ECONSTOR
Make Your Publications Visible.

A Service of

ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Herwartz, Helmut; Theilen, Bernd

Working Paper

The determinants of health care expenditure: Testing pooling restrictions in small samples

SFB 373 Discussion Paper, No. 2000,78

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Herwartz, Helmut; Theilen, Bernd (2000) : The determinants of health care expenditure: Testing pooling restrictions in small samples, SFB 373 Discussion Paper, No. 2000,78, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10048048

This Version is available at:
https://hdl.handle.net/10419/62216

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

The determinants of health care expenditure: Testing pooling restrictions in small samples*

Helmut Herwartz
Institute for Statistics and Econometrics, Sonderforschungsbereich 373
Humboldt University Berlin, †

Bernd Theilen
Depart. de Economia
Universitat Rovira i Virgili ‡

July 2000

Abstract

Health care expenditure has increased substantially in all western industrialized countries in the last decades. The necessity to contain the increase in health care expenditure has motivated the analysis of its determinants to explain differences across countries and health systems. However, recent studies have questioned the use of cross–section data arguing that health systems are too different to allow for such comparisons. In this paper we investigate whether this criticism is really justified. We analyze the variations of health care expenditure in OECD countries relative to income, population aging and technological change. Our analysis is based on pooled cross-section data and time series. Firstly, formulating an error correction model for individual countries we demonstrate that for almost all of the countries investigated the variables are cointegrated. Secondly, we use a bootstrap framework for inference and examine whether the influence of the explanatory variables is unique across countries. Applying recursive estimation procedures we find that there is evidence for a common influence in the period 1961 to 1979 while in the last two decades explanatory variables have different impacts on health care expenditure across countries. This indicates the divergence of health systems and the growing importance of country-specific effects in the explanation of differences in health care expenditure across countries.

Journal of Economic Literature Classification Numbers: I11, C22, C23.
Keywords: Health care expenditure; Cointegration; OECD Panel Data.

*We gratefully acknowledge the comments participants of the EEA99 congress in Santiago de Compostela and the XXIV Simposium de Análisis Económico in Barcelona.

†Spandauer Str. 1, D-10178 Berlin, e-mail: helmut@wiwi.hu-berlin.de, Tel. (030) 2093-5725, Fax. (030) 2093-5712, Financial support by the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

‡Avinguda de la Universitat 1, E-43204 Reus, e-mail: bt@fcee.urv.es, Tel. (977) 759800, Fax. (977) 759810.

1 Introduction

The relationship between national health care expenditure (HCE) and gross domestic product (GDP) has become the subject of numerous empirical studies. The growing interest in this field mirrors the importance of its implications for instance with respect to the role of institutional factors [1] or the integration process of European countries [2]. As another but related issue building on such a relationship one may regard the design of a fair health care financing system for different regions within a country [3]. While earlier studies were based on a cross–sectional approach with observations for OECD countries for particular years, more recently, the use of panel data has allowed to increase the sample size and thus the power of econometric procedures. A specific result often obtained from cross-section analyses is that the income elasticity of HCEs is larger than unity [1,4-8]. While these contributions differ in the inclusion of explanatory variables other than income, like public expenditure, age structure, relative price of health care, urbanization, and others, the investigated samples are rather small throughout. The available sample sizes are in no case greater than 25, thus, the power of statistical inference is presumably low.

Instead of cross-sectional analysis some authors have applied dynamic models to characterize HCEs. For instance, Murillo et al. [9] analyze the effects of price variations on HCE for 9 European countries. Murthy and Ukpolo [10] discuss the determinants of health spending for the US. Both investigations confirm an income elasticity larger than unity. However, due to short time spans of available data the low power argument given above still applies.

Numerous recent contributions provide estimates of the parameters of interest obtained from an analysis of time series panel data. In these studies cross–section data and time series are pooled. While the results of some authors [2,3,11-13], confirm an income elasticity exceeding unity, Blomqvist and Carter [14] and O'Connell [15] find an income elasticity less than one. In Blomqvist and Carter this is due to the inclusion of a trend variable which captures changes in medical technology, while O'Connell obtains this result by including country–specific age coefficients.

With the application of time series panel data two main problems arise. First, the series should be identified as stationary (possibly including a deterministic trend) or nonstationary. In the latter case it is of importance whether the investigated series exhibit a long run equilibrium relationship, i.e. whether or not the nonstationary variables cointegrate. Hansen and King [16] showed that variables like GDP or HCE are mostly nonstationary. In no case these two variables turned out to be cointegrated. McCoskey and Selden [17] test the unit root hypothesis within a system of pooled time series and find that GPD and HCE are stationary. However, Hansen and King [18] argue that this result may be due to outliers and heteroskedasticity. Blomqvist and Carter [14] confirm that GDP and HCE are nonstationary. In contrast to Hansen and King [16], however, they conclude that both variables are cointegrated around a linear trend.

In all previous contributions the issue of unit root testing is performed under the assumption of structural invariance of the underlying data generating processes. In practice of analyzing macroeconomic data, however, deterministic shifts for instance in growth rates are often observed. Perron [19] provides the first treatment of unit root testing assuming shifts in the deterministic component of a time series process. In the most simple framework the time point of a structural shift is assumed to be given exogeneously. Therefore taking account of structural breaks requires to estimate critical values of the popular ADF statistic by simulations. In this paper unit root tests are performed under the impression that the speed of economic growth during the last two decades differs considerably from the high growth rates typically observed in the 1960s and early 70s. Following this approach we draw rather similar conclusions concerning the order of integration of GDP and HCE. Even with respect to the employed population variable we find that this series is integrated of order one for a considerable number of OECD members.

Second, inference in pooled systems of single country models is often performed via two step procedures. For instance, Pedroni [20] proposes a test to infer against cointegration in heterogeneous panels based on first step OLS residuals. Blomqvist and Carter [14] employ a multi step procedure to test pooling restrictions in a system of 18 OECD members. Their results indicate that a common trend and a common income elasticity cannot be assumed across countries. In light of this result pooling time series and cross-sectional data appears to be inconvenient and conclusions from the comparison of different health care systems cannot be drawn.

The purpose of this paper is to revisit especially the latter result. We adopt inference procedures to test linear restrictions across single country equations which are equivalent to maximum likelihood inference under a specific set of assumptions. Often maximum likelihood methods are preferable to feasible two step procedures due to asymptotic efficiency which is often achieved by the former approach. To our knowledge maximum likelihood or equivalent methods are not yet well developed for the analysis of panels of time series data. Critical values for the employed likelihood ratio statistics are estimated by means of a particular bootstrap scheme, namely the wild bootstrap. Herwartz and Neumann [21] show that this approach allows efficient asymptotic inference even in presence of cross-sectional error correlation. In addition, the small sample performance of the bootstrap approach is shown to outperform first order asymptotic approximations. Heteroskedastic error distributions are also treated conveniently. Opposite to feasible two step procedures, as e.g. GLS methods, the advocated approach does not require any parametric representation of cross correlation or heteroskedasticity.

