed effects estimators all produce price estimates indicating that the average price of safety is negative (it is a disamenity), similar to what was reported by Brown (1980). On the other hand, the proposed estimator for the price of safety when the \textit{h} variable is set to be AFQT clearly lies on the other side of 0 and yields a result contrary to all other established methods in observational data. The point estimate indicates that on average, workers must give up \textasciitilde 0.31 standard deviations of income to obtain a typical job with a standard deviation of safety.\textsuperscript{25} Still, similar descriptive compensating differentials can be found for jobs that have more regular schedules (\textasciitilde 0.29 SD less income), those with more freedom to make decisions (\textasciitilde 0.07 SD less income), and jobs that tend to be in the non-profit sector (\textasciitilde 0.42 SD less income).

To emphasize, the proposed estimator in this case is using data that is a strict subset of the data used by what I refer to here as the Mincerian approach. The only difference is that

\textsuperscript{24}Related to the problem of imprecision, the potential for finite-sample bias when summarizing the data linearly can also be detected in Figure 5. Because our data has sampling variation, some of the points in Panel B appear to lie off trend. When viewing this graph, the noise in the measure of the offer curve level is apparent. Now, consider a regression of income on safety controlling for the estimated level of the offer curve, which contains sampling variation. For any level of the noisily estimated offer curve, there is a weaker relationship between wage and amenity due to the sampling variation, so the linear estimate of the slope of interest looks more like a vertical line that it should. This is the graphical interpretation of the finite-sample bias that was shown algebraically in Section IV.B.

\textsuperscript{25}This is not precisely the “price” of safety, as it does not control for any other amenities. The price of safety would describe the trade-off holding all other attributes fixed. Rather, this describes the trade-off between income and a typical safer job.
it treats AFQT as a locational shifter of ability, rather than a precise conditioning variable. Next, I turn to examine how results from the proposed estimator compare when proxies of workers’ productive abilities other than AFQT are used.

\section*{V.C Sensitivity to $h$ variable normalization}

A strength of the NLSY is its preponderance of measures of skills. There is no way to prove empirically that any potential $h$ variable fits the identification assumption. Whereas each potential ability proxy may always be subject to its own nuanced critiques as to why it may not fit the identification assumption, an opportunity exists to compare results from many different independent measures of skill. If very different proxies for the ability channel yield similar results, this can lend some credibility.

\subsection*{V.C.1 Data on $h$ variables}

In section II.E, I argued that a cognitive test such as the AFQT, though by no means “innate,” may be an ability proxy not contaminated by wage-amenity sorting. This would be plausible so long as workers do not systematically manipulate their AFQT in order to gain particularly more of either wage or amenity, rather than total compensation. The AFQT was administered to NLSY respondents in the 1981 survey year.

A variable that may be more pre-determined is the worker’s height, which is surveyed in the NLSY in several years including 2012. The positive relationship between height and earnings, which holds even in developed nations, has been well documented though its source is a topic for debate. The cause of this association does not seem to be that height is a skill that is useful in certain physically demanding jobs – i.e., it does not seem to be the case that warehouse firms compete for taller workers who can reach higher boxes. Evidence against this explanation is presented by Persico et al. (2004), who show that the height premium is entirely distinct from a worker’s contemporaneous height once height during adolescence is controlled. Thus, the returns to height must be running through an
unobserved variable correlated with adolescent height. Persico et al. (2004) propose that the skills and confidence stemming from increased social and athletic experiences during high school may partly be at play. However, more recent work by Case and Paxson (2008) makes a convincing case that taller adults earn more because they have higher cognitive skills. Since physical development is known to be influenced partly by in utero and early childhood environmental factors such as disease and nutrition, such environmental factors would also act in the same direction on cognitive development. Whichever combination of social, cognitive, or even taste-based discriminatory explanations one wishes to believe, the evidence indicates that height is proxying for an ability channel that is likely to be irrelevant to where workers locate along the wage-amenity frontier. To better isolate the component of height accounted for by environmental rather than genetic factors, I follow the recent literature by residualizing height on sex.\(^{26}\)

Other plausibly exogenous variables include education and measures of what might be called social skills. Education, measured in the NLSY as highest grade completed by 2012, is a variable often used as one of several controls for ability in the labor economics literature, and has often been included as a control when researching compensating differentials. As described in Section II.D, any valid ability control from the typical OLS or “Mincerian” specification is a valid \( h \) variable here. I also use three scores contained in the NLSY from tests that have been referred to by previous literature as “social” or “non-cognitive” skills that are predictive of compensation (e.g., Heckman et al. (2006); Deming (2017)). These variables are self-esteem, mastery, and locus of control. The self-esteem measure, pioneered by Rosenberg (1965), is based on a 10-item survey that gauges agreement with statements about attitudes toward oneself (e.g., “I have a number of good qualities.”). Mastery is a similarly constructed variable based on a survey developed by Pearlin et al. (1981) to study factors that mediate stressful life events such as job losses; it is a self-reported measure of a worker’s ability to solve problems that affect him or her. Finally, locus of control is a variable

\(^{26}\)Results are similar when height is also residualized on race.
constructed from an abbreviated version of the 60-item scale originally developed by Rotter (1966). It is higher when individuals agree with statements that their life is determined primarily by fate or factors outside of their control (external control) and lower when people respond that they direct their own lives (internal control). I use the measure of self-esteem from 1980, mastery from 1992, and locus of control from 1979.

Table 1 displays pairwise correlations among income and the proposed $h$ variables. All correlations are significant when run as regressions between income and the $h$ variable. Years of education and AFQT have the strongest correlations with income. Height has the weakest, though still positive. Higher self-esteem and mastery are predictive of higher income, and workers with a more external locus of control have lower income.

