To Randomize or to not Randomize? 1970s Princetonian Solutions for Selection Bias

Arthur Brackmann Netto[1]

Abstract

This paper aims to explain how Orley Ashenfelter and James Heckman "discovered" selection bias and dealt with the problem in their own ways, explaining how the first simulated randomization and the second "modeled out" the bias for creating new econometric estimators. Historically, it is interesting to notice that, although nowadays their solutions are in opposed sides of the contemporary debate (randomistas vs. structural modelers), Heckman and Ashenfelter have quite similar backgrounds. Both are labor economists with Ph.D. from Princeton. They are almost contemporaneous, the first is the youngest born in 1944, and the second is two years older, born in 1942. Consequently, during their initial careers, Heckman and Ashenfelter had similar knowledge and published similar research. They even published numerous papers together until 1974. From this first broad view, the papers will scrutinize what happened from 1968 - Heckman's entry in Princeton's Ph.D. - to 1978/79 when both presented their different solutions for sample-selection. For achieving this objective, the paper uses bibliometric and primary sources for scrutinizing the publications and events of the period.

1) Introduction

1971 was turmoiled year. 500.000 walked through the streets of Washington against the Vietnam War. At the same time, America's discrimination issues were on their peak. Conflicts were on the streets and affluency on the hand of a few. Inside academic walls, the war for changes was also being fought during this period. Discrimination was a palpable problem. You could sense it, but you could not measure it. The battle to measure discrimination and manifest it as a manageable problem was not as violent, widespread or poetical as that on the streets. Still, in the economic site of the academic war, this was the battle being fought.

For economists the war was answering societies questions. In 1971, the Civil Rights Act was active for seven years already. But had it changed something? Had Johnson's war on poverty worked? Did discrimination in the labor market reduce? Economists worried with discrimination should answer those questions to policymakers. They had to reach them out with comprehensible measures and models to make changes happen.

However, in 1971, discrimination in the labor markets was demonstrating to be stastically hard to be dealt with. Social programs had data, but only on those who had participated in the programs. There were no control groups. Wage and employment data were only about those who worked and not about those who did not work. How could economists deal with these samples if they did not describe the entire populations? What conclusions could be drawn from these biased samples? In contemporary econometric language, econometricians had still to discover a way to deal with selection bias.

In this context, 1971 was also a special year for two Princetonian young scholars. James Heckman was receiving his Ph.D. from the Industrial Relations Section of Princeton's Economics Department, while Orley Ashenfelter was becoming the director of this same section. This paper aims to explain *how Ashenfelter and Heckman "discovered" selection bias and dealt with the problem in their own ways, giving rise to two opposing communities: randomistas and structural modelers*.

Historically, it is interesting to notice that, although nowadays "randomistas" and structural modelers are in opposed sides of the contemporary debate, Heckman and Ashenfelter have quite similar backgrounds. Both are labor economists with Ph.D. from Princeton. They are almost contemporaneous, the first is the youngest born in 1944, and the second is two years older, born in 1942. Because of this minimum difference in age, common interests, and common alma mater, Heckman and Ashenfelter met at Princeton. They even published together during the first years of their careers (Ashenfelter and Heckman 1971,1972,1973,1974). Consequently, during their initial careers, Heckman and Ashenfelter had similar knowledge and published similar research. From this first broad view, it is natural to question: *What happened from 1971 to 1978/79 when both presented their different solutions for sample-selection?*

2) Preamble: Randomization and Selection Bias

Selection-bias, from the point of view of other disciplines, is a specific type of the more general problem of sampling and non-sampling errors. These are errors that occur in data sets created through interviews of sampled populations and are common in a wide array of areas. Sampling errors occur when the sample does

---

[1] Doctoral student at the University of São Paulo (USP), Brazil. E-mail: arthurbnetto@usp.br

not represent the aimed population. When this error occurs because the interviewee had an omitted reason to participate or not participate (missings and truncation), the problem is called selection-bias. On the other hand, non-sampling errors are the errors that occur in the interview process such as errors of the interviewer, respondents and the researcher that built the questionnaire. Interviewers may record and question wrongly, and they may even cheat when completing the interviews. Respondents may be unwilling or incapable of answering. Researchers may censor and truncate the possibility of answers when designing the answer pool.

In contemporary economics, sampling, non-sampling and selection biases all came to be known biases under the umbrella of omitted variables (see: Wooldridge 2002). Their consequence is identical to that of simultaneity. Omitted variables impair the assumption that the error term is random in the following manner. Observe these two equations:

$$y_i = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_i x_i + \beta k + \varepsilon_i \tag{5}$$

$$y_i = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_i x_i + u_i \tag{6}$$

In this situation, it is possible to see that in the second equation the term βk has been omitted. From the first equation w,e know that $cov(x_i, k) \neq 0$. As a consequence, in the second equation we must have $cov(x_i, u) \neq 0$. Therefore, the error term is correlated with the regressors. It is not possible to state that the regressors cause the explained variable because there is an omitted factor that may be the real cause.

In other disciplines, given the possibility of active engagement of the researcher in the collection of data, sampling and non-sampling error are corrected through random sampling[2]. Missingness at random does not bias the estimators, while any type of selection bis that omits characteristics does. This is observable through the following figures:

Fig. 1 – Missing at random vs. selection-bias
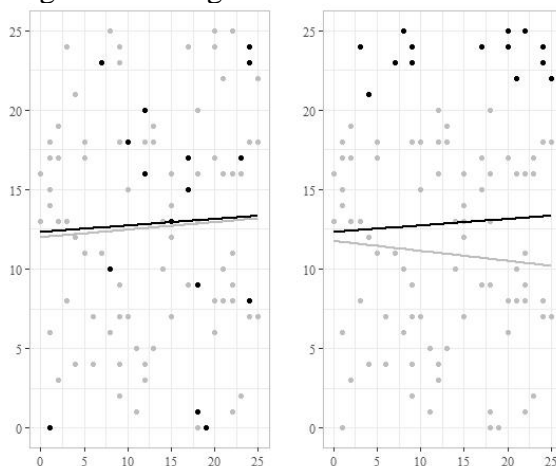
Source: Produced by the author

When data is missing at random, the estimated curves are almost identical. On the other hand, when data is missing for some omitted reason the curves have drastically different inclinations. Hence, conclusions are biased. In econometrics, econometricians do not have control over the collection of data and cannot create an experiment that collects random samples. Graphically, this means that econometric data is almost always as data of the figure in the right.

This is unfortunate, because random sampling is so common that ancient and renamed Statistics' Journals - such as the Journal of the Statistical Society of London and Biometrika – included mentions to "random sampling" already in their first volumes, in 1838 and 1901 respectively. As a result, statisticians have been dealing with methods to solve failures in the collection of data for a long time, which could serve as inspiration for microeconometricians.

Beyond theoretical research in statistics, as a practical tool, randomization and experimentation have always been applied in diverse fields. Experimentation and randomization have a long history that can be

---

[2] See antonakis for technical…

traced to the 17th century (Jamison 2017), with more relevant works being conducted in the 19th century. In this regard, Levitt and List (2008) point out Pasteur's experiments with sheep in the last quarter of the XIXth century. Ian Hacking (1988) and Favereau (2014), on the other hand, highlight Charles Pierce's role in conducting randomized experimentation in psychology during the same period. While paternity of experimentation and randomization is an unresolved issue, researchers tend to agree that the 1920s marked the outburst of these methodologies (Favereu 2014; Levitt and List 2008; Diaz, Jimenez-Buedo and Teira 2010). Two major actors arise here: Ronald Fisher and Philip Wright. The First was responsible for establishing the role of random assignment (Fisher 1926, 1935); the second played an important role in the definition of instrumental variables (Wright 1928, Stock and Trebbi 2003).

The usefulness of randomization, thus, allowed it to be implemented in distinct areas such as agriculture, education, and health. Randomization was early adopted in agriculture to test different cropping methods. Ronald Fisher himself was known as a biostatistician and worked on Rothamsted Experimental Station – an agricultural research institute – from 1919 to 1933. During this same period, Jerzy Neyman (1923[1990]), wrote "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles" in polish. According to the translators Dabrowska and Speed, Neyman "introduces a model for the analysis of field experiments conducted for the purpose of comparing a number of cropping varieties, which makes use of a double-indexed array of unknown potential yields, one index corresponding to varieties and the other to plots" (Neyman 1923[1990], p.1)

The same early movement can be seen in the health sciences. From Pasteur's vaccination experiments in the 19th century (see Latour 1982), randomization gradually became a functional tool for researchers in health areas. Claridge et al. (2005) affirm that during the first half of the 20th century randomized controlled trials (RCTs) became a reality in medical research.

In education, Oakley (1998) highlights that in 1901, Thorndike and Woodworth (1901) were realizing an experiment to test increases in mental ability, an intercept of psychology and education research. This intercept - educational psychology - has a long history of experimentation. Levitt and List (2008) and Oakley (1998) add the case of McCall (1923) as another relevant experiment in the area during the first quarter of the 20th century. As a result, in the 1940s, technical discussions about experimental designs and statistics in educational and psychological research were already a common reality. *The Journal of Experimental Education* is an interesting manifestation of this reality, having its first volume released in September of 1932.

