EXPLORING NATURAL EXPERIMENTS: GENDER OF OFFSPRING AND THE CHALLENGES OF THE STOPPING RULES[1]

Giambattista Salinari, Gianni Carboni, Virginia Zarulli

Abstract. The sex of children has frequently been used in the field of social sciences to conduct natural experiments. The key hypothesis behind this methodology is that the sex of the firstborn (or the first k births) is an exogenous variable, meaning it is not influenced by family characteristics observed before the birth of the children. Recent analyses have questioned the supposed exogeneity of the sex of children, proposing that stress experienced during pregnancy results in higher male embryo mortality, thereby leading to a higher probability of female births. This hypothesis casts doubt on the results reached by studies that have used the sex of children to conduct natural experiments. However, the analyses supporting this hypothesis have not properly considered the problems arising from stopping rules, specifically the tendency of some families to continue having children until a child of the desired sex is born. In this work, we show, using an indirect approach, that if stopping rules are properly taken into account, the sex of offspring is not associated to parental stress.

1. Introduction

Randomized experiments are generally considered the most reliable method for estimating a causal effect of interest (Imbens and Rubin 2015). The logic behind these experiments is very simple: a treatment is randomly administered to individuals in a study group. Individuals randomly selected to receive the treatment will form the treatment group. The remaining individuals will constitute the control group. The random assignment of the treatment ensures that the control group and the treatment group share, on average, the same pre-treatment characteristics, except for sampling errors. Thus, any observed differences in the outcome of interest between the control and the treatment group will reflect only the effect of the treatment itself, and not any pre-existing differences between the two groups.

In the field of social sciences, the possibility of conducting randomized experiments often clashes with ethical limits. For example, it is not possible to randomize

[1] Giambattista Salinari produced the research questions and the study design, conducted the analysis, provided methodological expertise for the study and drafted the first version of the manuscript. Gianni Carboni and Virginia Zarulli contributed significantly through providing overarching critical feedback, interpreting the results with insightful conceptual enhancements, editing and revising the final manuscript.

individuals based on different levels of education to then verify the effect of this variable, say, on fertility, because, obviously, one cannot impose on an individual the level of education they must have.

To overcome these difficulties, the social sciences have developed a set of techniques known as *natural experiments* (Dunning 2012). The basic idea is to use a "natural randomizer" to replace the random assignment of treatment used in randomized experiments. These "natural randomizers" allow us to study situations that would otherwise be hard to study because of the interconnection of two or more variables and the impossibility to detangle the causal effect. Today, numerous instances of this approach exist, with random variability often sourced from actual lotteries, as in the classic example of Angrist's (1990) study on the income of Vietnam War veterans and studies on migrations from Tonga (McKenzie et al. 2010) are such examples. The sex of children has also been considered as a random source of variability, for example in the study of the effect of fertility on female labor supply (Angrist and Evans 1998), the study of divorce and separation (Dahl and Moretti 2008) and the study of the effect of fertility on cognitive decline at old age (Bonsang and Skirbekk 2022).

Hamoudi and Nobles (2014), however, argued that the sex of children cannot be considered a random source of variability to be employed in natural experiments. To bolster their argument, they used biological considerations, claiming that stress during pregnancy increases male embryo mortality compared to females, thus suggesting that a lower probability of male births may occur in conflict-ridden families. To validate this hypothesis, these authors conducted analyses showing that the sex of the children was associated with the stress levels measured before birth.

With this work, we show that Hamoudi and Nobles (2014) might have underestimated the role played by the so-called stopping rules: those strategies by which parents continue to have children until the desired sex, or the desired combination of sexes is reached. Herein we demonstrate that in the presence of such strategies, the analysis proposed by Hamoudi and Nobles (2014) suffers from an insidious and little-known form of bias called *collider bias*.

