Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Gendered Returns to Cognitive Skills in Canada Appendix

Ashley Pullman, Arthur Sweetman and Ross Finnie

Published by

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

Education+Skills RESEARCH INITIATIVE

The Research Initiative on Education and Skills is an innovative and collaborative policy-research initiative led by the Mowat Centre and the Higher Education Quality Council of Ontario. The initiative is funded by Employment and Social Development Canada and the Ministry of Training, Colleges and Universities. Its purpose is to access, analyze and mobilize data relating to the education, skills and labour market outcomes of Canadians, and to disseminate the findings to inform policy development.

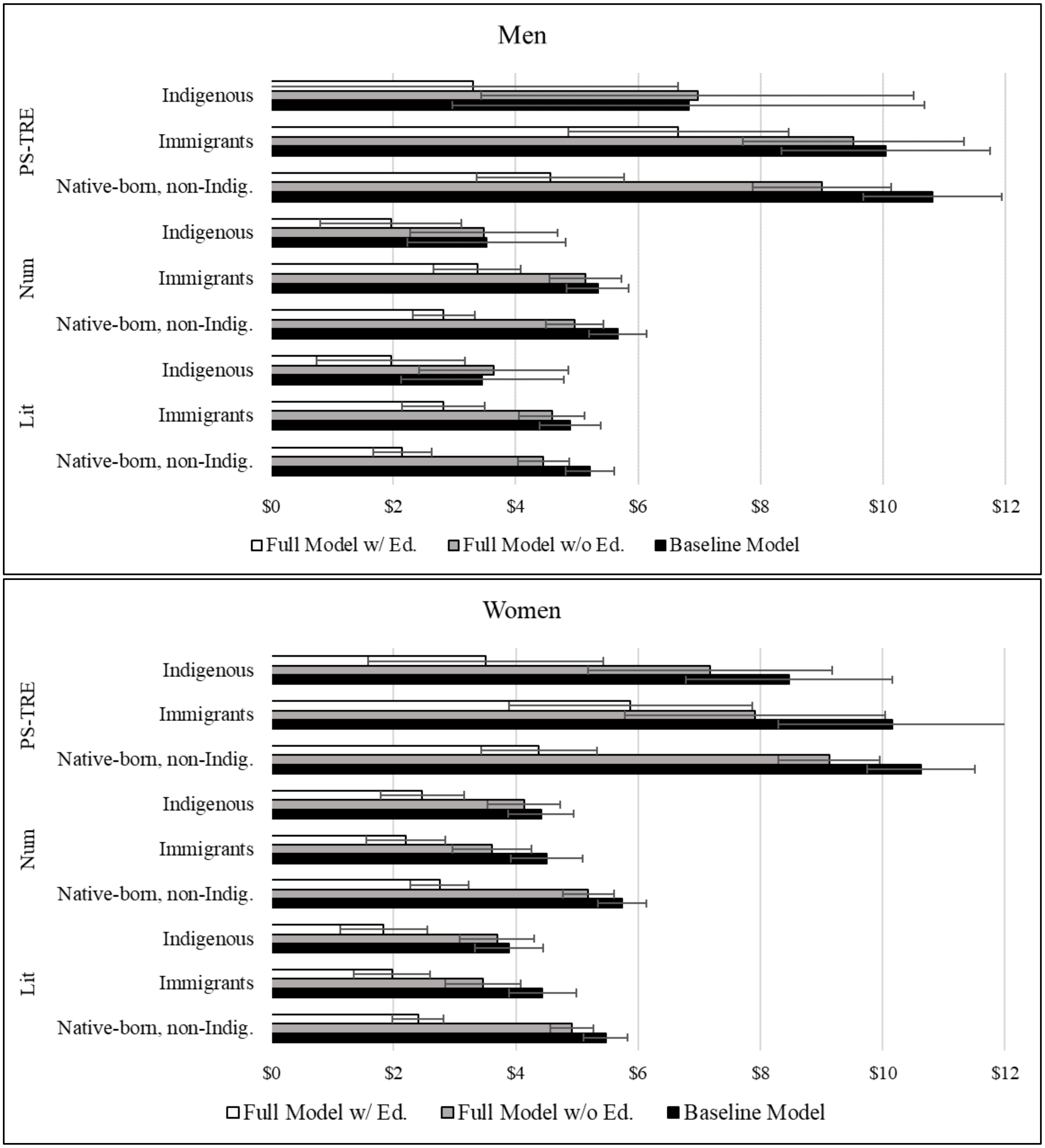
Cite this publication in the following format:

Pullman, A., Sweetman, A. & Finnie, R. (2019). *Gendered Returns to Cognitive Skills in Canada Appendix.* Toronto: Higher Education Quality Council of Ontario.

