Supplement to "The price of polarization: Estimating task prices under routine-biased technical change"

(*Quantitative Economics*, Vol. 11, No. 2, May 2020, 761–799)

Michael J. Böhm
Department of Economics, Institute for Applied Microeconomics, University of Bonn and IZA

Contents

Appendix A: Additional tables and figures

Table A.1. Male employment in the NLSY with respect to average demographics, early, and contemporary skill determinants.

	NLSY79	NLSY97
Nbr of observations	3054	1207
Percentage of observations	71.60	28.40
Demographics		
Age	27.00	27.00
White	0.80	0.72
Black	0.13	0.14
Hispanic	0.06	0.14

(*Continues*)

Michael J. Böhm: mboehm1@uni-bonn.de

Available at http://qeconomics.org. https://doi.org/10.3982/QE1031

TABLE A.1. *Continued.*

	NLSY79	NLSY97
Early skill determinants		
AFQT	167.31	167.65
Low AFQT Tercile	0.34	0.33
Middle AFQT Tercile	0.33	0.34
High AFQT Tercile	0.33	0.32
Math score (NCE)	50.45	50.73
Verbal score (NCE)	50.26	50.49
Mechanical score (NCE)	50.41	50.69
Illicit activities (NCE, Measured 1980)	49.98	50.01
Precocious sex (NCE, Measured 1983)	49.91	50.24
Mother's education (years)	11.86	13.11
Father's education (years)	10.83	13.09
Contemporary skill determinants		
High School Dropout (HSD)	0.12	0.07
High School Graduate (HSG)	0.43	0.58
Some College (SC)	0.20	0.06
College Graduate (CG)	0.19	0.24
Advanced Degree (AD)	0.06	0.04
North East	0.22	0.17
North Central	0.29	0.25
South	0.32	0.35
West	0.17	0.21

Note: The table shows average demographics and skill proxies in the NLSY79 and NLSY97 for all males weighted by hours worked. NCE indicates variables in the population (including nonworkers) are standardized to "normal curve equivalents" with mean 50 and standard deviation 21.06. This is done when absolute values of these variables cannot be compared over the two cohorts.

TABLE A.2. Sorting into tasks in the NLSY, multinomial logit regressions.

	(1) NLSY79	(2) NLSY79	(3) NLSY97	(4) NLSY97
Abstract				
Constant	−4.025 (0.265)	−1.665 (0.268)	−3.166 (0.373)	−1.386 (0.363)
Black	0.194 (0.165)	0.110 (0.167)	−0.153 (0.252)	−0.110 (0.261)
Hispanic	0.044 (1.63)	−0.030 (0.166)	−0.474 (2.61)	−0.459 (0.271)
Math (NCE)	0.047 (0.005)		0.034 (0.007)	
Verbal (NCE)	0.023 (0.0005)		0.032 (0.007)	
Mechanic (NCE)	−0.014 (0.004)		−0.019 (0.006)	
Middle math tercile		1.172 (0.197)		0.455 (0.256)
High math tercile		2.342 (0.224)		1.437 (0.272)
Middle verbal tercile		0.163 (0.190)		0.666 (0.274)
High verbal tercile		0.732 (0.209)		1.437 (0.308)
Middle mechanic tercile		−0.311 (0.180)		−0.257 (0.279)
High mechanic tercile		−0.603 (0.196)		−0.621 (0.305)
Illicit activities (NCE)		−0.008 (0.003)		−0.003 (0.005)
Precocious sex (NCE)		−0.005 (0.003)		−0.006 (0.004)
Manual				
Constant	−1.674 (0.273)	−1.570 (0.274)	−1.338 (0.340)	−2.032 (0.361)
Black	0.613 (0.174)	0.723 (0.176)	0.472 (0.267)	0.656 (0.283)
Hispanic	0.212 (0.182)	0.243 (0.182)	−0.217 (0.264)	−0.113 (0.263)
Math (NCE)	−0.003 (0.006)		−0.009 (0.007)	
Verbal (NCE)	0.018 (0.005)		0.021 (0.009)	
Mechanic (NCE)	−0.023 (0.005)		−0.017 (0.007)	
Middle math tercile		−0.322 (0.193)		−0.179 (0.249)
High math tercile		0.141 (0.266)		−0.472 (0.382)
Middle verbal tercile		0.303 (0.209)		0.305 (0.262)
High verbal tercile		0.470 (0.267)		0.840 (0.402)
Middle mechanic tercile		−0.351 (0.194)		−0.264 (0.247)
High mechanic tercile		−0.955 (0.250)		−0.586 (0.324)
Illicit activities (NCE)		−0.001 (0.003)		0.013 (0.007)
Precocious sex (NCE)		−0.004 (0.003)		−0.003 (0.006)
Pseudo R-squared	0.133	0.123	0.114	0.112
N	2944	2944	1210	1210

Note: Each column presents the results from a multinomial logit regression of task choice on demographics and talent proxies. The omitted group is the routine task. The first column uses only linear test scores in the NLSY79. The second column, which is the specification to estimate task propensities in the following, uses terciles of test scores and adds measures of risky behavior. Standard errors in parentheses behind the coefficients.

Table A.3. Task price changes with heterogeneous amenities (1984/92 to 2007/09).

	$\Delta(\pi_A - \pi_R)$ Log Points (s.e.)	$\Delta(\pi_M - \pi_R)$ Log Points (s.e.)	$\Delta\pi_R$ Log Points (s.e.)	$\bar{a}_A - \bar{a}_R$ Amenity (s.e.)	$\bar{a}_M - \bar{a}_R$ Amenity (s.e.)
OLS on propensities	21.5	−7.8	3.3	57.7	195.0
	(20.4)	(50.4)	(10.2)	(153.1)	(168.8)
OLS on propensities	49.4	28.1	−5.2	131.3	50.1
($t=0$: all $\boldsymbol{x}_i$ interactd)	(27.7)	(79.2)	(15.7)	(220.7)	(246.3)
OLS on propensities	22.8	20.4	−2.2	−26.0	156.3
(2nd-stage college)	(19.9)	(49.6)	(10.0)	(149.1)	(165.6)
OLS on propensities	25.0	24.2	1.6	13.7	156.7
(2nd-st degree dum.)	(20.2)	(49.7)	(11.0)	(147.9)	(162.9)
OLS on propensities	24.2	−7.8	1.7	43.1	211.6
(Adj. for min. wage)	(20.9)	(51.7)	(10.5)	(157.7)	(174.4)

Note: The table shows the results from the estimation method for the generalized Roy model with heterogeneous amenities by $\boldsymbol{x}$ characteristics. The first row presents task price changes and amenity intercepts from propensity regression (24) with further controls $\boldsymbol{x}_i \cdot \Delta p_A(\boldsymbol{x}_i)$ and $\boldsymbol{x}_i \cdot \Delta p_M(\boldsymbol{x}_i)$ interacted with time. Row two further adds $\boldsymbol{x}_i$ fully interacted with itself. The third and fourth row add college and detailed degree dummies interacted with time, respectively. The fifth row reports estimates when wages are first adjusted for the change in the real value of the minimum wage as in Lee (1999). The coefficients $b_k - b_R$ on the talents themselves are not reported in the interest of space. Bootstrapped standard errors (500 iterations) below the coefficients.

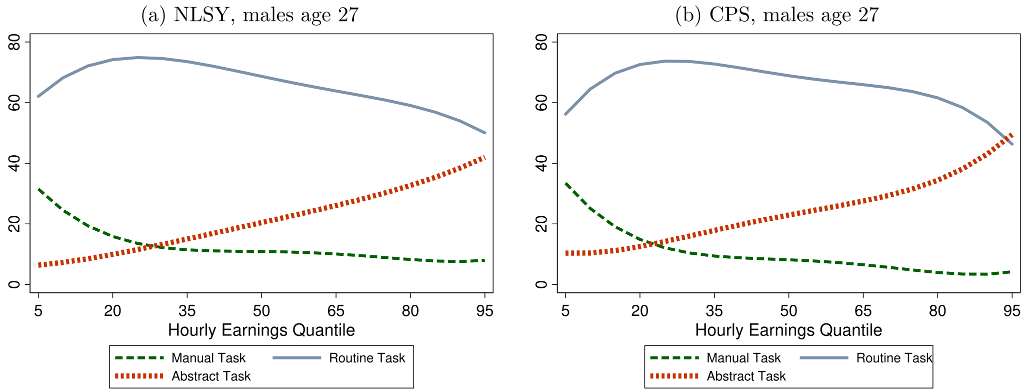

Figure A.1. Representation of tasks in the wage distribution, 1979 cohort. *Note*: The figure plots the smoothed employment share of the manual, routine, and abstract tasks within the quantiles of the wage distribution. Smoothing is done using the predicted values from a fourth-order polynomial regression of the employment shares on the quantiles.

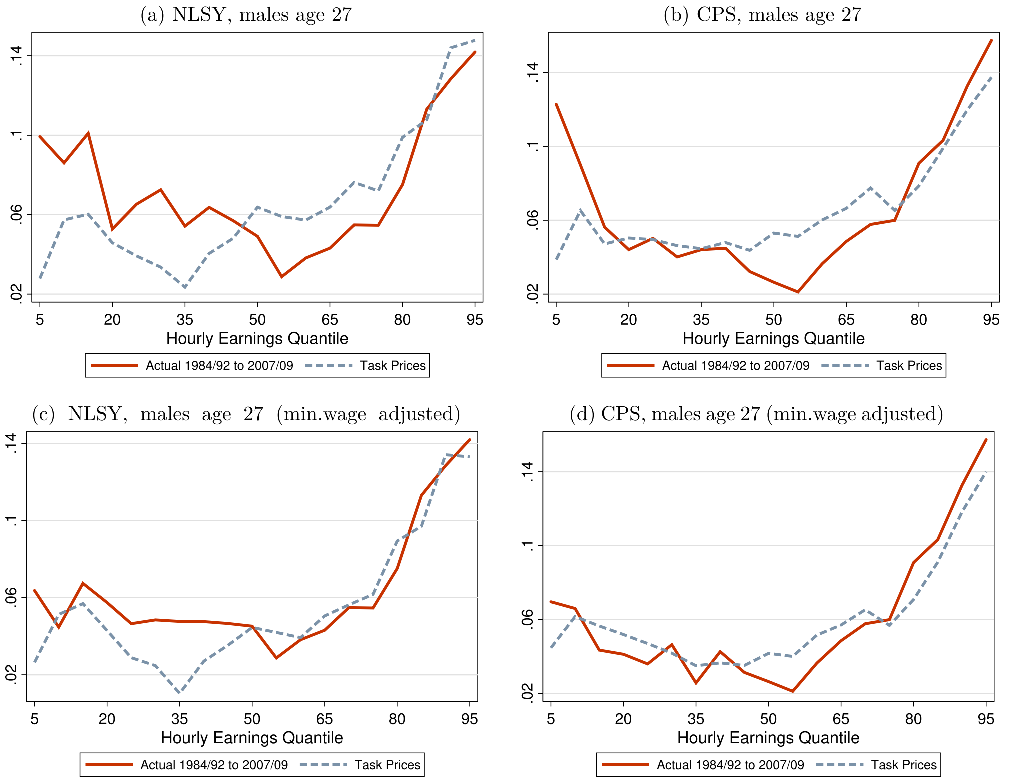

FIGURE A.2. Change in log real wages by quantile of the wage distribution, actual, and predicted due to changing task prices (homogeneous amenities).

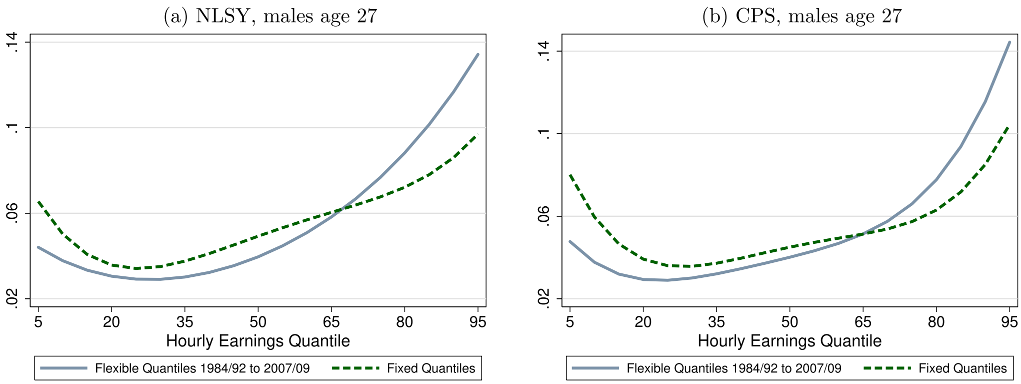

FIGURE A.3. Smoothed predicted change in the wage distribution due to changing task prices, flexible and fixed quantiles. *Note*: The solid line depicts the predicted change in log real wages along the quantiles of the wage distribution due to estimated changes in task prices. The dashed line depicts the same predicted change when individuals are fixed at their original quantiles in the wage distribution. The lines are smoothed because for the predicted under fixed quantiles the individuals who correspond to these quantiles exclusively determine their change. This would make the predicted change very spiky. Smoothing is done using the predicted values from a fourth order polynomial regression of average wage changes on the quantiles.

