University of Economics, Prague

Faculty of Economics
Field of study: Economic Analysis

MASTER THESIS

Economics of Family: The Effect of Biological and Socio-economical Status and Air Pollution on the Sex of Children

Author: **Marek Pažitka**

Supervisor: **Mgr. Ing. Dominik Stroukal**

Academic Year: **2013/2014**

Declaration of Authorship

I declare on my honor that I wrote this master's thesis independently, and I used no other sources and aids than those indicated.

Prague, August 18, 2014

Signature

Acknowledgments

The author is grateful especially to Mgr. Ing. Dominik Stroukal for his inspiring insights and thoughtful guidance. I feel extremely grateful for his time and energy. I am also thankful to Ing. Petr Houdek for the encouragement and the valuable comments, to Ján Selep for the help with a language side of the thesis and to Tomaš Drvoštěp for help with typographical tools. At last but not least, I want to thank my girlfriend and family for their continuous support.

Abstract

The Trivers-Willard hypothesis (TWH) states that parents in good conditions will bias the sex ratio toward sons and parents in poor conditions will bias the sex ratio toward daughters. The present study contributes to literature in several ways: a large, general, country population data set (N= 1 401 851) from modern contemporary society; first study in the Czech Republic; an inclusion of air pollution into the TWH estimation; and a more detailed focus on stillbirths. With the natality microdata from the Czech Statistical Office and data concerning the level of air pollution in the Czech Republic from the Czech Hydrometeorological Institute, I analyze if the biological and socio-economics status of mothers and the characteristics of our surroundings (air pollution) affect the sex of children. The results are insignificant or not robust across specifications. I identified three hypotheses which are most likely the reason for the insignificant results: a non-inclusion of the biological and socio-economical status of a father, insufficient diversity or evolutionarily novel environment in the Czech Republic. As a conclusion, the presented evidence suggests that stillbirths are random in the Czech Republic and that the sex ratio is not affected by the socio-economics status of mothers or the characteristics of our surroundings (pollution).

JEL Classification	J11, J13, J16, D01
Keywords	Trivers–Willard hypothesis; Sex ratio; Human population; Sex allocation theory

Author's e-mail	`marek.pazitka@gmail.com`
Supervisor's e-mail	`xstrd08@vse.cz`

Abstrakt

Trivers-Willardova hypotéza (TWH) uvádí, že rodiče v dobrých podmínkách budou vychylovat poměr pohlaví k synům a rodiče ve špatných podmínkách budou vychylovat poměr pohlaví k dcerám. Tato studie přispívá k dosavadní literatuře několika způsoby: využívá dostatečn? velký data set (N= 1 401 851) reprezentující populaci současné moderní společnosti; je první studie tématu v České republice; obsahuje zařazení znečištění vzduchu do odhadu TWH; a podrobněji se zaměřuje na děti narozené mrtvé. S porodními mikrodaty z Českého statistického úřadu a údaji ohledně úrovně znečištění ovzduší v České republice z Českého hydrometeorologického ústavu analyzuji, zda biologické a sociálně-ekonomické postavení matek a charakteristiky prostředí (znečištění vzduchu) ovlivňují pohlaví potomků. Výsledky nejsou statisticky signifikantní nebo postrádají robustnost napříč specifikacemi. Identifikoval jsem tři hypotézy, které jsou s největší pravděpodobností důvodem nevýznamnosti výsledků: chybějící informace ohledně biologického a socio-ekonomického postavení otce, nedostatečná diverzita v socio-ekonomickém statutu nebo nové evoluční prostředí v České republice. Zavěrem, předložené výsledky naznačují, že děti narozené mrtvé jsou v České republice náhodným jevem, a že na poměr pohlaví nemá vliv ani sociálně-ekonomické postavení matek, ani charakteristiky prostředí (znečištění vzduchu).

Klasifikace JEL	J11, J13, J16, D01
Klíčová slova	Trivers–Willardova hypotéza; Poměr pohlaví; Lidská populace; Teorie alokace pohlaví

E-mail autora	marek.pazitka@gmail.com
E-mail vedoucího práce	xstrd08@vse.cz

Contents

List of Tables

List of Figures

Acronyms

CHMI Czech Hydrometeorological Institute

CZSO Czech Statistical Office

DHS Demographic and Health Surveys

gTWH generalized Trivers-Willard hypothesis

PCA Principal component analysis

PM10 Particulate matter concentrations refer to fine suspended particulates less than 10 microns in diameter that are capable of penetrating deep into the respiratory tract and causing significant health damage.

RAB Resource allocation biasing

SRB Sex ratio biasing

TWH Trivers-Willard hypothesis

Master Thesis Proposal

Author	Marek Pažitka
Supervisor	Mgr. Ing. Dominik Stroukal
Proposed topic	Economics of Family: The Effect of Biological and Socio-economical Status and Air Pollution on the Sex of Children

Topic characteristics Trivers and Willard (1973) propose that parents might under some circumstances be able to vary the sex ratio of their offspring in order to maximize their reproductive success. The Trivers-Willard hypothesis (further stated TWH) predicts that mothers living in wealthy conditions (relatively big family income) will have more sons and mothers in poor conditions (relatively small family income) will have more daughters. This argument is based on the assumption, that mothers (parents) maximize the reproductive success of their offspring. Because the reproductive success of a male offspring tends to be more variable and resource sensitive it is more convenient to have a daughter in poor conditions and a son in wealthy conditions. In this case assortative mating does not hold.

Women prefer men with greater resources (more wealthy), so such men are able to attract a larger number of high-quality mates (Betzig, 1986) and simultaneously men prefer younger and more attractive women for their mates (Buss 1989; Kanazawa 2003), so women's reproductive success is largely orthogonal (not correlated) to parent's wealth. Further, women cannot produce a large number of offspring due to their greater obligatory in parental investment into each offspring (Trivers, 1972). As a consequence men with greater resources can afford more offspring, but for woman this does not hold. In other words, let's consider that parents maximize the number of grandchildren, so for the poor parents a girl is a safe bet (the reproductive success among women is quite stable) and for the wealthy ones it is more convenient to have a boy, because thanks to the resources provided by parents, they will tend to have more children than an average man. Trivers-Willard hypothesis has been further generalized by Kanazawa (2005).

A lot of research on TWH has been performed on nonhuman species. However, there are also studies, which test this hypothesis on humans. For example, Almond and Edlund (2007) use US natality data 1983-2001 (48 million births and 310 thousands infant deaths) and their findings confirms presence of patterns predicted by the TW hypothesis. They report two main findings: (i) married, better educated and younger mothers bore more sons and (ii) in case of the unmarried and young mother, male had higher probability of infant death.

In my thesis I will use data from the Czech Statistical Office (further stated CZSO) to test the above mentioned hypothesis in conditions of the Czech Republic. For my purposes I have following years of natality microdata available: 1992, 1994, 1996-2004, 2006, 2008 and 2010. On every child born in the Czech Republic in above mentioned time period, I posses several variables, for example, education of mother and father, marital status of mother, weight, gestation age, vitality or age of mother. From Czech Hydrometeorological Institute (further stated CHMI) I drew data concerning level of air pollution in the Czech Republic and data on prices of properties will be also drew from CZSO.

The thesis will be divided in two parts. The first part of the thesis will test the classical TWH. I will use only data from the CZSO, more specifically, I will use variables, which are connected with the socio-economic status (education of mother, marital status etc.) or with the characteristics of a child (weight, gestation age - because these could differ in case of a boy or a girl). More educated parents have more resources (higher income), so I expect that they will have more sons. Also marital status could influence sex of a child, for example, Darwin (1874) reports lower percentage of boys born among unmarried women or more recent study of Whiting (1993) reports a lower proportion of sons among polygynously married women in Kenya. Variables concerning characteristics of a child will be used as additional control variables. The test of TWH will be performed by multiple regression model.

The second part will test the TWH combined with the connection of air pollution and prices of properties in particular area. In more polluted areas, prices of properties are lower than prices of properties in less polluted areas (Polinsky and Shavell 1975), so as a consequence people have less money in more polluted areas (all else equal). As a result in more polluted areas, people live in "poorer conditions" (their wealth is lower) and because of that it is more convenient to have a daughter. The reason for testing this hypothesis on different data set is that the data from the CZSO and the data from CHMI overlap only in few years.

Core bibliography

Almond, D., and L. Edlund. 2007. "Trivers-Willard at Birth and One Year: Evidence from US Natality Data 1983-2001." *Proceedings of the Royal Society B: Biological Sciences* 274 (1624): 2491-2496.

Betzig, L. L. 1986. *Despotism and Differential Reproduction: A Darwinian View of History.* New York : Aldine Publishing Co.

Buss, D. M. 1989. "Sex Differences in Human Mate Preferences: Evolutionary Hypotheses Tested in 37 Cultures." *Behavioral and Brain Sciences* 12 (1): 1-49.

Darwin, C. 1874. *The Descent of Man.* 2nd ed. New York: Random House.

Kanazawa, S. 2003. "Can Evolutionary Psychology Explain Reproductive Behavior in the Contemporary United States?" *The Sociological Quarterly* 44 (2): 291-302.

Kanazawa, S. 2005. "Big and Tall Parents Have More Sons: Further Generalizations of the Trivers-Willard Hypothesis." *Journal of Theoretical Biology* 235 (4): 583-590.

Polinsky, A. M., and S. Shavell. 1975. "The Air Pollution and Property Value Debate." *The Review of Economics and Statistics* 57 (1): 100-104.

Trivers, R. L. 1972. *Parental Investment and Sexual Selection.* In Sexual Selection and the Descent of Man, 1871-1971, ed. B. G. Campbell, 136-179. 1st ed. Chicago: Aldine.

Trivers, R. L., and D. E. Willard. 1973. "Natural Selection of Parental

Ability to Vary the Sex Ratio of Offspring." *Science* 179 (1): 90-92. Whiting, John W. M. 1993. "The Effect of Polygyny on Sex Ratio at Birth." *American Anthropologist* 95 (2): 435-442.

Introduction

Trivers and Willard (1973) propose that if parents want to maximize their number of offspring, they should vary the sex ratio of their children in order to maximize their reproductive success. The Trivers-Willard hypothesis (further stated TWH) states that mothers living in good conditions (relatively big family income, high education of parents or a good environment) will have more sons and mothers in poor conditions (relatively small family income, low education of parents or a bad environment) will have more daughters. The reproductive success of a male offspring tends to have a bigger variance and it is more resource sensitive, so it is more convenient to have a daughter in poor conditions and a son in good conditions. In other words, women are taken as a "safe bet", since women almost always have some offspring in contrast to men who are taken only in the case when sufficient resources are available to ensure that they will have children.

The TWH poses several interesting questions. For example, if the TWH mechanism works, then there are reasons to believe that the effect is stronger in more unequal countries.[1] If this is the case, then the TWH can contribute into the social mobility, since parents in poor conditions will be more likely to have a girl and parents in good conditions will be more likely to have a boy and the probability that those two will marry will increase. In contrast, one may pose the question as to whether the TWH contributes to the selection of abortions? Moreover, when we turn to the after birth TWH, where more resources are allocated to a boy by parents in good conditions, meanwhile parents in poor conditions should invest more into a girl, we can raise other relevant questions like: Are mothers in good conditions with a girl, expected to return to the workforce earlier than those with a boy (or vice versa)? Or, is the TWH mechanism at least partly responsible for a gap between the education of men and women? Lastly, are fathers with a boy more interested in their

[1] I will discuss this matter further into the thesis.

paternal investments or in having the mother breastfeed longer?

There is a substantial amount of literature on the TWH. An excellent example is a study from Almond and Edlund (2007), on which I built on and expanded in multiple ways. Almond and Edlund use general samples - US natality data from 1983-2001 (48 million births and 310 thousands infant deaths) and they find results consistent with the TWH. The study provides two main results. The first being that, married, better educated and younger mothers bore more sons and, the second, infant deaths of males were greater if the mother was young and unmarried.

My thesis has several contributions to the contemporary literature. Firstly, the tests of the TWH are rare in modern contemporary societies, moreover, the test on general country population data sets are extremely uncommon, since most of the research is done on specific samples (not random samples).[2] Secondly, to my knowledge, I am the first in considering the TWH in the Czech Republic. Thirdly, I connected the TWH with the capitalization hypothesis. In other words, I tested air pollution as one of the determinants of the TWH.[3] And last but not least, I am focusing on stillbirths in more detail than most of the current papers.[4]

The main research question to be addressed is whether or not the biological and socio-economics status of mothers and the characteristics of surroundings (like air pollution) affect the sex of children. In order to test the hypothesis, I have drawn natality microdata from the Czech Statistical Office (further stated as CZSO) and data concerning the level of air pollution in the Czech Republic from the Czech Hydrometeorological Institute (further stated as CHMI). With the model based on the Ordinary Least Squares estimation, I did not find any strong statistically significant effects of biological and socio-economics status of mothers or the characteristics of surroundings (like air pollution). Moreover, according to my results, the stillbirths are literally random (or at least not affected by any variable which I possess), except for the weak association of the stillbirths with air pollution (but the statistical significance is low and the effect is not robust across specifications). I identified three hypotheses which are most likely to be the reason for the insignificant results. The first is a non-inclusion of the biological and socio-economical status of a father the second

[2]Later, in the thesis I use the term "general sample" as a denomination for a full sample of the general population.

[3]It should be noted, that there are two possible mechanisms connected with air pollution - I present them both in Section 1.3.

[4]The deaths of a child which occurred late in a pregnancy (28 weeks or more).

and the third identify the possible insufficient diversity or novel, evolutionary environment in the Czech Republic.

The balance of my work proceeds as follows. In Chapter 1, I present a theoretical framework for my thesis, which contains three parts - a presentation of the Trivers-Willard hypothesis, the research done so far on the Trivers-Willard hypothesis and the relationship between air pollution and real estate prices. In Chapter 2, I introduce my data set, the creation of the variable Index of property prices and descriptive statistics. Chapter 3 presents the estimation methodology and in Chapter 4, I present the results. The Chapter 5 discusses the most likely reasons for the non-significant results and concludes the thesis with potential fields for future research.

Chapter 1

Theoretical Background

1.1 The Trivers-Willard hypothesis

Trivers and Willard (1973) show that under certain well-defined conditions, natural selection favors systematic deviations from a natural sex ratio 105-106[5] at conception, and that these deviations tend to cancel out in the local breeding population.

