

Does Sport Affect Health and Well-Being or Is It the Other Way Around? A Note on Reverse-Causality in Empirical Applications

Journal of Sports Economics

2021, Vol. 22(2) 218-226

© The Author(s) 2020



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1527002520967400

journals.sagepub.com/home/jseJing Guan¹ and J. D. Tena^{2,3} 

Abstract

Estimating the causal impact of sport or physical activity on health and well-being is an issue of great relevance in the sport and health literature. The increasing availability of individual level data has encouraged this interest. However, this analysis requires dealing with two types of simultaneity problem: (1) between exercise and response variables; and (2) across the different response variables. This note discusses how the previous literature has dealt with these two questions with particular attention paid to the use of seemingly aseptic econometric models proposed by some recent empirical papers. Regardless of the approach, identification necessarily requires the use of untestable hypotheses. We provide some recommendations based on analyzing the robustness of the estimation results to changes in the adopted identification assumptions.

Keywords

simultaneity, SUR, structural models

¹ School of Economics, Beijing Technology and Business University, Beijing, China

² Management School, University of Liverpool, United Kingdom

³ CRENoS and Università di Sassari, Italy

Corresponding Author:

J. D. Tena, University of Liverpool, Chatham Street, Liverpool L19 7ZH, United Kingdom.

Email: j.tena-horrillo@liverpool.ac.uk

Introduction

The purpose of this note is to discuss the use of systems of simultaneous equations in the empirical literature to estimate the impact of health behavior variables, such as sport and/or physical activity, on health and well-being.¹ Recent years have witnessed the availability of surveys that allow for the observation of these variables together with other individual socio-economic characteristics. Although this information has generated a burgeoning research literature aiming to estimate the main determinants of health and wellbeing,² an important concern in this type of analysis regards the simultaneous observation of the different variables in the model. This issue becomes especially problematic as many of the databases are cross-sectional data which makes the identification of causal impacts a very difficult task due to the impossibility to identify whether life style variables affect health outcomes or it is the other way around.

Simultaneity is an old problem in many ambits of economics and, in fact, it has been an issue of research even before the creation of the Cowles research institute to the development of econometrics in 1932, see the canonical example in economics about the identification of demand and supply (Wright, 1928). In this context, it is generally accepted that identification can only be achieved by means of untestable assumptions. Some examples are the use of instrumental variables that are selected under the exclusion restriction and matching or propensity score regressions under the strong ignorability assumption, see Wooldridge (2003) and Imbens and Wooldridge (2009) and references therein for relevant examples of these methodologies. Even in the absence of specific instrumental variables, identification can still be achieved by other means, such as assuming a direction of causality among the endogenous variables (recursiveness assumption) or imposing the effect of some specific shocks to be negligible in the long term (long-run restrictions) among others, see Christiano et al. (1999) and Blanchard and Quah (1989) respectively.

In the recent health and sport literature, it has become fashionable the use of simultaneous equation models to deal with simultaneity in cross-sectional databases. These models typically do not use instruments to identify the direction of causality of simultaneous variables but, instead, this is imposed by the recursiveness assumption. We argue this is a sensible approach only when there are solid theoretical arguments to justify this restriction. However, this is not the case if there is double causality between health behavior and health outcome. We also discuss the case where simultaneity affects more than one response variable. In this case, contrary to the claims of previous papers in the literature, the estimation of a reduced form specification is generally not a valid alternative to deal with the simultaneity problem of response variables regardless of whether a seemingly unrelated regression (SUR henceforth) strategy is used or not to account for the fact that errors in the different equations are potentially correlated.

This brief note does not attempt to discuss the main econometric properties of the estimators. Our main purpose is to discuss the theoretical implications of the use of simultaneous equation models to deal with endogeneity adopted by an important strand of the sport economic literature

General Discussion

A common interest in the empirical literature discussed in this note is the causal impact estimation of health behavior, typically sport or physical activity, on a set of response variables which can include health or well-being measures. This implies to deal with the following two types of identification problems: 1) simultaneity between exercise and response variables; and 2) simultaneity across the different response variables. The identification problem can be similarly defined for both cases in the following way.

Let's define three simultaneous scalar variables $y_{1,i,t}$, $y_{2,i,t}$ and $x_{i,t}$. For our purpose $x_{i,t}$ is assumed to be exogenous. This variable does not play a relevant role for the first discussion about the impact of physical activity on a single response variable but it is defined for convenience in the analysis of the impact of exercise on multiple response variables.

