HEALTH ECONOMICS
Health Econ. **10**: 669–671 (2001)
DOI: 10.1002/hec.638

ECONOMETRIC ISSUES IN TESTING THE AGE NEUTRALITY OF HEALTH CARE EXPENDITURE

CHRISTIAN SALAS* AND JAMES P. RAFTERY
Health Services Management Centre, University of Birmingham, Birmingham, UK

SUMMARY

A recent study by Zweifel *et al*. (Zweifel P, Felder S, Meiers M. Ageing of the population and health care expenditure: a red herring? *Health Economics* 1999; **8**: 485–496) suggests that age is not related to health care expenditure among the elderly once 'closeness to death' is controlled for. If correct, this finding has major policy implications, but flaws in the econometric analysis undermine its credibility. We highlight two in particular, and propose methods to deal with them. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS — ageing; health care expenditure; endogeneity; sample selectivity

INTRODUCTION

Population trends attributable to the combined effects of improvements in the survival of older persons, declining fertility among younger persons, and ageing of the 'baby-boom' cohorts have brought about large increases in the proportions of older people in industrialised countries. Recent projections of future changes in the age composition of OECD countries suggest that in some (e.g. Germany), well over 25% of the population will be aged 65 or older in 2030 [2]. Since the prevalence and incidence of most chronic diseases and disabilities seem to increase with age, there is widespread concern that per capita health care expenditures will have to increase commensurately. However, recent econometric evidence provided by Zweifel *et al*. [1] suggests that per capita health care expenditures are, in fact, independent of population ageing, because 'closeness to death' is the main cost driver at the level of the individual not ageing *per se*.

It is well known that dying patients can generate disproportionately large health care costs [3], but deficiencies in the paper's econometrics undermine its credibility with regard to age effects. Two particular problems are the potential *endogeneity* of closeness to death, and potential *sample selectivity* due to observations with zero health care use. The first is overlooked by Zweifel *et al*., and the second, we believe, is not properly controlled for, possibly leading to multicollinearity and incorrect inferences about age and selectivity. We discuss these issues, and suggest ways of dealing with them.

ENDOGENEITY OF CLOSENESS TO DEATH

Zweifel *et al*. use quarterly data and ordinary least squares (OLS) to estimate

* Correspondence to: Lecturer in Health Economics, University of Birmingham, Health Services Management Centre, Park House, 40 Edgbaston Park Road, Birmingham B15 2RT, UK. Tel.: +44 121 4147705; fax: +44 121 4147051; e-mail: c.p.salas@bham.ac.uk

Received 19 June 2000
Accepted 5 February 2001

$$\ln \text{HCE} = \beta_0 + \beta_1 A + \beta_2 A^2 + \beta_3 \text{SEXF} + \beta_4 (A\,\text{SEXF}) + \beta_5 \text{INS} + \beta_6 \lambda + \Sigma_q \gamma_q Q_q + \Sigma_t \delta_t Y_t + \varepsilon \quad (1)$$

where HCE is the health care expenditure, A is the calendar age in years, SEXF is a dichotomous sex indicator, INS is a dummy indicating an insurance policy providing for privileged treatment in hospital, λ is the hazard rate for the standard normal distribution ('Inverse Mills' ratio') to capture the impact of sample selectivity, Q_q is a dummy taking the value 1 in quarter q before death (0 otherwise), Y_t is a dummy taking the value 1 in year t (0 otherwise), and ε is a random error. To provide 'a stronger test of the age neutrality hypothesis', they also estimate OLS models treating HCE as a time series and regressing it against closeness to death. The coefficients are used to explore parameter stability.