The TWH predicts that mothers living in wealthy conditions (relatively big parental investments) will have more sons and the mothers in poor conditions (relatively small parental investments) will have more daughters.[6] This hypothesis is based on the assumption, that mothers (parents) maximize the reproductive success of their offspring. Because the reproductive success of a male offspring tends to be more variable and resource sensitive, it is more convenient to have a daughter in poor conditions and a son in wealthy conditions. For better understanding, I will introduce an analogy with financial markets. In good conditions (a "bull" market), is risky investment usually a better choice (stocks); in poor conditions (a "bear" market), can less risky option (bonds or bank accounts) protect investor against huge losses and so it is usually better. Financial market investment is guided by a conscious decision of the return maximization. In contrast, the TWH says nothing about the conscious motivations. Hopcroft (2005) comments this issue: "*It is simply that in animal and human evolution, those behavioral strategies that have in*

[5]The "natural" sex ratio in the human population at birth is about 105-106 boys for every 100 girls (Lazarus, 2002). Current sex ratio of 107 boys to 100 girls (CIA Fact Book) at birth is bias due to the sex-selective abortions and "gendercide", the killing of female infants, in countries such as China and India, where males are more desired.

[6]The term "parental investments" can denote many things, for example, the time spend with a child, a period of breastfeeding, the costs of education or health care, ect.

the past favored reproductive success have been selected and have proliferated in the population; those strategies that did not favor reproductive success have literally died out."(p. 1114).

There are at least two economic reasons, why should parents behave according to the TWH. The first, parents should invest resources with the highest rate of return. In this case, the rate of return is the offspring's reproductive success. When a variance in the reproductive success is greater for males than for females, and when the male's reproductive success is more resources sensitive, parents will invest more into males (or they will want a male rather than a female), because then they will have a greater return per investment (Clutton-Brock and Albon, 1982). The second reason is that, costs to the mother's future reproductive potential of producing a son versus a daughter vary with the maternal conditions, because the mother should invest more into sons. It is necessary to take into account the effect of a given unit of investment not only on the reproductive success of the offspring, but also on the mother's future reproductive success (Clutton-Brock and Iason, 1986). For example, a poor mother cannot invest many other things except time spent with a child. Moreover, because she comes from poor conditions she probably has to go back to work earlier after a pregnancy than a mother from good conditions.[7] So, a female child can be in some circumstances cheaper and more advantageous investment for a mother in poor conditions in terms of offspring's reproductive success, the mother's reproductive success and the cost of children.

The model of Trivers-Willard's depends on the three assumptions:

1) Condition of a child at the end of parental investment will be correlated with the conditions of parents during the parental investments. For example, more educated people will produce more educated children.[8] Similarly, children of wealthy people will be wealthier.

2) Differences in the children's conditions at the end of the parental investment will tend to endure into an adulthood. In the case of humans, the parental investments are quite long and also not pre-defined. In other words, parents can support their offspring for a whole life.

[7]Obviously, this does not have to be always true. It vitally depends on two things - family's income and women's income. Bigger the women's income, bigger the opportunity costs. But when family's income is very low, even small increment from mother could be a lot.

[8]Oreopoulos et al. (2006) show that there is causal effect of the parent's education on the children's education. Similar results provide Sewell and Shah (1968).

3) Resources have heterogeneous effects on the male's/female's reproductive success. In the Figure 1.1 and the Figure 1.2 I present those heterogeneous effects on the reproductive success of males and females. The pictures are only illustrative, they are not based on any data. In both cases I relate resources (available stock for parental investments) on the horizontal axis and a number of children on the vertical one. You can see that the reproductive success of males is much more resources-sensitive than the females one.

Reasoning for this assumption is following. Women prefer men with greater resources (wealthier), so such men are able to attract larger number of high-quality mates (Betzig, 1986) and simultaneously men prefer younger and more attractive women for their mates (Buss 1989; Kanazawa 2003). As a result the women's reproductive success is largely orthogonal (not correlated) to the parent's wealth. Further, women cannot produce large number of offspring due to their greater obligatory in parental investment into each offspring (Trivers, 1972). As a consequence, men with greater resources can afford more offspring, but for woman this does not hold.

One of the problems with this assumption could be paternal investments (investment of a father). Trivers and Willard (1973) comment this as following: "*The application of the model to humans is complicated by the tendency for males to invest parental effort in their young (which reduces variance in male reproductive success), and by the importance of kin interactions among adults. Despite these complications, the model can be applied to humans differentiated on a socioeconomic scale, as long as the reproductive success of a male at the upper end of the scale exceeds his sister's, while that of a female at the lower end of the scale exceeds her brother's.*" (p. 91). Further, Pérusse (1993) found that the male reproductive success in the Quebec in the range 1988-1989 did not increase with status (as measured by a composite of income, education, and occupational prestige), but it did lead to greater sexual access. So, it is not unlikely that in moder societies, this assumption does not hold. On the other hand, Cameron and Dalerum (2008), investigating the population of the wealthiest people on world (1000 people from the Forbes' billionare)[9], write: "*To test the assumptions of the TWH, we used only those billionaires that had children, and found that male billionaires had significantly more children and a more variable number of children than female billionaires (men: 1 to 61 children, women: 1 to 7 children).*" (p. 2).

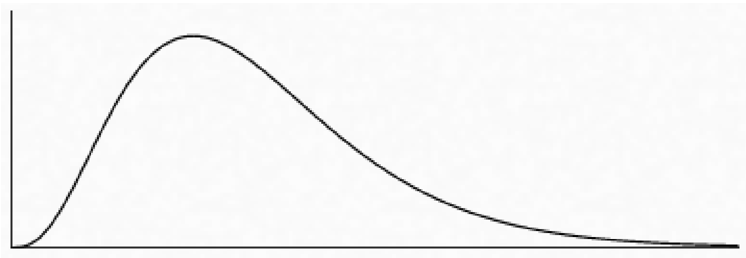
Nevertheless, Hopcroft (2005) argue that the differences in the reproductive success does not need to hold today in order to be the TWH valid: "*Given this situation in the evolutionary environment, by the logic of Trivers-Willard, there may exist evolved psychological and physiological mechanisms that promote high-status parents to invest more in sons and low-status parents to invest more in daughters,regardless of any contemporary sex differences in reproductive success. Given such evolved mechanisms, we can expect high-status parents to invest more in sons, and low-status parents to invest more in daughters, even if males are not actually more reproductively successful in the contemporary environment.*" (p. 1115). This statement is also supported by the study of Freese and Powell (1999): "*Once such a tendency has evolved, its influence on parental investment should persist even in evolutionary environments in which a Trivers-Willard effect does not contribute to greater fertility (e.g., in contemporary American society and others in which social status and number of offspring are not positively related).*" (p. 1710). Similarly, Wright (1994, p.173) claims that the TWH works "*by shaping human feelings, not by making humans conscious of its logic.*"

As a conclusion, we can state, that the third assumption can change to this form: Resources have or at least *had* heterogeneous effects on the male/female reproductive success. Supplementary, Hill (1984) and Irons (1993) study contemporary hunter-gatherer societies, they state that it is probable that a trustworthy relationship between reproductive success and male status existed among human ancestors. To support this claim even more I introduce the works of Nielsen (1994) and Crawford (1998), which claim that modern, developed societies have not existed long enough to reverse or substantively alter the cognitive mechanisms that have evolved over the last thousands or millions of years.

Before concluding this chapter, it should be noted that Trivers and Willard

[9]Due to some missing data, especially in the case of the sex of the children, the final number of the observations in their analysis is 399.

Figure 1.1: The reproductive success of males - illustrative

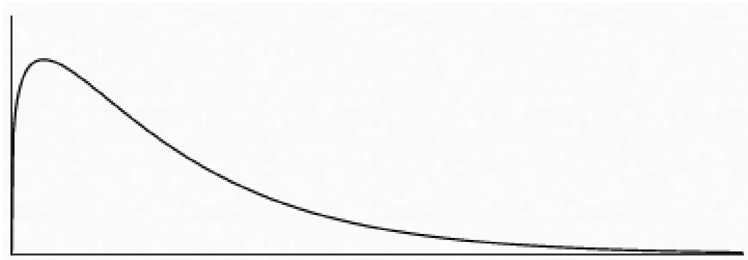


Note: The vertical axis depicts a number of children and the horizontal depicts an amount of resources.
Source: Picture is illustrative. It is not based on a real data.

(1973) claim that their hypothesis applies to parents' behavior toward children after the birth just as it applies to the sex ratio. They literally write: "*If the model is correct, natural selection favors deviations away from 50/50 investment in the sexes, rather than deviations in sex ratios per se. In species with a long period of* [parental investment] *after birth of young* [such as humans]*, one might expect biases in parental behavior toward offspring of different sex, according to parental condition; parents in better condition would be expected to show a bias toward male offspring*" (p. 91). We can expect that low-status parents will invest more into a female offspring than into a male offspring, while high-status parents will invest more into a male offspring than into a female offspring. To sum up, the TWH can be divided into the two separate hypotheses. The first, alteration of the sex ratio depending upon maternal conditions, both in utero or through the infanticide after the birth, is called the sex ratio biasing (further stated the SRB). The second, an allocation of more resources to offspring of one sex after the birth (a bigger amount of parental investments) depending upon parent's condition (family income, education, age or health) is called the resource allocation biasing (further stated the RAB) (Keller et al., 2001). In the case of humans, it is not only providing a nutrition to children while they are young, but it is also investments into education and the social and cultural development of children (Hopcroft, 2005).

Interestingly, there could be the distinction between the RAB and the SRB predictions. I am not the first to emphasize the distinction and its implication for predictions based on the TWH. Anderson and Crawford (1993) address the question with a simple model: "*Under what conditions do the parental behaviors that maximize numbers of grandchildren resemble the Trivers-Willard rules of thumb?*" (p. 151). Using data from !Kung of the southern Africa,

Figure 1.2: The reproductive success of females - illustrative



Note: The vertical axis depicts a number of children and the horizontal depicts an amount of resources.
Source: Picture is illustrative. It is not based on a real data.

they find that the optimal sex ratios are heavily influenced by the existing children of different ages and sexes in ways not predicted by the TWH. For this reasons some researches use only the first child in their analysis.[10] For the RAB, Anderson and Crawford state that an optimum parental behavior is sensitive to population dynamics, type of parental investment and, most importantly, relative ages of sons and daughters. Moreover, they conclude that it is doubtful whether the TWH rules (for the RAB) would maximize the number of descendants.

Another critique of the RAB is Keller (2001), who argues that the TWH does not, in fact, predict the RAB to the existing offspring. According to authors, the TWH should be limited for predicting those parental investments (e.g., the SRB, protection), which are related to a fitness value. They use following example: "*As a simple example, consider two mothers in equally poor condition, one that has a son and one a daughter. Given the TWH assumptions, the mother with the daughter should have fitness advantage over the mother with the son due to the different fitness values of the offspring; this is why selection favors a female-biased sex ratio for mothers in poor condition. However, the low-condition mother with the son should invest more in the son than the low condition mother does in the daughter if the marginal benefit of investing additional resources is greater for sons. The bias for SRB in this case is in the opposite direction from the RAB bias.*". (p. 357). As a conclusion, a prediction about the RAB are far more complex and a simple prediction of the TWH about this matter can be often non-maximization strategy. From now on, I will focus on the SRB only.

[10]For example, Kanazawa and Apari (2009).

1.2 Research on the Trivers-Willard hypothesis

In the following lines I will present the TWH research, I will start with the animal research, continue with the one on humans and conclude with the generalized Trivers-Willard hypothesis (further stated the gTWH).

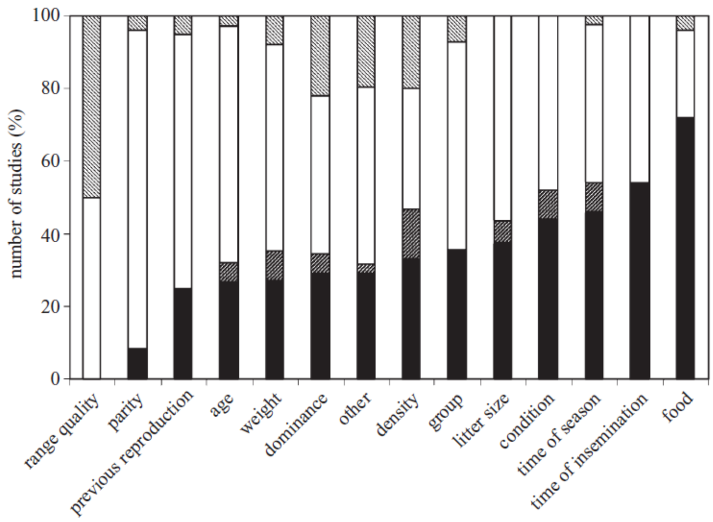
1.2.1 The Trivers-Willard hypothesis - research on nonhuman species

Probably the most classic example of the TWH on nonhuman species is research on red deer (Clutton-Brock, Albon, and Guinness, 1984, 1986). Clutton-Brock, Albon, and Guinness's (1984) show that, in polygynous red deer, dominant mothers produce significantly more sons than subordinates and that a maternal rank has a greater effect on the breeding success of males than females. In another words, red deer tend to inherit their ranks and the ranks are more important to the RS of males compared to females. Clutton-Brock, Albon, and Guinness's (1986) extends the analysis and provide the robustness check.

Symington (1987) use data from a long-term field study of the spider monkey in Peru. He reports that from 46 infants born between July 1981 and June 1986, 12 were male, 32 were female and 2 were of an undetermined sex. According to the author, the bias is caused by the fact that low-ranking females produce daughters almost exclusively and although high-ranking females produce males, the bias towards sons is not so strong as the bias towards daughters in the case of low-ranking females. Another example, Rivers and Crawford (1974) study the changes in nutritions on mice and they find out results consistent with the TWH.

Keller et al. (2001) map the nonhuman literature and find 62 studies related to the SRB. They comment their finding:"*Most published studies on SRB do show the theoretically predicted sex ratio bias - mothers in good condition tend to give birth to more sons, and vice versa for mothers in poor condition. Some studies, however, show no effect or show the opposite pattern in the sex ratio.*" (p. 344). However, authors are cautious about drawing some strong conclusion from the above-mentioned survey for two reasons. The first, whole science suffers from a publication bias - there is a tendency towards publishing positive results. Festa-Bianchet (1996) suggests that the publication bias can be especially large in this kind of literature. The second, positive results do not necessarily refute nonadaptive explanations, Clutton-Brock (1991) reports that sons are constitutionally weaker, and thus are more prone than daughters to die under adverse conditions, either in utero or postnatally. This can cause the positive results in case of the TWH, but Keller et al. (2001) note one last thing to this matter: "*To further complicate matters, "contradictory" findings are not necessarily contradictory: if maternal status correlates more positively with daughters' than sons' RS, this would reverse the prediction of the TWH.*" (p. 344).