Even social scientists had their randomization solution for sampling and non-sampling error. Since the 1930s sociologists already advocated in favor of experimental methods in social sciences. Chapin (1931) and Greenwood (1945) are the most cited cases. However, Gosnell's (1927) experiment about voter turnout can also be added to the hall of well succeeded social science's experiments. In the 1930s, as demonstrated by Brearley (1931)[3], at least 13 universities had courses in experimental sociology. Hence, sociology's early entrance in experimentation allowed researchers to be aware of self-selection and the problems related to individual heterogeneity. In this respect, Paul Wallin in *The Journal of American Sociology*, preceding the discussion in economics in almost 30 years, states in the late 1940s:

"Many studies in psychology and sociology, as well as surveys of opinions, attitudes, or consumer preferences, require a sample of volunteers to serve as subjects or informants. This dependence on volunteer subjects is a problem because not all persons whose participation is solicited will consent. Some do so and others do not, the proportions falling in the two groups varying from study to study. Dependence on volunteers, therefore, has as a consequence the *self-selection* of the units comprising the sample. This violates the fundamental principles of sampling that the method used in selecting the sample be such that each person in the universe have an equal chance of being a part of it." (Wallin 1949, p. 539).

Still, from statistics to these applied experiments, they all dealt with imperfections in the collection of data. Hence, they coped with failures in active observation. Self-selection in econometrics is something that biases estimators in the exact same manner as non-random sampling in other disciplines, but the difference is that the problem could not have been controlled by the researcher. It was not the econometrician who made a mistake in the sampling procedure, but some external institution. As Haavelmo (1944) claimed in the 1940s, econometricians are *passive observers* who borrow data from institutions to conduct their researches. They do not have control over their data. Econometricians are the ones accountable for nonexperimental data (see Wold 1969).

---

[3] Cited in Oakley (1998)

As follows, self-selection cannot be avoided in econometric data. Data is gathered externally and flows to economists' tables. Until the 1970s no explicit solution had been found in economics. Solely problems surrounding self-selection were having their solutions discovered. Censoring was solved by Tobin (1958) in the 1950s. Unobsorvables in panel data were being tackled since the 1960s (Dupont-Kieffer and Pirotte 2009). There were even economists leaving passive observations aside and engaging in the collection of data themselves, such as Heather Ross in the New Jersey Negative Income Experiments and experimental economists. However, there was still something missing for economist to claim victory over passive observations of individuals. There should be a manner of simulating randomization in their borrowed data. Orley Ashenfelter and James Heckman proposed two different solutions for fulfilling this gap in the 1970s.

3) 1970s Princetonian Solutions for the Selection Bias

Ashenfelter's difference-in-differences' method was presented in 1978 in "Estimating the Effect of Training Programs on Earnings". Intuitively, difference-in-differences "creates" a counterfactual group from observations of treated and non-treated individuals in two different time periods. The idea is to simulate a randomized assignment of treatment and control where this randomization process has in fact not occurred. As Ashenfelter points out when discussing the main problems of program evaluation, there is an "extreme difficulty of implementing an adequate experimental design so as to obtain a group against which to reliably compare trainee." (Ashenfelter 1978, p. 47). Therefore, he proposes a "data collection system" that copes with the problem. Graphically, this means:

Fig. 2 – Difference in Differences
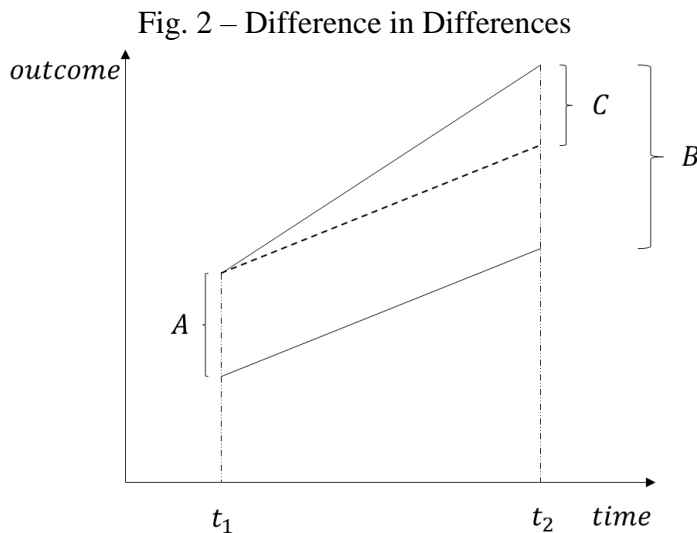
Source: produced by the author

What the figure demonstrates is the following definition. Without previous random assignment, it is possible to collect passive observations of two distinct groups -those who participated in a given program and those who not -, with different starting points. The interest, however, is not on the difference between these two groups, but in the difference of the counterfactual case, which is not apparently accessible. The diff-n-diff estimator allows to disclose this difference through the following equation:

$$Y_i = X_i'\alpha + \tau T_i + \varphi t_i + \beta(T_i t_i) + \varepsilon_i \qquad (10)$$

The interaction term β of the dummy variables time T_i and treatment t_i captures the effect C, which is the difference between B and A - and, consequently, a difference of two differences that names the methodology. Without this interaction term, only the effects A and B would be accessible, and those do not account for the true effect of the program, given that it is not possible to exclude the presence of omitted variables causing the differences with them. The diff-n-diff estimator, thus, assumes the parallel trend of the two groups in order to simulate randomization.

Therefore, the method is assuming out the presence of omitted variables in the collected passive observations through its design. This designed simulation of randomization is what has recently granted this kind of methodology the name of quasi-experiments. They do not control their data nor explicitly use any

form of randomization. However, through its design and assumptions, the methods can exclude omitted variables and allow causal claims.

Noticeably, Ashenfelter's insight draws a lot from the randomization literature. His method simulates randomization and is aimed towards the evaluation of programs. Heckman (1979), on the other hand, was concerned with a different kind of sample-selection, the misrepresentation of the aimed population in census' data. Technically, Heckman's first incursions with sample selection were concerns with missings and incidental truncation in U.S. labor statistics.

For motivating his research, he presented the problems of migrants, union members, and women's wage estimation. The example of women's wage estimation has since then become the canonical example in textbooks. Heckman observed that the wage is only observed for women who worked for wages[4]. Consequently, an omitted selection process of labor force participation biases the OLS estimators.

Intuitively, Heckman solved the problem with a two-step procedure. In the first step, Heckman proposes the use of an instrument to estimate a binary model of labor force participation, a probit model. In the second stage, a control variable is inserted in the original equation for removing the selection bias from the unobservable decision of participating in the labor market.

<mark>Technically</mark>

Heckman's solution was based on previous econometrics research, noticeably Gronau's (1974) and Tobin's (1958) papers. The first presented the motivation for researching truncated and missing data as selection-bias. The second worked on the problem of censored data using probit estimation of the latent variable, a procedure that resembles Heckman's and creates confusion even nowadays. Censoring and truncation are two common problems in econometric datasets, but only the second may create a relevant selection-bias. Therefore, Tobin's (1958) research can be said to be only insight into the use of binary models to model latent variables.

In contrast with Ashenfelter's proposal, Heckman's procedure does not simulate randomization, but "models out" the selection bias. The first stage of Heckman's procedure is the estimation of a model of decision-making. With the result of this model in hand, a second model for the observed variable is formulated. In the case of the labor market, for instance, the first model is a model of the decision of participating in the labor market and the second model a model of labor supply and demand equilibrium for determining optimum wage rates. From this point of view, it is interesting to observe that Heckman explicitly states that sample selection bias differs from other omitted variables in the sense that they can be theoretically formulated and distinctly estimated:

"This paper discusses the bias that results from using nonrandomly selected samples to estimate behavioral relationships as an ordinary specification bias that arises because of a missing data problem. *In contrast to the usual analysis of 'omitted variables' or specification error in econometrics, in the analysis of sample selection bias it is sometimes possible to estimate the variables which when omitted from a regression analysis give rise to the specification error.* The estimated values of the omitted variables can be used as regressors so that it is possible to estimate the behavioral functions of interest by simple methods." (Heckman 1979, p. 153)

Thus, intuitively and technically, the two solutions differ on their understanding of the cause of the omitted variable bias. Ashenfelter's does not assume the possibility of knowing the cause of the bias and opts to simulate randomization as a way of excluding all possible omitted variables. Heckman's, on the other hand, has an implicit assumption that the cause of the bias is known and can be modeled out.

