Causes of haze and its health effects in Singapore; a replication study

Kiviet, Jan

University of Amsterdam

11 November 2019

Causes of haze and its health effects in Singapore; a replication study

Jan F. Kiviet*

Version of 11 November 2019

JEL-Classifications: C12, C13, C26, I1, Q53.

Keywords: *endogeneity robust inference, environmental economics, health economics, instrument invalidity, sensitivity analysis.*

Abstract

Intermittently Singapore suffers from severe air pollution in periods of intense forest and peatland fires on neighboring South-Asian islands. A recent *American Economic Review* article modeled the causal relationships between fire intensity in Indonesia and air pollution (PSI) in Singapore, and between PSI and health clinic visits in Singapore. We find serious flaws in the quantitative assessment of these relationships. Attempts are made to repair these using the same classic methodology and data, but also by alternative methods requiring less speculative assumptions. Although actually more detailed data are required, also some results are produced which seem more credible.

1. Introduction

Over the past three decades Southeast Asia has suffered several times from severe smoky air pollution during periods of a few days or sometimes of several weeks. This particular type of aerosol is addressed as haze. It is caused by large scale burning of forests and carbon-rich peatlands, especially on the Indonesian islands Sumatra and Borneo, mainly to clear land for palm oil and pulpwood production. Such fires and their ill effects on neighboring countries seem worse during draught periods, which have become a more frequent phenomenon under the influence of climate change. Although all countries united in the Association of Southeast Asian Nations (ASEAN) agreed in 1997 on a Haze Action Plan, and signed in 2002 the Agreement on Transboudary Pollution, periods of serious haze still re-occur, most recently in Singapore during September 2015

*Professor of Econometrics, Amsterdam School of Economics, University of Amsterdam, PO Box 15867, 1001 NJ Amsterdam, The Netherlands (j.f.kiviet@uva.nl).

and September 2019. This phenomenon and the immense problems it generates calls for scientific investigation by many disciplines, such as: ecology, forestry, political sciences, economics, meteorology, toxicology, epidemiology, health sciences, cultural history, et cetera. Many aspects of its origins and scope, though, seem rather well understood already, and its occurrence and the assessment of many of its consequences have been closely monitored in many dimensions over many years. Also strategies to overcome or avoid them have been laid out decades ago already (see, for instance, Quah 2002). Nevertheless, successful actions (regulatory, legal, political, diplomatic, financial) to actually stop this from happening have not yet materialized, despite the intermittently ongoing immense ecological and health damage on top of serious loss of economic productivity. For Singapore alone, the costs of the 2015 haze period have been estimated to amount to US$1.33 billion (see Quah and Chia, 2019). To obtain this estimate they used the special tool of economic investigation called cost-benefit analysis. This aims to express in money terms all losses; not just those in productivity, such as in the tourist industry, but also those in health and in happiness.

Here we want to illustrate in what way another technique from the toolbox of economics can contribute to a better understanding of particular aspects of the haze problem. Econometric techniques can be used to quantify relationships between phenomena like the number of active bush fires in Indonesia, subsequent pollution index readings in Singapore, and resulting extra health clinic visits. The website of the Meteorological Service Singapore provides daily Regional Haze Maps, indicating the areas in Southeast Asia where the air pollution level is high in combination with the locations of hotspots. The latter represent observations by satellites of active fires. No specialized statistical training is required to be assured, after taking note of a series of such maps, that haze is directly caused by these fires. In addition, everybody who has ever actually experienced serious haze for several days, will have little doubts that haze may cause an upsurge of visits of patients to health clinics diagnosed with respiratory and eye ailments. However, as we shall see, it is quite difficult, and in fact seriously problematic, to accurately assess what pollution levels one may expect in Singapore, given the number and location of hotspots observed by satellite. Likewise, it is not easy to quantify how the number of patients diagnosed with particular haze-related ailments varies with the level of the pollution index.

It are precisely these relationships that have been investigated by Sheldon and Sankaran (2017), who (addressed as S&S below) measure the impact of Indonesian

forest fires on air pollution in Singapore, and also of this pollution on some Singaporean health indicators. For their quantitative causal analysis they use popular classic econometric methods, known under the name of Two-Stage Least-Squares (TSLS), which found their place in the standard econometrics toolbox shortly after the Second World War. In Section 2 we review the major aspects of this technique against the background of general methodological problems for this type of empirical statistical causal research, here applied to environmental and health issues. Next, in Section 3, we first critically discuss the way in which S&S chose to use these methods. Then, in a first subsection, we confront their results with more general implementations of the same TSLS technique while using the very same data. These strongly refute the S&S results. However, not for all four of the major relationships examined by S&S we can find fully satisfactory TSLS results ourselves. In part this seems due to limitations of the data set. Next, in subsection 3.2, we reconsider one of the number of health clinic visit relationships, but now applying a very recently developed alternative econometric technique, called KLS, which avoids making particular untested suppositions underlying TSLS. This all leads in the final Section 4 to some practical conclusions as well as to some suggestions for further research.[1]

2. Causal analysis complicated by endogenous regressors

Regression analysis forms the work-horse of econometricians; it is at the basis of most econometric techniques. At one extreme these may concern statistical forecasting techniques, which often use little or no economic or subject matter knowledge, but mainly try to extrapolate in a rather mechanical way regularities within and between variables as observed over the recent past. And at the other extreme, this may involve structural causal modeling, which aims to reveal from empirical observations the essential numerical parameters of complex economic (or more general natural scientific) relationships. In its nature the latter approach is often at least partly inspired by similarities with the decryption of the fundamental laws of physics.

Extra complications met in empirical structural causal research on interesting practical (social or natural lifelike) phenomena are that observations on these do as a rule not result from controlled laboratory experiments, but must directly be obtained from

[1] Shortly after we had started this replication study and had collected the critique on the S&S study as reported in the initial pages of Section 3 below, we contacted in May 2019 the authors and invited them to debate their approach. Also after a reminder a month later, no response was received.

what has actually happened in the complex world around us. This means that many disturbing incidental factors may seriously blur the essentials of the relationships under study. Moreover, mostly it are not just the parameter values that are yet to be disclosed. Also the set of determining factors as such of the phenomena under study, as well as the actual functional form of the relevant relationships, are as a rule yet unknown. Thus, these have to be modeled, together with the assessment of the relevant parameter values, in a recursive process of trial and error. In various rounds one specifies a set of assumptions, next exploits these to obtain empirical estimates, followed by confronting the obtained results with the assumptions made (misspecification testing), and in case of incongruities one should reformulate the assumptions to enter a next round. Especially when the available data are characterized by severe limitations, or the employed assumptions are too far from reality, there is no guarantee that this process will ever converge, and if it converges, whether it yields truthful outcomes. However, by being explicit about the road followed, one may hopefully challenge and facilitate future next attempts, possibly using better data sets, more realistic model assumptions, and perhaps also better newly developed statistical techniques.