Figure 1.3: The studies testing the TWH in the relation to a condition measure

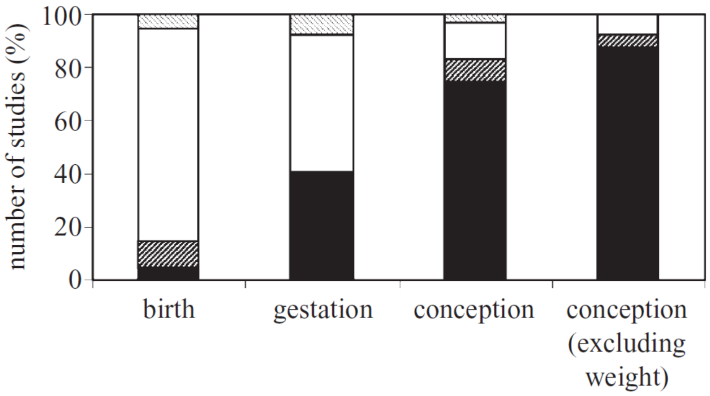


Note: A significant support (black), a non-significant trend to support (light hatching), no significant support (white) or a significant result in the opposite direction (dark hatching).
Source:Cameron (2004).

More recent study of Cameron (2004) performed a meta-analysis on the previous studies, investigating sex ratios biasing in all mammalian taxa (except humans).[11] The goal of the paper was to locate all the studies that investigated the sex-ratio variation in mammals. The method of the paper is quite sophisticated, it counts with the positive publication bias - studies that found results supporting the TWH are more likely to cite the TWH (Festa-Bianchet, 1996) and takes necessary steps in order to avoid it (author searches for all papers that mentioned "sex ratio" in the abstract in the Science Citation Index and the Current Contents). The study takes papers about the SRB in all mammalian taxa until the 15th April 2003.

[11]Cameron (2004) provide all papers used in an electronic appendix, which is available on the page: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1691777/bin/15306293s01.pdf.

Figure 1.4: The studies testing the TWH that measured a body condition at a specific time in the reproductive cycle



Note: A significant support (black), a non-significant trend to support (light hatching), no significant support (white) or a significant result in the opposite direction (dark hatching). The condition measures included are body weight, body condition and food availability. Source: Cameron (2004).

In the Figure 1.3 I present a graph from Cameron (2004), where particular measures are depict. So columns mean - a significant support (black), a non-significant trend to support (light hatching), no significant support (white) or a significant result in an opposite direction (dark hatching). There is a quite strong support for a claim that, for example, food affects the formation of the TWH. In the Figure 1.4 I depict a second graph from Cameron (2004). This one shows the time of measurement of a particular body condition. This procedure can indicate, if there was a stronger support, if the measure is taken close to the conception. You can see that if body condition, weight or food are measured or manipulated around the conception, the 74% of the included studies support the TWH, but if body condition, weight or food are measured or manipulated during the gestation (the 41% of included papers find the significant results) or at the birth (only the 5% find the significant results), the literature show a smaller evidence for the relationship. Authors conclude analysis: "*The TWH is consistently supported if the measure used relates to variation in maternal condition around conception.*" (p. 1725). The idea that the sex-ratio adjustment may occur around the conception was proposed by some of the previous works (Cameron et al., 1999; Enright et al., 2001), and an another meta-analysis on ungulates (Sheldon and West, 2004) also provide evidence that measures taken around conception are the most consistent support for the TWH.

1.2.2 The Trivers-Willard hypothesis - research on humans

Before proceeding towards particular papers, I should remind one challenge in the estimation of the TWH. It is probable that the TWH hypothesis may not be observed in societies that are resource-rich compared to ancestral environments. However, across time and area, the absolute level of resources varied considerably between ancestral groups, yet within each group, high-status males had the higher RS than low-status males (Keller et al., 2001). Keller et al. (2001) further comment on this issue: "*Because of this, a mechanism tracking absolute inputs (e.g., calories) would be disadvantageous compared to a mechanism that tracked relative inputs (e.g., status). Therefore, it is unlikely that the absolute level of resources has an effect on the putative TW mechanism.*" (p. 347). Koziel and Ulijaszek (2001) added that the TWH will probably reveal itself in the populations, where the extent of social stratification is sufficiently diverse.

The pattern, that mothers in good condition (married, better educated and younger) bore more sons than those in poor condition, was firstly noted by Darwin (1871, p. 281): "*mysterious fact that...the excess of male to female births is less when they are illegitimate than legitimate*". Based on the observation he hypothesizes that males suffer disproportionately from adverse conditions.

Perhaps the best-known report on the TWH on humans is the Dickemann's (1979) analysis of female infanticide in historically hypergynous societies.[12] Her main result is that the female infanticide was widespread among the upper classes in historically hypergynous societies, because higher-class daughters had smaller chance to "marry up" compared to higher-class sons. This fact can be interpreted as an indicator of an unwillingness to invest into the children of the murdered sex (higher-class daughters). However, other theoretical papers discuss that infanticide should rarely be an adaptive strategy (for example, Anderson and Crawford, 1993). Moreover, as Keller et al. (2001) remind that the sex bias produced by infanticide can be the opposite of that predicted by the TWH. As an example, we can take the study of Voland, Siegelkow, and Engel (1991), who investigate high-status 18th and 19th century Germans in Krummhorn. They report that the high-status Germans were more likely to perpetrate a male infanticide in that time. This strategy helped them to keep their property undivided. So infanticide can sometimes be consistent with the TWH, but it depends on a situation and on immediate goals of parents.

Since the original paper of Trivers and Willard (1973), there were a lot of at-

[12]After Dickemann also Betzig et al. (1988) similar results.

tempts to test the TWH on a real-world data. I will firstly present those, which are based on data from the less developed countries. Both Chacon-Puignau et al. (1996) and Zaldivar et al. (1991) test the TWH in Venezuela. Chacon-Puignau et al. (1996) with a great sample from the national birth registry of Venezuela (578,000 observations) observe only very small effect of marital status of women on the sex ratio. In contrast, Zaldivar et al. (1991) report no relationship between the status and the birth sex ratios or the sex ratios at later ages. Using data on the Gabbra pastoralists of Kenya, Mace (1996) also report no relationship between the status and the sex ratio at birth or among living children. On the other hand, Cronk (1989) investigating the Mukogodo children (ages 0-4) finds that this population was female-biased and since the Mukogodo are at the bottom of a regional hierarchy, the results are consistent with the TWH. Finally, more recent paper, Jha et al. (2006) investigate the sex ratio on population of India (N=133,738). Their main result in our context is that better educated women (grade 10 or more) had a significantly lower adjusted sex ratio (683 girls on 1000 boys, 99% CI 610-756) than illiterate women (869 girls on 1000 boys, 99% CI 820-917). We can see that the results of the TWH in the less developed countries are at least mixed.

From an overall view, test on the TWH in more developed societies are also inconclusive. Betzig and Weber (1995) used data on men in the U.S. executive branch, including presidents, vice presidents, and cabinet secretaries. They report that agents in their sample produce more sons than daughters in the first cohort (Presidents Washington through Garfield), however, in the second cohort (Presidents Arthur through Reagan) they produced roughly equal numbers of sons and daughters. Essock-Vitale (1984) provides a study of the number of children among the Forbes list of the 400 wealthiest Americans. She finds that they had on average more children than the general U.S. population and that the wealthy Americans appear equally likely to have sons as to have daughters. Cameron and Dalerum (2008) provide a similar study on the sample of the 1000 wealthiest people in the world (they also use the Forbe's billionaire list). However, since their original data are incomplete, they use just the 399 observation (350 male billionaires and 49 female billionaires). Their concluding the study with: "*Humans in the highest economic bracket leave more grandchildren through sons than through daughters. Therefore, adaptive variation in sex ratios is expected, and human mothers in the highest economic bracket do give birth to more sons, suggesting similar sex ratio manipulation as seen in other mammals.*" (p. 1).

Schnettler (2013) continues in the works of Essock-Vitale (1984) and Cameron, Dalerum (2008). His goal is to shed light on the matter of mixed results in the literature. He proposes two hypothesis - sample selection (mostly due to the missing data) and a lacking specification of the timing of the wealth accumulation. He corrects both problems. Firstly, the analysis is based on a data set of U.S. billionaires with a near-complete information on the sex of offspring. Secondly, subgroups of billionaires are distinguished according to the timing of the wealth accumulation. Although he finds that the result on the hypothesis that billionaires will have a higher share of male offspring than the general population are not consistent for all subgroups of billionaires, he also reports that heirs, but not self-made billionaires, have a higher share of male offspring than the U.S. population. Contrary to this finding, author also reports that heiresses have a much lower share of male offspring than the U.S. average. This results imply that there are other mechanisms affecting the sex birth ratio as well, nevertheless, they are not uncovered in the paper.

There are also several papers, which use a general population sample (not only some exclusive lists). Abernethy and Yip (1990) use linked birth-death records from years 1976 to 1983 for the state of Tennessee. After a stratification of a sample by socioeconomic indicators authors find the pattern for post-neonatal infant deaths to be supportive of the TWH. Norberg (2004) investigates a maternal partnership status at the time of conception, he finds that the status can be taken as a determinant of the sex ratio. Author reports: "*In a sample of 86 436 human births pooled from five US population-based surveys, I found 51.5% male births reported by respondents who were living with a spouse or partner before the child's conception or birth, and 49.9% male births reported by respondents who were not* (χ^2 *= 16.77, d.f. = 1,* $p < 0.0001$*).*" (p. 2403). Almond and Edlund (2007) analyze the US-linked births and infant deaths of the white mothers ranging from 1983 to 2001. Their total number of the observations are approximately 48 million births and 310 000 infant deaths. The study provides two main results. The first, married, better educated and younger mothers bore more sons and, the second, infant deaths were more male if the mother was unmarried and young. Finally, Guggenheim et al. (2007) provide a comprehensive analysis of nationally representative samples from 35 countries (survey data were collected by the Demographic and Health Surveys - DHS) and reports that the analyses do not support the TWH, but there is an evidence of regional and country level differences.

1.2.3 The generalized Trivers-Willard hypothesis

In his paper Kanazava (2005) suggests that the TWH is just a special case of the gTWH which has following form:

"*gTWH: Parents who possess any heritable trait which increases the male reproductive success at a greater rate (or decreases the male reproductive success at a smaller rate) than female reproductive success in a given environment will have a higher-than-expected offspring sex ratio (more males). Parents who possess any heritable trait which increases the female reproductive success at a greater rate (or decreases the female reproductive success at a smaller rate) than male reproductive success in a given environment will have a lower-than-expected offspring sex ratio (more females).*" (p. 585).

Parental wealth and status are two heritable (at least culturally, if not genetically) traits of parents which can increase the sons' reproductive success, but they are largely orthogonal to the daughters' reproductive success.

The research on the gTWH was mainly performed on humans, but there are exceptions. For example, the Burley's (1986) study investigating zebra finches finds that parents adjusted the sex ratio to sons or daughters in order to produce offspring they expect to be most attractive.

When we turn our interest toward the human literature, the natural starting point is the study of Kanazawa (2005). Kanazawa (2005) hypothesizes that parent who are taller have a higher-than-expected number of sons and similarly parents who are heavier have a higher-than-expected number of sons. Kanazawa (2005) explains the hypothesis: "*One highly heritable phenotype which influences sex-specific reproductive success is the body size. In the ancestral environment, where male intra-sexual competition was both fierce (in the absence of socially imposed monogamy) and largely if not entirely physical, big and tall men had particular advantages over smaller and shorter men. In contrast, large body size was not particularly adaptive for ancestral women.*". (p. 585). Nettle (2002a) or Pawlowski et al. (2000) document that to this day taller men has a greater reproductive success than shorter men and similarly Nettle (2002b) shows that for women the exactly opposite is true. Together with the fact that the body size is substantially heritable (Silventoinen et al., 2001), we have the complete rationale for the hypothesis. The study use the National Child Development Survey and the British Cohort Survey and those samples largely support the hypothesis.

Kanazawa and Vandermassen (2005) use the work of Baron-Cohen (1999, 2002, 2003; Baron-Cohen and Hammer, 1997; Baron-Cohen et al., 2004), who propose that there are "Type S brains" (ideal for systematizing used by engineers and mathematicians, for example) and "Type E brains" (ideal for empathizing used by nurses and school teachers, for example), which are substantially heritable. Based on this fact Kanazawa and Vandermassen suggest that people with strong male brains have more sons and people with strong female brains have more daughters. They used US General Social Survey from 1994 and report that the data support the hypothesis. Kanazawa (2006) investigates another heritable trait - tendency toward a violence and an aggression.[13] Authors report that women of violent men have significantly more sons than usual on data from the US General Social Survey (1994) and combined sample of the National Child Development Study (1999-2000) and the 1970 British Cohort Study. Finally, Kanazawa (2007a) investigates beauty, which is also heritable and increases the reproductive success of daughters (Buss 1989; Kanazawa 2003). The analysis suggests that the predictions of the gTWH are justified.

Gelman (2007), however, criticizes Kanazava's papers (Kanazawa, 2005, 2006, 2007a; Kanazawa and Vandermassen, 2005) for an incorrect use of statistical method and a bad interpretation of some results. Arguably the biggest problem of the Kanazava's papers is the usage of so called "endogenous variables", such as, the number of children. Gelman (2007) uses the paper Kanazawa and Vandermassen (2005) to demonstrate the incorrect usage - the total number of children can be affected by the predictor of interest (the parent's occupation), because different couples may try for a boy or for a girl.

Gelman explains the problem on a simple model. The assumptions are: 1) there are only engineers and nurses; 2) the probability of having a boy is exactly 50% for everybody; 3) engineers will stop at one child if the first child is a boy, but if the first child is a girl, they will have a second child and the 4) nurses will stop at one child with probability 30% and continue on to a second child with probability 70%, regardless of the sex of the first child.

