I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14019

Beyond the Origin Dummy: Heterogeneity of Ethnicity and Human Capital Accumulation

Agnieszka Postepska

JANUARY 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14019

Beyond the Origin Dummy: Heterogeneity of Ethnicity and Human Capital Accumulation

Agnieszka Postepska
University of Groningen and IZA

JANUARY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Beyond the Origin Dummy: Heterogeneity of Ethnicity and Human Capital Accumulation*

Ethnic background is a well recognized complementary factor in human capital accumulation process. This paper investigates how three aspects of ethnicity affect human capital formation: group's quality, size and closeness of ties. Relying on heteroskedasticity to identify parameters in the presence of endogenous regressors, I find evidence of heterogenous effects of ethnicity for men and women. The results show that women in groups characterized by close ties benefit from high-quality ethnic environment, regardless of the group's size. In contrast, among men, the group's size appears to be of importance, and men in large groups characterized by loose ties benefit the most from higher quality of the ethnic group. The results are consistent with different socialization patterns, as evidenced by the literature.

JEL Classification: J62, J16, J15

Keywords: human capital formation, ethnicity, endogamy, networks, identification through conditional second moments

Corresponding author:
Agnieszka Postepska
Department of Economics, Econometrics & Finance
University of Groningen
P.O. Box 800
9700 AV Groningen
The Netherlands
E-mail: a.postepska@rug.nl

* I would like to thank participants of the University of Groningen seminars series for useful comments. All errors and omissions are the sole responsibility of the author.

1 Introduction

Ethnicity is a well recognized complementary factor in the human capital accumulation process, and the ethnic hierarchy in educational attainment has been well documented in the literature (see Kao & Thompson (2003) for an excellent review). However, little is known about the mechanisms through which ethnicity affects schooling outcomes. With a couple of exceptions (Gang & Zimmermann, 2000), studies concerned with ethnicity and schooling outcomes typically focus on immigrants and their children assimilation in terms of human capital and find significant differences concerning the country of origin (Gang & Zimmermann, 2000; Riphahn, 2003). In that way, at best, they can establish systematic differences between individuals of different backgrounds but fail to shed light on the mechanisms behind the relationship between ethnicity and the human capital accumulation process and capture the vast within ethnic group heterogeneity. Moreover, the country of origin is considered exogenous to schooling decisions. However, this assumption might be violated if, for instance, there are systematic differences in the quality of immigrant groups. This study aims to fill this gap in the literature and show how three group's attributes (*size*, *tightness*, and *quality*) affect the human capital formation process.

The choice of the attributes is driven by the literature (and data availability). Group's *quality* has been proven as an important predictor of human capital accumulation. In the pioneering paper Borjas (1992), ethnicity is assumed to serve as an externality in the human accumulation process. The skills of the next generation depend on parental human capital and on the quality of the ethnic environment in which parents make their investment decisions. Borjas (1992) pins down the term *ethnic capital* and finds a strong positive correlation between the average educational attainment in the parents' generation and children's educational attainment. Recent replication (Postepska, 2019) finds a much smaller yet significant effect suggesting that an increase in average schooling in the parents' generation leads to a 0.08 of a year increase in children's schooling[1]. Group's *size* may affect human capital formation as the presence of ethnic networks may help the assimilation of immigrants. Indeed, Gang & Zimmermann (2000) find a positive effect of network size on educational attainment. Lastly, the impact of ethnicity can be affected by the group's ties. The stronger the ties, the stronger the effects of the group's attributes are expected to be.

[1]However, other studies found no effect of the ethnic capital in Switzerland (Bauer & Riphahn, 2007) Canada (Aydemir *et al.*, 2013) or Denmark (Nielsen *et al.*, 2003).

Moreover, the literature finds significant differences in ethnicity's effect on the schooling process among men and women. Ethnic socialization, a concept that describes maturing to ethnic identity, has been recognized to vary significantly between boys and girls. Shah *et al.* (2010) provides an excellent example of a qualitative study that shows that the role of social capital in the process of human capital accumulation differs between young British Pakistanis women and men. While all women shared the belief passed on to them from their parents that education is the best way to upward social mobility, men did not have such a unified view of schooling's role for achieving social status. They were more subject to the influence of the majority of peers who did not share that view. Since, in general, girls are more susceptible to social influences (Eagly & Carli, 1981), they might be more likely to be isolated due to parental fear of the "bad" influence of the majority, especially as women are often held accountable for preserving the ethnic traditions and identity (Mani, 1993; Bailey, 2001). Therefore, girls are prone to a much stricter control over their brothers (Djajić, 2003; Sung, 1987; Olsen, 1997; Agarwal, 1991) - a concept phrased as *double standard* in parental monitoring by Le Espiritu (2001). This finding is consistent over time and across many ethnic groups (Dasgupta, 1998; Gupta, 1997; Williams *et al.*, 2002; Yung, 1999; Sung, 1987). As a result, girls might be more likely to have peers of the same ethnicity than boys. Moreover, such vigilant supervision has proved to positively affect schooling among Vietnamese girls (Zhou & Bankston III, 2001). Also, this could lead to stronger importance of gender roles within one's ethnic group. A high correlation between parental education implies that high average education within the ethnic group is directly related to a high average education among women within this ethnic group. Even though women are seen as the culture keepers, girls are more likely to rebel against the traditions and values, and they are more flexible in choosing ethnic identity by bridging home and host country identities (Rumbaut, 1997; Olsen, 1997).

