Changes in Returns to Task-Specific Skills and Gender Wage Gap

(PRELIMINARY DRAFT)

Shintaro Yamaguchi*

April 23, 2012

Abstract

How did skilled-biased technological change affect wage inequality, particularly between men and women? To answer this question, this paper constructs a task-based Roy model in which workers possess a bundle of basic skills and occupations are characterized as a bundle of basic tasks. The model is estimated by a nonlinear correlated random effect regression using data from the Dictionary of Occupational Titles and the PSID. I find that men have more motor skills than women, but the returns to motor skills have dropped significantly, accounting for about a half of the narrowing gender wage gap. In addition, increases in women's cognitive and general skills contributed to the reduction in the gender wage gap significantly.

Keywords: Roy model, task-based approach, occupational choice, skill-biased technological change, soft skills.

JEL Codes: J24, J31

*Department of Economics, McMaster University, 1280 Main St. W., Hamilton, ON. Canada L8S 4M4. Email: yamtaro@mcmaster.ca. Website: socserv.mcmaster.ca/yamtaro. Supplementary appendix is available from the author's website. I am grateful to comments from participants of the seminars at Toronto and WashU, ASSA meeting, and Workshop on Labour Markets and the Macroeconomy. The use of SHARCNET computational facilities is gratefully acknowledged.

1 Introduction

This paper studies the contribution of technological change to the narrowing gender wage gap between 1979 and 1996 using data from the Panel Study of Income Dynamics (PSID) and the Dictionary of Occupational Titles (DOT). Figure 1 presents U.S. male wage inequality from 1975 to 2000. Inequality rapidly rises in the 1980's and continues to grow over the 1990s, but at a slower pace. Empirical evidence[1] shows that technological change since the early 1980's has raised the return to skills, increasing wage inequality. Figure 1 also demonstrates that women's wages relative to men's dramatically increased and that the two series of wage inequality measures exhibit similar changes during the same period. Despite the co-movement[2] of male wage inequality and women's relative wages, many papers, including Blau and Kahn (1997, 2000) and Card and DiNardo (2002), do not consider the role that technological change played in narrowing the gender wage gap, because men were more educated and experienced than women during the period. They argue that, despite technological change being unfavorable to women, the gender wage gap narrowed due to increases in the measured and unmeasured human capital possessed by women, and possibly a reduction in gender discrimination.

However, when one takes a closer look at the workplace, technological change seems to have reduced women's disadvantages. It is widely documented that computer-based technology has been adopted across plants, and is used to monitor the production process, making work much less arduous and physical (see Bureau of Labor Statistics (1994), for example). Weinberg (2000) provides empirical evidence that the de-emphasis in physical skills following computerization has increased the demand for female workers. Autor, Levy, and Murnane (2003) propose a theory which explains how the adoption of computer technology changes the workers' tasks and ultimately the demand for skills. In their model computers replace routine tasks, which decreases the demand for routine manual tasks. Black and Spitz-Oener (2010) and Bacolod and Blum (2010) apply the task-based approach to gender wage gap analysis and find evidence that the demand shift from manual tasks to analytical tasks plays an important role in the change in the gender wage gap. While these previous contributions provide suggestive evidence for the role of tasks in narrowing the gender wage gap, the literature lacks an explicit economic model to interpret the empirical findings. How and why are workers sorted across tasks? How are wages determined according to their skills and tasks? How should skills be measured? These basic questions remain unanswered.

The objective of this paper is to construct and estimate a Roy model of worker-task assignment that accounts for an equilibrium wage determination, in order to examine the sources of the narrowing gender wage gap from 1979 to 1996. I extend a standard Roy model of self-selection by

[1] See Katz and Murphy (1992), Autor, Katz, and Krueger (1998), and Krusell, Ohanian, Rios-Rull, and Violante (2000), among others.

[2] The correlation coefficient for the two series is 0.95.

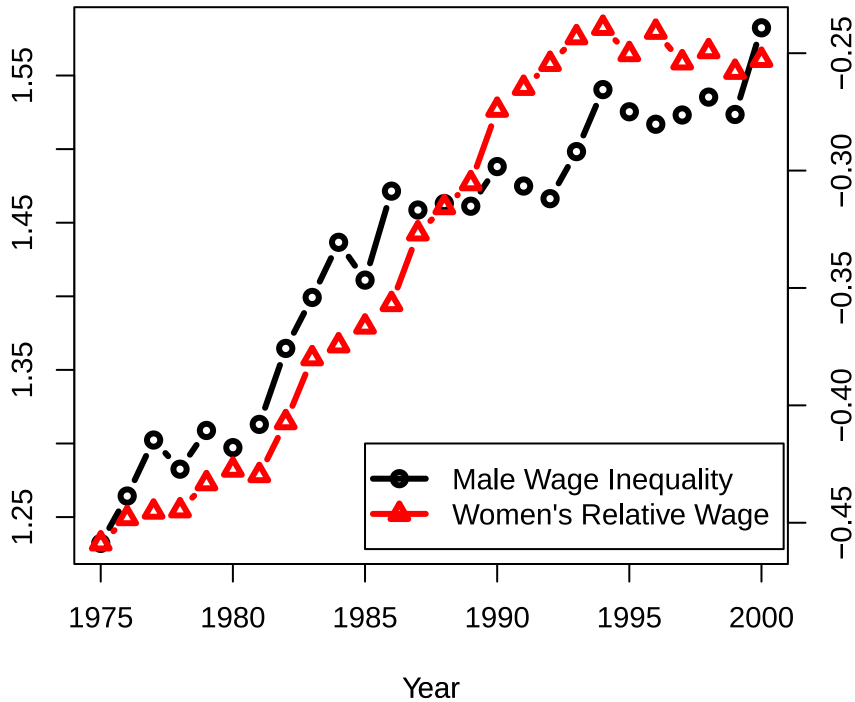


Figure 1: Wage Inequality Within and Between Genders (1975-2000)

Source: CPS 1976-2001. The reported wage in the CPS is the last year's. The sample includes civilian male and female non self-employed full-time ($\geq$ 1500 hours/year) workers in the non-agricultural sector between 18 and 65 years old. The wages are deflated by the 1983 PCE deflator. Wages less than \$1 per hour and more than \$250 per hour are dropped from the sample.

Note: Male wage inequality (left scale) is defined as a logwage difference between the 90th percentile and the 10th percentile. Women's relative wage (right scale) is on the right scale and given by the mean logwage of women less the mean logwage of men.

incorporating the recently popularized task-based approach. In the task-based approach, an occupation is viewed as a bundle of tasks, and differences between occupations are interpreted from the viewpoint of tasks. The task-based approach allows researchers to take a closer look at what workers do, which enables a deeper understanding of occupations that when they are treated as distinct categories.

The model focuses on two broadly defined task categories: cognitive and motor tasks. Jobs are characterized in terms of the level of complexity of cognitive and motor tasks. For example, tasks of a professional worker are characterized by a complex cognitive task and a simple motor task, while those of a craft worker are characterized by a simple cognitive task and a complex motor task. Workers possess cognitive and motor skills, and apply these task-specific skills to perform cognitive and motor tasks. The marginal products of the task-specific skills vary across occupations depending on the complexity of the tasks required of the occupation. Cognitive (motor) skills are more intensely used when the cognitive (motor) tasks are complex, but do not affect productivity substantially when tasks are simple. Workers develop their skills through education and work experience. Once workers enter the labor market, they acquire skills through learning-by-doing, and the speed of skill growth depends on the complexity of the tasks performed in their occupations. When a worker spends many years doing complex motor (cognitive) tasks, she will develop more motor (cognitive) skills than she would in occupations characterized by relatively simple motor tasks. In order to account for worker heterogeneity aside from task specific skills, the model also incorporates general skills whose productivity and growth rate do not vary across occupations.

Using data from the DOT and the PSID, I estimate the task-specific and general skill endowments for men and women, and their rate of return from 1979 to 1996 through a nonlinear correlated random effect regression (see Wooldridge (2009)). By taking advantage of the panel structure of the data, this method allows me to control for time-invariant unobserved skills. The model allows for heterogeneous skill and wage growth profiles across workers, because the accumulated skill endowments depend on the tasks of the previous jobs. By including a broad set of variables which control for worker skills, a possible endogeneity bias caused by unobserved worker heterogeneity is largely avoided. I provide evidence that the proposed method effectively takes care of the endogeneity problem at least to the same extent that known instrumental variables for occupations would.

The estimation results indicate that women have slightly more cognitive skills than men, while men have substantially more motor and general skills than women. I also find that, from 1979 to 1996, returns to motor skills dropped dramatically and returns to cognitive skills modestly increased. These changes in returns suggest that recent technological change worked in favor of women. The estimated model is used to construct an Oaxaca-Blinder decomposition of the change in the gender wage gap. The decomposition indicates that the decrease in returns to motor skills is mostly responsible for the narrowing gap, accounting for about half of the overall change. Increased cognitive and general skill endowments for women account for most of the remaining change. The rise in the returns to cognitive skills did not affect the gender wage gap directly, because the gender difference in cognitive skill endowments is small, although it may have encouraged women to invest in cognitive skills. These main results hold even when I take into account changes in gender discrimination and in the pattern of women's labor force participation.

Galor and Weil (1996) recognize the importance of gender differences in brains and brawn to explain the narrowing gender gap in the labor market. Welch (2000) also develops a brains-and-brawn model of earnings to explain the narrowing gender wage gap as well as the rising male wage inequality in a single framework. He argues that men's skills are brawn intensive relative to brains, while women's skills are brain intensive relative to brawn, and that the rise in the price of brains relative to brawn explains the rise in male wage inequality and the narrowing gender wage gap simultaneously. Using German data, Borghans, ter Weel, and Weinberg (2006) provide evidence for the increasing importance of people skills. They find that the employment share of people-skill intensive occupations increased significantly. Based on the Gorman-Lancaster characteristics model of earnings in Welch (1969, 2000), Rendall (2010) constructs a one sector general equilibrium model in which workers possess brains and brawn, and calibrate it to the U.S. economy to examine the sources of the narrowing gender wage gap.

The key feature of the Gorman-Lancaster characteristics model of earnings is that the returns to skills are uniform across subsectors. However, Heckman and Scheinkman (1987) prove that uniform returns to skills do not generally hold and soundly reject the uniform skill-price hypothesis using U.S. data. Moreover, Rosen (1978, 1983) points out that, under uniform skill prices, workers are indifferent across subsectors, and thus, a sectoral choice problem is precluded, which implies that the Gorman-Lancaster model cannot be applied to study worker allocation to tasks. Acemoglu and Autor (2011) and Firpo, Fortin, and Lemieux (2011) emphasize the role that occupational tasks played in the changes in the wage structure. Modeling worker assignment to tasks is an important step toward understanding workers' responses to changes in labor market conditions and technology.

Building on Heckman and Sedlacek (1985), I construct a Roy model in order to explain worker assignment and wage determination when workers possess multi-dimensional skills. Unlike the Gorman-Lancaster model, returns to skills are heterogeneous across occupations, and workers are sorted based on their skill endowments. While this paper focuses on the gender wage gap, the model provided in this paper is applicable to study other labor market changes in which occupational tasks play a key role. An obvious application is a study of labor market polarization in the U.S. since the 1990's.

The rest of the paper is structured as follows. Section 2 lays out the model and explains how workers are assigned to different tasks. Section 3 discusses the estimation strategy. Section 4 describes the data. The estimation results are presented in Section 5. Section 6 provides an empirical assessment of the identification strategy and robustness checks. The issues of gender discrimination and endogenous labor force participation of women are discussed. Section 7 concludes.

2 A Roy Model of Worker-Task Assignment

2.1 Technology

2.1.1 Aggregate Production Function

There are J occupations indexed by $j \in \{1, \cdots, J\}$. Products or services are heterogeneous across occupations. The aggregate output of occupation j produced in period t is denoted by Q_{jt}. These outputs are used by firms as intermediate inputs in the production of a homogeneous final consumption good Y_t. The aggregate production function F is given by

$$Y_t = F(Q_{1t}, \cdots, Q_{Jt}), \tag{1}$$

which is homogeneous of degree one in Q_{jt}. The aggregate output from occupation j is given by the sum of the outputs of workers in occupation j,

$$Q_{jt} = \sum_i q_j(s_{it}) d^o_{ijt}, \tag{2}$$

where $q_j(s_{it})$ is the output of a worker in occupation j when she has a skill vector s_{it} and d^o_{ijt} is an indicator variable equal to one if worker i is in occupation j in period t. Throughout, I choose the price of the final good as the numeraire.

2.1.2 Worker Productivity

Workers have a set of skills that consist of cognitive, motor, and general skills. The skill vector of worker i in year t is denoted by $s_{it} = (s_{C,it}, s_{M,it}, s_{G,it})$ where $s_{C,it}$, $s_{M,it}$, and $s_{G,it}$ are cognitive, motor, and general skills, respectively, and take non-negative values. Labor is the only factor of production. When occupation j is filled by a worker, she produces $q_j(s_{it})$ units of type j goods. The production function for occupation j is given by the following Cobb-Douglas form

$$\ln q_j(s_{it}) = \ln \alpha_j + \beta_{Cj} \ln s_{C,it} + \beta_{Mj} \ln s_{M,it} + \ln s_{G,it}, \tag{3}$$

where $\alpha_j, \beta_{Gj}, \beta_{Mj} > 0$ for all j. The parameters of the production function differ across occupations according to their tasks. Let $x_j = (x_{C,j}, x_{M,j})$ be a vector of the cognitive and motor task complexity indices that characterize the task of occupation j. The indices take non-negative values. For example, greater values of $x_{C,j}$ imply that the cognitive task of occupation j is more complex.

The parameters change with tasks such that $\alpha_j = a(x_j)$, $\beta_{k,j} = b_k(x_{k,j})$ for $k \in \{C,M\}$, and

$$\frac{\partial a}{\partial x_k} < 0 \tag{4}$$

$$\frac{\partial b_k}{\partial x_k} > 0. \tag{5}$$

These restrictions imply that output is sensitive to worker's cognitive and motor skills when tasks are complex, because the skills are intensely utilized. If tasks are simple, the outputs vary little across workers, although skilled workers always produce more than unskilled workers. Cognitive and motor skills can be considered as task-specific skills in the sense that their log marginal product of labor varies across occupations depending on task complexity. In contrast, the general skill is general in the sense that its log-marginal product of labor is constant across occupations.

This production function sorts workers across different tasks based on their skill endowments, and is an application of Gibbons, Katz, Lemieux, and Parent (2005) to multiple skill dimensions. Figure 2 illustrates how log output changes with a worker's cognitive task-specific skill between two occupations L and H that are ranked by the cognitive task complexity index such that $x_{C,L} < x_{C,H}$. For simplicity, assume that the motor task complexities of occupations L and H are identical. In both occupations, outputs increase with worker skills, but the slope of the output schedule is steeper for occupation H. Also note that the intercept of the output schedule is lower for occupation H. When the task is simple, the productivity difference between skilled and unskilled workers is small. In contrast, when the task is complex, output is sensitive to worker skills and the productivity difference between skilled and unskilled workers is amplified. Hence, in Figure 2, all workers with skills less than s^* are more productive in occupation L than in occupation H, while workers with skills more than s^* are more productive in occupation H than in occupation L.