We sketch here the characteristics of the type of model used in this and very many other empirical causal studies, focussing on the major complications of the prevailing technique used for its analysis. Let y, the particular variable of our primary interest (called the regressand), be causally determined by the observed linearly independent but usually correlated variables $x_1, ..., x_K$ (called the regressors), and by some other unobserved factors collected in component ε (the disturbance), according to the linear relationship

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K + \varepsilon. \tag{2.1}$$

Because nonlinear relationships between variables can often be modeled as linear in the unknown parameters after a nonlinear transformation of the observed variables, (2.1) also characterizes many situations in which the observed variables are nonlinearly causally related. The actual values of the unknown parameters $\beta_0, ..., \beta_K$ can rather easily be estimated if: (i) we have a sufficiently large sample of cross-section or time-series observations on these $K+1$ variables, to be denoted by $\{y_i, x_{i1}, ..., x_{iK}; i = 1, ..., n\}$; and (ii) it is reasonable to assume that the unobserved component ε, although affecting y, is not systematically related with the variables $x_1, ..., x_K$. Then we call the regressors exogenous with respect to ε in model (2.1). Such conditions can often easily be created for laboratory experiments, but adopting them when observational data are analyzed is

often doubtful.

To make assumption (ii) slightly more realistic, imagine that amongst the factors in ε there may in fact be some which show linear coherence with one or more of the regressors. Focussing on the simplest possible case, we could have

$$\varepsilon = \gamma_0 + \gamma_1 x_1 + \varepsilon^*, \tag{2.2}$$

with $\gamma_1 \neq 0$, while ε^* is now assumed to be not systematically related with $x_1, ..., x_K$. Then x_1 is no longer exogenous but endogenous with respect to ε in (2.1). However, substituting (2.2) in (2.1) makes clear that $x_1, ..., x_K$ are still all exogenous with respect to ε^*, so using the technique to analyze (2.1) which presupposes all regressors to be exogenous will actually yield estimates of the parameters $\beta_0^*, \beta_1^*, \beta_2, ..., \beta_K$, where

$$\beta_j^* = \beta_j + \gamma_j \text{ for } j = 0, 1. \tag{2.3}$$

If one wants to use the model for forecasting, this "automatic" correction for neglected systematics seems very welcome. However, when (2.1) has to serve a causal analysis then the purpose may be to identify β_1 instead of β_1^*.

Identifying β_1 would be easiest if one could augment the model with the explanatory variables that are correlated with already included variables. In case these variables are not fully known or simply not available identification of β_1 could also be achieved as follows. Suppose that it is possible to model variable x_1 according to

$$x_1 = \delta_0 + \delta_2 x_2 + ... + \delta_K x_K + \delta_{K+1} z_1 + ... + \delta_{K+L_1} z_{L_1} + \upsilon, \tag{2.4}$$

where $L_1 \geq 1$ and the variables $x_2, ..., x_K, z_1, ..., z_{L_1}$, which should be mutually linearly independent, are all unrelated with υ, while $z_1, ..., z_{L_1}$ are not systematically related with ε either. The latter means that $z_1, ..., z_{L_1}$ are validly excluded from (2.1). Thus, it is just the component υ from x_1 which is related with ε. Although the variables $z_1, ..., z_{L_1}$ do not have a direct effect on y, those for which $\delta_j \neq 0$ $(j = K+1, ..., K+L_1)$ do have an indirect effect on y via x_1.

Now defining $x_1^* = x_1 - \upsilon$ we can rewrite (2.1) as

$$y = \beta_0 + \beta_1 x_1^* + ... + \beta_K x_K + \eta, \text{ with } \eta = \varepsilon + \beta_1 \upsilon. \tag{2.5}$$

Since the variables $x_1^*, x_2, ..., x_K$ are all unrelated with η estimation of their coefficients, including β_1, would be viable if x_1^* could be observed. The coefficients of (2.4) being unknown, this is not the case. However, we can estimate (2.4) and calculate

$$\hat{x}_1 = \hat{\delta}_0 + \hat{\delta}_2 x_2 + ... + \hat{\delta}_K x_K + \hat{\delta}_{K+1} z_1 + ... + \hat{\delta}_{K+L_1} z_{L_1}$$

and next estimate

$$y = \beta_0 + \beta_1 \hat{x}_1 + ... + \beta_K x_K + \eta^*, \text{ where } \eta^* = \varepsilon + \beta_1 (x_1 - \hat{x}_1) \tag{2.6}$$

The approximation of x_1^* in (2.5) by $\hat{x}_1$ in (2.6) works only well when the sample is really large, and therefore the $\hat{\delta}_j$ get reasonably close to the δ_j for $j = 1, ..., K + L_1$, yielding $x_1 - \hat{x}_1$ to be close to υ.

The estimation of (2.4) to obtain $\hat{x}_1$ is called the first-stage regression, and that of (2.6) the second-stage regression. The technique applying both is therefore called two-stage least-squares (TSLS). The variables $z_1, ..., z_{L_1}$ are called the external instruments, whereas the exogenous regressors of (2.1), including the intercept, are called the internal instruments. So, in the special case where the right-hand side of (2.2) contains only one of the regressors, the total number of instruments is $L = L_1 + K$. If $L_1 = 1$ then β_1 is just-identified and over-identified when $L_1 > 1$. Only for $L_1 - 1$ of the variables $z_1, ..., z_{K_1}$ it can be tested statistically whether they are validly excluded from (2.1), assuming that the remaining one has been validly excluded. Therefore, one of the handicaps of TSLS is that it is based on an assumption which simply has to be adopted on other than statistical grounds. Another handicap can be that the joint contribution of the variables $z_1, ..., z_{L_1}$, to the description of x_1, additional to that of $x_2, ..., x_K$, may be so limited, that $\hat{x}_1$ is almost linearly dependent with $x_2, ..., x_K$, which leads to poor precision in the estimation of (2.6). In that case we call the external instruments weak.

Apart from not explicitly incorporating causal variables in a model like (2.1), which gives rise to endogeneity of regressors which are correlated with the excluded regressors, there are two other situations which will give rise to endogeneity of regressors, namely using proxies and simultaneity. When (2.1) has truly exogenous regressors, but x_1 has not been observed and is therefore replaced by a proxy variable $x_1^\dagger$, where $x_1^\dagger = x_1 + u$, then we obtain

$$y = \beta_0 + \beta_1 x_1^\dagger + ... + \beta_K x_K + \eta, \text{ with } \eta = \varepsilon - \beta_1 u. \tag{2.7}$$

In this equation $x_1^\dagger$ is endogenous with respect to η, because $x_1^\dagger$ and η are both determined by u, if $\beta_1 \neq 0$. Variable $x_1^\dagger$ is the only endogenous regressor, provided the proxy-error u is unrelated with $x_2, ..., x_K$.