Under this scenario, 75% of engineers will have at least one boy, but only

[13] Kanazava summarize his paper in subsequent work as: "*He first points out that violence and aggression were adaptive for men (but not for women) in the ancestral environment, where much of male intra-sexual competition for status and thus reproductive access to women was physical; violent and aggressive men may therefore have often had greater reproductive success in the ancestral environment than less violent and aggressive men. Kanazawa then notes that men's tendency toward violence and aggression, particularly, their tendency toward domestic violence, is a function of their baseline levels of testosterone (Booth and Osgood, 1993; Dabbs and Morris, 1990; Soler et al., 2000), and that testosterone levels are highly heritable (h^2 = 0.60) (Harris et al. 1998).*" (2005, p. 584).

50% of them will have at least one girl - family types and distribution: b 50%; gb 25%; gg 25%. Meanwhile, 57.5% of nurses have at least one boy and 57.5% have at least one girl - family types and distribution: b 15%; g 15%; bb 17.5%; bg 17.5%; gb 17.5%; gg 17.5%. The Table 1.1 is from the Gelman's study (2007) - he simulated 800 families (400 from engineers, 400 from nurses) under the probabilities described above and then he regress the number of boys on a parental occupation and the number of girls.

Table 1.1: The regression with a dependent of the number of boys

Variables	Coefficients	Standard error of coefficients
Intercept	1.18	0.02
Engineer*	0.14	0.02
Number of girls	0.56	0.02

Note: Variable with sign * is categorical with values from 0 to 1.
Source: Gelman (2007).

The problem is evident in the table. The coefficient for "engineer" is negative and statistically significant, but they are not more likely to have a son. Gelman (2007) explains: "*The nonzero coefficient is an artifact arising from controlling for the number of girls, a variable that is influenced by the other regression predictor.*" (p. 597). Yamaguchi and Ferguson (1995) provide evidence that couples with two boys or girls are more likely to have a third child than couples with a boy and a girl, so we can state that different couples have different stopping rules.

Kanazawa continued in his research and provided some evidence for the gTWH: the socio-sexually unrestricted parents have more sons (Kanazawa, 2009)[14] and again that big and tall parents are more likely to have sons (Kanazawa, 2007b). Admittedly, the gTWH have been a subject of criticism for a some time. For example, Denny (2008), Pollet and Nettle (2010) and Arcand and Rieger (2012) provide evidences that big and tall parent do not have more sons. Denny (2008) uses a British data, Pollet and Nettle (2010) usesa British and a Guatemalan data and finally Arcand and Rieger (2012) investigate a pooled sample of Demographic Health Surveys (DHS) from the 46 developing countries. Those studies cast some doubt on the result of the Kanazawa's papers (Kanazava. 2005; Kanazava 2007b).

[14] Using the US General Social Surveys and the National Longitudinal Study of Adolescent Health (Add Health).

1.2.4 Mechanism of the Trivers-Willard hypothesis

Most of the papers work with a simple mechanism for the TWH, for example, Hopcroft (2005):"*The mechanism typically proposed for the effect of status on sex ratio is the condition of the mother: high-status mothers are more likely to be in good condition than other mothers, and mothers in good condition are better able to carry a male fetus to term (Catalano 2003). High-status women are therefore more likely to have boys among their offspring.*". (p. 1117). Almond and Edlund (2007) argue in a similar way:"*TW conjectured that parental control over offspring mortality would be one mechanism. Obviously, the closer to conception, the lower the replacement cost of terminated offspring. Mortality in utero would for this reason be more advantageous than mortality after birth and most studies have focused on the sex ratio at birth.*" (p. 2491). However, as also Almond and Edlund (2007) add, the effects on barring prenatal sex determination or elective abortions are small (James, 1987).

Another branch of research focus on a maternal diet. Apparently, a diet high in saturated fats but low in carbohydrates results in the higher levels of circulating glucose (Folmer et al., 2003), which can lead to the birth of significantly more male than female offspring (on laboratory mice - Rosenfeld and Roberts, 2004). Mathews et al. (2008) investigate the effect on a human population. They use data on the 740 British women who were unaware of their foetus's gender. They report that:"*Fifty six per cent of women in the highest third of preconceptional energy intake bore boys, compared with 45% in the lowest third. Intakes during pregnancy were not associated with sex, suggesting that the foetus does not manipulate maternal diet.*" (p. 1661). Comparably, Gibson (2003) reports strong association between the sex of the most recent birth and a maternal-nutritional status in Kenya in 2000. Based on those evidences it seems likely that nutritional status of the mother (as an important part of the costs of reproduction) play a significant role in adjusting sex ratios.[15]

1.3 Air pollution and real estate prices

We will start with a simple yet informative model. Let's assume two identical communities, which differ only in the air pollution level. In the case of the zero moving costs and the same real estate prices everybody would want to live in the community with the lower level of air pollution. Such situation is

[15] For a more detailed description of the mechanism I suggest the study of Cameron (2004).

not sustainable - it is not an equilibrium. However, as people will move out from the more polluted community, the real estate prices will go down in the more polluted community. As a consequence of the moving people the real estate price in the less polluted community will increase. This mechanism will continue up to the point, where the two communities will be in the equilibrium - the difference between the real estate prices in the two communities will be greater than the marginal willingness to pay for reduction in air pollution of the people in the more polluted community. In this situation we can say that the pollution is capitalized in the real estate prices.

Polinsky and Shavell (1975) summarize the debate over the relationship between air pollution and the property values. They outline models and discuss the proper interpretations. Their more recent followers - Carriazo-Osorio (2001) provide the evidence for the hypothesis on data from Bogotá. They reports the following results:"*In the final analysis, estimations suggest that an increase of 1 per cent in the emission level of TSP decreases property values in 0.123 per cent. For the average housing price of Col.$37,506,800 (US$24,322), the marginal willingness to pay for a reduction of 1% in the emission levels is Col.$47,731 (US$31).*" (p. 1). Smith and Huang (1995) provide a meta-analysis. They include 37 studies with the 86 estimates for the marginal willingness to pay for reductions in air pollution. They report that the range for these estimated marginal values (measured as a change in asset prices) lies between zero and $98.52 (in 1982-84 dollars) for a one unit reduction in total suspended particulates (in micro grams per cubic meter). Furthermore, the mean of the marginal willingness to pay for the reductions in air pollution is about five times larger than the median ($109.90 vs. $22.40), so the outliers play an important role in any summary statistics for these estimates.

In addition to the fact that air pollution affect the real estate prices I include the study (Harrison et al., 1978) reporting examples of numerous cities, where suburban properties (with a smaller air pollution and an overall higher-level environment) are much more expensive than properties in the main core of the city.

The second mechanism, why air pollution can affect the sex ratio is motivated by the fact that male fetuses are more vulnerable and so more prone to miscarriages (Peterka, 2007). A lower quality of air can cause miscarriages and since boys are more likely to be miscarriage, they should be more often affected. In other words, we should observe more girls in more polluted areas.

Chapter 2

Data

2.1 The data set descriptions

2.1.1 The Czech Statistical Office data set

For my research I have used a micro-data on children and their parents from the CZSO. I have used the following years: 1992, 1994, 1996-2004, 2006, 2008 and 2010. Whole sample contains 1 355 902 observations with the below described variables.

Further, the data on prices of flats and houses are also from the CZSO. I posses the annual data on the prices of flats and houses for all districts in the time period of 2001 to 2011.I matched the prices to particular observations and the resulted number of observations is 712 746. The variables are included in the list below.

The list of variables from the CZSO

Birth date : The full date of birth (day, month and year).

Male : The dummy variable (male = 1, female = 0).

Vitality : The dummy variable, for a child born alive the variable is equal to 1 and 0 otherwise.

Multiple birth : A multiple birth occurs when more than one fetus is carried to term in a single pregnancy. In the case of a single birth, this variable is equal to 1, for twins it is 2, for triples it is 3 and so on.

Multiple birth - dummy : Equals 1, if a woman had multiple births and 0 otherwise.

Weight : The variable depicts a birth weight in grams.

Gestation age : The variable depicts a length of a pregnancy where the origin is the woman's last normal menstrual period and the end is birth. The variable is in weeks.

Czech citizen : Equal 1, if mother has Czech citizenship.

Number of previous children : The variable indicate how many children mother had before a current child. The total number of previous children is a value of the variable minus one, since for the first baby, the variable is equal to 1.

First or second child : Equal 1, if a mother born first or second child.

Single : 1 if a mother is single.

Divorced : 1 if a mother is divorced.

Married : 1 if a mother is married.

Widowed : 1 if a mother is widowed.

Education of father : The coding for this variable is as following: 1 = basic education; 2 = lower secondary education; 3 = secondary education with a completed maturity exam; 4 = first or a second stage of a tertiary education.

Education of mother : The coding is same as for the *Education of father*.

Mother's age : The age of a mother in years.

Flat prices : The average price per one square meter of a flat in a time of a child's birth in a district, where the child was born.

House prices : The average price per one square meter of a house in a time of a child's birth in a district, where the child was born.

2.1.2 The Czech Hydrometeorological Institute data

From the CHMI I drew daily data concerning the level of air pollution in the Czech Republic from 216 stations ranging from years 1993 to 2012. For my purpose, I do not need such detailed data set, because the prices of properties do not react on daily levels, but rather on a long term average levels. Because of this, I simplified my data on the pollution - I transformed them to the monthly averages.

The data set contains many missing observations. Ideally, the data should contain 216 (the number of stations) times 240 (the number of months in 20 years period) which is 51 840 measurements. In reality, the data contains only 28 513 measurements, so approximately 45 percent is missing. Nevertheless, the merged data set (the natality micro-data with the property prices and the pollution variable) contains 582 202 observations, so the decrease in the number of observation is not that drastic (around 18 percent of the sample).[16]

The air pollution variable is described below:

Air pollution (PM10) : Particulate matter concentrations refer to fine suspended particulates less than 10 microns in diameter (PM10) that are capable of penetrating deep into the respiratory tract and causing significant health damage (World Development Indicators, The World Bank).

2.2 The Index of property prices

The correlation between flats and houses prices is 0.93, which is quite large. Those data are ideal for a dimension reduction which in my case means that I will create a new variable - the *Index of property prices* representing the variables the *Flats prices* and the *Houses prices*. For such task I will use the principal component analysis, which is described in the upcoming paragraphs.

2.2.1 Principal component analysis - description of the method

In following section, I will explain the mechanism of the PCA. The intuition underlying the PCA is to capture most of the information in our observed variables $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p]$ with lesser number of new variables called principal components. This is done by finding a linear combination of original variables $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p]$ with maximum variance. For the PCA we need standardized $\boldsymbol{X}$, because some variables have a bigger variance then others and this difference could affect the results.[17] Let me denote the linear combination by

[16] This is also due to the usage of the monthly averages, since the daily data contain more than 49 percent of missing observables (there is 808 838 valid measurements from total 1 577 880).

[17] Think, for example, on variable property prices and number of children. I expect that in most cases variable property prices will have a much bigger variance and that would cause

vector $\boldsymbol{u}_i = (u_1, u_2, \ldots, u_p)$, then we can express the PCA goal for the first component as the maximizing variance of the elements of $\boldsymbol{z}_1 = \boldsymbol{X}\boldsymbol{u}_1$. The second component will then capture the largest amount of variance, which is not already captured by the first component and these new variables are orthogonal (uncorrelated). For other components, if there are any, the PCA will do exactly the same thing.

For deriving the PCA we will maximize $\boldsymbol{Z}$, which could be written as:

$$var(\boldsymbol{z}_i) = \frac{1}{n-1}\boldsymbol{u}_i'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{u}_i = \boldsymbol{u}_i'\boldsymbol{R}\boldsymbol{u}_i \tag{2.1}$$

where $\boldsymbol{R}$ is equal to $\frac{1}{n-1}\boldsymbol{X}'\boldsymbol{X}$, which is the sample correlation matrix. The Equation 2.1 now has a trivial solution - choose very large $\boldsymbol{u}_i$. For avoiding such a trivial solution we impose a constrain of the unit length on unit on the vector $\boldsymbol{u}_i'\boldsymbol{u}_i = 1$. Such constrain optimization problems could be solved by the method of Lagrange multipliers, such as:

$$\mathcal{L} = \boldsymbol{u}_i'\boldsymbol{R}\boldsymbol{u}_i - \lambda_i(\boldsymbol{u}_i'\boldsymbol{u}_i - 1) \tag{2.2}$$

where λ_i is called the Lagrange multiplier. The multiplier makes sure to penalize the objective function, if the equality constraint $\boldsymbol{u}_i'\boldsymbol{u}_i = 1$ is not met. Taking the derivative of $\mathcal{L}$ with respect to the elements of $\boldsymbol{u}_i$ yields:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{u}_i} = 2\boldsymbol{R}\boldsymbol{u}_i - 2\lambda_i\boldsymbol{u}_i \tag{2.3}$$

setting Equation 2.3 to zero and solving yields:

$$\boldsymbol{R}\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i \tag{2.4}$$

This equation has a special structure. It is known as an eigenvalue - eigenvector problem, where $\boldsymbol{u}_i$ is called the eigenvector and λ_i the eigenvalue. Provided that the correlation matrix is full rank, the solution will consist of p positive eigenvalues and associated eigenvectors, which have a special relationship with the variance of the principal components:

$$var(\boldsymbol{z}_i) = \boldsymbol{u}_i'\boldsymbol{R}\boldsymbol{u}_i = \boldsymbol{u}_i'\lambda_i\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i'\boldsymbol{u}_i = \lambda_i \tag{2.5}$$

the PCA to focus more on such a variable. Variables with relatively large variances could dominate results in the PCA and that is the reason for the standardization.

where we are using fact that $\boldsymbol{u}_i'\boldsymbol{u}_i = 1$. Thus, the eigenvalue λ_i is exactly the variance of associated principal component $\boldsymbol{z}_i$. Let $\boldsymbol{D}$ denote the diagonal covariance matrix of the principal components:

$$trace(\boldsymbol{D}) = trace(\boldsymbol{U'DU}) = trace(\boldsymbol{DU'U}) = trace(\boldsymbol{RI}) = trace(\boldsymbol{R}) = p \tag{2.6}$$

where we know that $trace(\boldsymbol{R})$ is simply the sum of the ones along the diagonal of the correlation matrix and p is the number of variables. Such property is useful for expressing a fraction of the total amount of variation accounted for by some subset of the principal components. Another useful by-product is a correlation matrix of the principal component scores $\boldsymbol{Z}$ with the original data $\boldsymbol{X}$. The correlation matrix is given by the following expression:

$$corr(\boldsymbol{X}, \boldsymbol{Z}) = \frac{1}{n-1}\boldsymbol{X}'\boldsymbol{Z}_s = \frac{1}{n-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{U}\boldsymbol{D}^{-\frac{1}{2}} = (\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}')\boldsymbol{U}\boldsymbol{D}^{-\frac{1}{2}} = \boldsymbol{U}\boldsymbol{D}^{\frac{1}{2}} = \boldsymbol{F} \tag{2.7}$$

where $\boldsymbol{F}$ is referred to component loadings, which are essential in interpretation of the PCA results.