Appendix B: RBTC's predictions for workers' wages

This section studies what RBTC implies for workers' observed wages in the Roy model. I first show that supposed predictions of RBTC at the aggregate level are not robust to different assumptions of how workers' skills are distributed across tasks. Instead, RBTC has unambiguous implications for individual workers' wage growth over time. Finally, I test these implications in the NLSY data.

As discussed in Section 4.1, the vast majority of RBTC models in the literature imply that task prices will polarize, that is,

$$\Delta(\pi_A - \pi_R) > 0 \quad \text{and} \quad \Delta(\pi_M - \pi_R) > 0, \tag{B.1}$$

with $\Delta\pi_k \equiv \pi_{k1} - \pi_{k0}$ and where $k \in \{A, R, M\}$ are the three task groups used in many studies and in this paper's empirical application. In the following, I take (B.1) as given and examine its implications for aggregate as well as individual workers' wage outcomes. Further, as in the main text and in most of the literature (e.g., Autor, Katz, and Kearney (2006), Acemoglu and Autor (2011), Firpo, Fortin, and Lemieux (2013), Autor and Dorn (2013), Cortes (2016)), I continue to assume that whereas task prices are changing, skills are not affected by RBTC.

The next section restricts itself to a pure Roy framework. This is what most papers in the RBTC literature use to model labor supply and I show that it is already flexible enough to generate the different aggregate wage outcomes that have been found in the data. Sections B.2 and B.3 continue to focus on the purely pecuniary model but I also discuss the generalization that includes amenities.

B.1 *RBTC has no unambiguous predictions at the aggregate level*

It is well established that in the two-sector Roy model rising task prices in one sector need not lead to increasing wages in that sector and that self-selection may have ambiguous effects on overall wage inequality (e.g., Heckman and Honoré (1990)). By extension, these (negative) results should carry over to the three sector case. The contribution of this section is to illustrate that, even under normality, RBTC (or task prices polarization) does not imply that average wages in tasks or the overall wage distribution will polarize. Therefore, several empirical findings in the literature that seemingly contradict RBTC are in fact potentially consistent with it.

The results in the following are illustrated with simulated data using a multivariate normal distribution of log skills across tasks in the pure Roy model. Normality implies that the variances and correlations of skill are the only parameters determining sectoral and aggregate outcomes. Under a different distribution, other parameters may matter (Heckman and Honoré (1990)). Hence, this should be understood as just one specific illustration of the more general results. Table B.1 reports the parameters for the simulations. The differences between the respective left and right panels are either with respect to the variances or the correlations of skills and highlighted in bold.

Prediction 1 (Negative). *Task price polarization has no clear implication on wages in tasks. In particular, it need not lead to the polarization of average wages in tasks.*

TABLE B.1. Parameter values for the simulations in Figure B.1.

	Figure B.1(a), (b)		Figure B.1(c), (d)	
	Left Panel	Right Panel	Left Panel	Right Panel
$\Delta(\pi_A - \pi_R)$	0.35	0.35	0.35	0.35
$\Delta(\pi_M - \pi_R)$	0.10	0.10	0.30	0.30
$\Delta\pi_R$	−0.05	−0.05	−0.20	−0.20
$\text{var}(s_{Ai})$	3.0	3.0	3.0	3.0
$\text{var}(s_{Ri})$	1.5	1.5	**1.7**	**1.1**
$\text{var}(s_{Mi})$	1.3	1.3	1.0	1.0
$\text{corr}(s_{Ai}, s_{Ri})$	0.3	0.3	0.5	0.5
$\text{corr}(s_{Ai}, s_{Mi})$	**0.0**	**0.7**	0.5	0.5
$\text{corr}(s_{Ri}, s_{Mi})$	0.0	0.0	0.5	0.5

Note: Skills are multivariate normal with mean zero. Variances and covariances are given in the table together with the task price changes. The parameter values that differ between the respective left and right panels are emphasized in bold. $N = 10{,}000$ observations were drawn for each panel.

SKETCH OF PROOF OF PREDICTION 1. The change of average wages in task $k \in \{A, R, M\}$ can be split into a price and a selection effect:

$$\begin{aligned} &\mathrm{E}\big(w_i \mid I_k(\boldsymbol{s_i}, \boldsymbol{\pi_1}) = 1\big) - \mathrm{E}\big(w_i \mid I_k(\boldsymbol{s_i}, \boldsymbol{\pi_0}) = 1\big) \\ &\quad = \pi_{k1} - \pi_{k0} + \mathrm{E}\big(s_{ki} \mid I_k(\boldsymbol{s_i}, \boldsymbol{\pi_1}) = 1\big) - \mathrm{E}\big(s_{ki} \mid I_k(\boldsymbol{s_i}, \boldsymbol{\pi_0}) = 1\big). \end{aligned}$$

While the (relative) prices $\pi_{k1} - \pi_{k0}$ may rise, the (relative) skills s_{ki} selected into k may fall, depending on the joint population distribution of worker skills in tasks. This is the classic idea of (changing) selection bias. In some cases, the overall change will be the inverse of the task price change. For details, see the formal proof in Young (2014) who also shows that $\mathrm{E}(s_{ki} \mid I_k(\boldsymbol{s_i}, \boldsymbol{\pi_1}) = 1) - \mathrm{E}(s_{ki} \mid I_k(\boldsymbol{s_i}, \boldsymbol{\pi_0}) = 1)$ can be positive as well as negative. For the purpose of this paper, the specific counterexample in Figure B.1, Panel (b) is sufficient as proof that there need not be wage polarization in tasks (parameters in Table B.1). □

The intuition behind Prediction 1 is that a changing selection bias in tasks may invert the direct effect of the task prices themselves. The top row of Figure B.1 illustrates this result for a case when the price of the abstract task rises more than of the manual task and the price of the routine task falls ($\Delta(\pi_A - \pi_R) = 0.35$, $\Delta(\pi_M - \pi_R) = 0.10$, $\Delta\pi_R = -0.05$).[1] In Panel (a), the correlation between abstract and manual skills in the population is low ($\text{corr}(s_{Ai}, s_{Mi}) = 0$) and average wages in tasks polarize. Conversely, in Panel (b) of Figure B.1, the correlation between abstract and manual skills is high ($\text{corr}(s_{Ai}, s_{Mi}) = 0.7$) and, rather than polarizing, average wages in manual tasks fall even more than average

[1] The weak increase of the manual task price admittedly makes it easier for a relatively moderate selection effect to overturn it. In the theoretical models of Autor, Katz, and Kearney (2006) and Autor and Dorn (2013), wages in the routine task may also either rise or fall, because the least able routine workers leave for the manual task.

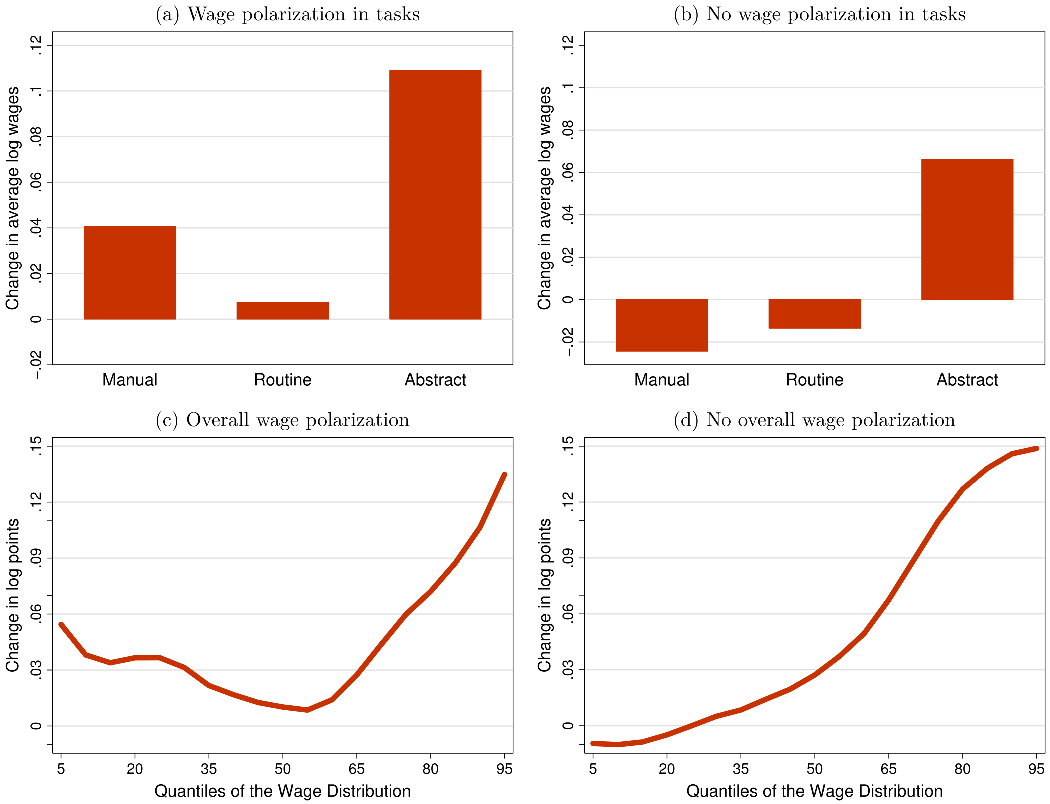

FIGURE B.1. Small differences in skill distributions can lead to qualitatively different outcomes, even under the same task price changes and a multivariate normal distribution of skills.

wages in routine tasks. The reason is that with a high correlation of abstract and manual skills, low-skill routine workers move into the manual task while high-skill manual workers move out into abstract tasks. The selection effect then dominates the price effect.

The result that RBTC need not lead to polarization of average wages in tasks is consistent with several empirical findings in the literature. In particular, during 1999–2007 employment in low-skill (service/manual-task-intensive) occupations rose strongly and at the same time wages dropped Autor (2015, Figures 2–4 and 6–7). Moreover, Autor and Dorn (2013) find that whereas employment contracted in routine-task-intensive clerical and sales occupations over 1980–2005, wages in these occupations increased. Mishel, Shierholz, and Schmitt (2013, p. 5) also conclude from their analysis that there is "little or no connection between decadal changes in occupational employment shares and occupational wage growth" in the U.S. over the last decades. Finally, in an international context, employment in low-skill occupations increased in the U.K. and Canada, while at the same time wages in these occupations dropped compared to routine occupations (Goos and Manning (2007), Green and Sand (2015)).[2]

[2]Adermon and Gustavsson (2015) find that RBTC cannot explain changing wages between occupations, but that it does have explanatory power for wages within occupations in Sweden.

Prediction 2 (Negative). *Task price polarization does not imply (overall) wage polarization.*

Sketch of proof of Prediction 2. Focus on the lower half of the wage distribution. Consider manual task worker m and routine task worker r who are initially located at the 10th and 50th percentile of the wage distribution. For simplicity, assume they do not switch tasks. If they stay at their original quantiles, the relative change in the quantiles becomes $\Delta w^{10} - \Delta w^{50} = \Delta \pi_M - \Delta \pi_R > 0$, that is, we observe wage polarization. However, suppose the manual worker overtakes the routine worker (he benefits from the higher price change for the manual task) and that they exchange positions in the wage distribution. In this case, the relative change in quantiles becomes $\Delta w^{10} - \Delta w^{50} = \underbrace{(w_{r1} - w_{m1})}_{-} + \underbrace{(w_{r0} - w_{m0})}_{+}$, which is flatter and may even be negative. Figure B.1 Panel (d) provides a counterexample where the overall wage distribution does not polarize (parameters in Table B.1). □

The idea behind Prediction 2 is that even if manual task workers are on average located at lower quantiles of the wage distribution than routine workers, it does not mean that these lower quantiles will rise more than the routine workers' quantiles. First, this is because, in the actual data, routine workers are also strongly represented in the lower end of the wage distribution (see Figure A.1). Second, with RBTC some of the manual task workers will move up in the wage distribution and overtake some of the routine task workers. Thus, not only manual task workers' initial quantiles will rise, but also the quantiles where they end up in (and vice versa for the routine task workers).[3] Empirically, this "overtaking effect" is to a greater or lesser degree always part of a change in the overall wage distribution. However, it is often assumed away in theoretical models by making workers' skill ranking one-dimensional. Such a restriction implies that wage polarization immediately follows from task price polarization.[4]

Generally, this is not the case, which is illustrated in the second row of Figure B.1. The relative increase in the manual task price is now assumed to be substantially higher ($\Delta(\pi_M - \pi_R) = 0.30$) than above ($\Delta(\pi_M - \pi_R) = 0.10$), since otherwise it would be hard to generate any wage polarization at all due to overtaking. In Panel (c), the variance of the routine skill in the population is high ($\text{var}(s_{Ri}) = 1.7$), which leads to a relatively large

[3] Rising abstract task prices have a compounding effect for inequality at the top of the wage distribution. They raise abstract workers' already high initial quantiles as well as the even higher quantiles that these workers end up in.