To reinforce the point that not any measure of skill will be an unbiased proxy for the Rosen frontier, I also re-run the analysis using several skills measured in the NLSY that are intended to measure occupation-specific skills rather than those useful in a broad variety of occupations. For these occupation-specific skills, I turn to the Armed Services Vocational Aptitude Battery, of which the AFQT is the primary composite score an I focus here on the individual sub-scores. The ASVAB is a battery of 10 tests that measure knowledge in skills ranging from word knowledge to automobile repair skills (e.g., “what type of wrench is pictured?”). The logic against using a variable like automobile repair skills as the direction by which to normalize the ability channel is that it is likely to contain information about a worker’s job preferences by way of specialization in job-specific skills. It thus violates the identification assumption because it is correlated with how workers split their compensation into wage and amenities.

V.C.2 Results

The main result of this section is Figure 7, which compares the results for the price of safety using each of the six proposed $h$ variables. These variables span widely different spaces of physical, cognitive, and social differences that workers bring to the labor market. Yet, they
all lead to very similar magnitudes of price estimates. All six confidence intervals overlap with each other, and none overlap with the OLS coefficient of income on safety.

In contrast, Figure A.1 additionally includes the 10 component scores of the ASVAB, and thus plots 16 estimates using the proposed estimator. These tests are developed to measure relative skills, rather than the generally ability channel that might proxy for total compensation. Evidently, not all of the ASVAB components even yield the same sign, with the ASVAB Auto & Shop score yielding an estimate even more extreme than the OLS estimate. Such a finding is to be expected given that this skill measures aptitude for a mechanic job, which is one of the most dangerous occupations by the O*NET measure. On the other end of the estimates is “coding speed,” a test which provides a table of words and numbers and asks which number corresponds to a certain word. Dropped from the ASVAB in 2002, the score was previously used to classify military recruits into clerical jobs (Held and Carretta (2013)).

Among several skills that are plausibly relevant to a general set of occupations, we obtain similar price estimates for safety. On the other hand, if the ability channel is normalized to a skill that is designed to measure aptitude in a narrowly defined set of occupations, we will obtain a different estimate. This is a normalization that the researcher must think critically about. Importantly, it is a strict sub-set of the questions a researcher must confront in finding ability controls for the typical approach to estimating compensating differentials in observational data.

To examine the extent to which different $h$ variables produce similar results for a wider range of amenities, Figure 8 zooms in on two leading candidate $h$ variables: height and AFQT. Estimates for several amenities, priced separately, are presented to investigate whether the two candidate $h$ variables, which share a correlation of only .1486 at the individual level, produce similar estimates. Although height is much less precise, there is no evidence to indicate that these variables are leading to different estimates. This is further evidence that several observed $h$ variables are reasonable proxies for the factors that determine a worker’s
level of total compensation.

V.D  Multivariate prices

For simplicity, all of the empirics so far have priced amenities against income in a bivariate nature. However, the estimator readily extends to a multivariate pricing equation as follows. In the approach using predicted values, all that is required is to include the additional amenities in both the first stage (generating predicted values) and in the second stage (calculating prices). Mechanically, this is equivalent to the ratio-of-coefficients approach in which $h$ is regressed on all forms of income and amenity compensation. The negated ratio of each amenity coefficient over the income coefficient is equivalent to the corresponding price estimate from the predicted values approach.

Figure 9 compares linear price estimates when amenities are priced in isolation of each other as opposed to in the multivariate framework. In general, the magnitudes of each amenity shrink when they are considered in light of each other rather than one-by-one. That the multivariate coefficients are less than the bivariate ones can be seen in the scatterplot of Figure 9 as a slope of the points less than the red 45° line. This finding is consistent with the thinking that in general, good job attributes come bundled together. That is also what would be expected if income and amenities are viewed as simply two forms of compensation firms use to attract more productive workers.

VI  Demographic Gaps

With estimates of the price of various amenities in the US labor market, I next turn to the question of quantifying the extent to which these amenities, at current prices, contribute to widely studied demographic pay gaps. This exercise is a decomposition of income inequality into an absolute component due to differences in offer curves and a relative component due to workers’ locations across the income-amenity frontier. Before accounting for a wide array
of amenities, it is instructive to first make the question concrete in the context of just one amenity.

VI.A Graphical Intuition: Regular Schedules

To begin with some graphical intuition, I focus on a variable describing how regular the job’s schedule is. Although safety seemed to be the quintessential job amenity in the days of Adam Smith, this variable seems to more closely mirror modern scholarship surrounding amenities in the gender gap.

The NLSY asks workers whether they “usually work a regular daytime schedule or some other schedule” at their current job. Of those who do not work a regular schedule, the most common alternatives are either irregular schedules arranged by employers or regular schedules containing evening or night shifts. About 80% of workers report working a regular, daytime schedule. Significant differences exist by race and gender; women and whites each work more regular schedules than their comparison groups. Admittedly, this variable may not be an ideal measure of the more nuanced job attributes relating to scheduling discussed in the literature on the gender gap, such as whether the nature of the work entails a convex return to hours as in Goldin (2014). However, if women tend to have higher valuation for a regular daytime schedule – for instance, to be home for children in the evenings – then we might still be interested to see how costly this noisy and narrowly defined amenity turns out to be in the data.

For the purposes of graphical exposition, let us assume that the true model is linear, and we are only interested in pricing this one amenity in relation to income. (That is, we will obtain the price of this amenity given how other unobserved amenities covary.) This price of the amenity was already reported in Figure 8: 1 SD of schedule regularity when AFQT is used as the $h$ variable costs .29 SD of income.\textsuperscript{27} This price estimate corresponds to the slope

\textsuperscript{27}In the previous section, I priced all amenities at the occupation-level because that was the level at which many were observed. This amenity is observed at the individual-level. Qualitatively, the result of this analysis are similar whether the amenity is priced at the occupation- or individual-level, although the price is smaller.
of the Rosen wage-amenity frontiers. That is, if we observe two workers who have different bundles of wage and amenity, the slope from the price estimate is enough to discern whether the workers lie on the same or different frontiers. We can thus decompose income inequality between men and women into a component that is due to sitting on different Rosen frontiers (differences in total compensation) versus differences in locations along the wage-amenity frontiers (differences in valuations).