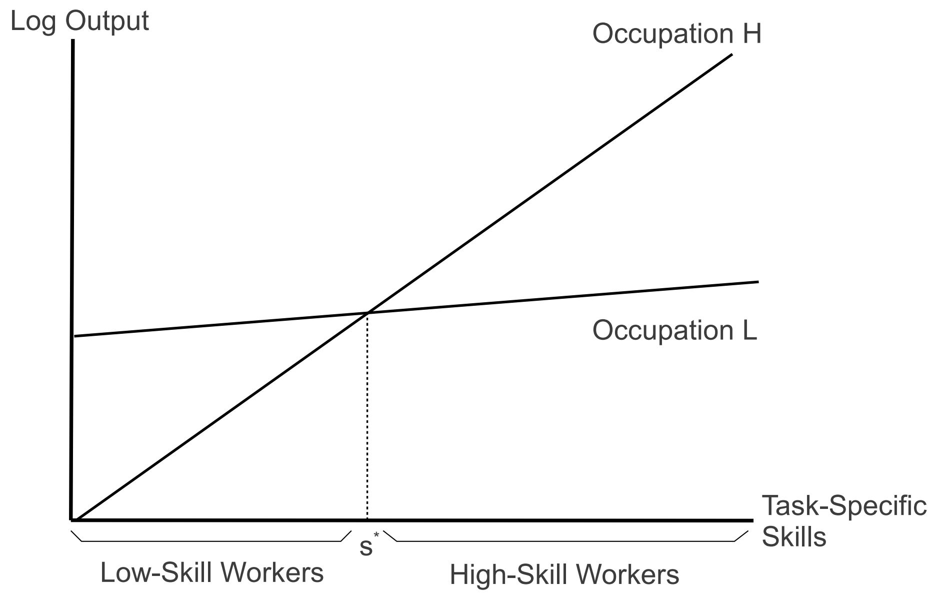


Figure 2: Occupational Sorting Based on Task-Specific Skill

Wages are paid according to the value of marginal product of labor

$$w_{ijt} \quad = \quad P_{jt} q_j(s_{it}), \tag{6}$$

where P_{jt} is the price of a type j intermediate good at period t. The intermediate output price is approximated by a time-varying function of tasks p_t,

$$P_{jt} \quad = \quad p_t(x_j). \tag{7}$$

This price function is differentiable in the task vector x_j. The logwage equation is given by

$$\ln w_{ijt} \quad = \quad \ln P_{jt} + \ln \alpha_j + \beta_{C,j} \ln s_{C,it} + \beta_{M,j} \ln s_{M,it} + \ln s_{G,it}. \tag{8}$$

With a fixed technology, the parameters α_j, $\beta_{C,j}$, and $\beta_{M,j}$ are time-invariant, and only the output price P_{jt} changes over time. Note that the intercept and returns to skills vary across occupations, which is an important feature of a Roy model. This is the key difference from the Gorman-Lancaster characteristics model in which returns to skills are uniform across subsectors of the economy.

Computerization has reshaped production technologies. One way to model technological change is to allow for changes in the aggregate production technology (Equation (1)) only, which is the approach taken by Lee and Wolpin (2006). This type of technological change affects the price of occupation-specific output P_{jt}, thus the effect appears in the occupation-specific intercept, and the returns to skills remain constant over time. However, empirical evidence presented in Firpo, Fortin, and Lemieux (2011) indicates that returns to skills within occupations have significantly changed over time. They find that this change is the key to understanding the source of wage structure changes such as the U.S. labor market polarization caused by technological change.

Prior to the introduction of sophisticated robots or other computer technologies, motor skills are highly valued in craft occupations, but not in sales occupations. When computers are adopted in the production process, the value of motor skills in craft occupations decreases, because complex (but routine) manual tasks are automated. On the other hand, the value of motor skills in sales occupations changes little, because manual tasks are not an important part of sales workers' jobs. Similarly, analytical skills are heavily used in professional occupations and returns to analytical skills in those occupations increase as computers are more and more widely used in the office. But, computers affect very little the returns to analytical skills in service occupations.

Given the empirical evidence and the theory, I incorporate the state of technology in the function governing the returns to skills. Ideally, computer investment at the occupational level should be included in the model, but the data are not readily available.[3] As an alternative, I allow for the function for returns to skills to change over time. The return to skill function b_k now depends on both task complexity x_k and time t and has the following properties,

$$\frac{\partial^2 b_C}{\partial x_C \partial t} > 0 \tag{9}$$

$$\frac{\partial^2 b_M}{\partial x_M \partial t} < 0. \tag{10}$$

The first inequality implies that the rate of the increase in returns to cognitive skills is greater for occupations that involve complex cognitive tasks. The second inequality implies that the rate of the decrease in returns to motor skills is greater for occupations that involve complex motor tasks.

2.1.3 Skill Growth

Workers enter the labor market with a vector of initial skill endowments $s_{i\underline{t}_i}$ where $\underline{t}_i$ is the year when worker i entered the labor market. The task-specific and general skills grow over time ac-

[3] As an exception, Beaudry, Doms, and Lewis (2010) use detailed data for computer capital at the establishment level.

cording to the following skill growth equations,

$$s_{k,it} = s_{k,it-1} \cdot g_k(e_{it}, x_{k,ijt}) \text{ for } k \in \{C, M\} \tag{11}$$

$$s_{G,it} = s_{G,it-1} \cdot g_G(e_{it}) \tag{12}$$

where e_{it} is general work experience and $x_{k,ijt}$ is the type k task complexity index for worker i in occupation j in period t. The skill growth functions take positive values and are differentiable with respect to experience and the task indices. These functions imply that the growth rates of task-specific skills depend on the level of complexity of the task, while those of general skills depends on experience only.

2.2 Individual Utility Maximization

Workers enter the labor market in year $\underline{t}_i$ and retire in year T_i. They supply labor inelastically and do not borrow or lend money, implying consumption equals wages $c_{it} = w_{ijt}$. In each period while in the labor market, a worker chooses an occupation to maximize her present value of lifetime utility. This is the only choice made by workers in this model. The value function of a worker with skill s_{it} is

$$V_{a,T_i} = \max_{j \in \{1,\cdots,J\}} u(c_{iT_i}) \tag{13}$$

$$V_{a,t}(s_{it}, a_{it}) = \max_{0 \le j \le J} u(c_{it}) + \rho V_{a+1,t+1}(s_{it+1}, a_{it+1}) \text{ for } t < T_i \tag{14}$$

where $c_{it} = w_{ijt}(x_{ijt}, s_{it})$,subscript a is for the age of a worker, and ρ is a discount factor. The constraints for the worker's maximization problem are given by the wage equation (8) and the skill growth equations (11) and (12).

To characterize a worker's optimal occupational choice easily, I assume that there is a continuum of occupations. Under this assumption, the choice of occupation is equivalent to the choice of task vector x. For analytical tractability, further assume $u(c_{it}) = \ln w_{ijt}$. Assuming an interior solution, the first order condition for optimality is given by differentiating the value function with respect to the task vector,

$$\frac{\partial \ln p_t(x_{ijt})}{\partial x_{ijt}} + \frac{\partial \ln q(s_{it}, x_{ijt})}{\partial x_{ijt}} + \rho \frac{\partial E[V_{a+1,t+1}(s_{it+1}, a_{it+1})|s_{it}]}{\partial s_{it+1}} \frac{\partial s_{it+1}}{\partial x_{ijt}} = 0. \tag{15}$$

In explaining how workers with different skills are self-sorted into different tasks, the second term of Equation (15) plays a central role. It captures the marginal effect of tasks on labor productivity. Note that the production function (3) expresses that labor productivity varies across occupations. As explained in Section 2.1.2, skilled workers are more productive in complex tasks, while unskilled workers are more productive in simple tasks. The first term of Equation (15) is the marginal effect of tasks on the output price. It does not change with worker skills, and thus, does not account for worker assignment to tasks. The third term captures the marginal effect of tasks on expected future utility generated through skill growth. In the estimated model, the skill growth rate increases with task complexity, which implies that all workers want to take complex tasks in order to improve their skills, and that the third term is positive for all workers. Although this skill growth effect is stronger for young workers than old workers, this term does not explain worker assignment to tasks.

A typical career progression pattern is where young workers start their careers in occupations requiring simple cognitive and complex motor tasks, and as their careers progress they gradually switch to jobs involving complex cognitive and simple motor tasks. Workers start their careers with low cognitive and high motor skills. Given that the production function is at the occupational level (see Equation (3)) , they self-select into simple cognitive and complex motor tasks. The parameter estimates indicate that workers develop cognitive skills through learning-by-doing, while their motor skills remain constant or slightly decrease due to depreciation. Later in their careers, workers possess high cognitive skills and low motor skills. Given their altered comparative advantage, older workers take complex cognitive and simple motor tasks.

Finally, the model laid out above does not incorporate preferences for tasks or a role for uncertainty over future skills, their returns, etc. Extensions in those directions are straightforward (see Yamaguchi (2012)).

3 Estimation Strategy

3.1 Parametrization

3.1.1 Wage Equation

The intermediate output price (7) and the parameters in the occupation-level production function (3) are functions of the task indices. Specifically, they are a linear or quadratic function of the task indices

$$
\begin{aligned}
\ln p_t(x_j) &= \pi_{0,t} + \pi_{1,t} x_{C,j} + \pi_{2,t} x_{M,j} + \pi_{3,t} x_{C,j} + \pi_{4,t} x_{M,j} && (16)\\
\ln a(x_j) &= \alpha_0 + \alpha_1 x_{C,j} + \alpha_2 x_{M,j} + \alpha_3 x_{C,j}^2 + \alpha_4 x_{M,j}^2 && (17)\\
b_{kt}(x_{k,j}) &= \beta_{k,0} + \beta_{k,1,t} x_{k,j} \; k \in \{C, M\}. && (18)
\end{aligned}
$$

No interaction term between cognitive and motor task indices ($x_{C,j}$ and $x_{M,j}$) is included in Equations (16) and (17) so that the logwage can be easily decomposed into cognitive skills/tasks and motor skills/tasks in Section 5.4. However, including the interaction term in these equations does not change the empirical results.

The changes in the time-varying parameters in the price function $\ln p_t(x)$ are approximated by a piecewise-linear function. For example, the parameter $\pi_{t,0}$ is given by

$$\pi_{0,t} = \pi_0 + \pi_{0,79} t_{79} + \pi_{0,89} t_{89}, \tag{19}$$

where $t_{79} = (t - 1979)$, $t_{89} = (t - 1989) \cdot I(t \geq 1989)$, and $I(\cdot)$ is an indicator function that takes the value one if the condition in the parenthesis is satisfied and zero otherwise. The prices of all the intermediate outputs in 1979 are normalized to be one, which implies $\pi_{l,t=1979} = 0$ for $l = \{0, 1, \ldots, 4\}$ in Equation (16). This normalization allows identification of the occupation-specific intercept $\ln a_j$.

In the returns to skills equation (18), the coefficient for the task index can change over time, reflecting technological change. Motivated by the argument in Section 2.1.2 and the restrictions expressed in Equations (9) and (10), this parameter is specified as

$$\beta_{k,1,t} = \beta_{k,1} + \beta_{k,2}(t - 1979) \; k \in \{C, M\} \tag{20}$$

The restrictions in Equations (9) imply that $\beta_{C,2} > 0$ and $\beta_{M,2} < 0$. I normalize the returns to task-specific skills to one for an average task in 1979. Specifically, I impose the restriction that

$$\beta_{k,0} + \beta_{k,1} \bar{x}_{k,t=79} = 1 \; k \in \{C, M\}, \tag{21}$$

where $\bar{x}_{k,t=79}$ is the mean task complexity index of the PSID sample in 1979.

3.1.2 Skill Equation

The initial skill endowment is specified as

$$s_{k,i\underline{t}_i} = \exp(\gamma'_{k,0} d_i + \sigma_{k,i} + \varepsilon_{k,i\underline{t}_i}), \tag{22}$$

where $\underline{t}_i$ is the year when worker i enters the labor market, d_i is a vector of time-invariant (after labor market entry) worker characteristics that includes education, race, and gender[4], $\sigma_{k,i}$ is a

[4]Many papers on gender wage gap do not include a gender dummy in their wage equations, but doing so raises concern over omitted variable bias. Fortin (2008) extensively discusses this issue and supports the use of a gender dummy variable.

time-invariant unobserved component, and $\varepsilon_{k,i\underline{t}_i}$ is a time-varying unobserved component.

The skill growth equations (11) and (12) as well as Equation (22) imply that log skills in year t can be written as

$$\ln s_{k,it} = \gamma'_{k,0}d_i + \gamma_{k,1}e_{it} + \gamma_{k,2}e_{it}^2 + \gamma_{k,3}\sum_{\tau=\underline{t}_i}^{t-1} x_{k,ij\tau} + \sigma_{k,i} + \varepsilon_{k,it} \; k \in \{C,M\} \tag{23}$$

$$\ln s_{G,it} = \gamma'_{G,0}d_i + \gamma_{G,1}e_{it} + \gamma_{G,2}e_{it}^2 + \varepsilon_{G,it}. \tag{24}$$

Note that none of the parameters in the skill equations are time-varying. The fourth term of Equation (23) is the sum of the task indices of the previous jobs. This term captures the skill component acquired through learning-by-doing. Because this term exhibits a heterogeneous skill growth profile depending on occupational experiences, this specification allows for heterogeneous wage growth profiles.

The time-invariant components of the unobserved task-specific skills are specified as

$$\sigma_{k,i} = h_k\left(\{x_{ijt}\}_{t=\underline{t}_i}^{T_i}\right) + \theta_{k,i} \tag{25}$$

$$E(\theta_{k,i}|\{x_{ijt}\}_{t=\underline{t}_i}^{T_i}, e_{it}, d_i) = 0, \tag{26}$$

where $k \in \{C,M\}$, the argument in the function h_k is the full history of task indices from the first period to the last period observed in the data, and $\bar{x}_i$ is the time-average of the task indices over years observed in the data, i.e.,

$$\bar{x}_{k,i} = (T_i - \underline{t}_i + 1)^{-1}\sum_{\tau=\underline{t}_i}^{T_i} x_{k,ij\tau}. \tag{27}$$

The function h_k is parametrized as

$$h_k\left(\{x_{ijt}\}_{t=\underline{t}_i}^{T_i}\right) = \gamma_{k,4}\bar{x}_{k,i} + \gamma_{k,5}\bar{x}_{k,i}^2. \tag{28}$$

Motivated by the worker assignment mechanism outlined in Section 2.2, I include the variable $\bar{x}_i$ to control for the time-invariant unobserved skills $\sigma_{k,i}$. In the model, workers who have high time-invariant unobserved cognitive (motor) skills choose complex cognitive (motor) tasks persistently throughout their careers, which is captured by the time-average of the task indices $\bar{x}_{k,i}$. This approach is known as the correlated random effect approach (see Wooldridge (2009) and Blundell and Powell (2003) and Altonji and Matzkin (2005) for related methods.) The basic idea of the correlated random effect approach is to put restrictions on the conditional distribution of unobserved heterogeneity given the entire history of the covariates, which is $\bar{x}_{k,i}$ in this paper.

Similarly, the time-varying components of the unobserved skills satisfy the conditional mean independence assumption,

$$E(\varepsilon_{k,it}|\{x_{ijt}\}_{t=\underline{t}_i}^{T_i}, e_{it}, d_i) = E(\varepsilon_{k,it}|x_{ijt}, \bar{x}_i, e_{it}, d_i) \tag{29}$$

$$= 0. \tag{30}$$

3.2 Estimation by Nonlinear Least Squares

Using the specifications laid out in Section 3.1, I estimate the wage equation (8) and the skill equations (23) and (24) conditional on tasks. Not modeling occupational choice may raise endogeneity bias concerns. I defer a discussion of this issue until the next subsection. While structural estimation of the full model including occupation choice seems feasible[5], I do not pursue this direction over concerns of possible misspecification in the occupational choice model. Notice that the true occupational decisions may be more complicated than those characterized by the first order condition (15) if preference for tasks and uncertainty affect workers' decisions. The approach taken in this paper is robust to different models of occupational choice. Another benefit is that a less complex approach makes the estimation process and the driving force of the result relatively transparent.

Estimation involves a single-equation, because the wage equation incorporates the skill equations. Notice that neither demeaning nor first-differencing eliminates the time-invariant unobserved skills $\sigma_{k,i}$, because they are interacted with the task index x. The quasi-differencing method used by Gibbons, Katz, Lemieux, and Parent (2005) does not work either, because there is more than one time-invariant unobserved variable.

I estimate the model using a nonlinear least square estimator with correlated random effects. Given the conditional mean independence assumptions (29) and (26), the parameters are consistently identified. I calculate the standard errors by clustering at the individual level so that they are robust to heteroskedasticity and serial correlation.

3.3 Identification

3.3.1 Addressing Endogeneity Bias

The key identifying assumption is that occupational choice is random, conditional on the present value and entire history of the observed variables. This assumption does not hold if unobserved skills affecting workers' occupational choices exist. The basic idea of this paper's approach is that a

[5]See Yamaguchi (2012) for the structural estimation of a similar model that incorporates occupational choice decisions.

broader set of control variables sufficiently captures heterogeneity amongst workers and effectively takes care of the endogeneity bias. I consider two components of unobserved skills: time-invariant and time-varying. The time-invariant component is largely controlled by the correlated random effect approach detailed in Section 3.1.2. The control variable, the time-average of task indices $\bar{x}$, is motivated by the theoretical argument over worker assignment to tasks in Section 2.2. I also argue that the time-varying component of unobserved skills is controlled for by including a rich set of observed variables in the skill equation (23). In particular, the key variable is the sum of the task indices of the past jobs. This variable can be viewed as task-complexity adjusted experience, and allows for heterogeneous skill growth and wage profiles across workers depending on their observed occupational histories.

