ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 235, 2010

Tadeusz Bednarski[*], *Filip Borowicz*[**]

ANALYSIS OF NON-RESPONSE CAUSALITY IN LABOR MARKET SURVEYS

Abstract. Non-response in labor force survey may result in biased estimation of unemployment duration distribution. Possible reason is that non-contact with respondents may be the effect of exit from unemployment before intended survey dates. It is then called the causal effect of non-response. Extending results by van den Berg et al. (2006), a simple statistical method to identify the casual effect of non-response is proposed.

Key words: non-response bias, unemployment duration.

I. INTRODUCTION

Distribution of unemployment duration with its dependence on such characteristics as social status, education, sex, age, etc. constitute a commonly accepted way to describe state transitions in labor markets. There is a vast literature on the methodology devoted to longitudinal labor market data analysis. Hackman and Singer (2008) give a review of the topic indicating benefits and possible drawbacks of longitudinal studies in economics.

One of the problems frequently arising in statistical inference from social data, and in particular in longitudinal labor market studies, is the impact of surveys high non-response rate. It may result in biased estimation and consequently in inadequate identification of unemployment determinants. A way to bypass this problem is better understanding of the non-response mechanisms. We refer to papers by Groves (2006), Boudarbat, Grenon (2000), Pedersen (2002), van den Berg et al. (1994, 2006), where extensive studies of non-response bias in longitudinal survey data are given.

Absence in surveying may arise in many ways. An unemployed person who spends much of his time searching for a job may not want to spend additional time on a survey interview. An economically inactive person also may not have a large interest in surveys no matter how close it is to finding a new job. In such cases we have a so called *selection effect* of non-response. On the other hand, transition from unemployment to employment may effect in respondent's change

[*] Professor, Institute of Economic Sciences, Wrocław University.
[**] MSc, Institute of Economic Sciences, Wrocław University.

of place he lives, may change his motivation to cooperate or simply may keep him busy enough to be unavailable to surveying agency. In other words, the acceptance of employment may make it more difficult for the survey organization to contact the individual. Therefore, if higher survey non-response results from unemployment exit before intended survey date, it is called the *causal effect* of non-response. Van den Berg et al. (2006) identifies these two main categories of non-response and propose a method of identifying them.

In this paper we propose a simple testing method of the presence of causal effects, based on the Cox model. The method is described heuristically and supplied with a Monte Carlo study.

II. THE VAN DEN BERG'S METHOD

As it was mentioned, an extensive empirical analysis of the impact of causal and selective effects on bias in statistical inference for unemployment data was presented by van den Berg, Lindeboom and Dolton (2006). The study was possible via access to two separate sources of data, based on the same group of unemployed persons. The first source of information were administrative records, originally collected by Policy Studies Institute for evaluating the effects of the policy program for unemployed workers “Restart”, introduced in the UK in 1987. The idea was to statistically control the efficiency of the policy program via a random sample of about nine thousands unemployed persons. Unemployed individuals were selected to the sample in March and April of 1989 in such a way that their sixth month of possible unemployment would be around May–June 1989. In fact two additional datasets were attached to the data collected by Policy Studies Institute: JOVOS (Joint Unemployment and Vacancies Operating System) and NOMIS (National Online Manpower System).

The second source of data, concerning the same individuals, came from a survey conducted by the Social and Community Planning Research (now known as the National Centre for Social Research). The survey was intended to supply additional information on background variables and job search behavior of the individuals. Interviews were conducted about 6 months after identification of the sample (in September and October 1989). It turned out that the survey response rate was only 56% (!).

Information on personal characteristics, such as sex, age, local unemployment rate and duration of unemployment was collected from the administrative records while survey data were used to determine the binary non-response indicator. After detailed analysis of the data van den Berg at al. suggested a method to distinguish empirically between selective and causal effects of non-response for a bias. Their method can be described as follows.

Let the binary non-response indicator Y be equal either to 0 or 1 depending on whether contact with the individual was established or not. Assume that the unemployment duration is described by a continuous random variable T and that the time interval between the moment of inflow into unemployment and the moment of interview is exactly c months. In the case studied by van den Berg et al. c corresponds to a period of twelve months - the survey was conducted in September and October 1989, approximately twelve months after each individual in the sample entered the unemployment state. The method is based on the following claim (Van den Berg et al. (2006)):

"… with a causal effect, non-respondents are more likely to have a duration outcome just before the survey date than just after, compared with respondents. A selection effect cannot give rise to such a strong local stochastic dependence of the duration outcome on the response status."