In case of simultaneity, next to the causal model (2.1) for y, it is the case that the causal model for, for instance, x_1 is given by

$$x_1 = \alpha_0 + \alpha_1 y + \alpha_2 x_2 + ... + \eta, \tag{2.8}$$

where we leave most of its explanatory variables implicit. Hence, variables y and x_1 are characterized by reciprocal causality, expressed through simultaneous equations. Substitution of (2.1) in (2.8) yields

$$x_1 = \alpha_0 + \alpha_1\beta_0 + (\alpha_2 + \alpha_1\beta_1)x_2 + ... + \phi, \text{ with } \phi = \eta + \alpha_1\varepsilon, \tag{2.9}$$

from which it is obvious that x_1 is again related with ε, provided $\alpha_1 \neq 0$.

The above can be generalized such that any explanatory variable in a causal model, which is correlated with wrongly excluded causal regressors, or is correlated with proxy-errors of one of the included explanatories, or is in its turn causally dependent on y, is endogenous. If this is the case for $K_1 \leq K$ of the explanatory variables, then at least K_1 external instruments are required to identify all the K coefficients of this relationship. From $L_1 \geq K_1$ external instruments, the validity of only $L_1 - K_1$ can be tested statistically, inevitably adopting the validity of K_1 instruments. Using the methodology sketched above is often addressed as a quasi-experimental technique. A strong case for following quasi-experimental approaches in environmental health studies is expressed by Dominici, Greenstone and Sunstein (2014).

The complication of explanatory variables to be endogenous rather than exogenous due to omitted explanatories or using proxies is also often addressed as confounding; see, for instance, Imbens (2014). A recent historical overview on approaches tackling confounding in causal environmental health studies is Bind (2019). That using instrumental variables may lead to poor inference when instruments are either invalid or weak has triggered a lot of research in econometric theory over the last decades, see for instance the overviews in Murray (2006) and Andrews, Stock and Sun (2019). This has brought techniques to disclose weakness of instruments and to adapt confidence regions such that they can still claim proper coverage probability when instruments are weak. However, when instruments are weak such confidence regions are very wide. Moreover, as explained above, rigorous statistical evidence on instrument validity is illusionary. Therefore, recently attempts have been made to develop a technique which avoids using instruments, see Kiviet (2019b).

3. Haze and health in Singapore

S&S estimated equations where either the pollution level PSI (Pollution Standards Index) is the regressand, or one of three different health indicators $H^{(j)}(j = 1, 2, 3)$, being polyclinic attendances for: acute upper respiratory tract infections ($j = 1$), acute

conjunctivitis i.e. inflammation of the upper tissue of the eye ($j = 2$), and for chickenpox ($j = 3$). They employed weekly data, from January 2010 until June 2016. Most of the variables that they used have been constructed by averaging daily data, sometimes measured just at a particular hour, or by aggregating over days and/or different locations. These data and a Stata-do-file can be downloaded from the *AER*-website.

By Ordinary Least-Squares (OLS) S&S estimate the first-stage equation

$$PSI_t = \alpha_1 FRP_t + \alpha_2 FRP_t \times SP_t + \alpha_3 PSIch_t + \alpha_4 D_t + \alpha_5' W_t + \alpha_6' F_t + v_t, \qquad (3.1)$$

and next used this to obtain Two-Stage Least-Squares (TSLS) estimates for the three second-stage equations ($j = 1, 2, 3$)

$$H_t^{(j)} = \beta_1^{(j)} \widehat{PSI_t} + \beta_2^{(j)} D_t + \beta_3^{(j)\prime} W_t + \beta_4^{(j)\prime} F_t + \varepsilon_t^{(j)}. \qquad (3.2)$$

So PSI (Pollution Standards Index) is supposed to be the one and only endogenous regressor in (3.2). For a precise definition of all their variables we refer to S&S. FRP denotes Fire Radiative Power in Indonesia measured by satellite. For the other variables we have compactified their notation: Here SP denotes *wind* (speed) in Singapore, $PSIch$ is the 0-1 dummy variable *PSIchange* (to account for the incorporation of smaller particulate matter, PM2.5, in the definition of *PSI* since April 2014), D denotes *diarrhea*, W is a vector of 6 *weather* variables, including rainfall, temperature (mean, maximum and minimum) and wind speed (mean and maximum) all measured near the center of Singapore, and F is a vector containing an intercept, 11 month dummies and 6 year dummies (just 4 year dummies when $j = 3$,, because of fewer available observations). F controls for seasonal patterns and for specific fixed year effects. Like S&S we consider rescaled variables for PSI, FRP and all $H^{(j)}$ such that their sample standard deviation is 1. Note that we use a different notation for the coefficients and the disturbance terms of both equations.

Equation (1) in S&S incorrectly misses regressor D, which controls for general health trends in the explanations of variables $H^{(j)}$ for $j = 1, 2, 3$. It should be included in the first-stage equation (3.1) simply because it is an exogenous regressor in (3.2), like the variables in vectors W and F are. We found out, though, that S&S too must in fact have included D correctly in the set of instruments in all regression results reported in their article. However, they do not when estimating (1) and the equation for $H^{(1)}$ in the Stata-do-file that they provide. S&S analyze the relationship for $H^{(3)}$ primarily as a check on whether $\beta_1^{(3)}$, unlike $\beta^{(1)}$ and $\beta^{(2)}$, will be insignificant, because haze is not supposed to have an effect on the prevalence of chickenpox.

In fact S&S do not use (3.1) just simply as a first-stage regression to obtain fitted values $\widehat{PSI}$ on the basis of all available instrumental variables: (i) the exogenous regressors D, W and F included in (3.2), and (ii) some further explanatories of PSI excluded (hopefully correctly) from (3.2), being FRP, $FRP \times SP$ and $FRPch$. They also interpret (3.1) as a causal regression model for PSI. And since they estimated it by OLS they apparently assume that (3.1) is specified such that none of its regressors is endogenous. Would that really be the case for FRP? That PSI depends on FRP seems evident. However, could at the same time FRP in Indonesia depend on PSI in Singapore? This is, of course, what Singapore should strive for! It would be reality when, as soon as PSI reaches critical levels, a message from Singapore to Indonesia would suffice to have sufficient fires extinguished by Indonesia in order to lower PSI in Singapore. Obviously, such reciprocal causality does not seem manifest yet. However, it does seem clear, that the FRP variable as constructed by S&S is just a proxy of the radiative power really relevant for haze in Singapore. S&S took for FRP radiative power integrated over all Indonesian latitudes and longitudes. Obviously, though, all these individual Indonesian hotspots should have been weighted in some way or another by factors including their distance to Singapore, and the actual wind speed and wind direction between their location and Singapore. This not being the case, with SP just referring to the wind speed in Singapore and W just characterizing the weather in Singapore, it seems most likely that equation (3.1) does suffer from a combination of measurement errors in FRP and omitted relevant regressors. Then (3.1) is not a proper reduced form equation and its OLS estimates will be inconsistent. Moreover, its substitution into (3.2) will not yield proper reduced form equations for $H^{(j)}$ either. This casts doubts on the first causal empirical findings reported in S&S. These are based on OLS estimates of these putative reduced form equations for $H^{(j)}$, stating that one standard deviation increase in FRP leads to a 0.7 standard deviation increase in both $H^{(1)}$ and $H^{(2)}$.