We investigate a system of 19 OECD countries for the sample period 1960-1997. For each economy we analyze the dynamic relationship between HCE, GDP, and the proportion of the population aged 65 and older (P65) by means of single equation error correction models (ECM). Furthermore, following Blomqvist and Carter [14] we include a deterministic trend as an explanatory variable which may account for cost increasing technical progress.

We discuss whether key parameters of the model are equal across countries, implying that pooling time series and cross-sectional data is appropriate in the context of international comparisons of HCEs. In addition, we test if there is evidence for an income elasticity exceeding unity, expenditure increasing technical change, or a positive influence of population aging. Finally, applying recursive estimation we analyze whether health care systems are converging or not.

2 Data and Methodology

The methodology applied is different to former studies in two aspects: First, we use a model combining the approaches of Blomqvist and Carter [14] and O'Connell [15]. Apart from real per capita income we include a deterministic trend to model HCE growth rates following Blomqvist and Carter [14]. As O'Connell [15] we augment our empirical model with the proportion of the population aged 65 and older. Second, we employ a pooled system of ECMs allowing equivalence between OLS and ML inference. Hypotheses of interest are tested by means of a likelihood-ratio (LR-) statistic. Critical values for this statistic are estimated via a bootstrap procedure which copes conveniently with heteroskedastic error distributions and cross sectional correlation.

2.1 The data

Our data source is OECD health data [22]. The data file contains information for all OECD member states. Greece, Luxemburg, New Zealand, Portugal, and Turkey are are excluded from the analysis since available time series for these countries cover a considerably smaller period compared to the remaining countries. The time series include annual data from 1960-1997. With respect to Denmark the investigated sample period is

1969-1997, since a few data points for the 1960s are not available for this country. We use the following variables for the empirical analysis: Health care expenditure (HCE), gross domestic product (GDP) and the percentage of the population aged 65 and over (P65). HCE and GDP are per capita measures at purchasing power parity values. Both variables are transformed into real values using the US GDP deflator (1995=100). All variables are mean–adjusted, i.e. we subtract from an observation for year t the cross sectional mean obtained for this particular period. The latter adjustment is applied to eliminate shocks that influence all countries in the same way, like world wide recessions or booms. An alternative would be to include time specific dummy variables as in O'Connell [15]. As Carter and Blomqvist [14] we include a deterministic trend to account for technological change which is expected to be cost increasing in the health care sector. Following most related empirical investigations we analyze all series in natural logarithms.

2.2 The econometric model

We formalize the following dynamic regression model:

$$\begin{aligned} \Delta \mathrm{HCE}_{n,t} &= \nu_n + \delta_n t/100 + \alpha_n \left(\mathrm{HCE}_{n,t-1} - \beta_{1n}\mathrm{GDP}_{n,t-1} - \beta_{2n}\mathrm{P65}_{n,t-1}\right) \\ &\quad + \gamma_{1n}\Delta\mathrm{GDP}_{n,t} + \gamma_{2n}\Delta\mathrm{P65}_{n,t} + \gamma_{3n}\Delta\mathrm{HCE}_{n,t-1} + u_{n,t}, \end{aligned} \tag{1}$$

where $n = 1, \ldots, N$ indicates the country and $t = 1, \ldots, T$ the time period. $u_{n,t}$ is a homoskedastic uncorrelated error sequence, i.e. $u_{n,t} \sim (0, \sigma_n^2)$. The parameters $\gamma_{jn}, j = 1, 2, 3$ govern short run dynamics of $\Delta\mathrm{HCE}_{n,t}$. Presample values necessary to estimate the model (1) are assumed to be given. The error correction coefficient α_n determines how current HCE reacts to lagged violations of the long run equilibrium relation, $eq_{n,t} = \text{HCE}_{n,t} - \beta_{1n}\text{GDP}_{n,t} - \beta_{2n}\text{P65}_{n,t}$. Note that α_n should be negative since we expect, for instance, positive equilibrium errors in $t-1$ to cause decreasing HCE in period t. The formulation in (1) allows simultaneous estimation of the long–run or cointegrating relationship, and short–run dynamics. In addition, the model is sensible to test the null hypothesis of cointegration between HCE, GDP and P65.

Estimation and inference within a single equation error correction model like (1) is asymptotically efficient and thus equivalent to (higher dimensional) full information maximum likelihood procedures if the following assumptions can be made [23]. First, $\text{HCE}_{n,t}$, $\text{GDP}_{n,t}$, and $\text{P65}_{n,t}$ have to be integrated of order one and cointegrated with cointegrating rank 1, i.e. there exists one and only one linear combination of the variables obtaining a stationary series. Second, the specification in (1) provides a valid conditioning scheme for $\Delta\text{HCE}_{n,t}$ in the sense that $u_{n,t}$ is a martingale difference sequence (mds), i.e. $E[u_{n,t}|u_{n,t-1},\ldots,u_{n,1}] = 0$. To guarantee the mds property of $u_{n,t}$ it may be necessary to augment the set of explanatory variables conveniently with further (lagged) stationary time series. Finally, $\text{GDP}_{n,t}$ and $\text{P65}_{n,t}$ are weakly exogenous for estimation of and inference on the long run parameters.

Under these assumptions the $t-$ratio of $\hat{\alpha}_n$ (t_{α_n}) obtained from OLS-routines is asymptotically normally distributed. If the variables are not cointegrated the error correction term, $eq_{n,t}$, is nonstationary and the regression in (1) is "unbalanced". In this case α_n should be zero. Therefore a test on significance of α_n is implicitly also a test of the null hypothesis of cointegration. Under the alternative of no cointegration, however, the $t-$ratio of $\hat{\alpha}_n$ is not asymptotically normally distributed. Kremers et al. [24] show that the distribution of t_{α_n} is somewhere between the distribution of the ADF t-statistic and the N(0,1)–distribution if $eq_{n,t}$ is nonstationary.

Furthermore, test-statistics on linear restrictions concerning the long run parameters β_{1n} and β_{2n} can be shown to follow their standard asymptotic distributions. In the following we briefly provide the LR–test since it allows easily to perform pooled tests if the model in (1) is just the $n-$th equation within a N-dimensional system of single equation models.