[4] For example, Acemoglu and Autor (2011) assumed a fixed ranking of skill between individuals whereby high-skill workers have an absolute advantage in all tasks over middle-skill workers who in turn have an absolute advantage in all tasks over low-skill workers. Cortes (2016) made a related assumption with a continuous distribution of skill. Focusing on the lower half of the wage distribution, Autor, Katz, and Kearney (2006) and Autor and Dorn (2013) assumed that high-school (or low-skill) workers all have homogenous skills in the manual task and thus are ranked one-dimensionally by their heterogenous skills in the routine task. In none of these papers, by assumption, can a worker who initially earned less than another worker overtake that latter worker in the wage distribution when the relative price of the task that he has a comparative advantage in rises.

difference in initial wages between routine and manual workers, and thus little overtaking when task prices change. Wages at the lowest quantiles of the wage distribution therefore increase compared to the quantiles located toward the middle. In Panel (d), because of lower routine-skill variance ($\text{var}(s_{Ri}) = 1.1$), initial wage differences between routine and manual workers are not as large. This leads to substantial overtaking when task prices change and an increase in wage inequality across-the-board instead of wage polarization. Hence, even when the task price changes are the same, one may obtain overall wage polarization or not with just a small modification of the skill distribution.

The result that RBTC may or may not lead to wage polarization is consistent with several empirical findings in the literature. Both employment and the wage distribution polarized in the United States over the 1990s and early 2000s. However, only employment in manual tasks expanded in the subperiod of the early 2000s, while relative wages only increased at the top of the wage distribution compared to the middle during that period (e.g., Acemoglu and Autor (2011, Figures 7–10), Autor (2015)). In addition, as mentioned above, a couple of recent papers find that job polarization already started in the 1980s in the United States (Mishel, Shierholz, and Schmitt (2013), Bárány and Siegel (2018)), although we know that wage inequality rose firmly across-the-board during this period (as in Panel (d) of Figure B.1). Finally, and probably most importantly, there is strong evidence for job polarization in other advanced countries, while there exists hardly any evidence for overall wage polarization in those countries.[5]

Earlier versions of this paper showed that even job polarization does not need to occur under RBTC. In fact, employment in only either abstract or manual tasks has to rise relative to routine tasks. The assumptions on the skill distribution to achieve this (some strong negative correlations between skills) are unrealistic, however, and job polarization is indeed a robust empirical fact across countries and time periods. Therefore, this more theoretical possibility is omitted from the current draft. Conversely, as seen in Figure B.1, it is relatively straightforward to generate a host of changes in average wages across tasks and the overall wage distribution. Many different parameter combinations are possible with three tasks, normality, and fixed task prices already, while it is ex ante not at all clear how skills in the population are distributed and that equilibrium task prices will be the same under different distributions.

Finally, it is also important to note that the argument made in Predictions 1 and 2 and in the respective simulation illustrations does not imply that differences in the effect of RBTC on the labor market need to be explained by differences in workers' skill endowments across countries and points in time. Task prices are an equilibrium outcome that depends on the interaction between production technologies, the extent and advancement of RBTC, the skill distributions, and workers' preferences (as in the generalized Roy model). All of these may differ across locations and will differ across time, and as one can verify in the simulated data, even small variations in these variables may

[5]For example, Dustmann, Ludsteck, and Schönberg (2009), Green and Sand (2015), Goos, Manning, and Salomons (2009), and Naticchioni, Ragusa, and Massari (2014) documented job polarization for Germany, Canada, and across European countries, while Dustmann, Ludsteck, and Schönberg (2009), Card, Heining, and Kline (2013), Green and Sand (2015), and Naticchioni, Ragusa, and Massari (2014) found an increase in wage inequality across-the-board for those same countries and time periods.

lead to large differences in employment, wages, and the task prices themselves. What is to be learned from Predictions 1 and 2 is therefore that RBTC and task price polarization are in principle consistent with a host of aggregate outcomes in labor markets over the last decades, while task price polarization itself is an implication that appears in all models of RBTC that have been proposed to date. The next section derives further robust implications at the individual level which follow from this.

B.2 *RBTC does make unambiguous predictions at the individual level*

The following shows that the pure Roy model of the previous section predicts higher wage growth of workers who choose abstract and manual tasks compared to workers who choose routine tasks. This result is closely related to the new method for estimating task price changes of the main text, and to the corresponding empirical tests in Acemoglu and Autor (2011) and Cortes (2016). The extension to the generalized Roy model does not affect this prediction given estimated amenity values in the data.

Consider the final worker-level Result (11) of Section 2

$$\Delta w_i \approx \Delta\pi_R + \bar{I}_{Ai}\Delta(\pi_A - \pi_R) + \bar{I}_{Mi}\Delta(\pi_M - \pi_R), \tag{B.2}$$

with tasks $k \in \{A, R, M\}$ from the empirical application and written relative to the routine task. I also start with the pure Roy model as in Acemoglu and Autor (2011) and Cortes (2016), that is, $\bar{v}_{Ai} = \bar{v}_{Mi} = \bar{v}_{Ri} = 0\ \forall i, k$. I continue to take task price polarization (B.1) as given and examine the effect on individual workers' wages in the following.

In equation (B.2), the wage growth for worker i solely depends on his initial task choice and the change in his task choice between $t = 0$ and $t = 1$. That is, relative skills in tasks determine the difference in workers' log wages at $\boldsymbol{\pi_0}$ and $\boldsymbol{\pi_1}$. Skill levels by themselves do not matter as $I_k(\boldsymbol{s_i}, \boldsymbol{\pi_t})$s are really only a function of relative skills (i.e., $s_{Mi} - s_{Ri}$ and $s_{Ai} - s_{Ri}$). Equation (B.2) therefore captures a key intuition from the Roy model: when task prices polarize, individuals who work in abstract and manual tasks gain relative to individuals who work in routine tasks. This is because the former can reap the benefits from higher comparative advantage in the rising tasks according to their skills. It is also intuitive that workers who start in abstract and manual from the beginning have the strongest comparative advantage in those tasks, but that switchers into abstract and manual have stronger comparative advantage than stayers in the routine task.

Prediction 3. *Task price polarization decreases the wages of workers who start out in the routine task compared to abstract or manual starters or both.*

Sketch of proof of Prediction 3. Without loss of generality, assume that abstract task prices rise the most ($\Delta\pi_A \geq \Delta\pi_M > \Delta\pi_R$) and consider the wage growth of a worker a who starts in the abstract task, that is, $I_A(\boldsymbol{s_a}, \boldsymbol{\pi_0}) \equiv \mathbf{1}[\pi_{A0} - \pi_{k0} + s_{Aa} - s_{ka} \geq 0\ \forall k \in \{A, R, M\}] = 1$, versus of a worker r who starts in routine, that is, $I_R(\boldsymbol{s_r}, \boldsymbol{\pi_0}) = 1$. This implies that $\bar{I}_{Aa} = 1$ since we also have $I_A(\boldsymbol{s_a}, \boldsymbol{\pi_1}) \equiv \mathbf{1}[\pi_{A0} + \Delta\pi_A - \pi_{k0} - \Delta\pi_k + s_{Aa} - s_{ka} \geq$

$0\ \forall k \in \{A, R, M\}] = 1$ and thus wage growth $\Delta w_a = \Delta\pi_A$ according to (B.2). For the routine-starting worker, depending on the choice in $t = 1$, we have

$$\Delta w_r \approx \begin{cases} \frac{1}{2}\Delta\pi_R + \frac{1}{2}\Delta\pi_A & \text{if } I_A(\boldsymbol{s_r}, \boldsymbol{\pi_1}) = 1, \\ \frac{1}{2}\Delta\pi_R + \frac{1}{2}\Delta\pi_M & \text{if } I_M(\boldsymbol{s_r}, \boldsymbol{\pi_1}) = 1, \\ \Delta\pi_R & \text{if } I_R(\boldsymbol{s_r}, \boldsymbol{\pi_1}) = 1, \end{cases}$$

which is larger for switchers (i.e., ranked from top to bottom in the equation) but always smaller than $\Delta w_a = \Delta\pi_A$. This completes the proof. Notice, however, that in some circumstances it may be that $\Delta w_r > \Delta w_m$ where m is a worker who starts in the manual task, that is, $I_M(\boldsymbol{s_m}, \boldsymbol{\pi_0}) = 1$ (in particular, if r switches to task A, m does not switch, and the difference $\Delta\pi_A - \Delta\pi_M$ is sufficiently large). □

Prediction 3 will be tested in the NLSY data below. The test is analogous to Acemoglu and Autor's (2011) regression of different demographic groups' wage growth onto their initial probabilities to work in the three tasks, but the NLSY is a conceptually attractive complement to the data used in that paper. Prediction 3 provides a theoretical foundation to Acemoglu and Autor's (2011) intuitive regression specification and demonstrates that it does not depend on a specific assumption about the underlying population distribution of workers' skills (with the caveat that strictly-speaking only the relative wages of those with higher initial probabilities to work in either abstract or manual have to rise).

Prediction 3 also shows that the empirical tests in Cortes (2016) persist under a general distribution of workers' skills. As mentioned above, a corollary of Prediction 3 is that switchers into abstract or manual tasks experience higher wage growth than stayers in the routine tasks. Using panel data at the individual level, Cortes found that the wage growth of stayers in the abstract and manual tasks is higher than the wage growth of stayers in routine tasks, but he also found some evidence that wage growth of switchers out of routine tasks is higher than of stayers in routine tasks. Prediction 3 implies that these are robust empirical results in favor of RBTC at the individual level under any arbitrary distribution of workers' skills in tasks.[6]

In general, Prediction 3 does not survive the extension to the Roy model with nonzero amenities

$$\Delta w_i \approx \Delta\pi_R + \bar{I}_{Ai}\Delta(\pi_A - \pi_R) + \bar{I}_{Mi}\Delta(\pi_M - \pi_R) - \Delta I_{Ai}(\bar{v}_{Ai} - \bar{v}_{Ri}) - \Delta I_{Mi}(\bar{v}_{Mi} - \bar{v}_{Ri}),$$

since movers out of routine and into abstract or manual tasks might lose substantial non-pecuniary benefits and thus experience higher wage growth than starters (and stayers) in either abstract or manual tasks (see also discussion in Section 2.2). However, this is not an empirically relevant case here, since the estimates in the NLSY data indicate

[6]The prediction in Cortes's (2016) model that the most and least skilled workers are leaving routine for the abstract and manual tasks, respectively, does not survive the general skill distribution. This is because the Roy model implies that marginal workers (switchers) are those with the lowest comparative advantage in their tasks, but in general it does not imply that comparative and absolute advantage align (in Cortes model they align perfectly).

that both $\bar{v}_{Ai} - \bar{v}_{Ri} > 0$ and $\bar{v}_{Mi} - \bar{v}_{Ri} > 0$. Finally, the sketch of proof for Prediction 3 used the approximated wage growth (B.2) instead of the precise equation (7). Nonetheless, the result goes through also in the precise case: starters in the task with the highest price increase will simply stay there and experience wage growth that cannot be exceeded by routine-task starters' wage growth no matter where they switch to.

B.3 *Workers' wage growth in the NLSY over the 1990s and 2000s*

Theoretical Prediction 3 states that relative wages of workers who start out in routine tasks will decline when task prices polarize. Since the same individual workers are not in the NLSY79 and NLSY97, Prediction 3 has to be evaluated using predicted probabilities based on talents $\boldsymbol{x_i}$ analogous to the propensity method for estimating task prices.

Focusing only on workers' initial task choices gives a reduced-form regression equation of the form:

$$w_{it} = \alpha_0 + \alpha_1 \cdot p_A(\boldsymbol{x_i}, \boldsymbol{\pi_0}) + \alpha_2 \cdot p_M(\boldsymbol{x_i}, \boldsymbol{\pi_0}) + \alpha_3 \cdot \mathbf{1}[t=1] + \alpha_4 \cdot p_A(\boldsymbol{x_i}, \boldsymbol{\pi_0}) \cdot \mathbf{1}[t=1] + \alpha_5 \cdot p_M(\boldsymbol{x_i}, \boldsymbol{\pi_0}) \cdot \mathbf{1}[t=1] + \alpha_6 \cdot c_i + \alpha_7 \cdot c_i \cdot \mathbf{1}[t=1] + \varepsilon_{it}. \quad \text{(B.3)}$$

This regression differs from the propensity regression (18) for estimating task prices in that it uses only the period $t = 0$ task choice probabilities (i.e., fitted values for the NLSY79 and counterfactual predicted probabilities for the NLSY97 data). The Monte Carlo simulations with confounders (see equation (18′) and Appendix E) show that a variable c_i and its time interaction can be included in the regression to control for changing returns to college that may affect workers' wages aside from RBTC. According to Prediction 3, at least one of the parameters α_4 and α_5 should be positive.