That (3.1) does not seem suitable to provide a serious causal model for PSI appears also from the following. For good reasons D may be assumed exogenous in (3.2). That makes it automatically one of the explanatories in the reduced form equation for PSI. However, when one wants to specify a causal model for PSI, one would never think of including D as one of its regressors. Otherwise one would put forward that in Singapore the number of diarrhea patients is one of the determinants of its air pollution levels, which seems hilarious. Nevertheless, the coefficient of D has p-values 0.19 and 0.06 in the samples of 325 and 220 observations respectively. Especially the latter p-value seems

to indicate that equation (3.1) does not really include all its direct genuine causes yet, giving rise to estimation bias.

An obvious candidate to improve the fit of equation (3.1) seems the interaction variable $FRP \times PSIch$. The inclusion of just $PSIch$ allows for a shift in the intercept of the equation due to the change in definition of PSI. It does not seem unlikely that also the slope effect of FRP on PSI will be affected after PSI has been redefined and includes very fine particulate matter as well. That would also suggest to use $FRP \times PSIch$ as an extra external instrument for equation (3.2). Equation (2) in S&S is invariant with respect to the change in definition of PSI, whereas it is said that very fine particulate matter has special negative effects on health. Therefore, it does not seem unlikely that $PSIch$ and $PSI \times PSIch$ should both be included as regressors in (3.2). Then $PSIch$ would no longer be validly excluded from (3.2) and ceases to be an external instrument, whereas $PSI \times PSIch$ would establish an extra endogenous regressor.

Yet another issue regarding the specifications of (3.1) and (3.2) is their static nature. All effects are modeled as instantaneous, whereas one would expect that it takes time before the exhaust of fires in Indonesia affects pollution levels in Singapore. Likewise, it may take time before hazardous pollution levels have actually caused ailments which force patients to visit clinics. Delays to visit clinics seem also likely because the Singaporean authorities advise its citizens, especially the more vulnerable, to stay indoors as much as possible as long as pollution levels are high. Of course, the aggregation of the data to weekly figures may accommodate these delay-issues, at least partly. However, it seems worthwhile to examine the significance of lagged regressors when added to both model specifications (3.1) and (3.2), also in the light of the residual serial correlation reported by S&S. Note that the Newey-West robustification by S&S of standard errors does not cure any omissions of relevant lagged regressors, because such omissions lead to inconsistency of the coefficient estimates, given their likely correlation with included regressors.

From the above we conclude that equation (1) in S&S is not a satisfactory structural form equation, and is no reduced form equation either, due to omitted and/or proxied explanatories. To repair it, the collection of more data seems unavoidable, at least on weather (rain, wind direction, wind speed) in areas at Indonesian hotspots and at intermediate locations by which the haze reaches Singapore. Collecting these data is beyond the purpose of this replication study.

Although from the current data set no reduced form equation for the pollution level

can be estimated, and therefore not for the health indicators either, this does not preclude the analysis of structural causal models for the health indicators. For their estimation by TSLS a completely satisfying specification of (3.2) is required, but not for (3.1); just a sufficient number of sufficiently strong external valid instruments will suffice. So, below we shall focus on analyzing whether equation (2) of S&S is a satisfying structural model for the health variables.

The endogeneity of PSI in (3.2) could be due to a combination of the consequences of simultaneity, using proxy variables, and confounders (wrongly omitted relevant explanatories). We discuss each in turn. There would be reciprocal causality, meaning that next to health indicators being caused by PSI, the health variables have an effect on PSI as well, when the Singaporean authorities, as soon as it notices critical health conditions, could successfully convince (or force) local industry and motorists to reduce exhaust and Indonesia to combat the fires. This may be the case to a (very) limited degree, and therefore it seems wise to allow for endogeneity of PSI. Measurement or proxy errors (called sorting in the environmental literature) in PSI seems less likely. PSI is obtained by averaging the pollution measurements at five locations spread over the relatively small Singapore island, and the correlation between these measurements is extremely high, as S&S report. However, PSI may also be endogenous due to omitted relevant explanatories. $PSIch$, which is correlated with PSI, may play a direct role in the proper causal specification of (3.2), and possibly too in interaction terms. The 23 exogenous controls in (3.2) are supposed to suppress any further problems regarding omitted regressors, though the investigation of possibly wrongly excluded lagged variables is certainly called for.

Using the same data and methodology as S&S we shall examine in Section 3.1 whether their equation (2) seems to suffer from using invalid instruments and wrongly omitted regressors. Next, in Section 3.2, we will confront these findings with a novel and yet nonstandard approach in which the use of instruments is avoided completely.

3.1. Re-analysis of the health equations using TSLS

We shall first just focus on the causal effects of PSI on health indicator $H^{(1)}$. Table 3.1.1 presents heteroskedasticity robust TSLS estimation and test results which have all been obtained by the Stata command ivreg2 and various of its available options, see Baum, Schaffer and Stillman (2007). The table does include just few of the regression coefficients; not presented are those associated with the vector of weather variables W,

and the month and year dummies with the intercept all collected in vector F. In the table N denotes the sample size, K the number of regressors, L the number of instruments, K_1 the number of endogenous regressors, L_1 the number of external instruments, and $\hat{\sigma}_\varepsilon$ is the estimate of the disturbance standard deviation.

Table 3.1.1	TSLS results (heteroskedasticity robust) for $H^{(1)}$				
	(A)	(B)	(C)	(D)	(E)
Major regressors:					
PSI	.351 (.070)	.389 (.067)	1.090 (1.310)	.333 (.080)	.462 (.073)
D	.010 (.001)	.010 (.001)	.011 (.001)	.011 (.001)	.011 (.001)
$PSIch$	-	-.489 (.178)	1.354 (3.118)	-.139 (.169)	-.168 (.156)
$PSI \times PSIch$	-	-	-.836 (1.452)	-	-
$L.H^{(1)}$	-	-	-	.602 (.047)	.531 (.064)
$L.PSI$	-	-	-	-.089 (.071)	.027 (.096)
$L.D$	-	-	-	-.006 (.001)	-.006 (.001)
$L2.H^{(1)}$	-	-	-	-	.107 (.067)
$L2.PSI$	-	-	-	-	-.256 (.057)
$L2.D$	-	-	-	-	.000 (.001)
External instr.:					
FRP	✓	✓	✓	✓	✓
$FRPsp$	✓	✓	✓	✓	✓
$PSIch$	✓				
$L.FRP$				✓	✓
$L.FRPsp$				✓	✓
$L2.FRP$					✓
$L2.FRPsp$					✓
Tests (p-value):					
SC1	.00	.00	.00	.04	.97
SC4	.00	.00	.00	.12	.32
Heterosk.	.01	.01	.44	.19	.51
Reset	.02	.02	.96	.00	.00
Over-id. rstr.	.03	.39	-	.48	.44
Exogeneity	.61	.87	.66	.63	.23
Weak Ident. (F)	57.6	74.7	.30	20.8	15.0
K_1; K; L_1	1; 26; 3	1; 27; 2	2; 28; 2	1; 36; 4	1; 45; 6
L; N	28; 325	28; 325	28; 325	39; 324	50; 323
$\hat{\sigma}_\varepsilon$	.606	.602	.670	.435	.415