Higher Education Quality Council of Ontario

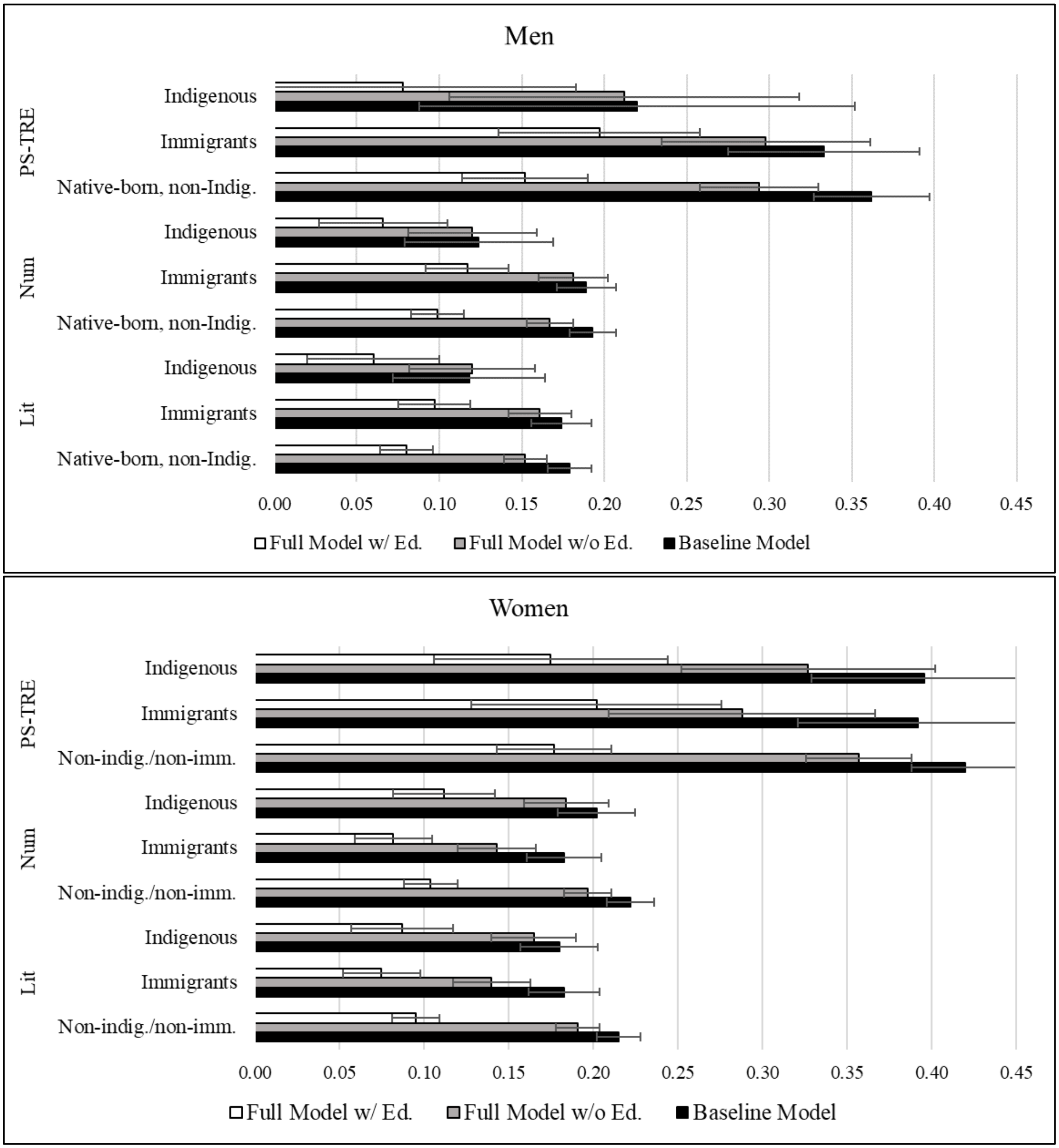
Appendix 1: Main Results

Figure 1: Wage Returns to Skills, OLS Regression Results



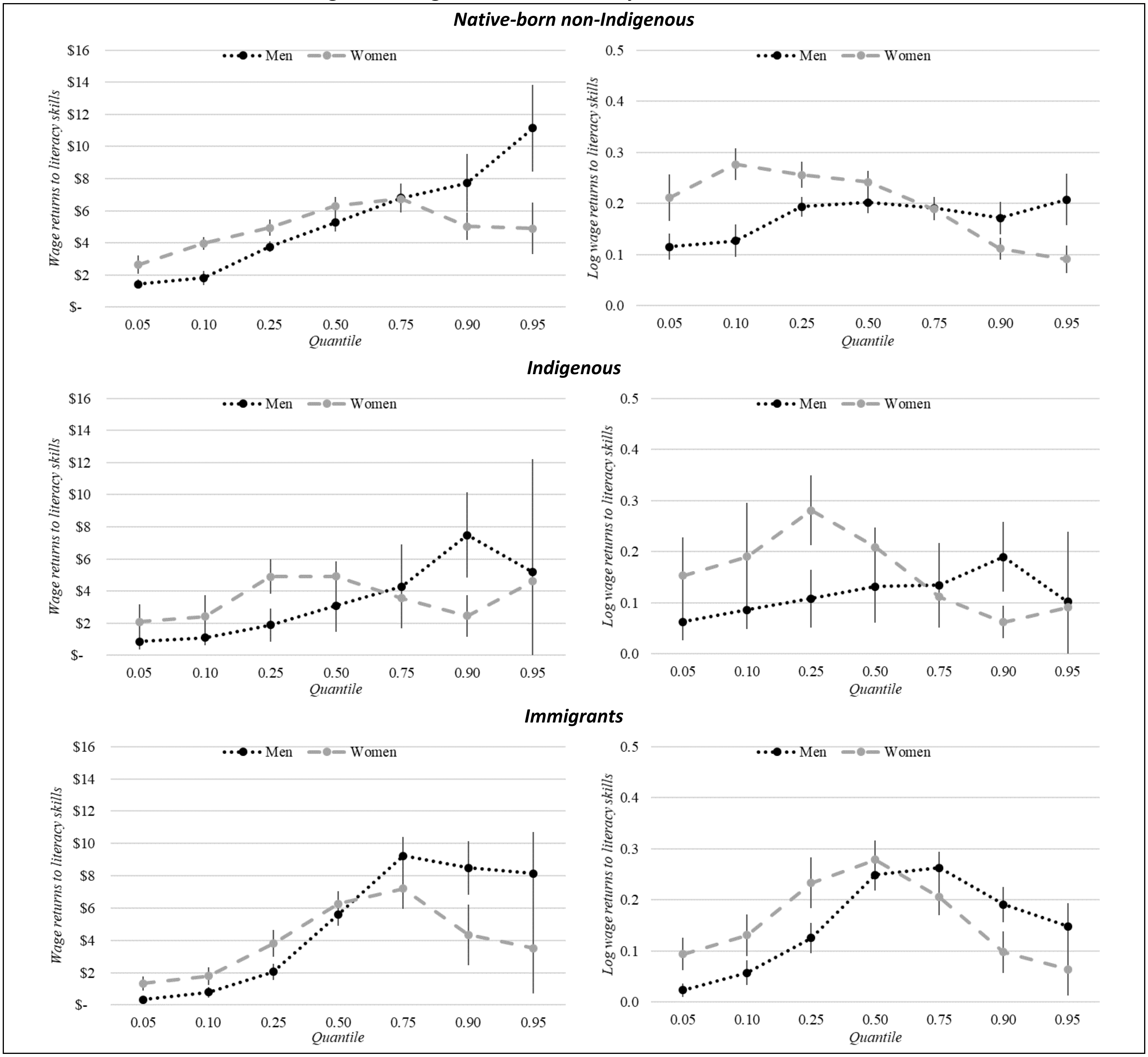
Models show untransformed wage return to literacy, numeracy and PS-TRE skills separately for men and women. There are three samples of employed PIAAC participants age 30–59 aged 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). The baseline model controls for years of full-time work experience (and its squared term), part-time employment, and (in the PS-TRE model) respondents who did not take the PS-TRE assessment. The full model controls for living with a spouse/partner, parental education, number of books at home at age 16, province of residence, rural/urban geographical location, and (in Model 3) education. The full immigrant-only models control for age of immigration to Canada and if the assessment test language was the same as a respondent's mother tongue. In the PS-TRE models, scores are recoded as for 0 for PIAAC participants who did not take the PS-TRE assessment test.