Analogous to the propensity method, the first-stage and comparability assumptions need to hold in order for the regression to identify these parameters correctly. The first-stage is estimated in a multinomial logit regression and presented in Column (2) of Table A.2 (see interpretation in Section 4.4). Multinomial probit or linear probability models give similar results. The first-stage multinomial choice regression and the second stage wage regression are bootstrapped in order to obtain the correct standard errors given that $p_A(\boldsymbol{x_i}, \boldsymbol{\pi_0})$ and $p_M(\boldsymbol{x_i}, \boldsymbol{\pi_0})$ are estimates with sampling variation.

Table B.2 displays the results from the second stage regression on task propensities (B.3). Unsurprisingly, in column one a higher propensity to enter the abstract task compared to the omitted routine task is associated with a significantly higher wage. A 10 percentage point higher probability to enter the abstract task (rather than the routine one) is associated with a 3.1 log points higher wage. The reverse is true for the propensity to enter the manual task. RBTC should however change the returns to propensities over time (i.e., α_4 and α_5), which are indicated in the table by "x NLSY97". Indeed the coefficients change strongly and significantly in the direction of Prediction 3. For the propensity to enter the abstract task, the coefficient almost doubles (from 0.31 to 0.60) while the coefficient for entering the manual task rises by more than a third (from −1.65 to −0.95).

Table B.2. Returns to NLSY79 task propensities over time.

	(1) Log Wage	(2) Log Wage	(3) Log Wage	(4) Log Wage
Constant	181.15	185.17	166.56	176.50
	(4.43)	(4.33)	(1.09)	(4.96)
Constant x NLSY97	−7.90	−10.27	3.36	20.75
	(6.67)	(6.67)	(2.22)	(13.35)
Propensity abstract task	0.31	0.03		−0.02
	(0.09)	(0.09)		(0.10)
Propensity abstract task x NLSY97	0.29	0.25		0.24
	(0.11)	(0.13)		(0.13)
Propensity manual task	−1.65	−1.80		−1.77
	(0.27)	(0.26)		(0.25)
Propensity manual task x NLSY97	0.70	0.86		0.93
	(0.39)	(0.39)		(0.39)
College		19.23	24.88	
		(3.07)	(2.58)	
College x NLSY97		4.04	8.44	
		(5.34)	(4.14)	
Observations	4154	4149	4260	4149
R^2	0.09	0.11	0.07	0.13
Degree dummies x NLSY97	No	No	No	Yes

Note: The table reports OLS wage regressions of 100 times the deflated log wage on predicted propensities to the abstract and manual tasks. The propensities are estimated from the NLSY79 only according to Column (2) in Table A.2. "x NLSY97" stands for the change in the coefficient between the NLSY79 and the NLSY97. Bootstrapped standard errors (500 iterations) below the coefficients.

For illustration of the effect of different propensities to enter the three tasks, Figure B.2 plots the predictions from linear wage regressions on each propensity at a time together with their probability densities. In the upper two subfigures, the positive wage effect of a higher propensity to enter the abstract task increases further while the negative wage effect of the propensity to enter the manual task attenuates. In contrast, for the propensity to enter the routine task the already slightly negative wage effect deteriorates substantially. For individuals with a high propensity to enter the routine task, which is quite frequent in the data, real wages even decline during the two decades between the NLSY79 and the NLSY97. This is indicated by the crossing of the two lines.

An important competitor hypothesis to RBTC, discussed in Section 4.2, is a combination of SBTC together with rising consumption demand for services. Since SBTC implies a rising return to college (e.g., Acemoglu and Autor (2011)), the remainder of Table B.2 examines theoretical Prediction 3 when changing returns to education are allowed for. First, column two inserts a 4-year college dummy and its time interaction into the estimation ($\alpha_6 \cdot c_i + \alpha_7 \cdot c_i \cdot \mathbf{1}[t=1]$ in equation (B.3)). On the one hand, the level of the coefficient on the propensity to enter the abstract task drops all the way to zero, but the changes in both coefficients are remarkably stable. On the other hand, the level of the

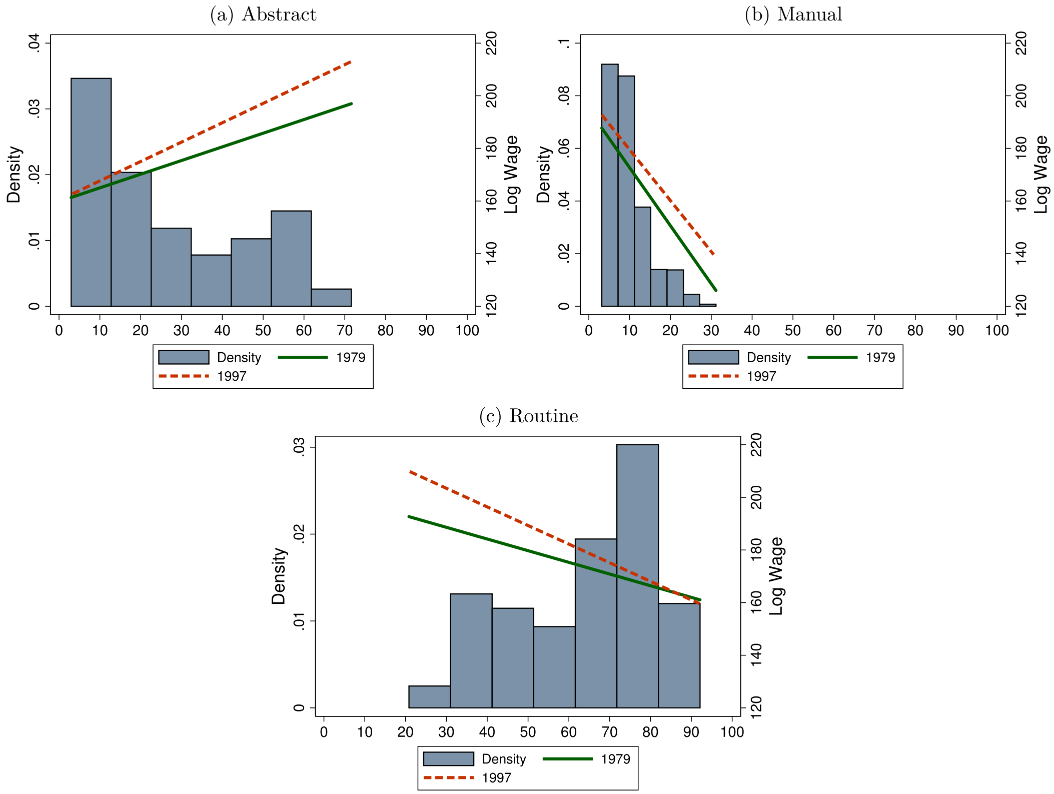

FIGURE B.2. Linear predicted returns to NLSY79 task propensities over time. *Note*: The figures plot the log wage returns to NLSY79 task propensities in the NLSY79 and the NLSY97 together with their empirical densities. The returns are estimated in regressions of log wages on a constant and the respective propensity together with an interaction term for the NLSY97 similar to equation (B.3) and Table B.2.

return to college is large and highly significant, while its change is not significantly positive once the propensities are accounted for (note that in column 3 the college dummy rises significantly).[7] The result is similar in the last column, which controls for four different degree dummies (high school dropout and graduate, some college, and at least 4-year college) and their time interactions.

These results suggests that Mincerian returns to education are important to explain wages in the cross-section, but that they (and SBTC) seem to have less power than relative skills in tasks to explain the change in wages that took place over the 20 years from the NLSY79 to the NLSY97. Prediction 3 from RBTC is therefore supported by the results in Table B.2, and rising returns to college may partly be a consequence of rising returns to abstract tasks.

[7]The increase in the college premium may even be overstated because of the net switching of workers into high college return tasks over time (see Appendix E).

Appendix C: A two tasks model for illustration

Another way of deriving equation (7) without using the envelope theorem is illustrative: for simplicity consider a case with only two tasks, abstract A and routine R. In that case, it is instructive to define choice indicators in terms of relative utilities: $I_A(U_{Ait} - U_{Rit}) \equiv \mathbf{1}[U_{Ait} - U_{Rit} > 0]$ and $I_B(U_{Ait} - U_{Rit}) = 1 - I_A(U_{Ait} - U_{Rit})$. Where there is no ambiguity, I may also write $I_{Ait} = I_A(U_{Ait} - U_{Rit})$ as a shorthand.

The change in i's realized utility $U_{i1} - U_{i0}$ when task prices (or amenities) change between $t = 0$ and $t = 1$ becomes

$$\Delta U_i = \Delta U_{Ri} + I_{Ai1} \cdot (U_{Ai1} - U_{Ri1}) - I_{Ai0} \cdot (U_{Ai0} - U_{Ri0})$$

$$= \Delta U_{Ri} + \begin{cases} (U_{Ai1} - U_{Ri1}) - (U_{Ai0} - U_{Ri0}) & \text{if } I_{Ai0} = 1, I_{Ai1} = 1, \\ (U_{Ai1} - U_{Ri1}) - 0 = U_{Ai1} - U_{Ri1} & \text{if } I_{Ai0} = 0, I_{Ai1} = 1, \\ 0 - (U_{Ai0} - U_{Ri0}) = U_{Ri0} - U_{Ai0} & \text{if } I_{Ai0} = 1, I_{Ai1} = 0, \\ 0 & \text{if } I_{Ai0} = 0, I_{Ai1} = 0 \end{cases} \tag{C.1}$$

$$= \Delta U_{Ri} + \int_{U_{Ai0}-U_{Ri0}}^{U_{Ai1}-U_{Ri1}} I_A(U_{Ait} - U_{Rit})\, \mathrm{d}(U_{Ait} - U_{Rit}). \tag{C.2}$$

The step from (C.1) to (C.2) is an educated guess. Since the relative gain in (C.1) depends on the utility range for which worker i chooses the abstract ($I_{Ait} = 1$) and routine ($I_{Ait} = 0$) tasks, respectively, this suggests an integral over the indicator function.

Check whether (C.2) is correct: if the worker always chooses abstract, he gets ΔU_{Ai} (first row of (C.1)). If he never chooses abstract, he gets ΔU_{Ri} (fourth row of (C.1)). If he switches from routine to abstract, he gets

$$\Delta U_{Ri} + \int_{U_{Ai0}-U_{Ri0}}^{0} 0\, \mathrm{d}(U_{Ait} - U_{Rit}) + \int_{0}^{U_{Ai1}-U_{Ri1}} 1\, \mathrm{d}(U_{Ait} - U_{Rit}) = U_{Ai1} - U_{Ri0},$$

which is the second row of equation (C.1). Equivalently, if he switches from abstract to routine, he gets the third row of (C.1). Hence, since within tasks the utility gain is constant, the gain (relative to the baseline ΔU_{Ri}) for a $R \to A$ switching worker depends solely on the "distance" of the adjustment that the worker is still in the routine $(0 - (U_{Ai0} - U_{Ri0}))$ and already in the abstract $((U_{Ai1} - U_{Ri1}) - 0)$ task. Notice that I never used the envelope theorem in this derivation. Also, the population distribution of skills does not appear here as equation (C.2) describes only one single worker's wage change when task prices change.

Finally, I can rewrite (C.2) as

$$\Delta U_i = \Delta U_{Ri} + \int_{U_{Ai0}-U_{Ri0}}^{U_{Ai0}-U_{Ri1}} I_A(U_{Ai0} - U_{Rit})\mathrm{d}(U_{Ai0} - U_{Rit})$$
$$+ \int_{U_{Ai0}-U_{Ri1}}^{U_{Ai1}-U_{Ri1}} I_A(U_{Ait} - U_{Ri1})\, \mathrm{d}(U_{Ait} - U_{Ri1})$$

$$= \int_{U_{Ri0}}^{U_{Ri1}} 1\,\mathrm{d}U_{Rit} + \int_{U_{Ri0}}^{U_{Ri1}} -I_A(U_{Ai0}, U_{Rit})\,\mathrm{d}U_{Rit} + \int_{U_{Ai0}}^{U_{Ai1}} I_A(U_{Ait}, U_{Ri1})\,\mathrm{d}U_{Ait}$$

$$= \int_{U_{Ri0}}^{U_{Ri1}} I_R(U_{Ai0}, U_{Rit})\,\mathrm{d}U_{Rit} + \int_{U_{Ai0}}^{U_{Ai1}} I_A(U_{Ait}, U_{Ri1})\,\mathrm{d}U_{Ait}, \tag{C.3}$$

where the last term is exactly the two-task analog of equation (7) in the main text.

