University of Zurich

Spurious correlation in estimation of the health production function:

A note

Sule Akkoyunlu, Frank R. Lichtenberg, Boriss Siliverstovs, Peter Zweifel

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

Author's address: Sule Akkoyunlu
                     E-mail: sule.akkoyunlu@soi.uzh.ch

                     Frank R. Lichtenberg
                     E-mail: frank.lichtenberg@columbia.edu

                     Boriss Siliverstovs
                     E-mail: boriss.siliverstovs@kof.ethz.ch

                     Peter Zweifel
                     E-mail: pzweifel@soi.uzh.ch

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

Sule Akkoyunlu[♥], Frank R. Lichtenberg[♣], Boriss Siliverstovs[♦] and Peter Zweifel[♠]

November 2008

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 linearly 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.

JEL Classification: H51, I12, J18, C22 (C32), O33

Key Words: Health; Life expectancy; Innovation; Pharmaceuticals; Health care expenditure; Cointegration

1. Introduction

Although empirical research into the production of health can benefit greatly from time series

data, there are only few examples [Katsouyanni et al. (1997), Woodruff et al. (1997), Samet et

al. (2000), Shmueli (2003), and Lichtenberg (2004, 2006)]. However, in these studies the

issue of spurious regression has never been addressed in a systematic way, in spite of its

particular relevance to time series analysis. This failure is important because it entails the risk

of linking health status to medical (and nonmedical) inputs when no links exist. It is even

gaining in importance in view of evidence presented by Lichtenberg (2004) suggesting that

recent increases in U.S. life expectancy are attributable to pharmaceutical innovation and

[♥] University of Zurich, Socioeconomic Institute, Hottingerstrasse. 10, CH-8032 Zurich, Switzerland, email: sule.akkoyunlu@soi.uzh.ch

[♣] Columbia University, Graduate School of Business, 614 Uris Hall, 3022 Broadway, New York, NY 10027, USA,
e-mail: frank.lichtenberg@columbia.edu

[♦] ETH Zurich, KOF Swiss Economic Institute, Weinbergstrasse 35, CH-8092 Zurich, Switzerland, e-mail: boriss.siliverstovs@kof.ethz.ch

[♠] University of Zurich, Socioeconomic Institute, Hottingerstrasse. 10, CH-8032 Zurich, Switzerland, email: pzweifel@soi.uzh.ch

public health care expenditure (HCEP) rather than health care expenditure (HCE) in general and nonmedical determinants such as real GDP *per capita* which are found to be insignificant.

In their pioneering analysis, Granger and Newbold (1974) showed that when the dependent and explanatory variables are integrated of order one [technically, I(1)], analysis of the relationship between these variables using the ordinary least squares is likely to be subject to spurious correlation. To exclude this possibility, one either has to prove that the variables in question are (trend-)stationary or that the I(1) variables are cointegrated. Alternatively, one has to use a method that works equally well irrespective of the underlying integration order of the variables. One such method is the bounds testing procedure suggested by Pesaran, Shin, and Smith (2001). 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.

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 I(1), 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 1 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 1 about here

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

4

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.

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 linearly related. This finding is robust both to lag length and the inclusion of a deterministic trend.

————————————

Table 2 about here

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),

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

(2)

$R^2 = 0.566$, 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.

---

[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."

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.

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., Dolado, J.J., Galbraith, J.W., and Hendry, D.F. (1993), *Cointegration, Error Correction, and the Econometric Analysis of Non-stationary Data*, Oxford University Press, Oxford.

Banerjee A., Dolado J., and Mestre R. (1998). Error-correction mechanism tests for cointegration in single-equation framework. *Journal of Time Series Analysis*, 19, 267-283.

Chow, G. C. (1960), Test of equality between sets of coefficients in two linear regressions, *Econometrica*, 28, 591-605.

Crémieux, P-Y, Ouellette P., Pilon C. (1999). Health care spending as determinants of health outcomes. *Health Economics*, 8, 627-639.