Panel A of Figure 10 shows the results of this analysis by binning the labor market into deciles of total compensation. There is a clear trend: at any level of compensation, men on average have higher earnings and lower regular schedules than women. The difference is substantial; averaging over all the bins men earn about .08 SD more than women conditional on total compensation. To put the number in perspective the overall gender income gap in this sample is .54 SD. This simple result implies that .08/.54 = 14% of the gender income gap can be attributed to this trend of costly substitution by women to jobs that, roughly measured, have regular schedules (and other features that go with it).

Panel B shows the identical analysis, but grouped by race rather than by gender. The pattern is vastly different. Regardless of the well-known income differences between racial groups, their mean locations are essentially indistinguishable once total compensation is conditioned on. On average, a black and a white worker who face the same job prospects do not make meaningfully different choices in the same way that men and women do. At least, blacks and whites do not meaningfully sort along the dimension measured by how regular the schedule of the job is. The analysis of Figure 8 is intended to serve mainly as a graphical introduction for the more comprehensive analysis of the next section.

and less significant when this amenity is priced at the individual-level. This difference is consistent with the binary amenity measure containing some measurement error at the individual-level, causing attenuation, which is mitigated when the analysis is aggregated to the occupation level.
VI.B Pay Gap Accounting

To get a more wholistic picture of the contribution of job amenities to the demographic pay gaps, I next consider a multivariate pricing equation of, in addition to regular schedule, the following job characteristics: nonprofit, sitting, unstructured, repetition, decision freedom, safety, and time pressure. These are the same characteristics priced separately and together in Figure 9. Two of the job characteristics (nonprofit and regular schedule) are not available in the PSID, so are not included for the analysis of inequality by parent background. For the analysis in the NLSY, I use AFQT as the \( h \) variable; as this variable is not available in the PSID, I use years of education to price amenities among that group. I re-price amenities in the PSID because after restricting the PSID to respondents for whom I observe parental earnings, I end up with a much younger sample than that of the NLSY; it may not be wise to assume the samples are representative of each other.

Effectively, the many dimensions of job heterogeneity allow me to construct multi-dimensional Rosen frontiers. If the result of the previous section was that men tend to sit on the income side of the income-regular schedule frontier, the question we now ask is which side of a high-dimensional income-amenity contour men sit on.

As usual, I define income gaps by sex (male minus female) and race (non-black/non-hispanic minus black). To operationalize inequality by parent income in binary terms, I define an income gap between those in the PSID based on whether one’s parents earned below or above the in-sample median. Figure 11 displays the results of this analysis, as well as the overall demographic gaps in the relevant sample. We first start in the main analysis sample, which has no full-time hours restriction, and then examine robustness of the results to alternative specifications. As reported in Table 2, in the NLSY analysis sample men earn on average almost $30,000 more than women, and the race gap is $22,000. The PSID reports a gap by parent income of $22,000.

Consistent with the graphical intuition of Figure 10, the race gap is not driven by differences in income-amenity substitution. Once offer curves are included as controls, there is
essentially no income divide between blacks and whites. Similar can be said of the parent income gap. Absent a method to calculate proper confidence intervals for the results, a great deal of caution is warranted in reading too much into the point estimates. Still, if one were to interpret the directions of the point estimates as signal rather than noise, the story would be that blacks, in addition to earning lower income, also earn slightly worse amenities at any level of compensation. That is, blacks substitute toward income, and thus the racial pay gap exists in spite of blacks’ higher willingness to take on worse work. The point estimate for the gap by parent income tells the opposite story, in which kids from poor families are more likely to substitute toward better amenities. Although quantitatively small and likely to be imprecise, the point estimate is consistent with the logic of Jencks and Tach (2005), who noted that if preferences for work characteristics are passed across generations, some part of the intergenerational income gradient is likely due to amenity substitution. Across alternative specifications to be discussed in Section VI.D, the directions in which amenities point for both the race and parent background gap seems to be relatively stable.

In contrast to the nature of inequality by race and parent background, which seem better attributed general-compensation factors such as skills or discrimination than to income-amenity substitution, the bulk of the gender gap still exists within observable Rosen frontiers. These findings are directionally similar to the analysis using only the regular schedule amenity, which found 14% of the gender gap due to the amenity, but are quantitatively much larger. The analysis with more amenities indicates that approximately two-thirds of the gender income gap in this dataset is due to variation along the wage-amenity frontier. Whether this result is large is a matter of perspective. On the one hand, if the remaining one-third of the gender gap were due to discrimination in the labor market, this would certainly be cause for concern. But on the other hand, given the limited amount of information observable about the job, it is possible that if further richness could be priced, even more of the gender pay gap might be explained.

At the very least, the findings indicate that the gender pay gap is to a large extent
a relative one, whereas those by race and parent background are absolute gaps in total compensation. The absolute differences in compensation may be due to differences in general-purpose skills such as education, which may themselves be due to discrimination, for instance along the lines of Neal and Johnson (1996). The absolute differences in job quality may also stem from discrimination in the labor market, as in search model of Black (1995). In contrast, much of the gender gap does not seem to be explained by such general channels. As I outline next, the gender gap appears to be best explained by the fact that the average man and average woman, in addition to having very different incomes, also have observable different jobs in other ways. An as we have already seen, differences in job characteristics command prices in the labor market.

VI.C Amenity quantities and prices

In principle, if the methodology put forward by this paper allows us to determine to what extent inequalities are generated by amenity substitution, the methodology should also be able to shed light on which amenities are important. However, it is worth reiterating that the methodology does not solve issues relating to omitted amenities that are correlated with observed ones, nor would it fix a problem that some may be measured so noisily as to be useless. While the results give credence to the idea that job features are important, the question of which job features matter is inherently subject to which job features the researcher chooses to include. The goal of this paper is not to argue that a definitive set of job features has been found, but rather to propose a methodology that many researchers can apply with other sets of job features. This exercise should be viewed as instructive rather than exhaustive.