2. Problems stemming from stopping rules

It is known that parents may have a preference regarding the sex of their children, like desiring boys or a mix of genders. When these preferences exist, parents may use stopping rules that consist of continuing to have children until they get the desired outcome. Such rules have been identified in many societies. Research by Angrist and Evans (1998) and Dahl and Moretti (2008) confirm this behaviour in the US population. For instance, Angrist and Evans, discovered that only 37.2% of

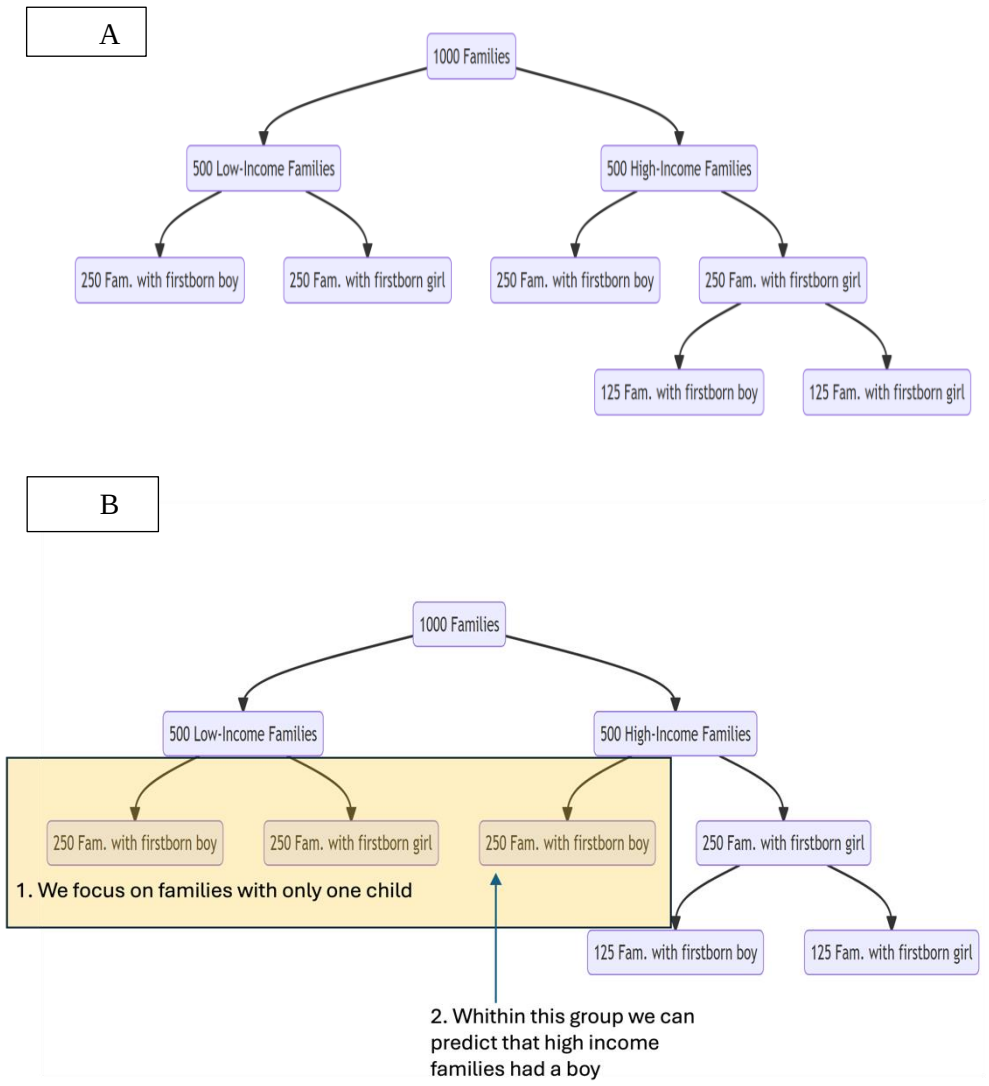
women with two children of different sexes go on to have a third child, compared to 43.2% of women who have two children of the same sex. In Denmark, it was found that families with two same-sex children had the highest fertility rates, those with two children of the same sex followed by a child of the opposite sex had the lowest rates, and that families with two boys showed a slight preference for girls, as they had higher fertility rates compared to families with two girls (Jacobsen et al. 1999).