Figure 2: Log Wage Returns to Skill, OLS Regression Results



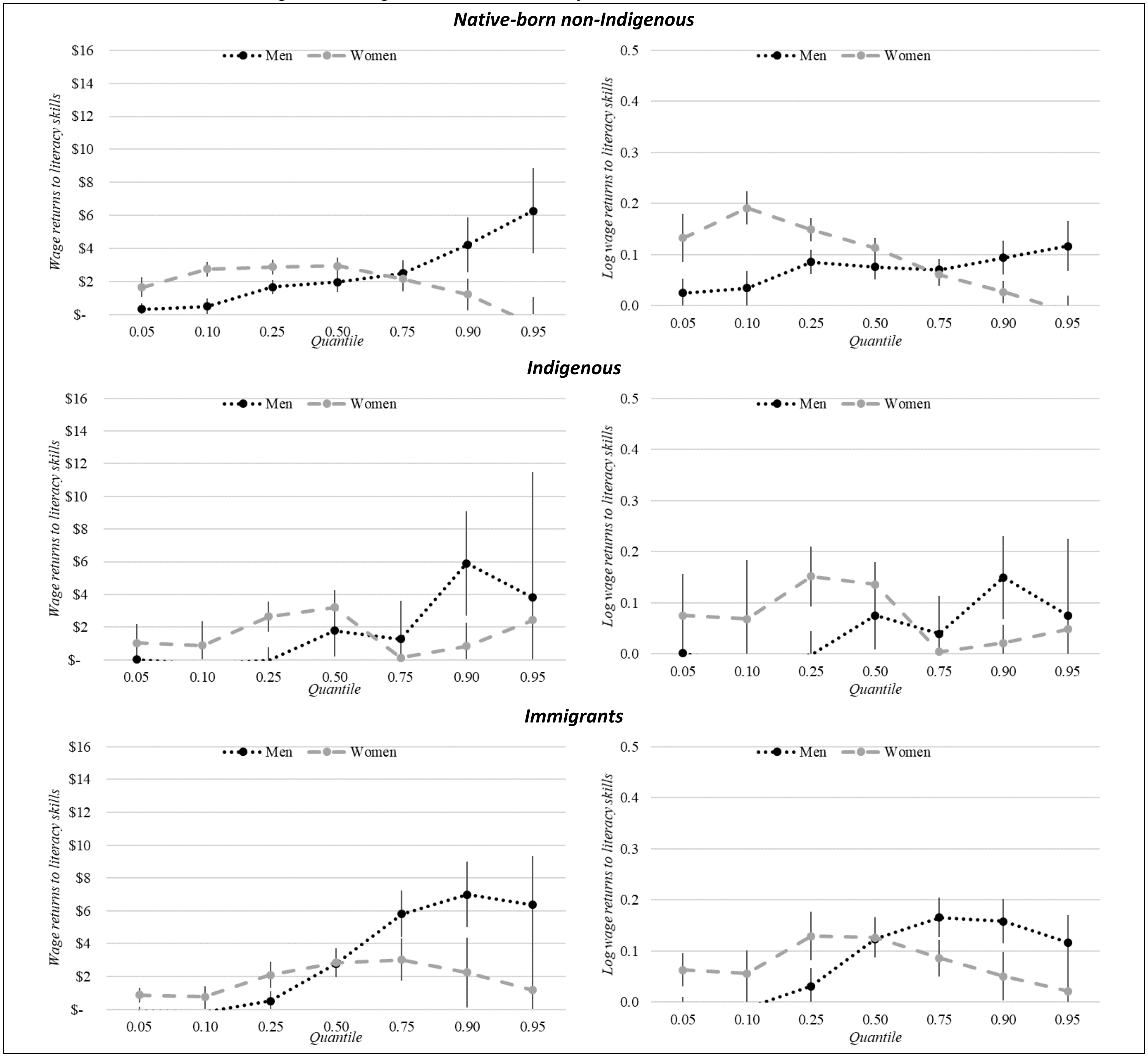
Models show log transformed wage return to literacy, numeracy and PS-TRE skills separately for men and women. There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). The baseline model controls for years of full-time work experience (and its squared term), part-time employment, and (in the PS-TRE model) respondents who did not take the PS-TRE assessment. The full model controls for living with a spouse/partner, parental education, number of books at home at age 16, province of residence, rural/urban geographical location, and (in Model 3) education. The full immigrant-only models control for age of immigration to Canada and if the assessment test language was the same as a respondent's mother tongue. In the PS-TRE models, scores are recoded as for 0 for PIAAC participants who did not take the PS-TRE assessment test.

Figure 3: Wage Returns to Literacy Skill, Baseline Model



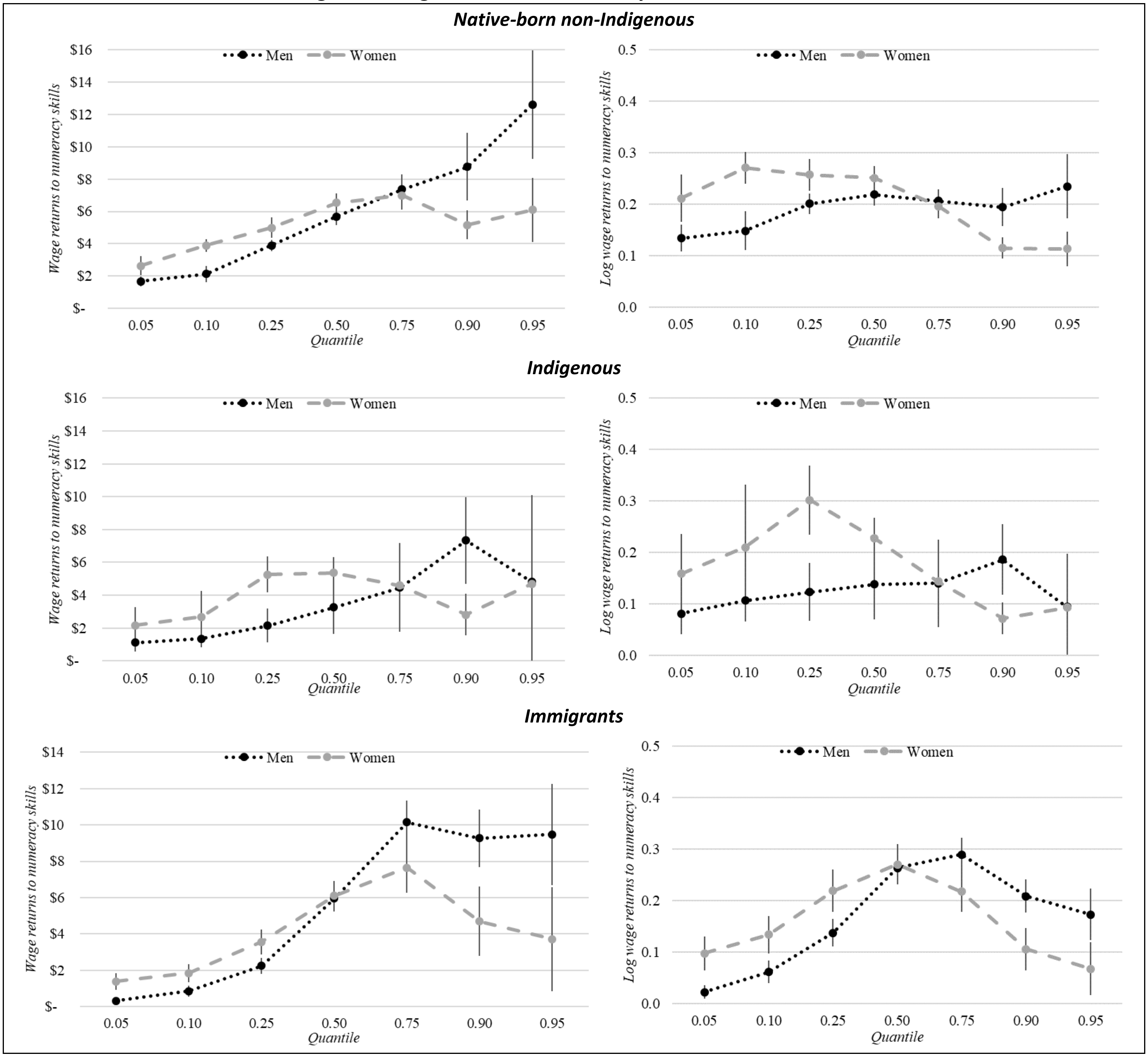
Each plot shows returns to skills for untransformed wages (left-hand side) and log transformed wages (right-hand side). There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). All models control for years of full-time work experience (and its squared term) and part-time employment status.

