Volume 30, Issue 3

Spurious correlation in estimation of the health production function: A note

Sule Akkoyunlu
Swiss Form for Migration and Population

Frank R. Lichtenberg
Columbia University

Boriss Siliverstovs
KOF Swiss Economic Institute

Peter Zweifel
University of Zurich

Abstract

In this paper, we address the issue of spurious correlation in the production of health in a systematic way. Spurious correlation entails the risk of linking health status to medical (and nonmedical) inputs when no links exist. This note first presents the bounds testing procedure as a method to detect and avoid spurious correlation. It then applies it to a recent contribution by Lichtenberg (2004), which relates longevity in the United States to pharmaceutical innovation and public health care expenditure. The results of the bounds testing procedure show longevity to be related to these two factors. Therefore, the estimates reported by Lichtenberg (2004) cannot be said to be result of spurious correlation, to the contrary, they very likely reflect an effective relationship, at least for the United States.

1. Introduction

The issue of spurious correlation in models involving time series data has a long history. Already in the third decade of the 20th century Yule pointed out on pitfalls of correlation analysis applied to trending data (Yule, 1926). The topic received further development in Granger and Newbold (1974) and Engle and Granger (1987) where the problem of spurious correlation has been formalized and the initial solution has been provided, respectively. In particular, it was emphasized that in order to avoid spurious correlation between the time series of interest one has to conduct pretesting for the presence of unit roots first and then apply cointegration analysis. Failure to do so entails the risk of finding a statistically significant relationship among the variables when in fact there is none. The classical example includes a very high, positive correlation—referred to as a nonsense correlation in Yule (1926)—between standardized mortality per 1,000 persons in England and Wales and the proportion of Church of England marriages per 1,000 of all marriages. In empirical analysis the problem of spurious regression is expected to arise very often, see Granger (1966), who showed that economic variables typically are non-stationary.

In Health Economics the empirical research greatly benefited from new developments in time series analysis. For example, investigating the relationship between health care expenditure and GDP one can distinguish between the following two stages (see Hartwig (2008) for an overview). The first stage is characterized mostly by cross-sectional studies (e.g., see Kleinman (1974), Newhouse (1977, 1987), Cullis and West (1979), Leu (1986), Parkin et al. (1987), Culyer (1990), Milne and Molana (1991), Getzen and Poullier (1991), Gertram and Jönsson (1991a,b)) such that the issue of spurious correlation could be avoided.

At the second stage the time series properties of the variables in question received more attention. The analyses include application of tests for the presence of unit roots and testing for cointegration between health care expenditure and GDP (e.g., see Murthy and Ukpolo (1994), Hansen and King (1996), Blomquist and Carter (1997), McCoskey and Selden (1998), Roberts (1999), and Gerdtham and Löthgren (2000)). Jewell et al. (2003) and Carrion-i-Silvestre (2005) extend the literature by addressing time series properties of the modelled variables allowing for structural breaks. See also an overview of the relevant literature as well as a brief description of unit root and cointegration methodology in Gerdtham and Jönsson (2000).

Nevertheless, the need for appropriate modelling of relationships involving time series variables so far has not been fully acknowledged in applied health economics. One example is the recent study of Lichtenberg (2004), which relates longevity in the United States to pharmaceutical innovation and public health care expenditure. The results of Lichtenberg (2004) suggest that recent increases in U.S. life expectancy are attributable to pharmaceutical innovation and public health care expenditure (HCEP) rather than health care expenditure (HCE) in general and non-medical determinants such as real GDP *per capita* which are found to be insignificant.

A potential problem with the conclusions reached in Lichtenberg (2004) is that the issue of spurious regression has not been addressed in a systematic way. This failure is important because it entails the risk of linking health status to medical (and non-medical) inputs when no links exist. Given the fact that the conclusions of Lichtenberg (2004) could have far-reaching consequences for conducting health policy in the United States, one has to verify his

results using a methodology that explicitly takes into account the issue of spurious correlation.

Contributing to the literature, we replicate the study of Lichtenberg (2004) using an appropriate econometric methodology that is robust against the spurious regression problem. To this end, we use the bounds testing procedure suggested by Pesaran, Shin, and Smith (2001), which has a number of advantages over other methods as outlined below in Section 2. We apply it to the Lichtenberg (2004) data in order to sort out whether the results reported therein are due to spurious correlation or reflect a genuine relationship between the variables in question.