Defining a vector of dependent variables, $y_n = (\Delta\mathrm{HCE}_{n,1}, \Delta\mathrm{HCE}_{n,2}, \ldots, \Delta\mathrm{HCE}_{n,T})'$ convenient design matrices X_n^0 and X_n^1 can be given to represent the model in (1) under the null and the alternative hypothesis as follows:

$$y_n = X_n^0 \delta_n^0 + u_n^0 \text{ and } y_n = X_n^1 \delta_n^1 + u_n^1. \tag{2}$$

The parameters in δ_n^0 match the restrictions imposed by a specific null hypothesis. δ_n^1 contains the coefficients of the unrestricted linear model. OLS–routines applied to the two alternative systems in (2) provide estimates of the residual sums of estimated squared error terms under the null and alternative hypothesis, respectively. The LR-statistic is T-times the log-ratio of the maximum likelihood variance estimators obtained under the two competing hypotheses, i.e.

$$\mathrm{LR}_n = T \ln \left(\frac{\tilde{\sigma}_{n0}^2}{\tilde{\sigma}_{n1}^2} \right), \tag{3}$$

where

$$\tilde{\sigma}^2_{n0} = \frac{1}{T}\hat{u}_n^{0\prime}\hat{u}_n^0,\ \tilde{\sigma}^2_{n1} = \frac{1}{T}\hat{u}_n^{1\prime}\hat{u}_n^1.$$

Under the null hypothesis LR_n is asymptotically $\chi^2(q)$ distributed with q denoting the number of restrictions imposed under the null hypothesis [25].

Now suppose the model in (1) to be the $n-$th equation in a cross sectional model of dimension N. To test a particular null hypothesis to hold in the pooled system it appears natural to apply the pooled LR-statistic,

$$\mathrm{LR}_p = \sum_{n=1}^{N} \mathrm{LR}_n. \tag{4}$$

The LR-statistic in (4) is provided by Lütkepohl [25] in a related context. By construction of the test, the error terms $u_{n,t}$ are allowed to be heteroskedastic across equations. LR_p is $\chi^2(qN)$ distributed under the null hypothesis if the single equation error terms are contemporaneously uncorrelated across different members of the pooled system. In presence of cross correlation the LR_p-statistic is not pivotal. For this case Herwartz and Neumann [21] show that the distribution of pooled LR-statistics can be conveniently evaluated by means of a particular bootstrap scheme, namely the wild bootstrap. In addition, the bootstrap approach is shown to have considerably better size properties in small samples even if the error terms $u_{n,t}$ are not contemporaneously correlated. Similarly, bootstrap inference turns out to outperform common asymptotic inference in single equation models like (1). Initially introduced by Wu [26] to account for heteroskedastic error processes the wild bootstrap scheme is also preferable to standard asymptotic inference if error processes have the mds property but fail to be independent and identically distributed. Since the reader may not be that familiar with this particular resampling scheme we give a brief description of the wild bootstrap below.

2.3 The wild bootstrap

As it is typical for bootstrap procedures the wild bootstrap is performed to mimic the distribution of a statistic of interest by resampling error terms u_n^* from their estimated counterparts, $\hat{u}_n^0$ say. Applying the wild bootstrap single elements of u_n^* are obtained by sampling from the corresponding elements of $\hat{u}_n^0$. A particular feature of this resampling scheme is that error terms u_{nt}^* match the low order moments of $\hat{u}_{nt}^0$. Given OLS-error estimates ($\hat{u}_{nt}^0$) a candidate procedure to obtain u_{nt}^* looks as follows [27]:

1. Generate standard normal and independent random variables V_t and W_t and compute $Z_t = V_t/\sqrt{2} + (W_t^2 - 1)/2$. By definition low order moments of this random variable are $E[Z_t] = 0$, $E[Z_t^2] = 1$, and $E[Z_t^3] = 1$.

2. Multiplying Z_t with $\hat{u}_{nt}^0$, one obtains random variates $u_{nt}^* = \hat{u}_{nt}^0 Z_t$ exhibiting the low order moments of $\hat{u}_{nt}^0$, i.e.

$$E[u_{nt}^*] = 0, \quad E[u_{nt}^*]^2 = (\hat{u}_{nt}^0)^2, \quad E[u_{nt}^*]^3 = (\hat{u}_{nt}^0)^3.$$

As provided above Z_t does not depend on n. Applying the same random variables Z_t to obtain error terms $\hat{u}_{nt}^0$, $n = 1, \ldots, N$, the former scheme also accounts for cross correlation, i.e.

$$E[u_{it}^* u_{jt}^*] = \hat{u}_{it}^0 \hat{u}_{jt}^0, \quad i, j = 1, \ldots, N, \quad i \neq j.$$

As an alternative approach to account for cross correlation of panel data one may regard feasible GLS methods. The advocated bootstrap scheme, however, is preferable to feasible GLS since it does not require any (estimated) parametric formalization of the cross correlation structure. In addition, the bootstrap scheme also copes with time varying patterns of cross equation correlation.

Having defined a sampling scheme for particular error variables we briefly sketch the entire procedure to estimate critical values for the LR_n or LR_p statistic for completeness.

1. Estimate the linear model(s) in (2) and obtain the test statistic of interest LR_n (LR_p), and estimates δ_n^0 and $\hat{u}_n^0$.

2. Compute a wild bootstrap sample as

$$y_n^* = X_n^0 \hat{\delta}_n^0 + u_n^*,$$

   where single error terms u_{nt}^* are generated as outlined above.

3. Compute a variant of LR_n (LR_p) from y_n^* and the initial design matrices X_n^0 and X_n^1.

4. Iterate R times on steps (2) and (3), $R = 1000$, say. The variants of the test statistics are recorded in a $R \times 1$ vector LR_n^* (LR_p^*).

5. The considered null hypothesis is rejected with significance level π if LR_n (LR_p) exceeds the $(1-\pi)$-quantile of LR_n^* (LR_p^*).

Note that it is not necessary for the wild bootstrap to generate the involved time series given in (1). Test-statistics of interest are resampled conditional on the initial realizations of explanatory variables. Thus it is not necessary to make specific assumptions on the data generating processes of $\mathrm{GDP}_{n,t}$ or $\mathrm{P65}_{n,t}$ for the present analysis.

3 Results

The discussion of our empirical results addresses three issues: First, we briefly discuss the results from unit root tests performed for the investigated time series. The second issue is to investigate the hypothesis of cointegration by means of single equation results. In addition, pooled estimates are used to infer against the cointegration assumption. Finally, given that the variables under study are cointegrated we discuss whether key parameters governing equilibrium relations between the series can reasonably be assumed to be unique within the pooled system.

3.1 Inference on unit roots

Table 1 displays the results of augmented Dickey-Fuller (ADF) tests for single country series. An intercept term is included in all ADF regressions. Additionally, if indicated by the BIC we also include a deterministic drift term [25]. The regression order varies to include lagged differences of order $k = 0, \ldots, 5$. The BIC is also used to determine the lag order. As a particular feature of the GDP and HCE series we observe that these measures exhibit different growth patterns over time. For almost all countries we find that GDP and HCE growth rates are considerably smaller in the last two decades of the sample compared to the period 1961-1979. Interestingly, a similar result also holds with respect to our population variable (P65) for the majority of considered countries. To account for a time varying drift term in the ADF regressions we include an additional dummy variable to indicate observations measured between 1980 and 1997. Critical values for the ADF-statistic are obtained from own simulations of ADF regressions (with a deterministic shift of the intercept term) applied to a random walk process.