Moreover, the literature focuses on immigrants (either first or second-generation) and the differences in schooling achievement gaps between natives and immigrants. In contrast, this study focuses on the long-term effects of native-born individuals' self-reported ethnicity to assume away the effects of immigration shock, whether of the individual herself or the parents on the schooling process. Specifically, immigrants are dropped from the sample to ensure equal access to educational institutions across individuals and level the grounds concerning barriers that especially first-generation immigrants face when navigating the

host country's education system, such as language barriers and asymmetry in information. The immigration shock has a detrimental effect on the children's schooling outcomes and the transmission of the human capital between generations (Gang & Zimmermann, 2000). However, the impact of ethnicity likely persists over subsequent generations of migrants. Studying the native-born children allows us to capture the long term effects of ethnicity net of the immigration shock.

The remaining of the paper is organized as follows. The next section discusses the data and summary statistics. Section 3 describes the model and identification strategy, while section 4 discusses the results. Section 5 concludes.

2 Data and summary statistics

I use the 1977-2018 General Social Survey data, and the final sample consists of 17021 individuals aged 18-64 born in the United States. I exclude individuals born abroad as well as native and African Americans. Since the literature found differential effects of parental education transfer on native and foreign-born children, I focus exclusively on native children. By doing so, the results of this analysis should be seen as a long-term effect of ethnic identity, and a larger role of unobserved characteristics is expected. The ethnicity of individuals was obtained through an open-ended question in which respondents were asked about the country or the part of the world that the ancestors came from. If a respondent named more than one country, the country to which the respondent was asked to name the county to which he/she feels closer to. The final sample contains individuals coming from 26 different origins, noting that descendent of immigrants from England and Wales, Germany, and Ireland constitute a great majority of the sample (almost 70% of all observations). Only individuals who grew up with both parents are included. Individuals for whom information about their own or their parents' education attainment is not available are omitted from the sample. Individuals in the sample are divided into five cohorts. Parental human capital is measured with both father's and mother's education.

Following Borjas (1992), group's quality is measured with average years of schooling in the fathers (or mothers) generation[2] within given cohort and region[3]. The group's *size* is

[2]Borjas refers to this as *ethnic capital*

[3]Since ethnic capital is measured as an average of educational attainment in the parents' generation within a cohort, only individuals for whom there is at least 30 other individuals in the same cohort of the same ethnic origin are included. Correlation between averages of the mother's and father's schooling within the

measured as the share of individuals belonging to the same cohort and ethnicity residing in the same region[4]). Lastly, the group's *tightness* is measured with endogamy rates, so the rates of intra-group marriages in a given region and cohort. Table 1 presents the summary

Table 1: Summary statistics of the main variables

	Men	Women
Years of schooling	13.96	13.73
	(2.86)	(2.62)
Father's years of schooling	11.38	11.21
	(4.04)	(3.99)
Mother's years of schooling	11.59	11.33
	(3.19)	(3.28)
Age	44.69	46.27
	(16.00)	(16.74)
Number of siblings	3.24	3.32
	(2.39)	(2.43)
Living in a city at 16	0.42	0.42
	(0.49)	(0.49)
Living in a Southern state at 16	0.23	0.23
	(0.42)	(0.42)
At least one parent born abroad	0.09	0.10
	(0.29)	(0.31)
Endogamy rates	0.35	0.35
	(0.15)	(0.15)
Group's quality	11.24	11.06
	(1.97)	(1.99)
Network size	0.11	0.11
	(0.07)	(0.07)
Number of observations	8032	9362

Notes: Standard deviations in brackets. Group's *quality* is measured with average years of schooling in the father's generation.

statistics for all variables used in this analysis separately for men and women, noting that 54 percent of the sample are women. The only noticeable difference between men and women appears along the age dimension, with women being on average two years older than men. The average individual has about three siblings and has completed almost 14 years of schooling[5]. The average parental and ethnic capital are approximately the same at 11 years of schooling. Forty-one percent of all individuals lived in urban settings at the age of 16, and 25 percent lived in the South at 16. Only 10 percent of individuals have at least one parent born abroad. The average network size is 11 percent and on average 35 percent of

ethnic group is 0.96, which makes identification of individual effects not feasible.

[4]region of the U.S. is the finest geographical division available in the data

[5]This might appear as surprisingly high. Cross-checking with US Census data on men born in the US between 1889 and 1995, the average years of schooling is about 13. Limiting the sample to individuals who grew up with both parents would likely result in a higher average. Therefore, GSS data appears to resemble a representative sample of the US population

marriages appear within the ethnic group.

3 Model, Identification and Estimation Strategy

Let *edu* denote the individual's education. Then the basic specification of the model is as follows:

$$edu_i = \gamma_1 eduf_i + \gamma_2 edum_i + \gamma_3 quality_i + \gamma_4 network + \gamma_5 tigntess + \delta_0 X_i + u_i \quad (1)$$

where *quality* denotes ethnic group's quality measured with average years of schooling in the father's generation[6]; *network* denotes network size measured with a share of individuals in a given cohort living in the same region, and *tightness* denotes the closeness of ties measured with endogamy rates. X denotes a vector of socioeconomic characteristics (age, age squared, number of siblings, indicators whether the individual lived in the South, lived in the city at the age of 15, and has a foreign-born parent) as well as time and region fixed effects.

