What explains the negative effect of unemployment on health? An analysis accounting for reverse causality

Gerhard Krug[a,*], Andreas Eberl[b]

[a] University of Erlangen-Nuremberg and Institute for Employment Research (IAB), Germany
[b] Institute for Employment Research (IAB), Germany

ABSTRACT

The unemployed are often in poorer health than their employed counterparts. This cross-sectional correlation is often attributed to a causal effect of unemployment on health. Recent research analyzing longitudinal data often supports alternative explanations, such as spurious correlation and/or selection of unhealthy workers into unemployment (i.e., reverse causality). In this paper, we apply a dynamic panel data estimator (system GMM) to account for both unobserved confounders and reverse causality. Despite some evidence for health selection, we still find strong support for the causality thesis. Furthermore, we show that the adverse health effect is partially explained by the loss of self-perceived social status due to unemployment but not by the loss of household income or social contacts.

1. Introduction

It is well documented that the unemployed are in worse health than their employed counterparts (Kroll, Müters, & Lampert, 2016). There are three distinct but not mutually exclusive explanations for this difference: social causation, indirect selection and direct (or health) selection. The first and long-standing social causation explanation assumes that becoming unemployed is damaging to an individual's health (for overviews see Brand, 2015; Wanberg, 2012). Unemployment is considered a stressful life event (cf., Pearlin, 1989) that creates distress and leads to health problems. In support of this explanation, a variety of mechanisms − again, none of them mutually exclusive − have been put forward. One of the most comprehensive discussions of this issue is provided by Jahoda (1981), who distinguished between the manifest and latent functions of work. In this framework, becoming unemployed leads to a loss of these functions and consequently to a deterioration of health. First and foremost, the unemployed are deprived of the manifest function of work, i.e., to provide the necessary financial means so that they can support themselves and their families. In addition, work fulfills a variety of latent functions. Work not only enables individuals to contribute to a collective purpose and structures their daily activities but also provides status and societal recognition and widens individuals' social networks beyond family. Despite a long research tradition, the literature is often criticized for not testing the pathways or mechanisms underlying the social causation thesis (Bartley, 1994; Janlert & Hammarström, 2009). This lack of testing is unfortunate because, as argued by Goldthorpe (2001), for example, a test of the

hypothetical mechanism underlying a causal relationship can improve the ability of statistical analysis to provide us with convincing evidence of the existence or absence of this causal relationship.

Testing the underlying mechanisms of social causation is even more important, given that social causation is, by far, not the only explanation for the unemployment-health relationship. A second explanation refers to "indirect selection", i.e., the assumption that the relationship between health and unemployment might be mainly due to common causes. For example, workers with lower education have a higher risk of becoming unemployed (Mincer, 1991; Wolbers, 2000). At the same time, education is positively related to health (Conti, Heckman, & Urzua, 2010; Grossman, 1972). As a consequence, the common cause creates compositional differences such that the unemployed tend to be in poorer health than their employed counterparts. Accounting for indirect selection in quantitative research is a matter of controlling for all common causes, including those unobserved in the data, by suitable statistical methods.

Third, other scholars argue that there is a direct selection of unhealthy workers into unemployment. Sometimes, the term "reverse causality" is used to indicate that, compared to the social causation thesis, the direct selection explanation reverses the causal order between unemployment and health. The theoretical argument behind direct selection is that those workers whose health deteriorates become less productive and are more likely to be laid off (Bartley, 1988; Cook, 1985; West, 1991). Whereas the main bulk of the literature focuses on social causation, several authors have also argued for the importance of direct selection (Andreeva, Magnusson Hanson, Westerlund, Theorell, &

Brenner, 2015; Heggebø, 2015; Korpi, 2001; McDonough & Amick, 2001 Stewart, 2001; Strully, 2009; Virtanen et al., 2006). Empirically, the social causation and direct selection explanations are not mutually exclusive. However, most research designs do not effectively distinguish empirically between the two. As a consequence, empirical results in support of social causation are often criticized by referring to the possibility of direct selection as an alternative explanation.

In the analysis below, we attempt to go beyond the current literature by making the following contributions. First, we aim to identify the net effect of social causation by explicitly ruling out both indirect and direct selection as alternative explanations. We do this by applying a dynamic panel data regression based on the generalized method of moments (GMM). Distinguishing between social causation and selection is important because the two explanations have very different implications for policymakers. In cases of social causation, it would make sense to take measures to avoid (prolonged) unemployment. In case of (direct) health selection, such measures will only induce costs and will be ineffective. Here, measures to foster the health of employed individuals would be more efficient. Second, we aim to further investigate the social causation effect by testing several causal mechanisms, namely those based on the reduction in financial resources, social network resources and self-perceived social status. Identifying the relevant mechanisms will help policy makers better focus potential measures against the adverse health effects of unemployment.

The article proceeds as follows: In the next section, we review the literature on social causation and derive our hypotheses. We then describe our data − the German Panel Study "Labour Market and Social Security" (PASS) − and the dynamic panel data estimator based on the generalized method of moments as our method of choice. After presenting our results, we conclude with a discussion.