In using OLS, the authors assume that closeness to death is weakly exogenous with respect to the parameters of (1) i.e. that (1) can be estimated without knowledge of the parameters of the data-generating process of closeness to death [4]. This implies that HCE in a given quarter cannot affect closeness to death in that quarter. No attempt is made to test or justify this, although it raises the question of why such care is sought (and provided) in the first place, and it is clearly plausible that HCE in a given quarter *does* contemporaneously affect health status and thence closeness to death. If so, weak exogeneity is wrong, OLS estimates may be biased and inconsistent, and their standard errors incorrect. If closeness to death is not strongly exogenous either, and lagged HCE values are uncorrelated with ε, the latter will 'Granger-cause' closeness to death [4] and may be used as 'instruments' in an instrumental variables (IV) routine to estimate (1) consistently. In the context of Zweifel *et al.*'s model, the procedure might be as follows:

(a) Run by OLS the following augmented version of model (1):

$$\ln \text{HCE} = \beta_0 + \beta_1 A + \beta_2 A^2 + \beta_3 \text{SEXF} + \beta_4 (A\,\text{SEXF}) + \beta_5 \text{INS} + \beta_6 \lambda + \Sigma_q \gamma_q Q_q + \Sigma_t \delta_t Y_t + \Sigma_i \rho_i \ln \text{HCE}_{-i} + \varepsilon \quad (2)$$

where the extra variables are lagged values of the dependent variable. The number of lags is limited by the number of repeated observations on individuals. (In Zweifel *et al.*, three lags on the right-hand side of (2) would reduce the 'sample size' from 4211 to around 2600 in their Sample 1, since the first three quarters' HCE values for each individual would have to be taken as given. The benchmark for the dummies Q_q would also change).

(b) In (2), test the null hypothesis that the coefficients of the lagged variables are jointly equal to zero. It can be shown [5, p. 148] that this is equivalent to a 'Hausman test' for equality of OLS and IV estimates of the parameters of (1), using the lagged variables as instruments. If the null cannot be rejected, then neither can the hypothesis that OLS is *at least* as good as IV for estimating (1) (it may be more efficient). Otherwise, inference should be based on the IV estimates, with the lagged dependent variables as instruments.

The above is valid only if the lagged dependent variables are orthogonal to ε. This may fail to hold, for example, if unobserved time-persistent characteristics relating to individuals' particular diseases/conditions also affect HCE, since these would be present in the lagged values of HCE as well as in ε. A superior model (requiring more data) would also test and control for such effects.

SAMPLE SELECTIVITY

To control for sample selectivity, Zweifel *et al.* employ the familiar 'Heckit' model, which uses the results from an initial probit analysis of zero/positive health care utilization to compute λ in (1). The probits are not presented, but we are told that there are significant age effects, with a negative coefficient of A^2 leading to a negative marginal impact of an extra year when evaluated at the mean of A (counterintuitively, this suggests that the probability of health care utilization declines with age). Since no other variables are mentioned, we assume that the probit model is

$$d = \mathbf{1}[\beta_0^{**} + \beta_1^{**} A + \beta_2^{**} A^2 + \eta > 0] \quad (3)$$

where $d = 1$ if health care is used during the observation period (0 otherwise), $\mathbf{1}[\cdot]$ is an indicator function of the event inside the brackets, the β_i^{**} are parameters to be estimated, and η is a standard normal random error term. It makes no

difference to our critique if more of the remaining variables in (1) also appear in (3), and there is clearly an additional specification issue here about the appropriateness of omitting them.

A Maclaurin series expansion of λ (evaluated at $\beta_0^{**} + \beta_1^{**}A + \beta_2^{**}A^2$) gives

$$\lambda = \beta_0^{*} + \beta_1^{*}A + \beta_2^{*}A^2 + \xi \tag{4}$$

where β_0^{*}, β_1^{*} and β_2^{*} are functions of β_0^{**}, β_1^{**}, β_2^{**} (and also of $\lambda(0)$, $\lambda'(0)$, $\lambda''(0)$ and $\lambda'''(0)$), and ξ involves higher-order powers of A and derivatives of λ. (4) in (1) gives