So far I have assumed that my data is appropriate for using the PCA. For assessing such a manner, we could use the Bartlett's sphericity test. This test directly addresses the question if correlation matrix should be factored. The test is an approximate chi-squared test:

$$\chi^2[\frac{(p^2-p)}{2}] = -[(n-1) - \frac{(2p+5)}{6}]ln|\boldsymbol{R}| \tag{2.8}$$

where ln $|\boldsymbol{R}|$ is a natural logarithm of the correlation matrix, $\frac{(p^2-p)}{2}$ is the number of degrees of freedom associated with the chi-square test statistic, p is the number of variables and n is the number of observations. A null hypothesis of the test is that the true correlation matrix of the underlying population is an identity matrix, so if we are unable to reject the null hypothesis, we can conclude that dimension reduction is inappropriate.

The important question in the PCA is how many components we should choose. The Kaiser's rule advises to take all eigenvalues bigger than 1. Intuition behind the rule reflects the common sense notion that any principal component should account for at least as much variation as any of the original variables in $\boldsymbol{X}$. The second method is a graphical one called a scree plot. The approach involves plotting the eigenvalues for each principal component in order from the largest to the lowest ones. Then we look for an “elbow” in the curve, in which is a point after which the remaining eigenvalues decline in approximately linear fashion. However, I am using only two variables in the analysis, so the scree plot makes little sense in such case.

Despite all the benefits, the PCA method has one crucial assumption - it can only work on variables with ratio variables.[18] For example, the PCA cannot work on categorical data.[19] Conveniently, both my variables are ratio ones.

2.2.2 The extraction from the principal component analysis

In this chapter, I will present the extraction from the PCA. The variable *Index of property prices* is constructed from the variables *Flats prices* and *Houses prices*. The Bartlett's test χ^2 with one degree of freedom (since I am using only two variables) has the critical value of 6.63 (based on a 99 percent level of confidence, i.e., $\alpha = 0.01$). The resulted statistic is around one and half million, which is much higher than the critical one. According to the standard tables the p-value is practically zero, so I am rejecting the null hypothesis. Based on the test, the dimension reduction is justified.

Table 2.1: The eigenvalues and total variance explained - The Index of property prices

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	1.938	96.877	96.877
2	0.062	3.123	100.000

Source: Authors' calculations based on property prices from the CZSO data.

In the PCA the number of variables is equal to the sum of eigenvalues. In the Table 2.1 we can see that the first component is able to explain almost 97 percent of the variation in the original two variables. The application of the Kaiser's rule is very straight forward for the two variables - just include the first component.[20] So, I will extract only the first component. This decision is further supported by the evidence in the Table 2.2, where I present the component loadings. The component loadings are correlation between the original

[18] A ratio variable, has a clear definition of zero and the difference between a value of 100 and 90 is the same difference as between 90 and 80. Examples of ratio variables are height, weight or enzyme activity.

[19] A categorical variable, also called a nominal variable, is for mutual exclusive, but not ordered, categories. For example, you can code regions of some country.

[20] Except, the case of two truly orthogonal variables. But in such case I would not be able to reject the null hypothesis in the Bartlett's test.

variables and the components. Moreover, their squares indicate what percentage of the variance in the original variable is explained by the component. The first component is very easy to interpret and clearly contains the most important information from the two source variables. Furthermore, when we look at communalities, which depict a shared common factor, we can see quite large numbers (as for the component loadings, for the communalities hold that 1 is maximum and 0 minimum in absolute values). Taking all together, the resulted variable - the *Index of property prices* represent the original variables very well.

Table 2.2: Component matrix and communalities

Component Loadings	1	2	Communalities
Flats prices	0.984	-0.177	0.969
Houses prices	0.984	-0.177	0.969

Source: Authors' calculations based on property prices from the CZSO data.

The component is extracted by method of Bartlett scores. This method considers the PCA equation as a system of regression equations, where the original variables are the dependent ones, the factor loadings are the explanatory ones and the factor scores are the unknown parameters. The estimation of the scores is obtained by weighted least squares in order to account for a heteroscedasticity.

2.3 Descriptive statistics

In the Table 2.3 I present the descriptive statistics of the selected variables.

In terms of the mother's characteristics, we can see most mothers to have a Czech citizenship, their average age in the time of birth is almost 27 and most of the mothers have a first or a second child (more precisely, 85.4 percent). More than 75 percent of the mothers are married and more than a half have secondary education with graduation or higher (university degree).

When we look at the children's characteristic, approximately 51.5 percent are boys, which is close to the natural ratio of 105 boys to 100 girls (51.2 percent). Majority of the babies are alive in the time of birth and part of multiple births is three percent. The gestations age is close to 40 weeks with standart deviation of 2.2[21] and the birth weight is 3 308 grams on average.

[21] Natural time of gestation is between 38 to 42 weeks.

The *Index of property prices* has zero mean and the standard deviation of one. This is not a coincidence, but it is caused by the design of the PCA.

Father's education Practically, twenty percent of values concerning fathers' education are missing on the grounds of the reason that the entry for fathers' education is voluntary. I cannot state, if there is a self-selection in not reporting the fathers' education or if it is random. As a regard to this I provide the descriptive statistic depending upon the validity of the variable father's education. In the Table A.1 I present the descriptive statistics of the selected variables, where only the cases with the missing father's education are presented and in the Table A.2 I present the descriptive statistics of the selected variables with the valid cases of the father's education. For deciding, if I am facing a non-responsive bias in my sample, I use the t-test for two unknown means, which we can see in the Equation 2.9:

$$t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{2.9}$$

where x_1 and x_2 are means in the particular samples, s_1 and s_2 are standard deviations of the samples, n_1 and n_2 are the number of observations and t is the resulted statistic, which is presented in the Table 2.4. You can see that apart from one t-statistic (a sex of a child) all are statistically significant at 1% level (the null hypothesis indicate that means are not statistically different). So based on the analysis, we can see that the variable fathers' education contain the non-responsive bias and should be therefore use with caution or not at all.

I use the same procedure for an assessment of the air pollution sample. In the Table A.3 and the Table A.4 I present the descriptive statistics of the selected variables, where the cases are included based on an availability of air pollution record. For the statistical assessment of the differences I include the Table 2.5. We can see that, unfortunately, most of the values are significant, so the sample is mostly different from the original one and as a result cannot be as a general sample.[22]

[22]I also performed the tests with the original (general) sample and the sample with the valid air pollution record - the results are mostly similar and so the conclusion is the same.

Table 2.3: The descriptive statistics of the selected variables

Variables	N	Mean	Std. Dev.
Mother's characteristics			
Czech citizen*	1 401 851	0.982	0.133
Mother's age	1 401 851	26.735	5.067
Number of previous children	1 401 851	1.737	0.933
First or second child*	1 401 851	0.854	0.352
Marital status			
Single*	1 401 851	0.192	0.394
Married*	1 401 851	0.755	0.430
Divorced*	1 401 851	0.051	0.219
Widowed*	1 401 851	0.003	0.054
Education			
No education*	1 401 851	0.004	0.064
Basic*	1 401 851	0.123	0.329
Secondary*	1 401 851	0.358	0.479
Secondary with graduation*	1 401 851	0.393	0.488
University degree*	1 401 851	0.122	0.327
Child's characteristics			
Male*	1 401 851	0.514	0.500
Vitality* (1 = alive)	1 401 851	0.997	0.054
Multiple births	1 401 851	1.033	0.181
Multiple births - dummy*	1 401 851	0.032	0.176
Gestation age (in weeks)	1 401 851	39.291	2.234
Weight in grams	1 401 851	3 308.187	561.622
Characteristics of the surroundings			
Air pollution (PM10)	935 014	35.765	19.916
Flat prices for m^2	712 746	14 368.226	10 622.466
House prices for m^2	715 107	2 302.357	1 793.854
Index of property prices	708 524	0.000	1.000

Note: Variables with sign * are dummies with values ranging from 0 to 1. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Table 2.4: The resulted t-statistic from the t-test for two unknown means - the father's education

Variables	t	Variables	t
Czech citizen*	-23.44***	University degree*	-276.71***
Mother's age	-119.24***	Male*	-0.87
Number of previous chil.	-25.69***	Vitality* (1 = alive)	-12.42***
First or second child*	-148.21***	Multiple births	-23.71***
Single*	862.30***	Multiple births - dummy*	-22.99***
Married*	-2931.57***	Gestation age (in weeks)	-58.33***
Divorced*	258.91***	Weight in grams	-120.55***
Widowed*	58.91***	Air pollution (PM10)	20.10***
No education	64.73***	Flat prices for m^2	-76.69***
Basic*	263.04***	House prices for m^2	-45.54***
Secondary*	21.60***	Index of property prices	-61.83***
Secondary with grad.*	-194.37***		

Note: Variables with sign * is dummy with values ranging from 0 to 1. ***, **, * significance at the 1, 5 and 10% level. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Table 2.5: The resulted t-statistic from the t-test for two unknown means - air pollution

Variables	t	Variables	t
Czech citizen*	31.18***	Secondary with graduation*	17.27***
Mother's age	199.05***	University degree*	-15.87***
Number of previous children	1.15	Male*	1.07
First or second child*	-5.30***	Vitality* (1 = alive)	3.40***
Single*	136.75***	Multiple births	24.58
Married*	-143.97***	Multiple births - dummy*	25.15***
Divorced*	39.26***	Gestation age (in weeks)	-43.27***
Widowed*	0.44	Weight in grams	-2.24**
No education	41.91***	Flat prices for m^2	194.08***
Basic*	9.29***	House prices for m^2	165.81***
Secondary*	-74.32***	Index of property prices	188.08***

Note: Variables with sign * is dummy with values ranging from 0 to 1. ***, **, * significance at the 1, 5 and 10% level. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Chapter 3

Estimation methodology

3.1 Methodology for the Trivers-Willard hypothesis estimation

I will test the TW sex ratio hypothesis in two ways. The first, given that the sample is the general sample for the Czech Republic than the high-status are more likely to have a male offspring (according to the hypothesis). I will provide the sex ratios conditional on the citizenship, the number of previous children, the marital status and the education of mother and father. The second way is inspired by Almond and Edlund (2007) and the rest of the chapter is dedicated to the description of it.

I adopt a perspective, where the sex of offspring is endogenous to the socio-economic and the biological characteristics of the mother. As I already presented - one of the TWH mechanism is a sex-bias mortality which is affected by the conditions of the mother in the time of the pregnancy. I posses the micro-data, so I use the sex of the child as a dependent variable. This specification will allow me to investigate if the sex ratio of the children is influenced by the mother's characteristic (or more general by the parental characteristic). One of the advantages of my data set is that I can be sure that all mothers are the biological mothers.

Although, in the theoretical part I presented also the mechanism of the variation in the sex ratio at a conception, I do not posses a data to test such variation. So I will focus on how the mortality in utero shapes the sex ratio at birth. I do not have a data on the early fetal deaths, but I do have data on the stillbirths.[23] With these data I can study the extent to which the

[23]By the early fetal death I mean a spontaneous intrauterine death of a fetus at any time

mortality in the late fetal period (more than 28 weeks) may be considered a proximate mechanism for a TW effect. My hypothesis is that there will be a larger mortality among males born by the mothers in poor conditions and this would lead to a positive association between the maternal condition and a male sex among the live births.

My first regression specification look as following:

$$male_i = \alpha_0 + \alpha_1 \mathbf{X}_i + \alpha_2 \mathbf{H}_i + \gamma_t + \epsilon_i \tag{3.1}$$

where $male_i$ is a dummy variable which takes value 1 if child i is a male; α_0 is intercept; $\mathbf{X}_i$ is a matrix of the socio-economic (citizenship, marital status and education) and the biological (age) characteristics of the mother which will capture the condition of the mother; H_i is a matrix of variables capturing the child's condition (gestation age and birth weight); γ_t is a vector of dummy variables for particular years which accounts for the annual differences in the sex ratio; and ϵ_i is an error term. The $\mathbf{H}$ matrix is included as a control of the infant's health status at birth. Theoretically, this information is already included in the matrix $\mathbf{X}$, because the more educated people usually have healthier children (measured by the gestation age and the birth weight) or they may be less likely to deliver pre-term or low-birth-weight babies.[24] I include these variables as the controls, for example, boys are heavier in average than girls, but they also suffer from the different mortality risk at a given birth weight.

I estimate the Equation 3.1 on three samples: 1) all live births; 2) all stillbirths; and 3) all births. Under the TWH we can expect α_1 to be positive (for the positive characteristics - such as education). Also we can expect the married mothers to be more likely to have a son, but only under the assumption that the married mothers are in better condition than the unmarried mothers. Of course, in the second sample we can expect different signs for the positive characteristics.

Unfortunately, the statistical significance of the comparisons of the α_1 estimates from the Equation 3.1 from the different samples cannot be determine. From this reason I provide the second specification which can formally test

during a pregnancy, but less then approximately 28 weeks. The deaths late in the pregnancy (28 weeks or more) I denote as the stillbirths.

[24]Generally, the birth weight is the best proxy variable for a newborns' health (Almond and Currie, 2011). Although, this indicator is not perfect, it is still very strong predictor of human capital and labor market outcomes (Royer, 2009‘ Black et al., 2007).

the hypothesis that stillbirths are a part of the TWH mechanism, it looks as following:

$$male_i = \beta_0 + \beta_1\mathbf{X}_i + \beta_2\mathbf{H}_i + \gamma_t + \beta_3 d_i + \beta_4 d_i\mathbf{H}_i + \beta_5 d_i\mathbf{H}_i + \epsilon_i \qquad (3.2)$$

where $male_i$ is a dummy variable which takes value 1 if child i is a male; β_0 is intercept; $\mathbf{X}_i$ is a matrix of the socio-economic (marital status, education) and the biological (age) characteristics of the mother which will capture the condition of the mother; $\mathbf{H}_i$ is a matrix of variables capturing the child's condition (gestation age and birth weight); γ_t is a vector of the dummy variables for particular years which accounts for the annual differences in the sex ratio; d_i is a dummy variable which takes value 1 if child i died; and ϵ_i is an error term.