The rest of the paper proceeds as follows. In Section 2 the bounds testing procedure of Pesaran et al. (2001) is described in detail. In Section 3 the results are presented and the final section concludes.

2. Method

In order to resolve the spurious correlation problem, the bounds testing procedure developed by Pesaran et al. (2001) will be presented in the following. This choice can be justified based on the following considerations. First, the bounds testing approach has broad applicability since the regressors can be of different and/or mixed order of integration, e.g., the variables can be either integrated of order one [technically, I(1)], of order zero [I(0)], or mutually cointegrated. This is a great advantage since the unit root tests regarding the order of integration of the relevant variables yield inconclusive results (as discussed in Lichtenberg, 2004). Second, the procedure is based on the unrestricted error correction model, which permits joint estimation of long and short-run effects. As pointed out in Banerjee et al. (1993, 1998), joint estimation has better statistical properties than the two-step Engle-Granger procedure that pushes the short-run dynamics into the error term. Third, as discussed in Narayan (2005), the bounds testing procedure performs rather well in small samples like the one employed in Lichtenberg (2004). In small samples, application of the more popular Full Information Maximum Likelihood method (Johansen, 1995) is problematic.

To implement the bounds testing procedure in the present context, one assumes that life expectancy (LE), new molecular entities (NME, a measure of pharmaceutical innovation), and public health care expenditures per capita (HCEP) are related according to a vector autoregressive (VAR) model. Dropping real GDP and health care expenditures in general that were found insignificant by Lichtenberg (2004) and also by us, the VAR model that can be reduced to the following conditional error correction model (ECM),

$$\Delta LE = \theta_0 LE_{t-1} + \theta_1 NME_{t-1} + \theta_2 HCEP_{t-1} + \sum_{i=1}^{p-1} \lambda_i \Delta LE_{t-i} +$$

$$+ \sum_{i=0}^{p-1} \beta_i \Delta NME_{t-i} + \sum_{i=0}^{p-1} \beta_i \Delta HCEP_{t-i} + \omega' DT_t + \varepsilon_t, \tag{1}$$

Here, p is the lag length of the underlying VAR model. The lagged values of *LE, NME* and *HCEP* (are in logs) form a long-run relationship in levels. The deterministic terms such as a constant, a linear trend, and dummy variables are denoted by DT_t. The short-run dynamics is

captured by means of lagged values of ΔLE_t and current and lagged values of ΔNME_t and $\Delta HCEP_t$. The data span the time period from 1960 to 2001. The conditional long-run elasticities of life expectancy with respect to *MNE* and *HCEP* are given by $-\theta_1/\theta_0$ and $-\theta_2/\theta_0$, respectively (Banerjee et al., 1998).

The bounds testing procedure uses the conventional F test for testing the null hypothesis H_0: $\theta_1=\theta_2=0$. Note that this statistic has a non-standard distribution which depends upon (i) the order of integration of the regressors; (ii) the number of regressors; (iii) the set of deterministic terms included in the model; and (iv) sample size. Pesaran et al. (2001) provide the set of asymptotic critical values, generated from 1,000 bootstraps. However, given a small sample size (41 observations in the present case), these values are likely to be inappropriate. Therefore, the critical values reported in Narayan (1995) for small sample sizes are used in the following.

There are two sets of critical values. The first set gives the lower bound, applicable when all regressors are I(0). The second gives the upper bound, applicable when all regressors are I(1). If the calculated F statistic falls below the lower bound, the null hypothesis of no relationship in levels cannot be rejected. Conversely, if the F statistic exceeds the upper bound, the null hypothesis of a level relationship can be accepted. Both rules hold irrespective of the order of integration of the regressors. Finally, if the F statistic falls within the critical bounds, the order of integration of the variables must be established in order to obtain conclusive inference.

3. Test results

Table I presents the results of the lag order p selection procedure for equation (1). The information criteria (Akaike, AIC and Schwarz, SIC) as well as the Langrange Multiplier statistic testing for remaining autocorrelation of order 1 and up to 4 in regression residuals are reported. As a robustness check, equation (1) is estimated without and with a linear trend.