3.3.2 What Features of the Data Identify Skill Endowments?

The model allows me to estimate how an observed characteristic such as education can be associated with three different types of skills: cognitive, motor, and general. This subsection explains which features of the data allow me to identify the relationship between a single variable and three different types of skills.

To simplify the discussion, assume all the parameters are time-invariant and set $t = 1979$. The conditional mean logwage is given by

$$\begin{aligned}
& E(\ln w_{ijt} | \{x_{ijt}\}_{t=\underline{t}_i}^{T_i}, e_{it}, d_i) \\
= \; & \alpha_0 + \alpha_1 x_{C,ijt} + \alpha_2 x_{M,ijt} + \alpha_3 x_{C,ijt}^2 + \alpha_4 x_{M,ijt}^2 \\
& (\beta_{C,0} + \beta_{C,1} x_{C,j}) \cdot (\gamma'_{C,0} d_i + \gamma_{C,1} e_{it} + \gamma_{C,2} e_{it}^2 + \gamma_{C,3} \sum_{\tau=\underline{t}_i}^{t-1} x_{C,ij\tau} + \gamma_{C,4} \bar{x}_{C,i}) \\
& (\beta_{M,0} + \beta_{M,1} x_{C,j}) \cdot ((\gamma'_{M,0} d_i + \gamma_{M,1} e_{it} + \gamma_{M,2} e_{it}^2 + \gamma_{M,3} \sum_{\tau=\underline{t}_i}^{t-1} x_{M,ij\tau} + \gamma_{M,4} \bar{x}_{M,i}) \\
& + (\gamma'_{G,0} d_i + \gamma_{G,1} e_{it} + \gamma_{G,2} e_{it}^2),
\end{aligned} \tag{31}$$

where the second line corresponds to $\ln a_j$ (the intercept of the production function), the third line is the product of the cognitive skill endowment and its return, the fourth line is the product of the motor skill endowment and its return, and the fifth line is general skills. The output price function is dropped due to the normalization that the price in 1979 is one, meaning that $\ln P_{jt=1979} = 0$.

Identification of the parameters for the occupation-specific intercept and general skills (α, and γ_G) is straightforward. The parameters in the function for returns to skills β and those in the skill equations γ are identified up to scale. In this argument, let us normalize the parameters by setting $\gamma_{k,4} = 1$ for $k \in \{C, M\}$. This normalization is different from the one used in the estimation, but it is useful for the discussion here. This normalization implies that task-specific skills are measured by the time-average of tasks $\bar{x}$. Given the normalization, the constant term $\beta_{k,0}$ for the returns to skills is identified by the variation of the time-average of the task complexity index, because

$$\beta_{k,0} = \partial E(\ln w_{ijt} | \{x_{ijt}\}_{t=\underline{t}_i}^{T_i}, e_{it}, d_i) / \partial \bar{x}_{k,i}.$$

Note that, in order to identify the returns to skills parameters, it is important to restrict that the cognitive (motor) task index be included only in the cognitive (motor) skill function. At least one of the variables in the skill functions must not appear in the functions for other types of skills.

The slope parameter $\beta_{k,1}$ for returns to skills is identified by the variation of the product of the current task complexity index and the time-average of the task complex index. This can be algebraically restated as $\beta_{k,1} = \partial^2 E(\ln w_{ijt} | \{x_{ijt}\}_{t=\underline{t}_i}^{T_i}, e_{it}, d_i) / [\partial x_{k,j} \partial \bar{x}_{k,i}]$ for $k \in \{C, M\}$.

The parameters for the task-specific skill production function (γ_k and δ_k for $k \in \{C, M\}$) are identified using the variation of the product of the current task complexity index and the worker characteristics d. To see this, observe that $\beta_{k,1}\gamma_{k,0} = \partial^2 E(\ln w_{ijt} | \{x_{ijt}\}_{t=\underline{t}_i}^{T_i}, e_i, d_i) / [\partial x_{k,j} \partial d_i]$, which implies that the parameter $\gamma_{k,0}$ is given by

$$\gamma_{k,0} = \{\partial^2 E(\ln w_{ijt} | \{x_{ijt}\}_{t=\underline{t}_i}^{T_i}, e_i, d_i) / [\partial x_{k,j} \partial d_i]\} / \{\partial^2 E(\ln w_{ijt} | \{x_{ijt}\}_{t=\underline{t}_i}^{T_i}, e_i, d_i) / [\partial x_{k,j} \partial \bar{x}_{k,i}]\} \tag{32}$$

This result indicates that the interaction term of tasks and worker characteristics is the main identification source for how a single worker characteristic can be associated with the task-specific skills.

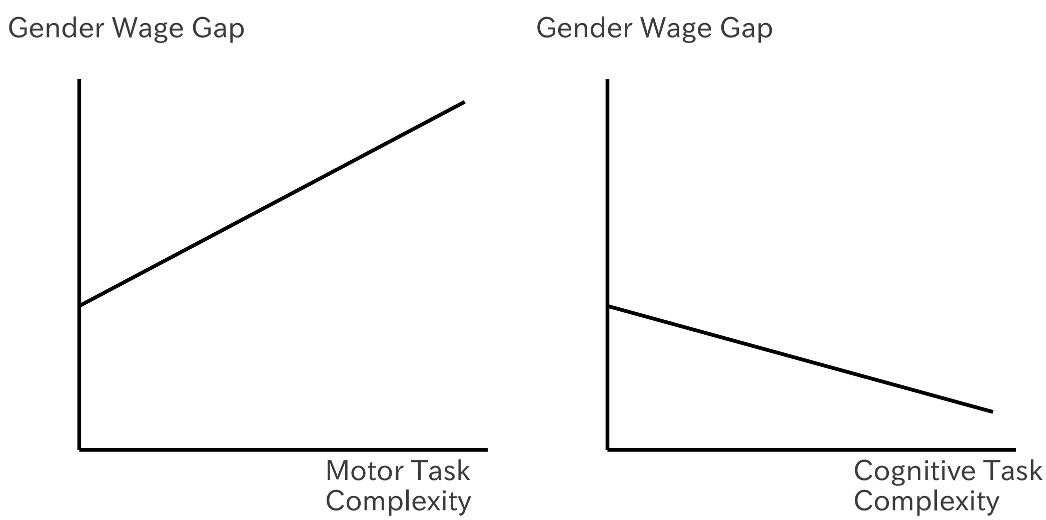


Figure 3: Identification of Skill Endowments
Note: Gender wage gap = average men's wage - average women's wage.

To understand the intuition behind this result, consider the identification of task-specific skill differences between men and women. Figure 3 shows how the gender wage gap varies across tasks. Workers are randomly assigned to different tasks given observed characteristics and the full history of occupational choices in Figure 3, which is consistent with the identification assumptions discussed in the previous section. In the left panel in Figure 3, the gender wage gap increases with motor task complexity, which implies that men have more motor skills than women. Remember that task-specific skills are more intensely used when tasks are complex. Men's advantage in motor skills is more strongly pronounced in complex motor tasks, which results in a larger gender wage gap. In contrast, in the right panel in Figure 3, the gender wage gap decreases with cognitive task complexity, which implies that women have more cognitive skills than men. Women's advantage in cognitive skills is amplified when cognitive tasks are complex, given the production function (3). In this way, the varying gender wage gap across tasks is the driving force behind the identification of gender differences in task-specific skills. The remaining gender wage gap not explained by task-specific skills is reflected in general skills. General skills can be regarded as residuals in the sense that they include all factors other than task-specific skills, and may include discriminatory factors. The issue of possible gender discrimination is extensively discussed in Section 6.2.1.

4 Data

4.1 Dictionary of Occupational Titles

The DOT contains information on 12,099 occupations defined by the tasks performed by workers in those individual occupations. The U.S. Department of Labor compiled the data to provide standardized occupational information for an employment service matching job applicants with job openings. The information included in the DOT is based on on-site observation of jobs as they are performed in diverse business establishments and, for jobs that are difficult to observe, on information obtained from professional and trade associations.[6] On this basis, in the fourth edition of the DOT, analysts rate each occupation with respect to about 50 characteristics including aptitudes, temperaments, and interests necessary for adequate performance; the training time necessary to prepare for an occupation; the physical demands of the occupation; and the working conditions under which work in the occupation typically occurs.

[6]One might be concerned that task complexity cannot be correctly measured by observing jobs performed, because what analysts observe is a realized combination of job tasks and worker skills in equilibrium: even when the task is simple, an analyst might consider it complex if the worker is skilled. This confusion should be at least partially avoided because job information is obtained from other sources as well (e.g. interviewing incumbents and supervisors.) It is also worth noting that the DOT explicitly states that each occupation is defined on the basis of the tasks performed. See Miller, Treiman, Cain, and Roos (1980) for a critical review of the DOT.

Many characteristics are measured by a multi-point scale and have detailed definitions. For example, the variable DATA measures the complexity of tasks in relation to information, knowledge, and conceptions by integers from 0 to 6. Tasks at the lowest level of complexity involve judging the readily observable characteristics of data. Examples include sorting hats according to color and size as specified, comparing invoices of incoming articles with the actual number and weights of articles, and so on. Tasks at the intermediate level of task complexity involve compiling information. Examples include summarizing details of transactions, collecting, classifying, and recording data, and receiving customer complaints to record and file them for future processing. Tasks at the highest level of complexity involve integrating analysis of data to discover facts or developing knowledge concepts of interpretations. Examples include formulating hypotheses and experimental designs, writing critical reviews of art for publication, and conducting research. Other tasks such as operating machines or equipment are also evaluated in a similar manner. Some tasks are measured by a binary variable that takes the value one if the occupation involves the task and zero otherwise.

To facilitate interpretation of the data, I summarize the detailed information in the DOT by constructing a low dimensional vector of occupational tasks using Multiple Correspondence Analysis (MCA).[7] MCA is similar to the frequently used Principal Component Analysis (PCA), but can deal with discrete variables. Because ratings in the DOT are ordinal, not cardinal, MCA is more suitable than PCA, which assumes cardinal variables.

In light of the job analysis literature, as well as previous economics papers using the DOT, this paper assumes that tasks are broadly categorized into either cognitive or motor tasks. By examining the textual definitions of the DOT variables, I assume which variables measure the complexity of which task type. The following choices seem reasonable, and I confirm that the constructed task complexity index is robust to the choice of the DOT variables. The DOT variables which measure cognitive task complexity consist of 2 worker function variables (DATA and PEOPLE), 3 General Educational Development variables for reasoning, mathematical, and language, 3 aptitude variables (intelligence, verbal, and numerical), and 2 temperament variables (influencing and dealing with people). Motor task complexity is measured by 1 worker function variable (THINGS) and 7 aptitude variables for motor coordination, manual dexterity, finger dexterity, eye-hand-foot coordination, spatial, form perception, and color discrimination.

I conduct the MCA using the April 1971 CPS augmented by the fourth edition of the DOT. This augmented CPS file contains the 1970 census occupation code, the DOT occupation code, and the DOT variables. The resulting indices at the individual level are aggregated into the level of the 1970 census occupation by taking the mean for each of the 1970 census 3-digit occupations using

[7] See supplementary appendix for an outline of the algorithm. Greenacre (2007) is a good introduction to the method.

the sampling weights so that they can be merged with the PSID. Details of the results of MCA are available in Appendix A.1.

Figure 4 plots the levels of cognitive and motor task complexity for each 1-digit occupation. Cognitive tasks of professionals and managers are the most complex, and are followed by sales, clerical, and craft workers. Cognitive task complexity is lowest for service workers, operators, and laborers. The complexity of motor tasks of craft workers are highest among all occupations, and are followed by professionals, clerical workers, and operators. A large degree of heterogeneity in motor task complexity exists for professionals and clerical workers. Among professional occupations, physicians' and engineers' motor tasks are highly complex, reflecting the finger and manual dexterity required by these jobs. Among clerical occupations, motor tasks of bank tellers, typists, and related workers are complex, reflecting the extensive use of hands and fingers when typing, counting bills, and operating office machines, for example. Service and sales workers and laborers are involved in simple motor tasks. Managers report the lowest level of motor task complexity.

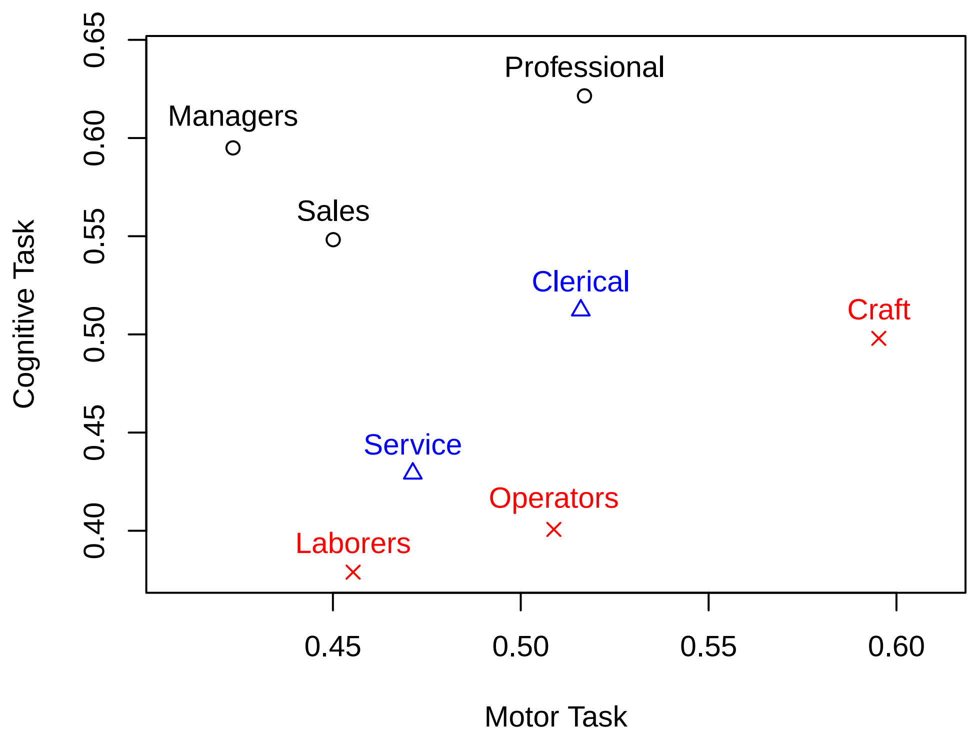


Figure 4: Occupations in the Task Space

Source: The 1971 Current Population Survey augmented by the fourth edition of the Dictionary of Occupational Titles.

Note: The task complexity indexes are normalized so that the mean and standard deviation for the 1971 CPS are 0.5 and 0.1, respectively.

4.2 PSID

The PSID is a nationally representative household panel survey that began in 1968. To study the evolutions of tasks, skills, and wages, I draw a sample of household heads[8] and wives who worked full-time (1,500 hours a year or more) from 1979 to 1996. I select these survey years, because all the variables are available, and more importantly, wage inequality increased and the gender wage gap narrowed rapidly during this period. I restrict that the sample individuals be between 18 and 65 years old. The sample does not include self-employed workers, because their wage determination mechanisms may be significantly different from those of the employed. The sample restrictions are comparable to those used in the literature such as in Blau and Kahn (1997, 2006).

Hourly wages are calculated by dividing labor income by hours of work, and are deflated by the 1983 PCE Index. Note that hours of work and labor income are for the previous year of the survey, just like the CPS. Years of experience is not available in every survey year. For years when experience is not recorded, it is imputed using hours of work. Occupations in the PSID are coded using the 1970 census scheme. This coding is used to merge the PSID sample with the task indices from DOT. Details of variable definitions are outlined in Appendix A.2.

Several papers point out that occupations are often misclassified, and address this issue by using other employment related variables rather than occupation.[9]One possible way to correct these errors is to assume that all occupation changes within the same employer are false. Neal (1999), Pavan (2011), and Yamaguchi (2010) take this approach. However, this edit is likely to result in a downward bias in the mean task complexity, because many occupation code changes within the same employer are promotions to managers. If an occupation is often mis-coded to other occupations that are similar in terms of job characteristics, the classification error may lead to only a small measurement error in the task complexity index. In order to avoid worsening the classification error bias, I do not overwrite the recorded occupation code. However, when an occupation code is missing, but no employer or position change is reported, I interpolate it using the last year's (or the next year's if not available) occupation code.

