DISCUSSION PAPER SERIES

IZA DP No. 7433

A Note on Height and Surnames: The Role of Networks

Wolter Hassink
Bas van Leeuwen

May 2013

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

A Note on Height and Surnames: The Role of Networks

Wolter Hassink
Utrecht University
and IZA

Bas van Leeuwen
Utrecht University
and Hangzho Normal University

Discussion Paper No. 7433
May 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 7433
May 2013

ABSTRACT

A Note on Height and Surnames: The Role of Networks

Many studies indicate that human height is determined largely by childhood circumstances, which in turn influences an adult's labor market opportunities. The aim of this note is to test this thesis by examining the correlation between childhood circumstances and labor market outcomes on the one hand, and heights on the other, when networks are included as proxied by surnames. The fact that, after the inclusion of this surname proxy, we find a correlation only between height and labor market outcomes suggests that, while childhood circumstances affect height largely via social status and networks as captured by surnames, the same does not apply for labor market outcomes.

JEL Classification: J01, N35, Z13

Keywords: Indonesia, networks, stature

Corresponding author:

Wolter Hassink
Utrecht University
School of Economics
P.O.Box 80125
3508 TC Utrecht
The Netherlands
E-mail: w.h.j.hassink@uu.nl

1. Introduction

Human height is generally accepted as an indicator of well-being. It is acknowledged that height is largely a function of the socio-economic conditions experienced by an individual during childhood, and similarly, a determinant of labor market opportunities encountered during adulthood. Some effects of socio-economic conditions during one's growth years on height and/or one's labor market opportunities may be direct; others may operate indirectly through the intermediary of one's networks (and social status[1]). In the pages that follow we argue that including surnames as a proxy for an individual's network weakens the effect of childhood circumstances on height, but it does not affect the relation between height and labor market outcomes.

The relationship between childhood circumstances / labor market opportunities and heights is discussed at length in the literature on human stature.[2] Height -- being a function of income, religion, ethnicity, and place of birth -- is generally accepted as a proxy of well-being (e.g., Eveleth and Tanner, 1976; Mandemakers and Van Zanden, 1993; Baten, 2000, 2006; A'Hearn, 2003; Maria-Dolores and Martinez-Carrion, 2011). In addition, it has been argued that height may be a determinant of educational, and hence occupational, opportunities (e.g., Persico, Postlewaite, and Silverman, 2004; Cinnirella, Piopiunik, and Winter, 2011).

Our contention is somewhat more nuanced: that childhood circumstances and labor market outcomes affect well-being through the intermediary of networks.[3]

---

[1] One may argue that social status and networks are the same, the former serving to valorize the latter. After all, social status as such cannot affect well-being unless it is validated by a social network. The simple fact that one is an aristocrat is meaningless when removed from the context of family, friends, and acquaintances.

[2] In this literature, childhood circumstances are called "socio-economic consequences" while labor market opportunities are called the "functional consequences" of height (Steckel, 1999, 1908).

[3] The close relation between networks and height has been demonstrated (e.g., Komlos, 1990; 1994). Among researchers in this field, there is a consensus that after controlling for socio-economic

Networks may exist both for childhood circumstances and labor market opportunities. The former network is relatively static. It consists of preexisting alliances; for instance, a couple's socio-professional status – which includes their own social network, or networks, which in turn may include distant ancestors and living relatives – will largely determine the social status of their offspring. It follows that a social-status and network patterns are likely to persist across generations, where individuals keep, and pass on the same surname to their offspring. Networks covering labor market opportunities might be described as dynamic. These networks feature the creation of alliances, sometimes through the intermediary of an established, static network, which serves to screen new connections by education or occupation, and therefore helps to explain the well-documented correlation between labor market outcomes and height.