Figure 4: Wage Returns to Literacy Skill, Full Model with Education



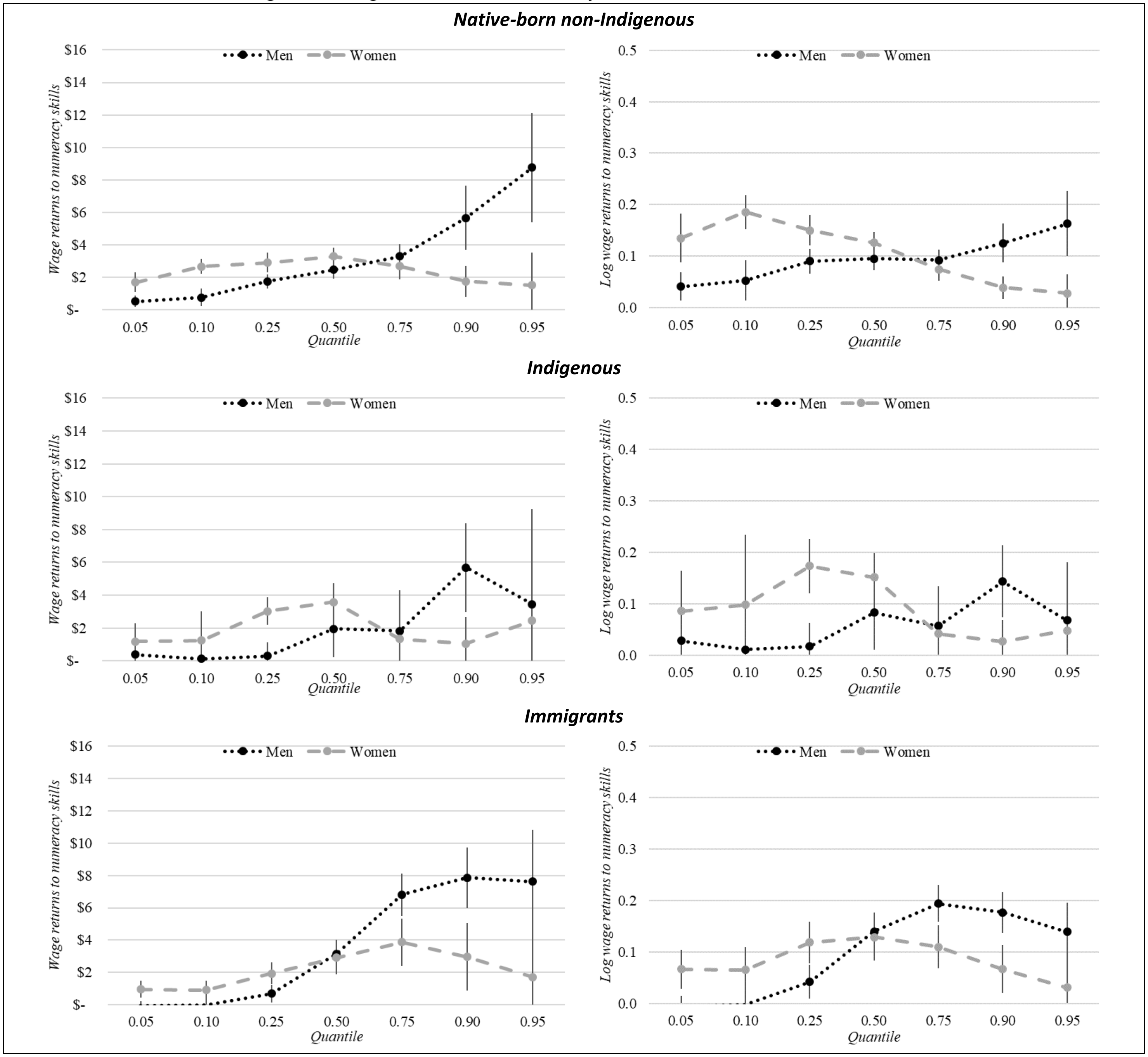
Each plot shows returns to skills for untransformed wages (left-hand side) and log transformed wages (right-hand side). There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). All models control for years of full-time work experience (and its squared term), living with a spouse/partner, part-time employment, parental education, number of books at home at age 16, province of residence, rural/urban geographical location, and education. Immigrant-only models control for age of immigration to Canada and if the assessment test language was the same as a respondent’s mother tongue.

Figure 5: Wage Returns to Numeracy Skill, Baseline Model



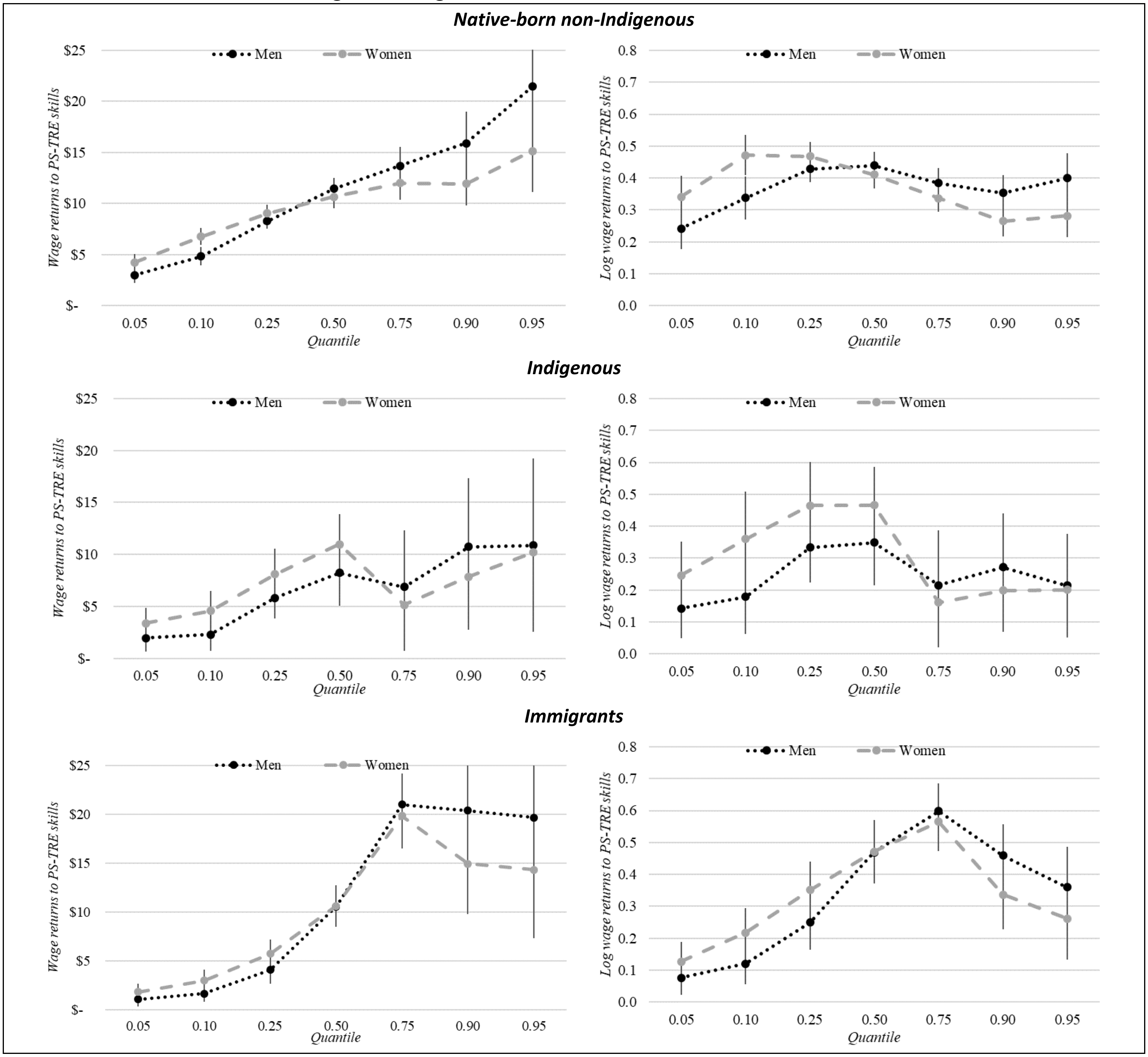
Each plot shows returns to skills for untransformed wages (left-hand side) and log transformed wages (right-hand side). There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). All models control for years of full-time work experience (and its squared term) and part-time employment status.