A linear structural system that describes this relationship can be defined as:

$$y_{1,i,t} = a + by_{2,i,t} + \delta x_{i,t} + \beta' z_{i,t} + \mathcal{G}' w_{1,i,t} + \varepsilon_{1,i,t} \quad (1)$$

$$y_{2,i,t} = c + dy_{1,i,t} + \phi x_{i,t} + \gamma' z_{i,t} + \phi' w_{2,i,t} + \varepsilon_{2,i,t} \quad (2)$$

where the vector of variables $w_{1,i,t}$ and $w_{2,i,t}$ are specific to each of the two equations while both equations are affected by a vector of common covariates $z_{i,t}$; and $\varepsilon_{1,i,t}$ and $\varepsilon_{2,i,t}$ are fundamental structural shocks, where $E(\varepsilon_{1,i,t}, \varepsilon_{2,i,t}) = 0$. Variables $z_{i,t}$, $w_{1,i,t}$ and $w_{2,i,t}$ can include exogenous and predetermined variables.

In this framework, identification of parameters in equations (1) and (2) is possible if variables $z_{i,t}$, $w_{1,i,t}$ and $w_{2,i,t}$ can be observed. If this is not the case, equations (1) and (2) are observationally equivalent and identification restrictions about the direction of causality between $y_{1,i,t}$, $y_{2,i,t}$ and $x_{i,t}$ are required to achieve identification.

Impact of Physical Activity on a Single Response Variable

If the estimation problem concerns the impact that physical exercise exerts on a single response variable, which could be denoted by $y_{1,i,t}$ and $y_{2,i,t}$ respectively in expression (1) and (2), the parameter of interest would be d . In this case, identification can still be achieved by finding instruments that only affect one of the two endogenous variables. If this is not possible, we require a restriction assumption on the direction of causality, i.e. setting $b = 0$. For example, Contoyannis and Jones (2004) and Balia and Jones (2008) consider recursive multivariate probit models

where reduced form specifications are considered for health behavior activities while a latent health stock is contemporaneously affected by health behavior. Similarly, Humphreys et al. (2014) also impose unidirectional causality from physical activity to health outcomes. They justify this assumption on theoretical arguments as physical activity is an input in the regression of health. Therefore, in all these papers, the use of a simultaneous system of equations can be useful to deal with the endogeneity problem due to omitted variables that can affect both variables of interest as some variables in the (1)-(2) specifications could be omitted. However, this approach does not solve the identification problem due to the simultaneous observation of physical exercise and health outcome which is achieved by the imposition of the recursiveness assumption.

Impact of Physical Activity on Multiple Response Variables

The simultaneous observation of different response variables creates an additional endogeneity problem as, in most cases, it is not possible to set a direction of causality across them based on economic theory. We consider in this case that physical exercise is denoted by $x_{i,t}$. For clarity of exposition, let's abstract from the endogeneity problem concerning physical exercise, which was previously discussed, and consider that the two response variables health and well-being are now denoted by $y_{1,i,t}$ and $y_{2,i,t}$ respectively in equations (1) and (2).

As in the previous section, the exclusion restriction is a well-known identification condition of the system above. It establishes that identification of our parameters of interest, δ and ϕ , together with all the other parameters in equations (1) and (2) is only possible if variables $w_{1,i,t}$ or $w_{2,i,t}$ can be observed. However, if this is not the case, equations (1) and (2) are observationally equivalent. In empirical applications, it is not always possible to find instruments that only affect one particular response variable but not the other and this necessarily means that a direction of causality must be imposed, i.e. either b or d must be set equal to zero, to achieve identification.

The aforementioned simultaneity problem cannot be solved by a joint estimation of the reduced form version of equations (1) and (2). This issue has been considered in Rasciute and Downward (2010) and references therein to estimate the impact of physical exercise on health and happiness. They claimed to solve the simultaneity problem between health and well-being estimating the following reduced form version of the system described by equations (1) and (2), when variables $w_{1,i,t}$ or $w_{2,i,t}$ cannot be observed, by means of a SUR model:

$$\begin{aligned} y_{1,i,t} &= e + \pi x_{i,t} + \tau' z_{i,t} + v_{1,i,t}, & (3) \\ y_{2,i,t} &= f + \rho x_{i,t} + \lambda' z_{i,t} + v_{2,i,t}. & (4) \end{aligned}$$

However, this estimation does not solve the simultaneity problem between $y_{1,i,t}$ and $y_{2,i,t}$. First, error terms in expressions (3) and (4) $v_{1,i,t} = \frac{\varepsilon_{1,i,t} + b\varepsilon_{2,i,t}}{1-db}$ and $v_{2,i,t} = \frac{d\varepsilon_{1,i,t} + \varepsilon_{2,i,t}}{1-db}$ are correlated, regardless of whether there is not any omitted variable in the model, just if either $b \neq 0$ or $d \neq 0$. More importantly, even if this

correlation is taken into account by means of a SUR model, $\pi = \frac{\delta + b\phi}{1 - bd}$ and $\rho = \frac{d\delta + \phi}{1 - bd}$ are unbiased estimates of the structural parameters of interest δ and ϕ if and only if there is no simultaneity between $y_{1,i,t}$ and $y_{2,i,t}$, i.e. when both $b = 0$ and $d = 0$. Therefore, taking into account the correlation between $v_{1,i,t}$ and $v_{2,i,t}$ can improve estimation efficiency but it will result in a biased causal estimation unless the simultaneity between $y_{1,i,t}$ and $y_{2,i,t}$ is properly addressed.