Column (A) of Table 3.1.1 conforms to the specification preferred by S&S, who conclude that (with a high level of significance) $H^{(1)}$ increases immediately by 0.35 standard errors when PSI increases by one standard error. They claim support for this conclusion because the three external instruments are found to be reasonably strong,

which we endorse on the basis of the Kleibergen-Paap F-value, which is appropriate for models based on time-series. And because they report a Sargan test for the two overidentifying restrictions with a p-value of 0.07. Apparently its small value does not worry S&S. For overidentifying restrictions we calculated the heteroskedasticity robust Hansen J test and find a really worrying p-value of 0.03. We also examined tests for serial correlation of first (SC1) and fourth (SC4) order and a Reset test on functional form (the one degree of freedom variant which adds squared optimal forecasts of the dependent variable to the model). They all strongly reject the specification chosen by S&S. Given these problems, interpretation of the heteroskedasticity test (the variant which regresses the squared residuals on all instruments and their squares) is premature, but its low p-value does encourage to continue using heteroskedasticity robust standard errors. Also, for this clearly misspecified model, the high p-value of the exogeneity test (which is based on the difference between two J tests) provides no license to declare PSI to be exogenous. Next we set out on attempts to improve the structural model for $H^{(1)}$ and obtaining a proper set of instrumental variables.

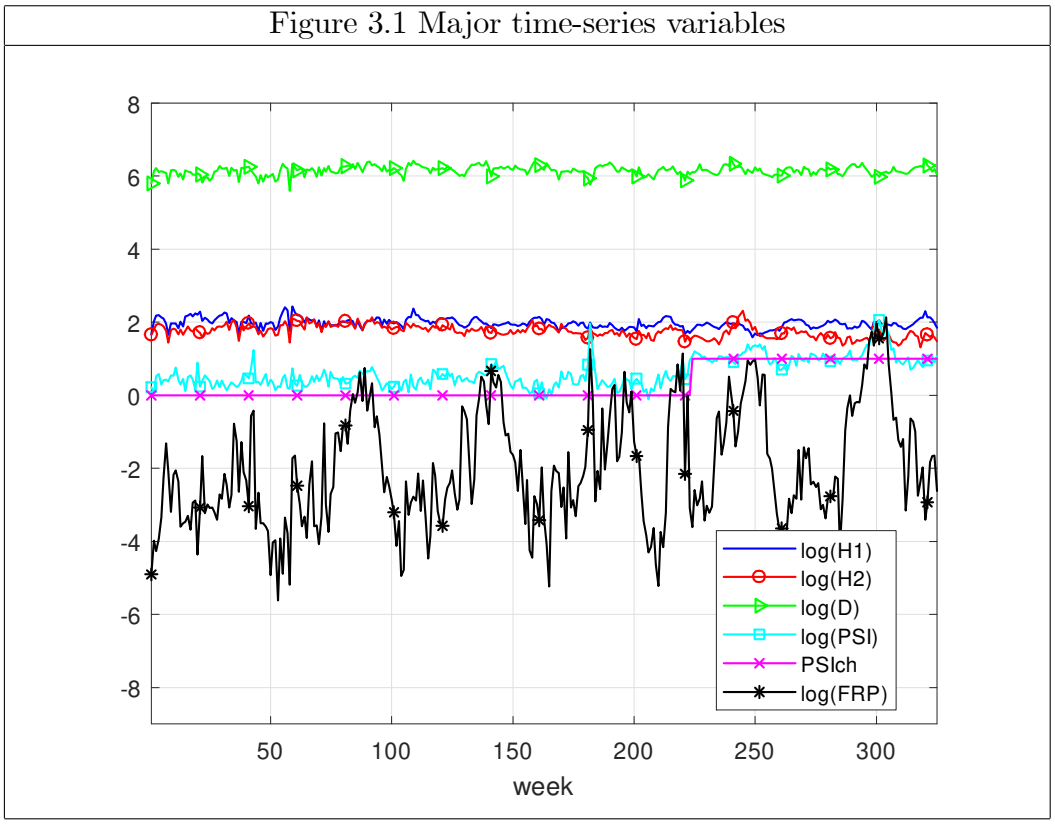
First we examine the doubtful validity of $PSIch$ as an external instrument. In column (B) we included it as an exogenous regressor and it proves to be significant, so in (A) it seems wrongly excluded. Although now using $PSIch$ in (B) as an internal instrument improves the test of the remaining one overidentifying restriction, this augmented specification is still strongly rejected by the serial correlation and reset tests. Therefore, in column (C), we examine whether the change in definition of PSI also affects its slope parameter, by including the endogenous interaction term $PSI \times PSIch$. This model is just identified, having as many endogenous regressors as external instruments, but proves to be actually very weakly identified given the small Kleibergen-Paap F value. This leads to large standard errors of the coefficients for all three regressors associated with PSI and to an increased estimate of $\hat{\sigma}_\varepsilon$.

In (D) we have removed regressor $PSI \times PSIch$ again, but aim to get rid of the serial correlation by including in the regression the first-order lags of particular explanatories. For obvious reasons we did not include lags of the regressors in the vector F, because that would lead to extreme multicollinearity. We also did not include the lag of dummy $PSIch$; its effect would involve the removal from the sample of the first week in which the definition changed, which causes difficulties for Stata to calculate and interpret some of the test statistics. At the same time we added the first lags of the external instruments to the set of instruments. Although the serial correlation tests are less critical now, the

Reset test is still highly unsatisfactory. Therefore, in (E) we included also the second-order lags, and now it is just the Reset test which is still rejecting the chosen linear functional form.

It is certainly not self-evident that the relationship between PSI and $H^{(j)}$ should (in the long-run) be linear. For instance, when PSI (before its standardization, which involved division by 18.9) increases by 10 to enter the unhealthy level (which is reached at 150), the increasing effect on $H^{(j)}$ could well be larger than when it increases by 10 while still in the good air quality zone (below 50). Hence (see also Dominici et al, 2014, p.259, first column, bottom line), there is good reason to embark on a functional specification where the derivative of $H^{(j)}$ with respect to PSI is not constant, as in S&S, but increasing with PSI. This would be enabled by including the square of PSI as a regressor. Because this would bring in a second endogenous regressor, it seems more attractive to have this derivative increase with either $H^{(j)}$ or with $H^{(j)}/PSI$. This will occur in a log-linear or log-log specification respectively.

Figure 3.1 Major time-series variables



Before we continue applying robust TSLS, we first inspect our major data series visually, as depicted in Figure 3.1. We see that PSI simply moved to a higher level since the definition change. We also note (from their logs) that the $H^{(j)}$ series show

some variation through time, but –remarkably– any association with the much more pronounced volatility of PSI and FRP is not evident from the depicted series.