2. Unemployment and health: literature review and hypotheses

With regard to indicators of physical and psychological health, the unemployed are in poorer health than employed individuals. The largest part of the literature on health and unemployment is based on cross-sectional data analysis[1] (cf., Paul & Moser, 2009). Cross-sectional data allow researchers to control for the observed part of those variables that induce indirect selection. In the last two decades or so, researchers have become more sensitive to the problem of unobserved confounders when making causal claims and to the advantages of longitudinal data. When using longitudinal data, the bias due to indirect selection based on unobserved variables can be eliminated by applying fixed-effects regressions or similar methods, such as difference-in-differences estimation, as long as the influence of these unobserved confounders is constant over the observation period (Andreß, Golsch, & Schmidt, 2013; Brüderl & Ludwig, 2015).

In the following paper, we review only studies that apply such stricter methods of causal estimation and thus have eliminated the influence of indirect selection more convincingly than cross-sectional studies. The results of these studies are more heterogeneous than the results of studies based on cross-sectional data. Some authors who use longitudinal data also report negative and statistically significant causal effects on health or health-related behavior. For example, Young (2012) report significant effects on mental health in the US, Mandemakers and Monden (2013) in the UK, and Cygan-Rehm, Kuehnle, and Oberfichtner

(2017) in the US, UK, Australia and Germany. Minelli et al. (2014) report significant effects for self-rated health in Italy, and Marcus (2014) reports significant effects on smoking behavior and body mass index in Germany. Tøge and Blekesaune (2015) find negative causal effects for self-rated health in Europe and identify no negative health trends preceding unemployment that might point towards direct selection. Other authors report results that are more ambiguous or even point to no causal effects of unemployment. For Germany, Gebel and Voßemer (2014) find statistically negative effects on life satisfaction but no effects on health satisfaction. For Finland, Böckerman and Ilmakunnas (2009) find no effects of unemployment on self-rated health, but they do find that those who eventually become unemployed are in worse health than those who do not. They interpret this as support for the direct selection thesis. For Sweden, Korpi (2001) finds no effects on physical health for current unemployment, but the effects of past unemployment duration are statistically significant. In addition, the author also finds that ill physical health increases the risk of becoming unemployed. Therefore, he finds support for both the direct selection and social causation theses. Salm (2009) investigates firm closures in the USA and also finds no effects on physical or mental health, concluding that the correlation of health and unemployment is due to indirect selection. Black, Devereux, and Salvanes (2015) use Norwegian data indicating that the effects of unemployment on coronary disease are significant but very small. For Germany, Schmitz (2011) finds negative effects of unemployment on health satisfaction, mental health and hospital visits, but not for the unemployed who lose their jobs due to firm closures. Schmitz (2011) interprets this finding as evidence for direct selection. If unemployment is due to firm closure, selection based on poor health is unlikely because all employees lose their jobs. If this group of unemployed shows no health effects, but other unemployed do show such effects, this points to pure selection effects.[2] Heggebø (2015) explicitly and exclusively focuses on direct selection and finds significant effects of limiting, long-standing illness on unemployment in Denmark, no effects in Norway, and significant negative effects of unemployment in Sweden.

In all, research that eliminates the bias that is due to unobserved confounders only partly supports the social causation thesis. Distinguishing between mental and physical health, those studies focusing on mental health mostly report significant negative health effects, whereas results from studies using physical health or overall health indicators are more mixed. However, for both mental and physical health, even if indirect selection is excluded as an alternative explanation, the empirical validity of the social causation effect is still largely open to debate. The reason is that even studies that apply fixed-effects or difference-in-differences estimators do not − per se − distinguish between social causation and direct selection[3] (i.e., reverse causality; see the method section below). Therefore, our first hypothesis is the following:

H1. There is a negative causal effect of unemployment on health (social causation)

Research on the effects of unemployment on health rarely also tests for one or more of the underlying mechanisms. Those who perform such tests mostly apply methods that do not account for indirect selection based on unobserved confounders (Kessler, Turner, & House, 1987; Kokko & Pulkkinen, 1998; Paul, Geithner, & Moser, 2009). One of the most common explanations for the social causation effect of

---

[1] The literature also uses so-called prospective studies, which are based on two-wave longitudinal data. Paul and Moser (2009) report that in general, such studies report smaller correlations than cross-sectional studies, but differences between the unemployed and employed are still statistically significant. These studies compare individuals who did and did not become unemployed between two points in time. Control variables are measured at time 1, including baseline health. Health outcomes, in contrast, are measured at time 2. Such studies are slightly better equipped to control for indirect selection than purely cross-sectional studies, but unobserved heterogeneity is only partly accounted for, because no actual panel data regression are used, as compared e.g. to lagged dependent fixed-effects regressions.

---

[2] Firm closures are often seen as a way to eliminate the bias due to reverse causality, but see Morris and Cook (1991) for a comprehensive critique of using firm closures to research unemployment effects on health.

[3] Among the studies cited above, only Tøge and Blekesaune (2015) explicitly apply dynamic panel data estimator to address reverse causality. Unfortunately, the paper is not clear on the specification of the model and test results (are independent variables instrumented and if so, with how many lags), a pivotal point in applying such models (cf., Piper, 2015).