From this previous exposition it is possible to assert that Ashenfelter advocates in favor of randomization, while Heckman in favor of structural modeling. What this means, according to Qin (2013) and Heckman (2010), is that in Heckman's structural case, the workflow consists in specifying a theoretical model, identifying the causal claim excluding endogeneity bias, estimating the econometric model and finally testing the model, while Ashenfelter's methodology may condone to some degree the theoretical aspect. As result, Ashenfelter's proposal gave rise to a community known as "Randomistas". In their point of view, randomization is the gold standard of the evaluation of programs. There is not the necessity of modeling the problems when it is possible to utilize or simulate randomization[5].

---

[4] A formal definition of the model can be found in Wooldrige (2002) section 17.4.

[5] This dichotomy between randomization and structural modeling has resulted in a heated contemporary debate (see: Deaton and Cartwright 2017, Heckman and Vitacyl 2005, 2007, Heckman 2010, Angrist and Pischke 2010). Usually, the most important point taken out of the debate is the timespan of application of the two methodologies. Structural models are said to be "ex-ante" evaluations, while randomistas apply solely "ex-post" evaluations.

Their options of vocabulary suggest this division of "modeling solution" versus "randomization" as well. In 1979 the cumulative research of Ashenfelter and Heckman[6] differed significantly in their most used words. Heckman pended in direction to modeling vocabulary, while Ashenfelter was prone to data terms. The following figure demonstrates the top words for each author in their papers from the beginning of their carreers to the point where they proposed their methodologies:

Fig. 3 – Heckman and Ashenfelter main words comparison
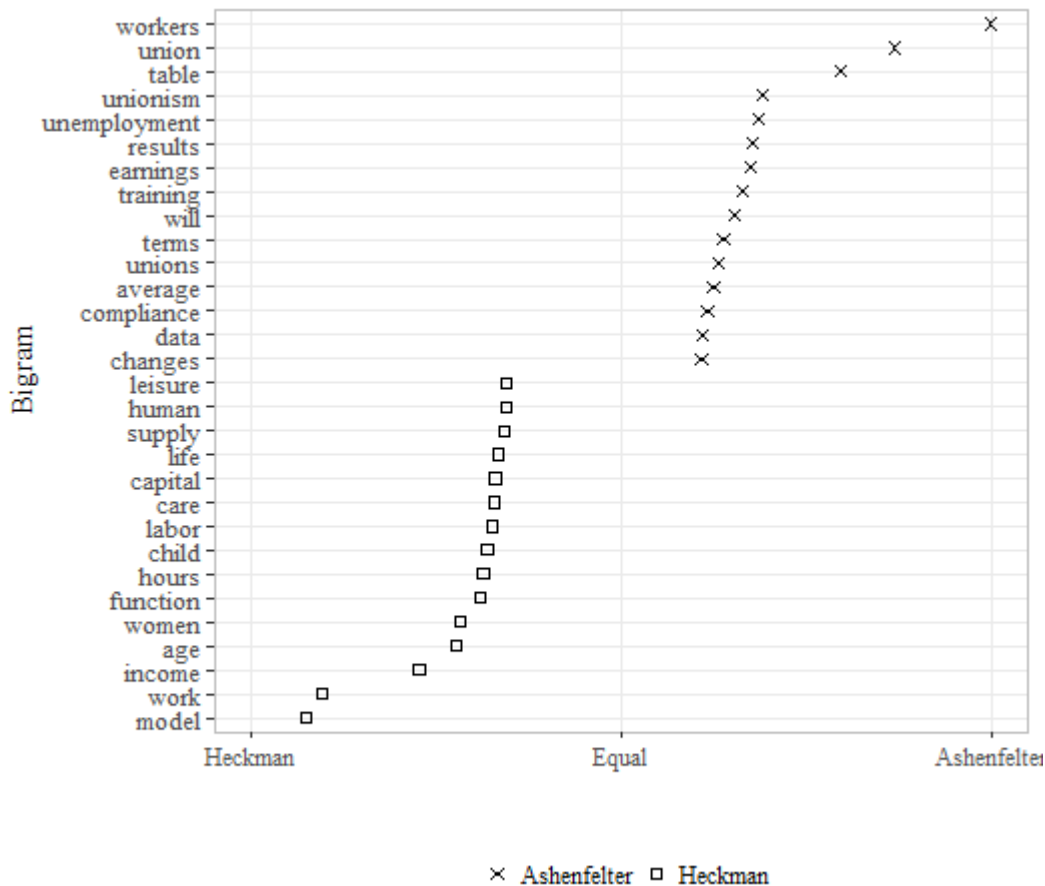
Source: Heckman and Ashenfelter's paper list[7]

It is possible to observe that Heckma used the words "model" and "function" in bigger proportion than Ashenfelter. On the other hand, Ashenfelter presents in his research a higher presence of the words "data" and "table". As follows, the bibliometric evidence endorses the argument presented earlier that the estimators differed in strategies. Heckman's estimator is a model out solution while Ashenfelter present a simulation of randomization. That is the conceptual difference, but how did both solved the same problem in two different versions? What happened in their carrers from their first publications together to the end of the 1970s?

## 4) The same problem, different bibliometric trends

The answer to the previous questions can be traced through their different scientific inspirations and institutional connection throughout the years. For now, with a simple intuitive exposition of the methods, it seems that Ashenfelter had influences from randomization and Heckman from econometrics[8]. However, this is a technical and ahistorical point of view. Is this perception confirmed in historical and sociological perspectives? The following figure demonstrates how their researches have differed over the years until the point they have proposed their methodologies:

Fig. 4 – Cumulative research differentiation

---

[6] This means all papers…
[7] See appendix 1
[8] Techcnically and ahistorically the sources of inspiration could be seen in other areas. Heckman (1996) explains the relationship between I.V. and Randomization, for instance.
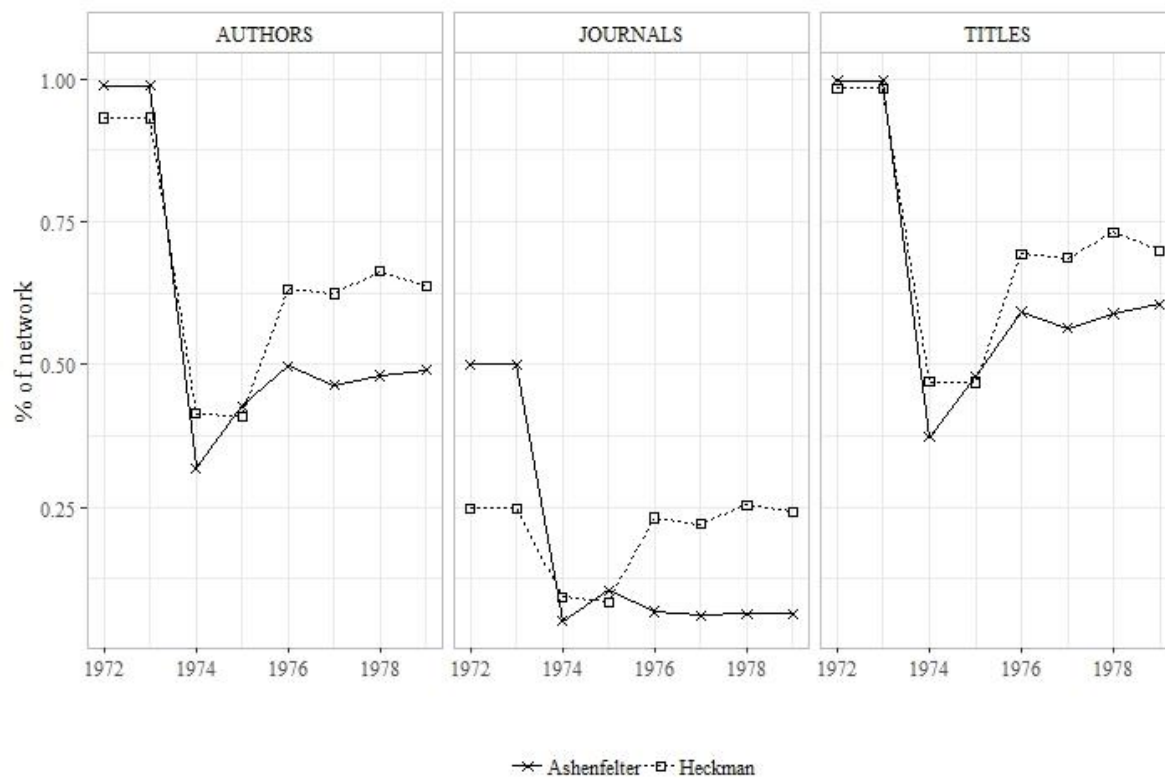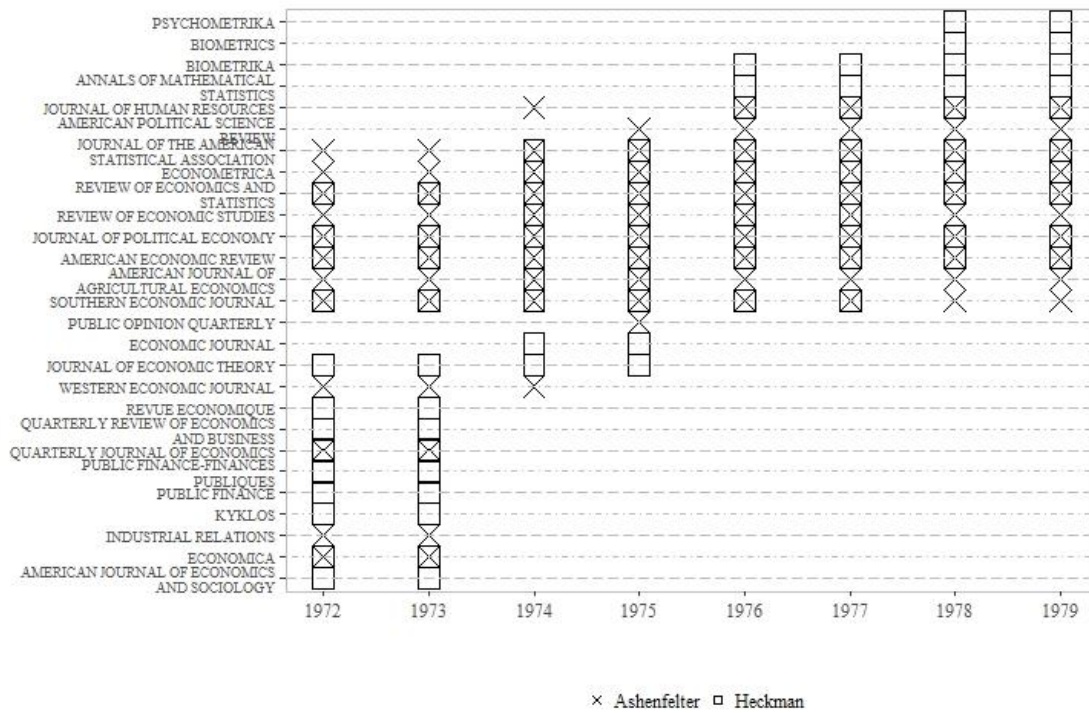
Source: Web of Science