Crémieux, P-Y., Mieilleur M-C., Ouellette P., Petit P., Zelder P., Potvin K. (2005). Public and private pharmaceutical spending as determinants of health outcomes in Canada. *Health Economics*, 14, 107–116.

Doornik, J. A. and D. F. Hendry (2001). *Modelling Dynamic Systems Using PcGive*. London: Timberlake Consultants Press.

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

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

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

Katsouyanni, K., Touloumi, G., Spix, C., Schwartz, J., Balducci, F., Medina, S., Rossi, G., Wojtyniak, B., Sunyer, J., Bacharova, L., Schouten, J. P., Ponka, A., and Anderson, H. R.

(1997). Short-term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: Results from time series data from the APHEA project. *British Medical Journal,* 314, 1658–1663.

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* 5418, January.

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 .

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

Narayan, P. K. (2005). The saving and investment nexus for China: Evidence from cointegration tests. *Applied Economics*, 37(17), 1979-1990.

Or, Z. (2001). Exploring the effects of health care on mortality across OECD countries. *OECD Labour Market and Social Policy Occasional Papers*, No. 46, OECD Publishing.

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

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

Samet, J. M., Dominici, F., Curriero, F. C., Coursac, I., and Zeger, S. L. (2000). Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. *New England Journal of Medicine*, 343, 1742–1749.

Schmueli, A. (2004). Population health and income inequality: New evidence from Israeli time-series analysis. *International Epidemiological Association*, 33, 311-317.

Soumerai, S. B., McLaughlin T. J., Ross-Degnan D., Casteris C. S., Bollini P. (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.

Woodruff, T. J., Grillo, J., and Schoendorf, K. C. (1997). The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States. *Environmental Health Perspectives*, 105, 608–612.

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

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

Table 1 Lag order selection

	Without deterministic trends				With deterministic trends			
p	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 Scwarz 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.

Table 2 *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 $T = 40$.
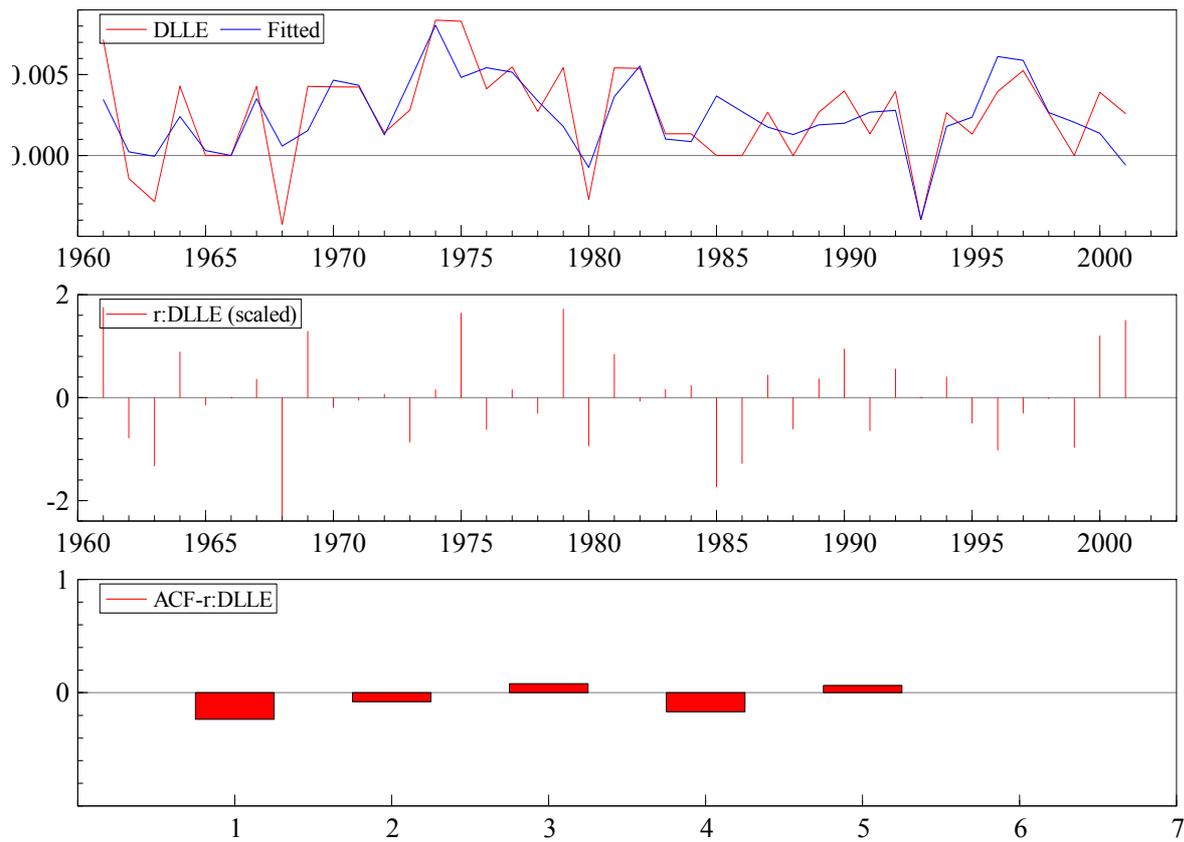
Figure 1: Actual and fitted values (the upper panel); Regression residuals (the middle panel);