Figure 12 provides intuition as to how one might decompose the result on the gender pay gap into prices and quantities of amenities. The amenity quantities in Panel A are a simple but instructive way to begin to think about the implications of this paper for the gender pay gap by describing the extent to which the jobs of men and women tend to differ among...
observable dimensions, without taking a stance on what types of jobs are “better” or “worse.” For instance, although average income of women is lower than men, the average quantity of on-the-job safety is .47 SD higher among women. There are two potential implications consistent with this fact. First, this fact would be consistent with the lower earnings of women being partly due to costly substitution toward the types of jobs that tend to provide workers with safer environments. That would be the case if safety is a costly amenity, but we should also consider the case that safety is actually a disamenity, which is the typical finding of observational approaches. If safety is a disamenity that workers get paid more to take on, then safety is an amenity that offsets the gender pay gap: i.e., women get paid more than men due to their increased safety, whereas countervailing differences in other amenities generate the overall trend that women as a group earn less than men. To interpret the descriptive amenity quantities reported in Panel A, we need to have some idea of what the price of amenities may be. Depending on the sign of the price, we would find that an amenity like safety either contributes to or mitigates income differences by gender.

Panel B displays multi-variate price estimates for the same bundle of amenities shown in Panel A. Prices allow us to interpret the differences in quantities. In the example of safety, we see this amenity is valued by the labor market – a positive price, displayed here as a negative compensating differential. Holding fixed other observed job characteristics, workers on average lose $6,622 of income to obtain 1 SD of safety. A simple back-of-the-envelope calculation thus suggests, from gender differences in safety alone, we would expect to see a pay gap of .47 x -$6,621 = -$3,103. Of course, if we had found that jobs with higher safety command a compensating differential of $0, then we would have been led to conclude that this amenity does not contribute to the gender pay gap. And if the price were of the opposite sign, then our analysis would have concluded that the amenity at current prices functions to close the gender pay gap.

Panel C systematically applies the same calculation to the rest of the amenities, and is ordered by the product of quantity differences times the price. Across many of the amenities
studied, the finding is that gender differences of the amenity – at current prices – generate substantial pay gaps in favor of men. Notably, not all amenities included in this analysis follow this trend. For instance, men on average tend to have jobs with more freedom to make decisions, and decision freedom is a job attribute estimated to be costly.

VI.D Further robustness checks

The first dimension of robustness I explore in the analysis of demographic pay gaps is the choice of $h$ variable that proxies for the level of the total compensation frontier. Figure 13 summarizes the main results for the gaps by gender and race for the six candidate $h$ variables described earlier. I examine only the gaps by gender and race here, as such a wide set of $h$ variables is not available in the dataset used to measure the gap by parent income. Between 50% and 75% of the gender pay gap in the main analysis sample persists within frontiers, depending on choice of $h$ variable. In contrast, the results by race straddle 0 fairly closely. Reassuringly, the findings on demographic pay gaps are relatively similar regardless of which $h$ variable is used.

The remaining robustness checks described below will also be summarized in Table 3, in the order in which they are discussed.

Much prior literature has emphasized the sensitivity of the gender pay gap to how one defines the sample. In particular, differences in hours worked seem to play important roles. For this reason, Figure A.4 reproduces Figure 11 when the entire analysis is run only on the sub-sample of workers who report working at least 40 hours per week. As expected, the gender gap in this sample is substantially smaller, at $24,273. The gap controlling for offer curves in this specification falls by a smaller amount $18,049. Thus, when a restriction on working hours imposed to make the comparison among full-time workers, the analysis now suggests amenities account for 74% of the gender pay gap. It is also worth noting that, whereas the gender gap falls considerably upon imposing the hours restriction, the race gap rises substantially, though it is still close to 0 when conditioning on frontiers.
Another concern particularly relevant to the gender pay gap is whether discrimination in the labor market may lead men and women to face different prices. Such a case seems unlikely if labor market discrimination functions in an absolute way to reduce women’s total compensation. On the other hand, discrimination might function in a more relative way, in which different genders must pay a different wage cost to enter certain types of occupations. In such a model, the Rosen frontiers for men and women would have different slopes. To explore this possibility, I calculate amenity prices separately by gender, and use sex-specific price estimates to build Rosen frontiers. Within frontiers of sex-specific slopes, 82% of the gender gap persists. I also perform additional robustness applying female-specific and male-specific prices to the entire sample, which yield respective estimates of 41% and 76% of the gender gap remaining.

More generally, major caveat worth reiterating is that the results are likely to be sensitive to which amenities are included, as discussed in the Introduction. I do not have a systematic solution to this problem. However, in an attempt to broaden the scope of the amenities beyond the types of measures that have typically been examined by economists, I turned to the sociological literature, in which several studies relating to occupational prestige have been conducted dating back to Duncan’s (1961) socioeconomic index and Blau and Duncan’s (1967) application of the index to occupational mobility across generations. A key input into the original prestige rankings was surveyed opinion scores. Davis et al. (1992) provide survey measures of prestige rankings from the 1989 General Social Survey, but updated for the 1980 Census occupation titles. Occupational prestige was measured among a national sample of 1,500 respondents asked to rank each occupation’s prestige on a scale from 1 to 9. Not every respondent had to rank every occupation, but sufficient overlap was included to normalize respondents’ scores. Figure A.5 modifies the analysis of Figure 11 to include the occupational prestige score as an amenity. In this case, the additional dimension of heterogeneity has magnified each bar: 87% of the gender gap and 30% of the gap by parent income are now attributed to costly amenity substitution, and the race gap is 8% lower due
Finally, although the main analysis has priced amenities at the occupation-level, the fact that some amenities are observed at the individual-level begs the question of whether the analysis would be much different if all amenities were just priced at the individual-level. In this case, pricing amenities at the individual level happens to have magnified the findings slightly by gender and parent background to 84% and 21% respectively, and the race result is still relatively small at -6%. The fact that some results are made larger when amenities are priced at the individual level may indicate that there is meaningful variation to exploit in the relationship between income, amenities, and skills at a more granular level than at the occupation.