The figure demonstrates the comparison of the cumulative related network of Heckman and Ashenfelter[9]. It demonstrates the percentage of related works included in each other's network. Therefore, as it can be seen, in 1972 and 1973, their networks were almost completely different. Then, in the following two years, when they worked together, their researches converged to similar related networks. Around 40% of the authors were distinct in the networks in those years, but the Journals covered were almost identical, given that only 10% of the Journals' network could not be found in one another's network.

More importantly, after the convergence, we observe a new period of differentiation and the gap in the figure represents that Heckman's related network was greater than Ashenfelter's. This means that Ashenfelter's related network reached only 35% of the same research scope of Heckman's, while Heckman reached 50% of Ashenfelter's research scope. This illustrates that Heckman was divergent, producing papers that escaped from their previous research topics. A more specific picture can be seen in the following figures:

---

[9] See appendix 2

Fig. 5 – comparison of Journals

× Ashenfelter □ Heckman

Source: Web of Science

In this figure, it is possible to observe that the journals covered by the related networks were considerably different in the two first years. The convergence is clear in the second period, from 1973 to 1974. However, more interestingly, is the case of differentiation. From 1976 forward, Heckman's main Journals started to include *Psychometrica*, *Biometrics*, and *Biometrika*. The first is a Journal concerned with the measurement and mathematization of human behavior. The second with the application of statistics to biology with dedicated papers on theoretical statistics, the third was mainly concerned with theoretical statistics.

Heckman presents 197 titles of these three Journals in his network, while Ashenfelter only has 9 titles in his related network. Hence, theoretical statistics is far more relevant in Heckman's research than in Ashenfelter's. However, Heckman was not alone in doing works similar to those present in these Journals. Goldberger (1971), in the beginning of the 70s, had already published a "survey of communalities" between econometrics and psychometrics in *Psychometrika*. Still, even though not alone, the evidence suggests that Heckman research was similar to two important domains, the modeling of human behavior and theoretical statistics.

It is interesting to notice that a common thread in these Journals during the 70s was the use of likelihood methods and binary models, especially the logit model. In this respect, table 1 of appendix 2 demonstrates that Heckman's research resembled several papers of the psychometric community about categorical and incomplete data written between 1970 and 1976. From this point of view, Heckman's research had since the begginings of the 70s, a relevant source of inspiration for the "model out" solution using the probit estimation that did not come from econometrics.

In contrast, Ashenfelter had a related network connected with political science and agriculture, as seen by the prevalence of *American Political Science Review* and *The American Journal of Agricultural Economics* in his network starting in 1976. Theoretical statistics was not absent, but the bibliometric evidence points out that his research was more closely related to practical fields. Ashenfelter has 191 articles in his related network written in journals about either political science or agricultural economics, while Heckman has 91.

In table 2 of appendix 2, it is possible to notice that Ashenfelter's research had strong similarities with empirical estimation of the equilibrium of numerous different markets, such as food, health facilities and even shrimps. Moreover, different from Heckman, Ashenfelter's research resembled works from the 50s and 60s, even though most similarities were found with works form the 70s as well.

This as interesting fact because, from the analysis of the diff and diff methodology, it seemed that Ashenfelter's research was embedded in researches using randomization in agriculture and sociology. However, the bibliometric evidence suggests that his research, firstly, had more contact with econometric research on those fields, with formulation of theoretical models and estimation of equilibrium. The following

figure demonstrates that in the bigrams of the titles of the related networks, "economtr model" and "econom theori" were among the most found cases in Ashenfelter's network during the whole period:

Fig. 6 – Comparison of Bigrams in the Titles of the Related Networks
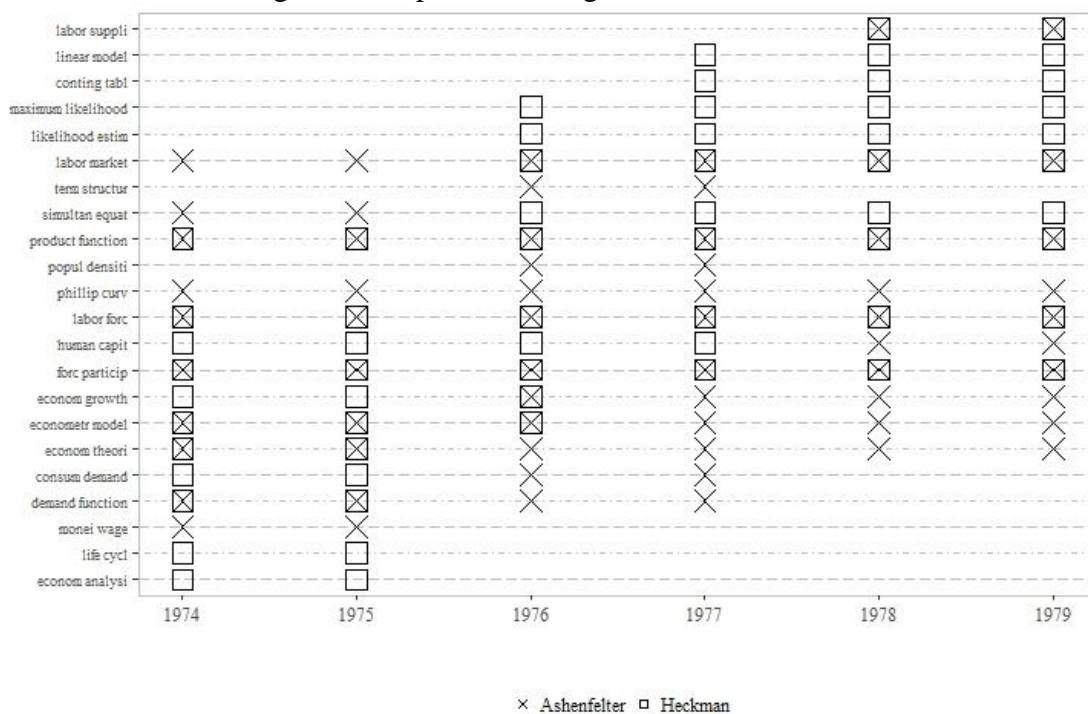
Source: Web of Science

As it can be observed, Ashenfelter's research, although based on practical fields such as agriculture and political science, was still economic in essence, dealing with themes such as the phillips curve and human capital. Heckman, as expected, has its related network connected with labor (as in "labor market" and "forc partic") and strongly connected with topics in theoretical statistics and econometrics, especially likelihood methods. The situation is similar in the case of authors as it can be seen in the next figure:

Fig. 7 – Authors Comparison
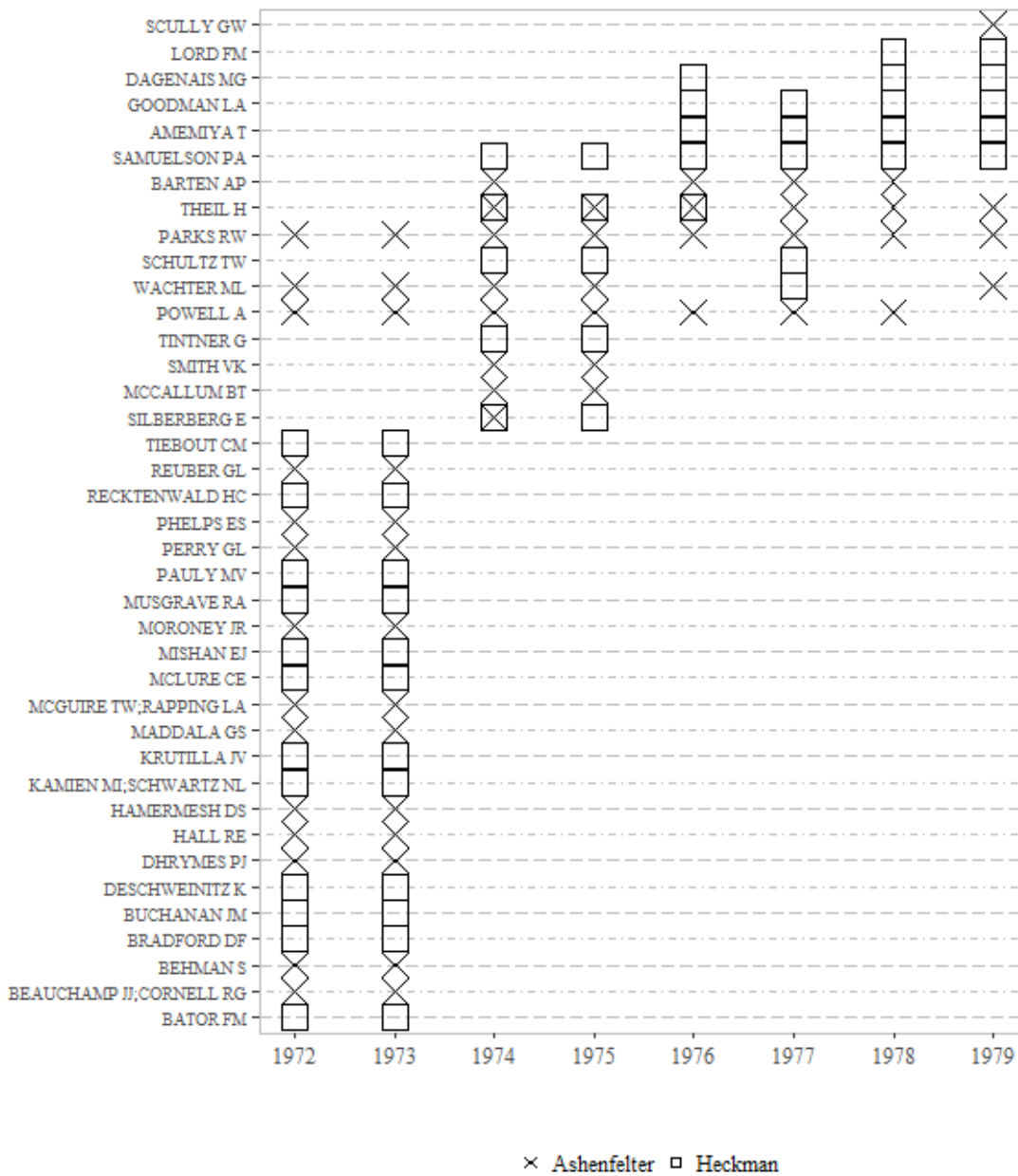
Source: Web of Science

In the case of authors, the trend of Heckman's network differentiation towards statistics is confirmed, since Heckman starts to present authors from statistics and mathematical statistics in his network. The most relevant Authors in Heckman's related network are: Amemyia, Goodman, Lord and Dagenais (see table 3 in appendix 2). In this regard, Lord and Goodman are two interesting appearances in Heckman's network. The first was famous for presenting the "Lord's paradox". His paradox defined the possibility of finding two different solutions with the same dataset depending on the initial adjustment of the data (Lord 1967). Heckman's method, it must be remembered, is exactly that: a correction of the dataset. Therefore, since 76 Heckman's research resembled statistics ideas.

The second, as observed in table 3 of appendix 2, worked with different kinds of data such as surveys and dichotomous variables. In the same vein, Heckman's related network contains works of Amemyia and Dagenais about different types of data that were written since the beginning of the 70s. Taking this into consideration, the evidence suggests that Heckman had also in econometrics inspiration for dealing with unusual types of data.

In contrast, bibliometric evidence on the authors of the related network of Ashenfelter suggests that his works resembled papers of Powell, Barten, Parks and Wachter (see table 4 in appendix 2). In this regard, Ashenfelter research was similar to researches on the estimation of demand and supply once again. Moreover, his research resembled papers on theoretical microeconomic problems, such as Slutsky conditions. From this point of view, Ashenfelter was embedded in a research that was far from being atheoretical.

This brief bibliometric research confirms the first impressions about Heckman's and Ashenfelter's syntheses. Ashenfelter had influences from randomization, but there is no evidence that he had them from theoretical discussions. His network is closely related to sociology and political science and from there is the most probable source of inspiration for the diff and diff estimator. Heckman, on the other side, was far from limited by econometrics discussions as was the initial appearance. His research network went further into technical statistical discussions.

5) 1968-1978: changing contexts

While bibliometrics is an interesting tool to understand *what* happened, it does not indicate *how* things have happened. The fact is that in the previous analysis solely with bibliometrics differentiation occurred gradually, given that both authors had the same Ph.D. formation and they even published together until 1974. But this change is more complex when the historical facts come to the narrative.

Heckman and Ashenfelter were quite different individuals who, yet, briefly enjoyed each other's company while at Princeton. Their exchanges had deep impact in their future publications. Still, their personal contact was brief and before the bibliometric data sugests, since Heckman stayed in Princeton only from 1968 to 1971. After that he left for Columbia and Chicago, while Ashenfelter stayed at Princeton, with a brief sojourn at the Office of Economic Opportunity in 1972.

5.1) Princeton years: Albert Rees and Late-night talks

In 1968, Chicago was already renamed. Economics was viewed as a science. Milton Freidman and George Stigler were the epitomes of the economist qua scientist. And the young Ph.D. student, James Heckman, opted to leave Chicago for Princeton. Put like that, it seems this young scholar had made a mistake. However, from his personal point of view it was the right move to make.

Chicago's atmosphere was hard for an outsider. Heckman had a strong mathematical background, far better than his economics skills. That in a place where students were many and attention for basic or even controversial questions was scarce, made him want to leave. Protests and lack of mathematical propensity in the department may have had their roles as well.

Be as it may, the fact is that Heckman choose Princeton as his new home. Princeton Industrial relations department was small and familiar, contrasting with Chicago. The department was renewing itself. Institutionalists were leaving their place for a young generation of empirical professors. In 1966 Harry Kalejian had come to the department after his Ph.D under the supervision of Marc Nerlove and Arnold Zellner. In 1968, Stanley Black – supervisee of James Tobin – and Albert Rees – supervisee of H. Gregg Lewis – were joining the department from Chicago and Yale respectively. Alan Blinder and Daniel Hamermesh would join the department in the beginning of the 70s after their Ph.Ds. There Heckman found not only a young faculty interested in empirical labor, but also a host of noteworthy students. Heckman enrolled by the side of Orley Ashenfelter, Ronald Oaxaca and John Pencavel who were then studying the new field of labor economics. It was a younger and exciting place to be Ph.D. student.

In this ambiance, not only Heckman was thrilled, but all members of the new community. Something new was being built. Frank Harbinson, who had been the director of the Industrial relations Department for 15 years, hence, left the post for the newcomer Albert Rees. Rees had arrived in Princeton in 1968, along with Heckman, and brought with him enthusiasm for building an empirical community. Microdata was emerging rapidly and had not yet been used appropriately. Rees engaged in transforming the department in the home of the application of this microdata. In this regard, Rees kept renewing Princeton's faculty towards empiricism, being the responsible for enlisting Alan Blinder and Daniel Hamermesh, for instance.

However, his real impact in the renewing of the department was through supervision. Ress has had important supervising impacts in the research of Oaxaca, Pencavel, Ashenfelter and Heckman, since he was part of the Ph.D. committee of all of them. In the words of his students:

"He [Rees] was a very conscientious and courteous adviser of many students. Teaching was important to him and, in the 1970s, he authored the major textbook at that time in Labor Economics (1973). His gracious manner made him a popular teacher and colleague." (Ashenfelter and Pencavel 2010,)

"Rees knew his price theory, but he rooted his research in data and encouraged students to take data very seriously. His vision of economics, like that of many Chicago economists of his generation, was that it is only of value if it is practically useful. He was an honest, open, friendly person who encouraged curiosity and imposed high standards on those around him." (Heckman 2014, p. 309)

Both Heckman and Ashenfelter knew Rees well. In the case of Heckman, Rees impact stems from the fact that, when Heckman enrolled in Rees labor economics class, he was the sole student to do so. Rees "wanted to cancel the class". Heckman "pleaded with him that [he] wanted to learn labor. So [they] agreed to have a tutorial." (Ginther 2010, p. 558). That created a bond between them. Heckman (2014, p. 309-310) remembers that Rees was important for his change from Chicago to Princeton, for presenting neoclassical labor economics in his third year of Ph.D. and introducing him to the new array of microdata emerging in the 1960s and 1970s. Heckman even participated in the New Jersey Income Maintenance Experiment captained by Rees to help "enroll some of the first participants in the experiment".