Autocorrelation function of regression residuals (bottom panel).
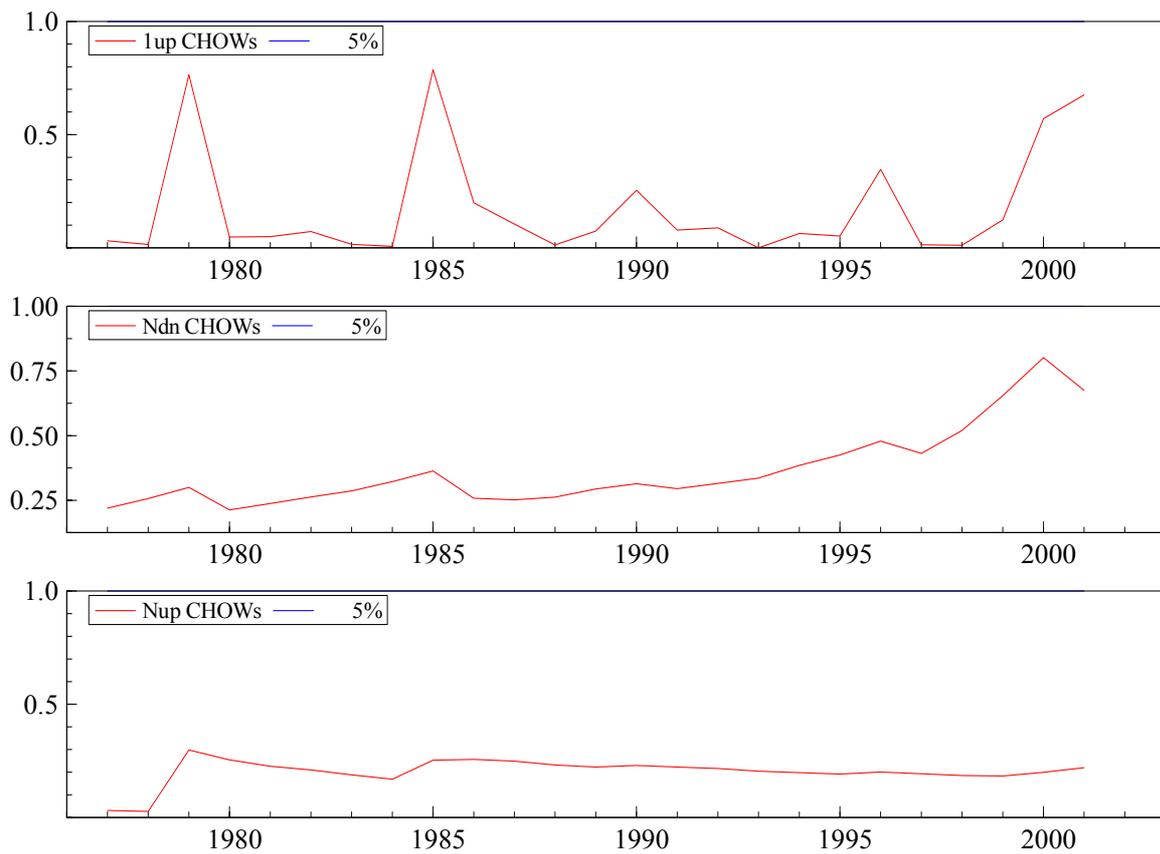
Figure 2: Recursive stability 1-step, breakpoint, and forecast Chow test statistics scaled by their respective 5% critical values.