Therefore the authors conclude that the time dependent conditional probability $P\left(Y=1|T=t,X\right)$, where X is an explanatory variable, has to jump downwards at time $t=c$, while $P\left(Y=0|T=t,X\right)$ has to jump upwards at the same time. Since in practice, as the authors admit, it might be difficult to statistically detect the discontinuity of a probability function they instead suggest to examine the following expression

$$\frac{P\left(Y=1\middle|\,t=c^{+},X\right)\Big/P\left(Y=1\middle|\,t=c^{-},X\right)}{P\left(Y=0\middle|\,t=c^{+},X\right)\Big/P\left(Y=0\middle|\,t=c^{-},X\right)}$$

which in presence of a causal effect is to be smaller than 1, and it is to be 1 if the causal effect does not occur.

III. STATISTICAL INFERENCE FOR THE CAUSALITY FACTOR

Basic tools for statistical duration analysis of labor data, due to frequent censoring effects, are the same as those of survival analysis. As the Kaplan-Meier estimator (1958) let us estimate the overall duration distribution of unemployment under censoring, the Cox's regression model (1972, 1975) provides a way to measure the quantitative impact of various social characteristics on the distribution of unemployment time under censoring. More precisely, the statistical relationship between unemployment duration T and the vector Z of explanatory variables such as sex, age, educational level etc. is described by the *hazard function*, which is defined conditionally for fixed $Z=z$ as

$$\lambda(t,z)=\lambda_0(t)\exp(z\beta),$$

where $\lambda_0(t)$ is the baseline hazard function and $\beta \in R^k$ is a vector of regression parameters.

The following simple statistical method to identify the causal effect of non-response, based on the Cox model, is proposed. Define Y to be an artificial explanatory variable identifying the presence or absence of the contact with an individual:

$$Y=\begin{cases} -1 & \text{contacted with respondent at time } c \\ 1 & \text{not contacted with respondent at time } c \end{cases}$$

where c is the survey moment. Notice that if the probability of non-response changes with time in a manner independent of the unemployment duration T, so to say Y is independent of T then the estimator of the regression coefficient corresponding to Y will tend in probability to zero. One can heuristically argue that dependence of the probability of survey non-response on unemployment time duration would lead to a coefficient value different than 0. Therefore we can use the standard Cox regression model testing of statistical hypotheses to identify the causality effect by verification of

$$H_o : \beta_Y = 0, \quad H_1 : \beta_Y \neq 0,$$

where β_Y is the regression parameter of the artificial variable Y.

Notice that the joint distribution of the unemployment duration time T and the variable Y need not come from the Cox model. These are purely analytical properties of the Cox model that validate the method. The formal argumentation about the consistency of the method are still at work and they will be given elsewhere. Here we give a Monte Carlo study of the proposed method under "practically typical" conditions. The method is in fact much simpler than van den Berg's proposal. It is also more flexible from the practical standpoint since the survey time can be fixed individually, not globally for the entire population.

IV. SIMULATIONS

A Monte Carlo experiment was carried out to evaluate the efficiency of the proposed method. The time T representing the unemployment spell was generated from the exponential distribution with parameter $\lambda = 2$ and independently the survey times c was generated from the uniform distribution on [1/2, 3/2]. The experiment was repeated 10000 times with the sample size 1000.

The causal effect is associated here with the change of two survey non-response probabilities, depending on whether $c < t$ or $c \geq t$. Two cases were studied. In the first one the two probabilities were assumed not to depend on survey time c. In the second case time dependence was allowed.

The first case. Two variants of the distribution of the explanatory variable Y, depending on the relation between unemployment duration and survey times, were taken into account. For $c < t$, the survey non-response probability was equal p_1 and for $c \geq t$ the probability p_2 of non-response was established. The equality of p_1 and p_2 is equivalent to absence of the causal effect, while their non-zero difference is equivalent to existence of the causal effect. For each values of p_1 and p_2 the null hypothesis $H_o : \beta_Y = 0$ was tested at the significance level 0.05. Results obtained for 10000 repetitions at every fixed parameter value are shown on the following two graphs (Fig. 1 and 2). The horizontal axis shows the value p_1. The vertical axis gives frequency of correctly detected cases of the causal effect. Fig. 1 demonstrates the change of correct detection of the causal effect depending on the value of p_1 when the difference between the probabilities is precisely 0.1. Fig. 2 shows the same effect when the difference is 0.2. The two graphs let us then compare the impact of differences between the two probabilities on correct detection of causal effect. For the difference equal to 0.3 the correct detection increased practically to 1.