In the case of Ashenfelter, the relationship was obvious and direct, given that Rees had been his Ph.D. supervisor. Their exchange during Ashenfelter's Ph.D. years even resulted in a co-organized conference about discrimination in the labor market in 1971 followed by its correspondent joint book a few years into the 1970s.

As follows, both have had intense contact with Rees' empirical mindset. Rees knowledge of statistics and economic theory made him soon aware of the puzzles behind microdata and the collection of data. In Rees and Jacobs (1961) he was already aware of the biases caused by different selections of items to include in price indexes. Few years later, in 1966 he was commenting on the new data collected by the Department of Labor Statistics claiming for information on 'omitted variables':

"[a] problem exists in the unemployment statistics we have used for twenty-five years. How many of the unemployed cannot find work because the wage they are asking is unreasonably high relative to their abilities? No one knows, yet I cannot recall this issue ever being seriously raised in connection with the collection of unemployment statistics." (Rees 1966, p.462).

This concern comes from the "empirical" Chicago school where Rees had had his formation. H. Gregg Lewis supervised Albert Reees in his years at Chicago. Lewis, differently from Friedman and Stigler who were famous theoreticians, was an empiricist concerned with careful collection and analysis of data (see Rees 1976). His most well-known works are two exhaustive reviews of the literature on unions, dissecting statistical methodologies over all the works he had found on the issue.

As a result, in the 1970s, Rees was simply transmitting the empirical-labor DNA ahead. All his students had to credibly collect data and overcome its biases. Not by accident, Heckman, Ashenfelter and Oaxaca presented techniques for dealing with omitted variables[10]. In this regard, until 1970, Ashenfelter and Heckman were on the same track: dealing with data-collection biases and applying labor economic theory to it. They discussed the problem in the hallways, with professors and even between them. As Heckman remembers:

"I was happy to be able to combine rigorous economic theory with microdata. […] I shared my enthusiasm for this general research program with Orley Ashenfelter, who was two years ahead of me in the graduate program and who took a job at Princeton in my final year of residence there. Ashenfelter's enthusiasm for economics, data and life was infectious. We would discuss topics into the night. We shared a common vision about applying economic theory to empirical work in labor economics. We wrote several papers estimating Slutsky income and substitution effects in labor supply. He also acquainted me with Becker's Economics of Discrimination and shared with me his research ideas on explaining the time series of racial wage and employment patterns."

Their connection went far beyon night chats. During their Princeton years, therefore, they worked together on empirical labor problems with strict connection to Rees interests. In Ashenfelter and Heckman (1971, 1972, 1974) they reviewed labor economic models motivated by the empirical research on the negative income experiments, in which Rees had had an important role. In Ashenfelter and Heckman (1973) they worked together on the measurement of the effects of antidiscrimination programs.

It is interesting to notice that in their work together they were already aware that data could hide information. Together in 1973, they highlighted that there are effects that "can be measured and effects that cannot". As a consequence, they engaged together in the formulation of framework "to measure and interpret program effects". While neither the difference in difference nor the two-step procedure has been advanced in the paper, it is clear that the intention to overcome hidden information was among their common interests.

5.2) Columbia and OEO: theory and practice

---

[10] (two step procedure, diff-n-diff, Oaxaca-Blinder)

Noticeably, up to 1971, both authors were in the same track. Things started to change during that year. In 1970 Heckman was finishing his Ph.D. and received an offer from Columbia. Given his interests in labor economic theory, Columbia seemed the perfect place to go. Gary Becker and Jacob Mincer should be there at the time he received the offer. Becker was coming back to Columbia after going to Chicago. However, as Heckman recalls:

"I went there hoping to work with Gary Becker; as of March 1970 when I had to decide where to go, he was going back to Columbia, but he changed his mind! He left me high and dry. But with Jacob Mincer and a ock of his first rate, highly motivated, and intellectually engaged students. So even though I did not then work with him directly, I saw his legacy and the devotion of his scholars and students first hand." (Heckman 2014)

Still, his option for Columbia demonstrates that Heckman was eager to follow an academic path on economic theory. He wanted to learn from the best in the rising applied fields of labor economics, "new home economics" and economics of education (see Grossbard 2006): Mincer and Becker. Columbia had a famous labor workshop led by the two renamed scholars, where young scholars went nervously present their fledgling researches. Heckman attended the workshop without missing, even though Becker was not there anymore.

During his first year in New York, Heckman also received a proposal to participate in the activities of the NBER. He gladly accepted the offer, since the NBER was a growing institution with a blooming array of young scholars. It was also a break from the labor workshops. Heckman recalls having informal but intense discussion that went into the night when at the NBER (Heckman 2014). At the institution, he also met Finis Welch and Robert Willis, with whom he had the opportunity of discussing his research. With Willis, Heckman wrote two papers during the 1970s.

Although always conscious of microdata problems, Heckman went to Columbia to develop his theoretical skills. It was by accident that he became an econometrician:

"At the end of my first year at Columbia, the chair of the department, Kelvin Lancaster, approached me to teach graduate econometrics because the resident econometrician was leaving and no one else in the department knew anything about econometrics. I took this as an occasion to teach myself Henri Theil's then-new Principles of Econometrics. His was a breakthrough book that introduced basic asymptotic theory into econometrics. Teaching this course enabled me to break into graduate teaching." (Heckman 2014, p.)

This coincidence put Heckman on the road to his two-step procedure. In studying econometrics, Heckman reviewed the problems that he had already encountered: "In my thesis I faced the recurring problem of having missing wages for about half the women whose labor supply I was seeking to determine. Moreover the missing wages were associated with women not working (zero hours of work)." (Heckman 2014). After some research on the topic, Heckman decided to overcome the problem he had rediscovered. In 1972 he worked on a specific case of the problem: nonworking women (see Heckman 1974). This was the first rough form of his "model out" solution. At the time, he was concerned with the lack of theoretical reasons behind the methods being implemented to measure wage differential in with U.S. labor data. For him, theoretically, all women whether they worked or not should be accounted in labor market analysis.

As it can be observed, it was a theoretician looking through the lenses of econometrics to solve a problem yet without solution. He included, thus, a first theoretical step in the estimation of women wages that regarded their decision to participate or not in the labor market. With positive feedback, Heckman noticed that his framework could be generalized and already in 1972-73 he worked on the generalization of his methodology:

"working at Columbia, I began to develop a general framework for organizing discrete, continuous, and joint discrete-continuous variables in a common framework. I realized that my method for using economic theory to produce counterfactual missing wages was more generally applicable. In early 1973, I wrote a paper called "Dummy Endogenous Variables" that spelled out a general framework for modeling interrelated discrete and continuous choice models." (Heckman 2014)

Heckman's unpublished version of his 2-step procedure circulated widely. However, It took him around seven years to take it into print. Hence, as soon as 1973, Heckman had already developed his theoretical methodology for dealing with missing information. He frequently acknowledges the relevance of New York's community for developing his method. The interaction between the NBER and Columbia's Labor workshop was unique and has never repeated itself after NBER moved to Cambridge in 1973 (see Grossbard 2006).

Ashenfelter, on the other hand, followed a different path, but also had an early version of his methodology. He was already part of Princeton's faculty since 1968, although he has finished his Ph.D. only in 1970. Given his connection to Albert Rees, being his supervisee and sharing his enthusiasm for empirical labor economics,

Ashenfelter became the director of Princeton's Industrial Relations Department in 1971. The choice was interesting. Ashenfelter was acquainted with the peculiarities of Princeton, having been its student since his B.A., but he was also was a young and enthusiastic researcher that could maintain the same empirical track that had became Princeton's gene in the 1960s.

Ashenfelter had to make substitutions in the department as he saw his Ph.D. colleagues and professors leaving the house. Hamermesh went to Michigan state and Kaleijian to New York University. Oaxaca and Pencavel, in the same manner as Heckman, left right after their Ph.Ds to Western Ontario and Stanford respectively. As result, Ashenfelter brought to the department Farrel Bloch and Sharon P. Smith. However, similarly to Albert Rees, Ashenfelter's most lasting contribution to the department was not on his substitutions, but in a different area. Ashenfelter, in 1972, had the opportunity to create a network with the U.S. Department of Labor.

During that year, Ashenfelter was on a leave as the director of a branch of the Office of Economic Opportunity. In his words: "in early 1972 I was offered a civil service position in the U.S. Department of Labor in which I was to direct an Office of Evaluation whose sole purpose was to ask and answer this [did training programs work?] and some related questions" (Ashenfelter 2014, p. 2).