One way to clarify the nature of the problems introduced by stopping rules is to resort to an example where the role of this type of rules is "extremized." Let's assume that the families of a given population conform to the following rule: a) all families have at least one child; b) all families with high income (or belonging to a certain ethnic group, or with low marital conflict, etc.) have an additional child if the firstborn is female; c) the probability of having a male child is 50%. Notice that nowhere in this rule it is stated that family income has a causal effect on the sex of children. We are thus assuming that the sex of a newborn and the family income are independent. Figure 1.A shows an example of this type of stopping rule, where it is assumed that there are 1000 families, divided into 500 high-income families and 500 low-income families. This example demonstrates that the stopping rule results in high-income families having a higher average number of children compared to low-income families. However, the sex ratio remains unaffected by the stopping rule. Among high-income families, there are 750 children, with an equal distribution of 350 females and 350 males, mirroring the distribution in low-income families, which have 250 males and 250 females. The fact that stopping rules do not alter the sex ratio has been mathematically proved by various articles (see Grech et al. 2018 for a summary). So, in general, we cannot predict the sex of a child based on the income of the family to which the child belongs. And this is consistent with the example assumption, according to which family income does not cause the sex of newborns.

Things change, however, if we condition (focus) the analysis on family size. This means that if we consider only those families with, say, only one child, then we can predict the sex of a newborn based on the information about family income.

To illustrate this point, let's consider families with only one child (see Figure 1.B). From this group, we select a family known to have a high income. We can infer that this family has a boy because, if they had a girl, they would have likely continued to have more children in pursuit of a male, given their high income. Therefore, conditioning on the number of children creates the impression that family income can predict whether a male child is present, even though there is no causal link between income and the sex of the children by design.

Figure 1 - *A hypothetical example of a stopping rule.*



Note: In this example, we suppose we observe 1000 families divided into 500 low-income families and 500 high-income families. These families conform to the following rules: a) all families have at least one child; b) all high-income families have an additional child if the firstborn is female; c) the probability of having a male child is 50%.

The fact that it seems possible to predict the sex of a child based on the family's income in the presence of stopping rules is a statistical artifact known as collider bias. In general, given three variables A, B, and C, B is said to be a collider (or collision node) when both A and C have a causal effect on B, i.e.:

$$A \rightarrow B \leftarrow C$$

In this simple causal graph,[2] it is possible to prove through a criterion called *d-separation* (Pearl 2009) that A and C are marginally independent. At the population level, the information about the value of A does not allow predicting the value of C. And this is correct, because it is evident that there is no causal effect of A on C or vice versa. With the d-separation criterion, it is also possible to show that, if the analysis is conditioned on B (for example, by selecting a subpopulation characterized by a specific value of B), within this subpopulation, A and C will be associated. Within the subgroups defined by the values of B, it is thus possible to predict the value of C from the knowledge of the value of A, even though A is not causing C or vice versa. A and C are thus linked by a spurious association (correlation), not produced by the existence of a causal connection.

In our example of Figure 1, the reason why it seems possible to predict the sex of a child based on family income, in the presence of stopping rules, is closely related to what we have just discussed. Let's denote B as the family size. In our example, B is determined (caused) by the sex of the first child (A) and the family's income level (C). When we condition the analysis on family size, a spurious association between the sex of the first child and family income is triggered.

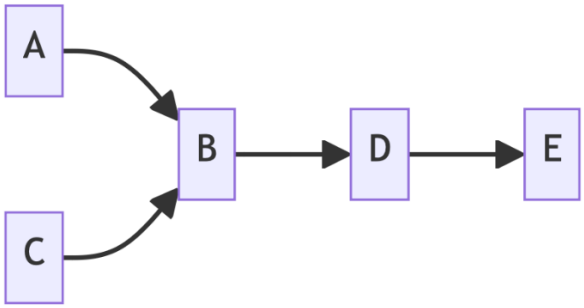
In reality, the conditions leading to the formation of collider bias are more complex than what discussed so far. Here it will suffice to notice that, in general, collider bias will occur whenever the analysis is conditioned on a collider or on a descendant of a collider, a variable that is directly or indirectly caused (through the mediation of other variables) by a collider. Referring to the causal diagram in Figure 2, we can say that a spurious association (bias) between A and C will emerge if the analysis is conditioned on B, or on D, or on E.

In our example, if the analysis is conditioned on any variable that is directly or indirectly influenced by family size, this will produce a spurious correlation between family income level and the sex of the newborn.

The existence of stopping rules can thus complicate the identification of the determinants of children's sex.

[2] In the context of causal graph theory, an arrow (→) indicates a direct causal effect, nodes represent variables, direction shows causality flow, and acyclicity ensures there are no loops.