Table I: Lag order selection

P	Without deterministic trends				With deterministic trends			
	AIC	SIC	AR(1)	AR(4)	AIC	SIC	AR(1)	AR(4)
1	-9.377	**-9.078**	0.158	0.448	-9.340	**-8.999**	0.190	0.480
2	**-9.382**	-8.956	0.357	0.533	**-9.344**	-8.875	0.370	0.540
3	-9.278	-8.723	0.818	0.469	-9.230	-8.633	0.792	0.459

Notes: p is the lag order of the underlying VAR model for the conditional ECM of equation (1). **Bold entries** indicate the preferred lag order. AIC and SIC are the Akaike and Schwarz Information Criteria, respectively. AR(1) and AR(4) are the LM test statistics for testing for residual autocorrelation of orders up to 1 and 4, respectively.

When no deterministic trend is included, the AIC selects $p=2$ whereas the SIC selects $p=1$. In both cases, there is no evidence of remaining autocorrelation in the regression residuals. Adding a linear deterministic term does not change these conclusions. Given the evidence of

no residual autocorrelation regardless of the value of p and a rather small number of observations, the parsimonious model with $p=1$ is preferred. Next, it seems that inclusion of a linear deterministic trend is not supported either by AIC or SIC, since both statistics take on a higher value (indicating better fit) without a linear trend. Therefore, the model with $p=1$ and without a linear trend in Table 1 will be used for the bounds testing procedure. Nevertheless, the results with a trend included will also be reported to provide some sensitivity analysis.

From the five cases distinguished by Pesaran et al. (2001), the preliminary results reported above calls for retaining the following three,

Case III: unrestricted constant and no trend;

Case IV: unrestricted constant and restricted trend;

Case V: unrestricted constant and unrestricted trend.

Table II: *F*-test statistics for testing the existence of a levels relationship

P	F_{III}	F_{IV}	F_V
1	10.437***	10.556***	10.385***
2	7.607***	7.839***	7.565**

Notes: F_{III}, F_{IV}, and F_V are the *F*-test statistics that correspond to the three cases mentioned in Pesaran et al. (2001): Case III – an unrestricted constant and no linear deterministic trend; Case IV – unrestricted constant and restricted linear deterministic trend; Case V – unrestricted constant and unrestricted linear deterministic trend. Symbols '***' and '**' indicate that the *F*-statistic exceeds the upper bound corresponding to the 1% and 5% significance levels, respectively, as reported in Narayan (2005) for $T = 40$.

The corresponding F test statistics are denoted as F_{III}, F_{IV}, and F_V in Table 2. They exceed the critical upper bound (1 percent significance level) in all cases except for F_V, which exceeds the critical upper bound at the 5 percent significance level for $p=2$. This finding indicates that the null hypothesis of no relationship in levels can be rejected. Thus, longevity, the number of new molecular entities, and public per capita health expenditure are related. This finding is robust both to lag length and the inclusion of a deterministic trend.

Next, one can capitalize on the existence of this relationship for the modelling of the health production function. The objective is to find the most parsimonious model passing the standard diagnostic tests. The implied values of the long-run impact coefficients of that model can then be compared with those reported by Lichtenberg (2004). The equation obtained reads (standard errors in parentheses, error probabilities in brackets):

$$\Delta LE_t = 0.770 + 0.00496 * \Delta NME_t - 0.196 * LLE_{t-1} + 0.00802 * NME_{t-1}$$
$$(0.140) \quad (0.0011) \quad\quad (0.037) \quad\quad (0.0016)$$
$$+ 0.00849 * HCEP_{t-1} - 0.00559 * D93_t \quad\quad\quad\quad (2)$$
$$(0.0019) \quad\quad\quad (0.0022)$$

$R^2 = 0.566$, $F(5,35) = 9.129$ [0.000], T = 41			
AR 1-2 test:	$F(2,33)$	=	1.651 [0.207]
ARCH 1-1 test:	$F(1,33)$	=	0.140 [0.710]
Normality test:	$\chi^2(2)$	=	0.241 [0.886]
Heteroskedasticity test:	$F(9,25)$	=	1.218 [0.329]
Heteroskedasticity-X test:	$F(15,19)$	=	1.135 [0.392]
RESET test:	$F(1,34)$	=	0.683 [0.415]

All coefficients are estimated with a high degree of precision. The model passes all standard specification tests, including tests of no residual autocorrelation, of no residual ARCH effects, of no residual heteroskedasticity, of residual normality, and of no functional form misspecification. These encouraging results are also supported by the close match between actual and the fitted values displayed in the top panel of Figure 1. The estimated regression residuals show no sign of misspecification (middle panel).[1] Finally, the autocorrelation function (bottom panel) takes on rather small values that moreover change signs for lags ranging from 1 to 5 years.