**VI.E The naïve analysis**

The methodological contribution of this paper is to clarify how to handle simultaneity from unobserved ability in estimating compensating differentials. Yet the question of this section relates solely to amenities. Could this question have been answered with data on amenities alone?

A seemingly more straightforward approach to understanding the role of amenities in the gender income gap would be to simply control for them. If amenities played a large role in the gender pay gap, which is what my prior results have indicated, we would expect to see little-to-no income differences between men and women once statistical controls for amenities are introduced into a regression of income on gender. However, Figure 14 shows that the opposite finding is true. Once my standard set of amenities is included as a control, income differences between men and women if anything slightly *widen*. Conventional wisdom might suggest that amenities, as they have been measured here, must not contribute to the gender pay gap because their inclusion as controls has not reduced the regression coefficient. Clearly, that conclusion would be in conflict with my main findings of this section.

To evaluate the credibility of the two approaches, one needs to specify what one thinks
is in the structural error term of that regression of income on amenities. In this paper, I have argued that the unobserved factors that lead workers to earn higher wages—such as unobserved ability—would also lead to better jobs in other dimensions. If this is the case, as is shown in Online Appendix D.B, even if the full set of amenities were included, the regression would still not be “identified,” in the sense that its structural error term will always be correlated with the regressors. The only case in which this regression would identify a parameter of interest is if the entire structural error term were also observed and could be controlled for. In that case, the regression should also have a perfect fit and $R^2$ of 1.

VI.F Confirming prior results

This paper puts forward a new way to look at typical data on income, amenities, and observed skills. And yet, the findings of this section might appear old, or even obvious to some. A great deal of literature on gender inequality has increasingly moved toward understanding compensating differentials. A sweeping analysis of the gender pay gap by Goldin (2014) pointed a finger toward pay schedules that are non-linearly increasing in hours, particularly in male-dominated occupations. In a recent review of literature on the gender wage gap, Blau and Kahn (2017) concluded that in the current labor market, with the gender education gap now reversed, conventional human capital variables explain “little of the gender wage gap” while “gender differences in location in the labor market—distribution by occupation and industry—continued to be important in explaining the gap.” These differences between where men and women tend sit in the labor market is a “factor long highlighted in research on the gender wage gap.” Much recent work using stated and real choices has also drawn attention to the fact that men and women have different distributions of valuations for various workplace amenities, which, along with prices, drives my results.

Less work has sought to analyze the effect of amenity substitution in other demographic pay gaps. A notable exception is work by Jencks and Tach (2005) to systematically review available evidence on the sources of intergenerational income inequality. Some evidence does
exist on the propensity for kids to take on similar roles in the workplace to their parents – for instance, the intergenerational correlation in hours worked documented by Altonji and Dunn (2000) and studies documenting trends in self-employment across generations. Jencks and Tach (2005) offer an educated guess that at least 10% of the intergenerational income correlation might be due to the effect of a broad class of values passed across generations, including hours worked, location, and choice of non-monetary work characteristics. In other words, they argue that part of the intergenerational income gap is simply due to the tendency of children from lower-income backgrounds to choose lower-paying jobs with better amenities. My results are most consistent with this theory. A contrary theory, which I have not encountered in the literature, is that children from low-income families would substitute toward higher-paying jobs with worse amenities so as to compensate for lost resources of parent background. Under such a theory, the intergenerational income gap in income terms would be even larger if kids from poor backgrounds had the same preferences as those from richer ones; my results are less consistent with that theory.

Although the results of this section are not entirely new, there may be much to learn in the procedure by which we obtained them. The mounting scholarship on the role of job amenities in the gender pay gap stands in discord with a literature focused on estimating the prices of such amenities. The central finding of this literature, to echo the words of Bonhomme and Jolivet (2009), appears to be a “pervasive absence of compensating differentials.” But these findings demonstrate that when more versatile econometric techniques are used to estimate compensating differentials, a key finding on the importance of job amenities in the gender pay gap in the current US labor market can also be obtained from widely available datasets. The econometric techniques can now be applied to datasets of other countries and other decades in order to better understand the role of compensating differentials in other income inequalities of interest.
VII Conclusion

In their history of instrumental variables regression, Stock and Trebbi (2003) write about an influential economic text written prior to the discovery of instrumental variables. Based on the strong positive correlation between price and quantity of iron, economist Henry Moore had concluded in 1914 he discovered a “new type” of upward sloping demand curve. Since that time, the vocabulary and procedure of instrumental variables has allowed economists to more precisely separate the structural price-quantity relationships of supply and demand.

Calling a regression of wage on amenities with controls a “compensating differential” is as crude as calling a regression of price on quantity a “demand curve.” Both approaches neglect the underlying economic structure of the data. Prices and quantities are each caused by demand shocks and supply shocks; it is necessary to specify which variation we will isolate. We now know that a few observed supply shocks would more-than identify the demand curve, but that would not be achieved just by including the supply shifters as controls when regressing prices on quantities. Similarly, a basic economic model of compensating differentials would specify that wages and amenities are each caused by some notion of absolute ability and relative valuations. Here, too, I have shown that a single ability shifter can identify the wage-amenity variation due to preferences, but that is not achieved by the typical approach of including it as a control in a regression of wage on amenity. Instead, this paper develops a new estimator to fit the empirical context that seemed to already be on the tip of many researchers’ tongues: that of having some rough measures of the ability channel, but not the full picture. The estimator has a similar motivation to instrumental variables, but differs in that the variable being “instrumented” for – the ability channel – is unobserved. Additional applications of the estimator likely exist in other fields of economics. Unlike the typical quasi-experimental framework for instrumental variables, a necessary feature of problems for which the estimator is useful seems to be that the unobserved variable for which a noisy proxy exists is the only unobserved variable in the structural equation. This feature of the problem allows the unobserved variable to be written as some weighted combination
of observed variables, although it is counter-intuitively a causal component of both.