With respect to GDP and HCE we find almost unique evidence for these series to be integrated of order one. The unit root hypothesis is not rejected for any GDP level series at the 10% significance level. For growth rates, however, the unit root hypothesis is rejected for 15 (11) countries at the 10% (5%) level. The Danish HCE series is the only health care variable that we find to be stationary at the 5% level. Investigating first differences of HCE (ΔHCE) we reject the unit root hypothesis with 10% significance for all countries except Canada and the US. At the 5% level we still obtain 13 rejections of the unit root hypothesis for ΔHCE.

For the P65 series our results are not that robust across countries. For instance, P65 appears to be stationary for Italy since we reject the unit root hypothesis for levels and first differences of this variable. For 8 countries, however, we find the employed population variable to be integrated of order one, since the unit root hypothesis cannot be rejected with 10% signficance for the level series whereas it is rejected for the growth rates.

Our results essentially confirm Blomqvist and Carter [14] who infer on unit roots in GDP and HCE for the period 1960 to 1991 using the (nonparametric) Phillips Perron [28] test. Applying (parametric) ADF type tests as adopted in this paper Hansen and King [16] find considerably less support for both variables to be integrated of order one within samples covering the period 1960 to 1987. Since parametric test procedures are easily implemented using common model selection criteria we prefer the ADF testing procedure. It is worth mentioning that our results would come closer to the Hansen and King [16] study if a structural shift in empirical growth rates is ruled out a priori.

3.2 Testing for cointegration

Table 2 displays estimation results for single country error correction models. In addition we show the Ljung-Box statistic against joint error autocorrelation up to order 20. The latter diagnostic statistics indicate the potential of serial error correlation at the 5% significance level for 3 of the 19 estimated equations. As mentioned before, a possible strategy to improve the adopted empirical specifications is to augment the employed set of explanatory variables by additional lagged differences of the investigated time series. We refrain from such a strategy to keep the set of explanatory variables identical across equations. We further conjecture that the actual sample size is hardly large enough to obtain valid critical values for the Ljung Box statistic from the asymptotic distribution. In effect we consider the underlying error terms to have the mds property.

We already pointed out that the ECM in (1) is unbalanced if the right hand side level variables do not cointegrate. In this case α_n must be zero. Thus cointegration is indicated if the null hypothesis $H_0 : \alpha_n \geq 0$ is rejected against $H_1 : \alpha_n < 0$. All estimated error correction coefficients $(\hat{\alpha}_n)$ are less than zero. Regarding deviations from the equilibrium $(eq_{n,t})$ to be stationary the latter null hypothesis is rejected in all but three (France, Japan and Sweden) single equation models at the 5% significance level.

Turning to the pooled system of dynamic models one may think of a common error correction coefficient α to govern the adjustment of equilibrium errors in all members of the pool. Now, each estimate $\hat{\alpha}_n$ may be interpreted as an estimator of α. We take the mean of estimates $\hat{\alpha}_n$ as an estimator of α, i.e.

$$\hat{\alpha} = \frac{1}{N} \sum_{n=1}^{N} \alpha_n.$$

For the pooled system we obtain $\hat{\alpha} = -0.31$. To provide a $t-$ratio for this quantity one may regard single estimates $\hat{\alpha}_n$ to be independent with common variance. Alternatively, one may compute a $t-$ratio using the empirical variances of $\hat{\alpha}_n$. Irrespective of the adopted procedure we reject $H_0 : \alpha \geq 0$ against $H_1 : \alpha < 0$ at any reasonable significance level.

Summarizing our results we conclude that there exists a long run equilibrium relating HCE, GDP and P65. These results coincide with the results of Blomqvist and Carter [14], who used the Shine [29] test in the context of a dynamic regression model. They found that cointegration holds for almost all of the investigated countries. However, our results are in contrast to Hansen and King [16]. Using the Engle-Granger test Hansen and King [16] found no evidence in favor of cointegration for 17 of 20 countries under consideration. The latter finding, however, may be addressed to the weak performance of a two step, residual based approach in small samples.

An alternative procedure to infer against the cointegration hypothesis in panel data would be to apply a two step residual based procedure introduced by Pedroni [20]. Such a device is preferable in dynamic systems with more than one cointegrating relationship. As pointed out before, single equation ECMs provide an asymptotically efficient framework for estimation and inference in our case. Therefore we refrain from testing panel cointegration by means of residuals obtained from static regressions of nonstationary variables.

3.3 Investigating long run equilibrium relations

Regarding the investigated time series variables to be cointegrated the ECM in (1) provides an asymptotically efficient means to estimate the parameters governing the long run equilibrium relationship between HCE, GDP, and P65. Due to the inclusion of a deterministic trend variable which accounts for technical progress HCEs may not increase with income throughout. In 15 of 19 empirical single country models the estimator $\hat{\beta}_{1n}$ has the expected sign, for six economies we find this coefficient to be significant at the 5% level. For Germany and Sweden we obtain insignificant but rather large estimates $\hat{\beta}_{1n}$. With respect to Germany the empirical results are essentially the same if a dummy variable is added to the model to account for the post German unification period. Note that the $t-$ratios provided in Table 2 allow for efficient inference only asymptotically. For the latter argument it is interesting to perform the empirical analysis on the level of pooled equations.

Applying OLS to a pooled system of equations we obtain the following restricted ECM ($t-$ratios in parentheses):

$$\begin{aligned}\Delta\mathrm{HCE}_{n,t} &= \hat{\nu}_n + \underset{(0.04)}{6.1\text{E-}04}t/100 - \underset{(9.04)}{.138}\left(\mathrm{HCE}_{n,t-1} - \underset{(5.65)}{.740}\,\mathrm{GDP}_{n,t-1} + \underset{(0.63)}{.095}\,\mathrm{P65}_{n,t-1}\right)\\ &+ \underset{(6.52)}{.437}\,\Delta\mathrm{GDP}_{n,t} - \underset{(0.02)}{.003}\,\Delta\mathrm{P65}_{n,t} + \underset{(1.99)}{.069}\,\Delta\mathrm{HCE}_{n,t-1} + \hat{u}_{n,t}. \end{aligned} \tag{5}$$

The intercept terms are unrestricted to allow for country-specific effects. Estimated parameters are reported in Table 3 as deviations from $\hat{\nu}_{\mathrm{US}}$ which we regard as a benchmark estimate. The estimates given in (5) provide further support of the cointegration hypothesis, since the OLS t-ratio of $\hat{\alpha}$ is significant at any reasonable level. The estimated income elasticity is significant and less than unity. With respect to the trend and to the impact of P65 we do not obtain significant estimates for the pooled model.

From the results of the single country models given in Table 2 we obtain that the P65-elasticity of HCE is significant for 8 (6) countries at the 10% (5%) level. Therefore we conjecture that the population variable has country-specific impacts. A similar argument applies to the deterministic trend parameter which is found to be significant within 3 (2) country specific equations at the 10% (5%) significance level. In equation (5) $\hat{\delta}$ has the expected positive sign but fails to be significant questioning to some extent the argument of Blomqvist and Carter [14] that one of the most important determinants of the growth in HCE is the advent of more costly technology. Note, however, that the $t-$ratio of $\hat{\delta}$ may actually be biased in presence of cross equation correlation and multicollinearity.