Finally, the results of the Chow tests for recursive stability (Chow, 1960) are shown in Figure 2. The values of the 1-step, breakpoint, and forecast Chow test statistics are compared to their respective 5% critical values (Doornik and Hendry, 2001). None of the tests show any sign of model instability.

According to equation (2), the long-run elasticities of life expectancy with respect to pharmaceutical innovation and public health expenditure are 0.0408 and 0.0432, respectively.

These values are close to those reported in Lichtenberg (2004, Table 3), i.e. 0.0390 and 0.0434. In sum, the estimates reported by Lichtenberg (2004) cannot be said to result from spurious correlation; to the contrary, they very likely reflect an effective relationship between life expectancy on the one hand and, pharmaceutical innovation and public health care expenditure on the other in the United States.

---

[1] In Equation (2), *D93* denotes a dummy that is equal to one in 1993 and zero otherwise. The dummy in 1993 accounts for a decrease in life-expectancy in that year that cannot be explained by developments neither in pharmaceutical innovation nor in public health care expenditures. According to Kranczer (1994), "In 1993, a decrease in life expectation was primarily brought about by the considerable rise in the absolute number of deaths and the corresponding increase in mortality rates. Indications are that mortality increased from all major forms of death as well as from AIDS. It is estimated that there were 2,260,000 deaths in 1993 compared with the provisional count of 2,177,000 in 1992 and the final figure of 2,169,518 in 1991."
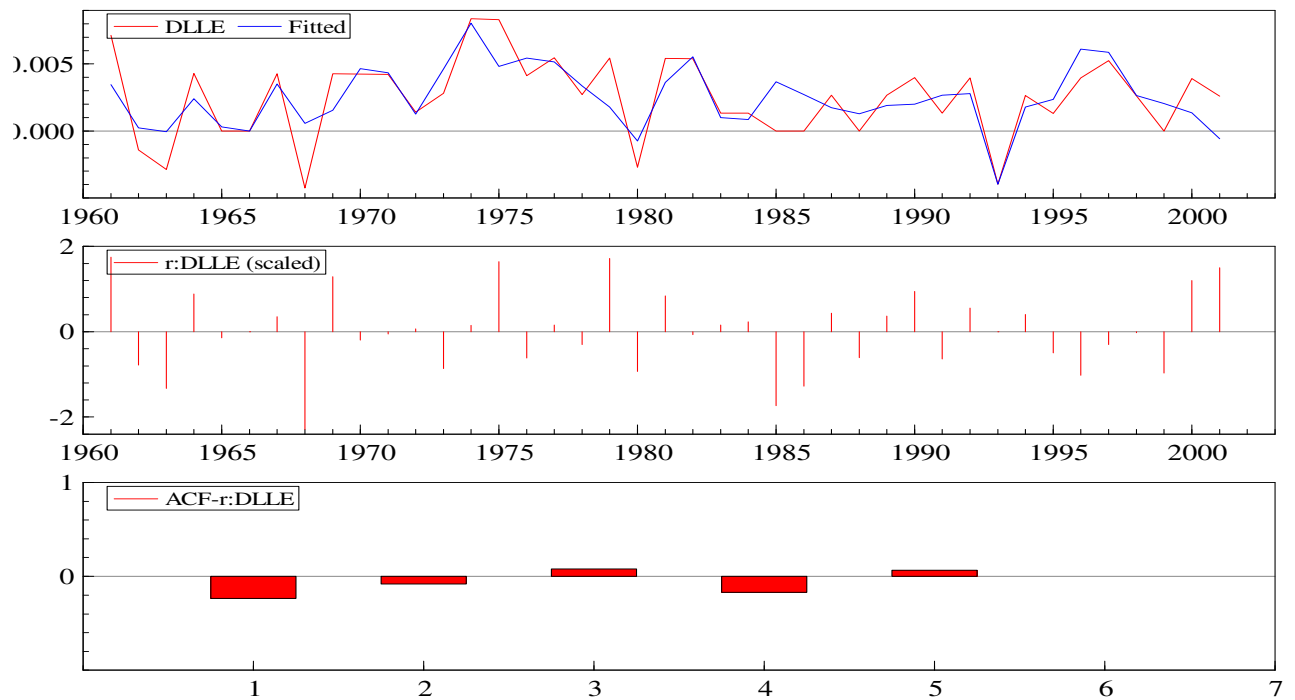
Figure 1: Actual and fitted values (the upper panel); Regression residuals (the middle panel); Autocorrelation function of regression residuals (bottom panel).
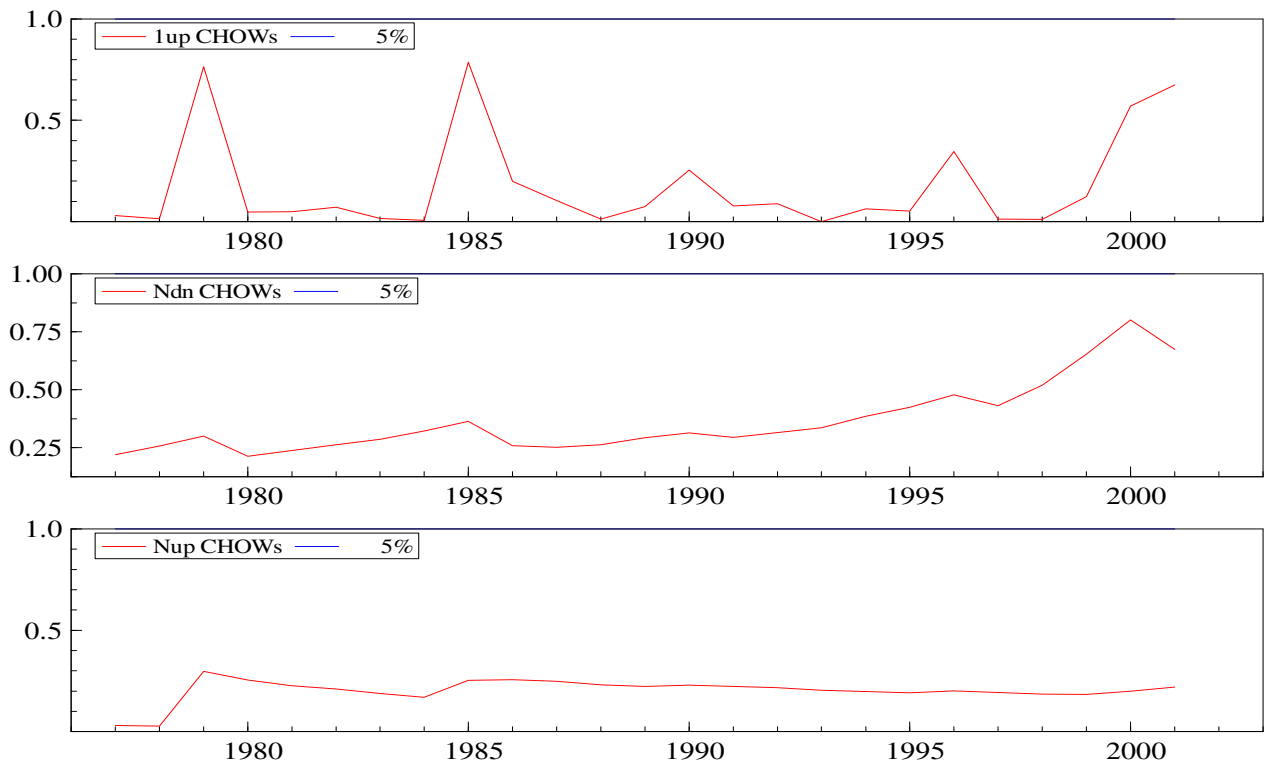

Figure 2: Recursive stability 1-step, breakpoint, and forecast Chow test statistics scaled by their respective 5% critical values.