If we apply OLS to equation 1, the obtained estimates will be biased due to confounding effects of unobserved ability affecting both parents and children schooling outcomes, noting that *quality* of the group is measured with the average education in the father's or mother's generation[7]. Therefore, I follow Postepska (2019) and apply Klein & Vella (2010) estimation strategy. I extend the model in Postepska (2019) to allow for differential effects of maternal and paternal schooling on men's and women's educational attainment. This extension is crucial when analyzing the heterogeneity of gender as mothers' and fathers' schooling have been found to have differential effects on men's and women's schooling outcomes. Especially for girls, an advantageous ethnic environment can have a very different effect on their educational attainment depending on the mother's educational attainment. For the cohorts considered, it is reasonable to assume that fathers' schooling serves as a good proxy of socioeconomic status, while mothers' education affects schooling decisions through different channels, such as role models. As the conceptual framework mimics the one in Postepska

[6] Alternatively, quality could be measured with average mothers schooling within a cohort. Results are basically the same and are available upon request. Inclusion of both variables is not feasible due to very high correlation between average schooling among mothers and fathers (0.96)

[7] Since ethnic capital is measured as an average of educational attainment in the parents' generation within a cohort, only individuals for whom there is at least 30 other individuals in the same cohort of the same ethnic origin are included. Correlation between averages of the mother's and father's schooling within the ethnic group is 0.96, which makes identification of the effect of individual effects not feasible.

(2019), below, I focus on the key features of the extended model and refer the reader to the above manuscript for details.

The three endogenous regressors are *edum*, *eduf* and *quality*. All other regresors are considered exogenous. In absence of exclusion restrictions to identify the parameters of equation 1, identification of the parameters relies on assumptions about the structure of the error term and heteroskedasticity in the model. The three endogenous equations are of the following form:

$$\begin{aligned} eduf_i &= \delta_2 X_i + v_i^f \\ edum_i &= \delta_2 X_i + v_i^m \\ quality_i &= \delta_3 X_i + v_i^q \end{aligned} \tag{2}$$

As given in Klein & Vella (2010), there are two conditions necessary for identification. The first condition states that there has to be heteroskedasticity in the error term of the main equation and/or in the error terms of the three auxiliary equations. In other words we can write the error terms of the main equation and of the three endogenous equations as:

$$u_i = H_u(X_i)u_i^*, \quad v_i^f = H_v^f(X_i)v_i^{f*}, v_i^m = H_v^m(X_i)v_i^{m*} \text{ and } v_i^q = H_v^q(X_i)v_i^{q*} \tag{3}$$

where u_i^*, v_i^{f*}, v_i^{m*}, v_i^{q*} are correlated homoskedastic error terms and $H_u^2(X_i)$, $H_v^{f2}(X_i)$, $H_v^{m2}(X_i)$ and $H_v^{q2}(X_i)$ denote the conditional variance functions for u_i, v_i^f, v_i^m and v_i^q, respectively.

The second assumption states that the correlation between the homoskedastic part of the error terms are constant:

$$\begin{aligned} E[u_i^* v_i^{f*}|X_i] &= E[u_i^* v_i^{f*}] = \rho^f \\ E[u_i^* v_i^{m*}|X_i] &= E[u_i^* v_i^{m*}] = \rho^f \\ E[u_i^* v_i^{q*}|X_i] &= E[u_i^* v_i^{q*}] = \rho^q \end{aligned} \tag{4}$$

To summarize, this assumption implies that there exists a constant transfer that is not affected by any socioe-conomic characteristics or behavior of the individual or the parents.

As argued in details in Postepska (2019) both assumptions appear plausible in the context

of this paper.[8] The conditional constant correlation assumption implies that after controlling for all the exogenous variables in the model, the correlation between the unobserved factors affecting individual's educational attainment and father's, mother's educational attainment or average educational attainment in the ethnic group remains constant. In the case of parental schooling, the genetic transmission of ability yields a natural interpretation of this error structure. In case of the group's *quality*, Postepska (2019) argues that aside from average ability, transfer of cultural capital can be deemed consistent with this assumption. Furthermore, the second assumption of heteroskedasticity in the child's schooling and/or the three endogenous equations is guaranteed. The contribution of the unobservables to educational attainment formation differs depending on the child's characteristics such as age, gender, or place of residence.

This assumed error structure allows construction of control functions which inclusion in equation 1 makes estimation of all parameters in equation 1 feasible by estimation of the following model:

$$\begin{aligned} edu_i &= \gamma_1 eduf_i + \gamma_2 edum_i + \gamma_3 quality_i + \gamma_4 size_i + \gamma_5 tightness_i + \delta_0 X_i \\ &+ \rho^f \frac{H_u(X_i)}{H_v^f(X_i)} v_i^f + \rho^m \frac{H_u(X_i)}{H_v^m(X_i)} v_i^m + \rho^{av} \frac{H_u(X_i)}{H_v^q(X_i)} v_i^q + \epsilon_i \end{aligned} \tag{5}$$

The estimation strategy follows closely Farré *et al.* (2012) and Postepska (2019).[9] I first estimate the three endogenous equations using OLS. Next, the conditional variance in all three equations is estimated using non linear least squares. I use exponential function to model the conditional variance in all equations. The last step involves simultaneous estimation of the three heteroskedastic index and the coefficients of the main equation. The results of the estimation of the auxiliary equation including the estimation of the heteroskedasticity equations and construction of the correction term is not of a core importance for our analysis and thus the results are discussed in Appendix B.

[8] Extending the model with one additional endogenous equation does not alter the interpretation of the identifying assumptions.

[9] For details of estimation I refer the reader to Appendix A and the aforementioned manuscripts.

4 Estimation results

Table 2 presents the summary for the estimation results. All models are estimated separately for men and women. For the sake of brevity, OLS results are only presented for the base model as specified in equation 1. Since the literature delivers suggestive evidence that the group's attributes can interact in how they affect human capital accumulation, interactions between group's *quality* and *size*, and group's *quality* and *tightness* are added to the base specification, noting that interactions between group's *tightness* and group's *size* are insignificant in all models, so they are not included.[10] First, each interaction term is added in isolation (columns 2, 3, 8 and 9) and finally a model with both interaction terms is estimated (columns 5 and 10). Even though the estimates concerning parental schooling effect on educational attainment are not of core importance, they are included for two reasons: to show that the estimation strategy is successful at accounting for endogeneity of these two regressors, and to serve as a reference point for the assessment of the magnitude of the estimated effects. Overall, the results deliver evidence that the ethnic groups' attributes, considered in this paper, play a significant role in the human capital accumulation process, regardless of controlling for the country of origin. Second, substantial gender heterogeneity is apparent, as evident in figure 1, which summarizes the main results.