When I apply the proposed estimator to datasets used for decades to study job amenities, I find price estimates that, relative to typical observational estimates, are more in line with both quasi-experimental evidence and common sense. This finding confirms the analysis of Hwang et al. (1992) that the bias from the unobserved ability channel in the typical observational approach was substantial. It also answers their call for new econometric methods to deal with it. With improved estimates of the prices workers must pay for some of the observable dimensions in which jobs differ, I turn to an analysis of demographic pay gaps. At the estimated prices, I find that job amenities play a prominent role in income inequality by gender, and essentially no role in the gaps by race or by parent income. This finding highlights that in the current US labor market, income inequality by gender has very different roots compared to inequalities by race or by parent background.

In speaking to policy relevance, it is important to note that this paper does not find that the gender gap is “caused by” differing valuations of work by men and women. Differences in valuations alone cannot generate pay gaps unless the amenity has a non-zero price. Rather, the analysis indicates that it is the combination of differing valuations and corresponding prices of job amenities that generates a substantial portion of the gender gap. A natural economic approach to reducing the gender gap would be to consider whether the prices of amenities can be changed. In fact, the government already regulates many aspects of the workplace, including through OSHA and EEOC. My findings imply that efforts to regulate cheap provision of amenities that are valued by women could have large potential to reduce the gender pay gap.

An alternative implication, at least at face value, seems more controversial: If women’s valuations of certain job attributes were more similar to men’s, then there would be less income inequality by gender. A nuance of this conclusion is that it is still open for debate the extent to which differing valuations of job characteristics stem from preferences or from constraints many women may face outside of the labor market, such as those relating to
childcare. Individual workers may be unable to remove such constraints.

Finally, the empirical exercises provided in this paper are limited in scope. Viewing these exercises as a proof in concept, there is a great deal more the new estimation strategy can teach us about compensating differentials. More research should be undertaken to explore the contributions of amenities to other pay gaps, as well as to the intersections of the identities studied here. Whereas my analysis has focused on a cohort of older workers, questions remain about how valuations of job features change over the lifecycle, particularly around life events such as childbirth. The estimation strategy is simple enough that researchers can test their own hypotheses about the role of their preferred set of job amenities in income inequality. The estimator can be applied to other datasets, such as the data on sexual harassment claims that Hersch (2011) used to study compensating differentials. Bigger datasets with more individual-level variation would yield more precision. Future work can exploit individual-level measures of job amenities commonly available, such as paid vacation days, child care, and health insurance. The trade-offs between power and noise when aggregating amenities to the occupation-level should be better explored, as well as methods for inference relating to demographic gaps. Lastly, other proprietary and publicly available datasets can help shed light on how the price of amenities has changed over time and how prices differ across countries’ labor markets. For instance, how have compensating differentials for jobs providing flexible hours, childcare, or paid maternity leave changed in the US as more women have entered the labor market? How did the price of jobs with certain schedules change following France’s adoption of the 35-hour work week in 2000? Such analyses can help reveal the responsiveness of amenity prices both to the changing composition of the labor pool and to differences in government regulations.
References


VIII Figures

Figure 1: The Classic Problem

(a) Income by Safety

![Graph showing the relationship between income and safety with a slope of 0.1506 (0.0345).]

(b) With Ability Controls

![Graph showing the relationship between income and safety with a slope of 0.0235 (0.0254).]

Notes: Panel A is a binned scatterplot of income on ten bins of safety. The red line is the best-fit line on the unbinned data. Its slope and standard error are noted on the bottom left of the graph. Panel B includes controls that attempt to proxy for worker ability: years of education, age, and AFQT. Data are NLSY respondents in 2012 and standard errors are clustered at the occupation level; additional information on data is contained in Section V.A.
Figure 2: Graphical Intuition

(a) OLS

(b) Proposed Estimator

Notes: This figure describes two estimation strategies on hypothetical data in which higher-ability workers get more wage and amenity, such that a trade-off exists among workers of a given ability. True ability is imperfectly observed by the researcher in only 3 rough bins, which I call “observed ability.” The dots represent workers and are shaded in accordance with observed skill level, with blue representing lowest, then red, then green. Panel A describes a regression of wage on amenity with fixed effects for 3 levels of observed ability. The fixed effects estimator here is depicted as the within-ability estimator of wage on amenity, and the final average over all ability levels is given by the black line. The upward-sloping OLS estimate mirrors the finding of Figure 1. Panel B describes the proposed estimator of observed ability on wage and amenity, described in more detail in the text. The slope of the regression line labeled Step 1, which tells us the direction in which observed ability increases, is \( \hat{\beta} \). Of greater interest, the line orthogonal to that regression line, labeled Step 2, corresponds to wage-amenity variation when skill is held constant. The formula for the slope of that line is the negative inverse of the regression line, \(-\frac{\hat{\delta}}{\hat{\beta}}\), which corresponds to the true price of the amenity in this model. Alternatively, the exact estimate could have been found by obtaining predicted values from the regression of observed ability on wage and amenity (denoted ability), then regressing wage on amenity with those predicted values as a control. Approximate levels of ability appear on the graph as colored bandwidths.
Figure 3: The “Kissing Equilibrium”