The office had been created during Johnson's war on Poverty with the sole purpose of evaluating the overwhelming amount of data accumulated by its programs – mainly programs related to the Manpower Development and Training Act (MDTA) such as Head Start and Job corps. According to Ashenfelter (2014), the office was a place where there was no political interference, and the sole purpose was to find a transparent manner of evaluating programs. There, randomization definitely entered in Ashenfelter's research agenda.

Ashenfelter joined civil service for a year, where, beyond economists, there were several other researchers – especially from political science, education and sociology. All of them had the common notion that the active collection of data and randomization was the way to go. As Ashenfelter points out from his experience - and Latour (1982) and Druckman et al. (2006) discuss sociologically -, he discovered during that year that randomization is an interesting tool to "whisper in the ears of kings":

"a key reason why this procedure [difference-in-differences as a simulation of randomization] was so attractive to a bureaucrat in Washington, D.C., was that it was a transparent method that did not require elaborate explanation and was therefore an extremely credible way to report the results of what, in fact, was a complicated and difficult study" (Ashenfelter 2014)

This was the case because it does not make sense neither for researchers nor for policy makers to compare participants of training programs with comparison groups that are not counterfactual. There is no "credibility". In the words of Ashenfelter, comparing with a non-counterfactual group made him aware of the necessity of simulating randomization with longitudinal data:

"The key thing learned from this comparison was that the program participants had lower earnings, both before and after the program, than the comparison group. This automatically made it clear that the analysis would not meet the highest standards for credibility. This also suggested that the participants should be compared with themselves instead of with the comparison group alone, and with longitudinal data that is precisely what was possible." (Ashenfelter 2014)

As a consequence, Ashenfelter developed his Difference in difference estimator already in 1973-74 and circulated it in the labor department and used it to inform policymakers. He, as well, took some time to put his work on print. The now widely cited estimator was only published in 1978.

5.3) Establishing difference: Chicago and Princeton

In accordance with the bibliometric evidence, 1974 is an important year for the authors. Both had already concluded most part of the research related to their estimators and were establishing themselves as researchers. They had already finished their Ph.Ds for some time and had now to establish themselves academically to continue their research agendas.

In the case of Heckman, this meant leaving New York. In 1973, when the NBER was going to Cambridge, Heckman received a proposal from Becker to assume a position in the university of Chicago. As stated, Chicago was among the most renamed institutions in economics at the time and Becker, a guiding a exemplar for Heckman, was there. Considering that the NBER-Columbia cluster was dismantling, the decision was simple for Heckman. As a consequence, he joined the economics faculty of Chicago by the side of Milton

Friedman, George Stigler, Gary Becker, Robert Lucas and a host of memorable economists. According to Heckman:

At Chicago, Heckman got tenured and received the task of teaching Gregg Lewis' class on unionism already in 1974, only one year after his arrival. In the research side, Heckman expanded his research agenda quickly. He soon got his 1974 paper published and started to work on application of the methodology for different areas. At the same time, he formulated the general framework that would be published in 1979. Still, Heckman found time to research different topics such as life cycles and human capital for instance. As a result, from 1974 to 1979 Heckman published more than twenty papers – most of them as a single author. In Heckman's words:

"Chicago in the 1970s was an ideal environment for me to conduct research. One-on-one interactions and seminars improved my thinking on any project I undertook. I had developed an agenda in my first three years at Columbia, and I expanded on it and generalized it in my early years at Chicago." (Heckman 2014, p. 322)

Either for its quantity or quality, Heckman's research circulated widely. His novel ideas were applied in an array of different topics and areas. The next table demonstrates the amount of JSTOR's mentions for "heckman" by subperiods, showing Heckman's reach in sociology, mathematics and statistics.

Table 1 – Mentions to Heckman by subperiods

Discipline	1969-1974	1974-1979
Economics	22	309
Sociology	7	58
Statistics	1	30
Mathematics	4	59

Source: JSTOR

As it can be observed, Heckman readily got studied for his technical model in theoretical areas such as mathematics, statistics and mathematical sociology. In this regard, Heckman (2014) remembers meting Burt Singer and James Coleman - eminent mathematical sociologists - in the 1970s, with whom he worked together. After his initial contact with mathematical sociology, Heckman developed a lifelong interest by the subject, participating frequently in mathematical sociology workshops.

Being studied in these diverse fields became a challenge that pu Heckman on the track of developing his model. In the 1980s, for instance, after frustrated attempts to apply Heckman's framework to problems where normality could not be assumed, Heckman developed a nonparametric formulation of his method.

Columbia, hence, was the place where the idea was born, but Chicago was where Heckman disseminated his model and created a network outside economics. Chicago offered him the opportunity of developing his research freely and even interacting with renamed scholars from other areas. Columbia, on the toher hand, offered him the opportunity of linving intensely the field of labor economics and new home economics within the Columbia-NBER cluster. Both played essential roles in the development of the tow step procedure.

However, it is interesting to notice that was in Chicago that Heckman established and generalized his framework as a "modeling out" procedure for the identification of causal relationships. In this sense, Chicago's context may have played an special role in this definition. As Heckman remembers: "The entire department [Chicago's] seemed to embrace Friedman's methodology of positive economics. […] No theory was worthwhile unless it survived a reality check.". In this context, Heckman was developing a research essential for the department. He was allowing theoretical questions that were before unanswerable to face a reality check for the first time. Heckman's framework implicit stated that no labor theory could be confronted with econometric data without consulting it first. Friedman's positive method could only be applied in labor economics through Heckman.

Ashenfelter, on the other hand, finished his sojourn at the office of economic opportunity and came back to his position as the director of Princeton's Industrial relations section in 1973. Starting that year, he centralized Princeton as the place of empirical labor economics. The section's research seminars were renewed after his comeback. Whereas during Albert Rees directorship most seminars were held by Princetonian Scholars, from 1973 forward, research seminars started to count with several renamed scholars from outside

Princeton. Presenters came from several places, but specially from University of California LA and the University of Chicago.

Table 2 – Seminar presentation by institutions

Institution	Presentations (1973-1979)
University of California, LA	6
University of Chicago	5
University of Massachussets	4
Harvard University	4
Cornell University	4
University of Wisconsin	4

Source: Princeton's Industrial Relations Section Reports

Ashenfelter managed to centralize the discussion of labor economics and evaluation of Programs in the university of Princeton. Beyond renewing the seminars, Ashenfelter organized two conferences on the evalution of Programs in 1974 and 1976 respectively, counting with the presence of numerous renamed scholars and researchers form the U.S. labor department. These conferences ended in the publication of books such as Ashenfelter and Blum (1976).

It is interesting to observe that both conferences were co-sponsored by the governmental organs. In 1974, James Blum from the Office of the Assistant Secretary for Policy, Evaluation and Research of the U.S. department of labor co-organized the conference with Ashenfelter. In 1976, it was the time of Ernst Stromsdorfer, fomr the same office, to help in the organization of the conference. In both conferences, therefore, Ashenfelter made an effort to integrate policymakers and academics. Among the participants and discussers there were several former academics that were at the time working for governmental institutions. The list of invitations was equally divided between academic and government invitees. These efforts of creating bonds with governmental organs were relevant for two main reasons: first, access to data; second, learning how to translate academic research to practical purposes.

These intentions were well succeeded. Data became abundant in the Industrial Relations Section. According to Robinson (2016), the industrial relations section became one of the main users of the University's Computer Center. Both graduate students and the faculty benefited from the growing computerization. The increasing ammount of microdata, such as the 1960s census and the 1970s first U.S. Labor datasets could only be grasped in a computer and Ashenfelter was quick in maintaining a relationship with the computer center.

Amidst his responsibilities as Director of the section, Ashenfelter managed to position himself as an important researcher as well. From 1974 forward, he specialized himself in the evaluation of programs, becoming one of the first economists of the area. This meant embracing randomization and creating a "credible" language for communicating academic discovers to policymakers.

From 74 to 78, Ashenfelter published six papers in the working papers series of the industrial relations section that would sooner or later become important published researches. Among the papers, in 74 and 76, Ashenfelter published his works on the evaluation of MTDA programs. Also, in 76 and 77 he published evaluation of other programs such as employment tax credit and the income maintenance experiments. In all works, simple and understandable economic analysis were Ashenfelter's effort. In his 1977 evaluation of the employment tax credit, for instance, his primary concern was to translate the confuse pages of the employment tax act to a more simple economic language: "To do this I first set out in simplified form the accounting details of how ETC is designed to operate in terms of the change in the wage rate it may be expected to induce." (Ashenfelter 1977, p. 1)

6) **Concluding Remarks**

Historical facts confirm what has been demonstrated by the bibliometric and technical evidence: Heckman and Ashenfelter differed their researches due to their different point of views, theoretical and applied respectively. However, while technical and bibliometric evidence point to a gradual modification of ideas,

historical facts demonstrate that already in the fisrts years after their Ph.D. they had already developed their estimators.