Consider group's *size* first. In columns (2) and (7) it is evident that individuals from larger networks accumulate more human capital, a results consistent with the literature (Gang & Zimmermann, 2000). However, the effect is twice as large for men than for women. A 1% increase in the share of one's ethnicity in the population composition of a region yields a 0.05 increase in years of schooling for men. At face value, this is not a large effect. However, it is comparable to the magnitude of the effect of an additional year of mother's schooling. Allowing for an interaction between group's *size* and *quality* (column 3 and 8) shows that a larger network benefits men from the least advantageous ethnic groups, while for women, the opposite relationship is observed suggesting two attributes act as complements in the human capital process and women in most advantaged groups benefit the most from larger presence of their ethnic group. However, columns (5) and (10) further highlight that it is necessary to take into account the interaction between the group's *size* and *quality* as well as the interaction between group's *tightness* and *quality*. In the model with both interaction

[10]Note that once endoegenity of group's quality is accounted for by the control term in equation 5, the interaction terms with quality are assumed to be exogenous

terms, the effect of the group's *size* for women is small and not significant as depicted on the top panel of figure 1. The contrast between the results in column (8) and (10) highlights the interplay between the group's attributes in the human capital process and stresses the importance of the group's *tightness* for how ethnicity affects women's educational attainment.

While group's *tightness* appears insignificant in the base specification, results in columns (4) and (9) suggest that ties' closeness affects men and women in an almost identical way but its effect is not constant with respect to group's *quality*. Holding the group's *quality* constant at "at least high school" level, individuals from groups characterized by closer ties accumulate more human capital. The magnitude of this effect increases with the group's quality, suggesting that the group's *tightness* and *quality* act as complements in the human capital accumulation process. In contrast, among individuals of ethnic groups with low average schooling levels, closer ties decrease human capital accumulation, noting that the average educational attainment below ten years is rather low (about 15 percent of the sample). The results also deliver evidence of gender heterogeneity. Men benefit more from closer ties in more advantaged groups (higher *quality*), and women lose more because of closer ties in less advantaged groups. These observed patterns are consistent with parents relying more on the role of social networks in the human capital accumulation process of their female offspring, resulting in under-investment in the schooling process.

Last, consider the group's *quality*. Bottom panel of figure 1 further confirms the importance of group's *tightness* among women and group's *size* among men. A "tight group" is defined to have an endogamy rate of 0.6 (meaning that 60 percent of individuals marry within the same ethnicity), and a "loose" group has an endogamy rate at 0.2. Women in tight groups benefit from increasing *quality* of the ethnic environment at a constant rate, regardless of group size. The magnitude is quite small, though - approximately 0.02 years per one standard deviation increase in the group's endogamy rate. The opposite is true for women in groups characterized by looser ties. It could be that women in tight groups benefit a lot from the group's capital due to socialization styles and through role models. The correlation between mothers and fathers average schooling in parents generation is very high in the data implying that higher *quality* implies higher educational attainment among mothers. For men, the effect of group's *quality* is larger in magnitude (2-3 fold) and in addition to group's *tightness*, group's *size* plays a significant role. Men in small, tight groups appear to enjoy the same benefits as women in tight groups (of any size). It might

be the case that they are also subject to tighter upbringing compared to their counterparts in large groups and are more likely to have peers from the same ethnic group. In contrast, men in large groups, characterized by close ties, decrease educational attainment when the group's quality increases. It is consistent with parents putting too much confidence in the role of social capital when they belong to a large and close group and investing less in the human capital accumulation process of their male offspring in a more advantageous ethnic environment.

Table 2: Ethnicity and human capital accumulation

	Men					Women				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS Eq. 1	CF*	CF	CF	CF	OLS Eq. 1	CF	CF	CF	CF
Father schooling	0.174	0.120	0.119	0.126	0.131	0.151	0.173	0.172	0.177	0.177
	(0.009)	(0.010)	(0.007)	(0.011)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
Mother schooling	0.164	0.073	0.072	0.060	0.059	0.192	0.078	0.073	0.067	0.067
	(0.012)	(0.013)	(0.015)	(0.014)	(0.013)	(0.010)	(0.008)	(0.010)	(0.011)	(0.012)
Quality	0.015	-0.055	-0.013	-0.161	-0.139	0.003	-0.027	-0.082	-0.186	-0.184
	(0.039)	(0.045)	(0.048)	(-0.049)	(0.054)	(0.032)	(0.033)	(0.035)	(0.047)	(0.043)
Network	4.619	5.426	10.522	4.527	14.625	1.947	2.635	-3.302	1.750	1.404
	(1.001)	(1.170)	(3.292)	(1.098)	(4.167)	(0.792)	(1.005)	(2.420)	(0.990)	(2.827)
Tightness	-0.809	-1.098	-0.873	-4.514	-5.154	-0.614	-0.760	-0.991	-5.031	-4.988
	(0.563)	(0.600)	(0.573)	(1.112)	(1.325)	(0.437)	(0.552)	(0.510)	(0.943)	(1.239)
Quality×network			-0.480		-1.007			0.560		0.033
			(0.260)		(0.338)			(0.208)		0.245
Quality×tightness				0.359	0.499				0.447	0.441
				(0.090)	(0.119)				(0.084)	(0.12)
ρ_{father}		0.082	0.084	0.075	0.068		-0.047	-0.043	-0.052	-0.052
		(0.014)	(0.014)	(0.016)	(0.012)		(0.014)	(0.017)	(0.017)	(0.013)
ρ_{mother}		0.101	0.101	0.112	0.111		0.180	0.183	0.191	0.191
		(0.011)	(0.012)	(0.013)	(0.013)		(0.012)	(0.013)	(0.013)	(0.012)
$\rho_{quality}$		0.010	0.011	0.006	0.015		0.007	0.007	0.006	0.006
		(0.003)	(0.003)	(0.003)	(0.004)		(0.003)	0.004	(0.004)	(0.003)
Ethnicity dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Columns 1-5 present the results for men and columns 6-10 for women. *columns 2-5 and 7-10 present the control function estimates (CF). Standard errors bootstrapped. Other controls include age, age squared, dummies indicating whether the person lived in a city and in the South at the age of 16, number of siblings and a dummy variable indicating whether at least one parent is foreign born, and year and region dummies.