Literature Discussion and Recommendations

Table 1 shows an overview of the strategies adopted in different papers regarding identification of the causal impact of physical activity on health and well-being. The two approaches considered in this literature to achieve identification are the use of the recursive system of equations discussed in previous section, regarding a single response variable, and the exclusion restriction. The former is based on a recursive system of equations adopted in Humphreys et al. (2014) which does not require instrumental variables but instead, it is based on the hypothesis that exercise causes health but not the other way around. The exclusion restriction is the most common strategy and it is based on the untestable hypothesis that there is an instrumental variable which only indirectly affects the response variable (health and/or happiness) through its effect on the decisions to practice exercise. The most common type of instrument is some sport access indicators such as the presence of sport facilities nearby. This strategy was adopted, for example, in Forrest and McHale (2011), Ruseski et al. (2014), Wicker and Frick (2015, 2016), Downward and Dawson (2016), Brechot et al. (2017), and dos Santos et al. (2019). There are other examples of more original instrumental variables that account for sports supply. Thus, Forrest and McHale (2011) consider parental encouragement to play sport during childhood, Sarma et al. (2015) and Downward and Dawson (2016) temperature and month of the year respectively, Wicker and Frick (2015, 2016) club membership and Ruseski et al. (2014) beliefs about the importance of sport participation.

The different approaches and instruments used in these papers are not aseptic as these methodologies are based on untestable assumptions. Thus, a sensible strategy would be testing the robustness of the results under different instrumental variables and assumptions. For example, Humphreys et al. (2014) consider both a recursive system of equations and the exclusion restriction to estimate the impact of physical exercise on health outcome finding similar results. Others based their identification strategy on the exclusion restriction but considering more than just one instrumental variable, see Forrest and McHale (2011); Ruseski et al. (2014); Pawlowski et al. (2011); Wicker and Frick (2015, 2016) and Downward and Dawson (2016).

As discussed in previous section, regarding multiple response variables, identification becomes even more problematic when it concerns the simultaneity of health and well-being as it is hard to imagine a variable which only affects one of them, but not the other. Even under the absence of an instrument, it is still possible to achieve identification by using a class of triangular simultaneous equation model as

Table 1. Research Papers Studying the Causal Impact of Physical Exercise on Health and/or Well-being.

Paper	Response variable	Identification strategy	Instrument variable(s)
Humphreys et al. (2014)	Different health outcomes and self-reported health	(1) Recursive system of equations (2) Exclusion restriction	Sense of belonging to social community
Sarma et al. (2015)	Different health outcomes	Exclusion restriction	Average temperatures surrounding the respondents' interview month
dos Santos et al. (2019)	Health outcome (BMI)	Exclusion restriction	Presence of a public place to practice physical activity near home
Brechot et al. (2017)	Different health outcomes and subjective health	Exclusion restriction	Number of sports facilities within 6 miles.
Pawlowski et al. (2011)	Happiness	Exclusion restriction	(1) Frequency of attending sports events (2) Frequency of participation in a sports association/group
Wicker and Frick (2015)	Happiness	Exclusion restriction	(1) Sports opportunities (2) Club membership
Wicker and Frick (2016)	Happiness	Exclusion restriction	(3) Time spent sitting (1) Whether sports facilities were available within 20 minutes traveling distance of home
Forrest and McHale (2011)	Happiness	Exclusion restriction	(2) Parental encouragement to play sport during childhood County sports establishment count
Huang and Humphreys (2012)	Happiness	Exclusion restriction	(1) Distance between an individual's home and the nearest sports facility.
Rusecki et al. (2014)	Happiness	Exclusion restriction	(2) Answer to the survey question asking if they believe that individual participation is important.
Downward and Dawson (2016)	Happiness	Exclusion restriction	(1) Sports supply (whether can get to a sports facility within 20 minutes) (2) Month in which the respondent answered the survey

suggested by Klein and Vella (2010). The norm in economics is to study how results depend upon the direction of causality imposed by the recursiveness assumption (Christiano et al., 1999). Thus, in our specific case this would amount to analysing the robustness of the estimation results to changes in the identification assumptions about the causality between health and well-being.