Panel A: Worker and Firm Tangency

Panel B: The Frontier

Panel C: Multiple Frontiers

Notes: This figure graphically represents the equilibrium between labor supply and demand discussed in Section II, and illustrates how the case of a binary amenity can be thought of as a discretization of a continuous model. In Panel A, an indifference curve for a particular worker is drawn in green, such that $C_i$ represents the amount of wage required to make the worker indifferent with a one-unit change of amenity (a change from 1 to 0, following the binary treatment of Online Appendix B). The blue curve depicts the analogous level-set of profits for the firm that has matched with the worker. The particular level sets of utility and production are chosen such that they meet at a tangency. Panel B shows the tangency for this pair and for an additional worker-firm pair. The frontier $\Psi$ is drawn in black such that it passes through the tangencies for all pairs, and such that $\Psi$ is tangent to each worker/firm indifference curve (which restricts which level-sets each pair chooses). Appendix Figure A.6 illustrates that if a worker-firm pair chose a level of wages and amenities such that the frontier did not satisfy the latter tangency condition, then the match would not be stable. Finally, Panel C demonstrates the same sorting equilibrium for multiple sub-markets that are ordered by how much compensation they entail. For a single worker of a particular ability level $\eta$, the corresponding $\Psi$ frontier constitutes the set of obtainable jobs, similar to a budget set. The local slope of $\Psi$ is viewed by the worker as the price of the amenity, in wage units.
This figure discretizes a hypothetical labor market into three segments of high, medium, and low compensation, represented by linear Rosen frontiers drawn as heavy black lines. Colors are used to depict levels of observed $h$; in ascending order blue then red then green. No bias is present in Panel A because within each $\eta$-level frontier, $h$ is distributed independently of $w$ and $z$. When the ability channel is normalized to this $h$ variable, the estimated price would lie on-top of the true frontiers as desired. In Panel B, additional observations have been added at the edges of the frontiers such that $h$ is clearly not independent of $w$ and $z$ conditional on the frontier. Matching the example discussed in Section IV.A, on each frontier workers with the highest $h$ now also have the highest $z$, and likewise those with the lowest $h$ have the highest $w$. The positive conditional covariance of $h$ and $z$ leads to a rotation of the price estimate, shown in dashed lines. The slope estimate is more negative than the true one. In other words, we would conclude that workers give up more wage for this amenity than is truly the case.
Notes: Panel A is a binned scatterplot of income by bins of safety, separately by terciles of AFQT, in the NLSY analysis sample at the occupation level. Green diamonds are workers with the highest AFQT, then red squares, and blue circles are the lowest-AFQT workers. The black line is the linear best-fit: on average, workers with more safety also earn more income, even within skill bins. Panel B plots mean AFQT by income and safety; these means have been discretized into terciles for visualization. The black line is the linear best-fit of income on safety controlling for the bin’s mean AFQT. Panel C is similar to Panel B, except the dots are shaded in the following way. First, the center dot is made dark to indicate it is the reference point. Then, two-sided t-tests are performed to test whether the mean in the reference bin is significantly different from each of the other bins. Bins in which mean AFQT is indistinguishable at the $\alpha = .05$ level are colored in blue, and the rest in grey. For all panels of this figure, income and safety are in percentiles rather than standard deviations. All panels of this figure use the odd number of 9 bins so that a center dot can be chosen in Panel C. Additional sample details are found in Section V.A.
Figure 6: Raw, Mincerian, Fixed-Effects, and Proposed Estimates

Notes: Each data point on the bar graph is a separate (bivariate) regression of the price of the amenity versus wage, controlling for no other amenities, in the NLSY sample at the occupation level. Income and all amenities are in standard deviation units. Mincerian includes indicators for years of education, exact AFQT score, and age. 95% CI’s are calculated at the occupation level, and are identification-robust for the proposed estimates using AFQT. All series use the NLSY sample in 2012 except for the individual fixed effects specification, which extends the years to 2004-2014. Additional sample details are found in Section V.A.
Figure 7: Revisiting Safety: Comparison of $h$ Variables

Notes: The horizontal axis represents the coefficient on the amenity in a regression of income on the one amenity controlling for $\hat{h}$, the predicted values from a regression of $h$ on income and the amenity. Negative values indicate that workers on average lose income to acquire the amenity. All regressions are on the NLSY sample and run at the occupation level, weighted by number of workers. Anderson-Rubin 95% confidence intervals shown at the occupation level. Additional sample details are found in Section V.A.
Notes: This graph compares the results when the $h$ variable used is AFQT versus height residualized by sex. The correlation between residualized height and AFQT in the NLSY is 0.1504 at the individual level and 0.3974 at the occupation level. The horizontal axis represents the coefficient on the amenity in a regression of income on the one amenity controlling for mean $h$ in the NLSY sample at the occupation level. Additional sample details are found in Section V.A.
Figure 9: Separate & Multivariate Amenity Prices

Notes: The scatterplot compares linear price estimates when amenities are priced separately (on the horizontal axis) versus in a multivariate framework (on the vertical axis). The \( h \) variable is AFQT. The red line is the 45\(^\circ\) line. Additional sample details are found in Section V.A.
Notes: Both panels depict superimposed gradients of slope -.29, corresponding to the linear price estimate of this amenity when AFQT is used as the $h$ variable; both axes are in standard deviations. The widths of each of the 10 bandwidths are not drawn to scale, but loosely corresponds to the region of the graph in which workers who obtain a particular decile of compensation would fall. The exact bandwidths used for calculations were calculated in the following way. First, AFQT was regressed on income and regular schedule at the occupation level weighted by number of workers; then, at the individual level, predicted values were calculated. The predicted values were grouped into 12 quantiles, and the top and bottom quantiles were dropped. This left 10 quantiles of predicted values, which correspond to (winsorized) quantiles of total compensation. The data points were constructed in the following way using the exact quantiles. For Panel A, I plot the average income and amenity values for men and women, separately, by decile of total compensation. Mens’ locations are summarized by blue dots, and women’s locations by red diamonds. Panel B is constructed in the exact same way, but separately by race rather than by gender. Red diamonds summarize the locations of “black” respondents, and blue dots refer to respondents who identify as “non-black, non-hispanic.” The data underlying the figure is the main NLSY analysis sample; additional sample details are found in Section V.A.
Figure 11: Income Gaps vs. Gaps in Frontiers