Appendix D: Optimal minimum distance estimation

One alternative approach to identifying the task prices is based on minimum distance estimation of equation (16). This approach considers each talent in turn. I use the fact that I observe $\boldsymbol{x_i} = [x_{1i}\ \ x_{2i} \cdots x_{Ji}]'$ and that individuals have comparative advantages in occupations varying with each x_j in order to obtain overidentifying restrictions implied by changing task prices. The intuition is that the return to a talent should change depending on which task choice it predicts and how that changes.

Consider equation (12) and multiply on both sides by an element j of the talent vector, which is projected off the other $J-1$ talents:

$$w_{i1}\tilde{x}_{ji} - w_{i0}\tilde{x}_{ji} \approx \sum_k \frac{I_k(\mathbf{s_i}, \pi_1)\tilde{x}_{ji} + I_k(\mathbf{s_i}, \boldsymbol{\pi_0})\tilde{x}_{ji}}{2}\Delta\pi_k. \tag{D.1}$$

The projection residual $\tilde{x}_{ji}$ carries the unique information in talent j and it is mean zero. Taking expectations on both sides over all i and dividing by the variance of $\tilde{x}_{ji}$ gives

$$\Delta\gamma_j \approx \sum_{k=1}^{K} \frac{\delta_{kj0} + \delta_{kj1}}{2}\Delta\pi_k = \Delta\pi_1 + \sum_{k=2}^{K} \frac{\delta_{kj0} + \delta_{kj1}}{2}\Delta(\pi_k - \pi_1), \tag{D.2}$$

where $\delta_{kjt} = \frac{\mathrm{cov}(I_k(\mathbf{s_i}, \boldsymbol{\pi_t}), \tilde{x}_{ji})}{\mathrm{var}(\tilde{x}_{ji})}$, $\gamma_{jt} = \frac{\mathrm{cov}(w_{it}, \tilde{x}_{ji})}{\mathrm{var}(\tilde{x}_{ji})}$, and the second equality using $\sum_{k=1}^{K} \delta_{kjt} = 1$. These parameters can be recovered from OLS allocation

$$I_k(\mathbf{s_i}, \boldsymbol{\pi_t}) = \delta_{k0t} + \delta_{k1t}x_{1i} + \delta_{k2t}x_{2i} + \cdots + \delta_{kJt}x_{Ji} + v_{kit} \tag{D.3}$$

and wage regressions

$$w_{it} = \gamma_{0t} + \gamma_{1t}x_{1i} + \gamma_{2t}x_{2i} + \cdots + \gamma_{Jt}x_{Ji} + u_{it}. \tag{D.4}$$

Therefore, Result (D.2) provides an alternative approach for estimating task price changes to the propensity method of the main text. The paper's empirical application uses data on individuals' talents, their choices of entering abstract, routine, or manual tasks, and their wages in period $t=0$ (NLSY79) and $t=1$ (NLSY97). First, I run $K-1$ $(=2$ here) allocation regressions (D.3) in each period, which recover the partial correlations of the observed talents and task choices δ_{kjt}. Second, I run $K-1$ wage regressions (D.4) for $t=0$ and $t=1$, which recover the partial correlations of the observed talents and wages γ_{jt} in each period. Then, according to condition (D.2), the change of a talent's effect on the wage equals its average effect in the allocation regressions times the change in relative prices.[8]

Condition (D.2) is in fact very intuitive. The return to a talent x_j should change by the extent to which, conditional on the other talents, it increases the probability to work in tasks for which prices increase (i.e., δ_{Aj0} and δ_{Mj0} in the application), and the extent to which this association increases over time (i.e., $\delta_{Aj1}-\delta_{Aj0}$ and $\delta_{Mj1}-\delta_{Mj0}$). To assess the validity of the RBTC hypothesis in the data, one could thus simply check whether the returns changes to individual talents line up with what their allocation coefficients imply.

However, a more encompassing estimation and test of the model recognizes that condition (D.2) should hold for all J talents at the same time. Thus, as long as there are at least as many talents that predict (relative) choices as there tasks (i.e., $J \geq K-1$) the task prices can be estimated. If there are more talents than that which fulfill the first-stage Assumption (i) (i.e., $J > K-1$), the resulting overidentification yields a straightforward test of the model's restrictions in (D.2) for talent allocation and returns.

The first step in such a test is to implement a minimum distance estimator for the implied relative task price changes. Define $\bar{\delta}_{kj} \equiv \frac{\delta_{kj0}+\delta_{kj1}}{2}$, and stack $\Delta\gamma_j$ and $\bar{\delta}_{kj}$ into $J \times 1$ vectors. Then, using the allocation and wage regression estimates $\hat{\Delta\gamma}$ and $\hat{\bar{\delta}}_k$ and defining the $J \times 1$ vector $m(\Delta\boldsymbol{\pi}) = \hat{\Delta\gamma} - \iota\Delta\pi_1 - \sum_{k=2}^{K} \hat{\bar{\delta}}_k \Delta(\pi_k - \pi_1)$, this estimator minimizes

$$Q(\Delta\boldsymbol{\pi}) = m(\Delta\boldsymbol{\pi})' W m(\Delta\boldsymbol{\pi}) \tag{D.5}$$

with respect to the $\Delta\pi_k$s. The weighting matrix $W = [\text{Var}(m(\Delta\boldsymbol{\pi}))]^{-1}$ yields the Optimal Minimum Distance (OMD) estimator of the task prices. The OMD can be implemented by a (feasible) GLS regression. Just as GLS the OMD is asymptotically optimal and it yields consistent estimates of the task price changes $\Delta\pi_k$. Moreover, the objective function (D.5) in optimum can be shown to be asymptotically chi-squared distributed with $J-K$ degrees of freedom:

$$Q(\hat{\Delta\boldsymbol{\pi}}) = m(\hat{\Delta\boldsymbol{\pi}})' \big[\text{Var}\big(m(\hat{\Delta\boldsymbol{\pi}})\big)\big]^{-1} m(\hat{\Delta\boldsymbol{\pi}}) \overset{a}{\sim} \chi^2(J-K).$$

This provides an overall test of the cross-equation restrictions implied by the model ("J-test"), the results of which for the empirical application are reported in the main text.

In the generalized Roy model, the individual wage change equation (11) multiplied by projection residual $\tilde{x}_{ji}$ becomes

$$w_{i1}\tilde{x}_{ji} - w_{i0}\tilde{x}_{ji} \approx \sum_k \bar{I}_{ki}\tilde{x}_{ji}\Delta\pi_k - \sum_k \tilde{x}_{ji}\Delta I_{ki}\bar{v}_{ki}. \tag{D.6}$$

[8] Note that the literature on SBTC has also run linear wage regressions on test scores (e.g., Murnane, Willett, and Levy (1995)). The difference here is that the drivers of returns changes are explicitly examined in the allocation regressions and that the results are interpreted within an explicit Roy model of sorting and task prices.

When amenities are homogenous (i.e., $\bar{v}_{ki} = \bar{a}_k$), this yields the extended moment condition:

$$m(\Delta\boldsymbol{\pi}) = \hat{\Delta\gamma} - \iota\Delta\pi_1 - \sum_{k=2}^{K} \hat{\bar{\delta}}_k \Delta(\pi_k - \pi_1) + \sum_{k=2}^{K} \Delta\hat{\delta}_k(\bar{a}_k - \bar{a}_1), \tag{D.7}$$

where the $J \times 1$ vectors $\hat{\bar{\delta}}_k = \frac{\delta_{k0}+\delta_{k1}}{2}$, $\Delta\hat{\delta}_k = \hat{\delta}_{k1} - \hat{\delta}_{k0}$, and $\hat{\Delta\gamma}$ are still estimated in the same OLS allocation and wage regressions. The construction of the quadratic form (D.5) and "J-test" is then analogous to before, with the exception that now $2K - 1$ ($= 5$ in the empirical application) parameters are estimated and the degrees of freedom reduce accordingly (i.e., the test statistic under the null is asymptotically $\chi^2(J - 2K + 1)$ distributed).

Finally, when the amenities are heterogeneous by $\boldsymbol{x}_i$-types yields

$$\begin{aligned}
\mathrm{E}(\Delta w_i \tilde{x}_{ji}) &\approx \sum_{k=1}^{K} \mathrm{E}(\bar{I}_{ki}\tilde{x}_{ji})\Delta\pi_k - \sum_{k=1}^{K} \mathrm{E}(\Delta I_{ki}\tilde{x}_{ji})\bar{a}_k - \sum_k E\left(\tilde{x}_{ji}\Delta I_{ki}\sum_{j'} \bar{b}_{kj'}x_{j'i}\right) \\
&= \sum_{k=1}^{K} \mathrm{E}(\bar{I}_{ki}\tilde{x}_{ji})\Delta\pi_k - \sum_{k=1}^{K} \mathrm{E}(\Delta I_{ki}\tilde{x}_{ji})\bar{a}_k - \sum_k \sum_{j'} \bar{b}_{kj'}\mathrm{E}(x_{j'i}\tilde{x}_{ji}\Delta I_{ki}).
\end{aligned}$$

Dividing this by $\mathrm{var}(\tilde{x}_{ji})$ would aim to again write it as regression coefficients for element j of the talent vector. However, the last term on the right-hand side turns out unwieldy. First, this the expectation of the product of unresidualized $x_{j'i}$ with residualized $\tilde{x}_{ji}$ ($j' = j$ only for one of the $j' \in \{1, \ldots, J\}$ elements) and with the changing choice ΔI_{ki}. It is not clear to me how to exactly obtain this from an extension of the allocation regression (D.3), since it is not simply the interaction term of the talents as an additional regressor.[9] Moreover, even if such an extended regression were to give the correct coefficients, there would be a lot of them: J additional ones per task and element of the talent vector, that is, $J \times J \times (K - 1)$. In sum, constructing the moment condition (D.7) would be very complicated and possibly fraught with error in the case of heterogeneous amenities. I therefore refrain from it in this paper and only report the results from the propensity regression estimates in the main text.

Appendix E: Monte Carlo simulations

This section conducts Monte Carlo simulations in order to assess the ability of the propensity method to recover the correct task prices. It also summarizes the results from introducing rising returns to college, increasing college attainment, and a changing minimum wage as potential confounders. Finally, I extend the Monte Carlos to the generalized Roy model with homogeneous and heterogeneous nonpecuniary amenities. Easy-to-use Stata dofiles are posted on my website for interested readers to replicate the results and to try out alternative parametrizations (e.g., for the unobserved skills).[10]

[9] Also, dividing by $\mathrm{var}(\tilde{x}_{ji})$ does not work for this term, as this is not the variance of the regressor anymore.

[10] Go to "Code for Monte Carlo Simulations" on https://sites.google.com/site/michaelboehm1/research.

E.1 *Pure Roy model without confounders*

I simulate data that resembles what is observed in the actual NLSY. Workers possess a vector of observed talents $\boldsymbol{x_i}$, including math, verbal, and mechanical talent, which are jointly normally distributed and positively correlated among each other. A linear combination of these talents map into the (log) observable skill component in abstract, routine, and manual tasks. In particular, math and verbal load highly on abstract, mechanical loads highly on routine, and verbal loads *relatively* highly on manual. The unobservable skills in Roy-type models are often taken as normally or extreme value distributed (e.g., Heckman and Sedlacek (1985), Hsieh, Hurst, Jones, and Klenow (2019)). Therefore, I let (log) unobservable skill components be either distributed multivariate normal, again with a positive correlation among each other, or type 1 (Gumbel) extreme value. The latter is a special case of the type 2 (Frechet) extreme value distribution for which relative task prices can be estimated using multinomial logit regressions.[11]

The model has a period $t = 0$ (corresponding to the NLSY79) and a period $t = 1$ (NLSY97). Abstract and manual task prices are set to rise between these two periods by 40 and 50 log points, respectively, and routine task prices decline by 15 log points. Workers' potential wages in each task and period are the sum of the log task prices, observable skill components, and unobservable skill components as in equation (13). Workers choose the task that offers the highest potential wage in each period. I simulate the model using a small sample with 4000 individuals (2000 in each period), to mimic the NLSY, and a larger sample with 40,000 individuals (20,000 in each period), to have enough statistical power in order to identify (the likely size of) any potential bias.

There are 100 iterations. In each iteration, I estimate the first stage in each period separately using a multinomial logit regression of task choice indicators on the math, verbal, and mechanical talents $\boldsymbol{x_i}$. I then construct the average task choice propensities $\overline{p}_k(\boldsymbol{x_i}) \equiv \frac{p_k(\boldsymbol{x_i},\pi_1)+p_k(\boldsymbol{x_i},\boldsymbol{\pi_0})}{2}$ with $k \in \{A, M\}$ from the predicted values of the multinomial regressions. Finally, I regress workers' observed wages onto these propensities interacted with time (i.e., propensity regression (18)).