Bibliometric and technical evidence point to this distinct conclusion especially for one main reason: print time. While in the end of the 1960s Ashenfelter and Heckman were already writing their papers together, they would come to print only in 1974. The same is true for their estimators, both had already develo9ped their ideas in 1974 but would publish their works in 1978 and 1979.

Nevertheless, although contrasting on their conclusions about timing, all evidences point to the same history. Both Authors were concerned with the same problem during part of their careers: biases in the collection of econometric microdata. This is seen in the bibliometric evidence in their almost identical related networks during the period they published together. This is also seen by the historical fact that both learned about this issue through albert Rees, with whom they had important contact.

In this regard, it is interesting to notice that what is now a clear clash , was historically only a change of context. Both authors had the same concerns, but during the period from 1968 to 1974 they moved to completely different contexts. Heckman went to New York where he worked on his theoretical skills and became an econometrician by accident. Ashenfelter stayed at Princeton and briefly worked at the Office of Economic Opportunity where he had contact with the credibility of randomization. Quickly, in their new contexts, they proposed their solutions that took seven to five years to come to print.

Ashenfelter and Heckman established their research agendas from 1974 to 1979. Heckman engaged in the task of conciliating heterogeneity in economic behavior with econometrics and Ashenfelter engaged in the consolidation of the field oe evaluation of economic programs. This is contrasts in some degree with the bibkiometric data that indicates that the authors were differentiating their researches during this period. However, as already seen, bibliometric data presents a delay especially due to print and editorial times, but also for personal deicsions about the gestation period of the papers. Thus, the bibliometric results remain but can be read as delayed versions of the historical facts.

References

Angrist, J. D. et Pischke J-S, (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking Con out of Econometrics", Journal of Economic Perspectives, 24(2): 3-30.

Antonakis, John; Samuel Bendahan, Philippe Jacquart, Rafael Lalive, (2010) On making causal claims: A review and recommendations, The Leadership Quarterly, Volume 21, Issue 6, Pages 1086-1120

Ashenfelter and Heckman (1971) - The Estimation of Income and Substitution Effects in a Model of Family Labor Supply. Working Papers (Princeton University. Industrial Relations Section) 29.

Ashenfelter and Heckman (1972) - Estimating Labor Supply functions. Working Papers (Princeton University. Industrial Relations Section) 34

Ashenfelter and Heckman (1973) - Measuring the Effect of Antidiscrimination Program. Working Papers (Princeton University. Industrial Relations Section) 52

Ashenfelter O. (2014) The Early Historyof Program Evaluation and the Department of Labor. Working Paper (Princeton's Industrial Relations Section) 580.

Ashenfelter, O and Heckman,J (1974) The Estimation of Income and Substitution Effects in a Model of Family Labor Supply Econometrica, Vol. 42, No. 1 pp. 73-85

Ashenfelter, O. & Pencavel, J. (2010). Albert Rees. in: The Elgar Companion to the Chicago School of Economics, chapter 12 Edward Elgar Publishing.

Ashenfelter, O. and Blum, J. (1976) Evaluating the Labor-Market Effects of Social Programs. Princeton University, Industrial Relations Section; 1st edition. 238 p.

Brearley HC. Experimental sociology in the United States. Social Forces 193; Dec:196-9

Chapin, F. Stuart (1931) The Problem of Controls in Experimental Sociology. The Journal of Educational Sociology, Vol. 4, No. 9 pp. 541-551

Deaton, A., and Nancy C. (2018). "Understanding and misunderstanding randomized controlled trials." Social Science & Medicine 210 (August 2018): 2-21.

Díaz, M. Jiménez-Buedo, D. Teira, (2015). Quasi- and Field Experiments. In: James D. Wright (editor-in-chief), International Encyclopedia of the Social & Behavioral Sciences, 2nd edition, Vol 19. Oxford: Elsevier, pp. 736–741.

Druckman, J., Green, D., Kuklinski, J., & Lupia, A. (2006). The Growth and Development of Experimental Research in Political Science. American Political Science Review, 100(4), 627-635.

Favereau, J. (2014). L'approche expérimentale du J-Pal en économie du développement: un tournant épistémologique? (Doctoral dissertation) Retrieved from: http://www.theses.fr/2014PA010010

Fisher, R. A. (1926). The arrangement of field trials. Journal of the Ministry of Agriculture of Great Britain 33, 503-513.

Fisher, R. A. (1935). The Design of Experiments. Oliver and Boyd: Edinburgh.

Ginther, D. (2010). AN INTERVIEW WITH JAMES J. HECKMAN. Macroeconomic Dynamics, 14(4), 548-584. doi:10.1017/S1365100510000611

Gosnell HF. Getting out the vote: an experiment in the stimulation of voting. Westport, CT: Greenwood Press, 1977. (Originally published 1927.

Greenwood E. Experimental sociology. New York:Octagon Books, 1976:72.(Originally published 1945).

Gronau, R. (1974) Wage Comparisons-A Selectivity Bias. Journal of Political Economy, 82 (1974), 1119-1144

Grossbard S. (2006) The New Home Economics at Columbia and Chicago. In: Grossbard S. (eds) Jacob Mincer A Pioneer of Modern Labor Economics. Springer, Boston, MA

Haavelmo, T. (1944) The probability approach in econometrics. Econometrica, 12, supplement; mimeograph (1941) at Harvard University.

Hacking, Ian, (1988). Telepathy: Origins of Randomization in Experimental Design. Isis, 79(3): 427-451.

Heckman, J (2014). James J. Heckman. In Spencer R. & Macpherson D. (Eds.), Lives of the Laureates: Twenty-three Nobel Economists (pp. 233-266). MIT Press.

Heckman, J. (1974) Shadow Prices, Market Wages, and Labor Supply Econometrica, Vol. 42, No. 4 (Jul., 1974), pp. 679-694

Heckman, J. and Vitacyl, E (2005). Structural Equations, Treatment Effects and Econometric Policy Evaluation. Econometrica, 2005, 73(3): 669-738,

Heckman, J. J. (1996). Randomization as an Instrumental Variable. The Review of Economics and Statistics, 78(2), 336-341.

Heckman, J. J. (2000) Causal parameters and policy analysis in economics: A twentieth century retrospective, Quarterly Journal of Economics, 115, 45–97.

Heckman, J. J. (2010). Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy. Journal of Economic Literature 48 (2): 356–98.

Heckman, J.J. (1979), 'Sample selection bias as a specifi cation error', Econometrica, 47 (1), 153–61.

Jamison, J. (2017). The Entry of Randomized Assignment into the Social Sciences. World Bank Policy Research Working Paper 8062.

Latour, B (1982), Give me a laboratory and I will move the world In K. Knorr et M. Mulkay (editors) Science Observed, Sage, 1983, pp.141-170 [New edition slighly abridged in Mario Biagioli (editor) Science Studies Reader, London Routledge, 1999]

Levitt, S., List, J. (2008). Field Experiments in Economics: The Past, The Present, and The Future. NBER Working Paper, 14356.

Lord, E. M. (1967). A paradox in the interpretation of group comparisons. Psychological Bulletin, 68, 304–305

McCall WA. (1923) How to experiment in education. New York: Macmillan, pp. 38-41

Neyman, Jerzy. 1923 [1990]. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." Statistical Science 5 (4): 465–472. Trans. Dorota M. Dabrowska and Terence P. Speed

Oakley, A (1998) Experimentation and social interventions: a forgotten but important history BMJ; 317 :1239

Published by:Wold, H. O (1969). Econometrics as Pioneering in Nonexperimental Model Building. Econometrica, Econometric Society, vol. 37(3), pages 369-381,

Qin, D. (2013). A History of Econometrics: the reformation from the 1970s. Oxford: Oxford University Press.

Rees, A. (1966) Now Is the Time to Lick Hard-Core Unemployment, Challenge, 14:6, 29-41.

Rees, A. (1976), 'H. Gregg Lewis and the development of analytical labor economics', Journal of Political Economy, 84 (4, part 2: Essays in labor economics in honor of H. Gregg Lewis), S3–8.

Rees, A., & Jacobs, D. (1961). Real Wages in Manufacturing, 1890-1914. Princeton University Press.

Robinson, L. (2016) Princeton's University's Industrial relations Section in Historical Perspective. Industrial relations Section.

Stock, J. H., Trebbi, F. (2003). Retrospectives: Who Invented Instrumental Variable Regression? Journal of Economic Perspectives 17 (3): 177–94.

Thorndike EL, Woodworth RS. The influence of improvement in one mental function upon the efficiency of other functions. Psychol Rev 190; 8:247-61, 384-95, 553-64.

Tobin, J. (1958) Estimation of relationships of limited dependent variables Econometrica, 26, 24–36.

Wallin, P. (1949) Volunteer Subjects as a Source of Sampling Bias. American Journal of Sociology, Vol. 54, No. 6, pp. 539-544

Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. Cambridge, Mass: MIT Press.

Wright, P. (1928). The Tariff on Animal and Vegetable Oils: Appendix B. Reproduced in Stock and Trebbi (2003).