Empir Econ (2011) 40:95–118
DOI 10.1007/s00181-010-0375-6

Testing for structural breaks in panel varying coefficient models: with an application to OECD health expenditure

Dandan Liu · Rui Li · Zijun Wang

Received: 28 July 2009 / Accepted: 10 March 2010 / Published online: 17 June 2010

Abstract In this article, we propose a model selection approach for testing structural breaks in a semiparametric panel varying coefficient model. Monte Carlo evidence shows that the proposed model selection approach performs well in finite sample settings. Applying the method to an empirical data, we find evidence of structural breaks in Organisation for Economic Co-operation and Development (OECD) health expenditure data by allowing for income elasticity to be state (income)-dependent. The relationship between health expenditure and income is subject to two types of structural changes: smooth changes over income and structural breaks in the time dimension. The findings hold for both foreign exchange rate-converted and Purchasing Power Parity-converted expenditure and GDP.

Keywords Panel data · Structural break · Varying coefficient model · Model selection · OECD health expenditure

JEL Classification C14 · C23 · I10

D. Liu (✉)
Department of Economics, Kent State University, P.O. BOX 5190, Kent, OH 44242-0001, USA
e-mail: dliu1@kent.edu

R. Li
School of Economics & Management, Beijing University of Aeronautics & Astronautics, Beijing, China
e-mail: lirui@buaa.edu.cn

Z. Wang
Private Enterprise Research Center, Texas A&M University, College Station, TX, USA
e-mail: z-wang@tamu.edu

1 Introduction

There is a rich literature on testing for the constancy of regression relationships. Two major forms of structural changes (parameter instability) have been studied.[1] (i) The first one is structural breaks (regime shifts), which are sudden and rare. Most of the earlier studies in this line of research are designed for the case of a single break with a known or unknown break date. Recent studies are more concerned with multiple structural breaks. Related literature includes Andrews et al. (1996), Liu et al. (1997), Bai and Perron (1998), and Pesaran and Timmermann (2007).[2] This form of parameter instability (structural breaks) has typically been studied in linear models. (ii) More recently, Cai et al. (2000a), Cai et al. (2000b), Li et al. (2002), and Ahmad et al. (2005) considered a new form of parameter instability in a semiparametric setting, namely, varying coefficient models where model parameters are unknown functions of some state variables. Some familiar nonlinear time series models, such as threshold autoregressive (TAR) models, smooth transition autoregressive (STAR) models, and other regime-switching models, are special cases of the flexible semiparametric varying coefficient models.

While empirical evidence suggests that many economic time series are subject to structural breaks, there are also some interesting applications of the semiparametric varying coefficient models in financial economics and health economics (e.g., Chou et al. 2004, Cai Z, and Kuan CM (2005), Jansen et al. 2008). Combining the literature of testing for structural breaks with that of estimating a varying coefficient model, in this article, we propose a procedure of testing structure breaks in a panel varying coefficient model. For example, consumers' preferences for health care may change when their incomes change from one level to another.[3] Accumulation of knowledge about medical technology may also cause smooth changes in consumers' preferences for health care at the same level of income. On the other hand, health care systems are socialized to various degrees in developed countries. Therefore, one-time changes in institutional factors (such as reimbursement policies of health spending and regulations on health insurance markets) can cause some dramatic changes in aggregate health expenditure. The annual growth rates of health care spending in the U.S. and many other OECD countries were all subjected to such structural breaks (Narayan 2006).

To formally test for structural breaks in a semiparametric varying coefficient model, we recommend the use of the model selection approach that is based on a corrected version of Akaike information criterion (AIC) (Hurvich et al. 1998). We use Monte Carlo experiments to evaluate the finite sample performance of the suggested procedure. The results show that the approach performs well in detecting structural breaks and performs reasonably well in demarcating breaking points. We then apply the model selection procedure to study the long-standing issue of the relationship between

[1] Structural breaks and structural changes are sometimes used interchangeably. Here, we broadly define structural changes as any forms of unstable relationships between the response variable and its covariates, whether the changes in model parameters are sudden or smooth.

[2] See Perron (2006) for a comprehensive review of the recent contributions.