Table E.1 reports the results from this exercise. Under both the extreme value (upper part) and the multivariate normal distribution (lower part) of unobserved skills, the propensity method recovers the true task prices very well. The multinomial logit regression gets even marginally closer to the actual task prices if the true unobservable

[11] To be precise, the assumptions about skills are:

$$s_A = 6x_{\text{math}} + 2x_{\text{verb}} + 0x_{\text{mech}} + u_A,$$

$$s_R = 0x_{\text{math}} + 0x_{\text{verb}} + 4x_{\text{mech}} + u_R,$$

$$s_M = 0x_{\text{math}} + 1x_{\text{verb}} + 0x_{\text{mech}} + u_M$$

with

$$\begin{pmatrix} x_{\text{math}} \\ x_{\text{verb}} \\ x_{\text{mech}} \end{pmatrix} \sim N\begin{pmatrix} 0 & 1 & 0.7 & 0.5 \\ 0, & 0.7 & 1 & 0.6 \\ 0 & 0.5 & 0.6 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} u_A \\ u_R \\ u_M \end{pmatrix} \sim N\begin{pmatrix} 0 & 1 & 0.5 & 0.5 \\ 0, & 0.5 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 1 \end{pmatrix} \quad \text{or} \quad \text{type 1 extreme value.}$$

Period 0 task prices are $\pi_{A0} = 1$, $\pi_{R0} = 3$, $\pi_{M0} = 0.5$ in order to match the high and low initial employment share of the routine and manual tasks, respectively.

TABLE E.1. Monte Carlo simulations for the pure Roy model.

	Propensities Method			Multinomial Logit	
	$\Delta\pi_R$	$\Delta(\pi_A-\pi_R)$	$\Delta(\pi_M-\pi_R)$	$\Delta(\pi_A-\pi_R)$	$\Delta(\pi_M-\pi_R)$
TRUE	−15.0	40.0	50.0	40.0	50.0
Extr. value; 4000 *individ.*	−15.73	40.45	51.07	41.21	48.55
St. error of mean	(1.37)	(2.88)	(2.40)	(1.28)	(1.32)
Avg. std. error	(1.73)	(27.09)	(37.65)		
Extr. value; 40,000 *ind.*	−14.53	39.09	49.62	39.46	49.84
St. error of mean	(0.47)	(1.03)	(0.96)	(0.42)	(0.40)
Avg. std. error	(5.49)	(8.57)	(11.89)		
Multiv. norm; 4000 *ind.*	−14.91	37.13	51.25	72.39	93.99
St. Error of Mean	(1.51)	(2.97)	(2.46)	(1.98)	(1.85)
Avg. std. error	(16.78)	(26.02)	(34.91)		
Multiv. norm; 40,000 *ind.*	−14.71	40.67	49.83	74.70	91.86
St. Error of Mean	(0.50)	(1.06)	(0.72)	(0.65)	(0.60)
Avg. std. error	(5.30)	(8.23)	(10.99)		

Note: The table reports the mean estimated task prices from the propensity regression (samples with 4000 and 40,000 individuals) under extreme value type 1 and normally distributed unobservables. The standard error of the mean estimate over 100 iterations (in parentheses) and the average estimated standard error (in parentheses and italics) are also shown. For parametrization details, refer to the text and footnotes.

skills are distributed extreme value type 1. But it is off by orders of magnitude if true task prices are distributed multivariate normal. In previous versions of this paper, I also showed the converse finding, that is, that assuming multivariate normal unobservables yields incorrect task prices when true unobservables are extreme value distributed. In the propensity method, the standard error of the mean in the simulations (directly below the point estimate) does not let one reject the null hypothesis that the estimator recovers the true prices exactly, even in the large sample. This statistically supports the economic argument from Section 2.2 that the linear interpolation of workers' choices is not a problem.[12]

I have checked the robustness of these results using different loadings of the talent vector into observable skill components, different joint distributions of unobserved skills (e.g., type 2 extreme value and uniform), different levels and changes of task prices, and using multinomial probits and linear probability specifications in the first stage of the propensity regression. The simulated model also yields reasonable aggregate out-

[12]The standard deviations of the price estimates for the small sample with 4000 individuals are quite large (with 100 iterations, $\sqrt{100}$ times the standard error of the mean), and they are larger than for the multinomial logit. However, for 40,000 individuals the estimation becomes reasonably precise already, and the average standard error (reported one line below in italics) is at the same order of magnitude. In fact, in an earlier version of the paper with 2000 simulation iterations, both are very similar. This suggests that inference even in the small sample of the NLSY is not compromised.

The average standard errors here are directly from the second-stage regression (18) without bootstrapping the first and second stage together. In the actual NLSY data, I also find that these "naive" standard errors are generally very similar to the bootstrapped standard errors, so that the bootstrapping correction does not make a big difference in practice.

comes, such as polarizing employment, and, under these specific parameters and evolution of task prices, a polarization of the overall wage distribution.

E.2 *Pure Roy model with potential confounders*

The second part of the Monte Carlo simulations examines to what extent the propensity method is compromised when potentially confounding forces impact the wage distribution. A rigorous formal analysis of these confounders' impact on the estimator is outside the scope of this paper, so the Monte Carlos will be used to notify about potential threats to identification and adjustments to the propensity method that may account for them. I focus on three prominent such confounding forces here, rising returns to college, increasing college attainment, and a changing minimum wage. Labor market experience would be another skill, the accumulation or returns to which may have changed. Since this is similar to college returns and attainment, I do not report those simulations separately but mention some results in the text. Table E.2 shows the results of the propensity method with different adjustments, in order to save space only for the large sample (to precisely assess any bias of the estimator) and multivariate normality of the unobserved skills. The rest is the same as in Table E.1.

The first two panels of Table E.2 focus on changing returns to education, in particular to college attainment. Including college as a skill modifies potential wages (1) to

$$w_{kit} = \pi_{kt} + s_{ki} + \lambda_t \cdot c_i \quad \text{with } k \in \{A, R, M\}. \tag{1'}$$

c_i is a college dummy for individual i and λ_t the time-varying return to college. Assume first that c_i may be a function of observable or unobservable talents, but that the selection of talents into c_i does not change over time (i.e., it is not a function of task prices π_{kt}). Then equation (16) becomes

$$\mathrm{E}(w_{i1} - w_{i0} | \boldsymbol{x_i} = \boldsymbol{x}) \approx \Delta\pi_R + \overline{p}_A(\boldsymbol{x})\Delta(\pi_A - \pi_R) + \overline{p}_M(\boldsymbol{x})\Delta(\pi_M - \pi_R) + \mathrm{E}(c_i|\boldsymbol{x})\Delta\lambda.$$

One way to estimate these parameters is an augmented propensity regression (18):

$$\begin{aligned} w_{it} = {} & \theta_0 + (\theta_1 - \theta_0) \cdot \overline{p}_A(\boldsymbol{x_i}) + (\theta_2 - \theta_0) \cdot \overline{p}_M(\boldsymbol{x_i}) \\ & + \Delta\pi_R \cdot \mathbf{1}[t=1] + \Delta(\pi_A - \pi_R) \cdot \overline{p}_A(\boldsymbol{x_i}) \\ & \cdot \mathbf{1}[t=1] + \Delta(\pi_M - \pi_R) \cdot \overline{p}_M(\boldsymbol{x_i}) \cdot \mathbf{1}[t=1] \\ & + (\theta_3 - \theta_0) \cdot c_i + \Delta\lambda \cdot c_i \times \mathbf{1}[t=1] + \varepsilon_{it}, \end{aligned} \tag{18'}$$

where the task prices are identified by $\widehat{\Delta\pi_R}$, $\Delta\widehat{(\pi_A - \pi_R)}$, $\Delta\widehat{(\pi_M - \pi_R)}$ again, and the change in the college premium by $\widehat{\Delta\lambda}$. The first panel of Table E.2 reports the results of this estimation when college is assumed to be dependent on observables $\boldsymbol{x_i}$ and some other random factors (e.g., credit constraints), but this relationship is not changing between period 0 and period 1.[13] The table shows that not controlling for college overes-

[13] Specifically, I generate an index $c^* = f(x) + v_c = 9.0x_{\text{math}} + 4.8x_{\text{verb}} - 3.5x_{\text{mech}} + v_c$ with $v_c \sim N(0, 1 * \text{s.d.}[f(x)])$. The college indicator is set as $c = \mathbf{1}[c^* > Q_{c^*}(0.75)]$, where $Q_{c^*}(0.75)$ is the 75th percentile of the c^* distribution. This matches well the cross-sectional relationship between c_i and $\boldsymbol{x_i}$ in the actual NLSY data. The college premium rises from $\lambda_0 = 0.2$ to $\lambda_1 = 0.4$.

TABLE E.2. Monte Carlo simulations with college and minimum wage as confounders.

Confounder	Adjustment	$\Delta\pi_R$ (s.e.m.)	$\Delta(\pi_A - \pi_R)$ (s.e.m.)	$\Delta(\pi_M - \pi_R)$ (s.e.m.)	$\Delta\lambda$ (s.e.m.)
TRUE		−15.0	40.0	50.0	
Collg return rises	No adjustment	−12.36 (0.48)	50.32 (0.95)	48.39 (0.72)	
$\Delta\lambda = 20.0$	2nd-stage collg	−14.50 (0.48)	40.60 (0.97)	49.38 (0.71)	21.36 (0.85)
Collg return occ-spec	2nd-stage collg	−14.59 (0.48)	40.56 (0.97)	49.40 (0.72)	22.46 (0.86)
& $\Delta\lambda = 20.0$	2nd-st collg X occ	−13.47 (0.48)	41.37 (0.94)	49.62 (0.70)	19.18 (0.64)
Collg attnm rises	No adjustment	−14.30 (0.48)	41.16 (0.94)	49.30 (0.72)	
but $\lambda_1 = \lambda_0 = 0$	2nd-stage collg	−35.26 (0.50)	72.58 (0.94)	51.21 (0.73)	−87.04 (0.80)
Collg attnm rises	No adjustment	−2.98 (0.48)	53.05 (0.94)	46.52 (0.72)	
& $\Delta\lambda = 20.0$	2nd-st pred collg	−1.27 (0.01)	40.33 (0.13)	49.35 (0.95)	−7.74 (0.03)
Min. wage rises	No adjustment	−13.71 (0.49)	39.23 (1.06)	75.97 (0.70)	
& no disempl.	Latent wage distr	−15.22 (0.42)	39.73 (0.94)	51.31 (0.62)	

Note: The table reports the mean estimated task prices (100 iterations) under different confounding factors and adjustments made to the propensity method in the sample with 40,000 individuals and normally distributed unobservables. The standard error of the mean estimate is shown in parentheses. For detail about parametrization and the assumed confounders, refer to the text and footnotes.

timates the rising abstract task prices, because some of the rising college premium is attributed to that task. Including college in the wage regression as in (18′) solves this problem and also yields the correct college premium.

It is realistic to assume that returns to college differ across tasks. The second panel of Table E.2 imposes that $\lambda_{kt} = \lambda_k + \lambda_t$, with λ_k highest in the abstract and lowest in the routine task. This implies that workers gain from switching into tasks with higher college returns over time.[14] The second panel in Table E.2 runs regression (18′) and an augmented version, allowing for different college returns in levels $\lambda_k \cdot c_i$, $k \in \{A, R, M\}$. For the purpose of identifying the task prices, regression (18′) performs well. The changing returns to college are unsurprisingly overstated because of the gains from switching. Regression (18′) also does not do worse considering the task prices than the version accounting for $\lambda_k \cdot c_i$. This indicates once more that correct modeling of wage *levels* in (18) or (18′) is not important for estimating the changing task prices, and that including college into the wage regression will suffice even when returns to college are task specific. These conclusions that changing returns or task-specific returns can be accounted for are similar in the corresponding model with labor market history. In particular, adding years of actual work experience as a regressor into (18′) identifies the correct task prices as well as the returns to experience changes.

The bigger challenge to the propensity method arises in the case of changing selection into college over time. From now assume that c_i is still dependent on $\boldsymbol{x_i}$ but that

[14]In particular, $\lambda_A = 0.4$, $\lambda_R = 0$, $\lambda_M = 0.2$, and $\Delta\lambda = 0.2$ as above. Complete flexibility of λ_{kt} is not allowed because of the maintained assumption that the production of skills in tasks s_{ki} is time-invariant and only the task prices change (see Section B.2). The gain for a (college) worker from switching is $[(I_{Ai1} - I_{Ai0})(\lambda_A - \lambda_R) + (I_{Mi1} - I_{Mi0})(\lambda_M - \lambda_R)]c_i$.

more individuals go to college in period 0 than in period 1.[15] The model in the third panel of Table E.2 generates the (indeed plausible) case that the direct return to college is zero ($\lambda_t = 0$, $t \in \{0, 1\}$), and that all of the positive and rising college wage premium found in the data comes from it being associated with abstract tasks (i.e., the strict task model where worker characteristics are not priced directly). Intuitively, not controlling for college as in regression (18) should identify the correct task prices, because the true model return to college is zero. This is confirmed in the table, but when controlling for college the coefficient on $\Delta(\pi_{At} - \pi_{Rt})$ is severely upward-biased.[16] Therefore, the college dummy does not belong into the regression in this setting.