Figure 6: Wage Returns to Numeracy Skill, Full Model with Education



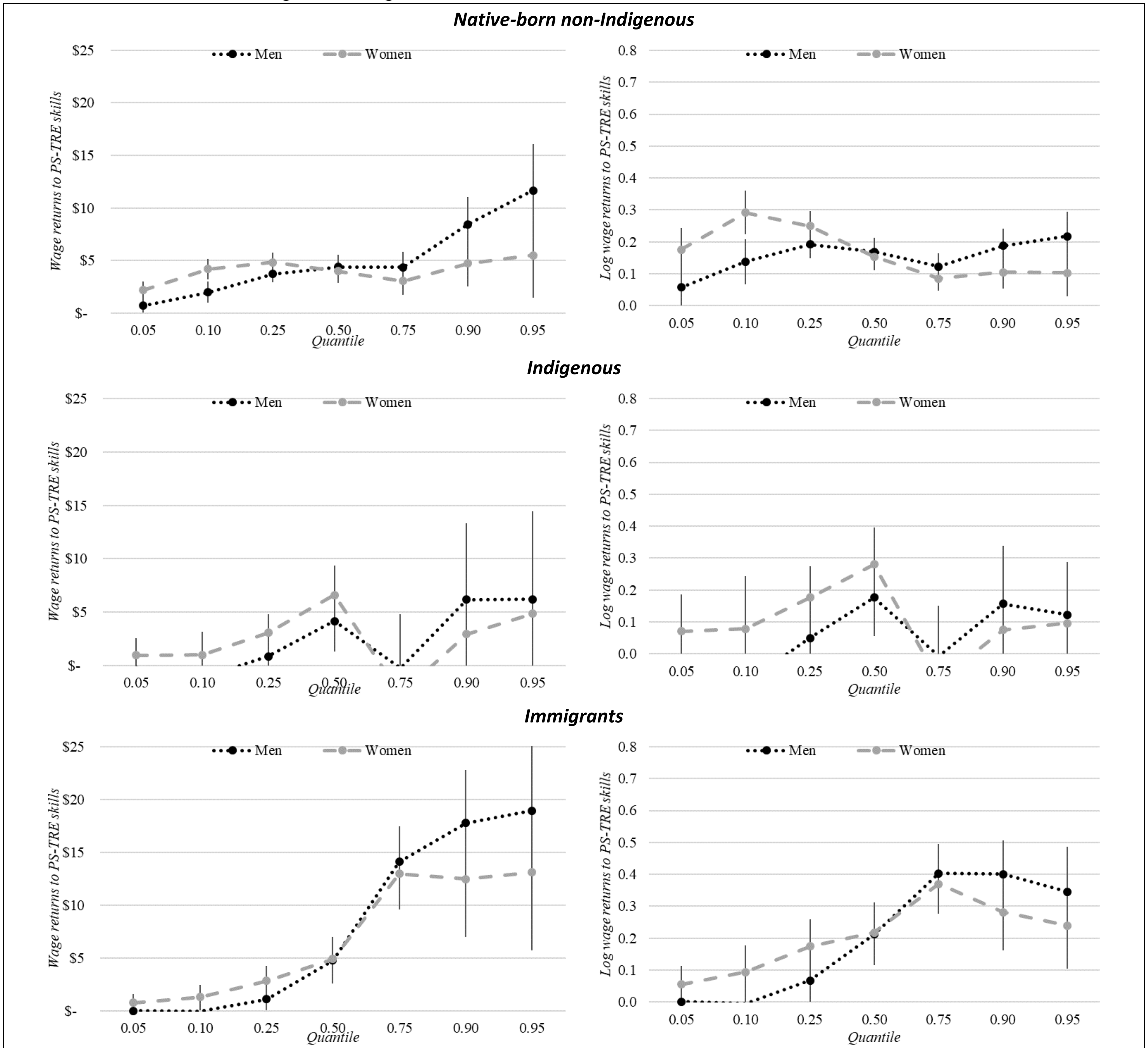
Each plot shows returns to skills for untransformed wages (left-hand side) and log transformed wages (right-hand side). There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). All models control for years of full-time work experience (and its squared term), living with a spouse/partner, part-time employment, parental education, number of books at home at age 16, province of residence, rural/urban geographical location, and education. Immigrant-only models control for age of immigration to Canada and if the assessment test language was the same as a respondent's mother tongue.

Figure 7: Wage Returns to PS-TRE Skill, Baseline Model



Each plot shows returns to skills for untransformed wages (left-hand side) and log transformed wages (right-hand side). Scores are recoded as 0 for PIAAC participants who did not take the PS-TRE assessment test. There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). All models control for years of full-time work experience (and its squared term), respondents who did not take the PS-TRE assessment, and part-time employment status.

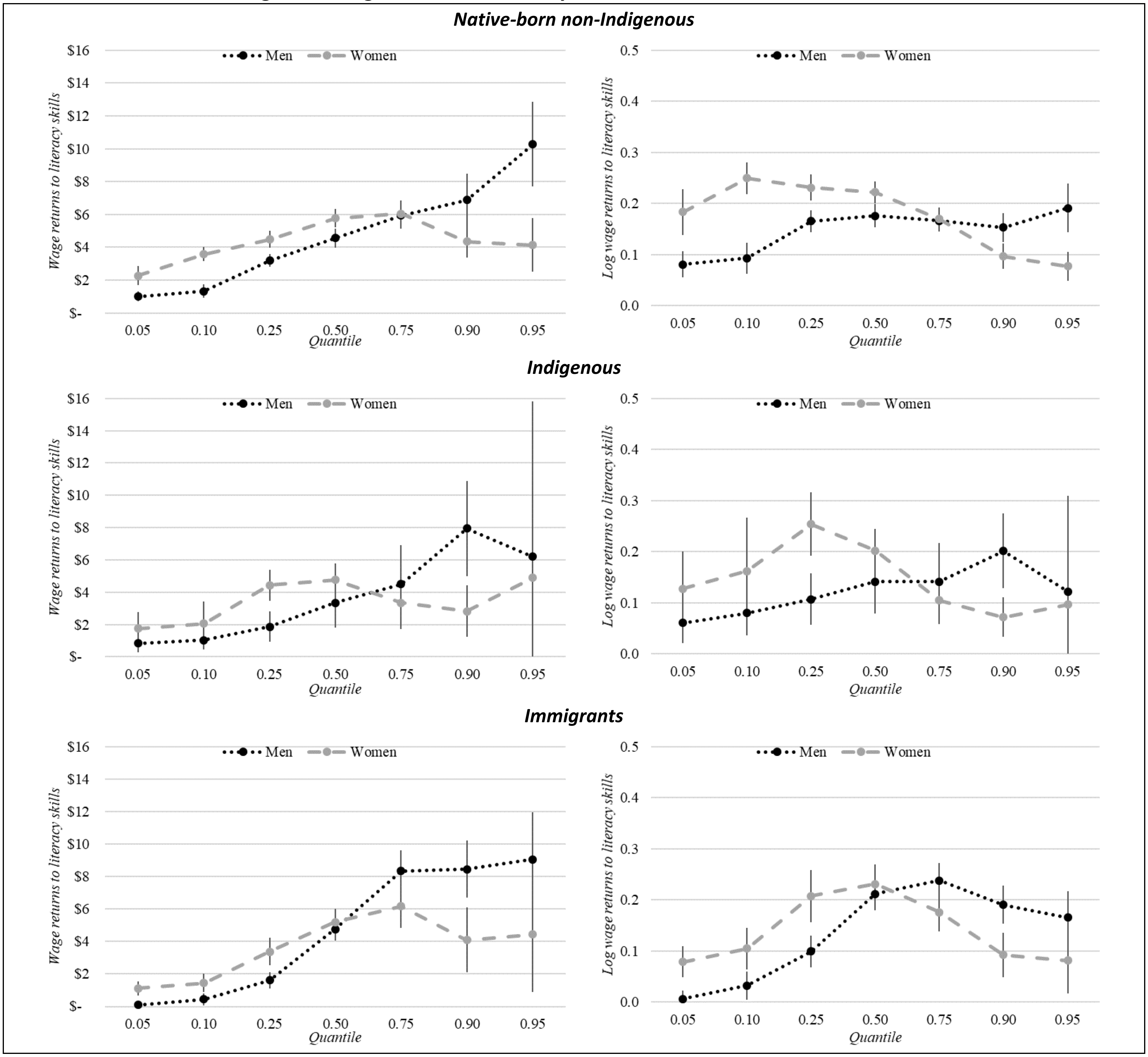
Figure 8: Wage Returns to PS-TRE Skill, Full Model with Education