Table 3.1.2	TSLS results (heteroskedasticity robust) for $lH1$, $H2$ and $lH2$			
	$lH1$	$H2$	$lH2$	$lH2$
Major regressors:				
$L.lHj$	.540 (.062)	-	.514 (.064)	.529 (.064)
$L2.lHj$	.100 (.059)	-	.149 (.056)	.123 (.058)
$L.Hj$	-	.496 (.076)	-	-
$L2.Hj$	-	.169 (.055)	-	-
PSI	.065 (.010)	.407 (.086)	.073 (.015)	-
$L.PSI$	.001 (.013)	-.056 (.079)	-.011 (.014)	-
$L2.PSI$	-.034 (.008)	-.232 (.050)	-.041 (.009)	-
$lPSI$	-	-	-	.130 (.116)
$L.lPSI$	-	-	-	.004 (.054)
$L2.lPSI$	-	-	-	-.079 (.028)
D	.0016 (.0001)	.0080 (.0008)	.0014 (.0001)	-
$L.D$	-.0008 (.0001)	-.0044 (.0010)	-.0008 (.0002)	-
$L2.D$	.0001 (.0001)	-.0009 (.0007)	-.0001 (.0001)	-
lD	-	-	-	.589 (.064)
$L.lD$	-	-	-	-.372 (.069)
$L2.lD$	-	-	-	-.028 (.055)
$PSIch$	-.025 (.021)	.129 (.184)	.019 (.031)	.017 (.038)
Tests (p-value):				
SC1	.87	.01	.00	.04
SC4	.20	.03	.02	.07
Heterosk.	.43	.29	.78	.84
Reset	.13	.14	.52	.36
Over-id. rstr.	.41	.01	.01	.05
Exogeneity	.10	.20	.20	.66
Weak Ident. (F)	15.0	14.6	14.6	2.45
K_1; K; L_1	1; 45; 6	1; 45; 6	1; 45; 6	1; 45; 6
L; N	50; 323	50; 323	50; 323	50; 323
$\hat{\sigma}_\varepsilon$	.055	.502	.083	.084

After some experimentation we chose for the log-linear specification for the dependent variable $H^{(1)}$, which implies the form of nonlinearity between the variables where the fixed regression coefficients $\beta_k^{(j)}$ $(k = 1, ..., K)$ represent the proportional change in $H^{(j)}$ due to a one unit change in the k^{th} regressor (ceteris paribus). Table 3.1.2's first column with results has $lH1 = \log(H^{(1)})$ as the dependent variable with its first and second lag as regressors plus the 43 other regressors in (E) of Table 3.1.1, while still using the

same 6 external instruments. This leads to statistically satisfactory results, although the p-value of the Reset test is still a bit smaller than we would like. The low p-value of the exogeneity test supports using TSLS instead of OLS.

From the 45 regressors 14 have a t-ratio below one in absolute value. Most probably, some of these regressors may actually be redundant. Although removing some will yield cosmetic improvements, we did not pursue this road. Testing whether the sum of the three coefficients of (lags of) PSI equals zero is strongly rejected. Hence, this supports to conclude that the long-run effect of $H^{(1)}$ with respect to PSI, which is estimated to be $(0.065+0.001-0.034)/(1-0.54-0.10) = 0.032/0.36 = 0.09$, is distinctly positive too. The immediate effect of an increase in PSI by one standard deviation (which is 18.9) on health clinic visits for respiratory illnesses is estimated to be 6.5%, with a relatively small standard deviation. Hence, and increase of actual PSI by 100 is estimated to leads to a $(6.5/18.9) \times 100\% = 34\%$ increase of visits by patients to a clinic in the same week, and an increase of $(9/18.9) \times 100\% = 48\%$ in the long run. The static S&S analysis, which adopts a linear relationship between $H^{(1)}$ and PSI, suggests an immediate and non-lagging absolute increase of clinic visits of $0.35 \times 376/18.9 = 6.98$ for each unit increase of actual PSI. So, for an increase by 100 it suggests 698 extra clinic visits. Since its average has been 2715, at that level an increase by 698 is 26%, but is would be a larger percentage increase at a lower initial level and a smaller one at a larger level.

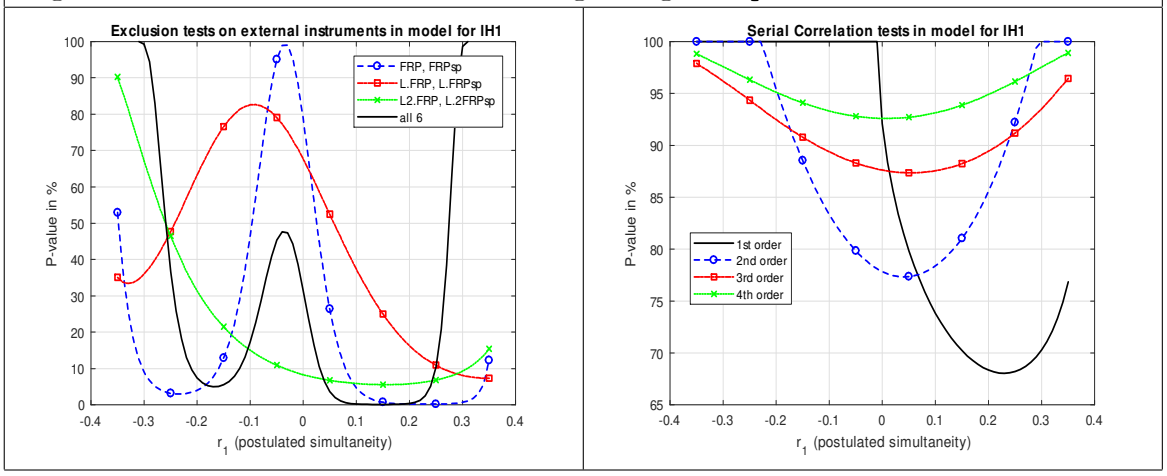
For $H^{(2)}$ we experienced difficulties to find a satisfactory model specification using the TSLS methodology. The S&S specification proved to lack dynamics here too. The equivalent of the result for the dynamic linear model for $H^{(1)}$ in column (E) of Table 3.1.1 is presented in the second column with results of Table 3.1.2. Note that some instruments seem invalid and the residuals are still serially correlated. Neither the log-linear nor the log-log model, presented in the remaining two columns, are really satisfactory either. Of course, in the log-log model we took the log of the 6 external instruments, but note that these appeared to be very weak. So, the above TSLS analysis is certainly unsatisfactory for the analysis of $H^{(2)}$, but also the results for $H^{(1)}$ remain doubtful, simply because the validity of external instruments is always speculative, even if the test for overidentification restrictions produces a high p-value. Hence, the obtained TSLS estimates, both of the coefficients and of their standard errors, may be seriously biased. On the other hand, it could also be the case that the degree of endogeneity of the regressors is in fact so futile, that the use of TSLS is in fact not called for and OLS would suffice. In the next subsection we will therefore examine the sensitivity of

the TSLS results with respect to the possible degree of endogeneity, by using recently developed instrument-free methods.