Figure 2 – *Collider bias.*



3. The association between marital stress and the sex of children

Hamoudi and Nobles (2014) argued that marital stress experienced during pregnancy can lead to higher mortality for male embryos compared to female embryos, thereby altering the sex ratio at birth. The authors use panel data from the National Longitudinal Study of Youth (1979) that reports the level of marital stress in 1992 and estimate a linear probability model of the following kind:

$$E(Y|x,w) = \alpha + \beta X + \gamma W, \qquad (1)$$

where Y is a binary variable representing the birth of a male child after 1992, X measures marital stress, and W is the total number of children had after 1992. The authors justify the use of a linear probability model stating that the choice of the functional form (e.g. logit, probit, etc.) does not affect the result of their analysis. The authors include additional "control" variables in their model, such as birth order and age at marriage, but for the moment, we ignore these additional variables, because what is relevant here is the effect produced by the inclusion in the model of W.

We can first suppose that the variable X (marital stress) is a determinant of family size (W). For example, in the analysis conducted by Mencarini and Tanturri (2006) on a sample of Italian women, family relationship instability is cited as a reason for not having (another) child by 34% of childless women and by 24% of women with only one child. Depending on whether we consider women with parity 0 or 1, this is the second and third most important reason given for not having (more) children.

We can also suppose that the sex of a child is a determinant of family size if there are groups within the population that are applying stopping rules.

Finally, if we assume that marital stress does not influence the sex of the children (the null hypothesis within the test being conducted), we conclude that the causal diagram representing the causal connections between the variables in model (1) is:

$$Y \rightarrow W \leftarrow X.$$

In this diagram, the variable W turns out to be a collider. The fact that this collider has been included as a control in the model is equivalent to conditioning the analysis on this variable. But as we discussed in the previous paragraph, conditioning the analysis on a collider triggers a spurious correlation between X and Y. This means that we do not know if the significant correlation found between marital stress and the sex of the child resulted from the biological process imagined by Hamoudi and Nobles (2014) or from the spurious correlation introduced by the collider bias. Additionally, note how the inclusion of other variables worsens the situation. The variable birth order is clearly a descendant of family size (for example, if W=1, then birth order will necessarily be 1). Thus, the model is also conditioning the analysis on a descendant of a collider.

The problem with model (1) lies in the inclusion of a bad control, namely W. Cinelli et al. (2022) recently published a list of the most frequently encountered bad controls in empirical analyses, showing the solutions to adopt to avoid distortions in the results. In the case considered here, regarding the relationship between stress and the sex of newborns, the solution is simply to remove the bad controls from the model.

4. Testing the association between stress and the sex of children

Verifying the link between stress during pregnancy and the sex of newborns is not straightforward, primarily due to the lack of adequate data. When a healthy individual faces a sudden challenge, such as a threat to their physical integrity, or during a natural disaster like an earthquake, there is an activation of their hypothalamus-pituitary-adrenal (HPA) axis resulting in the production of cortisol (the stress hormone). The concentrations of this hormone in blood, saliva, or even hair can indicate the level of stress in an individual. However, the samples of individuals for whom these measures are taken and for whom we know the sex of their offspring are generally small, making it challenging to identify the effect of cortisol on the sex of births. If such an effect exists, it is considered small and therefore requires large samples to be detected (Hamoudy and Nobles 2014).

To address this issue, we decided to test the effect of stress on the sex of offspring indirectly. It is known that exposure to adverse conditions during childhood can lead to chronic stress conditions during adulthood, when cortisol levels remain elevated even in the absence of external threats or challenges. Conditions most commonly

associated with chronic stress include financial hardship during infancy, poor quality of the relationship with the parents, absence of one or both parents, parental union dissolution, and experiences of parental violence (Lupien et al., 2009; Miller et al., 2007; Sassler et al., 2009). Miller et al. (2007) further analyse how the chronic stress induced by these adverse childhood and youth conditions is associated with a range of negative outcomes in adulthood, including heavy alcohol use, drug use, sexual risk-taking, anxiety, depression, interpersonal violence, and self-oriented violence.

We then used retrospective information collected during the seventh wave (2017) of the Survey on Health Ageing and Retirement in Europe (SHARE) to obtain information on the conditions experienced by the respondents during childhood and the sex of their children throughout their lives. Additionally, since this survey focuses on the population aged 50 and over, we have information on the entire reproductive history of these women.