β_1 has now following interpretation - it is the effect of the matrix $\mathbf{X}_i$ on the sex of offspring among the live births. However, the most important coefficients are in the β_4 on the interaction terms $d_i\mathbf{H}_i$. The specification permit to the stillbirths to change the relationship between the matrix $\mathbf{H}_i$ and $male_i$. The vector of coefficients β_4 can statically compare $\hat{\alpha_1}$ and $\bar{\alpha_1}$. More formally, it is $\beta_4 = \hat{\alpha_1} - \bar{\alpha_1}$, where $\hat{\alpha_1}$ is the vector of coefficients obtained from the Equation 3.1 on the sample of the stillbirths, and $\bar{\alpha_1}$ is the vector of coefficients obtained from the Equation 3.1 on the sample of the live births.

Under the TWH we can expect to β_4 be negative in the case that the matrix $\mathbf{X}_i$ is positive and vice versa if it is negative. So under the TWH the children in good conditions would be less likely to be a stillbirth.

The β_4 should be positive, since males suffer higher infant mortality rates than females. Also α_0 and β_0 should be positive, because there is more males born than females.

3.2 The air pollution inclusion

As I already described - I will include air pollution to extend the TWH. The regressions specification look as following:

$$male_i = \alpha_0 + \alpha_1\mathbf{X}_i + \alpha_2\mathbf{H}_i + \alpha_3 k + \gamma_t + \epsilon_i \qquad (3.3)$$

$$male_i = \beta_0 + \beta_1\mathbf{X}_i + \beta_2\mathbf{H}_i + \beta_3 k + \gamma_t + \beta_4 d_i + \beta_5 d_i\mathbf{H}_i + \beta_6 d_i\mathbf{H}_i + \beta_7 k + \epsilon_i \qquad (3.4)$$

where k denotes air pollution and the rest is same as described in previous chapter.

Chapter 4

Results

4.1 Correlations between a sex of a child and the selected variables

In the first test of the TWH-SRH I will correlate on child's sex with the selected variables (the socio-economic and biological characteristics of a mother and the characteristic of the surroundings). I can use this procedure, since I possess the general sample. The resulted correlations are presented in Table 4.1. As you can see, 57 correlations are insignificant and only 3 are statistically significant on the 5% level. Using simple math we know that three out of 60 (the total number of the correlations) is exactly 5%, so it is not unlikely that we are facing an error of the first kind (occurs when the null hypothesis H_0 is true, but is rejected). Moreover, the significant correlations are in the smallest sample - stillbirths (N = 4 144). To conclude, the first method did not provide any evidence of a link between child's sex and the selected variables.

4.2 Regression results using a general sample from the Czech Republic

In Table 4.2 and Table 4.3 we can see the results from Equation 3.1 and Equation 3.2, respectively. The results are calculated by the method of Ordinary least squares.[25] They are not very supportive of the TWH. In terms of the TWH, we would expect that married, better educated or younger women would bear more sons.

[25]The results are not sensitive for a usage of another method.

Table 4.1: The correlations between a sex of a child and the selected variables

Variables	All	Live births	Stillbirths	First child
Mother's characteristics				
Czech citizen*	.000	.000	-.027	-.001
Mother's age	.000	.000	.014	.001
Marital status				
Single*	-.001	-.001	.009	-.001
Married*	.001	.001	-.001	.001
Divorced*	.000	.000	-.007	-.001
Widowed*	.000	.000	-.022	-.001
Education				
No education*	.000	.000	-.007	.000
Basic*	.000	.000	.021	.000
Secondary*	-.001	-.001	-.013	-.002
Secondary with graduation*	.000	.000	.002	.001
University degree*	.000	.000	-.020	.000
Char.s of the surroundings				
Air pollution (PM10)	.000	.000	.031	.000
Flat prices for m2	.000	.000	.058*	.002
House prices for m2	.000	.000	.047*	.001
Index of property prices	.000	.000	.056*	.002

Note: * Correlation is significant at the 0.05 level (two-tailed). Columns represent different samples. Variables with sign * are dummies with values ranging from 0 to 1. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

The only consistent with the TWH is the age's dummies. Younger mothers (17-21 years) bear more sons than the baseline mothers (22-28 years) and the older mothers (more than 29 years) are less likely to bear a son. The case of education is different - although, educated mothers are more likely to have a boy (compared to the mothers without an education - baseline), the strongest coefficient has mothers with a basic education. This particular result may seem strange, but there is probably a very simple explanation behind it. The less educated mothers (for example, with basic education) are also quite often the youngest. The data supports this claim - the values of a mother's education based on age are: 1.03 for the mothers under 17 years; 1.96 for the mothers aged 17 to 21; 2.52 for the mothers aged between 21 and 29 and 2.75 for the mothers over 29 years old.[26]

When we look at the coefficients of marital status, we can see that they are directly against the TWH. Although, the effect is not large, being married is associated with about 2.51% lower probability of having a boy, so the effect is driven by a different mechanism. It should be noted, that similarly to the case of education, single mothers are the youngest as well.

If stillbirths operate in a manner consistent with the TWH, we would expect that the coefficients in the sample "live births" would be reinforced (to the side consistent with the TWH, of course). But as we can see Table 4.2 does not confirm this hypothesis. Moreover, if we look at the third column - there is not even one significant coefficient (except constant). This evidence suggests that the stillbirths are random, or at least not affected by the variables I possess and in other words, high-status women have the same likelihood to experience a stillbirth.

In Table 4.3, I present the regression results from Equation 3.2, which serves as a tool for the statistical testing of the differences between the coefficients from the different samples in Table 4.2. In the first column we have virtually the same estimates as in the first column in Table 4.2, so our focus should be directed to the second column - $Variable * d_i$. The only significant coefficient is for citizenship, the interpretation is as follows: being a Czech citizen is associated with a reduced risk of stillbirths. However, the coefficient is significant only at the 10% level and together with the number of coefficients in the table, it is not unlikely that we are facing the type one error again.

[26]As a reminder, the coding for the variable "mother's education" is as follows: 1 = a basic education; 2 = a lower secondary education; 3 = a secondary education with a completed maturity exam; 4 = a first or a second stage of a tertiary education.

Table 4.2: Regression estimates of the Trivers-Willard hypothesis (high-status mothers bear more sons)

Variables	Whole sample	Live births	Stillbirths
Constant	1.064***	1.056***	0.816***
	(104.83)	(103.35)	(5.23)
Mother's characteristics			
Czech citizen*	-0.000626	0.00255	0.0151
	(-0.20)	(0.80)	(0.20)
Under 17 years*	0.00440	0.00468	-0.0385
	(1.09)	(1.16)	(-0.57)
17 - 21 years*	0.00928***	0.00942***	-0.0145
	(7.18)	(7.28)	(-0.61)
More than 29 years*	-0.00631***	-0.00646***	0.0242
	(-6.17)	(-6.30)	(1.34)
Marital status			
Married*	-0.0129***	-0.0129***	-0.0182
	(-10.95)	(-10.93)	(-0.86)
Divorced*	-0.00690***	-0.00677***	-0.0461
	(-3.18)	(-3.12)	(-1.27)
Widowed*	-0.00788	-0.00721	-0.161
	(-1.01)	(-0.92)	(-1.61)
Education			
Basic*	0.0994***	0.0987***	0.0624
	(14.62)	(14.49)	(0.55)
Secondary*	0.0824***	0.0816***	0.0332
	(12.25)	(12.11)	(0.30)
Secondary with graduation*	0.0791***	0.0783***	0.0455
	(11.78)	(11.64)	(0.41)
University degree*	0.0809***	0.0802***	0.0205
	(11.94)	(11.82)	(0.18)
N	1 401 851	1 397 707	4 144
F	1556.6	1564.2	2.371
R^2	0.0281	0.0283	0.0148

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. Columns represent different samples. Variables with sign * are dummies with values ranging from 0 to 1. All regressions contain a vector of dummy variables which controls for particular years, birth weight and gestation age.
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Table 4.3: Regression estimates of the Trivers-Willard hypothesis (high-status mothers bear more sons) - pattern reinforced by post-neonatal mortality

Variables	*Original Var.*	$Variable * d_i$
Constant	1.065***	
	(104.58)	
Stillbirth	-0.167	
	(-1.25)	
Mother's characteristics		
Czech citizen*	0.00255	-0.0388*
	(0.80)	(-1.93)
Under 17 years*	0.00466	-0.0396
	(1.16)	(-0.59)
17 - 21 years*	0.00941***	-0.0211
	(7.27)	(-0.90)
More than 29 years*	-0.00645***	0.0279
	(-6.30)	(1.58)
Marital status		
Married*	-0.0130***	-0.00137
	(-10.95)	(-0.07)
Divorced*	-0.00678***	-0.0322
	(-3.12)	(-0.90)
Widowed*	-0.00722	-0.160
	(-0.92)	(-1.62)
Education		
Basic*	0.0987***	0.00087
	(14.48)	(0.01)
Secondary*	0.0816***	-0.00931
	(12.10)	(-0.09)
Secondary with graduation*	0.0782***	0.00439
	(11.63)	(0.04)
University degree*	0.0801***	-0.0245
	(11.81)	(-0.22)
N	1 401 851	
F	1017.6	
R^2	0.0282	

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. Variables with sign * are dummies with values ranging from 0 to 1. All regressions contain a vector of dummy variables which controls for particular years, birth weight and gestation age.
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Interestingly, I have no support for the often observed pattern of the ”competing forces” (biological and socio-economical). Royer (2004) noted that the effect of a maternal age on an infant’s health comprises a tension between the biological and socio-economical conditions. Younger mothers tend to be healthier than older mothers, but their socio-economic status were worse (similar result obtained Almond and Edlund, 2007).

4.3 Regression results using the sample with air pollution and the real estate prices

In the following parts of this section I will present the results from Equation 3.3 and Equation 3.4. For the first equation, I provide three different specifications in Table 4.4, Table 4.5 and Table 4.6. The reason for providing three different specifications is to provide evidence of the inclusion sensitivity of air pollution and the Index of property prices. As we can see, the variable’s resulted coefficients and the t-statistic are very similar across the specifications.

The data does not support the hypothesis that the people who live in the less polluted areas have more sons. However, the Table 4.6 show little evidence that in the more polluted area, more girls are born, since the male stillbirth is positively associated with air pollution. But the statistical significance is low - only the 10% level and when we look into the Table 4.7, we can see that the result is not robust. Also, given the number of observations and the statistical significance of the Index of property prices, we should not draw any strong conclusion from the weak significance of the variable. With this number of observations we would expect much larger significance (at least at the 1% level).

For the sake of completeness, in Table 4.7 I present the regression results from Equation 3.2. The second column of the table can statistically compare the coefficients between the second and the third column in Table 4.4. Again, most of the results are insignificant and those which are, are not significant enough to draw any conclusion (given the number of observations).[27]

I provide a discussion of the results in the next chapter and I am particularly focused on the reason for the insignificant results.

[27]The results for the second and the third specification are also insignificant. I do not present them in the paper.

Table 4.4: Regression estimates of the Trivers-Willard hypothesis with inclusion of air pollution and property prices

Variables	All	Live births	Stillbirths
Constant	0.904***	0.900***	0.869***
	(70.99)	(70.45)	(4.29)
Mother's characteristics			
Czech citizen*	-0.00215	-0.00208	-0.0399
	(-0.44)	(-0.42)	(-0.39)
Under 17 years*	0.000488	0.000460	0.0255
	(0.07)	(0.06)	(0.22)
17 - 21 years*	0.00570**	0.00594**	-0.0510
	(2.16)	(2.24)	(-1.06)
More than 29 years*	-0.00568***	-0.00579***	0.0305
	(-3.92)	(-3.99)	(1.08)
Marital status			
Married*	-0.0127***	-0.0127***	-0.00933
	(-7.74)	(-7.75)	(-0.30)
Divorced*	-0.00635**	-0.00629**	-0.0411
	(-2.09)	(-2.07)	(-0.73)
Widowed*	-0.0373***	-0.0363***	-0.314
	(-2.98)	(-2.89)	(-1.52)
Education			
Basic*	0.0760***	0.0756***	0.0734
	(10.10)	(10.02)	(0.57)
Secondary*	0.0588***	0.0584***	0.0552
	(8.02)	(7.94)	(0.43)
Secondary with graduation*	0.0535***	0.0532***	0.0427
	(7.35)	(7.28)	(0.34)
University degree*	0.0577***	0.0573***	0.0896
	(7.84)	(7.77)	(0.69)
Char.s of the surroundings			
Air pollution	-0.0000407	-0.0000430	0.000428
	(-1.17)	(-1.24)	(0.63)
Index of property prices	-0.00125*	-0.00129*	0.0199
	(-1.74)	(-1.80)	(1.29)
N	582 202	580 649	1 553
F	723.0	725.4	1.658
R^2	0.0254	0.0256	0.0222

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. Columns represents different samples. Variables with sign * are dummies with values ranging from 0 to 1. All regressions contain a vector of dummy variables which controls for particular years, birth weight and gestation age. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Table 4.5: Regression estimates of the Trivers-Willard hypothesis with inclusion of property prices

Variables	All	Live births	Stillbirths
Constant	0.919***	0.916***	0.793***
	(79.59)	(79.07)	(4.37)
Mother's characteristics			
Czech citizen*	-0.00131	-0.00133	-0.00863
	(-0.28)	(-0.29)	(-0.09)
Under 17 years*	-0.000527	-0.000626	0.0455
	(-0.08)	(-0.09)	(0.42)
17 - 21 years*	0.00506**	0.00531**	-0.0584
	(2.12)	(2.22)	(-1.33)
More than 29 years*	-0.00633***	-0.00647***	0.0393
	(-4.80)	(-4.90)	(1.53)
Marital status			
Married*	-0.0121***	-0.0121***	-0.00380
	(-8.04)	(-8.05)	(-0.13)
Divorced*	-0.00562**	-0.00561**	-0.0170
	(-2.03)	(-2.02)	(-0.33)
Widowed*	-0.0266**	-0.0258**	-0.226
	(-2.36)	(-2.28)	(-1.27)
Education			
Basic*	0.0766***	0.0762***	0.0867
	(10.89)	(10.80)	(0.75)
Secondary*	0.0602***	0.0597***	0.0664
	(8.75)	(8.67)	(0.58)
Secondary with graduation*	0.0555***	0.0551***	0.0362
	(8.13)	(8.06)	(0.32)
University degree*	0.0595***	0.0591***	0.0715
	(8.62)	(8.55)	(0.61)
Char.s of the surroundings			
Index of property prices	-0.00110	-0.00116*	0.0219
	(-1.59)	(-1.67)	(1.48)
N	708 524	706 635	1 889
F	930.9	934.3	1.869
R^2	0.0256	0.0258	0.0196