[3] See Di Matteo (2003) for empirical evidence.

health expenditure and income in OECD countries. Our empirical results show that the relationship was subjected to both types of structural changes during the period from 1960 to 2002: smooth changes occurred over state variables (income) and structural breaks arose in the time dimension. Therefore, the often-used log linear model with or without structural breaks is misspecified. We further find that the parameters of other non-income variables (age structures and financing of health care) are not constant; instead they also vary with income. These results hold true whether expenditure and income variables are converted according to foreign exchange rates or purchasing power parities.

The remainder of the article is organized as follows. In Sect. 2, we introduce the semiparametric varying coefficient model and outline the model selection procedures to conduct structural break tests in the semiparametric model setting. Section 3 reports results of a small Monte Carlo experiment that evaluates the performance of the proposed testing procedure. In Sect. 4, we apply the proposed estimation/testing procedure to the OECD health data and discuss policy implications of the empirical findings. Section 6 concludes the paper.

2 Econometric models and structural breaks

2.1 A semiparametric panel varying coefficient model

We consider the following semiparametric panel varying coefficient model (e.g., Cai et al. 2000a; Li et al. 2002)

$$Y_{it} = X_{it}'\beta(Z_{it}) + \varepsilon_{it}, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \tag{1}$$

where X_{it} is a $(p \times 1)$ vector, Z_{it} a vector of dimension $(q \times 1)$, the coefficient function $\beta(z)$ is a $(p \times 1)$ vector of unspecified smooth functions of z. Under the assumption that model (1) is the correct specification, then $E(\varepsilon_{it}|X_{it}, Z_{it}) = 0$.

Pre-multiplying both sides of (1) by X_{it} and taking conditional expectation (conditional on Z_{it}) leads to $E(X_{it}Y_{it}|Z_{it}) = E\left(X_{it}X_{it}'|Z_{it}\right)\beta(Z_{it})$. Solving for $\beta(Z_{it})$ yields

$$\beta(z) = \left(E\left(X_{it}X_{it}'|Z_{it} = z\right)\right)^{-1} E(X_{it}Y_{it}|Z_{it} = z). \tag{2}$$

Replacing the two conditional mean functions in Eq. 2 by some nonparametric estimators, say by the local constant or local linear kernel estimators, one obtains a feasible estimator of $\beta(z)$. Estimation methods as well as the asymptotic distributions of various kernel-based estimators for varying coefficient models have been considered by the above-cited studies.

A useful generalization of model (1) is the so-called partially linear varying coefficient model (hereafter PLVC). It takes the form

$$Y_{it} = X_{1,it}'\beta_0 + X_{2,it}'\beta(Z_{it}) + \varepsilon_{it}, \quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T, \tag{3}$$

where $X_{it} = \left(X'_{1,it}, X'_{2,it}\right)'$. In order to estimate the PLVC model, Fan and Huang (2005) proposed a procedure, which combines Robinson (1988) two-step procedure of estimating a partially linear model and the above estimation method for the basic varying coefficient model (2).

To test the PLVC model (3) against the basic varying coefficient model (1), or to test models (1) and (3) against a restrictive linear model (linearity test), we consider a bootstrap version of the Ullah (1985)-type goodness of fit test which is due to Cai et al. (2000a). This test is based on the difference of the sums of squared residuals between the two competing models, which can be understood as a generalization of the likelihood ratio test in linear regressions (hereafter GLR):

$$T_n = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2 - \sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\varepsilon}_{it}^2\right)\Big/\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\varepsilon}_{it}^2, \tag{4}$$

where $\hat{\varepsilon}_{it}$ is the residual from the linear (null) model, and $\tilde{\varepsilon}_{it}$ is the residual from the alternative (partially linear) varying coefficient model. One rejects the null model for large values of T_n. Cai et al. (2000a) also suggested using the bootstrap approach to evaluate the p-value of the test. In particular, they bootstrap the centralized residuals from the alternative model instead of the null model, because the alternative model residuals are consistent under both the null and alternative hypotheses. To improve the finite sample performance of the test in the presence of heteroskedasticity, we compute the wild bootstrap statistics of Liu (1988), as advocated by Li and Wang (1998).