In order to determine whether networks play a role in the correlation between childhood circumstances and height (static network) or between labor market outcomes and height (dynamic network), we perform a regression featuring, for the first time in the height literature, surname (a standard proxy for networks) as well as the two standard factors considered in previous empirical studies of height: childhood circumstances and labor market outcomes. This approach has enabled us to confirm that surname is (that is, that networks are) indeed a reliable predictor, among others, of height.

2. Data

Our analysis is based on three categories of data: childhood background (place of birth, ethnicity, and religion), labor market outcome (education and occupation), and

factors such as parents' education, height differences across occupations persist (e.g., Lantzsch and Schuster, 2009). In other words, social networks seem to have a greater impact on childhood circumstances than on subsequent labor market outcomes.

economic status (height). As for the use of surnames as a proxy for social networks, there are a number of precedents (Pollitzer, Smith, and Williams, 1988; Smith, 2002; Smith and MacRaild, 2009). An individual's surname not only functions as a label indicating membership of one or more socio-economic networks, such as family, ethnic group, and religious faith (Colantonio, Fuster, and Küffer, 2007) but also serves to track the evolution of societal patterns in migration, marriage (Smith and MacRaild, 2009), and labor market demographics.[4]

Our dataset consists of pension-list information on a sample of 9,085 Dutch-Indies Army (KNIL) Indonesia and ethnic-Chinese recruits. European recruits were excluded as they constitute a distinctly different sector of Indonesian society. It is because the recruitment data (the standard source of information on army recruits) attached to these individuals has not survived that we exploit these pension lists, for the years 1946-1947 (Nationaal Archief, arch. nr. 2.10.50), which were sent to the Netherlands after Indonesia gained independence.

The fact that during this postwar period the Dutch administration had has slightly more restricted access to Indonesia did not affect recruitment practices, including minimum and maximum age thresholds. Although the war had ended, the Dutch still needed recruits, for two "police actions," (i.e. the repression of the paramilitary independence movement that sprung up in the wake of the Netherlands' attempts to oust the nationalist Sukarno, president from August 1945**)** and many veterans chose to reenlist. As a result, our sample spans an unusually broad swath of birth cohorts: from 1881 to 1932, as well as one with as broad a geographical basis, including a large number from the Moluccas and Java (consistent with historical

[4] For information on education, see Case and Paxon (2008); on informal recruitment, Pellizzari (2010); and on discrimination, Fryer, Jr., and Levitt (2004), Bertrand and Mullainathan (2004), and Foster (2008).

trends). In addition, because the Dutch no longer had unrestricted access to Java, they were obliged to extend their reach to the Outer Provinces and Madura (Bouman, 1995); with the result that the army was now more representative of Indonesian society, at least with respect to demographic/geographic spreads (Zwitzer and Heshusius, 1977).

The dataset features each recruit's name, date of birth, place of birth, place of last residence, ethnicity, religion, education, and occupation. Surnames presented two problems, requiring two modifications. First, before the war, names were spelled (presumably) phonetically. For instance, we find not only "Muhammad" but also “Muhammat,” “Muhamad,” Moehamed,” “Moehamad,” and so on. Our approach has been to treat all such versions as a single surname. Second, Muslim and Hindu names (the latter largely from the island of Madura) were often combined with a patronymic or the like. For example, Muhamad bin Achmad, means "Muhamad, son of Achmad." The Hindu name I Wajan Dama means, literally, "firstborn son, family name of Dama." In those cases in which one element of the name identifies the individual's caste or sex or birth order, we chose to use the name unique to the individual, rather than the element identifying his caste or his father.

<Table 1 about here>

The pension lists a total of 9,085 individual records, but some of these were incomplete, and were excluded from the sample. Of the excluded records, there were 789 that omitted height, 1,946 that omitted education, 225 that omitted occupation, 1 that omitted place of birth, and 1 that omitted birth year. Our sample consists of the remaining 6,125 complete files. These files comprise 3,931 different surnames, 3,028

(or 77%) of which occur only once and thus were omitted as well. This results in a total sample of 3,097 persons, 52.8% of which share the same surname, 20.8% of which the surname is shared by three recruits, and 1% where the surname is shared by at least 16 individuals (Table 1).