4. Concluding remarks

As already noted by Gerdtham and Jönsson (2000), health economists until recently have been paying rather little attention to the fact that some of their data are of the time series type. However, there are modern econometric techniques for detecting the existence of a relationship between variables that can be integrated either of order one or zero, or mutually cointegrated. Applying these techniques to the paper by Lichtenberg (2004), this note finds that the null hypothesis of no relationship between life expectancy, pharmaceutical innovation, and public health care expenditure in the United States can be rejected.

However, there are other recent works that might be subjected to a similar scrutiny. For example, Zweifel et al. (2005) employ panel data for revisiting the so-called Sisyphus syndrome (Zweifel and Ferrari, 2002), which involves a production function (HCE resulting in more elderly survivors) and a HCE function (where elderly survivors exert public pressure for more HCE). Other panel data applications that investigate the relation between life expectancy and health care spending include Crémieux et al. (1999), and Or (2001). More specifically, the findings by Peltzman (1987), Lichtenberg (1996), Moore and Newman (1993), Soumerai et al. (1994), and Crémieux et al. (2005), that pharmaceutical consumption contribute life expectancy could be checked using the testing procedures described in this note.

References:

Banerjee, A., J. J. Dolado, J. W. Galbraith and D. F. Hendry (1993) *Cointegration, Error Correction, and the Econometric Analysis of Non-stationary Data*, Oxford University Press: Oxford.
Banerjee, A., J. J. Dolado, and R. Mestre (1998) "Error-correction mechanism tests for cointegration in single-equation framework" *Journal of Time Series Analysis* **19**, 267-283.
Blomqvist, Å. G. and R. A. L. Carter (1997) "Is health care really a luxury?" *Journal of Health Economics* **16**, 207–229.
Carrion-i-Silvestre, J. L. (2005) "Health care expenditure and GDP: Are they broken stationary?" *Journal of Health Economics* **24**, 839–854.
Chow, G. C. (1960) "Test of equality between sets of coefficients in two linear regressions" *Econometrica* **28**, 591-605.
Crémieux, P-Y, P. Ouellette and C. Pilon (1999) "Health care spending as determinants of health outcomes" *Health Economics* **8**, 627-639.
Crémieux, P-Y., M-C. Mieilleur, P. Ouellette, P. Petit, P. Zelder and K. Potvin (2005) "Public and private pharmaceutical spending as determinants of health outcomes in Canada" *Health Economics* **14**, 107–116.
Cullis, J. G. and P. A. West (1979) *The Economics of Health: An Introduction*, Martin Robertson: Oxford.
Culyer, A. J. (1990) *Cost containment in Europe, Health Care Systems in Transition*, OECD: Paris.
Doornik, J. A. and D. F. Hendry (2001) *Modelling Dynamic Systems Using PcGive*, Timberlake Consultants Press: London.
Engle, R. F. and C. W. J. Granger (1987) "Cointegration and Error-Correction: Representation, estimation, and testing" *Econometrica* **55**, 251-276.
Gerdtham, U.-G. and B. Jönsson (1991) "Price and quantity in international comparisons of health care expenditure" *Applied Economics* **23**, 1519–1528.

Gerdtham, U.-G. and B. Jönsson (1991) "Conversion factor instability in international comparisons of health care expenditure" *Journal of Health Economics* **10**, 227–234.

Gerdtham, U.-G. and B. Jönsson (2000) "International comparisons of health expenditure: theory, data and econometric analysis" in *Handbook of Health Economics* by A. J. Culyer and J. P. Newhouse, Eds., Vol. 1A, Elsevier Publishers: Amsterdam, 11-53.

Gerdtham, U.-G. and M. Löthgren (2000) "On stationarity and cointegration of international health expenditure and GDP" *Journal of Health Economics* **19**, 461–475.

Getzen, T. E. and J.-P. Poullier (1991) "An income-weighted international average for comparative analysis of health expenditures" *International Journal of Health Planning and Management* **6**, 3–22.

Granger, C. W. J. (1966) "The typical spectral shape of an economic variable" *Econometrica* **34**, 150–161.

Granger, C. W. J. and P. Newbold (1974) "Spurious regression in econometrics" *Journal of Econometrics* **2**, 111–120.

Hansen, P. and A. King (1998) "Health care expenditure and GDP: Panel data unit root test results—comment" *Journal of Health Economics* **17**, 377–381.

Hartwig, J. (2008) "What drives health care expenditure? Baumol's model of 'unbalanced growth' revisited" *Journal of Health Economics* **27**, 603-623.