Concluding Remarks

Dealing with simultaneity in cross-sectional databases is a difficult task which requires the use of untestable identification assumptions either in the choice of instruments or the direction of causality. Therefore, when an exclusion restriction is not found, this problem can only be solved by a joint estimation of equations for each of the simultaneously observed variables if we have strong arguments to accept that health behavior affects health outcome but not the other way around. The problem of simultaneity can also regard the estimation of the effect of health behavior on several outcome variables. In this case, contrary to what is claimed by some papers in the literature, if response variables are simultaneously related, in the absence of exclusion restrictions, a SUR reduced form specification produces biased estimation of causal effects. A more sensible approach in this circumstance would be to study the robustness of the results to changes in the direction of causality imposed by the recursiveness assumption.

Acknowledgments

The authors thank two anonymous referees for their helpful comments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

J. D. Tena  <https://orcid.org/0000-0002-0036-9735>

Notes

1. See, for example, Contoyannis and Jones (2004), Balia and Jones (2008) and Humphreys et al. (2014) among others.
2. Just to cite two examples, see Forrest and McHale (2011) and Wicker and Frick (2015).

References

- Balia, S., & Jones, A. M. (2008). Mortality, lifestyle and socio-economic status. *Journal of Health Economics*, 27(1), 1–26.
- Blanchard, O. J., & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *The American Economic Review*, 79(4), 655–673.
- Brechot, M., Nüesch, S., & Franck, E. (2017). Does sports activity improve health? Representative evidence using local density of sports facilities as an instrument. *Applied Economics*, 49(48), 4871–4884.
- Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of macroeconomics*, 1, 65–148.
- Contoyannis, P., & Jones, A. M. (2004). Socio-economic status, health and lifestyle. *Journal of Health Economics*, 23(5), 965–995.
- dos Santos, A. M., Jacinto, P. d. A., & Victor, R. (2019). The Impact of physical activity on womens health: Evidences for Brazil. *Economics Bulletin*, 39(4), 2381–2401.
- Downward, P., & Dawson, P. (2016). Is it pleasure or health from leisure that we benefit from most? An analysis of well-being alternatives and implications for policy. *Social Indicators Research*, 126(1), 443–465.
- Forrest, D., & McHale, I. (2011). *Subjective well-being and engagement in sport: evidence from England. The economics of sport, health and happiness: The promotion of well-being through sporting activities* (pp. 184–199). Edward Elgar Publishing.
- Huang, H., & Humphreys, B. R. (2012). Sports participation and happiness: Evidence from US microdata. *Journal of Economic Psychology*, 33(4), 776–793.
- Humphreys, B. R., McLeod, L., & Ruseski, J. E. (2014). Physical activity and health outcomes: Evidence from Canada. *Health Economics*, 23(1), 33–54.
- Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5–86.
- Klein, R., & Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, 154(2), 154–164.
- Pawlowski, T., Downward, P., & Rasciute, S. (2011). Subjective well-being in European countries—On the age-specific impact of physical activity. *European Review of Aging and Physical Activity*, 8(2), 93–102.
- Rasciute, S., & Downward, P. (2010). Health or happiness? What is the impact of physical activity on the individual? *Kyklos*, 63(2), 256–270.
- Ruseski, J. E., Humphreys, B. R., Hallman, K., Wicker, P., & Breuer, C. (2014). Sport participation and subjective well-being: Instrumental variable results from German survey data. *Journal of Physical Activity & Health*, 11(2), 396–403.
- Sarma, S., Devlin, R. A., Gilliland, J., Campbell, M. K., & Zaric, G. S. (2015). The Effect of leisure-time physical activity on obesity, diabetes, high BP and heart disease among Canadians: Evidence from 2000/2001 to 2005/2006. *Health Economics*, 24(12), 1531–1547.
- Wicker, P., & Frick, B. (2015). The relationship between intensity and duration of physical activity and subjective well-being. *The European Journal of Public Health*, 25(5), 868–872.

- Wicker, P., & Frick, B. (2016). Intensity of physical activity and subjective well-being: An empirical analysis of the WHO recommendations. *Journal of Public Health, 39*(2), e19–e26.
- Wooldridge, J. M. (2003). Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics letters, 79*(2), 185–191.
- Wright, P. G. (1928). *Tariff on animal and vegetable oils*. Macmillan Company.

Author Biographies

Jing Guan is a lecturer at Beijing Technology and Business University. Her research interests include insurance economics and sports economics.

J. D. Tena (PhD, Economics, University of Newcastle-upon-Tyne) is a senior lecturer at University of Liverpool, UK, and a researcher at Università di Sassari and CRENoS, Italy. His research focuses on applied econometrics, with an emphasis on sports economics, forecasting and operational research.