Of course, not including the college college dummy is also not an option when both, selection into college and the returns to college λ_t, change over time (fourth panel of Table E.2). Neither controlling for college nor not controlling for college (not reported) work here. It turns out that empirically modeling the relationship between c_i and $\boldsymbol{x_i}$ and, just as with the task propensities, using predicted college $\hat{c}_i(\boldsymbol{x_i})$ identifies the relative task prices correctly. The reason is that the relationship between $\hat{c}_i(\boldsymbol{x_i})$ and $\boldsymbol{x_i}$ does not change over time other than a level shift in $\hat{c}_i(\boldsymbol{x_i})$ from period 0 to period 1. However, this depends on the assumed relationship between c_i and $\boldsymbol{x_i}$ here. In fact, additional unreported simulations show that if the college dummy c_i depends on unobserved skills $\boldsymbol{z_i}$ instead of $\boldsymbol{x_i}$, this identification breaks down, while regressions (18) and (18′) work. My takeaway from these different settings is that the propensity method is at least partly able to account for rising returns to college and rising college attainment. What seems important is to run both specifications, including and not including college, in regression (18) and (18′) to see whether differences arise. Explicit modeling of the (endogenous) selection of observable skills into college may also be able to deal with this threat. But it is beyond the scope of this paper, which focuses on endogenous selection and changing returns to tasks, not college. Again, the results for changing actual labor market experience between the two cohorts of 27 year olds would be similar.

The last type of confounder considered is an increase in the minimum wage. This is of empirical relevance because the real minimum wage was raised substantially in the beginning of the 1990s. I follow Lee's (1999) work here. In period 0, wages are censored at the 5th percentile of the overall wage distribution and in period 1 they are censored at the 10th percentile (*Case 1, "Censoring": no spillovers, no disemployment* in Lee (1999)). Lee estimates the latent wage distribution before censoring and adjusts workers' wages below the 50th percentile to match that latent distribution. The simulations do the same,

[15]Referring back to footnote 13, now $c = \mathbf{1}[c^* > Q_{c^*}(0.75)]$ in period 0 and $c = \mathbf{1}[c^* > Q_{c^*}(0.50)]$ in period 1.

[16]To see the difference between estimation of (18) and (18′), note that $\hat{\alpha}_4 = \frac{\text{cov}(w_{i1}, \tilde{\overline{p}}_A^{(1)}(\boldsymbol{x_i}))}{\text{var}(\tilde{\overline{p}}_A^{(1)}(\boldsymbol{x_i}))} - \frac{\text{cov}(w_{i0}, \tilde{\overline{p}}_A^{(0)}(\boldsymbol{x_i}))}{\text{var}(\tilde{\overline{p}}_A^{(0)}(\boldsymbol{x_i}))}$ where $\tilde{\overline{p}}_A^{(t)}(\boldsymbol{x_i}) = \overline{p}_A^{(t)}(\boldsymbol{x_i})[1 - \frac{\text{cov}(\overline{p}_A^{(t)}(\boldsymbol{x_i}), c_i^{(t)})}{\text{var}(c_i^{(t)})}]$ is the residual of regressing $\overline{p}_A^{(t)}(\boldsymbol{x_i})$ onto the college dummy and the other regressors in (18′) (omitted for brevity). If, as is the case here, $\text{E}(c_i^{(0)} \mid \boldsymbol{x_i}^{(0)}) \neq \text{E}(c_j^{(1)} \mid x_j^{(1)})$ with $\boldsymbol{x_i}^{(0)} = x_j^{(1)}$, then the comparability Assumption (ii) with a college regressor $\tilde{\overline{p}}_A^{(0)}(\boldsymbol{x_i}) \neq \tilde{\overline{p}}_A^{(1)}(x_j)$ is violated, although without a college regressor $\overline{p}_A^{(0)}(\boldsymbol{x_i}) = \overline{p}_A^{(1)}(x_j)$ it is not. Intuitively, comparability is violated because talent selection into the residual task propensity $\tilde{\overline{p}}_A^{(t)}(\boldsymbol{x_i})$ changes between the two periods, which induces a selection bias into $\hat{\alpha}_4$.

assuming that the estimation recovers the true latent distribution. Note that workers do not get back their individual latent wages, only the overall distribution is matched. The last panel of Table E.2 reports that the estimates for the manual task price change are severely overstated without the adjustment. However, when the minimum wage is correctly adjusted for, all the task price estimates are reasonably close to the truth again. More intricate effects of the minimum wage, like cases *2 "Spillovers"* and *3 "Truncation"* in Lee (1999), are beyond the scope of this paper.

To conclude from these Monte Carlo simulations under pure Roy, the basic version of the model without confounders recovers the changing task prices correctly under multiple tasks and for different joint distributions of workers' skills. Reassuringly, the linear interpolation approximation (9) is quantitatively unimportant and the OLS propensity regression (18) achieves identification. Moreover, while this paper analyses a task model and task prices, the potential confounders of rising returns to college (years of experience), increasing college attainment (changing labor market experience), and changes in the minimum wage can also partly be accounted for.

E.3 *Generalized Roy model with nonpecuniary amenities*

The last part of the Monte Carlo simulations analyzes the propensity method's performance in the generalized Roy model (without confounders). I start with the homogeneous case and assign non-pecuniary values of 50 log points to the abstract as well as 80 log points to the manual task, comparable to the estimates from the empirical application.[17] Since the amenities (but not the task prices) turn out to be quite imprecise, I report on the large sample of 40,000 individuals and 300 iterations in order to be able to identify any clear biases. The rest of the specification and parametrization are the same as in the pure Roy Section E.1.

The upper two panels of Table E.3 report the results for homogeneous amenities. In the first three columns, the task price changes are correctly and precisely estimated using generalized propensity regression (20) under both multivariate normal and extreme value distributions for the unobserved skills. In fact, even with the large sample and 300 iterations, the true parameters lie comfortably within the confidence intervals implied by the small standard errors of the mean estimates. On the other hand, the amenity estimates turn out to be rather imprecise. Particularly in column four the estimated abstract task amenity is on average 8.5 log points below and 4.5 log points above the truth in the extreme value and the multivariate normal case, respectively, but with a lot of variation around this mean. Therefore, even in 300 iterations, one cannot reject the null that the correct relative abstract task amenity is identified on average. The average estimated standard errors are also wide whereas those for the relative task prices are sufficiently tight to be useful for inference.[18]

[17]The routine task amenity is set to zero without loss of generality because only relative amenities affect workers' task choices and wages (and can thus be identified), as discussed in the main text.

[18]The amenities' standard errors are more precise with multivariate normally distributed unobservable skills and comparable to those in the empirical application (Table 3).

TABLE E.3. Monte Carlo simulations for the generalized Roy model.

	Propensities Method				
	$\Delta\pi_R$	$\Delta(\pi_A - \pi_R)$	$\Delta(\pi_M - \pi_R)$	$\bar{a}_A - \bar{a}_R$	$\bar{a}_M - \bar{a}_R$
TRUE	−15.0	40.0	50.0	50.0	80.0
Homog. Amen.; Extr. Value	−15.34	39.55	50.60	41.59	76.70
Std. error of mean	(0.29)	(0.69)	(0.34)	(8.10)	(2.63)
Avg. std. error	(5.16)	(8.06)	(6.99)	(117.88)	(60.82)
Basic prop. regr.	−19.18	42.48	50.36		
Homog. Amen.; Multiv. Norm.	−14.68	40.68	49.52	54.44	78.55
Std. error of mean	(0.27)	(0.63)	(0.29)	(5.06)	(1.49)
Avg. std. error	(4.42)	(7.37)	(6.26)	(76.97)	(37.38)
Basic prop. regr.	−19.04	43.60	49.47		
Heterog. Am.; Extr. Value	−14.65	40.05	50.03	150.13	92.94
Std. error of mean	(0.49)	(0.74)	(0.79)	(35.31)	(4.75)
Avg. std. error	(7.17)	(9.74)	(13.65)	(616.46)	(92.93)
Basic prop. regr.	−18.36	39.43	54.43		
Heterog. Am.; Multiv. Norm.	−15.19	40.37	50.51	295.59	99.94
Std. error of mean	(0.38)	(0.68)	(0.54)	(32.89)	(2.99)
Avg. std. error	(5.90)	(8.62)	(10.68)	(600.49)	(67.81)
Basic prop. regr.	−19.82	41.49	56.64		

Note: The table reports the mean estimated task prices from the generalized propensity regression in samples with 40,000 individuals and 300 iterations under extreme value type 1 and normally distributed unobservables. The standard error of the mean estimate (in parentheses), the average estimated standard error (in parentheses and italics), and estimates using the basic propensity regression (18) for comparison are also shown. The upper two panels report the case with homogeneous amenities using propensity regression (20). The bottom two panels show the case with heterogeneous amenities (idiosyncratic preferences are normally distributed) using propensity regression based on (21). For parametrization details, refer to the text and footnotes.

One reason for the dichotomy in the precision between the estimates is that the task prices are using variation in *average* sorting for identification while the amenities are using *changes* of sorting. That is, the former exploits information on every individual in the sample, regardless of whether they switch tasks, whereas the latter only uses information from the subsample of switchers. In terms of the propensities in equation (20), there is hence more informative variation in $\bar{p}_k(\boldsymbol{x})$ than in $\Delta p_k(\boldsymbol{x})$ in the data. This is also consistent with the (unreported) fact that more workers reallocate from routine to manual than from routine to the abstract task between the two periods in the simulated data,[19] and that at the same time the manual task amenity is estimated substantially more precisely (and closer to the truth on average).

The respective last rows in each panel show the results when estimating the basic propensity regression (18) for the pure Roy model in the presence of amenities. It turns out that the relative task prices are not very far off, with abstract modestly (but actually highly significantly) overestimated and manual relatively close to its true value. The estimated $\Delta\pi_R$ however is clearly lower than the true change of the routine task price, which

[19] The reason is that task prices in manual are assumed to rise more than in abstract but also that skill differences between routine and manual are not as large as between routine and abstract (see footnote 11).

suggests that the negative wage changes that are associated with flows from routine to the higher-amenity abstract and manual tasks are largely picked up by the intercept. This may be the case when there is not that much variation across worker types in the extent of these moves, in line with the above discussion about the imprecision of the amenity estimates but also with the empirical application where $\Delta p_k(\boldsymbol{x})$ does not vary that much by $\boldsymbol{x}$. These results differ, and the bias in the task prices can get substantial, in parameterizations with more extreme amenities. Nonetheless, it seems fair to conclude that the basic propensity regression may often yield qualitatively similar results to the one that accounts for the generalized Roy model, as we also see in the empirical results for the actual NLSY data.

The bottom two panels of Table E.3 show the case when amenities are heterogeneous across individuals, estimating the generalized propensity regression which additionally controls for $\sum_k \boldsymbol{b}_{\boldsymbol{k}}'\boldsymbol{x} \cdot \Delta p_k(\boldsymbol{x})$ and its time interaction implied by equation (21). I parameterize these task-specific mappings from workers' characteristics to utility such that math and verbal talents load strongly on the abstract task, and verbal loads strongly on manual tasks. Idiosyncratic preferences are distributed multivariate normal (but unreported results with extreme value idiosyncratic preferences are very similar).[20]

The task prices resulting from this estimation turn out to be precise and correct, even despite the substantial idiosyncratic preferences in tasks that cannot be controlled for directly. The relative amenity terms are now far from the truth, however, and with very large standard errors, especially for the $\bar{a}_A - \bar{a}_R$ estimate. This may not be very surprising given that the regressors $\Delta p_k(\boldsymbol{x})$ by themselves (identifying $\bar{a}_k - \bar{a}_R$) and interacted with talents $\boldsymbol{x} \cdot \Delta p_k(\boldsymbol{x})$ (identifying $\boldsymbol{b_k} - \boldsymbol{b_R}$) are both based on changes of workers' sorting and thus potentially highly correlated. The idiosyncratic preference term in equation (21), which is also a function of changes in sorting, further impedes consistent estimation of the amenity coefficients. This, and the limited variation in $\Delta p_k(\boldsymbol{x})$ discussed above, make it hard to separately identify $\bar{a}_k - \bar{a}_R$ and $\boldsymbol{b_k} - \boldsymbol{b_R}$ (not reported because of too many parameters) in finite data.