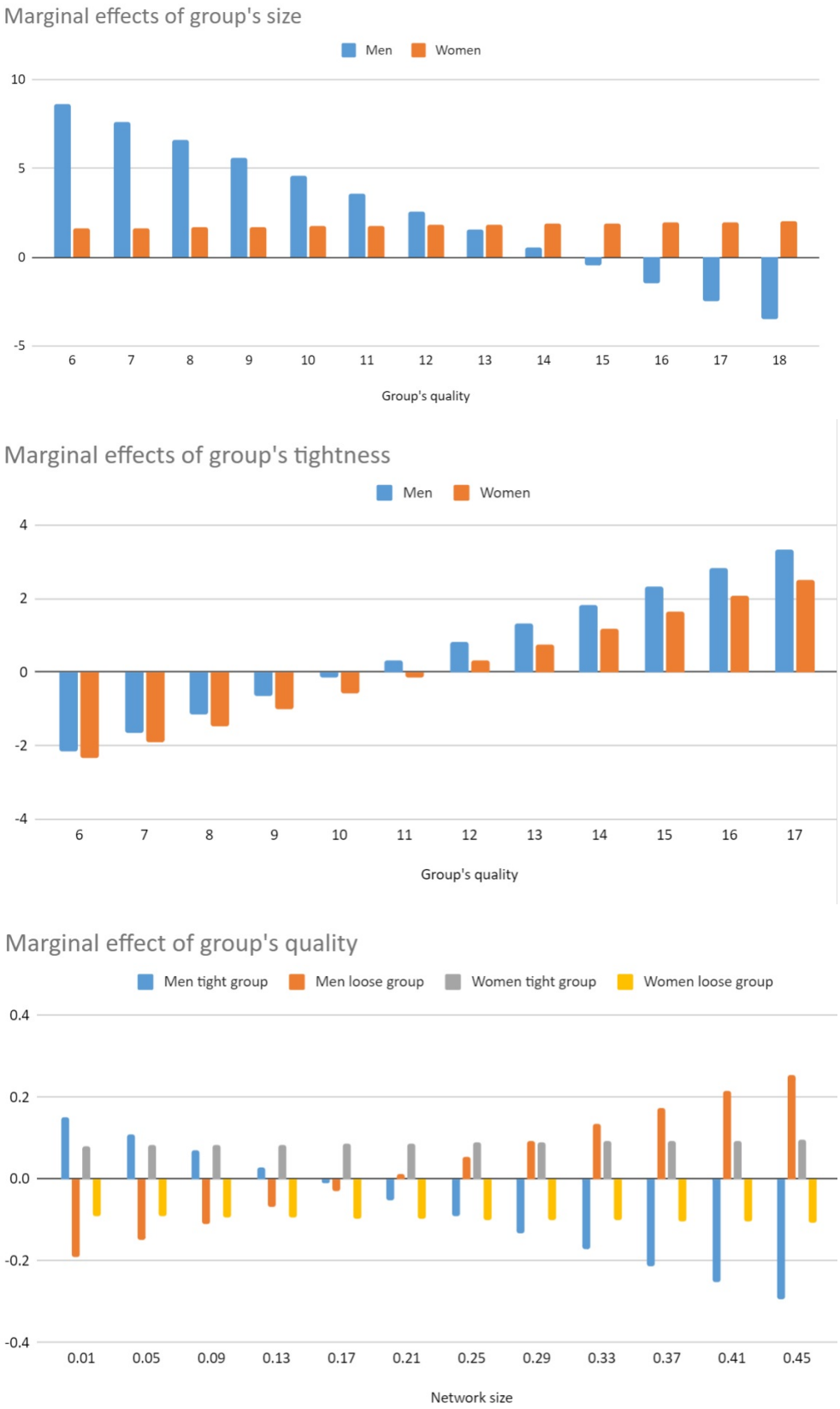

Figure 1: Marginal effects of group's *size*, *tightness* and *quality* on human capital accumulation computed based on results in column 5 and 10 in table 2

5 Conclusion

This paper extends the framework developed in Postepska (2019) to analyze the effects of different ethnic groups' attributes on the human capital formation process. I find evidence of substantial gender heterogeneity consistent with the literature on the socialization of minority youth. A high-quality ethnic group affects men and women differently, depending on the group's size and ties' closeness. Women in groups characterized by close ties benefit from the high-quality ethnic environment, regardless of the group's *size*. In contrast, among men, the group's *size* appears to be of importance, and men in large groups characterized by loose ties benefit the most from higher *quality* of the ethnic group. The results are consistent with different socialization patterns, as evidenced by the literature.