<Figure 1 about here>

The height of each recruit is that which was attained after adolescence (about 16 years, which may vary culturally of course). Figure 1 provides the histogram and kernel density. The mode is at 160 cm.

<Tables 2 - 4 about here>

3. Estimates

We regress height on a number of two groups of explanatory variables. We distinguish between factors indicating childhood circumstances and factors indicating labor market opportunities. The childhood circumstances indicators include: religion (7 indicators)[5]; ethnicity (14 indicators)[6] and place of birth (23 indicators)[7]. Factors indicating labor market opportunities include: level of education (14 indicators)[8];

[5] Muslim, Protestant, Catholic, Buddhism, Confucianism, Hinduism, other (including non-religious), unknown.

[6] Amboinees, Balinees, Boeginees, Chinese, Javanese, Madoerees, Maleier, Manadonees, Papoeanees, Sangirees, Soendanees, Timorees, Toradjanees, and other.

[7] West Java, Central Java, East Java, Amboina, Aceh, Bali and Lombok, Bali, Bengkulu, Belitung, South Sulawesi, Jambi, Indragiri, Lampung, Manado (North Sulawesi), New Guinea, East Sumatra, Palembang, Riau, Tapanuli, Ternate, Timor, West Kalimantan, West Sumatra, (South)east Kalimantan, Madura, Batavia, and Nusa Tenggara.

[8] Vervolgschool, Ambonse school, Hollands-Inlandsche school (HIS), Inlandsche school, Maleische school, volksschool/desa school, zendingsschool, ambachtsschool, arabische school, Ardjoena school, middelbaar onderwijs (including MULO), Chinese school, Europees Lager Onderwijs, Korpsschool, schakelschool, no education.

occupation (9 indicators)[9]; place last lived (24 indicators); and an indicator capturing change over time (birth year – 43 indicators).

In Table 2 the estimation results are provided for each group of indicators specifying the F-statistic - of a statistical test of joint significance of the group of explanatory variables - and the corresponding p-value. The estimation results reveal that both the childhood circumstances and the labor market opportunities are related to height. This changes, however, when we add surname as an explanatory variable to the regression equation (Table 3). We not only observe that all factors related to childhood circumstances become jointly insignificant while labor market opportunities remain jointly significant, but that its inclusion causes a significant increase in the R-squared.

Next, we include all explanatory variables in one regression equation. In Table 4 the estimation results are reported. In each column we remove groups of explanatory variables that are not jointly significant. The key specification is the one that includes birth year, education, surname, and occupation (Column (7)), because only labor market-opportunity indicators and surnames remain, while all childhood factors are shown to be insignificant, or become insignificant when surnames are included. All of the variables are jointly statistically significant at the 5% level. Surname is significant at the 1% level.

4. Placebo test

One could argue that the estimation results presented in Table 4 could be the result of some random application of a certain surname to certain individuals. In this case, the network of surnames could be coincidental. To test for this possibility, we applied a

[9] Agriculture, industry, trade, other services, labor, coolie, day laborer, djongo, no occupation, and unknown.

"placebo test", or falsification test. We simulated comparable networks sharing the same surname distribution (Table 1) by randomly assigning 3,097 observations to a surname but with the prerequisite that the distribution of names be identical to the distribution of names in the sample (Table 1). For each of the random assignments, we re-estimated the specification of column (7) of Table 4 and reported the p-value of the F-test of surname.

We consider 550 estimates of the height equation, where each estimate is based on a different assignment of last name, focusing on the distribution of the p-value of the F-statistic of surname. We report a p-value of height of 0.003 (Table 4 (column 7)), whereas the average p-value of the F-statistic of surname is 0.505 (550 estimates, and the median is 0.09. For 0.0018%, the p-value of surname is below the value of 0.003. Our conclusion is that the close correlation between surname and height is anything but coincidental.