Each plot shows returns to skills for untransformed wages (left-hand side) and log transformed wages (right-hand side). Scores are recoded as for 0 for PIAAC participants who did not take the PS-TRE assessment test. There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). All models control for years of full-time work experience (and its squared term), respondents who did not take the PS-TRE assessment, living with a spouse/partner, part-time employment, parental education, number of books at home at age 16, province of residence, rural/urban geographical location, and education. Immigrant-only models control for age of immigration to Canada and if the assessment test language was the same as a respondent’s mother tongue.

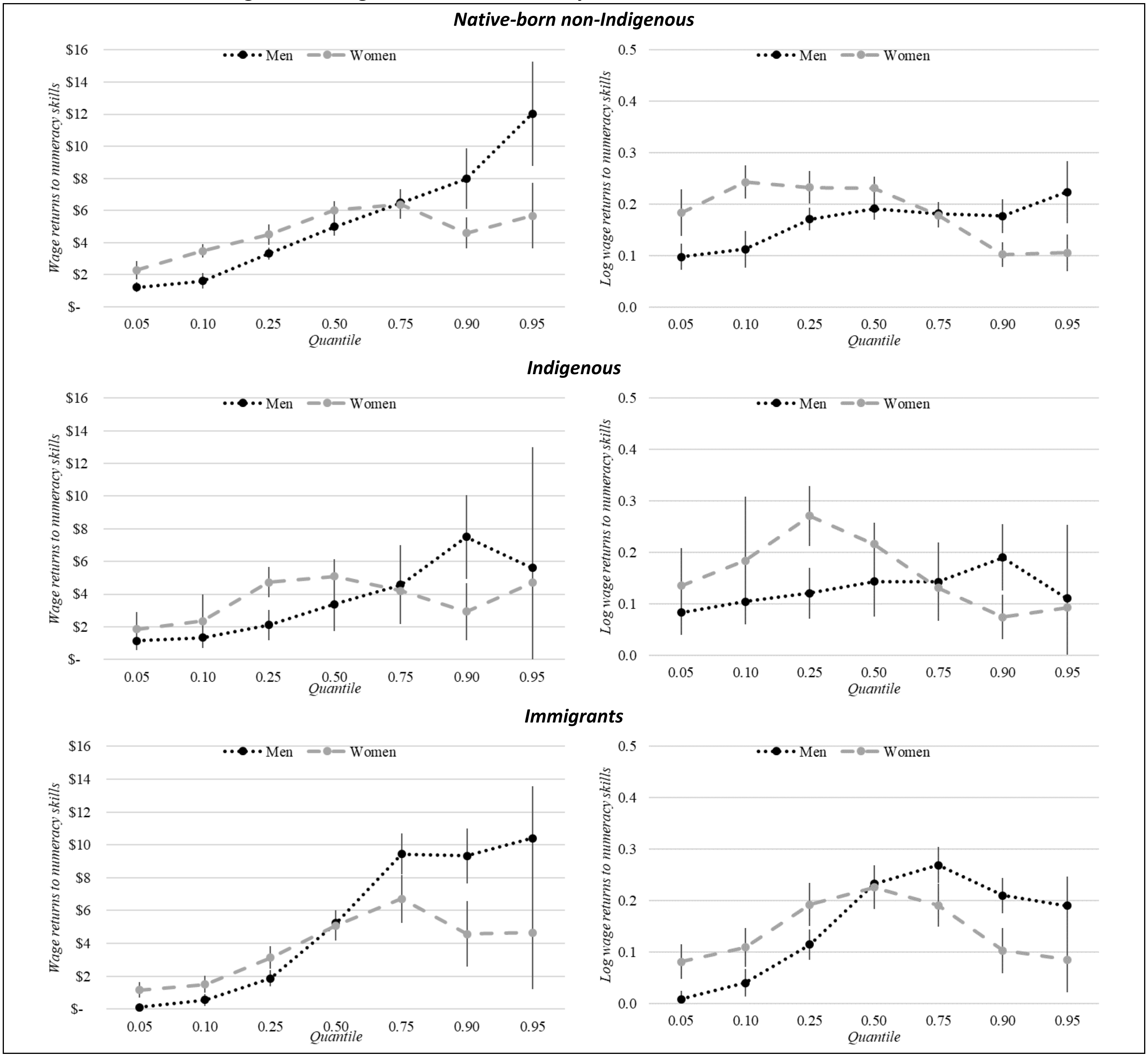
Appendix 2: Results for Full Model without Education