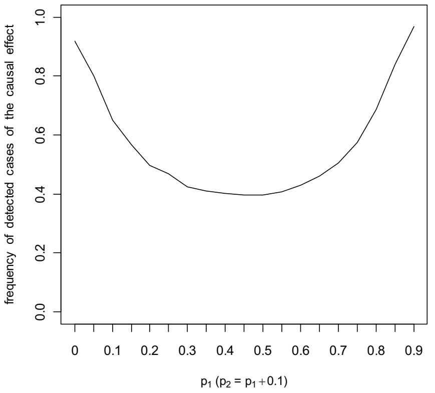

Fig. 1

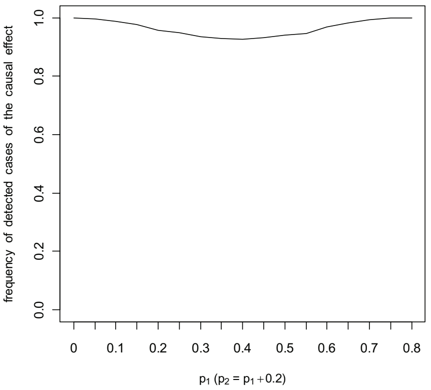

Fig. 2

Second case. Again two variants of the distributions of the artificial variable are possible. Now however they depend on survey time c in the following way

$$P(c) = e^{dc},$$

where *d* is the fixed value. This time the causal effect is associated with change of *d*, depending on whether $c < t$ or $c \geq t$. The two parameters vary in the interval [-3,-0.2]. The exemplary change of probabilities is shown on Fig. 3 and 4 where the parameter differences are either 0.1 or 1.5 and the separate curves correspond to c=0.1, 0.2, …, 1.

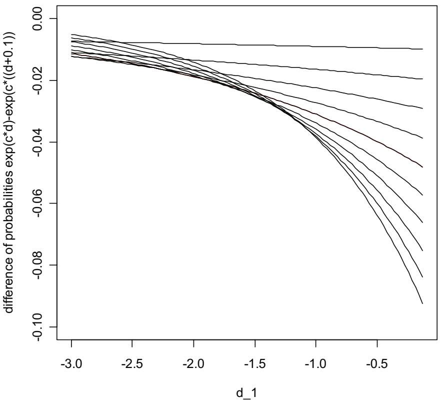

Fig. 3

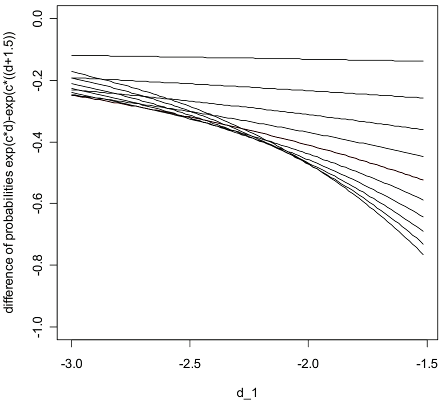

Fig. 4

Simulation results are presented on Fig. 5 – Fig. 8. The horizontal axis shows change in d_1 under constant difference between d_1 and d_2. It is visible that the method looses slightly its efficiency when the probabilities of non-response depend so strongly on time. However as in the case of constant probabilities the situation improves very much when the difference between the probabilities exceeds 0.2 under sample size equal 1000.