4.2.1 Summary Statistics of the Sample

Table 1 reports the means of selected labor market outcomes for men and women for the years 1979, 1989, and 1996. The mean log hourly wage exhibits well-known patterns. Men's wages are higher than women's, but the gender wage gap has been steadily decreasing, although at a slower

[8]In the PSID, a man is by construction a household head for a married or cohabiting couple. A women can be a household head in other types of households such as a single household.

[9]The classification error in occupations is a common problem across surveys. Kambourov and Manovskii (2008, 2009) report the issue for PSID. Neal (1999), Pavan (2011), Sullivan (2009), and Yamaguchi (2010) find evidence for occupation classification error in the National Longitudinal Survey of Youth 1979.

pace in the 1990's. Both men and women became better educated over time at about the same rate. The gender gap in work experience has significantly decreased from 6.5 years to 4 years.

Changes in task complexity indices are graphically presented in Figure 5. In 1979, men's cognitive tasks are more complex than women's. While men's cognitive task complexity indices remained roughly constant during the period, women's cognitive task complexity indices grew over time. Women surpassed men in cognitive task complexity in early 80's, and the difference between sexes has increased since then. Men's motor tasks are more complex than women's throughout the period. Over time, both men and women shifted to occupations that involve less complex motor tasks, but the pace of the shift for women is faster. These patterns exhibit the shift from motor to cognitive tasks, and are consistent with the nuanced view of technological change proposed by Autor, Levy, and Murnane (2003).

Like other panel data, sample attrition is a possible concern for the PSID. Blau and Kahn (1997) extensively examine this issue by comparing various statistics from the PSID and those from the CPS using variables including hourly wages, experience, occupation, industry, education, and other demographic variables. Their exercise is a valid assessment for an attrition bias, because the CPS is a representative data set and does not have a sample attrition problem. Blau and Kahn (1997) find no evidence of an attrition bias when weights provided in the PSID are correctly applied.

Table 1: Mean of Selected Labor Market Outcomes in PSID

	1979		1989		1996	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Logwage						
Men	2.394	0.011	2.362	0.013	2.393	0.016
Women	1.910	0.012	2.007	0.013	2.096	0.015
Cognitive Task Index						
Men	0.518	0.002	0.525	0.002	0.529	0.002
Women	0.513	0.002	0.530	0.002	0.542	0.002
Motor Task Index						
Men	0.511	0.002	0.506	0.002	0.502	0.002
Women	0.499	0.002	0.491	0.002	0.483	0.002
Years of Education						
Men	12.610	0.060	13.467	0.055	13.632	0.059
Women	12.433	0.059	13.232	0.052	13.506	0.058
Years of Full-time Experience						
Men	18.243	0.274	18.344	0.250	19.040	0.282
Women	11.698	0.252	13.645	0.214	15.088	0.245
			Number of Observations			
Men	1927		1874		1355	
Women	1359		1735		1305	

Source: PSID 1979-1996.
Note: Wages are deflated by 1983 PCE Index. The sample includes household heads and wives who worked full-time (1,500 hours a year or more). Self-employed workers and those who are younger than 18 or older than 65 are excluded from the sample.

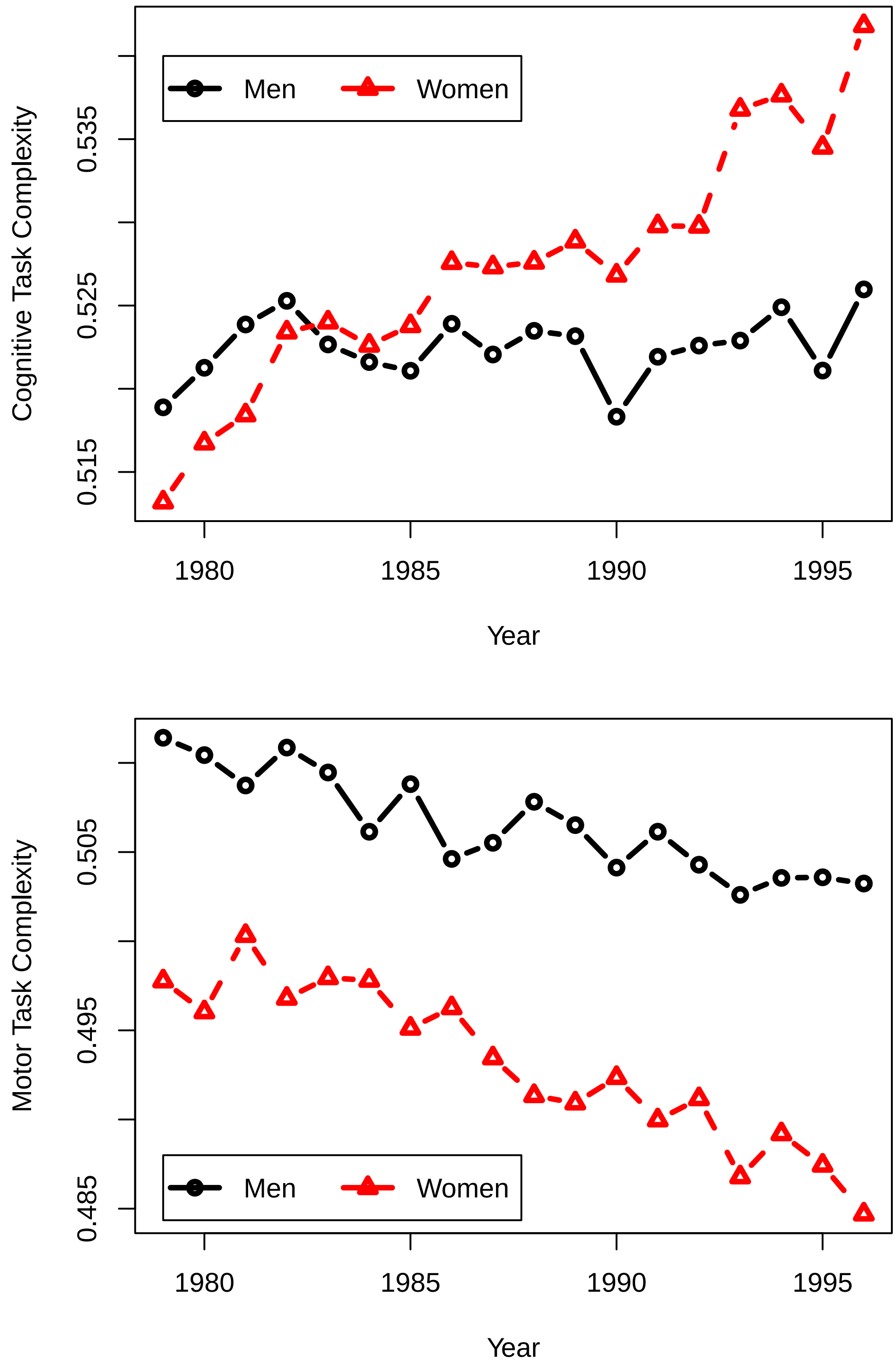


Figure 5: Task Complexity Indexes
Source: PSID 1979-1996 and DOT
Note: The task complexity indexes are normalized so that the mean is 0.5 and the standard deviation is 0.1 for the working population in the 1971 CPS.

5 Empirical Results

Before I present estimation results for the structural model, I provide empirical evidence that occupational tasks are the key to understanding how technological changes affected the wages of men and women differently in Section 5.1. The evidence is based on reduced-form estimation and descriptive statistics, and thus, the main features of the data are uncovered in a simple way. This simple empirical analysis also helps us understand the driving forces behind the more complicated structural estimation results.

5.1 U.S. Labor Market Facts

While I use the PSID to estimate the structural model, I use the 1980 and 1997 CPS for the descriptive analysis in this section because of its representativeness and large sample size. The CPS survey years were selected to match the sample period of the PSID data used. The CPS sample consists of civilian male and female non self-employed full-time workers in the non-agricultural sector between 18 and 65 years old. Full-time work is defined as 1,500 hours of work per year or more. Hourly wages are deflated by the 1983 PCE deflator. I exclude wages less than \$1 per hour and more than \$250 per hour from the sample. Note that wages and hours reported in the CPS, as well as in the PSID, are the previous years'. Thus, the 1980 and 1997 surveys report wages and hours worked in 1979 and 1996, respectively. The sample restrictions imposed in this paper are comparable to Blau and Kahn (1997).

Based on evidence offered by organizational theorists and computer scientists, Autor, Levy, and Murnane (2003) develop a theory that explains how automation and computerization has reshaped the demand for occupational tasks by replacing routine tasks. While my task complexity measures are not identical to their routine and nonroutine measures, complex motor tasks are largely routine manual tasks, and complex cognitive tasks are largely nonroutine analytical/interactive tasks in their measures. Autor, Levy, and Murnane (2003) argue that strong complementarities exist between computers and complex cognitive tasks, while substantial substitution exists between computers and complex motor tasks. Their theory predicts that occupations that involve complex motor tasks were harmed by computerization, while those involving complex cognitive tasks have benefited. In Figure 4, occupations in the bottom right region (craft workers, operators, and laborers) should have suffered the most from computerization, because they are relatively motor task intensive occupations. Occupations in the top left region (professionals, managers, and sales) should have benefited from computerization, because they are relatively cognitive task intensive occupations. Occupations near the diagonal (clerical and service) should not have been strongly affected by computerization.

To examine if this prediction is consistent with the data, I calculate the changes in the composition-adjusted[10] mean wages from 1980 to 1997 for each 1-digit occupation[11] in Column (1) of Table 2. Operators, laborers, and craft workers suffered large wage losses of 18% during this period. Clerical and service workers had modest wage losses from 8% to 5%. Wages of sales workers and managers remained almost constant, and professionals enjoyed 6% wage growth. Note that wage growth during this period did not monotonically change with the skill level of occupations. For example, craft workers experienced a large wage loss of 18%, but their average wage in 1979 is the third highest among 1-digit occupations, and immediately follows those of professionals and managers (see Column (2) of Table 2). Clerical and service workers had modest wage losses, and their wages are the first and second lowest, respectively. The wage growth patterns presented here cannot be explained by a simple model in which workers possess a single dimensional skill and its return increases over time. Instead, they are consistent with the labor market polarization story presented in Goos and Manning (2007) and Autor, Katz, and Kearney (2008).

These wage changes have different consequences for men and women. The occupations which suffered the largest wage losses are male-dominated. In 1979, only 10% of laborers and 5% of craft workers were female. Female operators were also uncommon; their employment share was 28%. In contrast to men, the majority of women were in occupations that did not experience a significant wage loss. Also, the share of female workers has increased significantly in professionals, managers, and sales workers: from 41% to 53% in professionals, from 24% to 40% in managers, and from 32% to 41% in sales workers. These occupations also did not experience a wage loss during the time period.

The descriptive analysis presented in this subsection strongly suggests that occupational tasks are the key to understanding the driving forces behind the narrowing gender wage gap. In the following, I provide more detailed empirical analysis using the structural model.

[10]See the note of Table 2 for the method to adjust worker compositions.

[11]The 1950 census occupation classification is used. Although the 1990 coding scheme is also available, the 1950 coding scheme is much more similar to the 1970 coding scheme, which is used in the PSID.

Table 2: Composition-Adjusted Wages and Share of Female Workers

	Wage Growth Rate in %	Logwage in 79	% Women	Change in % Women
Operatives	−18	1.99	28	−2
Laborer	−18	1.93	10	4
Craftsmen	−18	2.25	5	2
Service	−8	1.70	49	3
Clerical	−6	1.89	77	0
Managers	−1	2.36	24	16
Sales	−1	2.11	32	9
Professional	6	2.31	41	12

Source: CPS 1980 and 1997. The reported wage in the CPS is the last year's. The sample includes civilian male and female non self-employed full-time ($\geq$ 1500 hours/year) workers in the non-agricultural sector between 18 and 65 years old. The wages are deflated by the 1983 PCE deflator. Wages less than \$1 per hour and more than \$250 per hour are dropped from the sample.

Note: The conditional mean wage function for 1997 is estimated by a nonparametric regression spline method using the CPS. Covariates include age, gender, five education levels (high school dropouts, high school graduates, some college, college graduates, and advanced degree), and 1-digit occupations. The composition-adjusted wage is calculated by applying the estimated wage equation to the 1980 CPS data. See Ma, Racine, and Yang (2011) for details of the theory and Racine and Nie (2011) for an implementation in the R language.

5.2 Estimates of Structural Parameters

Many structural parameters are included in the model to control for worker heterogeneity. I restrict the discussion to the parameter estimates concerned with unobserved task-specific skills. Tables 8 to 10 in Appendix B report all parameter estimates with standard errors clustered at the individual level.

The results in Table 8 indicate that the variables added to control for the time-invariant component of unobserved task-specific skills work well. For cognitive skills, the time-average of the cognitive task indices and its square are highly significant. For motor skills, these variables are jointly, but not individually, significant. Moreover, the sum of the task indices for past jobs is significant for both cognitive and motor skills. It is included to control for time-varying components of task-specific skills. The significance implies that the model successfully accounts for heterogeneous skill and wage growth profiles.

5.3 Changes in Skill Endowments and Their Returns

5.3.1 Skill Endowments

Figures 6 and 7 show the average skill endowments for men and women from 1979 to 1996. Note that skills are identified up to a linear transformation, because there exists no natural metric for skills. This implies that the skill level reported in the graphs does not have any meaning on its own, while the difference between the two series does. Returns to skills are normalized to one for the average tasks in the PSID sample from 1979. Hence, the skill difference between men and women is translated to the logwage difference when their task complexity indices are at the 1979 average.

The top panel of Figure 6 shows the mean cognitive skill endowments. Interestingly, women have more cognitive skills than men. This finding is very different from that of previous papers, where men always possess more skills than women. However, this is not necessarily the case if skills are multidimensional. Evaluated at the 1979 average tasks, women's cognitive skill advantage over men leads to women earning about 4% more, holding all else equal. Women's mean cognitive skills grow over time due to better education and more experience in complex cognitive tasks, while men's decrease, resulting in a greater advantage for women in cognitive skills. Again, evaluated at the 1979 average tasks, women's cognitive skill advantage over men in 1996 results in women earning about 7% more, holding all else equal.

The bottom panel of Figure 6 reports mean motor skill endowments. Men have more motor skills than women throughout the period, and average motor skills decrease over time for both men and women. The motor skill gap between men and women leads to a sizable gender wage gap. Evaluated at the 1979 average tasks, men's motor skill advantage over women leads to men earning about 23% more, holding all else equal. The gender gap in motor skills slightly increased over time, and accounts for a 25% wage difference when evaluated at the 1979 average tasks.

Intrinsic gender differences in cognitive skills is captured by the female dummy in the cognitive skill function, and the estimate is 0.042 in Table 8. Almost all of the motor skill gap is due to intrinsic gender differences. The estimate in Table 8 indicates that women have less motor skills than comparable men, accounting for a 25% lower wage at the 1979 average tasks. No gender difference in cognitive skills and a significant gender difference in motor skills are often assumed (see Galor and Weil (1996) and Rendall (2010)), but not estimated. Black and Spitz-Oener (2010) and Bacolod and Blum (2010) measure the gender difference in these skills by the task indices alone, but do not find a difference large enough to lead to a significant wage gap. This paper takes one step further and estimates the gender gap in skill endowments by the identification approach outlined in Section 3.3.2. I find the estimates are intuitive, and are similar to the speculations by Galor and Weil (1996) and Rendall (2010).

Finally, Figure 7 reports mean general skill endowments. Note that general skills account for everything aside from the task-specific skills. The gender difference in this component may include time-invariant gender discrimination effects. A large gender gap in general skills exists throughout the period. In 1979, the general skill gap accounts for a 28% wage difference. Notice that the intrinsic gender difference in general skills is not the whole difference. The estimate in Table 8 indicates that the intrinsic general skill difference accounts for a 15% wage gap. The remaining difference is explained by differences in education and experience, in which men have advantage over women throughout the sample period (see Table 1). Since then, women have accumulated education and experience, and the general skill gap has narrowed between the sexes. In 1996, the general skill gap accounts for a 22% wage difference.