Figure 9: Wage Returns to Literacy Skill, Full Model without Education



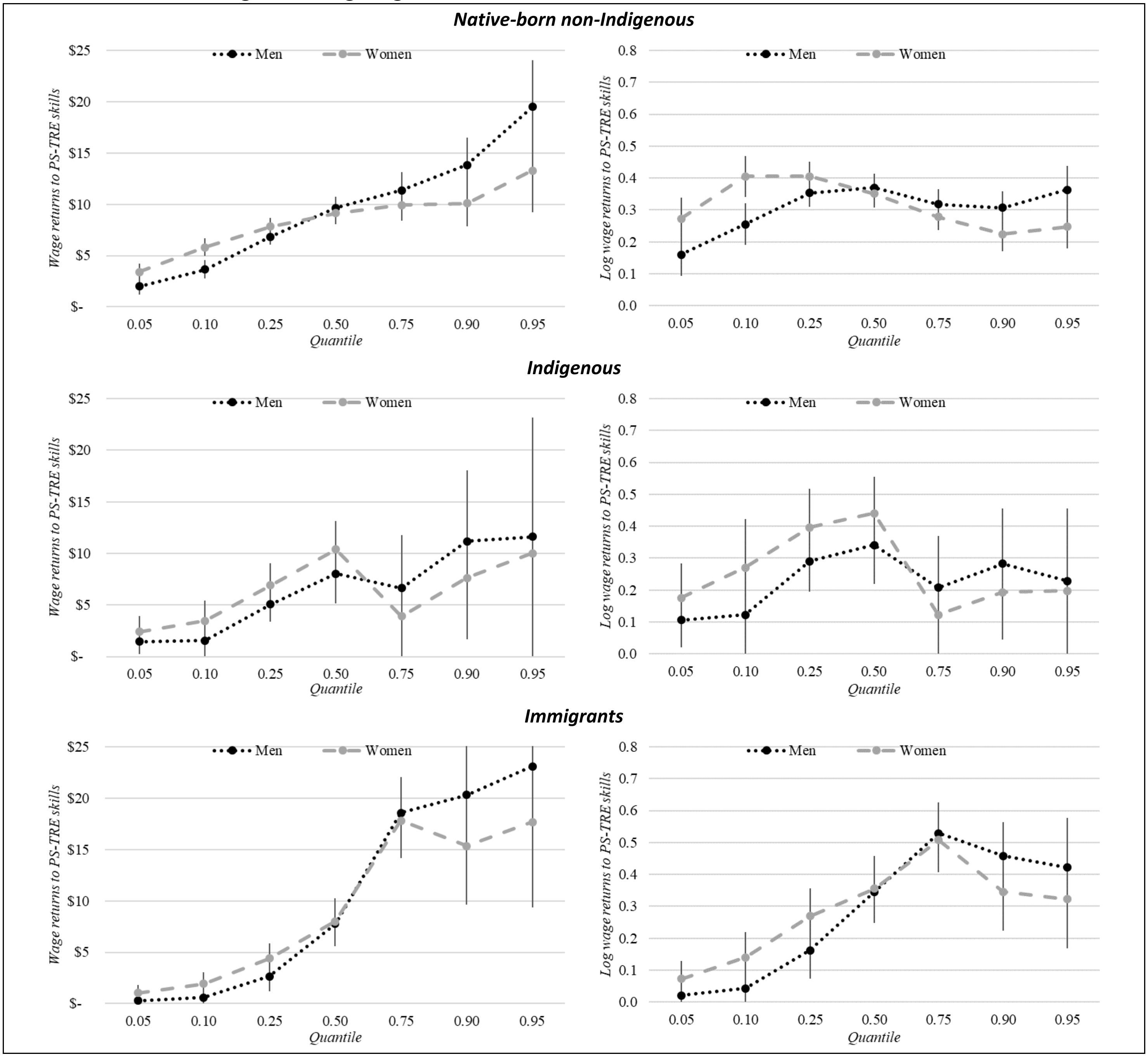
Each plot shows returns to skills for untransformed wages (left-hand side) and log transformed wages (right-hand side). There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). All models control for years of full-time work experience (and its squared term), living with a spouse/partner, part-time employment, parental education, number of books at home at age 16, province of residence, and rural/urban geographical location. Immigrant-only models control for age of immigration to Canada and if the assessment test language was the same as a respondent’s mother tongue.

Figure 10: Wage Returns to Numeracy Skill, Full Model without Education



Each plot shows returns to skills for untransformed wages (left-hand side) and log transformed wages (right-hand side). There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). All models control for years of full-time work experience (and its squared term), living with a spouse/partner, part-time employment, parental education, number of books at home at age 16, province of residence, and rural/urban geographical location. Immigrant-only models control for age of immigration to Canada and if the assessment test language was the same as a respondent's mother tongue.

Figure 11: Log Wage Returns to PS-TRE Skill, Full Model without Education



Each plot shows returns to skills for untransformed wages (left-hand side) and log transformed wages (right-hand side). Scores are recoded as for 0 for PIAAC participants who did not take the PS-TRE assessment test. There are three samples of employed PIAAC participants age 30–59: native-born non-Indigenous people (3,200 men and 3,775 women); Indigenous peoples (922 men and 1,075 women); and immigrants (1,022 men and 1,026 women). All models control for years of full-time work experience (and its squared term), respondents who did not take the PS-TRE assessment, living with a spouse/partner, part-time employment, parental education, number of books at home at age 16, province of residence, and rural/urban geographical location. Immigrant-only models control for age of immigration to Canada and if the assessment test language was the same as a respondent's mother tongue.

Appendix 3: Methodology

Data and Sample Selection

Between November 2011 and June 2012, Canada participated in round one of the PIAAC. The Canadian portion surveyed more than 27,000 residents between the ages of 16 and 65, irrespective of legal status or nationality. The target population was based on the 2011 short-form Census and National Household Survey and the response rate was approximately 58% (OECD, 2013a). Several groups were excluded from the sampling frame, although together they fell well under the OECD's maximum non-coverage rate for the target population, which was set at a maximum level of 5%.[1]

The PIAAC survey involves a computer-assisted in-person interview and a one-hour skill assessment module in the domains of literacy, numeracy and problem solving in technology-rich environments (PS-TRE). The 45-minute background questionnaire captures wages, employment status, background characteristics, and a range of social and economic information. After being sorted into computer- or paper-based assessment mediums,[2] each survey respondent then receives each skill assessment test. The questions a respondent receives are dependent on successful or unsuccessful completion of the previous item — an assessment strategy termed multi-stage adaptive testing.

Given that each person only answers a subset of assessment test items, final scores for each domain are based on the characteristics of the items answered and calculated using ten plausible values that estimate assessment score distributions in reference to both the background characteristics of an individual and other respondents who answered similar test items.[3] Calculations using plausible values account for the variance in proficiency scores that is due to imputation and the sampling error component (OECD, 2013a). In this sense, plausible values represent "the range of abilities" a person "might reasonably have" (Wu & Adams, 2002, n.p.) and thus can only measure skills at the population level. All analysis follows the OECD (2013a) standards for estimating assessment scores using plausible values and replicate weights.

[1] The sampling frame excluded individuals living on First Nations reserves, in small and remote communities, in non-institutional collective dwellings, on military bases, or in institutions (OECD, 2013a).

[2] PIAAC channels individuals who self-report no computer experience in the background questionnaire directly to paper-based literacy and numeracy modules. Those with computer experience undergo a short assessment, which, if completed unsuccessfully, also directs them into a paper-based survey. The group exempt from the computer-based assessment does not partake in the problem-solving assessment. To ensure our analysis includes this group, those with missing PS-TRE scores are given a score of 0 and a control variable representing this group (i.e., 1=missing score, 0=score not missing) is included in all PS-TRE analysis. The PS-TRE results without this modification were also produced by the authors and are available upon request.

[3] For the OLS regression analysis, plausible values, the final survey weight, and the replicate weights were handled using the svyset and miset commands in Stata and run using a user-defined program. For the quantile regression analysis, these same elements of the complex survey design were handled using the repest command in Stata.

With a focus on prime-age employees, PIAAC respondents who were not aged 30 to 59 in 2012 are excluded from the analysis. Other respondents are also removed from the analysis due to being self-employed or unemployed — two groups for which no wage information is available. Among those aged 30 to 59, 1,390 people are omitted due to missing wage information or by intentional exclusion.[4] Finally, a small number of people (*n*=79) are removed from the analysis due to missing information on the other indicators included in the models.

Because the Canadian PIAAC survey design included a large sample size overall and oversampled Indigenous peoples and immigrants, separate sociodemographic groups can be analyzed.[5] With all exclusions, all analyses examining returns to literacy, numeracy and problem solving in technology-rich environments (PS-TRE) are based on survey responses and assessment data from 6,975 native-born non-Indigenous people (3,200 men and 3,775 women), 1,997 Indigenous peoples (922 men and 1,075 women), and 2,048 immigrants (1,022 men and 1,026 women). After taking the survey design and exclusions into account, the analysis in this report is generalizable to adults age 30 to 59 who: were active in the labour force for any number of hours at the time of being surveyed; were willing to self-report their wages; did not live in locations excluded by the sampling frame; and did not earn wages that were in the top or bottom 1%.