Applying feasible GLS-methods might provide slightly different $t-$ratios or parameter estimates. In effect a LR-type statistic is more convenient to shed light on the role of technical progress in determining health care spending. Therefore we turn now to LR-inference on the level of pooled equations that can be shown to be asymptotically efficient without requiring a parametric estimate of the pattern of contemporaneous correlation.

Regarding the restricted parameter estimates in (5) we test the following hypotheses to hold on the pooled level: $H_0^1 : \delta = 6.1\text{E-}04$, $H_0^2 : \beta_1 = 0.740$ and $H_0^3 : \beta_2 = -0.095$. The corresponding LR-statistics are denoted as LR^1 to LR^3. We also consider tests of the following (unconditional) hypotheses: $H_0^4 : \delta = 0$, $H_0^5 : \beta_1 = 1$ and $H_0^6 : \beta_2 = 0$ obtaining the LR-statistics LR^4 to LR^6. Since we allow for country specific deterministic effects throughout we refrain from testing any restriction concerning ν_n. Critical values for all LR-statistics are estimated by means of the wild bootstrap. Test results are displayed in Table 4.

The obtained LR_p statistics vary between 41.16 ($H_0^1 : \delta = 6.1\text{E-}04$) and 44.07 ($H_0^6 : \beta_2 = 0$). Taking critical values from a $\chi^2(19)$ distribution all hypotheses are rejected at least with 0.22% significance. The latter distribution, however, provides asymptotically valid critical values only in the case of zero cross correlation. Estimating critical values by means of the bootstrap we obtain that both hypotheses involving $\hat{\beta}_2$ cannot be rejected at the 5% significance level within the pooled system. Both hypotheses concerning the trend parameter δ and the income elasticity of HCE are rejected at the 5% level. It is interesting to note that the hypothesis of a unit income elasticity of HCE involves a considerably higher marginal significance level compared to the $H_0^2 : \beta_1 = 0.740$. This result may be regarded as evidence in favor of an income elasticity of HCE close to or even larger than unity.

Testing pooling restrictions concerning the trend parameter supports the argument raised by Blomqvist and Carter [14] that technical change has a positive impact on HCE growth. They adress this to the fact that technical change in the production of health takes the form of progress in the ability to transform health services into 'good health' rather than in lowering the effective production cost. Note however that the magnitude of this impact may still be country specific since pooling restrictions concerning δ are clearly rejected at the 5% level. Therefore though technical change is an important determinant of the growth in the cost of health care in all OECD countries institutional differences are strongly related to the form in which new medical technology is introduced.

As discussed above our empirical results are obtained from evaluating a sample period ending in 1997. To address issues like convergence of health care systems among OECD countries it appears reasonable to perform the foregoing inference exercises recursively. Doing so we start with a basic sample period ending 1981 which we extent iteratively until 1997. For each recursive sample we estimate a restricted ECM as given in (5) and perform hypothesis tests LR^1 through LR^6.

Figure 1 displays marginal significance levels obtained recursively for the six test statistics given above. Obviously and similar to the results obtained for the entire sample both hypotheses concerning a particular parameter of interest involve very similar p-values. In addition, the time paths of the p-values show a similar pattern for all three parameters of interest. In the period 1981 to 1987 p-values decrease almost uniformly for all investigated hypotheses indicating that OECD health care systems failed to converge in this period. For samples ending in the late 1980s the income elasticity of HCE differs significantly at the 5% level across OECD members. In addition, the hypothesis of a joint unit income elasticity is rejected. Samples ending in the early 1990s support the assumption of converging health care systems since during this period recursive p-values obtained for all hypothesis tests of interest show an increasing pattern. However, samples ending later than 1994 again revert the latter result.

This result suggests that the introduction of new medical technology, the evolution of the population structure and income development had similar influences on HCE in different countries during the decades 60 and 70. During the last two decades the influence of these variables on HCE differs across countries. During this period HCE in most countries was submitted to much stronger cost control. Then the questions of how and which medical technology is introduced, and how the access to this technology is regulated become more important, that is, the structure of health systems becomes particularly relevant. Therefore we would expect variables like P65 and GDP to be closer related to health care system-specific effects like the importance of rationing by waiting lists, the existence and extend of private and public health insurance and health care provision. The results in Figure 1 support this view. Actual differences in HCE across countries must be explained much more by country-specific effects than by differences in income or in the age structure of the population.

The results displayed in Figure 1 indicate that the sample period investigated has considerable influence on the conclusions to be drawn. This allows us to integrate results of former studies. For the sample period of our study we confirm the result of Blomqvist and Carter [14] that there is considerable evidence against the hypotheses of an equal income elasticity and a common trend. However, for the sample period 1960-1991 investigated in their study we cannot reject the first hypothesis at the 5 percent level while the second one is rejected. Though the employed samples differ somewhat with respect to the countries included we address this difference to the small sample performance of the multistep estimation procedure followed by Blomqvist and Carter [14]. Additionally, the results show that contradictory results of former studies which apply cross-section data and use different base years may be due to the divergence of health care systems. The underlying assumption of a common income elasticity in these contributions is especially justified for those with a base year in the early 1980s like in the one of Parkin et al. [6].

4 Conclusions

The starting point of our analysis is the efficient modulation of the long-run relationship between HCE, GDP, and a population variable for 19 OECD countries. We investigate whether the involved variables are integrated of order one or stationary. Applying ADF tests that allow for a structural shift in the deterministic drift term we find almost unique evidence for stationarity in first differences for these variables and thereby confirm the results Blomqvist and Carter [14] who apply nonparametric unit root tests. Formulating an error correction model, we show that variables are cointegrated. This indicates that a long–run cointegration relationship between HCE, GDP and P65 exists. We thereby refuse the first criticism on using pooled cross-section data and time series stating that the variables of our model are neither stationary nor cointegrated.

We apply methods of statistical inference which take account of cross-sectional error correlation and heteroskedasticity and which perform particularly well in small samples. We find that dynamic relationships governing HCE may fail to be constant across countries. Applying recursive estimation we show that this result is sensitive to the analyzed sample period. For the period 1960-1981 the hypothesis of a common trend, income elasticity, and influence of population aging on HCE across countries cannot be rejected. During this period HCE displays (on average) higher growth rates than GDP. Therefore the argument of Blomqvist and Carter [14] that technical change is responsible for cost increases in health care may be valid for this period.

If the sample period is increased up to 1997 evidence for a constant influence of the variables of our model on HCE across countries is almost uniquely decreasing, and for the entire sample period the hypothesis of a common income elasticity and a common deterministic trend is rejected. We address this result to the increasing importance of institutional factors during the last two decades. Our findings suggest that these factors cannot be captured only by country-specific dummy variables but also influence the income elasticity and the impact of technological change on HCE. This casts doubt upon the view that mainly differences in income are responsible for differences in HCE across countries.