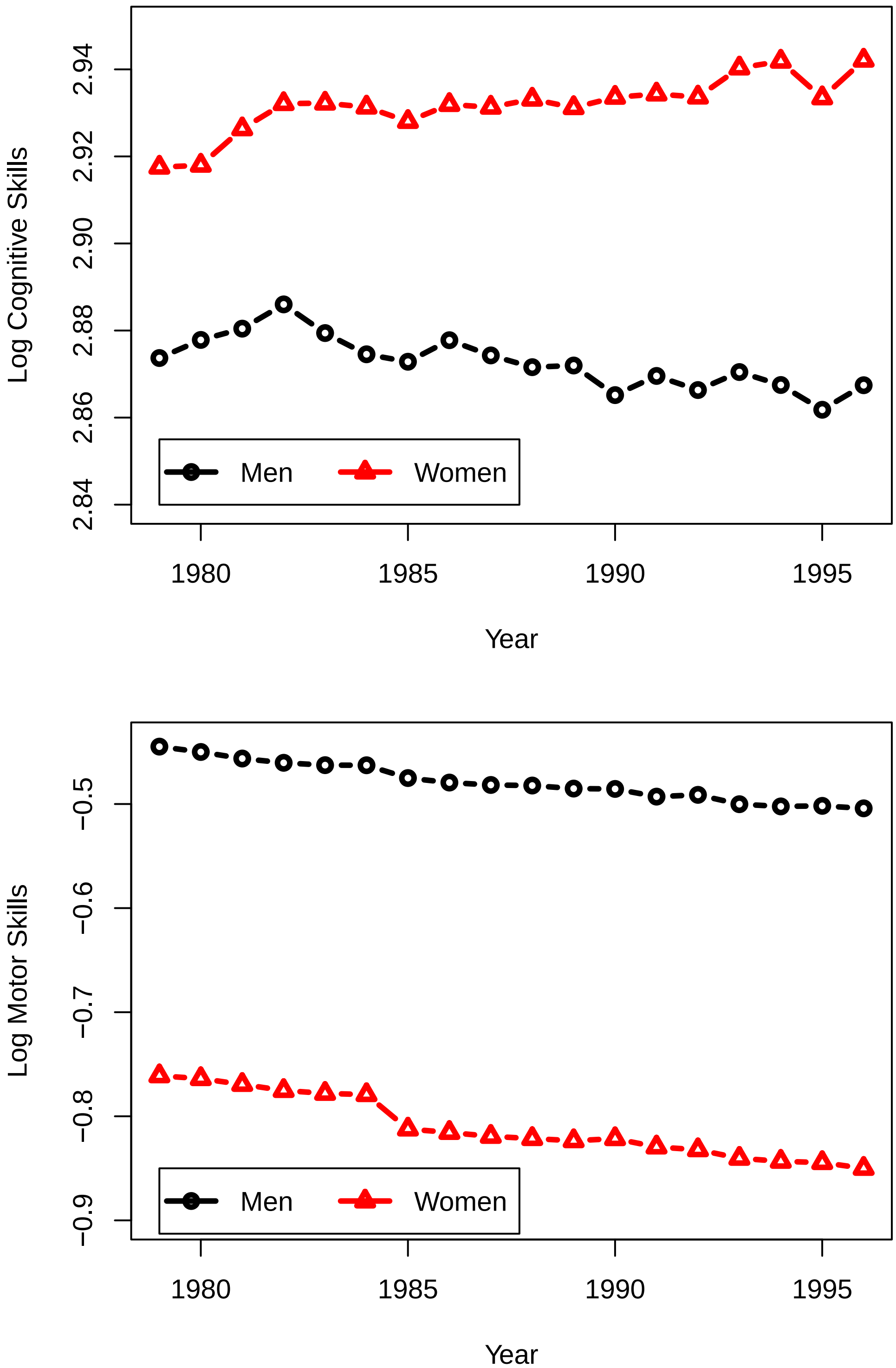


Figure 6: Task-Specific Skills

Note: Skills are identified up to a linear transformation. The returns to task-specific skills are normalized to one for the average task in 1979. Hence, the skill level does not have any meaning, but the skill difference is equivalent to the logwage difference when evaluated at the average task in 1979.

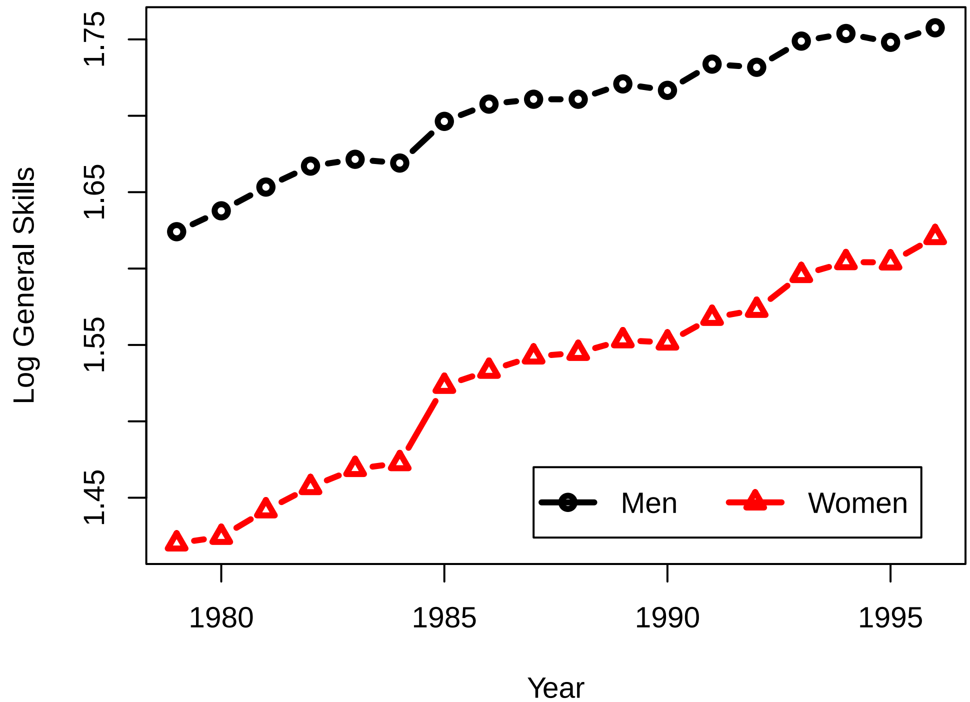


Figure 7: General Skills

5.3.2 Returns to Skills

In a Roy model, returns to skills vary across occupations. I calculate the average of the returns to skills over all workers in the sample, in order to provide a general picture of the evolution of returns to skills. The average returns to skills change over time for two reasons. First, workers undertake different tasks over time. If workers complete more and more complex tasks over time, the returns to skills increase over time, because complex tasks utilize their skills more intensely, resulting in a higher rate of return. Second, technological change directly affects the returns to skills, which is captured by changes in the parameters over time.

Figure 8 illustrates the estimated average returns to task-specific skills and the counterfactual average returns to skills over time. The counterfactual series are the returns that would prevail if technology remained at the 1979 level. They are calculated by holding the parameters at their 1979 levels. The top panel shows the returns to cognitive skills. They have increased by about 50% from 1979 to 1996. This increase was caused by both changes in tasks and changes in technology. If technology remained at its 1979 level, the average returns to cognitive skills would have increased by about 15% during the same period. The remaining 35% increase is due to technological change.

The bottom panel shows the average returns to motor skills. They dropped about 40% by 1996. However, if technology remained at its 1979 level, changes in motor task complexity would have caused only a 3% decrease in average returns. Thus, technological change lowered the average returns to motor skills by 37%.

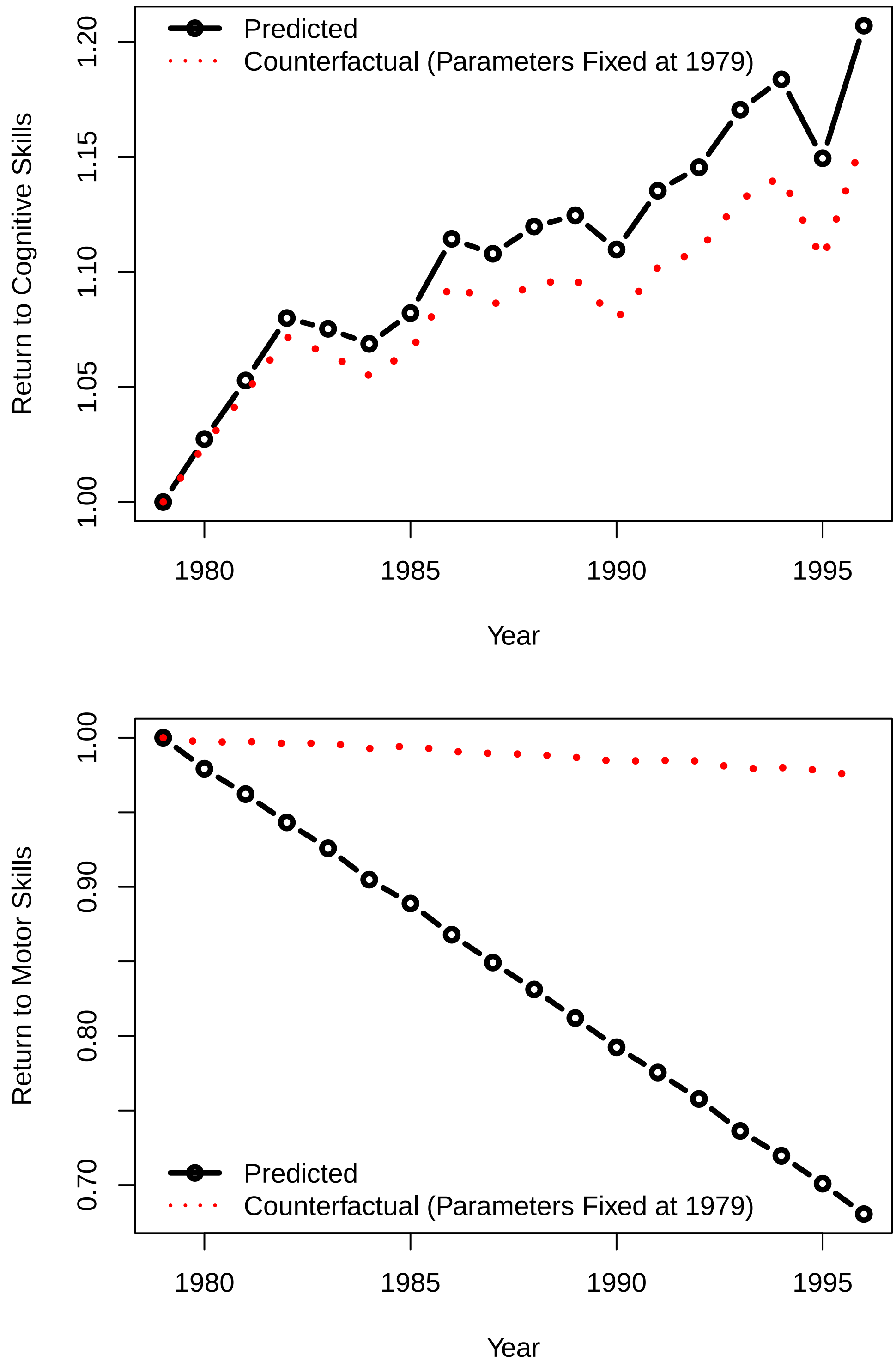


Figure 8: Returns to Task-Specific Skills

Note: The returns to task-specific skills are normalized to one for the average task in 1979.

5.4 Gender Wage Gap Decomposition

The results reported in the last subsection suggest that technological change reduced the gender wage gap. Women have more cognitive skills than men, and technological change raised the returns to these skills. Moreover, men have more motor skills than women, and technological change reduced returns to motor skills. This subsection assesses the degree to which each of these factors contributed to the narrowing gender wage gap.

I decompose changes in the gender wage gap using a method similar to Oaxaca-Blinder decomposition. Namely, changes in the gap are decomposed into (1) a composition effect that reflects the changes in tasks and skills and (2) a wage structure effect that reflects changes in parameter values. To derive the composition effect, I fix the parameters (i.e., the wage structure) at either the 1979 or 1996 level and calculate counterfactual wage gaps. The difference between the predicted wage gap of the estimated model and the composition effect identifies the wage structure effect. The decomposition exercise is carried out as follows. The gender wage gap in year t is given by

$$G_t(\Theta_t) \quad = \quad E(\ln w_{ijt} | i \in \text{Men},\ \Theta_t) - E(\ln w_{ijt} | i \in \text{Women},\ \Theta_t),$$

where Θ_t is a set of parameters in year t. The change in the gender wage gap from 1979 and 1996 is $\Delta G_{96-79} = G_{96}(\Theta_{96}) - G_{79}(\Theta_{79})$. The composition effect from 1979 and 1996 is given by

$$\Delta^C G_{96-79} \quad = \quad G_{96}(\Theta_{79}) - G_{79}(\Theta_{79}),$$

where the parameters are fixed at the 1979 level. The wage structure effect for the same period is given by $\Delta^{WS} G_{96-79} = \Delta G_{96-79} - \Delta^C G_{96-79}$. Also, notice that the logwage is linearly separable in the cognitive, motor, and general skill components. Hence, composition and wage structure effects are further decomposed into these different skill components.[12]

Table 3 contains the results for the gender logwage gap decomposition. From 1979 to 1996, the observed gender logwage gap has decreased by 0.187 log points. Out of 0.187 log points, changes in returns to skills and tasks (wage structure effects) account for 0.087-0.108, changes in skill endowments (composition effects) account for 0.067, and 0.011 remain unexplained by the model.

The most notable feature in the table is that the drop in returns to motor skills accounts for about half of the narrowing gender wage gap: 0.091-0.097 log points. This result is along the same line as the empirical analysis in Section 5.1 in which motor task intensive jobs suffer a large wage loss and are male-dominated. Devaluation of motor skills and tasks hurt men, but not women,

[12] Although it is technically feasible, I do not decompose into the effects of skills and those of tasks, because tasks are endogenously determined. Given the theoretical model, changes of tasks are interpreted as consequences of changes in skills.

resulting a narrowing gender wage gap. Changes in returns to cognitive skills and tasks affected the gender wage gap little, because the changes and women's advantages in cognitive skills are both modest.

Another important point is that women's growth in general and cognitive skills greatly narrowed the gender wage gap, although this is slightly offset by their loss of motor skills: women's growth in cognitive skills account for 0.019-0.034 lot points, and their growth in general skills account for 0.067 log points. During the period, women became more educated and experienced, and undertook more complex cognitive tasks than they used to, resulting in larger cognitive and general skill endowments. Again, this result is along the same line as the increased share of female workers in managerial and professional occupations as seen in Section 5.1.

Previous papers found that the wage structure effect is small or leads to a widening gender wage gap. Blau and Kahn (1997) found that changes in the wage structure should have widened the gender wage gap, because they measure skills by education and experience, and the returns to those variables have increased. Bacolod and Blum (2010) include occupational task variables in their wage regression and find that the wage structure effect narrowed the gender wage gap by only 0.02 log points. Their results indicate that the composition effect is the major factor in explaining the narrowing gender wage gap. In contrast, this paper finds a large wage structure effect: the decrease in the rate of return to motor skills has narrowed the gender wage gap significantly.

Table 3: Decomposition of Changes in Gender Logwage Gap

	Specification 1		Specification 2	
	Estimate	Std. Error	Estimate	Std. Error
Change in Returns to Skills and Tasks				
Cognitive	−0.012	0.006	0.004	0.005
Motor	−0.097	0.017	−0.091	0.017
Subtotal	−0.108	0.016	−0.087	0.016
Change in Skill Endowments				
Cognitive	−0.019	0.005	−0.034	0.005
Motor	0.018	0.013	0.012	0.012
General Skills	−0.067	0.013	−0.067	0.013
Subtotal	−0.068	0.004	−0.089	0.003
Unexplained	−0.011	0.016	−0.011	0.016
Overall Change in Gap	−0.187	0.028	−0.187	0.028

Note: Standard errors are calculated by the delta method and clustered at the individual level. In specification 1 (2), the quantity effect is calculated by fixing the parameters at the 1979 (1996) level.

I also conduct the same analysis for male high school graduates and male college graduates, in order to see the sources of the rise in male wage inequality as measured by the college logwage premium. The results are reported in a supplementary appendix. Not surprisingly, changes in returns to skills and tasks account for about two thirds of the rise in the college wage premium.[13] These results indicate that changes in the value of cognitive and motor skills account for a large part of both the narrowing gender wage gap and the rise in male wage inequality.

The method adopted in this paper has a limitation in that it cannot identify the sources of the changes in returns to skills and tasks without a strong restriction. Namely, it is not quite clear to what extent the technological change is responsible for the changes in the rate of return relative to supply and demand. This limitation is due to the fact that a change in output prices (P_{jt} in Equation (8)) and that in an occupation-specific intercept (α_{jt} in Equation (8)) cannot be separately identified in general. However, I provide evidence that the technological change caused a modest increase in returns to cognitive skills and a large decrease in returns to motor skills (see Figure 8). In addition, the evidence that the supply of cognitive skills increased while the supply of motor skills decreased (see Figure 6) implies a decrease in the value of cognitive skills and an increase in the value of motor skills if demand remains the same. Hence, demand factors, not supply factors, are responsible for the changes in the value of cognitive and motor skills. Among others, the nuanced view of skill biased technological change is consistent with empirical evidence provided in this paper, and computerization seems to be the main factor that changed labor demand.