3.2. Sensitivity analysis avoiding the use of instruments

Endogeneity of PSI in models for $H^{(j)}$ could be due to simultaneity, as we argued above. In that case, taking it for granted that the equivalents of β_1 and α_1 in (2.1) and (2.8) have $\beta_1 > 0$ and $\alpha_1 < 0$, that would imply ρ_1 to be mildly negative. Measurement errors in PSI seem unlikely, but would lead to negative endogeneity too. The inclusion of the many regressors in F and W, and our abundant use of lagged regressors, are all meant to forestall endogeneity of PSI due to omitted regressors. So, we should focus on situations where the correlation between PSI and the model disturbance term is around zero with special interest in mildly negative values. For that we use Matlab code for the methods only very recently put forward in Kiviet (2019b). These generalize those for (almost) normally distributed cross-section data presented in Kiviet (2019a), to make them applicable to general time-series relationships such as we have in the present study. We will just examine the sensitivity with respect to possible endogeneity of TSLS-based inferences on the effect of PSI on $lH1$ as assessed in the (in its own context reasonably satisfactory) first column of Table 3.2.1.

Figure 3.2.1 Instrument-free inference regarding the specification of model for $lH1$

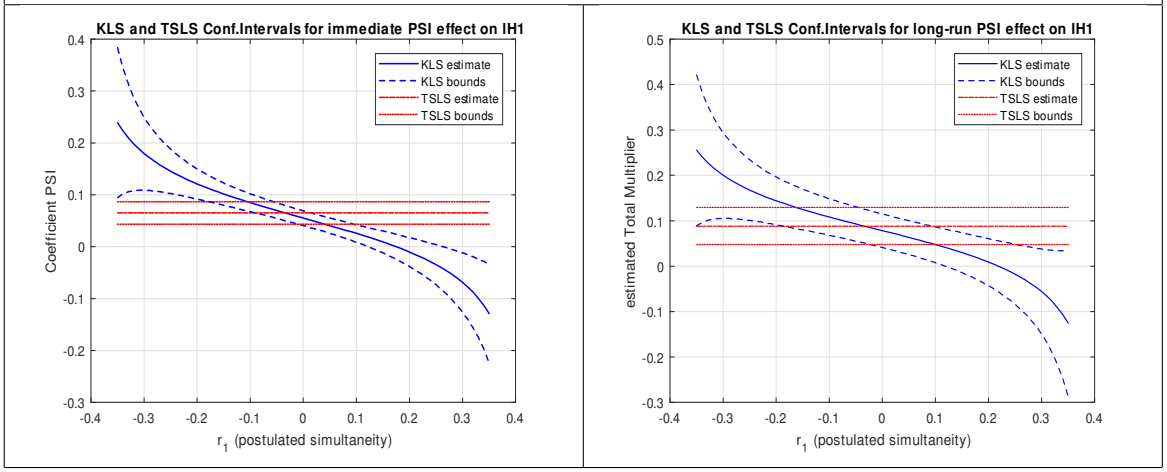


The left-hand graph in Figure 3.2.1 presents p-values for exclusion restrictions tests, when applied to the model for $lH1$, while assuming regarding simultaneity a particular numerical value of the correlation between endogenous regressor PSI and the disturbance term. This value r_1 is indicated on the horizontal axis. For $r_1 = 0$ the model (with the earlier excluded regressors) has been estimated by OLS, and for $r_1 \neq 0$ by a

technique called KLS, which basically corrects OLS for its simultaneity bias, and adapts the estimated variance of the corrected coefficient estimates accordingly. The resulting p-values of the test statistics suggest that when ρ_{xu} is actually close to -0.2, for instance, some of the instruments used to obtain our TSLS estimates seem actually invalid, given the low p-values for two of the curves, despite the value of 0.41 for the overidentification restrictions test obtained earlier by TSLS. Possibly, (lagged) FRP measurements in Indonesia do directly contribute, in addition to (lagged) Singaporean PSI measurements, to explain $H^{(1)}$. In what follows, we will not try to dig deeper into finding a possibly improved specification of a model for $lH1$, because we suspect that would require collecting more data. We shall just focus on using KLS to disclose the sensitivity of the obtained TSLS results regarding the degree of simultaneity.

The right-hand graph of Figure 3.2.1 shows that the adopted second-order dynamics of the model rendered it immune regarding serial correlation, irrespective of the degree of simultaneity, provided that this is mild (smaller than 0.35 in absolute value). These tests have been obtained by applying KLS to the original model augmented by lagged KLS-residuals, and then testing their significance.

Figure 3.2.2 TSLS and KLS inference regarding dynamic PSI effects on $lH1$

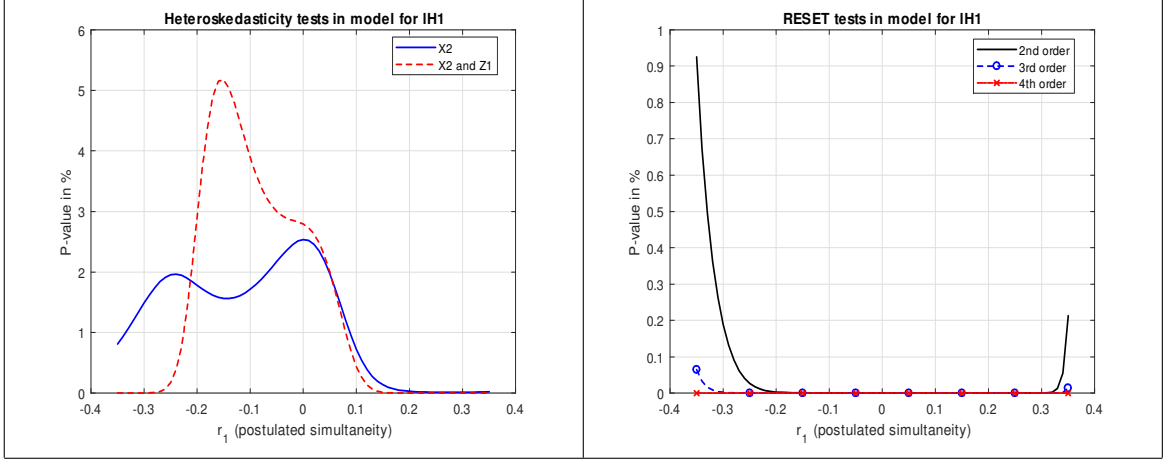


In Figure 3.2.2 the major inference regarding the effect of PSI on $H^{(1)}$ is presented as obtained by KLS, together with the earlier TSLS findings, but now (because of the acceptable p-value for the heteroskedasticity test in Table 3.1.2) not robustifying the standard errors with respect to heteroskedasticity. In the left-hand graph we see confidence intervals for the true coefficient of PSI (unlagged) with asymptotic confidence level 95% (assuming validity of all adopted underlying assumptions). Note that the

TSLS confidence interval is invariant regarding the degree of simultaneity and is centered around the estimated value of 0.065 already given in Table 3.1.2. The KLS interval is not invariant regarding endogeneity. For $r_1 = 0$ it presents the OLS interval, which has always a smaller width than the TSLS interval, although here the difference is quite modest, because the instruments used by TSLS are reasonably strong here. If it happens to be the case that actually $\rho_{xu} = -0.2$ then we find that the TSLS and KLS intervals have no overlap. Under the assumption that the employed model specification is adequate, and given the fact that the actual accuracy in finite samples of KLS has been proved to be much better than that of TSLS, we conclude that in the present situation it can only be the case that TSLS is reasonably accurate as well, if the actual simultaneity is very modest, namely somewhere between -0.06 and 0.0. If it is really -0.2 or even smaller the immediate effect of PSI is much larger than 0.065.