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. Columns represents different samples. Variables with sign * are dummies with values ranging from 0 to 1. All regressions contain a vector of dummy variables which controls for particular years, birth weight and gestation age. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Table 4.6: Regression estimates of the Trivers-Willard hypothesis with inclusion of air pollution

Variables	All	Live births	Stillbirths
Constant	1.006***	1.011	0.975***
	(87.00)	(0.01)	(4.67)
Mother's characteristics			
Czech citizen*	-0.00396	-0.00383	0.0191
	(-0.95)	(-0.92)	(0.24)
Under 17 years*	-0.000559	-0.000421	-0.00436
	(-0.11)	(-0.08)	(-0.05)
17 - 21 years*	0.00720***	0.00726***	0.00643
	(4.11)	(4.13)	(0.20)
More than 29 years*	-0.00533***	-0.00545***	0.0205
	(-4.43)	(-4.53)	(0.93)
Marital status			
Married*	-0.0130***	-0.0130***	-0.0178
	(-9.46)	(-9.44)	(-0.70)
Divorced*	-0.00774***	-0.00764***	-0.0390
	(-3.06)	(-3.02)	(-0.89)
Widowed*	-0.0196**	-0.0197**	-0.0385
	(-2.05)	(-2.06)	(-0.29)
Education			
Basic*	0.0939***	0.0935***	0.0324
	(12.86)	(12.79)	(0.26)
Secondary*	0.0758***	0.0754***	0.0193
	(10.53)	(10.46)	(0.15)
Secondary with graduation*	0.0719***	0.0715***	0.0393
	(10.02)	(9.95)	(0.32)
University degree*	0.0733***	0.0729***	0.00385
	(10.11)	(10.05)	(0.03)
Char.s of the surroundings			
Air pollution	-0.0000320	-0.0000357	0.000971*
	(-1.13)	(-1.26)	(1.80)
N	935 014	932 355	2 659
F	1001	967.6	1.847
R^2	0.0271	0.0273	0.0179

Note: t-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. Columns represents different samples. Variables with sign * are dummies with values ranging from 0 to 1. All regressions contain a vector of dummy variables which controls for particular years, birth weight and gestation age.

Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Table 4.7: Regression estimates of the Trivers-Willard hypothesis with inclusion of air pollution and property prices - pattern reinforced by post-neonatal mortality

Variables	*Original Var.*	*Variable* $*d_i$
Constant	0.900***	
	(70.46)	
Stillbirth	-0.135	
	(-0.68)	
Mother's characteristics		
Czech citizen*	-0.00209	-0.0264
	(-0.43)	(-0.26)
Under 17 years*	0.000459	0.0323
	(0.06)	(0.28)
17 - 21 years*	0.00594**	-0.0602
	(2.24)	(-1.26)
More than 29 years*	-0.00580***	0.0355
	(-4.00)	(1.28)
Marital status		
Married*	-0.0127***	-0.000248
	(-7.74)	(-0.01)
Divorced*	-0.00628**	-0.0321
	(-2.07)	(-0.58)
Widowed*	-0.0363***	-0.295
	(-2.89)	(-1.44)
Education		
Basic*	0.0755***	0.0387
	(10.01)	(0.30)
Secondary*	0.0583***	0.0424
	(7.93)	(0.34)
Secondary with graduation*	0.0531***	0.0320
	(7.27)	(0.26)
University degree*	0.0572***	0.0715
	(7.76)	(0.56)
Char.s of the surroundings		
Air pollution	-0.0000428	0.000499
	(-1.23)	(0.77)
Index of property prices	-0.00131*	0.0255*
	(-1.82)	(1.84)
N	582 202	
F	412.3	
R^2	0.0255	

Note: T-statistics are in parentheses; ***, **, * significance at the 1, 5 and 10% level. Variables with sign * are dummies with values ranging from 0 to 1. All regressions contain a vector of dummy variables which controls for particular years, birth weight and gestation age.
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

4.4 The possible causes of the insignificant results

The findings in the lines above suggest that the TWH mechanism is not present in the Czech Republic. I identified three possible hypotheses which can explain the results:

Biological and socio-economical status of a father : Probably one of the most severe problems in my data set is the absence of information on the biological and socio-economical characteristics of a father. I possess only information about mothers and by using only those I assume that the biological and socio-economical conditions of parents (on aggregate) are correlated with the mother' ones. If this assumption does not hold, then the data is not sufficient for the proper verification or rejection of the TWH.

Insufficient diversity : I already discussed that the absolute wealth (conditions) are unlikely to be the driver behind the TWH mechanism. This said, the relative differences matter. However, imagine the case, where the relative differences are too small and the TWH mechanism cannot operate. In other words, to assess the size of the relative differences in the Czech Republic, we can use a proxy, for example, the Gini coefficient.[28] More precisely, I used the Gini coefficient after taxes and transfers for 31 European countries drawn from the Eurostat. The Czech Republic with a value of 24.9 is the fourth most egalitarian country in the year 2012 (Eurostat, 2012). In previous years the situation was similar.

Evolutionarily novel environment : This hypothesis is related to the previous one. It is possible that our culture's egalitarian values mask the TWH or the TWH is not culturally compatible with our society - as I already introduced the study of Guggenheim et al. (2007) who provided a comprehensive analysis of nationally representative samples from 35 countries and found only regional and country level evidence. Further research is needed to answer the question, why the TWH mechanism is active in some countries and inactive in others.

[28]Scale fo Gini coefficient: 0 is an absolute equity and 100 an inequity.

Conclusion

This paper investigates the effect of the biological and socio-economics status of mothers and the characteristics of surroundings (like air pollution) on the sex of children. Despite my best attempts, I was unsuccessful in my search for significant results.

I identified three hypotheses focused on the reason for the insignificant results. Unfortunately, they are all untestable with my current data set, so I can only mention them. The three possible problems are non-inclusion of the biological and socio-economical status of a father,[29] possible insufficient diversity or evolutionarily novel environment in the Czech Republic.

There are a lot of potential topics for future research. An obvious one would be to extend my data set - especially, for the father's biological and socio-economical characteristics. Secondly, it should be determined why the TWH is active in some areas, whilst inactive in others. Thirdly, the TWH research should be connected to the (income) inequality research. The hypothesis would be that the TWH mechanism is masked or does not work at all in more equal countries.

Generally, the most interesting question in the context of the TWH is, if parents ever maximized their number of offspring and if yes, under what conditions did they do it.

[29] Arguably, the biggest problem of the thesis.

Bibliography

Abernethy, Virginia, and Ray Yip. 1990. "Parent Characteristics and Sex Differential Infant Mortality: The Case in Tennessee." *Human Biology*, 279-90.

Almond, Douglas, and Janet Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *The Journal of Economic Perspectives* 25 (3): 153-72.

Almond, D., and L. Edlund. 2007. "Trivers-Willard at Birth and One Year: Evidence from US Natality Data 1983-2001." *Proceedings of the Royal Society B: Biological Sciences* 274 (1624): 2491-2496.

Anderson, Judith L., and Charles B. Crawford. 1993. "Trivers-Willard Rules for Sex Allocation: when do they maximize expected grandchildren in humans?" *Human Nature* 4 (2): 137-74.

Arcand, Jean-Louis, and Matthias Rieger. 2012. "Maternal Height and the Sex Ratio." *Journal of Evolutionary Psychology* 10 (2): 69-81.

Baron-Cohen, Simon. 1999. "The Extreme Male Brain Theory of Autism." In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders.* Cambridge, MA: The MIT Press.

——. 2002. "The Extreme Male Brain Theory of Autism." *Trends in Cognitive Sciences* 6 (6): 248-54.

——. 2003. *Essential Difference: Male and Female Brains and the Truth about Autism.* Penguin, London.

Baron-Cohen, Simon, and Jessica Hammer. 1997. "Is Autism an Extreme Form of the 'Male Brain'?" *Advances in Infancy Research* 11: 193-218.

Baron-Cohen, S., Lutchmaya, S., Knickmeyer, R., 2004. *Prenatal Testosterone in Mind: Amniotic Fluid Studies.* MIT Press.

Betzig, L. L. 1986. *Despotism and Differential Reproduction: A Darwinian View of History.* New York : Aldine Publishing Co.

Betzig, Laura L., Monique Borgerhoff Mulder, and Paul Turke. 1988. *Human Reproductive Behaviour: A Darwinian Perspective.* CUP Archive.

Betzig, Laura, and Samantha Weber. 1995. "Presidents Preferred Sons." *Politics and the Life Sciences*, 61-64.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2007. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes." *The Quarterly Journal of Economics* 122 (1): 409-439.

Booth, Alan, and D. Wayne Osgood. 1993. "The Influence of Testosterone on Deviance in Adulthood: Assessing and Explaining the Relationship." *Criminology* 31 (1): 93-117.

Burley, Nancy. 1986. "Sex-Ratio Manipulation in Color-Banded Populations of Zebra Finches." *Evolution*, 1191-1206.

Buss, D. M. 1989. "Sex Differences in Human Mate Preferences: Evolutionary Hypotheses Tested in 37 Cultures." *Behavioral and Brain Sciences* 12 (1): 1-49.

Cameron, Elissa Z., Wayne L. Linklater, Kevin J. Stafford, and Clare J. Veltman. 1999. "Birth Sex Ratios Relate to Mare Condition at Conception in Kaimanawa Horses." *Behavioral Ecology* 10 (5): 472-75.

Cameron, Elissa Z. 2004. "Facultative Adjustment of Mammalian Sex Ratios in Support of the Trivers-Willard Hypothesis: Evidence for a Mechanism." *Proceedings of the Royal Society of London, Series B: Biological Sciences* 271 (1549): 1723-28.

Cameron, Elissa Z., and Fredrik Dalerum. 2009. "A Trivers-Willard Effect in Contemporary Humans: Male-Biased Sex Ratios among Billionaires." *PLoS One* 4 (1): e4195.

Carriazo-Osorio, Fernando. 2001. "Impacts of Air Pollution on Property Values: An Economic Valuation for Bogota, Colombia." Paper presented at the 2nd Workshop on Population, Economy and the Environment: Modeling and Simulating Their Complex Interaction at the Max Plant Institute for Demographic Research, Germany.

Catalano, Ralph A. 2003. "*Sex Ratios in the Two Germanies: A Test of the Economic Stress Hypothesis.*" *Human Reproduction* 18, (9): 1972-75.

Central Intelligence Agency. *The World Factbook.*

Chacon-Puignau, Grace Cristina, and Klaus Jaffe. 1996. "Sex Ratio at Birth Deviations in Modern Venezuela: The Trivers-Willard Effect." *Biodemography and Social Biology* 43 (3-4): 257-70.

Clutton-Brock, Tim H. 1991. *The Evolution of Parental Care.* Princeton University Press.

Clutton-Brock, Tim H., S. D. Albon, and F. E. Guinness. 1982. "Parental Investment in Male and Female Offspring in Polygynous Mammals." *Nature* 289 (5797): 487-89.

——. 1984. "Maternal Dominance, Breeding Success and Birth Sex Ratios in Red Deer." *Nature* 308 (5957): 358-60.

——. 1986. "Great Expectations: Dominance, Breeding Success and Offspring Sex Ratios in Red Deer." *Animal Behaviour* 34 (2): 460-71.

Clutton-Brock, Tim H., and G. R. Iason. 1986. "Sex Ratio Variation in Mammals." *Quarterly Review of Biology*, 339-74.

Crawford, Charles. 1998. "The Theory of Evolution in the Study of Human Behavior: An Introduction and Overview." *Handbook of Evolutionary Psychology: Ideas, Issues, and Applications*, 3-43.

Cronk, Lee. 1989. "Low Socioeconomic Status and Female-Biased Parental Investment: The Mukogodo Example." *American Anthropologist* 91 (2): 414-29.

Dabbs, James M., and Robin Morris. 1990. "Testosterone, Social Class, and Antisocial Behavior in a Sample of 4,462 Men." *Psychological Science* 1 (3): 209-11.

Darwin, C. 1874. *The Descent of Man.* 2nd ed. New York: Random House.

Denny, K. 2008. "Big and tall parents do not have more sons." *Journal of Theoretical Biology* 250 (4): 752-753.

Enright, W. J., L. J. Spicer, M. Kelly, N. Culleton, and D. J. Prendiville. 2001. "Energy Level in Winter Diets of Fallow Deer: Effect on Plasma Levels of Insulin-like Growth Factor-I and Sex Ratio of Their Offspring." *Small Ruminant Research* 39 (3): 253-59.

Essock-Vitale, Susan M. 1984. "The Reproductive Success of Wealthy Americans." *Ethology and Sociobiology* 5 (1): 45-49.

Eurostat, *Gini coefficient.* 2012.

Festa-Bianchet, Marco. 1996. "Offspring Sex Ratio Studies of Mammals: Does Publication Depend upon the Quality of the Research or the Direction of the Results." *Ecoscience* 3 (1): 42-44.

Folmer, Vanderlei, Júlio C. M. Soares, Diogo Gabriel, and Joao B. T. Rocha. 2003."A High Fat Diet Inhibits δ-Aminolevulinate Dehydratase and Increases Lipid Peroxidation in Mice (*Mus Musculus*)." *The Journal of Nutrition* 133, (7): 2165-70.

Gelman, Andrew. 2007. "Letter to the Editors Regarding Some Papers of Dr. Satoshi Kanazawa." *Journal of Theoretical Biology* 245 (3): 597-99.

Guggenheim, Cordelia B., Melinda F. Davis, and Aurelio José Figueredo. 2007. "Sons or Daughters: A Cross-Cultural Study of Sex Ratio Biasing and Differential Parental Investment." *Journal of the Arizona-Nevada Academy of Science* 39: 73-90.