Jewell, T., J. Lee, M. Tieslau and M. C. Strazicich (2003) "Stationarity of health expenditures and GDP: Evidence from panel unit root tests with heterogeneous structural breaks" *Journal of Health Economics* **22**, 313–323.

Johansen, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press: Oxford.

Kleinman, E. (1974) "The determinants of national outlays on health" in *The Economics of Health and Medical Care* by M. Perlman, Ed. , Macmillan: London, 66-81.

Kranczer, S. (1994) "Changes in U.S. life expectancy" *Statistical Bulletin Metropolitan Life Insurance Company*, **75**, 11-7.

Lichtenberg, F. (1996) "The effect of pharmaceutical utilization and innovation on hospitalisation and mortality" NBER Working Paper number 5418.

Lichtenberg, F. (2004) "Sources of U.S. longevity increase" *The Quarterly Review of Economics and Finance* **44**, 369-389.

Lichtenberg, F. (2006) "The impact of increased utilization of HIV drugs on longevity and medical expenditure: an assessment based on aggregate U.S. time-series data" *Expert Review of Pharmacoeconomics and Outcomes Research* **6**, 425-436.

Leu, R. E. (1986) "The public–private mix and international health care costs" in *Public and Private Health Care Services: Complementarities and Conflicts* by A.J. Culyer and B. Jönsson, Eds., Basil Blackwell: Oxford, 41-63.

McCoskey, S. K. and T. M. Selden (1998) "Health care expenditures and GDP: Panel data unit root test results" *Journal of Health Economics* **17**, 369–376.

Milne, R. and H. Molana (1991) "On the effect of income and relative price on the demand for health care: EC evidence" *Applied Economics* **23**, 1221–1226.

Moore, W. J., and R. J. Newman (1993) "Drug formulary restrictions as a cost-containment policy in medicaid programs" *Journal of Law and Economics* **16**, 143-162.

Murthy, N. R. V. and V. Ukpolo (1994) "Aggregate health care expenditure in the United States: Evidence from cointegration tests" *Applied Economics* **26**, 797–802.

Narayan, P. K. (2005) "The saving and investment nexus for China: Evidence from cointegration tests" *Applied Economics* **37**, 1979-1990.

Newhouse, J. P. (1977) "Medical-care expenditure: a cross-national survey" *Journal of Human Resources* **12**, 115–125.

Newhouse, J. P. (1987) "Cross-national differences in health spending: what do they mean?" *Journal of Health Economics* **6**, 159–162

Or, Z. (2001) "Exploring the effects of health care on mortality across OECD countries" OECD Labour Market and Social Policy Occasional Papers number 46.

Parkin, D., A. McGuire and B. Yule (1987) "Aggregate health care expenditures and national income: Is health care a luxury good?" *Journal of Health Economics* **6**, 109–127.

Peltzman, S. (1987) "Regulation and health: the case of mandatory prescriptions and extension" *Managerial and Decision Economics* **8**, 41-46.

Pesaran, M. H., Y. Shin and R. J. Smith (2001) "Bounds testing approaches to the analysis of level relationships" *Journal of Applied Econometrics* **16**, 289–326.

Roberts, J. (1999) "Sensitivity of elasticity estimates for OECD health care spending: Analysis of a dynamic heterogeneous data field" *Health Economics* **8**, 459–472.

Soumerai, S. B., T. J. McLaughlin, D. Ross-Degnan, C. S. Casteris and P. Bollini (1994) "Effects of a limit on medicaid drug-reimbursement benefits on the use of psychotropic agents and acute mental health services by patients with schizophrenia" *The New England Journal of Medicine* **331**, 650-655.

Yule, G. U. (1926) "Why do we sometimes get nonsense correlations between time-series? A study in sampling and the nature of time-series" *Journal of Royal Statistical Society. Series A* **89**, 1-69.

Zweifel, P., and M. Ferrari (1992) "Is there a Sisyphus syndrome in health care?" in *Health Economics Worldwide* by P. Zweifel and H. E. Frech III, Eds., Kluwer Academic Publishers: Amsterdam, 311-330.

Zweifel, P., L. Steinmann and P. Eugster (2005) "The Sisyphus syndrome in health revisited" *International Journal of Health-Care Finance and Economics* **5**, 127-145.