6 Discussions

6.1 Empirical Assessment of Identification Strategy

This subsection empirically assesses the validity of my identification strategy to supplement the discussion in Section 3.3.1. Previous papers use instrumental variables to address the endogeneity bias following from selection into occupations based on unobserved worker characteristics. For example, Gould (2002) and Fletcher and Sindelar (2009) instrument occupations using the occupations of the worker's father. The rationale is that they affect the job preference of children, but not wages directly. Another set of instrumental variables extracts the variation in local labor market conditions. Fletcher and Sindelar (2009) use the fractions of blue-collar workers in a given state to measure availability of blue-collar and white-collar jobs for their study of the causal effect of occupations on health. The fact that a substantial proportion of jobs in Michigan are blue collar increases the likelihood that a new entrant will start working in a blue collar job. Devereux (2002)

[13]The male college wage premium increased by 0.239 log points from 1979 to 1996. Out of 0.239, changes in returns to cognitive skills and tasks account for 0.115-0.117, and changes in returns to motor skills and tasks account for 0.039-0.046 log points.

finds that the assignment of workers to occupations changes over the business cycle. Expansions allow workers to upgrade to occupations that pay higher wages and require more skill. These variables are potential candidates for instruments for occupations.

However, I do not instrument the task indices for occupations by these variables. First, they may not be valid because of their potential direct effects on wages. Father's occupations plausibly affect children's job preferences, but children are likely to invest in relevant skills in order to be qualified for their preferred jobs.[14] Local labor market conditions certainly affect workers' occupational choices by shifting the availability of particular jobs, but they are also likely to affect the wages of those who do not change occupations, which means the labor market conditions directly affect wages.

Second, my control variables for skills seem to work well, and father's occupation and local labor market conditions do not provide much additional identification power. I run the instrumental variable first-stage regressions to see how strongly the instruments are correlated with the task indices. The dependent variable is the cognitive (motor) task index, and the instruments used include 9 dummies for father's occupation, the state-level employment shares for 8 occupations (1-digit), and the state-level unemployment rate. Along with these instruments, I include two sets of covariates. In model 1, I include variables that are routinely used for wage regressions: education, dummies for college and advanced degrees, experience, experience-squared, and dummies for whites and females. In model 2, in addition to the covariates in model 1, I include the sum of the task indices for the past jobs and the time-average of the task indices. These additional variables in model 2 are not used by previous papers, and I expect them to control for worker heterogeneity. I run the first-stage regressions for models 1 and 2 by pooling the PSID sample, eliminating observations with missing variables.

Table 4 reports F-statistics. Staiger and Stock (1997) and Stock, Wright, and Yogo (2002) find that the F-statistic must be greater than 10 to reject the null hypothesis of weak instruments in cross-section data with homoskedastic error terms, which is widely known as a rule-of-thumb for applied econometricians. Olea and Pflueger (2011) argue that an even larger value of the F-statistic is necessary to reject the null in the presence of heteroskedasticity and autocorrelation, which is likely to be found in panel data. In model 1, the instrumental variables are strongly correlated with the cognitive task index with the F-statistic being 64. I cannot reject that they are weak instruments for the motor task index as the F-statistic is 10. However, in model 2 with the skill control variables, the instruments are no longer strong. For the cognitive task index, the F-statistic is only 10, and it is 3 for the motor task index. This exercise suggests, but by no means proves, that the additional

[14] Note that validity of the instrument depends on what other covariates are also included in a regression. Fletcher and Sindelar (2009) are aware of the potential direct pathways linking father's occupation and child's adult health. They claim that the direct effects are controlled by including child's health through adolescence, child's educational attainment, and child's risk preference (proxied by smoking status).

skill control variables effectively take care of the endogeneity problem at least to the extent that the previous papers do using instrumental variables.

Table 4: Significance Test for Instruments for Task Indices

Dependent Variable	DF	F-statistic
Model 1		
Cognitive Task	18	63.514
Motor Task	18	9.608
Model 2		
Cognitive Task	18	9.710
Motor Task	18	2.670

Source: PSID 1979-1996. Sample size is 61,311.
Note: IVs include 9 dummies for father's occupation, the state-level employment shares for 8 occupations (1-digit), and the state-level unemployment rate. In model 1, I also include education, dummies for college and advanced degrees, experience, experience-squared, and dummies for whites and females. In model 2, in addition to the covariates in model 1, I include the sum of the task indexes for the past jobs and the time-average of the task indexes.

6.2 Robustness

6.2.1 Skills or Discrimination?

In the model, wages are the value of marginal product of labor, and the gender wage gap is due to differences in skills and tasks. Gender discrimination is not explicitly in the model, but discriminatory factors may potentially be picked up by female dummy variables in the the skill functions (23) and (24). However, the estimates of task-specific skill endowments do not suffer from a bias arising from potential gender discrimination unless the level of discrimination is correlated with task complexity. Recall the discussion in Section 3.3.2 over which features of the data identify the task-specific skill endowments. When men have more motor skills than women, the gender wage gap increases with motor task complexity, because men's motor skills are more intensely used in complex motor tasks. This variation in the gender wage gap across tasks is the source of identification for the task-specific skills. If gender discrimination exists and is not correlated with task complexity, discrimination is absorbed by general skills.

Not many theories of gender discrimination suggest a correlation between tasks and discrimination. An exception is the "glass ceiling" that keeps women from rising to upper rungs of the corporate ladder. This glass ceiling story implies that women are more strongly discriminated against in high-paying occupations such as managers. These occupations are usually characterized by a high level of cognitive task complexity. If this is the case, I would underestimate women's cognitive skill endowments. However, the glass ceiling theory is only relevant for high-rank positions, not for the occupations of average female workers. Bertrand and Hallock (2001) study gender differences among top executives using the ExecuComp data set containing information on the five highest-paid executives in each of a large number of U.S. firms for the years 1992–97 . While they find a significant gender gap that may or may not be due to discrimination, their result may not be generalizable to average workers, who are exactly the focus of this paper. This is not to say that gender discrimination such as the glass ceiling does not exist. Instead, I claim that the effects of discrimination on wages are absorbed into the general skill effect, and that task-specific skills and their returns are consistently estimated.

Note that women's growth of general skills are driven by better education and more experience. This statement is true even if the estimates for general skill endowments are biased due to gender discrimination, because the coefficient for a female dummy is time-invariant in the skill production function. This assumption may be restrictive, and a change in the level of gender discrimination over time could be a source of the narrowing gender wage gap. I relax the restriction and allow for the coefficient to change over time. However, no statistically significant change of the coefficient over time is found. This empirical exercise also indicates that a change in gender discrimination is not an important source of the narrowing gender wage gap.

6.2.2 Selection into Labor Force

Mulligan and Rubinstein (2008) find that the rise in returns to skills which changed the labor force participation patterns of women, ultimately changed the observed gender wage gap. During the 70's, less skilled women participated in the labor market, while skilled women remained in home. In the 80's and 90's, returns to skills rose, increasing the opportunity cost of staying at home for skilled women. During the 90's, skilled women participated in the labor market, while unskilled women remained at home. They claim that this change in the labor force participation patterns of women largely accounts for the narrowing gender wage gap.

This is a potentially important issue not addressed by the model. To account for changes in the skill composition of workers, I allow for cohort effects in the general skills function by including a birth year variable and its interaction with a female dummy. These additional variables capture changes in conditional mean unobserved skills given full-time labor force participation, which may include the selection effect mentioned above and pure cohort effects. Note that these effects are identified separately for men and women. Cohort effects could be included in the task-specific skill functions as well, but the standard errors too large to interpret the result in a meaningful way. In addition, the covariates in the task-specific functions already seem to control for unobserved heterogeneity, and thus, a potential cohort effect may have already been captured by the broad set of control variables.

Table 5 reports the Oaxaca-Blinder decomposition result under the augmented specification.

Compare this with the result reported in Table 3 When cohort effects are included, a change in general skill endowment accounts for a greater part of the narrowing gender wage gap as one can expect. It accounts for 0.095 out of 0.187 log points, which is an increase from 0.067. In contrast, the contribution of a decrease in returns to motor skills and tasks decrease from 0.091-0.097 to 0.064-0.069 when cohort effects are included. Contributions of other factors to the narrowing gender wage gap remain almost the same. The results indicate that changes in selection and cohort effects are not negligible, but are not large enough to alter the main message of the paper: the drop in the returns to motor skills is the main factor which accounts for the narrowing gender wage gap..

While this paper addresses the consequence of the selection problem by allowing for cohort effects, readers may consider it not the best approach, and that the selection problem should be addressed by modeling the selection process. However, constructing counterfactual wages using a selection model is not a straightforward exercise and needs to rely on strong assumptions. In particular, identification of the intercept, which is needed to compare mens' and womens' mean wages, is known to be difficult. The well-known Heckman selection model (Heckman (1979)) relies on the normality assumption. A credible exclusion restriction is also necessary. It is true that understanding the selection mechanism (possibly using a structural model) is a useful step toward a thorough understanding of the relationship between technological change and female labor market outcomes, but this paper takes a less ambitious approach to avoid biases stemming from potential misspecification of the selection process.

Table 5: Decomposition of Changes in Gender Logwage Gap (Accounting for Cohort Effects)

	Specification 1		Specification 2	
	Estimate	Std. Error	Estimate	Std. Error
Change in Returns to Skills and Tasks				
Cognitive	−0.014	0.005	0.002	0.005
Motor	−0.069	0.017	−0.064	0.016
Subtotal	−0.083	0.016	−0.062	0.016
Change in Skill Endowments				
Cognitive	−0.017	0.005	−0.034	0.005
Motor	0.014	0.010	0.010	0.009
General Skills	−0.095	0.016	−0.095	0.016
Subtotal	−0.098	0.013	−0.119	0.013
Unexplained	−0.007	0.017	−0.007	0.017
Overall Change in Gap	−0.187	0.028	−0.187	0.028

Note: Standard errors are calculated by the delta method and clustered at the individual level. In specification 1 (2), the quantity effect is calculated by fixing the parameters at the 1979 (1996) level.

7 Conclusion

I estimate the endowments of cognitive and motor skills for men and women, and the returns to these skills. The proposed model applies the task-based approach to the Roy model, in order to explain how workers are sorted across occupations depending on their skill endowments. It allows me to understand better what an occupation is and account for a greater degree of heterogeneity in tasks and skills, relative to models in which an occupation is treated as a distinct category as often done in empirical papers.

The empirical results show that women have more cognitive skills than men, while men have more motor and general skills. The returns to motor skills dropped dramatically, while those to cognitive skills modestly increased, over 1979 to 1996, which suggests that technological change is responsible for the narrowing gender wage gap. In particular, the significant drop in returns to motor skills accounts for as much as a half of the narrowing gap. Increases in women's cognitive and general skills also significantly narrowed the gender wage gap.

In the results reported in supplementary appendix, I find that the changes in the value of cognitive and motor skills also largely account for the rise in male wage inequality. These results indicate that the proposed task-based Roy model provides a unified framework to study two different changes in the wage structure.

Although I provide evidence that taking occupational choice and labor force participation as exogenous, conditional on the broad set of control variables, does not largely bias the main results, incorporating these features in the model is a promising direction for future research. Changes in returns to skills should have affected workers' skill investment as well as occupational choice and labor force participation decisions. Studying these changes in worker behavior is an important step toward a thorough understanding of how technological change has reshaped the labor market.

References

ACEMOGLU, D., AND D. AUTOR (2011): *Skills, Tasks and Technologies: Implications for Employment and Earnings*Elsevier, vol. 4 of *Handbook of Labor Economics*, chap. 12, pp. 1043–1171.

ALTONJI, J. J., AND R. L. MATZKIN (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053–102.

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2008): "Trends in U.S. Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, 90(2), 300–323.

AUTOR, D. H., L. F. KATZ, AND A. B. KRUEGER (1998): "Computing Inequality: Have Computers Changed The Labor Market?," *The Quarterly Journal of Economics*, 113(4), 1169–1213.

AUTOR, D. H., F. LEVY, AND R. J. MURNANE (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics*, 118(4), 1279–1333.

BACOLOD, M., AND B. S. BLUM (2010): "Two Sides of the Same Coin: U.S. "Residual" Inequality and the Gender Gap," *Journal of Human Resources*, 45(1), 197–242.

BEAUDRY, P., M. DOMS, AND E. LEWIS (2010): "Should the Personal Computer Be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas," *Journal of Political Economy*, 118(5), 988 – 1036.

BERTRAND, M., AND K. F. HALLOCK (2001): "The Gender gap in top corporate jobs," *Industrial and Labor Relations Review*, 55(1), 3–21.

BLACK, S. E., AND A. SPITZ-OENER (2010): "Explaining Women's Success: Technological Change and the Skill Content of Women's Work," *The Review of Economics and Statistics*, 92(1), 187–194.

BLAU, F. D., AND L. M. KAHN (1997): "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s," *Journal of Labor Economics*, 15(1), 1–42.

——— (2000): "Gender Differences in Pay," *Journal of Economic Perspectives*, 14(4), 75–99.

——— (2006): "The U.S. gender pay gap in the 1990s: slowing convergence," *Industrial and Labor Relations Review*, 60(1), 45–66.

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econonometrics: Theory and Applications*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky. Cambridge University Press, vol. 2, pp. 312–357.

BORGHANS, L., B. TER WEEL, AND B. A. WEINBERG (2006): "People People: Social Capital and The Labor-Market Outcomes of Underrepresented Groups," NBER Working Paper 11985.

BUREAU OF LABOR STATISTICS (1994): *Technology and Labor in Pulp, Paper,Paperboard and Selected Converting Industries*Washington D.C., vol. BLS Bulletin 2443.

CARD, D., AND J. E. DINARDO (2002): "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *Journal of Labor Economics*, 20(4), 733–783.

DEVEREUX, P. J. (2002): "Occupational Upgrading and the Business Cycle," *LABOUR*, 16(3), 423–452.

FIRPO, S., N. FORTIN, AND T. LEMIEUX (2011): "Occupational Tasks and Changes in the Wage Structure," University of British Columbia.

FLETCHER, J., AND J. L. SINDELAR (2009): "Estimating Causal Effects of Early Occupational Choice on Later Health: Evidence Using the PSID," NBER Working Paper no. 15256.

FORTIN, N. (2008): "The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People," *Journal of Human Resources*, 43(4), 884–918.

GALOR, O., AND D. N. WEIL (1996): "The Gender Gap, Fertility, and Growth," *American Economic Review*, 86(3), 374–87.

GIBBONS, R., L. F. KATZ, T. LEMIEUX, AND D. PARENT (2005): "Comparative Advantage, Learning and Sectoral Wage Determination," *Journal of Labor Economics*, 23(4), 681–724.

GOOS, M., AND A. MANNING (2007): "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *Review of Economics and Statistics*, 89(1), 118–133.

GOULD, E. D. (2002): “Rising Wage Inequality, Comparative Advantage, and the Growing Importance of General Skills in the United States,” *Journal of Labor Economics*, 20(1), 105–47.

GREENACRE, M. (2007): *Correspondence Analysis in Practice*. Chapman & Hall/CRC.

HECKMAN, J., AND J. SCHEINKMAN (1987): “The Importance of Bundling in a Gorman-Lancaster Model of Earnings,” *Review of Economic Studies*, 54(2), 243–55.

HECKMAN, J. J. (1979): “Sample Selection Bias as a Specification Error,” *Econometrica*, 47(1), 153–61.

HECKMAN, J. J., AND G. SEDLACEK (1985): “Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market,” *Journal of Political Economy*, 93(6), 1077–1125.

KAMBOUROV, G., AND I. MANOVSKII (2008): “Rising Occupational and Industry Mobility in the United States: 1968-1997,” *International Economic Review*, 49(1), 41–79.

——— (2009): “Occupational Specificity of Human Capital,” *International Economic Review*, 50(1), 63–115.

KATZ, L. F., AND K. M. MURPHY (1992): “Changes in Relative Wages, 1963-1987: Supply and Demand Factors,” *The Quarterly Journal of Economics*, 107(1), 35–78.