We proceeded in two different phases. Firstly, we sought to verify if the conditions experienced during childhood were indeed significantly associated with depression at the time of the interview. This preliminary analysis serves to confirm that the hypothesized association between adverse childhood conditions, chronic stress, and depression (explicitly formulated in Hamoudi and Nobles 2014) was also supported in our data. To perform this test, we applied a chi-square test. First, we considered experiencing financial hardship during childhood. (the data are described in Table 1).

Table 1 – *Financial hardship suffered by a woman during childhood and the likelihood of depression.*

	Financial hardship during childhood	
Depression	Yes	No
1-2	129	4444
3-4	112	3302
5-6	61	1662
7-8	33	688
9-10	20	239

Note: data derived from wave 7 of SHARE.

The chi-squared statistic for this table is 23 with 4 degrees of freedom, which leads to a rejection of the null hypothesis of independence between these two variables (p-value < 0.001). The average score on the depression severity scale is 3.35 for individuals who did not experience financial difficulties during childhood, and 3.91 for those who did. Therefore, financial hardships experienced during childhood appear to have a long-term association with depression, consistent with the hypothesis that financial hardship during childhood is associated with a condition

of chronic stress. We can thus conclude that at least a fraction of people who experienced financial hardship during childhood have developed a chronic stress condition.

We should then expect that the sex distribution of births in this subpopulation is skewed in favour of female births.

Table 2 – *Financial hardship suffered by a woman during childhood and the sex of her children.*

	Financial hardship during childhood	
Sex	Yes	No
Male	184	5413
Female	177	5044

Note: data derived from wave 7 of SHARE.

We can, again, test this hypothesis by applying a second chi-square test of independence to the data shown in Table 2. This time, the value of the chi-squared statistic is 0.59 with only one degree of freedom. This test does not reject the null hypothesis of independence (p-value = 0.81). So, experiencing financial hardship during childhood does not seem to produce any detectable effect on the sex distribution of children born to a woman. Note that in testing the association between financial hardship experienced by the mothers and the sex distribution of the children, we have not controlled for the existence of potential confounders. This is for several reasons. First, the (null) hypothesis that we are considering is that the sex distribution of the children is an exogenous variable (because this is the role given to this variable in the natural experiments that we cited in the introduction). As such, the sex distribution of the children, being a pure source of random variability, is supposed to not be affected by any variable. Similarly, it cannot be confounded by any other variable, because a confounder, by definition, is a variable that affects simultaneously the treatment, in this case stress, and the outcome, in this case the sex of the child. Second, by avoiding conditioning the analysis according to the value of other variables, we protect our results from a collider bias. As we have seen in the two previous sections, it suffices to include in a model a variable which is indirectly affected by family size to trigger such a bias. Since we do not really know the complete set of variables indirectly affected by family size, the safest approach is to avoid conditioning the analysis on any other variables. Based on these considerations, we thus believe that the approach based on a simple chi-squared test is also the safest way to test the association between chronic stress and the sex distribution of children.

Table 3 – *Testing the association between adverse condition experienced during childhood with depression (Test 1) and the sex distribution of children (Test 2).*

	Test 1		Test 2	
Adverse conditions experienced during childhood	**Sample size**	**Depression**	**Sample size**	**Sex imbalance**
Father no job	10667	yes	10667	no
Lived without mother at 10	11228	yes	11228	no
Lived without father at 10	11228	yes	11228	no
Mother understand worries	11017	yes	11017	no
Father understand worries	10515	yes	10515	no
Relationship with mothers	11070	yes	11070	no
Relationship with father	10575	yes	10575	no
Mother physical harm	11070	yes	11070	no
Father physical harm	10643	yes	10643	no

Note: The first three adverse conditions experienced during childhood were coded as binary variables. For the next six conditions, the variables are ordinal and represent a scale ranging from 1 to 5. In the two columns labelled "Depression" and "Sex imbalance," "yes" and "no" indicate whether the null hypothesis of independence was rejected (yes) or not (no).

We repeated the two types of tests shown earlier for nine additional adverse conditions experienced during childhood. The results of these tests are shown in Table 3. For all the adverse conditions examined, an association with a state of depression in adulthood was identified (the women considered were interviewed at the age of 50 or older). The data presented here thus seem to confirm the causal pathway that links adverse conditions experienced during childhood and adolescence to the onset of a chronic stress condition and a depressive state during adulthood. Nevertheless, when comparing the sex distribution of the children born to women who suffered adverse conditions during childhood with those who did not suffer from these conditions, we find that these distributions are indistinguishable from each other. Therefore, we do not find evidence of the causal pathway proposed by Hamoudi and Nobles (2014), which suggests that adverse conditions experienced during childhood lead to a chronic stress condition that, in turn, would lead to an imbalance in the sex distribution of offspring.