The right-hand graph of Figure 3.2.2 produces asymptotic 95% confidence intervals for the effect of PSI on $lH1$ in the long-run, which is estimated by the ratio with in the numerator the sum of the coefficients for PSI, $L.PSI$ and $L2.PSI$, and in the denominator 1 minus the sum of the coefficients for $L.lH1$ and $L2.lH1$. To estimate the variance of this ratio we used the so-called delta-method; see, for instance, Cameron and Trivedi (2005, p. 231). We see that TSLS and KLS almost agree on a value of this long-run effect being between 0.05 and 0.13 when ρ_{xu} is slightly below -0.05. Interestingly, the correlation between endogenous regressor PSI and the TSLS residuals is -0.034 (from simulations it is known, though, that such an estimator usually has quite a large standard error; moreover, it is only consistent when all instruments are valid).

Figure 3.2.3 Further findings on the specification of the model for $lH1$



The implications of a long-run effect of (standardized) PSI on $lH1$ of 0.09 have already been indicated at the end of subsection 3.1; those for the boundary values (0.04 larger or smaller) can easily be obtained in the same way. However, as was already concluded after interpreting the exclusion restrictions test results, it seems that firm conclusions should not be drawn yet from the present model. In line with that, we see from the left-hand graph in Figure 3.2.3 that, irrespective of the actual seriousness of the simultaneity, heteroskedasticity is detected, despite the reassuring p-value of 0.43 in TSLS Table 3.1.2. As the theory for robustifying KLS standard errors still has to be developed, we cannot correct the results in Figure 3.2.2 for that yet. Moreover, the heteroskedasticity could also be due to omitted regressors, as the exclusion restrictions tests already indicated. Also, our choice for a log-linear specification is possibly not fully satisfactory either, as the RESET test results in the right-hand graph of Figure 3.2.3 object very strongly against the chosen functional form.

4. Concluding remarks

Sheldon and Sankaran (2017), referred to as S&S here, aim at obtaining separate causal structural relationships which explain the air pollution index level (PSI) in Singapore, and the effect of PSI on particular health indicators in Singapore. They use a data set of 325 weekly observations over 2010 through mid-2016 on a (necessarily limited) set of particular variables. It is well-known that such research based on non-experimental data is as a rule complicated by possible endogeneity of explanatory variables due to either (or a combination of) mismeasured explanatories, omitted explanatories, or simultaneity (especially in case of using discrete time-series obtained by integrating mostly intrinsically continuous data).

That a genuine causal structural model for the level of the air pollution index in Singapore cannot be obtained just on the basis of integrated data on all bush and peat fires all over Indonesia, weather variables just in Singapore, and simple fixed year and month effects should of course not surprise. Local weather at the individual fire zones in Indonesia, and at locations between these fires and Singapore, and on the distance of these locations from Singapore, seem indispensable too. Nevertheless S&S model PSI without taking endogeneity of explanatories into account. Because the available data set does not allow to overcome this, in this replication study we focus on the relationships for health indicators only.

By just using and interpreting the broad gamut of model validation tests directly

available in the same software as used by S&S, we show that the models they selected for health indicators wrongly omit particular significant regressors, in particular lagged regressors, and also that they used a functional form which seems inadequate for rather obvious reasons. We aimed at repairing their results, while allowing like S&S for endogeneity of the explanatory variable PSI. This endogeneity may stem from simultaneity, due to possibly successful interventions by the Singaporean authorities, or from omitted explanatories, since measurement errors of PSI seem less likely. However, we did book only limited success, mainly because it seems that also for these health relationships really satisfactory specifications require a more extensive data set. For the specific ailments analyzed (respiratory and eye inflammations) the available data set is poor on explanatory variables for these ailments other than local weather and mechanical time-effects. The major challenge with causal structural analysis on the basis of non-experimental data is, however, that in principle one has to incorporate into the model all aspects explaining the regressand. In experiment-based Randomized Control Trials, on the other hand, it is just the marginal effect of the treatment that has to be modeled, because the effect of all other factors can be eliminated due to the randomization. As the available data set just allows to model the ailments (apart from the pollution effect) by the general health trend represented by clinic visits for diarrhea, weather and fixed seasonal and year effects, there will be omitted relevant explanatories. This may lead, for instance, to diarrhea to become another endogenous explanatory, next to PSI.

Self-evidently, trying to obtain data on relevant further explanatory variables, both for PSI and for the health variables, and next produce further analysis, would establish a useful follow-up project.

Only for the logarithm of the respiratory health indicator we seem to have come rather closely to a satisfactory TSLS-based result. Because TSLS results always remain doubtful, as in the TSLS context the validity of instruments is untestable, we supplement these TSLS results with findings obtained by recently developed instrument-free methods, which require to make an assumption (possibly in the form of an interval) on the likely degree of simultaneity. At the one hand these are found to provide some backing of the TSLS inference on the pattern of the numerical effects of PSI on health, but on the other hand these techniques also allow new forms of model evaluation tests. Those tests seem to be much stricter than those based on TSLS, because unlike the TSLS validation checks, they reject the TSLS-based preferred specification.

References

Andrews, I., Stock, J., Sun, L., 2019. Weak instruments in IV regression: Theory and practice. *Annual Review of Economics* 11, 727–753.

Baum, C.F., Schaffer, M.E., Stillman, S., 2007. Enhanced routines for instrumental variables / generalized method of moments estimation and testing. *The Stata Journal* 7, 465-506.

Cameron, A.C., Trivedi, P.K., 2005. *Microeconometrics; Methods and Applications.* Cambridge University Press.

Dominici, F., Greenstone, M., Sunstein, C.R., 2014. Particulate matter matters. *Science* 344, 257-259.

Imbens, G.W., 2014. Instrumental Variables: An econometrician's perspective. *Statistical Science* 29, 323-358.

Kiviet, J.F., 2019a. Testing the impossible: identifying exclusion restrictions. To appear in *The Journal of Econometrics.*

Kiviet, J.F., 2019b. Instrument-free inference under confined regressor endogeneity; derivations and applications. *Mimeo.*

Murray, M.P., 2006. Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives* 20, 111-132.

Sheldon, T.L., Sankaran, C., 2017. The impact of Indonesian forest fires on Singaporean pollution and health. *American Economic Review* 107, 526-529.

Quah, E., 2002. Transboundary pollution in Southeast Asia: The Indonesian fires. *World Development* 30, 429-441.

Quah, E., Chia, W.M., 2019. What the 2015 haze cost S'pore: $1.83 billion. *The Straits Times*, September 17.