The results of this paper are important for two reasons. First, they show that ethnic background affects human capital accumulation in a complex way, and studies interested in understanding the differences in human capital accumulation should account for various group attributes. Second of all, by looking at native-born individuals, this paper establishes a long-term link between ethnicity and human capital accumulation, net of the migration shock. In most studies, native-born individuals are considered natives (or at best second-generation migrants), concealing the effects studied in this paper. I show that ethnic background has an impact that lasts over generations and under certain circumstances, gives individuals from more advantaged backgrounds a head start in schooling outcomes.

References

Agarwal, P.: *Passage from India: Post 1965 Indian immigrants and their children: Conflicts, concerns, and solutions*, Yuvati Publications, 1991.

Aydemir, A., Chen, W.-H., & Corak, M.: Intergenerational education mobility among the children of Canadian immigrants, *Canadian Public Policy*, 39(Supplement 1), (2013), S107–S122.

Bailey, B.: Dominican-american etbnic/racial identities and united states social categories, *International Migration Review*, 35(3), (2001), 677–708.

Bauer, P., & Riphahn, R. T.: Heterogeneity in the intergenerational transmission of educational attainment: evidence from Switzerland on natives and second-generation immigrants, *Journal of Population Economics*, 20(1), (2007), 121–148.

Borjas, G. J.: Ethnic capital and intergenerational mobility, *The Quarterly Journal of Economics*, (pp. 123–150).

Dasgupta, S. D.: Gender roles and cultural continuity in the Asian Indian immigrant community in the US, *Sex roles*, 38(11-12), (1998), 953–974.

Djajić, S.: Assimilation of immigrants: Implications for human capital accumulation of the second generation, *Journal of Population Economics*, 16(4), (2003), 831–845.

Eagly, A. H., & Carli, L. L.: Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: a meta-analysis of social influence studies., *Psychological Bulletin*, 90(1), (1981), 1.

Farré, L., Klein, R., & Vella, F.: Does increasing parents' schooling raise the schooling of the next generation? Evidence based on conditional second moments, *Oxford Bulletin of Economics and Statistics*, 74(5), (2012), 676–690.

Gang, I. N., & Zimmermann, K. F.: Is child like parent? educational attainment and ethnic origin, *Journal of Human Resources*, (pp. 550–569).

Gupta, M. D.: What is Indian about you? A gendered, transnational approach to ethnicity, *Gender & Society*, 11(5), (1997), 572–596.

Kao, G., & Thompson, J. S.: Racial and ethnic stratification in educational achievement and attainment, *Annual Review of Sociology*, 29(1), (2003), 417–442.

Klein, R., & Vella, F.: Estimating a class of triangular simultaneous equations models without exclusion restrictions, *Journal of Econometrics*, 154(2), (2010), 154–164.

Le Espiritu, Y.: " we don't sleep around like white girls do": Family, culture, and gender in filipina american lives, *Signs: Journal of Women in Culture and Society*, 26(2), (2001), 415–440.

Mani, L.: Gender, class and cultural conflict: Indu krishnan's knowing her place, *Our feet walk the sky: Women of the South Asian diaspora*, (pp. 32–36).

Nielsen, H. S., Rosholm, M., Smith, N., & Husted, L.: The school-to-work transition of 2^{nd} generation immigrants in Denmark, *Journal of Population Economics*, 16(4), (2003), 755–786.

Olsen, L.: *Made in America: Immigrant students in our public schools.*, New York: New Press, 1997.

Postepska, A.: Ethnic capital and intergenerational transmission of educational attainment, *Journal of Applied Econometrics*, 34(4), (2019), 606–611.

Riphahn, R. T.: Cohort effects in the educational attainment of second generation immigrants in germany: An analysis of census data, *Journal of Population Economics*, 16(4), (2003), 711–737.

Rumbaut, R. G.: Ties that bind: Immigration and immigrant families, *Immigration and the family: Research and policy on US immigrants*, (pp. 3–46).

Shah, B., Dwyer, C., & Modood, T.: Explaining educational achievement and career aspirations among young british pakistanis: Mobilizing 'ethnic capital'?, *Sociology*, 44(6), (2010), 1109–1127.

Sung, B. L.: *The Adjustment Experience of Chinese Immigrant Children in New York City.*, Staten Island, NY: Center for Migration Studies., 1987.

Williams, L. S., Alvarez, S. D., & Hauck, K. S. A.: My name is not Maria: Young Latinas seeking home in the heartland, *Social Problems*, 49(4), (2002), 563–584.

Yung, J.: *Unbound voices: A documentary history of Chinese women in San Francisco*, Univ of California Press, 1999.

Zhou, M., & Bankston III, C. L.: Family pressure and the educational experience of the daughters of Vietnamese refugees, *International Migration*, 39(4), (2001), 133–151.

Appendix

A Estimation procedure

This section outlines the two step procedure to estimate the model. First, regress $eduf_i$, $edum_i$ and $quality_i$ on X_i and obtain α^f, α^m and α^q. Then define the residuals from these two regressions as follows:

$$\begin{aligned} \hat{v}_i^f &= eduf_i - X_i\hat{\alpha}^f \\ \hat{v}_i^m &= edum_i - X_i\hat{\alpha}^m \\ \hat{v}_i^q &= quality_i - X_i\hat{\alpha}^q \end{aligned}$$

The conditional variances of the parental education and average education (group's quality) errors can be estimated using both parametric and non-parametric methods. In this paper I employ parametric approach and assume the following functional form of the heteroskedasticity:

$$\begin{aligned} H_{vi}^{f2} &= \exp(Z_i\theta^f) \\ H_{vi}^{m2} &= \exp(Z_i\theta^m) \\ H_{vi}^{q2} &= \exp(Z_i\theta^q) \end{aligned}$$

where Z_i is a vector of variables responsible for the heteroskedasticty of the errors. The set of variables Z_i can vary between the three equations. Note that there are no restrictions imposed over the relationship between Z_i and X_i, i.e. model is identified even if $Z_i = X_i$. If, however, there are variables that appear in Z_i but not in X_i, they do not help identify the model in a standard way. Since it is the movement in the variances that grants identification in the model, variables in Z_i aid identification only if they can explain the differences in the variance across observations.