5. Conclusion

There is a large body of research that identifies correlations between childhood circumstances and height, and lifetime opportunities and height respectively. In the height literature, this dual link has often been described in terms of a direct relation between an individual's height and either a) childhood circumstances such as family income, religion, and ethnicity or b) labor market opportunities such as place last lived, education, and occupation.

Our height study differs from its predecessors in that (through the use of surname as a proxy for networks) we applied network analysis to height patterns, and found that when surnames are included in a height equation, the effects of the explanatory variables that are related to childhood circumstances become statistically

insignificant. In other words, while the link between height and childhood circumstances is governed by networks as proxied by last names, the same does not apply for the link between height and labor market opportunities. This suggests that, even in a colonial society such as Indonesia, with considerable religious, ethnic and social diversity, social background and networks have only a limited impact on access to the labor market and, hence, on one's future welfare development through occupation and education.

References

A'Hearn, B., 2003. Anthropometric evidence on living standards in northern Italy, 1730-1860. Journal of Economic History 63 (2), 351-381.

Baten, J., 2000. Height and real wages: an international comparison. Jahrbuch für Wirtschaftsgeschichte 2000-1, 17-32.

Baten, J., 2006. Global height trends in industrial and developing countries: an overview. Tuebingen University working paper.

Bertrand, M., Mullainathan, S., 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. The American Economic Review 94 (4), 991-1013.

Bouman, B., 1995. Van driekleur tot rood-wit: de Indonesische officieren van het KNIL1900-1950. 's-Gravenhage: Sectie Militaire Geschiedenis.

Case, A., Paxon, Ch., 2008. Stature and status: height, ability, and labor market outcomes. Journal of Political Economy 116 (3), 499-532.

Cinnirella, F., Piopiunik, M., Winter, J., 2011. Why does height matter for educational attainment? Evidence from German children. Economics and Human Biology 9 (4), 407-418.

Colantonio, S., Fuster, V., Küffer, C., 2007. Isonymous structure in the white population of Córdoba, Argentina, in 1813. Human Biology 79 (5), 491-500.

Eveleth, Ph., Tanner, J., 1990. Worldwide Variation in Human Growth, third ed. Cambridge: Cambridge University Press.

Foster, G. 2008. Names will never hurt me: Racially distinct names and identity in the undergraduate classroom. Social Science Research 37 (3), 934-952.

Fryer, Jr., R. G., Levitt, S. D., 2004. The causes and consequences of distinctly black names. Quarterly Journal of Economics 119 (3), 767-805.

Komlos, J., 1990. Height and social status in eighteenth-century Germany. Journal of Interdisciplinary History 20 (4), 607-621.

Komlos, J., 1994. The nutritional status of French students. Journal of Interdisciplinary History 24 (3), 493-508.

Lantzsch, J., Schuster, K., 2009. Socioeconomic status and physical stature in 19th-century Bavaria. Economics and Human Biology 7 (1), 46-54.

Mandemakers, C. A., Van Zanden, J. L., 1993. The height of conscripts and national income: apparent relations and misconceptions. Explorations in Economic History, 30 (1), 81-97.

Mario-Dolores, R., Martinez-Carrion, J. M., 2011. The relationship between height and economic development in Spain, 1850-1958. Economics and Human Biology 9 (1), 30-44.

Nationaal Archief, Den Haag, Ministerie van Koloniën: Stamboeken en pensioenregisters Militairen Oost-Indië en West-Indië, 1815-1954, nummer toegang 2.10.50.

Pellizzari, M., 2010. Do friends and relatives really help in getting a good job? Industrial & Labor Relations Review 63 (3).

Persico, N., Postlewaite, A., Silverman, D., 2004. The effect of adolescent experience on labor market outcomes: the case of height. Journal of Political Economy 112 (5), 1019-1053.