We attribute the growing divergence of health care systems to stronger efforts in cost containment in HCE with health care system-specific effects in western industrialized economies. The challenging future task is to analyze how this policy has influenced the quality and quantity of health care services in different countries and which health systems perform better in terms of cost-effectiveness.

References

1. Gerdtham U-G, Søgaard J, Andersson F, Jönsson B. An econometric analysis of health care expenditure: A cross–section study of the OECD countries, *Journal of Health Economics* 1992; **11**: 63-84.

2. Hitiris T. Health care expenditure and integration in the countries of the European Union, *Applied Economics* 1997; **29**: 1-6.

3. Theilen B. El gasto sanitario en los países de la C.E., *Working Paper*. Universitat Rovira i Virgili: Reus, 1998.

4. Newhouse JP. Medical-care expenditure: a cross–national survey, *Journal of Human Resources* 1977; **12**: 115-125.

5. Leu RE. The public–private mix and international health care costs. In *Public and Private Health Services*, Cluyer AJ. and Jönsson B (eds.). Basil Blackwell: Oxford, 1986; 41-63.

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

7. Gerdtham U-G, Jönsson B. Conversion factor instability in international comparisons of health care expenditure, *Journal of Health Economics* 1991; **10**: 227-234.

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

9. Murillo C, Piatecki C, Saez M. Health Care Expenditure and Income in Europe, *Health Economics* 1993; **2**: 127-138.

10. Murthy NRV, Ukpolo V. Aggregate health care expenditure in the United States, *Applied Economics* 1994; **26**: 797-802.

11. Gerdtham U-G. Pooling international health care expenditure data, *Health Economics* 1992, **1**: 217-231.

12. Hitiris T, Posnett J. The determinants and effects of health expenditure in developed countries, *Journal of Health Economics* 1992; **11**: 173-181.

13. Søgaard J. Functional form of the aggregate cross-section relationship between income and health care expenditure, *Health Expenditure Anal. Lett.* Summer 1992; 4-8.

14. Blomqvist ÅG, Carter RAL. Is health care really a luxury?, *Journal of Health Economics* 1997; **16**: 207-229.

15. O'Connell JM. The relationship between health expenditure and the age structure of the population in OECD countries, *Health Economics* 1996; **5**: 573-578.

16. Hansen P, King A. The determinants of health care expenditure: A cointegration approach, *Journal of Health Economics* 1996; **15**: 127-137.

17. McCoskey SK, Selden TM. Health care expenditure and GDP: panel data unit root test results, *Journal of Health Economics* 1998, **17**: 369-376.

18. Hansen P, King A. Health care expenditure and GDP: panel data unit root test results–comment, *Journal of Health Economics* 1998; **17**: 377-381.

19. Perron P. The great crash, the oil price shock and the unit root hypothesis, *Econometrica* 1989; **57**: 1361-1401.

20. Pedroni P. Critical values for Cointegration tests in heterogeneous panels with multiple regressors, *Oxford Bulletin of Economics and Statistics* 1998, forthcoming.

21. Herwartz H, Neumann M. Bootstrap Inference in Single Equation Error Correction Models, *mimeo*. Humboldt Universität: Berlin, 2000.

22. OECD. *OECD health data. Comparative analysis of health systems.* CREDES, OECD: Paris, 1999.

23. Banerjee A, Dolado JD, Galbraith JW, Hendry DF. *Co-Integration, Error Correction, And the Econometric Analysis of Non-Stationary Data*, Oxford University Press: Oxford, 1993.

24. Kremers JJM, Ericsson NR, Dolado JJ. The Power of Cointegration Tests, *Oxford bulletin of Economics and Statistics* 1992; **54**: 325–348.

25. Lütkepohl H. *Introduction to Multiple Time Series Analysis.* Springer Verlag: Berlin, 1991.

26. Wu CFJ. Jackknife, Bootstrap, and Other Resampling Methods in Regression Analysis (with discussion), *Annals of Statistics* 1986; **14**: 1261–1295.

27. Mammen E. Bootstrap and Wild Bootstrap for High Dimensional Linear Models, *The Annals of Statistics* 1993; **21**: 255–285.

28. Phillips PCB, Perron P. Testing for a unit root in time series regression, *Biometrika* 1988; **75**: 335-346.

29. Shine Y. A residual-based test of the null of cointegration against the alternative of no cointegration, *Econometric Theory* 1994; **10**: 91-115.

	HCE			GDP			P65		
	ADF	TR	k	ADF	TR	k	ADF	TR	k
Country	Level Series								
Australia	-0.925	0	0	-1.244	0	0	-0.483	0	2
Austria	0.575	0	0	-0.161	0	1	-2.515	0	4
Belgium	1.824	1	0	-0.782	0	1	-6.555^c	1	2
Canada	-2.132	0	1	-0.222	0	0	0.567	0	3
Denmark	-3.807^b	0	0	-2.432	0	0	0.519	1	0
Finland	-2.727	0	3	-0.265	0	1	-0.971	0	1
France	-1.610	0	0	-0.742	0	1	-1.015	0	2
Germany	-1.132	0	0	-0.372	0	0	-3.251	1	3
Iceland	-0.943	0	0	0.206	0	1	-4.492^b	1	0
Ireland	-0.649	0	0	1.095	0	0	-1.490	0	1
Italy	-2.998	0	0	1.745	1	0	-5.563^c	1	2
Japan	-1.859	0	0	-2.638	0	1	2.585	0	1
Netherlands	-2.365	0	1	-0.015	0	0	-0.887	0	0
Norway	-1.530	0	0	0.250	0	1	3.411	1	0
Spain	-1.974	0	1	-1.688	0	1	-1.800	1	2
Sweden	-1.523	0	1	-0.258	0	5	-2.756	0	2
Switzerland	-1.421	1	0	0.736	1	0	0.950	0	2
UK	-1.242	0	0	-0.220	0	0	0.947	1	0
USA	-1.828	0	1	-0.089	0	0	-0.555	0	5
	First Differences								
Australia	-4.994^c	0	0	-4.654^c	0	0	-0.870	0	1
Austria	-5.035^c	0	0	-3.819^b	0	0	-5.133^c	0	3
Belgium	-3.548^b	0	0	-3.902^b	0	0	-2.875	0	2
Canada	-2.448	0	0	-4.228^b	0	0	-0.981	0	2
Denmark	-3.933^b	0	0	-3.410^a	0	0	-1.286	0	0
Finland	-3.161^a	0	0	-3.200^a	0	0	-4.635^c	0	0
France	-3.799^b	0	0	-3.291^a	0	0	-1.601	0	0
Germany	-5.341^c	0	0	-5.361^c	0	0	-3.228^a	1	2
Iceland	-4.523^c	0	0	-3.478^b	0	0	-6.169^c	0	3
Ireland	-4.982^c	0	0	-6.053^c	0	0	-2.888	0	0
Italy	-3.547^b	0	0	-3.882^b	0	0	-3.198^a	0	3
Japan	-3.643^b	0	0	-2.941	0	0	-6.282^c	0	0
Netherlands	-3.288^a	0	0	-4.963^c	0	0	-5.604^c	0	0
Norway	-5.269^c	0	0	-3.178^a	0	0	-0.563	0	1
Spain	-3.160^a	0	0	-2.906	0	0	-4.250^b	0	4
Sweden	-3.397^a	0	0	-1.413	0	4	-0.817	0	1
Switzerland	-5.863^c	0	0	-3.538^b	0	0	-4.500^c	0	2
UK	-4.134^b	0	0	-2.742	0	1	-4.014^b	0	0
USA	-1.768	0	0	-4.151^b	0	4	0.631	0	4