KRUSELL, P., L. E. OHANIAN, J.-V. RIOS-RULL, AND G. L. VIOLANTE (2000): “Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis,” *Econometrica*, 68(5), 1029–1054.

LEE, D., AND K. I. WOLPIN (2006): “Intersectoral Labor Mobility and the Growth of the Service Sector,” *Econometrica*, 74(1), 1–46.

MA, S., J. S. RACINE, AND L. YANG (2011): “Spline Regression in the Presence of Categorical Predictors,” McMaster University.

MILLER, A. R., D. J. TREIMAN, P. S. CAIN, AND P. A. ROOS (eds.) (1980): *Work, Jobs, and Occupations: A Critical Review of the Dictionary of Occupational Titles*. National Academy Press.

MULLIGAN, C. B., AND Y. RUBINSTEIN (2008): “Selection, Investment, and Women’s Relative Wages Over Time,” *Quarterly Journal of Economics*, 123(3), 1061–110.

NEAL, D. (1999): “The Complexity of Job Mobility among Young Men,” *Journal of Labor Economics*, 17, 237–61.

OLEA, J. L. M., AND C. PFLUEGER (2011): "A Heteroskedasticity and Autocorrelation Robust Pre-Test for Weak Instruments," Harvard University.

PAVAN, R. (2011): "Career Choice and Wage Growth," *Journal of Labor Economics*, 29(3), 549 – 587.

RACINE, J. S., AND Z. NIE (2011): *crs: Categorical Regression Splines*R package version 0.15-12.

RENDALL, M. (2010): "Brain versus Brawn: The Realization of Women's Comparative Advantage," University of Zurich.

ROSEN, S. (1978): "Substitution and Division of Labour," *Economica*, 45(179), 235–50.

——— (1983): "A Note on Aggregation of Skills and Labor Quality," *Journal of Human Resources*, 18(3), 425–31.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20(4), 518–29.

SULLIVAN, P. J. (2009): "Estimation of an Occupational Choice Model when Occupations are Misclassified," *Journal of Human Resources*, 44(2), 495–535.

WEINBERG, B. A. (2000): "Computer Use and the Demand for Female Workers," *Industrial and Labor Relations Review*, 53(2), 290–308.

WELCH, F. (1969): "Linear Synthesis of Skill Distribution," *Journal of Human Resources*, 4(3), 311–27.

——— (2000): "Growth in Women's Relative Wages and in Inequality among Men: One Phenomenon or Two?," *American Economic Review*, 90(2), 444–49.

WOOLDRIDGE, J. M. (2009): "Correlated Random Effects Models With Unbalanced Panels," Michigan State University.

YAMAGUCHI, S. (2010): "The Effect of Match Quality and Specific Experience on Career Decisions and Wage Growth," *Labour Economics*, 17(2), 407–23.

Table 6: DOT Variables for Cognitive Task Complexity Index

Variable	No of Categories/Levels
Worker function: data	7
Worker function: people	9
General educational development: reasoning	6
General educational development: mathematics	6
General educational development: language	6
Aptitude: intelligence	4
Aptitude: verbal	5
Aptitude: numerical	5
Temperament: Influencing people	2
Temperament: Dealing with people	2

——— (2012): "Tasks and Heterogeneous Human Capital," *Journal of Labor Economics*, 30(1), 1 – 53.

A Details of Dataset Construction

A.1 Task Complexity Indices

Correspondence analysis is a generalized principal component analysis tailored for the analysis of qualitative data. Multiple correspondence analysis (MCA) is an extension of correspondence analysis which allows one to analyze the pattern of relationships of several categorical dependent variables. In other words, MCA is a dimension reduction technique for categorical variables including ordered ones. A brief description of the computing algorithm MCA can be found in the supplementary appendix.

MCA is applied to the 1971 augmented CPS to construct cognitive and motor task complexity indices.[15] Tables 6 and 7 list the DOT variables used for MCA. There are 10 categorical variables for the cognitive task complexity index and 8 categorical variables for the motor task complexity index. These variables are converted into 52 and 43 indicator variables, respectively. In MCA, variation of the data is called inertia, which is the sum of chi-square distances to the centroid. The constructed task complexity indices account for 50% and 39% of the inertia, respectively. In calculating inertia, I account for off-diagonal subtables of the Burt matrix only. The indices are then normalized such that the mean is 0.5 and the standard deviation is 0.1.

The task complexity indices at the 1970 census 3-digit level are calculated by taking average over individuals in a given 3-digit occupation. For a small number of the 3-digit occupations,

[15] The data file is available at the ICPSR website (http://dx.doi.org/10.3886/ICPSR07845.v2).

Table 7: DOT Variables for Motor Task Complexity Index

Variable	No of Categories/Levels
Worker function: things	8
Aptitude: motor coordination	5
Aptitude: finger dexterity	5
Aptitude: manual dexterity	5
Aptitude: eye-hand-foot coordination	5
Aptitude: spatial	5
Aptitude: form perception	5
Aptitude: color discrimination	5

no individuals are surveyed in the CPS. The task complexity indices for these occupations are constructed as follows. Using a crosswalk file[16] between the 1970 census occupation code and the 4th edition DOT occupation code, I take the relevant task variables from DOT for a given 1970 census occupation. There may be more than one DOT occupations for a census occupation. For all of the relevant DOT occupations, I calculate the predicted factor scores using the MCA results and apply the normalization outlined above. Finally, I construct the task indices for a census occupation by taking an unweighted average over the corresponding DOT occupations.

A.2 PSID Variable Definitions

Education Education is reported in the PSID in 1968, 1975, and 1985 for existing heads of households, and every year for the new entrants into the sample only. When education is missing, I first use education reported in the survey prior to the year when education is missing. Then, I use education reported in the survey after the year when education is missing, if necessary.

Demographic Variables Age, gender, and race are used in this paper. Race of a wife is not reported between 1968 and 1984. I assume a wife's race is same as that of her husband during this period. If more than one race is reported throughout the survey years, the most often reported answer is used.

Wages Hourly wages are calculated by dividing total labor earnings by hours worked. They are then deflated by deflated by 1983 PCE Index. Hourly wages less than \$1 or more than \$250 are treated as missing.

[16]Provided by the National Crosswalk Service Center. The file is downloadable from ftp://ftp.xwalkcenter.org/download/xwalks/cen70dot.xls.

Hours Worked When missing, I impute hours of work by taking an average of those in the previous and next years. The imputed values are used to determine whether one works full-time or not. A full-time work consists of 1,500 hours of work in a year. This variable is subsequently used to construct experience and task indices. The imputed values are not used to calculate hourly wages.

Experience Experience is reported in 1974, 1975, 1976, 1985, and years when a household is interviewed for the first time. For all other years, experience is imputed, using the indicators for full-time work and experience reported in the earlier survey closest to the year when experience is missing. For example, to impute experience in 1980, I add years of full-time work from 1976 to 1979 to experience reported in 1976. Similarly, experience in 1990 is calculated by adding years of full-time work from 1985 to 1989 to experience reported in 1990.

Occupation and Industry For years between 1968 and 1980, I use occupation and industry code from the retrospective supplemental data files. From 1981 on, the code reported in each survey year is used.

Task Index The task indices constructed from the DOT are merged with the PSID sample using the 1970 Census 3-digit occupation code. When occupation is missing, but those in the previous and next years are available, the task index is imputed by taking an average of them. When occupation is missing, but that in the previous (next) year is available and the individual does not work full-time in the next (previous) year, the task index in the current year is assumed to be same as that in the previous (next) year. No imputation is conducted in all other cases.

Cumulated Task Index One key control variable is the sum of task indices over the past jobs. However, the PSID does not include the whole occupational history since entry to the labor market. The cumulative task index is imputed as follows. I first run a fixed-effect regression of the task index on experience and its square. Here, I assume that individuals have a common slope for the profile of task index and experience, but they have different intercepts. The individual intercept is given by an average of residuals from the fixed-effect regression. The missing values are imputed using the estimated task-experience profile with heterogeneous intercepts.

B Structural Parameter Estimates

Table 8: Skill Equations

Variables	Estimate	Std. Error
Cognitive Skill		
Women	0.042	0.010
White	−0.012	0.010
Education	−0.010	0.003
College	0.035	0.022
Adv. Degree	0.003	0.055
Exprience	−0.001	0.002
Experience-sq/100	−0.002	0.003
Sum of COG Task	0.005	0.002
Ave. COG Task	8.249	0.963
Sq. of Ave. COG Task	−6.311	0.853
Motor Skill		
Women	−0.251	0.096
White	−0.032	0.037
Education	−0.016	0.011
College	−0.015	0.057
Adv. Degree	0.072	0.095
Exprience	−0.032	0.008
Experience-sq/100	0.038	0.016
Sum of MTR Task	0.023	0.003
Ave. MTR Task	0.186	0.677
Sq. of Ave. MTR Task	0.131	0.605
General Skill		
Women	−0.152	0.095
White	0.081	0.034
Education	0.068	0.010
College	0.017	0.064
Adv. Degree	0.092	0.150
Exprience	0.052	0.008
Experience-sq/100	−0.088	0.015

Note: Standard errors are clustered at individual level. See Section 3.1.2 for detail.

Table 9: Production Function at Occupation Level

Variables	Estimate	Std. Error
	Intercept: $\ln a_j$	
Intercept	7.164	0.842
COG	−15.606	2.084
MTR	4.618	1.343
COG-sq.	−6.664	1.012
MTR-sq.	−3.532	1.025
	Returns to Cognitive Skill	
$\beta_{C,0}$	−3.800	0.314
$\beta_{C,1}$	9.289	0.608
$\beta_{C,2}$	0.035	0.016
	Returns to Motor Skill	
$\beta_{M,0}$	−0.324	0.445
$\beta_{M,1}$	2.563	0.862
$\beta_{M,2}$	−0.041	0.014

Note: Standard errors are clustered at individual level. The production function is $\ln q_j(s_{it}) = \ln \alpha_j + \beta_{Cjt} \ln s_{C,it} + \beta_{Mjt} \ln s_{M,it} + \ln s_{G,it}$. The production function parameters are given by $\ln \alpha_j = \alpha_0 + \alpha_1 x_{C,j} + \alpha_2 x_{M,j} + \alpha_3 x_{C,j}^2 + \alpha_4 x_{M,j}^2$ and $\beta_{k,jt} = \beta_{k,0} + [\beta_{k,1} + \beta_{k,2}(t-1979)]x_{k,j}$ $k \in \{C, M\}$. The parameters for the returns to skills are normalized to one in 1979 for the average task in the PSID sample. Namely, it is imposed that $\beta_{k,0} + \beta_{k,1}\bar{x}_{k,t=79} = 1$ where $\bar{x}_{k,t=79}$ is the average task over individuals in the PSID sample in 1979. See Section 3.1.1 for detail.

Table 10: Output Price Function

Variables	Estimate	Std. Error
Intercept: $\pi_{0,t}$		
t_{79}	-0.042	0.041
t_{89}	0.211	0.082
Cognitive Task: $\pi_{1,t}$		
t_{79}	-0.047	0.117
t_{89}	-0.317	0.225
Motor Task: $\pi_{2,t}$		
t_{79}	-0.008	0.129
t_{89}	-0.358	0.249
Cognitive Task Sq.: $\pi_{3,t}$		
t_{79}	0.041	0.108
t_{89}	0.187	0.224
Motor Task: $\pi_{4,t}$		
t_{79}	0.008	0.121
t_{89}	0.335	0.234

Note: Standard errors are clustered at individual level. Note that the log output price in 1979 $\ln P_{jt=1979}$ is zero for normalization. The output price function is $\ln P_{jt} = \pi_{0,t} + \pi_{1,t} x_{C,j} + \pi_{2,t} x_{M,j} + \pi_{3,t} x_{C,j} + \pi_{4,t} x_{M,j}$. The changes of the time-varying parameters are approximated by a piecewise-linear function. For example, the parameter $\pi_{t,0}$ is given by $\pi_{0,t} = \pi_{0,79} t_{79} + \pi_{0,89} t_{89}$ where $t_{79} = (t-1979) \cdot I(t \geq 1979)$, $t_{89} = (t-1989) \cdot I(t \geq 1989)$, and $I(\cdot)$ is an indicator function that takes one if the condition in the parenthesis is satisfied and takes zero otherwise.

Supplementary Appendix to "Changes in Returns to Task-Specific Skills and Gender Wage Gap"

Shintaro Yamaguchi*

April 23, 2012

A Decomposition of Male College Wage Premium

In this section I decompose changes in the male college wage premium, in order to assess the extent to which changes in returns to skills and tasks account for rising male wage inequality as well as the narrowing gender wage gap. The male college wage premium is measured by the mean logwage difference between high school and college graduate workers. They are subpopulations of the main analysis, and thus, the model and its parameter estimates are identical to those used in the main analysis of the gender wage gap.

Table 1 reports the means of selected labor market outcomes for male high school and college graduates for the years 1979, 1989, and 1996. The mean log hourly wage exhibits well-known patterns. College graduates earn more than high school graduates, and the gap between the two groups keeps expanding, but at a slower pace in the 1990's. In 1979, the college wage premium was 0.290 log points. It quickly rose to 0.470 in 1989, and the premium was 0.528 log points in 1996. Experience increases over time for both groups, but the pace is faster for college graduates in my PSID sample.

The table also reports cognitive and motor task complexity indices, and Figure 1 illustrates the time profiles of these variables. The top panel exhibits large differences in cognitive task complexity. The cognitive task complexity index for college graduate workers is greater than that of high school graduates by as much as 0.10, which is equal to the standard deviation for the working population in the 1971 CPS. The gap is much larger than the gender gap in cognitive task

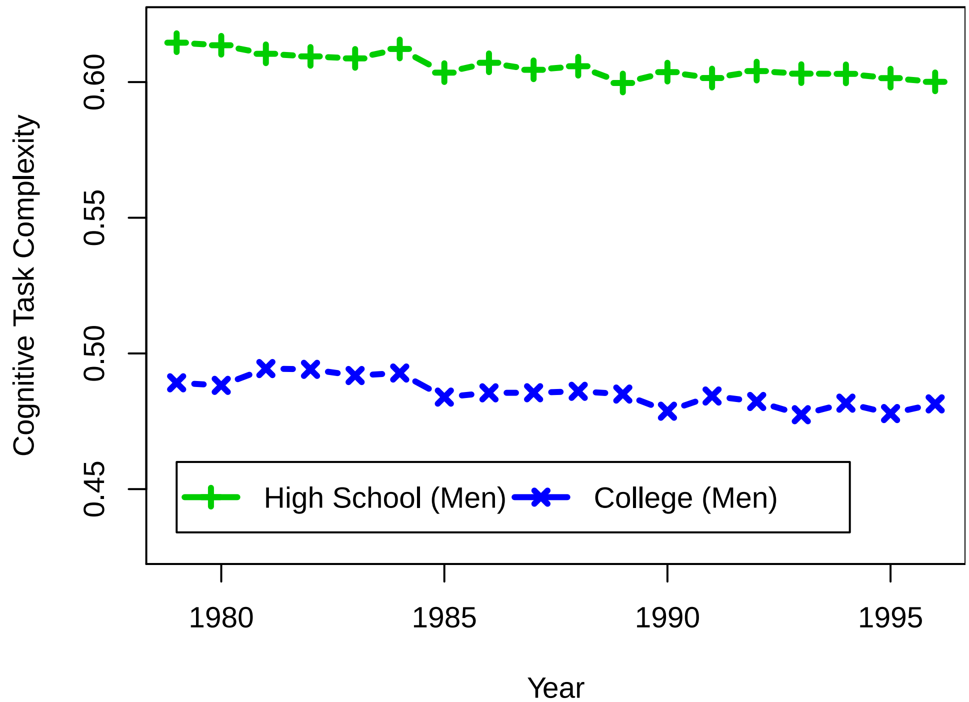
*Department of Economics, McMaster University, 1280 Main St. W., Hamilton, ON. Canada L8S 4M4. Email: yamtaro@mcmaster.ca. Website: socserv.mcmaster.ca/yamtaro.

complexity, which is between 0.01 and 0.02. The bottom panel in Figure 1 shows the differences in motor task complexity. The motor task complexity index for high school graduates is about 0.05 points greater than that for college graduates. Again, the difference between the education groups is greater than the corresponding gender wage gap, which is between 0.01 and 0.02. The differences in task complexity indices between high school and college graduates are large, and remain relatively constant throughout the sample period.