5. Conclusions

In their analysis, Hamoudi and Nobles (2014) argued that the sex of newborns is influenced by maternal stress levels and, therefore, cannot be considered an exogenous source of variability. This claim challenges a long-standing tradition of

research (Angrist and Evans 1998; Dahl and Moretti 2008; Bonsang and Skirbekk 2022) that has used the sex of children to conduct natural experiments. In this paper, we aim to assess the robustness of Hamoudi and Nobles' claim from both a theoretical and empirical perspective.

From a theoretical point of view, we first aimed to demonstrate how stopping rules can produce subtle distortions. We showed how such behaviours can create the illusion that variables such as income or stress determines the sex of children. This phenomenon occurs when the analysis is conditioned on family size or any variable that is directly or indirectly influenced by family size. It is thus possible that the innovative analysis conducted by Hamoudi and Nobles might suffer from this type of problem. However, it should be noted that this is not necessarily the case. The bias might still be small and not significantly alter the results of the analysis. For this reason, we believe this analysis should be repeated by removing bad controls from the model.

From an empirical point of view, we employed an indirect method to verify the effect of stress on the sex ratio of births. Hamoudi and Nobles (2014) assert, based on extensive literature, that adverse conditions experienced during childhood and adolescence increase the likelihood of suffering from chronic stress. Therefore, we sought to verify the existence of an association between a set of these adverse conditions and the sex ratio of children for a sample of approximately 10,000 women. However, our analyses were unable to identify any sex imbalance among women who suffered from these conditions. It is possible that this is due to the (indirect) method used, or the fact that the sample size is still too small. Whatever the reason, the hypothesis of an association between stress and the sex ratio of births seems to require further research work.

References

ANGRIST J.D. 1990. Lifetime Earnings and Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *The American Economic Review*, Vol. 80, No. 3, pp.313-336.

ANGRIST J.D., EVANS W.N. 1998. Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review*, Vol. 88, No. 3, pp. 450-477.

BONSANG E., SKIRBEKK V. 2022. Does Childbearing Affect Cognitive Health in Later Life? Evidence From an Instrumental Variable Approach. *Demography*, Vol. 59, pp. 975-994.

CINELLI C., FORNEY A., PEARL J. 2022. A Crash Course in Good and Bad Controls. *Sociological Methods & Research* https://doi.org/10.1177/00491241221099552.

DAHL G.B., MORETTI E. 2008. The demand for sons. *The review of economic studies*, Vol. 75, pp. 1085-1120.

DUNNING T. 2012. *Natural experiments in the social sciences: a design-based approach*, Cambridge University Press.

GRECH V., JAMESB W.H., LAURI J. 2018. On stopping rules and the sex ratio at birth. *Early Human Development*, Vol. 127, pp.15-20.

HAMOUDI A., NOBLES J. 2014. Do Daughters Really Cause Divorce? Stress, Pregnancy, and Family Composition. *Demography*, Vol. 51, No. 4, pp. 1423-49.

IMBENS G.W., RUBIN D.B. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences. An Introduction*. New York, Cambridge University Press.

JACOBSEN R., MØLLER H., ENGHOLM G. 1999. Fertility rates in Denmark in relation to the sexes of preceding children in the family, Human Reproduction Vol. 14, No. 4, pp.1127-1130.

MENCARINI L., TANTURRI M.L. (2006). High Fertility or Childlessness: Micro-Level Determinants of Reproductive Behaviour in Italy. *Population*, Vol. 61, No. 4, pp. 389-415.

MCKENZIE D., GIBSON J., STILLMAN S. 2011. How Important is Selection. Experimental vs. Non-Experimental Measures of the Income Gains from Migration. *Journal of the European Economic Association*, Vol. 8, No. 4, pp. 913–945.

PEARL J. 2009. *Causality. Models, Reasoning and Inference*. Cambridge, Cambridge University Press.

Giambattista SALINARI, University of Sassari, gsalinari@uniss.it
Gianni CARBONI, University of Sassari, gcarboni3@uniss.it
Virginia ZARULLI, University of Padua, virginia.zarulli@unipd.it