Table 1: Augmented Dickey Fuller Tests for level series and first differences. TR=1 indicates inclusion of a deterministic trend in the test regression. Lag orders were alternatively $k = 0, .., 5$. Order selection and selection of deterministic terms by means of the BIC criterion. Critical values from own simulations using 10000 replications of a random walk series $x_t = x_{t-1} + u_t$, $u_t \sim N(0,1)$. a, b, c indicate significance at level $0.10, 0.05, 0.01$, respectively. For the simulation we specified the ADF test regressions with additonal dummy variables to allow the last 18 observations to exhibit a different drift compared to former observations.

Country	$\hat{\nu}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\sigma}$, Q(20)
Australia	-0.065	0.145	-0.273^{b}	1.851	-0.401	0.499^{a}	-0.841	0.019	0.029
	(-0.906)	(0.749)	(-2.294)	(1.533)	(-0.566)	(1.785)	(-0.896)	(0.114)	10.74
Austria	-0.052	0.098	-0.303^{b}	1.462	0.738	0.365	1.942	0.177	0.044
	(-0.375)	(0.245)	(-2.524)	(0.882)	(0.629)	(0.524)	(1.350)	(0.923)	16.68
Belgium	0.201^{a}	-0.423^{a}	-0.365^{c}	2.896^{b}	-2.863^{c}	0.088	-0.069	0.152	0.032
	(1.834)	(-1.692)	(-3.357)	(2.503)	(2.864)	(-0.181)	(-0.061)	(0.854)	25.23^{b}
Canada	0.204	-0.302	-0.230^{b}	0.410	1.710^{c}	0.144	-0.090	0.353^{b}	0.019
	(1.538)	(-1.387)	(-2.213)	(0.471)	(3.718)	(0.634)	(-0.137)	(2.101)	13.30
Denmark	0.020	0.514	-0.537^{c}	0.029	-0.058	-0.165	8.139^{b}	0.050	0.069
	(0.132)	(1.637)	(-2.605)	(0.021)	(-0.038)	(-0.170)	(2.556)	(0.223)	7.82
Finland	-0.069	0.232	-0.298^{c}	1.914^{c}	-1.567^{b}	0.375^{b}	1.101	0.109	0.025
	(-1.540)	(1.623)	(-2.795)	(2.619)	(-2.385)	(1.964)	(1.203)	(0.657)	17.29
France	0.102	-0.199	-0.290	-0.259	-0.626	0.175	-0.191	-0.217	0.021
	(1.539)	(-1.416)	(-1.532)	(-0.444)	(-1.293)	(0.492)	(-0.393)	(-1.180)	10.65
Germany	0.119^{a}	-0.185	-0.253^{b}	-1.696	0.193	0.521^{b}	-0.802	0.060	0.024
	(1.946)	(-1.265)	(-2.244)	(-1.148)	(0.339)	(2.128)	(-1.120)	(0.370)	17.35
Iceland	0.132^{a}	0.061	-0.464^{c}	1.562^{c}	1.610^{b}	0.322	0.406	-0.240^{a}	0.043
	(1.802)	(0.406)	(-3.220)	(3.197)	(2.082)	(1.215)	(1.352)	(-1.698)	13.44
Ireland	0.080	-1.230	-0.443^{b}	0.764^{b}	-2.046^{b}	0.810^{b}	-1.813	0.229	0.042
	(0.502)	(-1.429)	(-2.164)	(2.246)	(-2.391)	(2.462)	(-1.246)	(1.131)	18.74
Italy	0.003	0.001	-0.300^{b}	1.669	-0.783	0.742	0.134	0.142	0.035
	(0.053)	(0.003)	(-2.213)	(1.184)	(-0.910)	(1.898)	(0.181)	(0.784)	14.29
Japan	-0.037	0.004	-0.189	-0.566	0.408	-0.179	1.579	-0.334^{b}	0.037
	(-0.197)	(0.007)	(1.438)	(-0.498)	(-0.280)	(-0.535)	(1.493)	(-2.051)	16.35
Netherlands	-0.017	-0.009	-0.186^{a}	1.196	-2.373	0.585^{b}	-0.960	0.305^{a}	0.017
	(-0.641)	(-0.103)	(-1.696)	(1.068)	(-0.916)	(2.308)	(-1.256)	(1.784)	19.97
Norway	-0.147^{b}	0.406^{b}	-0.372^{b}	0.296	0.681	0.508	0.998	0.068	0.035
	(-2.062)	(2.042)	(-2.307)	(0.634)	(0.860)	(1.626)	(0.906)	(0.374)	17.15
Spain	0.029	0.270	-0.344^{c}	2.335^{c}	-0.087	1.010^{c}	0.057	0.067	0.032
	(0.393)	(1.304)	(-3.343)	(4.949)	(-0.160)	(2.815)	(0.089)	(0.480)	38.39^{c}
Sweden	0.084^{a}	-0.444	-0.136	-1.360	1.079	0.266	-0.374	0.155	0.027
	(1.782)	(-1.490)	(-1.126)	(-0.463)	(0.654)	(0.737)	(-0.546)	(0.800)	16.12
Switzerland	0.099	0.130	-0.538^{c}	0.456	-0.940^{a}	0.205	0.189	-0.014	0.029
	(0.476)	(0.388)	(-2.835)	(0.781)	(-1.763)	(0.581)	(0.197)	(-0.071)	13.11
UK	0.094	-0.298	-0.276^{b}	0.627	-2.384^{b}	0.593^{c}	-2.738^{b}	0.122	0.021
	(1.088)	(-1.535)	(-2.542)	(1.182)	(-2.191)	(3.178)	(-2.530)	(0.758)	16.86
USA	-0.035	0.394^{b}	-0.157^{c}	1.984^{b}	3.052^{a}	0.452^{b}	0.478	0.053	0.017
	(-0.408)	(2.540)	(-2.676)	(2.196)	(1.732)	(2.180)	(0.848)	(0.318)	26.49^{b}

Table 2: Single equation results, unrestricted estimation of

$$\begin{aligned}\Delta\mathrm{HCE}_{n,t} &= \nu_n + \delta_n t/100 + \alpha(\mathrm{HCE}_{n,t-1} - \beta_{1n}\mathrm{GDP}_{n,t-1} - \beta_{2n}\mathrm{P65}_{n,t-1}) \\ &+ \gamma_{1n}\Delta\mathrm{GDP}_{n,t} + \gamma_{2n}\Delta\mathrm{P65}_{n,t} + \gamma_{3n}\Delta\mathrm{HCE}_{n,t-1} + u_{n,t}.\end{aligned}$$

t-values in parentheses, $Q(20)$ is the Ljung-Box statistic testing for joint residual autocorrelation up to order 15. Critical values are taken from the $\chi^2(14)$-distribution. $\hat{\sigma}$ is the estimated standard error of $u_{n,t}$. a, b, c indicate significance at level $0.10, 0.05, 0.01$, respectively.