Variables Included in the Analysis

Following Hanushek et al. (2015), the dependent variable is a continuous measure of hourly earnings plus bonuses, with the weighted top and bottom 1% of the distribution excluded. This measure is derived from questions in the PIAAC background questionnaire that asks interviewees to report their gross earnings in their current and main job at an interval of their own selection (i.e., per year, month, fortnight, week, day or hour). Interviewees who do not wish to disclose their exact earnings can also answer in broad categories. Because the questionnaire allows for different metrics of earnings, we use a derived variable that combines this information — an indicator that was generated by the PIAAC consortium. In the main findings section, we present the results for wages and log wages, with the latter measuring the exponential returns to skill rather than an absolute value.

[4] The top and bottom 1% of the weighted continuous distribution of wages are excluded from the analysis to reduce the influence of outliers and people with unusual wages (i.e., extremely low or high). People with missing wage information are also excluded, which may be due to non-response (i.e., unable or unwilling to answer), unrecorded responses (i.e., individuals who will return to work at a future date), or no earnings (i.e., unwaged family workers).

[5] Due to intra-regional data collection agreements, oversamples of Indigenous peoples only include those living off-reserve in Ontario, Manitoba, Saskatchewan, British Columbia, Yukon, Northwest Territories and Nunavut (Statistics Canada, 2013). Oversampling of foreign-born populations was also limited to Ontario, British Columbia, and Quebec. Although there is limited research examining how well these groups have been represented, the oversampling is known to draw more heavily upon those living in urban centers.

The main independent variables of interest are standardized measures of assessed scores in literacy, numeracy, and PS-TRE.[6] Because there is a strong correlation among all three assessment areas — ranging from 0.74 to 0.87 in Canada (OECD, 2013a, Table 18.5) — each domain is separately modeled. Although there is understood to be overlap, each area measures different essential information processing skills, each of which may be necessary to successfully complete everyday activities.[7]

All estimations in this report use three sets of additional independent variables. The first baseline model only includes a continuous measure of years of full-time work experience (and its corresponding squared term) and a dummy variable capturing if the respondent was employed part time (i.e., 29 hours or less per week). The second model adds a dummy variable measuring if a respondent lived with a partner or spouse; a dummy variable measuring if a respondent's mother and/or father (or female/male guardian) had a postsecondary bachelor's degree or higher; a categorical variable measuring the number of self-reported books at home at age 16 (25 books or less, 26–100 books, or 100 books or more); a categorical variable measuring a respondent's province of residence in 2012; and a categorical variable measuring if a respondent lived in a rural location in 2012 (rural, urban or missing).[8] Finally, in the third model, all control variables are retained and a categorical measure of each respondent's highest education level is added to the analysis (less than a high school diploma, high school diploma, vocational or college diploma, bachelor's degree, professional degree and graduate degree).

In general, each model specification is similar across the three sociodemographic groups (i.e., Indigenous peoples, immigrants and native-born non-Indigenous individuals) and three assessment areas; however, there are two differences to highlight. First, the OLS and quantile regression analyses measuring returns to skills for PIAAC immigrant respondents include a continuous measure of the age of immigration to Canada and a dummy variable measuring if the assessment test language was the same as a respondent's mother tongue. Second, analysis measuring PS-TRE scores includes a dummy variable capturing if a respondent's score is missing and thus recoded as zero.[9]

---

[6] Before transformation, the assessment scores, in principle, can range from 0 (low) to 500 (high). In the analysis, this scale is standardized by subtracting the mean from each scale and dividing it by its standard deviation — a common transformation in PIAAC-based research.

[7] As outlined in detail in the technical report (OECD, 2013a), each skill domain was developed by a separate working group. Although the intention was to measure distinct skills, previous research demonstrates there is a high level of correlation among large-scale assessment test domains, a finding that suggests they also likely measure general cognitive ability (Rindermann, 2007). Nonetheless, previous research generally finds a nested-factor structure best represents multiple test domains and that a single factor does not fully explain individual differences between domain scores (Gustafsson, 2016). Although there is clearly a need to further understand the PIAAC-domain intercorrelations, the majority of research models each skill area separately.

[8] The forward sortation area (i.e., the first three digits of their postal code) is used to construct an indicator representing area size. A postal code with a "0" in the second character is classified as a rural location by Canada Post. Rather than based on a strict population size, these postal locations are serviced by rural route postal drivers and/or outlets.

[9] Although prior research also recodes non-participants as having a score of zero in the PS-TRE domain (e.g., Hu et al., 2019; Truong & Sweetman, 2018), it is an imperfect procedure with limitations that are necessary to highlight. Mainly, recoding all non-participants as having a score of zero increases the overall variance in the PS-TRE domain while simultaneously decreasing variability in scores among all non-participants. That is, if non-participants did take the test, their scores would not vary and not equal zero. A methodological approach that takes selection into account may be able to assess the censoring produced by PS-TRE non-participants.

Analytical Approach

Part one of this study replicates and extend Hanushek et al.'s (2015) modelling strategy based on ordinary least squares (OLS) regression. In this model, the dependent variable (i.e., hourly wages) is related to a set of independent variables (e.g., years of full-time work experience). As the main independent variable of interest, each model includes a single measure of standardized assessment scores in either literacy, numeracy or PS-TRE — each modelled separately for men and women.[10] For each continuous skill measure, each significant and positive coefficient is interpreted as the extent to which wages increase for each additional standard deviation in assessment scores, holding all other variables constant (e.g., for a hypothetical individual with mean characteristics and mean wages).

Three separate models assess the impact of additional control variables. After replicating Hanushek et al.'s (2015) baseline model, the second linear regression model adds household, social capital and geographic indicators and the third model retains all variables and adds highest education level. This approach allows for an examination of how the returns to each skill domain change once other factors are taken into account (i.e., the "net" difference).

To understand if there is heterogeneity in returns to skills across the wage distribution, the second analysis uses *unconditional* quantile regression (Firpo, Fortin & Lemieux, 2009) to estimate how returns to all three skill domains change at select wage quantiles. For *conditional* quantile regression, the independent variables included in the model "condition" are the distribution of wages. As an example, conditional quantile regression would measure individuals in the lowest conditional quantile (i.e., at the 5th quantile) at a higher conditional wage position when the covariates included in the model largely explain their wage position.

In contrast, unconditional quantile regression defines each wage quantile without accounting for independent variables in the model using recentered influence function (RIF) regression that transforms the dependent variable. This initial step estimates a new dependent variable based on the probability that each individual in the sample earns less than the average wage amount at a selected quantile.[11] Because this first estimation step is not affected by the independent variables in the model, it addresses the critique of conditional quantile regression as based on the conditional distribution that changes with what variables are included in the model.

[10] Additional models, not shown but available upon request, formally test gender differences by combining men and women into a single model and including an interaction term between skill and gender. These results are also discussed in the findings section.

[11] This probability is divided by density at the selected quantile.

Using unconditional quantile regression, the analysis estimates returns to skills across the distribution of wages for the 5th, 10th, 25th, 50th, 75th, 90th and 95th quantile for both men and women. Separate intercepts (i.e., skill * men and skill * women) measure returns by gender in a single model. This approach is preferred to models that analyze the results for separate samples of men and women as it ensures the estimated wage quantile is the same for men and women. Namely, because of the gender wage gap, a model estimated for a female-only sample would have lower average wages for each quantile compared to a male-only sample.

Finally, while both modelling strategies are widely used, it is important to point out that the estimates they produce cannot be interpreted as representing the causal effect of a given independent variable on wages without further consideration. Notably, if they are correlated with the error term the causal effects may be biased. Thus, the present study generates important descriptive insight into the wage returns to skills, but should not be interpreted as estimating a causal relationship. That is, we have no evidence regarding the degree to which higher skills *cause* wages to increase.