$$\begin{aligned}\ln \mathrm{HCE} = {} & \beta_0 + \beta_1 A + \beta_2 A^2 + \beta_3 \mathrm{SEXF} \\ & + \beta_4 (A\,\mathrm{SEXF}) + \beta_5 \mathrm{INS} \\ & + \beta_6(\beta_0^{*} + \beta_1^{*}A + \beta_2^{*}A^2 + \xi) + \Sigma_q \gamma_q Q_q \\ & + \Sigma_t \delta_t Y_t + \varepsilon \end{aligned} \tag{5}$$

which shows that if ξ is small, λ in (1) is almost collinear with the age variables. Plotting λ shows that it is almost linear over much of its range [6], so ξ *is* likely to be small. Thus, Zweifel *et al*.'s inferences based on (1) and (3) may be unreliable because of potential multicollinearity between A, A^2 and λ (in addition to the potential endogeneity of closeness to death). This could make the coefficients of A and A^2 seem less significant than they are (due to inflated standard errors), and also appear with the 'wrong' signs [7].

To examine this issue, conventional tests for multicollinearity are applicable. A common way to deal with it is to exclude some variables from the probit model, but this is inappropriate here since HCE and the utilization decision depend on similar factors. Another way might be to abandon the additive model $\beta_0^{**} + \beta_1^{**}A + \beta_2^{**}A^2$ altogether in (3) (since this gives the almost linear expansion of λ in (4)), and respecify (3) as

$$d = \mathbf{1}[m(X) + \eta > 0] \tag{6}$$

where m is some (unspecified) function, X is a vector of weakly exogenous variables in (1) (including A and A^2), and η is as in (3). Then $\mathrm{Prob}(d = 1) = E(d|X) = \Phi(m(X))$, where Φ denotes the standard normal c.d.f., so

$$m(X) = \Phi^{-1}(E(d|X)) \tag{7}$$

Thus, $m(X)$ is estimable by evaluating $\Phi^{-1}(\cdot)$ at a nonparametric estimate of the conditional mean $E(d|X)$, which can be obtained as a weighted average of the selection variables d. Formulas and other issues such as computing the marginal effects of the variables in $m(X)$ are discussed in [5, p. 288] and [8]. Also see [6] and references therein. $\lambda(\cdot)$ in (1) is computed as $\lambda(\Phi^{-1}(E(d|X)))$, which is arguably less prone to be collinear with A and A^2 than $\lambda(\beta_0^{**} + \beta_1^{**}A + \beta_2^{**}A^2)$. This should, of course, be tested.

REFERENCES

1. Zweifel P, Felder S, Meiers M. Ageing of the population and health care expenditure: a red herring? *Health Economics* 1999; **8**: 485–496.
2. Organization for Economic Co-operation and Development (OECD). *Ageing in OECD Countries: A Critical Policy Challenge*. OECD: Paris, 1996.
3. Garber A, McCurdy T, McClellan M. *Medical care at the end of life: diseases, treatment patterns and costs*, 1998. National Bureau of Economic Research (NBER) Working Paper No. 6748.
4. Hendry D. *Dynamic Econometrics*. Oxford University Press: Oxford, UK, 1995.
5. Kennedy P. *A Guide to Econometrics* (3rd edn). Blackwell: Oxford, UK and Cambridge, USA, 1992.
6. Jones AM. Health Econometrics. In *North-Holland Handbook in Health Economics*, Newhouse JP, Culyer AJ (eds). Elsevier: Amsterdam, 2000; 23.
7. Greene W. *Econometric Analysis* (3rd edn). Prentice-Hall: New Jersey, 1997; 420.
8. Delgado M, Robinson P. Nonparametric and semiparametric methods for economic research. In *Surveys in Econometrics*, Oxley L, George D, Roberts C, Sayer S (eds). Blackwell: Oxford, UK and Cambridge, USA, 1995; 350–396.