Country	$\hat{\nu}$	t-values
Australia	-0.061 ^{c}	-5.461
Austria	-0.062 ^{c}	-4.330
Belgium	-0.070 ^{c}	-4.836
Canada	-0.044 ^{c}	-4.360
Denmark	-0.050 ^{c}	-3.896
Finland	-0.066 ^{c}	-4.915
France	-0.043 ^{c}	-3.565
Germany	-0.029 ^{b}	-2.382
Iceland	-0.064 ^{c}	-5.076
Ireland	-0.077 ^{c}	-4.167
Italy	-0.067 ^{c}	-4.870
Japan	-0.080 ^{c}	-5.385
Netherlands	-0.052 ^{c}	-4.421
Norway	-0.065 ^{c}	-4.516
Spain	-0.096 ^{c}	-4.903
Sweden	-0.034 ^{c}	-2.614
Switzerland	-0.043 ^{c}	-4.032
UK	-0.093 ^{c}	-6.007
USA		
constant	0.057 ^{c}	5.126

Table 3: Estimates of long-run country intercepts ν_i relative to USA, a, b, c indicate significance at level $0.10, 0.05, 0.01$, respectively.

Parameter	Hypothesis H_0	LR-statistic	p-value
δ	$H_0^1 : \delta = 6.1\text{E-}04$	41.16	0.036
	$H_0^4 : \delta = 0$	41.17	0.036
β_1	$H_0^2 : \beta_1 = 0.740$	44.68	0.023
	$H_0^5 : \beta_1 = 1$	41.71	0.049
β_2	$H_0^3 : \beta_2 = -0.095$	43.00	0.074
	$H_0^6 : \beta_2 = 0$	44.07	0.058

Table 4: Hypothesis tests of pooling restrictions. Critical values are estimated by means of the wild bootstrap.

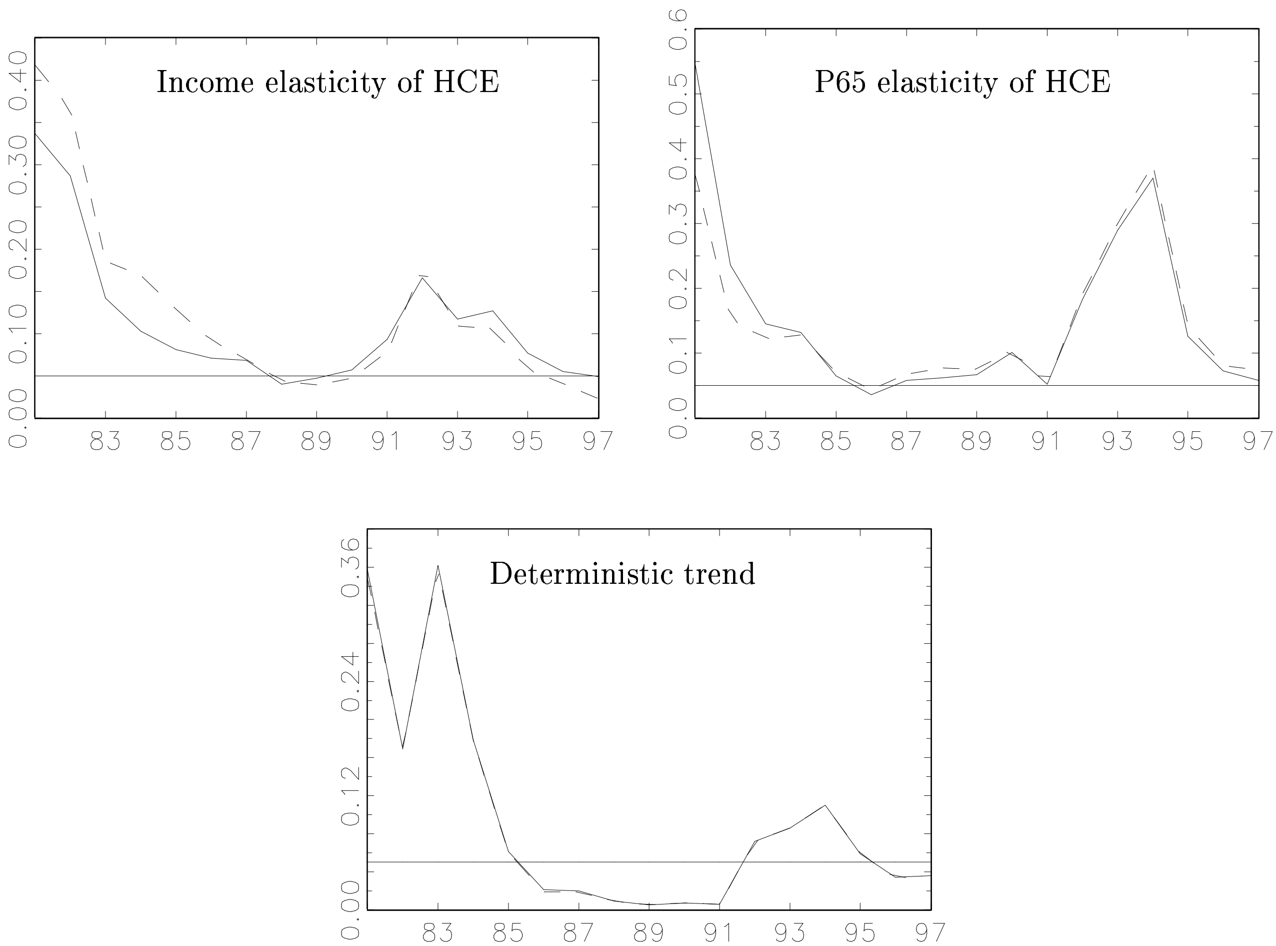

Figure 1: Marginal significance levels for recursive LR-tests. Sampling period is 1961-T^*, $T^* = 1981, \ldots, 1997$. Dashed lines: Tests of restricted estimates (H_0^1 to H_0^3), solid lines: a priori hypothesis H_0^4 to H_0^6