Notes: The height of the blue bar is the unadjusted income gap among all employed respondents. Data by gender (men minus women) and race (non-black/non-hispanic minus black) are from the NLSY in 2012, and data by parents (above minus below median parent income) are from PSID in 2013. The red bar is the coefficient on the demographic group in a regression of income on the group with a linear control for the multi-dimensional Rosen frontier. The frontier prices the following amenities: nonprofit (NLSY only), sitting, unstructured, repetition, decision freedom, regular schedule (NLSY only), safety, and time pressure. In the NLSY, AFQT is used as the $h$ variable; for the PSID results by parent background, the $h$ variable is education. Additional sample details are found in Section V.A.
Figure 12: Gender Pay Gap in Quantities and Prices

(a) Quantities

(b) Prices

(c) Quantities x Prices

Notes: Panel A plots the coefficients from separate regressions of the amenity on an indicator for female in the NLSY-O*NET sample. The ordering is in accordance with the result in Panel C. To construct Panel B, I first obtained predicted values from a regression of AFQT on income and all amenities shown at the occupation level, weighted by number of workers. The bars in Panel B correspond to the coefficients from a regression of income on these amenities controlling for predicted values. For this figure, yearly income is in dollars rather than standard deviations. Panel C plots the data in Panels A and B multiplied. Additional sample details are found in Section V.A.
Figure 13: Pay Gap Accounting in the NLSY, by h variable

(a) Gender Gap

(b) Race Gap

Notes: This figure repeats parts of the analysis of Figure 11 with different h variables. Panel A repeats the result by gender, and Panel B repeats the result by race. A similar re-analysis is not performed for result by parent background because the same set of h variables is not available in the PSID. Additional sample details are found in Section V.A.
Notes: This graph mirrors the construction of Figure 2, in which the height of the blue bar is the unadjusted income gap among all employed respondents. Whereas the red bars of Figure 2 represented the income gaps controlling for Rosen frontiers (the ability channel), this graph simply plots the coefficient on the demographic group variable on a regression of income on the group controlling for amenities. The sample definition and set of amenities are exactly the same as Figure 2.
IX Tables

Table 1: Correlations Among Income & Proposed $h$ Variables

<table>
<thead>
<tr>
<th></th>
<th>Income</th>
<th>Educ.</th>
<th>AFQT</th>
<th>Self-Est</th>
<th>Mastery</th>
<th>Loc. C.</th>
<th>Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years Educ.</td>
<td>0.3685</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFQT</td>
<td>0.3579</td>
<td>0.5426</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Esteem</td>
<td>0.1871</td>
<td>0.2803</td>
<td>0.3175</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery</td>
<td>0.1816</td>
<td>0.2392</td>
<td>0.2084</td>
<td>0.2922</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus of Control</td>
<td>-0.1501</td>
<td>-0.2117</td>
<td>-0.3095</td>
<td>-0.2814</td>
<td>-0.1605</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Height (res. sex)</td>
<td>0.066</td>
<td>0.1185</td>
<td>0.1486</td>
<td>0.0593</td>
<td>0.0577</td>
<td>-0.0492</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: This correlation matrix summarizes the main NLSY analysis sample with all variables non-missing (4,257 workers). Additional sample details and definitions of $h$ variables are contained in Section V.A.
Table 2: Income Gaps

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>29296.4</td>
<td>19396.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td></td>
<td></td>
<td>22407.1</td>
<td>-1959.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above-Median Parents</td>
<td></td>
<td></td>
<td></td>
<td>21813.6</td>
<td>1409.2</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>41250.9</td>
<td>41992.6</td>
<td></td>
<td></td>
<td>39988.8</td>
<td></td>
</tr>
<tr>
<td>Frontier Control</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>N</td>
<td>4868</td>
<td>4868</td>
<td>3977</td>
<td>3977</td>
<td>929</td>
<td>929</td>
</tr>
</tbody>
</table>

Notes: All columns report linear regressions in which the dependent variable is income in 2012 (or 2013, for the parental sample in columns 5 and 6). The sample is the main NLSY analysis sample with no hours restrictions for columns 1-4, and the PSID analysis sample for columns 5 and 6. Additional sample details are found in Section V.A. Standard errors are not included for this table due to the difficulty of calculating them for the specifications controlling for the frontier. The controls for Rosen frontiers in columns 2, 4, and 6 are constructed as described in Section VI.B, in which the frontier prices the following amenities: nonprofit (NLSY only), sitting, unstructured, repetition, decision freedom, regular schedule (NLSY only), safety, and time pressure. In the NLSY, AFQT is used as the $h$ variable; for the PSID results by parent background, the $h$ variable is education.
Table 3: Additional Specifications for Demographic Gaps

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>baseline</td>
<td>full-time</td>
<td>own-sex prices</td>
<td>female prices</td>
<td>male prices</td>
<td>including prestige</td>
<td>individual prices</td>
</tr>
<tr>
<td>Gender</td>
<td>.662</td>
<td>.744</td>
<td>.815</td>
<td>.405</td>
<td>.756</td>
<td>.873</td>
<td>.844</td>
</tr>
<tr>
<td>Race</td>
<td>-.087</td>
<td>-.052</td>
<td>-.085</td>
<td>-.126</td>
<td>-.105</td>
<td>-.088</td>
<td>-.056</td>
</tr>
<tr>
<td>Parents</td>
<td>.065</td>
<td>.068</td>
<td>.157</td>
<td>-.001</td>
<td>.082</td>
<td>.296</td>
<td>.213</td>
</tr>
</tbody>
</table>

Notes: All columns report the share of the income gap remaining after Rosen frontiers are included as controls, and are described in more detail in Section VI.D. Column 1 reports these numbers for the specification of Figure 11, and all other columns are perturbations of this exercise as follows. Column 2 restricts the sample to workers who report at least 40 hours of work per week. Column 3 estimates prices separately by gender and applies the relevant gender’s prices. Columns 4 applies to the whole sample the prices estimated among only women, and Column 5 applies prices estimated among men. Column 6 additionally includes the amenity of prestige sourced from the sociology literature, as discussed in Section VI.D. Column 7 prices amenities without aggregating to the occupation level. Additional sample details are found in Section V.A.
The Online Appendix to this paper is available at http://alexbell.net/jmp.html