2.2 The test of structural breaks in semiparametric models

The selection of the number of breaks is essentially a model specification problem just like the choice/testing of linear and nonlinear functional forms. In linear regressions, Yao (1988), Bai and Perron (2006), and Wang (2006) proposed the use of information criteria, as an alternative to parametric tests, to determine whether structural breaks occur in the regression relationships. Extending this line of research, here we recommend the use of information criteria to study structural breaks within the framework of the above varying coefficient models. Specifically, one can minimize an information criterion over a domain of models with zero, one, …, and up to a predetermined maximum number of breaks. The model with the minimum loss value is selected. If the maximum number of breaks is 1, then it is trivial to search for the break point that minimizes the information criterion. If the number of breaks is larger than 1, then the search algorithm suggested by Bai and Perron (1998) can be used.

The information criterion that we consider is a corrected version of AIC. AIC was originally designed for parametric models as an approximately unbiased estimate of the expected Kullback–Leibler information. However, it has been well documented that AIC can lead to significant bias in finite samples. Various methods to correct for the bias have been proposed by researchers, among whom Hurvich and Tsai (1989) proposed the following bias-corrected AIC (AICc) criterion for linear regressions:

$$\text{AICc} = \ln\left(\hat{\sigma}^2\right) + 1 + 2(k+1)/(T-k-2), \tag{5}$$

where $\hat{\sigma}^2$ is the estimated variance of the innovation term, and k is the number of free parameters in a linear regression model. Paralleling Hurvich and Tsai (1989), Hurvich et al. (1998) proposed the following nonparametric version of AICc:

$$\text{AICc} = \ln\left(\hat{\sigma}^2\right) + 1 + 2(tr(H)+1)/(NT - tr(H) - 2), \tag{6}$$

where *tr*(.) is the trace operator, and H is the smoother matrix associated with the above varying coefficient model. Let $\hat{y}$ be the vector of kernel estimate of E(y|x,z), since kernel estimate is linear in y, we have $\hat{y} = H\text{y}$, where H is a n by n squared ($n =$ NT) matrix. Specifically, let H_j (of dimension 1 $\times n$) be the *j*th row of H, then $H_j = X_j'(XKX)^{-1}X'K$, where X_j' is $1 \times p$, X is $n \times p$, K is a $n \times n$ diagonal matrix with the *l*th diagonal element equals to $K\left(\left(X_l - X_j\right)/h\right)$, here $K(.)$ is the (product) kernel function, h is the smoothing parameter. According to Hastie and Tibshirani (1999, section 3.5), $tr(H)$ can be interpreted as the effective number of parameters used in the smoothing fit, reflecting the complexity of the model. To determine whether there is a break for a given time period versus no break, we make the following calculations: (i) we compute AICc using the full sample, and denote this as AICc_0, (ii) we use the sample before (and up to) the possible break point to compute $\hat{\sigma}^2$ and $Tr(H)$, denote them by $\hat{\sigma}_b^2$ and $Tr(H_b)$, respectively; similarly we compute these quantities using the after break sample and denote them by $\hat{\sigma}_a^2$ and $Tr(H_a)$, respectively. The AICc is computed with $\hat{\sigma}^2 = \hat{\sigma}_b^2 + \hat{\sigma}_a^2$ and $Tr(H) = Tr(H_b) + Tr(H_b)$, the resulting AICc is denoted as AICc_1, we select the model with smaller AICc, i.e., we select the model of no break is $\text{AICc}_0 < \text{AICc}_1$, and we select the model with one break if $\text{AICc}_0 > \text{AICc}_1$. AICc for multiple break points model is computed similarly and we select the number of breaks by minimizing AICc. In practice, once the numbers of break points are determined, one should use each subsample (within which no break occurs) to estimate a varying coefficient model.

3 A Monte Carlo experiment

In this section, simulation studies are used to assess the finite-sample behavior of the model selection approach for testing structural breaks in semiparametric regressions. Let $v_{1,it}$, $v_{2,it}$, $v_{3,it}$ and $v_{4,it}$ all be uniform [0, 2] random variables, define $x_{1,it} = 2v_{1,it} + v_{2,it}$, $x_{2,it} = 2v_{3,it} + v_{2,it}$, and $z_{it} = 2v_{4,it}$, $\varepsilon_{it} = \mu_i + v_{it}$ where μ_i is i.i.d. N $\left(0, \sigma^2\right)$ and $v_{it} = \left(\sqrt{x_{1,it}} + \sqrt{x_{2,it}}\right)\theta_{it}$ with θ_{it} being an i.i.d. N $\left(0, \sigma^2\right)$ random variable. We choose $\sigma^2 = 1, 1.5, 2$ and consider three data generating processes (DGPs). The first one is a varying coefficient model without structural breaks:

$$y_{it} = x_{1,it}\beta_1(z_t) + x_{2,it}\beta_2(z_{it}) + \varepsilon_{it}, \tag{DGP1}$$