Pollitzer, W. S., Smith, M. T., Williams, R.W., 1988. A study of isonymic relationships in Fylingdales parish from marriage records from 1654 through 1916. Human Biology 60 (3), 363–382.

Smith, M. T., 2002. Isonymy analysis: the potential for application of quantitative

analysis of surname distributions to problems in historical research. In: Smith, M.T. (ed.). Human Biology and History. London: Taylor and Francis, 112-133.
Smith, M., MacRaild, D., 2009. Nineteenth-century population structure of Ireland and of the Irish in England and Wales: an analysis by isonymy. American Journal of Human Biology 21 (3), 283-289.
Steckel, R., 1995. Stature and the standard of living. Journal of Economic Literature 33 (4), 1903-1940
Zwitzer, H. L., Herhusius, C.A., 1977. Het Koninklijk Nederlands-Indisch Leger 1830-1950. ‘s-Gravenhage: Staatsuitgeverij.

Figure 1: Height (in cm)

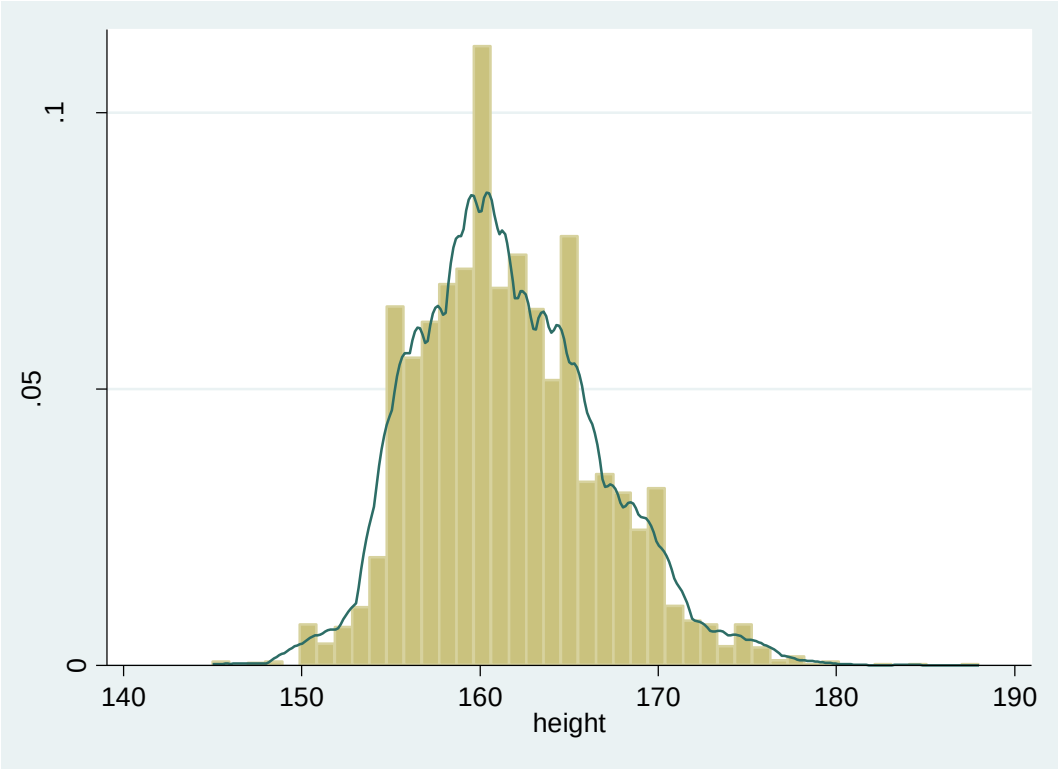

Table 1: Frequency of distribution of surnames

Incidence of surname	Number of surnames (%)	Number of persons involved [a)]
2	477 (52.8)	954
3	188 (20.8)	564
4	84 (9.3)	336
5	43 (4.8)	215
6	32 (3.5)	192
7	22 (2.4)	154
8	11 (1.2)	88
9	9 (1.0)	81
10	10 (1.1)	100
11	5 (0.6)	55
12	3 (0.3)	36
13	7 (0.8)	91
14	2 (0.2)	28
15	1 (0.1)	15
16 and more	9 (1.0)	190
Total	903 (100.0)	3,097

a) 3,097 persons. At least two persons with same surname.