In fact, one might not be too concerned that the amenity coefficients are biased and imprecise in the bottom panels of Table E.3. The reason is that, in this most flexible model, $\bar{a}_k - \bar{a}_R$ represents only the intercept of heterogeneous worker preferences (equation (3)) and not generally the average relative amenity in task k anymore. Second, the propensity method's main focus is to correctly estimate the changing task prices, which succeeds very well in the propensity regression based on equation (21). In unreported estimations, the generalized propensity regression (20), which does not include task-specific mappings from workers' talents as controls, returns substantially biased task prices even if the mappings are actually zero (i.e., $(\boldsymbol{b}_{\boldsymbol{k}}' - \boldsymbol{b}_{\boldsymbol{R}}')\boldsymbol{x} \cdot \Delta p_k(\boldsymbol{x}) = 0,\ \forall k$). The task-specific mapping control thus helps account for the conditional expectation

[20]To be precise, the parameterizations of the talent loadings (to be interpreted relative to mechanical talent and the routine task) and the idiosyncratic preferences are

$$\begin{aligned} v_{Ai} &= 1 \cdot x_{\text{math}} + 1 \cdot x_{\text{verb}} + 0 \cdot x,_{\text{mech}} + e_{Ai} \\ v_{Mi} &= 0 \cdot x_{\text{math}} + 2 \cdot x_{\text{verb}} + 0 \cdot x_{\text{mech}} + e_{Mi} \end{aligned} \quad \text{and} \quad \begin{pmatrix} e_{Ai} \\ e_{Mi} \end{pmatrix} \sim N\begin{pmatrix} 0 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}.$$

$\sum_k E[e_{ki}\Delta I_{ki} \mid \boldsymbol{x}]$ in equation (21) because it also interacts $\Delta p_k(\boldsymbol{x})$ with $\boldsymbol{x}$. One can therefore estimate the correct task price changes even in the most general version of the Roy model with idiosyncratic amenities.[21]

Appendix F: Detailed NLSY sample construction

I use data from the National Longitudinal Survey of Youth (NLSY) cohort of 1979 and 1997. The strength of the NLSY is that it provides detailed information about individuals' background and test scores in addition to education and labor market outcomes.

Individuals' labor market outcomes are evaluated at age 27 with the NLSY79 birth cohorts of 1957–64 reaching that age in 1984–92 and the NLSY97 birth cohorts of 1980–82 reaching it in 2007–09. Table F.1 summarizes how the sample restrictions, attrition, and labor market participation reduce the sample size from 6403 to 3054 and from 4599 to 1207 males in the NLSY79 and the NLSY97, respectively. I restrict the sample to individuals who participated in the Armed Services Vocational Aptitude Battery of tests (ASVAB) in the first survey year. This restriction is necessary because the ASVAB provides measures of different dimensions of talent for each individual that are comparable over the two cohorts.

Table F.1. From the full NLSY to the analysis sample.

	NLSY79 (Birth Years 1956–1964)	NLSY97 (Birth Years 1980–1984)
Reason for exclusion		
Total males	6403	4599
Excluded oversampled white and older arrivers in U.S. than age 16	4585	4599
Birthyear > 1982	4585	2754
Type of attrition		
Ought to be present with ASVAB at age 27	4585	2754
No ASVAB excluded	4299	2081
%	94	76
Not present at age 27 excluded	3939	1737
%	86	63
Conditioned on working		
Excluded who report no or farm occupation, self-employed, and those with no wage income	3054	1207

Note: The table reports how the analysis sample is constructed from the full NLSY 1979 and 1997, and where observations are lost or need to be dropped.

[21] Notice also that the Frisch–Waugh–Lovell theorem (sometimes called regression anatomy) indeed implies that, despite their potential multicollinearity, the regressors $\Delta p_k(\boldsymbol{x})$ and $\boldsymbol{x} \cdot \Delta p_k(\boldsymbol{x})$ may sufficiently account for the amenity part of equation (21) and thus allow the correct estimation of $\Delta\pi_k$.

The participation in ASVAB is substantially lower in the NLSY97 than the NLSY79 where almost everyone participated.[22] Moreover, sample attrition at age 27 is higher in the NLSY97 than the NLSY79 and overall only 63% of the NLSY79 participated in ASVAB and are also present at age 27. This problem is known (e.g., Altonji, Bharadwaj, and Lange (2012), Aughinbaugh and Gardecki (2007)) and the attrition and nontest-participation rates in the data closely line up with those reported in the study by Altonji, Bharadwaj, and Lange (2012, henceforth ABL). The only difference is that ABL consider outcomes at the younger age of 22, and thus have slightly lower attrition rates.

In their paper, ABL note that the higher attrition rate in the NLSY97 may be partly due to NLSY97 respondents being first interviewed at ages 12–16 versus ages 14–21 for the NLSY79, and thus had more time to attrit. ABL further extensively examine the potential non-randomness of attrition and nontest-participation and its likely impact in biasing important labor market outcomes. Aughinbaugh and Gardecki (2007) did a similar exercise but focused on social and educational outcomes. Both studies find evidence that attrition is not random with respect to youths' outcomes and their backgrounds. However, Aughinbaugh and Gardecki (2007) concluded that attrition from the NLSY97 does not appear to affect inference when estimating the three outcomes at age 20 that they are considering and ABL decide that the differences between nonattriters and the whole sample are not forbidding.

Moreover, ABL carefully select the samples of NLSY79 and NLSY97 to make them comparable to one another and compute weights that adjust for attrition and nontest-participation on observable characteristics. I closely follow their procedures for constructing my own sample.[23] First, immigrants who arrived in the United States after age 16 are excluded from the NLSY79. This is done because the scope of the NLSY97 (ages 12–16) also does not include older than age 16 arrivals. Second, I exclude the economically disadvantaged whites and military supplemental samples from the NLSY79 because they were discontinued early on in the survey and thus do not provide labor market outcomes at age 27 (or for ABL's purposes). Table F.1 reports that 1818 observations are dropped by making these restrictions to the sample. For each individual, the observation that is closest to 27 years and 6 months of age is retained and labor market and final educational outcomes are measured from this observation.

ABL use a probit model to adjust the NLSY79 and NLSY97 base year sample weights to account for attrition and nontest-participation according to several observable characteristics, such as parental education, parental presence at age 14, indicators by birth year, urban and SMSA residence status, indicator variables for race and gender, and

[22] According to the NLSY97 technical sampling report (Moore, Pedlow, Krishnamurty, and Wolter (2000)), nonrespondents to the ASVAB include ineligibles, refusals, breakoffs, and computer crashes, as well as individuals who are too ill or handicapped or with a language barrier. Moore et al. (2000) found that this is higher among metropolitan youths, nonwhites, males, and 16 year olds. They argue that there is a substantial impact of nonresponse only if the proportion of nonrespondents is high and if the differences between respondents and nonrespondents are high. Sampling weights, as used in this study, can account for differences in response rates between observable characteristics like the ones mentioned above.

[23] Thus, for more information on the sample construction and for statistics on the effects of attrition, please refer to ABL in addition to the description provided here. I would like to thank Prashant Bharadwaj for providing me with their data and do-files.

an interviewer coded variable describing the attitude of the respondent during the interview. I also employ a probit model to adjust weights for attrition and nontest-participation and use the same specification and variables as ABL apart from leaving out parental presence at age 14. Alternatively, a fully stratified set of indicators for birthyear, year, sex, and race, as employed by the Bureau of Labor Statistics for weighting, yields very similar results.[24] As ABL do in their paper, I proceed from this point with the assumption that, after attrition weighting, the two NLSY samples are representative of the population of young Americans that they are supposed to cover. These samples have the size of 3939 and 1737 individuals in the NLSY79 and the NLSY97, respectively.

I follow Lemieux (2006), who uses CPS Outgoing Rotation Group data, in how I compute wages and in defining the sample of working individuals (henceforth labor supply). Hourly wages reported for the current main job are used and normalized to 1979 real values by adjusting with the PCE deflator provided by the St.Louis Federal Reserve Bank.[25] While Lemieux (2006) removed outliers with 1979 real hourly wages below $1 and above $100, I remove the high wages from $40 onward because the NLSY wage data is very inaccurate for values above this threshold.

Finally, in order to condition on working individuals, all individuals who report not to be self-employed, and who are employed in a nonfarm, nonfishing, and nonforestry occupation according to the Census 1990 three-digit occupation classification are left in the sample. This leaves me with an analysis sample of 3054 and 1207 males in the NLSY79 and NLSY97, respectively (compare Table F.1 again). As in Lemieux (2006), all of those individuals are weighted by the number of hours that they work per week on top of the sample weights that are adjusted for test participation and attrition.

References

Acemoglu, D. and D. Autor (2011), "Chapter 12—Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of Labor Economics*, Vol. 4, Part B, 1043–1171. [6, 9, 10, 11, 12, 14]

Adermon, A. and M. Gustavsson (2015), "Job polarization and task-biased technological change: Evidence from Sweden, 1975–2005." *The Scandinavian Journal of Economics*, 117, 878–917. [8]

Altonji, J. G., P. Bharadwaj, and F. Lange (2012), "Changes in the characteristics of American youth: Implications for adult outcomes." *Journal of Labor Economics*, 30, 783–828. [29]

Aughinbaugh, A. and R. M. Gardecki (2007), *Attrition in the National Longitudinal Survey of Youth 1997*. Bureau of Labor Statistics, Washington DC. [29]

[24] I thank Steve McClaskie and Jay Zagorsky for providing me with the official attrition-adjusted sample weighting program for the NLSY.

[25] Source: "Personal Consumption Expenditures: Chain-type Price Index (PCECTPI)", accessed 2012-8-14, http://research.stlouisfed.org/fred2/series/PCECTPI.

Autor, D. H. (2015), "Polanyi's paradox and the shape of employment growth." In *Federal Reserve Bank of St. Louis: Economic Policy Proceedings, Reevaluating Labor Market Dynamics*, 129–177. [8, 10]

Autor, D. H. and D. Dorn (2013), "The growth of low-skill service jobs and the polarization of the US labor market." *American Economic Review*, 103, 1553–1597. [6, 7, 8, 9]

Autor, D. H., L. F. Katz, and M. S. Kearney (2006), "The polarization of the U.S. labor market." *American Economic Review*, 96, 189–194. [6, 7, 9]

Bárány, Z. L. and C. Siegel (2018), "Job polarization and structural change." *American Economic Journal: Macroeconomics*, 10, 57–89. [10]

Card, D., J. Heining, and P. Kline (2013), "Workplace heterogeneity and the rise of West German wage inequality*." *The Quarterly Journal of Economics*, 128, 967–1015. [10]

Cortes, G. M. (2016), "Where have the middle-wage workers gone? A study of polarization using panel data." *Journal of Labor Economics*, 34, 63–105. [6, 9, 11, 12]

Dustmann, C., J. Ludsteck, and U. Schönberg (2009), "Revisiting the German wage structure." *The Quarterly Journal of Economics*, 124, 843–881. [10]

Firpo, S., N. M. Fortin, and T. Lemieux (2013), "Occupational tasks and changes in the wage structure." Working Paper. [6]

Goos, M. and A. Manning (2007), "Lousy and lovely jobs: The rising polarization of work in Britain." *The Review of Economics and Statistics*, 89, 118–133. [8]

Goos, M., A. Manning, and A. Salomons (2009), "Job polarization in Europe." *American Economic Review*, 99, 58–63. [10]

Green, D. A. and B. M. Sand (2015), "Has the Canadian labour market polarized?" *Canadian Journal of Economics/Revue canadienne d'économique*, 48, 612–646. [8, 10]

Heckman, J. J. and B. E. Honoré (1990), "The empirical content of the Roy model." *Econometrica*, 58, 1121–1149. [6]

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, 1077–1125. [20]

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019), "The allocation of talent and U.S. economic growth." *Econometrica*, 87, 1439–1474. [20]

Lee, D. S. (1999), "Wage inequality in the United States during the 1980s: Rising dispersion or falling minimum wage?" *The Quarterly Journal of Economics*, 114, 977–1023. [4, 24, 25]

Lemieux, T. (2006), "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, 96, 461–498. [30]

Mishel, L., H. Shierholz, and J. Schmitt (2013), "Don't blame the robots: Assessing the job polarization explanation of growing wage inequality." EPI Working Paper. [8, 10]

Moore, W., S. Pedlow, P. Krishnamurty, and K. Wolter (2000), "National longitudinal survey of youth 1997: Technical sampling report." Technical report. [29]

Murnane, R. J., J. B. Willett, and F. Levy (1995), "The growing importance of cognitive skills in wage determination." *The Review of Economics and Statistics*, 77, 251–266. [18]

Naticchioni, P., G. Ragusa, and R. Massari (2014), "Unconditional and conditional wage polarization in Europe." IZA Discussion Paper. [10]

Young, A. (2014), "Structural transformation, the mismeasurement of productivity growth, and the cost disease of services." *American Economic Review*, 104, 3635–3667. [7]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 11 December, 2017; final version accepted 10 October, 2019; available online 5 November, 2019.