Table 1 The performance of the model selection approach via AICc for structural break tests in semiparametric models (DGP1: No breaks)

σ^2 -value	(T,N)=(30,20) $\sigma^2=1$	$\sigma^2=1.5$	$\sigma^2=2$	(T,N)=(60,10) $\sigma^2=1$	$\sigma^2=1.5$	$\sigma^2=2$
% of no break	0.965	0.945	0.910	0.975	0.920	0.895

Note: Each entry is the percentage of time AICc selects correct models (no break).

where $\beta_1(z_{it}) = af_1(z_{it}) + (1-a)f_2(z_{it})$, $\beta_2(z_{it}) = af_3(z_{it}) + (1-a)f_4(z_{it})$, $f_1(z_{it}) = (1/2)z_{it} - \sqrt{z_{it}}$, $f_2(z_{it}) = 1+\cos(z_{it})$, $f_3(z_{it}) = \sin(z_{it})$, $f_4(z_{it}) = z_{it}^2 - (1/2)z_{it}$ with $a = 0.5$.

The second DGP is a varying coefficient model with one break:

$$y_{it} = x_{1,it}\beta_1(z_{it}) + x_{2,it}\beta_2(z_{it}) + \varepsilon_{it}, \tag{DGP2}$$

where $\beta_1(z_{it})$ and $\beta_2(z_{it})$ are the same as defined in (DGP1) except that a differs from 0.5.

We choose $a = 0.4$ (and $a = 0.45$) for before the break, and $a = 0.6$ (and $a = 0.55$) for after the break β functions. For example, when $a = 0.4$, we have $\beta_{1,\text{before}}(z_{it}) = 0.4f_1(z_{it}) + 0.6f_2(z_{it})$, $\beta_{1,\text{after}}(z_{it}) = 0.6f_1(z_{it}) + 0.4f_2(z_{it})$. Similarly, replacing a by 0.4 in $\beta_2(z_{it})$ gives $\beta_{2,\text{before}}(z_{it})$ and replacing a by 0.6 in $\beta_2(z_{it})$ gives $\beta_{2,\text{after}}(z_{it})$.

We consider three possible break dates, $\tau = 0.3T, 0.5T$, and $0.7T$.

The third DGP is a varying coefficient model with two breaks. We use subscript I, II, and III to denote the β functions that correspond to before any break, after the first but before the second break, and after the second break.

$$y_{it} = x_{1,it}\beta_1(z_{it}) + \varepsilon_{it}, \tag{DGP3}$$

where $\varepsilon_{it} = \sqrt{x_{1,it}}u_{it}$, and $\beta_{1,\text{I}}(z_{it}) = 0.7g_1(z_{it}) + 0.2g_2(z_{it}) + 0.1g_3(z_{it})$, $\beta_{1,\text{II}}(z_{it}) = [g_1(z_{it}) + g_2(z_{it}) + g_3(z_{it})]/3$, $\beta_{1,\text{III}}(z_{it}) = 0.1g_1(z_{it}) + 0.2g_2(z_{it}) + 0.7g_3(z_{it})$, with $g_1(z_{it}) = z_{it} + \sin(z_{it})$, $g_2(z_{it}) = 1+\cos(z_{it})$, $g_3(z_{it}) = \sqrt{z_{it}}$, that is,

$$\beta_1(z_{it}) = \begin{cases} \beta_{1,\text{I}}(z_{it}) & \text{if} \quad t \le \tau_1, \\ \beta_{1,\text{II}}(z_{it}) & \text{if} \quad \tau_1 < t \le \tau_2, \\ \beta_{1,\text{III}}(z_{it}) & \text{if} \quad t > \tau_2. \end{cases}$$

The first break date $\tau_1 = T/3$ and the second break date $\tau_2 = (2T)/3$.

Table 1 summarizes the simulation results for DGP1.[4] When there is no structural break in the data, AICc performs quite well for all parameter choices in the two sample sizes: (T,N) = (30, 20), (60, 10), selecting the correct model with over 89% accuracy.