Gibson, Mhairi A., and Ruth Mace. 2003. "Strong Mothers Bear More Sons in Rural Ethiopia." *Proceedings of the Royal Society of London. Series B: Biological Sciences*: 270 (1): S108-S109.

Harris, Julie Aitken, Philip A. Vernon, and Dorret I. Boomsma. 1998. "The Heritability of Testosterone: A Study of Dutch Adolescent Twins and Their Parents." *Behavior Genetics* 28 (3): 165-71.

Harrison, David, and Daniel L. Rubinfeld. 1978. "The Air Pollution and Property Value Debate: Some Empirical Evidence." *The Review of Economics and Statistics* 60 (4): 635-38.

Hill, Jack. 1984. "Prestige and Reproductive Success in Man." *Ethology and Sociobiology* 5 (2): 77-95.

Hopcroft, Rosemary L. 2005. "Parental Status and Differential Investment in Sons and Daughters: Trivers-Willard Revisited." *Social Forces* 83 (3): 1111-36.

Irons, William. 1979. "Cultural and Biological Success" Pp. 257-72 in *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective*, Duxbury Press.

James, William H. 1987. "The Human Sex Ratio. Part 1: A Review of the Literature." *Human Biology* 59: 721-52.

Jha, Prabhat, Rajesh Kumar, Priya Vasa, Neeraj Dhingra, Deva Thiruchelvam, and Rahim Moineddin. 2006. "Low Male-to-Female Sex Ratio of Children Born in India: National Survey of 1.1 Million Households." *The Lancet* 367 (9506): 211-18.

Kanazawa, Satoshi, 2003. "Can Evolutionary Psychology Explain Reproductive Behavior in the Contemporary United States?" *The Sociological Quarterly* 44 (2): 291-302.

——. 2005. "Big and Tall Parents Have More Sons: Further Generalizations of the Trivers-Willard Hypothesis." *Journal of Theoretical Biology* 235 (4): 583-590.

——. 2006. "Violent Men Have More Sons: Further Evidence for the Generalized Trivers-Willard Hypothesis (gTWH)." *Journal of Theoretical Biology* 239 (4): 450-59.

——. 2007a. "Beautiful Parents Have More Daughters: A Further Implication of the Generalized Trivers-Willard Hypothesis (gTWH)." *Journal of Theoretical Biology* 244 (1): 133-40.

——. 2007b. "Big and Tall Soldiers Are More Likely to Survive Battle: A Possible Explanation for the 'returning Soldier Effect' on the Secondary Sex Ratio." *Human Reproduction* 22 (11): 3002-8.

Kanazawa, Satoshi, and Griet Vandermassen. 2005. "Engineers Have More Sons, Nurses Have More Daughters: An Evolutionary Psychological Extension of Baron-Cohen's Extreme Male Brain Theory of Autism." *Journal of Theoretical Biology* 233 (4): 589-99.

Kanazawa, Satoshi, and Péter Apari. 2009. "Sociosexually Unrestricted Parents Have More Sons: A Further Application of the Generalized Trivers-Willard Hypothesis (gTWH)." *Annals of Human Biology* 36 (3): 320-30.

Keller, Matthew C., Randolph M. Nesse, and Sandra Hofferth. 2001. "The Trivers-Willard Hypothesis of Parental Investment: No Effect in the Contemporary United States." *Evolution and Human Behavior* 22 (5): 343-60.

Koziel, Slawomir, and Stanley J. Ulijaszek. 2001. "Waiting for Trivers and Willard: Do the Rich Really Favor Sons?" *American Journal of Physical*

Anthropology 115 (1): 71-79.

Lazarus, John. 2002. “Human Sex Ratios: Adaptations and Mechanisms, Problems and Prospects.” *Sex Ratios: Concepts and Research Methods*, 287-311.

Mace, Ruth. 1996. “Biased Parental Investment and Reproductive Success in Gabbra Pastoralists.” *Behavioral Ecology and Sociobiology* 38 (2): 75-81.

Nettle, Daniel. 2002a. “Height and Reproductive Success in a Cohort of British Men.” *Human Nature* 13 (4): 473-91.

——. 2002b. “Women’s Height, Reproductive Success and the Evolution of Sexual Dimorphism in Modern Humans.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269 (1503): 1919-23.

Nielsen, Francois. 1994. “Sociobiology and Sociology.” *Annual Review of Sociology* 20, 267-303.

Norberg, Karen. 2004. “Partnership Status and the Human Sex Ratio at Birth.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271 (1555): 2403-10.

Oreopoulos, Philip, Marianne E. Page, and Ann Huff Stevens. 2006. “The Intergenerational Effects of Compulsory Schooling.” *Journal of Labor Economics* 24 (4): 729-760.

Pawlowski, Boguslaw, R. I. M. Dunbar, and A. Lipowicz. 2000. “Evolutionary Fitness: Tall Men Have More Reproductive Success.” *Nature* 403 (6766): 156-156.

Pérusse, Daniel. 1993. “Cultural and Reproductive Success in Industrial Societies: Testing the Relationship at the Proximate and Ultimate Levels.” *Behavioral and Brain Sciences* 16 (02): 267-322.

Polinsky, A. M., and S. Shavell. 1975. “The Air Pollution and Property Value Debate.” *The Review of Economics and Statistics* 57 (1): 100-104.

Pollet, Thomas V., and Daniel Nettle. 2010. “No Evidence for the Generalized Trivers-Willard Hypothesis from British and Rural Guatemalan Data.” *Journal of Evolutionary Psychology* 8 (1): 57-74.

Rivers, J. P. W., and M. A. Crawford. 1974. “Maternal Nutrition and the Sex Ratio at Birth.” *Nature* 252: 297-98.

Royer, Heather. 2004. *What All Women (and Some Men) Want to Know: Does Maternal Age Affect Infant Health?* Center for Labor Economics, University of California.

——. 2009. "Separated at Girth: US Twin Estimates of the Effects of Birth Weight." *America Economic Journal: Applied Economics*: 49-85.

Rosenfeld, Cheryl S., and R. Michael Roberts. 2004. "Maternal Diet and Other Factors Affecting Offspring Sex Ratio: A Review." *Biology of Reproduction* 71, (4): 1063-70.

Schnettler, Sebastian. 2013. "Revisiting a Sample of US Billionaires: How Sample Selection and Timing of Maternal Condition Influence Findings on the Trivers-Willard Effect." *PloS One* 8 (2): e57446.

Sewell, William H., and Vimal P. Shah. 1968. "Parents' Education and Children's Educational Aspirations and Achievements." *American Sociological Review*: 191-209.

Sheldon, Ben C., and Stuart A. West. 2004. "Maternal Dominance, Maternal Condition, and Offspring Sex Ratio in Ungulate Mammals." *The American Naturalist* 163 (1): 40-54.

Silventoinen, Karri, Jaakko Kaprio, Eero Lahelma, Richard J. Viken, and Richard J. Rose. 2001. "Sex Differences in Genetic and Environmental Factors Contributing to Body-Height." *Twin Research* 4 (01): 25-29.

Smith, V. Kerry, and Ju-Chin Huang. 1995. "Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models." *Journal of Political Economy* 103 (1): 209-27.

Soler, Hosanna, Preeti Vinayak, and David Quadagno. 2000. "Biosocial Aspects of Domestic Violence." *Psychoneuroendocrinology* 25 (7): 721-39.

Symington, M. McFarland. 1987. "Sex Ratio and Maternal Rank in Wild Spider Monkeys: When Daughters Disperse." *Behavioral Ecology and Sociobiology* 20 (6): 421-25.

Trivers, R. L. 1972. *Parental Investment and Sexual Selection.* In Sexual Selection and the Descent of Man, 1871-1971, ed. B. G. Campbell, 136-179. 1st ed. Chicago: Aldine.

Trivers, R. L., and D. E. Willard. 1973. "Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring." *Science* 179 (1): 90-92.

Whiting, John W. M. 1993. "The Effect of Polygyny on Sex Ratio at Birth." *American Anthropologist* 95 (2): 435-442.

World Bank. 2013. *World Development Indicators 2013*. Washington, DC: World Bank.

Wright, Robert. 1994. *The Moral Animal*. New York: Pantheon.

Yamaguchi, Kazuo, and Linda R. Ferguson. 1995. "The Stopping and Spacing of Childbirths and Their Birth-History Predictors: Rational-Choice Theory and Event-History Analysis." *American Sociological Review* 6): 272-98.

Zaldívar, Maria E., Roberto Lizarralde, and Stephen Beckerman. 1991. "Unbiased Sex Ratios among the Bari: An Evolutionary Interpretation." *Human Ecology* 19 (4): 469-98.

Appendix

Table A.1: The descriptive statistics of the selected variables, Included cases only if the father's education is missing

Variables	N	Mean	Std. Dev.
Mother's characteristics			
Czech citizen*	281 755	.98	.15
Mother's age	281 755	25.61	5.78
Number of previous children	281 755	1.69	1.09
First or second child*	281 755	.38	.49
Marital status			
Single*	281 755	.76	.43
Married*	281 755	.01	.12
Divorced*	281 755	.21	.41
Widowed*	281 755	.01	.11
Education			
No education*	281 755	.02	.13
Basic*	281 755	.31	.46
Secondary*	281 755	.38	.48
Secondary with graduation*	281 755	.25	.43
University degree*	281 755	.67	.47
Child's characteristics			
Male*	281 755	.51	.50
Vitality* (1 = alive)	281 755	1.00	.07
Multiple births	281 755	1.03	.16
Multiple births - dummy*	281 755	.03	.16
Gestation age (in weeks)	281 755	39.01	3.00
Weight in grams	281 755	3187.92	603.44
Characteristics of the surroundings			
Air pollution (PM10)	207 890	36.55	20.29
Flat prices for m^2	164 445	12660.48	10170.77
House prices for m^2	164 532	2130.20	1728.28
Index of property prices	163 119	-.13	.96

Note: Variables with sign * are dummies with values ranging from 0 to 1. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Table A.2: The descriptive statistics of the selected variables, Included cases only if the father's education is recorded

Variables	N	Mean	Std. Dev.
Mother's characteristics			
Czech citizen*	1 120 096	0.98	0.13
Mother's age	1 120 096	27.02	4.83
Number of previous children	1 120 096	1.75	0.89
First or second child*	1 120 096	0.53	0.50
Marital status			
Single*	1 120 096	0.05	0.21
Married*	1 120 096	0.94	0.24
Divorced*	1 120 096	0.01	0.10
Widowed*	1 120 096	0.00	0.02
Education			
No education*	1 120 096	0.00	0.03
Basic*	1 120 096	0.08	0.26
Secondary*	1 120 096	0.35	0.48
Secondary with graduation*	1 120 096	0.43	0.50
University degree*	1 120 096	0.92	0.27
Child's characteristics			
Male*	1 120 096	0.51	0.50
Vitality* (1 = alive)	1 120 096	1.00	0.05
Multiple births	1 120 096	1.03	0.19
Multiple births - dummy*	1 120 096	0.03	0.18
Gestation age (in weeks)	1 120 096	39.36	1.99
Weight in grams	1 120 096	3338.44	546.45
Characteristics of the surroundings			
Air pollution (PM10)	727 124	35.54	19.80
Flat prices for m^2	548 301	14880.41	10701.25
House prices for m^2	550 575	2353.80	1809.81
Index of property prices	545 405	0.04	1.01

Note: Variables with sign * are dummies with values ranging from 0 to 1. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).

Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Table A.3: The descriptive statistics of the selected variables, Included cases only if air pollution is recorded

Variables	N	Mean	Std. Dev.
Mother's characteristics			
Czech citizen*	935 014	0.98	0.12
Mother's age	935 014	27.33	5.00
Number of previous children	935 014	1.74	0.94
First or second child*	935 014	0.50	0.50
Marital status			
Single*	935 014	0.22	0.42
Married*	935 014	0.72	0.45
Divorced*	935 014	0.06	0.23
Widowed*	935 014	0.00	0.05
Education			
No education	935 014	0.01	0.07
Basic*	935 014	0.13	0.33
Secondary*	935 014	0.34	0.47
Secondary with graduation*	935 014	0.40	0.49
University degree*	935 014	0.87	0.34
Child's characteristics			
Male*	935 014	0.51	0.50
Vitality* (1 = alive)	935 014	1.00	0.05
Multiple births	935 014	1.04	0.19
Multiple births - dummy*	935 014	0.03	0.18
Gestation age (in weeks)	935 014	39.24	2.34
Weight in grams	935 014	3307.45	568.50
Characteristics of the surroundings			
Air pollution (PM10)	935 014	35.77	19.92
House prices for m^2	585 574	15170.54	11171.19
Flat prices for m^2	585 032	2427.25	1877.77
Index of property prices	582 202	0.07	1.05

Note: Variables with sign * are dummies with values ranging from 0 to 1. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.

Table A.4: The descriptive statistics of the selected variables, Included cases only if air pollution is missing

Variables	N	Mean	Std. Dev.
Mother's characteristics			
Czech citizen*	466 837	0.98	0.15
Mother's age	466 837	25.55	4.99
Number of previous children	466 837	1.74	0.92
First or second child*	466 837	0.51	0.50
Marital status			
Single*	466 837	0.13	0.34
Married*	466 837	0.82	0.38
Divorced*	466 837	0.04	0.20
Widowed*	466 837	0.00	0.05
Education			
No education	466 837	0.00	0.04
Basic*	466 837	0.12	0.32
Secondary*	466 837	0.40	0.49
Secondary with graduation*	466 837	0.38	0.49
University degree*	466 837	0.88	0.33
Child's characteristics			
Male*	466 837	0.51	0.50
Vitality* (1 = alive)	466 837	1.00	0.06
Multiple births	466 837	1.03	0.17
Multiple births - dummy*	466 837	0.03	0.16
Gestation age (in weeks)	466 837	39.40	1.99
Weight in grams	466 837	3309.67	547.59
Characteristics of the surroundings			
Air pollution (PM10)	0	-	-
Flat prices for m^2	127 172	10673.91	6415.68
House prices for m^2	130 075	1740.64	1202.73
Index of property prices	126 322	-0.34	0.62

Note: Variables with sign * are dummies with values ranging from 0 to 1. The Index of property prices is extracted with the principal component analysis and it represents the variables: the Flat prices for m^2 and the House prices for m^2 (all in CZK).
Source: Authors' calculations based on the natality data set from the CZSO, the property prices also from the CZSO and the air pollution data set from the CHMI.