Table 1: Mean of Selected Labor Market Outcomes in PSID

	1979		1989		1996	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Logwage						
College	2.644	0.024	2.691	0.027	2.742	0.031
High School	2.354	0.015	2.221	0.017	2.214	0.020
Cognitive Task Index						
College	0.611	0.003	0.601	0.003	0.604	0.003
High School	0.488	0.003	0.484	0.003	0.482	0.003
Motor Task Index						
College	0.484	0.006	0.480	0.005	0.471	0.005
High School	0.528	0.003	0.525	0.003	0.522	0.003
Years of Full-time Experience						
College	15.659	0.588	16.655	0.473	18.880	0.523
High School	17.993	0.454	18.687	0.415	18.910	0.443
			Number of Observations			
College	333		424		346	
High School	710		720		540	

Source: PSID 1979-1996.
Note: Wages are deflated by 1983 PCE Index. The sample includes male full-time (1,500 hours a year or more) high school and college graduate workers. Self-employed workers and those who are younger than 18 or older than 65 are excluded from the sample.



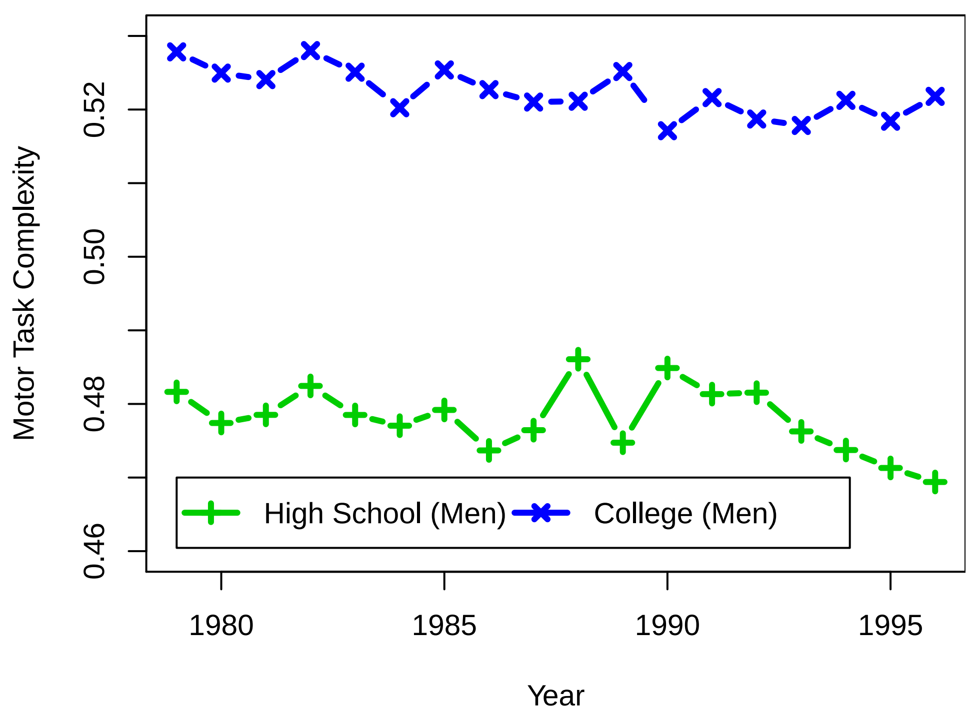
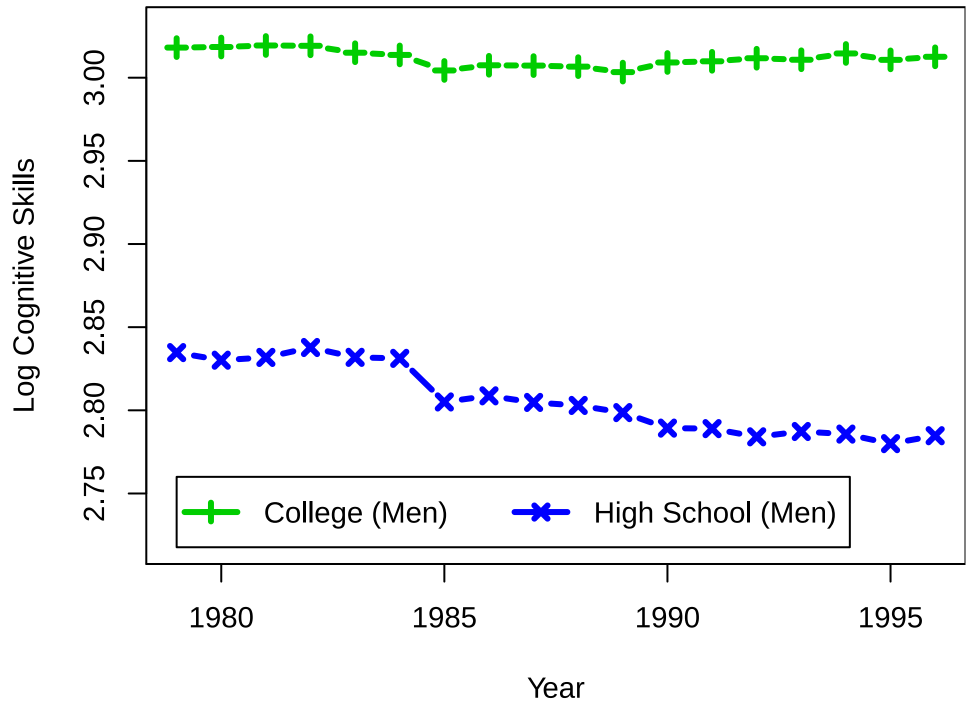


Figure 1: Task Complexity Indexes for Male High School and College Graduates
Source: PSID 1979-1996 and DOT
Note: The task complexity indexes are normalized so that the mean is 0.5 and the standard deviation is 0.1 for the working population in the 1971 CPS.

Figure 2 illustrates the cognitive and motor skill profiles. Remember that skills are identified up to a linear transformation. The returns to task-specific skills are normalized to one for the average task in 1979. Hence, the skill level does not have any meaning on its own, but the skill difference is equivalent to the logwage difference when evaluated at the 1979 average task. The top panel indicates that college graduates had more cognitive skills than high school graduates, with the difference increasing over time. In 1979, the cognitive skill wage gap allows college graduates to earn a 19% higher wage than high school graduates if all else is equal. The gap increases over time and is worth about 24% higher wages in 1996. The bottom panel shows that high school graduates have more motor skills than college graduates. In 1979, the motor skill wage gap is worth about 11% higher wages, and about 14% higher wages in 1996. Figure 2 contains the time profiles of general skills. College graduates possess more general skills, with the difference leading to about 31% higher wages. The difference continues to grow over time, and leads to about 37% higher wages in 1996.



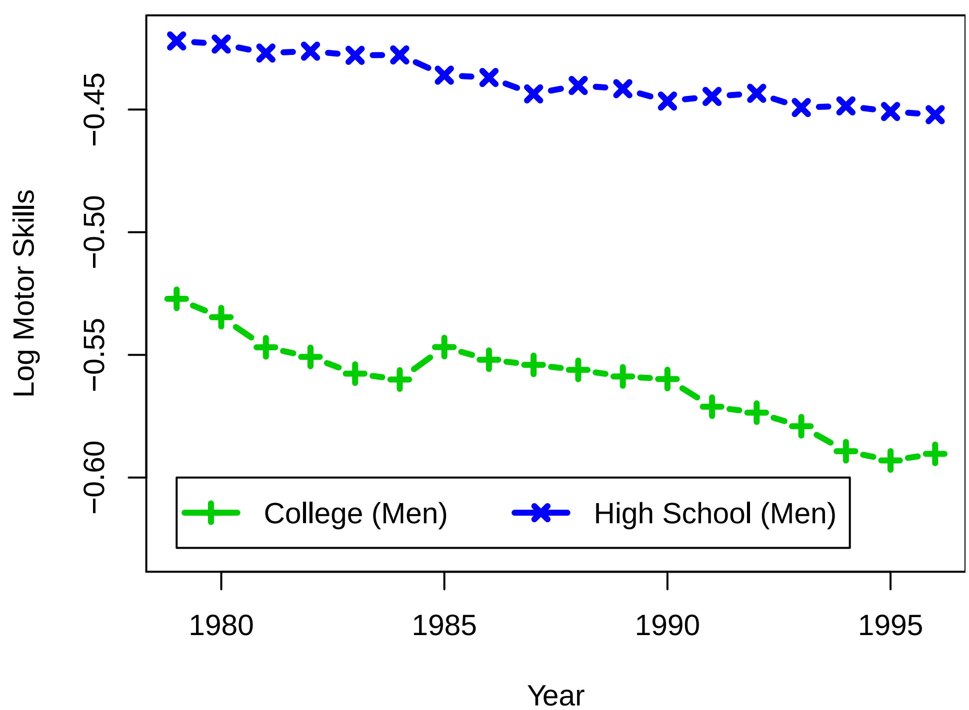


Figure 2: Task-Specific Skills for Male High School and College Graduates
Note: Skills are identified up to a linear transformation. The returns to task-specific skills are normalized to one for the average task in 1979. Hence, the skill level does not have any meaning, but the skill difference is equivalent to the logwage difference when evaluated at the average task in 1979.

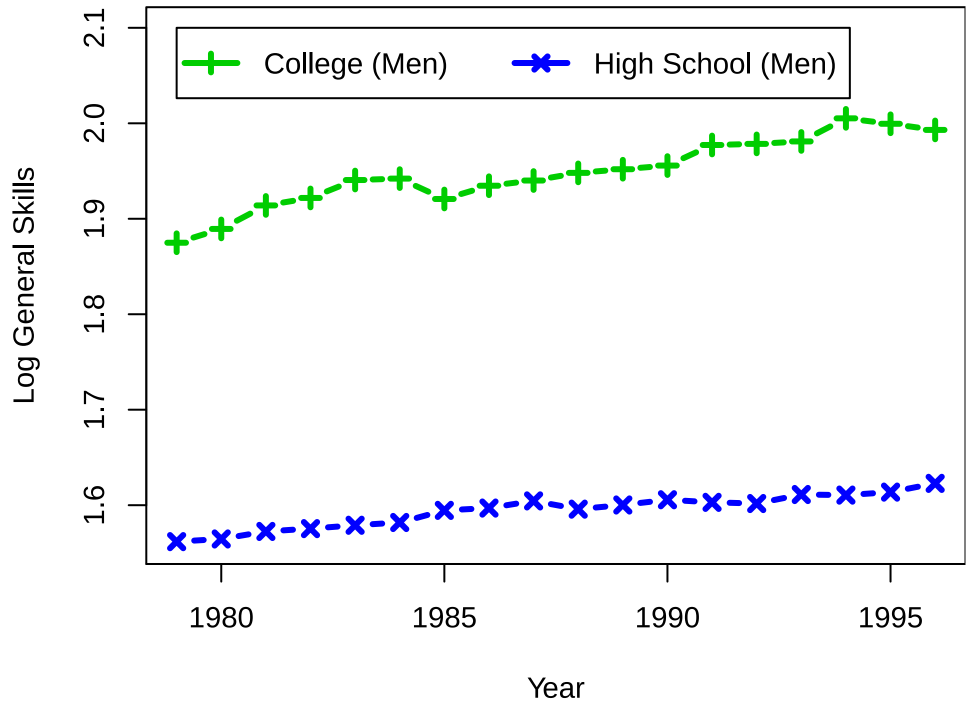


Figure 3: General Skills for Male High School and College Graduates

Table 2 presents the results of the Oaxaca-Blinder decomposition for changes in the logwage gap between male high school and college graduates. From 1979 to 1996, the male college wage premium rose by 0.239 log points. Changes in returns to skills and tasks account for as much as 0.161 points, about two thirds of the change. In particular, the change in return to cognitive skills and tasks accounts for a greater part at 0.115-0.117 log points, while the change in motor skills and tasks accounts for 0.039-0.046, reflecting a large cognitive skill gap and a modest motor skill gap between the two groups. The gap in skill endowments increases over time, accounting for 0.064-0.069 out of 0.239 log points. The growth of cognitive and motor skills offset each other, but the college graduates' faster growth of general skills leads to a rise in the college wage premium of 0.057 points. As seen in Table 1, years of work experience grow faster for college graduates than high school graduates.

The decomposition exercise reveals that changes in values of cognitive and motor skills are the main driver behind the rise in the male college wage gap as well as that of the narrowing gender wage gap. Combined with the analysis of the gender wage gap in the main body of the paper, this section demonstrates that the proposed task-based Roy model accounts for two different changes in the wage structure within a single framework.

Table 2: Decomposition of Male College Wage Premium

	Specification 1		Specification 2	
	Estimate	Std. Error	Estimate	Std. Error
Change in Returns to Skills and Tasks				
Cognitive	0.115	0.020	0.117	0.021
Motor	0.046	0.021	0.039	0.021
Subtotal	0.161	0.020	0.157	0.021
Change in Skill Endowments				
Cognitive	0.035	0.011	0.034	0.012
Motor	−0.028	0.017	−0.022	0.015
General Skills	0.057	0.018	0.057	0.018
Subtotal	0.064	0.006	0.069	0.006
Unexplained	0.013	0.020	0.013	0.020
Overall Change in Premium	0.239	0.044	0.239	0.044

Note: Standard errors are calculated by the delta method and clustered at the individual level. In specification 1 (2), the quantity effect is calculated by fixing the parameters at the 1979 (1996) level.

B Multiple Correspondence Analysis

Correspondence analysis is a generalized principal component analysis tailored for the analysis of qualitative data. Multiple correspondence analysis (MCA) is an extension of correspondence analysis which allows one to analyze the pattern of relationships of several categorical dependent variables. In other words, MCA is a dimension reduction technique for categorical variables including ordered ones. The purpose of this section is to give a brief description of the computing algorithm MCA. As such, notation in the following is not consistent with that in the main text. This section should be treated independently from the other parts of the paper. Readers who are already familiar with MCA can skip this and start from the next subsection. The following heavily relies on Greenacre (2007).

Suppose I have data for N individuals indexed by $i = 1, \cdots, N$. For each individual, there are Q variables that are categorized into a broadly defined task group such as a cognitive task group. They are denoted by y_{iq} ($q = 1, \ldots, Q$). The cognitive task group includes 10 variables such as mathematical development and intelligence. The variable y_{iq} has J_q categories (or ranks for an ordered variable). This categorical variable y_{iq} can be represented by a J_q dimensional vector of indicator variables z_{iq}. Suppose that y_{iq} falls in the category "good" when other categories are "bad" and "average." Then, y_{iq} can be represented by a vector of indicator variables

$z_{iq} = (d_{i,\text{bad}},\ d_{i,\text{average}},\ d_{i,\text{good}}\ )$, where $d_{i,-}$ is an indicator variable that takes one if y_{iq} falls in the category and takes zero otherwise. Namely, $y_{iq} =$ "good" can be converted into $z_{iq} = (0,0,1)$. For each individual, all Q task categorical variables can be converted into a vector of indicator variables $z_i = (z_{i1,} \cdots, z_{iQ})$. The dimension of z_i is $J = \sum_q J_q$. The data are compactly denoted by a $N \times J$ matrix Z whose row consists of z_i'.

MCA can be performed by applying CA to this data matrix Z. Let P be a correspondence matrix $P = \frac{1}{NQ}Z$. Let p_{ij} be the (i,j) element of the matrix P. Then, the row and column masses are $r_i = \sum_{j=1}^{J} p_{ij}$ and $c_j = \sum_{i=1}^{N} p_{ij}$. Vectors of row and column masses are given by $r = (r_1, \cdots, r_N)$ and $c = (c_1, \cdots, c_J)$. Define diagonal matrices of row and column masses $D_r = \text{diag}(r)$ and $D_c = \text{diag}(c)$. The standardized residual matrix S is given by

$$S = D_r^{-\frac{1}{2}}(P - rc')D_c^{-\frac{1}{2}}.$$

Calculate the singular value decomposition of S such that

$$S = UD_\alpha V',$$

where $U'U = V'V = I$ and D_α is the diagonal matrix of singular values. The standard coordinate Φ of rows is given by

$$\Phi = D_r^{-\frac{1}{2}}U.$$

The first column of Φ is the factor score used as a task complexity index for the task group (after normalization). The method can be easily modified to the data with sampling weights. Define a diagonal matrix of weights $D_w = \text{diag(w)}$, where w is a N-dimensional vector of weights. Let $\tilde{Z} = D_w^{\frac{1}{2}}Z$. Applying the CA to this converted data matrix $\tilde{Z}$, one can obtain the task index for the data with sampling weights. Sampling weights are used everywhere relevant in the following.

References

GREENACRE, M. (2007): *Correspondence Analysis in Practice*. Chapman & Hall/CRC.