The conditional variances are estimated using non linear least squares using $\ln(\hat{v_i}^f)$, $\ln(\hat{v_i}^m)$ and $\ln(\hat{v_i}^q)$ as dependent variables. Then we can compute the standard deviation of the error terms associated with the two reduced forms: $\hat{H}_{vi}^f = \sqrt{\exp(Z_i\hat{\theta}^f)}$, $\hat{H}_{vi}^m = \sqrt{\exp(Z_i\hat{\theta}^m)}$ and $\hat{H}_{vi}^q = \sqrt{\exp(Z_i\hat{\theta}^q)}$.

Last element necessary to estimate the parameters of the main equation is the standard deviation of the child's education error (so the error term of the main equation). Since consistent residuals are not readily available, it is estimated simultaneously with the parameters of the main equation in an iterative procedure. Let $\beta = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \delta_0, \theta_u\}$. Parameters in equation**??** are found using a non linear least squares:

$$\begin{aligned}\min_{\beta,\rho^f,\rho^m,\rho^q} \sum_{i=1}^{n} (edu_i &- \gamma_1 eduf_i - \gamma_3 edum_i + \gamma_3 eduav_i + \gamma_4 network_i + \gamma_5 tightness_i \\ &- \delta_0 X_i - \rho^f \frac{H_{ui}}{\hat{H}^f_{vi}} \hat{v}^f_i - \rho^m \frac{H_{ui}}{\hat{H}^m_{vi}} \hat{v}^m_i - \rho^q \frac{H_{ui}}{\hat{H}^q_{vi}} \hat{v}^q_i)^2 \end{aligned} \tag{6}$$

where H^2_{ui} denotes the conditional variance of the child's education equation. In a fully parametric specification, assume $H^2_{ui} = exp(Z_i\theta_u)$.

To simplify the computations, Klein and Vella (2010) suggest a two step procedure. First, for a given value of $\beta = \tilde{\beta}$, define the residuals $u_i(\tilde{\beta})$ and compute the standard deviation of the child's education error in the same way as $\hat{H}^f_{vi}$, $\hat{H}^m_{vi}$ and $\hat{H}^q_{vi}$, so $\hat{H}_{ui} = \sqrt{\exp(Z_i\tilde{\theta}_u)}$. Second, estimate ρ^f, ρ^m and ρ^q by minimizing the sum of the squared residuals of the child's education equation:

$$\min_{\rho^f,\rho^m,\rho^q} \sum_{i=1}^{n} \left(u_i(\tilde{\beta}) - \rho^f \frac{\hat{H}_{ui}(\tilde{\beta})}{\hat{H}^f_{vi}} \hat{v}^f_i - \rho^m \frac{\hat{H}_{ui}(\tilde{\beta})}{\hat{H}^m_{vi}} \hat{v}^m_i - \rho^q \frac{\hat{H}_{ui}(\tilde{\beta})}{\hat{H}^q_{vi}} \hat{v}^q_i \right)^2 \tag{7}$$

Repeat the last two steps until the minimum of (7) is found.

B Auxiliary equations

This appendix discusses the results of the estimation of the auxiliary equations. For the sake of brevity, I only present the results for the base specification (equation 1), including the estimation of the heteroskedastcity indexes of the three endogenous equations and the construction of the correction terms. All other results can be obtained on request. The sets of variables included in father's, mother's and the group's quality equations are almost the same, so I discuss them together. Since I do not have information about the age of the parents, I include the age of the children (and age squared) in both equations. This, together with the dummy variable indicating the cross section, controls for the age of the fathers and mothers. Dummy variables for regions control for geographic differences in educational

attainment that might result from labor market specific needs of given region or different access to educational institutions. I also include a dummy variable indicating whether the child was living in the South or in the city at the age of 16. Unfortunately, this information is not available for the parents so I use the information for the children as proxies. In the group's quality equation, equation I also include a dummy indicating whether at least one of the parents is foreign born and in both of the parents schooling equation a dummy variable indicating whether that parent is foreign born.

The OLS estimates of the conditional means of the parental capital and group's quality equations presented in Table **??** are in line with the literature. Most of the year dummies are significant and indicate an upward trend in educational attainment among parents and increasing average schooling among all ethnic groups. The results in the men and women sub-sample are alike. Individuals of ethnic origins with bigger networks have substantially better educated parents and have more favorable ethnic environment. The opposite is true for individuals of ethnic origin with closer ties. Younger individuals and those with fewer siblings have not only better educated parents but also more favorable ethnic environment. Individuals living in the city at 16 have, on average, better educated parents than their counterparts residing outside of the cities. The correlation with the group's quality is also positive but it's of much smaller magnitude. Residence in southern states lowers the average educational attainment of both parents as well as the average ethnic capital individuals are exposed to by almost one year. Parents born outside of the US have on average 2.7 years less of schooling than US born fathers. Also, average group's quality decreases for individuals with at least one parent born abroad by about 0.8 on average. This might reflect the fact that more recent immigrant groups are on average less educated than members of established groups within the US. There is also some evidence of regional differences for both parental capital as well as for the group's quality.

Lower panel of Table **??** presents the test statistics for White and Breush-Pagan tests for heteroskedasticity in all three equations. The null hypothesis of homoskedastic errors is strongly rejected in all equations confirming presence of heteroskedasticity in both of the parental as well as in the group's quality equations necessary for identification.