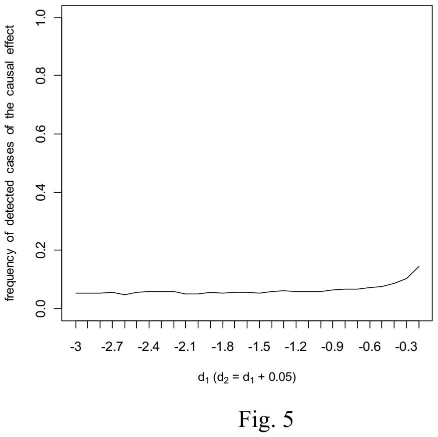

Fig. 5

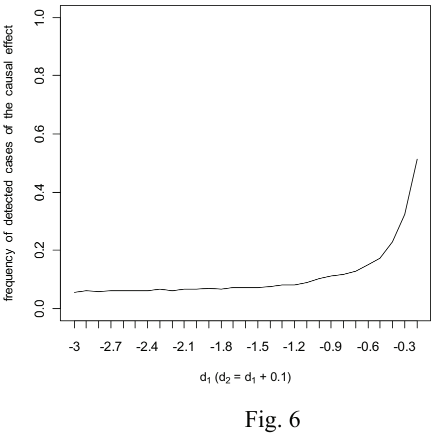

Fig. 6

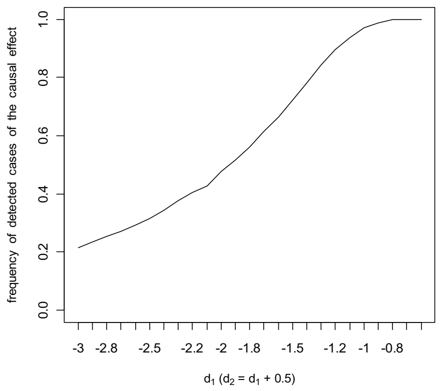

Fig. 7

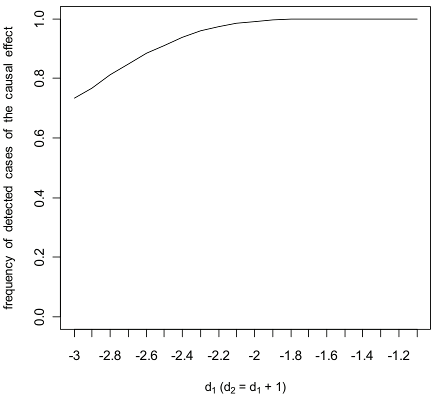

Fig. 8

Simulations performed for larger sample size (5000) showed higher detection rate of the causal effect – probability difference equal 0.05 resulted in detection rate roughly corresponding to the one shown on Fig. 1. A similar study was also carried out in the presence of explanatory variables, independent of the time c. Results are consistent with those presented above.

REFERENCES

Boudarbat B., Grenon L. (2007), Attrition and Non-Response in Panel Data: The Case of the Canadian Survey of Labor and Income Dynamics, submitted to the *Electronic Journal of Statistics*.

Groves R. (2006), Nonresponse Rates and Nonresponse Bias in Household, *Surveys Public Opinion Quarterly*, Vol. 70, No. 5, Special Issue 2006, 646–67.

Heckman J. J., Singer B. S. (2008), *Longitudinal Analysis of Labor Market Data*, Econometric Society Monographs 10. Cambridge University Press 2008.

Little R. J. A., Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, 2nd edn. New York: Wiley.

O'Muircheartaigh C., Campanelli, P. (1999), A multilevel exploration of the role of interviewers in survey non-response, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Volume 162 Issue 3, 437–446.

Pedersen J. P. (2002), Non-Response Bias - A Study Using Matched Survey-register Labour Market Data, Working Paper.

Romeo, C.J. (1997), Measuring information loss due to inconsistencies induration data from longitudinal surveys, *Journal of Econometrics*, Volume 78, Issue 2, June 1997, 159–177.

van den Berg G. J., Lindeboom M., Ridder G. (1994) Attrition in longitudinal panel data and the empirical analysis of dynamic labour market behaviour, *J. Appl. Econometr.*, 9, 421–435.

van den Berg, G. J., Lindeboom M. M., Dolton P. (2006), Survey non-response and the duration of unemployment, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 169, Number 3, 585–604.

Tadeusz Bednarski, *Filip Borowicz*

ANALIZA PRZYCZYNOWOŚCI ABSENCJI W BADANIACH RYNKU PRACY

Absencja respondentów w ankietowych badaniach rynku pracy może mieć wpływ na obciążoność estymacji rozkładu czasu poszukiwania pracy przez osoby bezrobotne. Brak kontaktu z ankietowanym może być skutkiem jego zatrudnienia przed przeprowadzeniem badania i jest to tzw. przyczynowy efekt absencji. W nawiązaniu do wyników przedstawionych przez van den Berga i innych (2006) proponowana jest prosta metoda umożliwiająca identyfikację efektu przyczynowego absencji respondentów.