Table 2: Regressions: height as dependent variable [a)]

	Childhood circumstances			*Labor market opportunities*		
	(1)	(2)	(3)	(4)	(5)	(6)
	Place of birth	Ethnicity	Religion	Place last lived	Education	Occupation
Variable [b)]	2.34 (0.0003)	4.36 (0.000)	5.72 (0.000)	3.54 (0.000)	3.83 (0.000)	4.13 (0.000)
Number of explanatory vars.	23	14	7	24	14	9
Number of obs.	3097	3097	3097	2811	3097	3097
Adjusted R-squared	0.010	0.015	0.011	0.021	0.013	0.009
R-squared	0.017	0.019	0.013	0.030	0.017	0.012

a) F-statistics. p-value of F-statistic in parentheses.
b) For each column a different class of explanatory variables: place of birth, ethnicity, religion, place last lived, education or occupation.

Table 3: Regressions including dynamic factors: height as dependent variable [a)]

	Childhood circumstances			*Labor market opportunities*		
	(1)	(2)	(3)	(4)	(5)	(6)
	Place of birth	Ethnicity	Religion	Place last lived	Education	Occupation
Variable [b)]	1.17 (0.258)	0.71 (0.771)	0.63 (0.728)	1.99 (0.003)	2.50 (0.002)	2.86 (0.002)
Surname	1.14 (0.008)	1.11 (0.028)	1.13 (0.013)	1.12 (0.023)	1.16 (0.004)	1.17 (0.003)
Number of explanatory vars.	925	916	909	916	916	911
Number of obs.	3097	3097	3097	2811	3097	3097
Adjusted R-squared	0.050	0.046	0.047	0.058	0.057	0.055
R-squared	0.334	0.328	0.327	0.365	0.336	0.333

a) F-statistics. p-value of F-statistic in parentheses.
b) For each column a different group of explanatory variables: place of birth, ethnicity, religion, place last lived, education or occupation.

Table 4: Regressions: height as dependent variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Childhood circumstances							
place of birth	1.59 (0.04)	1.87 (0.01)	-	1.43 (0.09)	1.39 (0.10)	1.22 (0.22)	-
Ethnicity	1.17 (0.29)	-	-	0.83 (0.63)	0.92 (0.54)	-	-
Religion	1.12 (0.35)	-	-	0.49 (0.85)	-	-	-
Labor market opportunities							
Education	1.76 (0.04)	2.24 (0.01)	2.59 (0.001)	1.53 (0.09)	1.59 (0.08)	1.61 (0.07)	1.68 (0.05)
Occupation	1.81 (0.06)	2.23 (0.02)	2.58 (0.01)	2.30 (0.01)	2.28 (0.02)	2.46 (0.01)	2.16 (0.02)
Other variables							
Surname	-	-	-	1.11 (0.03)	1.12 (0.02)	1.14 (0.01)	1.16 (0.004)
Birth year	2.45 (0.000)	2.46 (0.000)	2.51 (0.000)	2.00 (0.000)	2.03 (0.000)	2.05 (0.000)	2.06 (0.000)
Number of explanatory vars.	110	89	66	1012	1005	991	968
Adjusted R-squared	0.051	0.044	0.038	0.082	0.083	0.084	0.076
R-squared	0.08	0.07	0.06	0.38	0.38	0.38	0.43

a) Table of F-statistics. p-value of F-statistic in parentheses. 3,097 observations.

Note: We excluded "place last lived" since that variable reduces our sample substantially.