Table B.4: Conditional means: parental capital and group's quality

	Men			Women		
	Father	Mother	Quality	Father	Mother	Quality
Network size	6.25***	6.54***	9.39***	5.74***	7.03***	10.11***
	(0.69)	(0.60)	(0.34)	(0.68)	(0.58)	(0.30)
Group's tightness	-4.97***	-4.93***	-7.02***	-4.17***	-4.68***	-7.01***
	(0.37)	(0.33)	(0.19)	(0.37)	(0.32)	(0.17)
Number of siblings	-0.29***	-0.26***	-0.08***	-0.28***	-0.24***	-0.09***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
Lived in a city at 16	1.21***	0.53***	0.15***	1.11***	0.61***	0.10***
	(0.08)	(0.06)	(0.03)	(0.08)	(0.06)	(0.03)
Lived in the South at 16	-0.50***	-0.30**	-0.21***	-0.38***	-0.17	-0.09*
	(0.15)	(0.12)	(0.06)	(0.14)	(0.11)	(0.05)
Age	-0.10***	-0.05***	-0.06***	-0.08***	-0.06***	-0.05***
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
Age^2	0.00***	0.00	0.00	0.00	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Foreign born father	-2.45***	-1.91***		-2.44***	-2.22***	
	(0.17)	(0.15)		(0.16)	(0.14)	
At least one parent			-0.83***			-0.73***
born abroad			(0.06)			(0.05)
Constant	13.30***	13.35***	15.87***	12.93***	12.61***	15.47***
	(0.44)	(0.35)	(0.13)	(0.42)	(0.35)	(0.12)
Observations	8032	8032	8032	9362	9362	9362
Breush-Pagan test	146.88	486.09	742.98	111.81	426.97	650.49
White test	913.13	1115.24	2039.32	951.35	1160.08	2252.42

Robust standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Having established the presence of heteroskedasticity in all three endogenous equations, I proceed with estimation of the form of the heteroskedasticity and further construct the three control functions[11]. Results of the non linear least squares estimation of the conditional variances are presented in Table B.5. Given the assumed exponential form of heteroskedasticity, I can directly interpret the coefficients. Older individuals are exposed to

[11] In the paper results using the preferred specification are discussed. The results are qualitatively unaffected by the choice of the form of heterskedasticity. However, some small quantitative differences are present.

a smaller variation in average education among individuals from the same origin as well as their fathers and mothers have smaller residual variance. This could result from increasing heterogeneity of immigrants coming from the same origin as well as easier access to education. Living in the city at the age of 16 decreases the dispersion in father's schooling among men and in the group's quality for women, while it increases the dispersion in mother's schooling among men. Living in the South at 16 increases dispersion in fathers schooling among men and women and decreases dispersion in group;s quality in the men sub-sample. Parents born abroad appear to have lower residual variance variance in educational attainment. I also find smaller dispersion in group's quality for individuals with at least one parent born abroad. Larger groups also appear to be more homogeneous with respect to the groups' quality and mothers belonging to groups with closer ties appear to be more heterogenous with respect to their schooling outcomes. To construct the correction terms we still need the

Table B.5: Parental capital and group's quality - heteroskedastic indexes

	Men			Women		
	Father's schooling	Mother's schooling	Group's quality	Father's schooling	Mother's schooling	Group's quality
Age	-0.007	-0.069	-0.523	0.003	-0.058	-0.498
	(0.008)	(0.008)	(0.020)	(0.008)	(0.006)	(0.013)
Living in a city at	-0.400	0.291	-0.051	0.213	-0.012	-0.342
the age of 16	(0.047)	(0.063)	(0.109)	(0.040)	(0.045)	(0.093)
Living in the South	0.262	0.060	-0.530	0.206	0.051	-0.028
at the age of 16	(0.041)	(0.104)	(0.178)	(0.050)	(0.097)	(0.175)
Number of siblings	0.032	0.000	0.000	0.017	0.029	-0.039
	(0.008)	(0.12)	(0.020)	(0.009)	(0.009)	(0.022)
Father born abroad	-0.305	-0.329		-0.453	-0.636	
	(0.048)	(0.075)		(0.047)	(0.051)	
At least one parent			-0.185			-0.601
born abroad			(0.227)			(0.256)
Network size	-0.142	-2.465	-3.142	-0.544	-0.591	-3.580
	(0.330)	(0.492)	(1.189)	(0.324)	(0.558)	(1.029)
Group's tightness	0.279	1.208	-0.796	0.476	0.691	0.201
	(0.157)	(0.159)	(0.737)	(0.144)	(0.222)	(0.490)
Constant	0.571	2.212	9.745	0.643	1.834	9.336
	(0.225)	(0.242)	(0.352)	(0.218)	(0.177)	(0.237)
Number of observations		8032			9362	

Notes: Standard errors bootstrapped. All regressions also include age squared, dummy variables for region of residence. The regressions for father and mother schooling also include year dummies.

estimates of the heteroskesdticity index in the primary equation (equation 1). It is estimated

simultaneously with the coefficients of the same equation. The results are presented in Table B.6. It appears that older individuals and men whose at least one parent is born abroad have higher residual variance.

Table B.6: Heteroskedastic indexes for children's human capital equation

	Men	Women
Age	0.113	0.150
	(0.021)	(0.038)
Age^2	-0.001	-0.001
	(0.000)	(0.000)
At least one parent born abroad	0.205	0.073
	(0.113)	(0.258)
Network size	0.457	-0.757
	(0.732)	(1.032)
Number of siblings	-0.017	0.078
	(0.020)	(0.060)
Constant	-3.340	-4.870
	(0.513)	(0.931)

Notes: Standard errors bootstrapped.