Working Papers of the Socioeconomic Institute at the University of Zurich

0903	Spurious correlation in estimation of the health production function: A note, Sule Akkoyunlu, Frank R. Lichtenberg, Boriss Siliverstovs, Peter Zweifel, February 2009, 13 p.
0902	Making Sense of Non-Binding Retail-Price Recommendations, Stefan Bühler, Dennis L. Gärtner, February 2009, 30 p.
0901	Flat-of-the-Curve Medicine – A New Perspective on the Production of Health, Johnnes Schoder, Peter Zweifel, January 2009, 35 p.
0816	Relative status and satisfaction, Stefan Boes, Kevin E. Staub, Rainer Winkelmann, December 2008, 11 p.
0815	Delay and Deservingness after Winning the Lottery, Andrew J. Oswald, Rainer Winkelmann, December 2008, 29 p.
0814	Competitive Markets without Commitment, Nick Netzer, Florian Scheuer, November 2008, 65 p.
0813	Scope of Electricity Efficiency Improvement in Switzerland until 2035, Boris Krey, October 2008, 25 p.
0812	Efficient Electricity Portfolios for the United States and Switzerland: An Investor View, Boris Krey, Peter Zweifel, October 2008, 26 p.
0811	A welfare analysis of "junk" information and spam filters; Josef Falkinger, October 2008, 33 p.
0810	Why does the amount of income redistribution differ between United States and Europe? The Janus face of Switzerland; Sule Akkoyunlu, Ilja Neustadt, Peter Zweifel, September 2008, 32 p.
0809	Promoting Renewable Electricity Generation in Imperfect Markets: Price vs. Quantity Policies; Reinhard Madlener, Weiyu Gao, Ilja Neustadt, Peter Zweifel, July 2008, 34p.
0808	Is there a U-shaped Relation between Competition and Investment? Dario Sacco, July 2008, 26p.
0807	Competition and Innovation: An Experimental Investigation, May 2008, 20 p.
0806	All-Pay Auctions with Negative Prize Externalities: Theory and Experimental Evidence, May 2008, 31 p.
0805	Between Agora and Shopping Mall, Josef Falkinger, May 2008, 31 p.
0804	Provision of Public Goods in a Federalist Country: Tiebout Competition, Fiscal Equalization, and Incentives for Efficiency in Switzerland, Philippe Widmer, Peter Zweifel, April 2008, 22 p.
0803	Stochastic Expected Utility and Prospect Theory in a Horse Race: A Finite Mixture Approach, Adrian Bruhin, March 2008, 25 p.
0802	The effect of trade openness on optimal government size under endogenous firm entry, Sandra Hanslin, March 2008, 31 p.
0801	Managed Care Konzepte und Lösungsansätze – Ein internationaler Vergleich aus schweizerischer Sicht, Johannes Schoder, Peter Zweifel, February 2008, 23 p.
0719	Why Bayes Rules: A Note on Bayesian vs. Classical Inference in Regime Switching Models, Dennis Gärtner, December 2007, 8 p.
0718	Monoplistic Screening under Learning by Doing, Dennis Gärtner, December 2007, 29 p.
0717	An analysis of the Swiss vote on the use of genetically modified crops, Felix Schläpfer, November 2007, 23 p.