[4] The number of replications is 400. In simulations, we use the local constant estimator and the optimal smoothing parameter (bandwidth), $h = c_0 s T^{-1/5}$, where c_0 is a constant, s is the sample standard deviation of the state variable z_t. We set c_0 to be 1. We also use $c_0 = 0.8$ and $c_0 = 1.2$ the results are similar and are not reported here to save space.

Table 2 reports results that AICc finds one break when the DGP is subject to one break. For (T,N) = (30, 20) and $a = 0.4$, it finds one break point 100% of times. Specifically, when $a = 0.4$, $\sigma^2 = 1.5$, and $\tau = 9$, the AICc finds one break 100% times, it finds the break at exactly $\tau = 88\%$ of times, and it finds one break between $\tau = 9$ and $\tau = 10$, 95% of times. As expected the performance of the AICc deteriorates as the error variance increases. Also, as a increases (from 0.4 to 0.45), the break becomes less significant and the power of the AICc test deteriorates.

Finally, the proposed procedure also performs well when the varying coefficient model is subject to two structural breaks (Table 3). For example, when (T,N) = (30,20), the first break occurred at $\tau 1 = 10$ and the second break occurred at $\tau_2 = 20$, 95% of times our test finds the first break between $\tau_1 = 9$ to 10, and finds the second break between $\tau_2 = 19$ to 20.

We only report simulations for testing for the maximum number of two breaks in this article. Although it is straightforward to test multiple break points using our proposed method, testing a large number of breaks requires one to split the sample into many non-overlapping subsamples, and estimate a semiparametric model using each sub-sample data. The small sample size in each subsample can lead to unreliable estimation result. Therefore, the proposed method may not be suitable for testing a large number of breaks if one does not have a long panel over time dimension. On

Table 2 The performance of the model selection approach via AICc for structural break tests in semiparametric models (DGP2: one break)

Break time	$\tau = 0.3T$ $\tau = 9$		$\tau = 0.5T$ $\tau = 15$		$\tau = 0.7T$ $\tau = 21$	
(T,N) = (30, 20)	$\sigma^2 = 1$					
$a = 0.4$	100	94.0	100	95.5	100	93.0
		(9, 9)		(15, 15)		(21, 21)
	$\sigma^2 = 1.5$					
	100	88.0	100	87.0	100	89.5
		(9, 10)		(14, 16)		(20, 21)
	$\sigma^2 = 2$					
	100	85.0	100	80.0	100	83.5
		(9, 10)		(14,16)		(20, 21)
$a = 0.45$	$\sigma^2 = 1$					
	98.5	60.0	99.5	60.5	99.5	67.0
		(7, 13)		(13,17)		(19, 23)
	$\sigma^2 = 1.5$					
	95.5	46.0	99.0	46.5	97.5	54.0
		(7, 16)		(11,19)		(17, 23)
	$\sigma^2 = 2$					
	92.5	36.5	96.0	37.0	95.0	44.0
		(12, 20)		(10, 22)		(14, 24)

Table 2 continued

Break time	$\tau = 18$		$\tau = 30$		$\tau = 42$	
(T,N)=(60, 10)	$\sigma^2 = 1.0$					
$a = 0.4$	100	83.5	100	84.0	100	82.5
		(17, 19)		(29, 31)		(40, 44)
	$\sigma^2 = 1.5$					
	100	70.0	100	71.0	100	76.5
		(17, 20)		(28, 32)		(41,43)
	$\sigma^2 = 2$					
	100	62.5	100	64.0	100	68.0
		(17, 20)		(27, 32)		(40, 44)
$a = 0.45$	$\sigma^2 = 1$					
	98.5	33.0	99.0	41.0	99.5	48.5
		(16, 27)		(26, 34)		(38, 46)
	$\sigma^2 = 1.5$					
	97.0	23.5	98.0	29.0	98.0	38.0
		(14, 32)		(22, 37)		(36, 47)
	$\sigma^2 = 2$					
	95.0	23.5	96.5	22.0	97.0	30.5
		(14, 40)		(22, 43)		(27, 47)

Note: Each entry is the percentage of time that AICc selects the correct number of breaks (one break), or that it finds the exact break point. The *numbers in parentheses* are the two sample observations that correspond to the 5th and 95th percentiles of the empirical distribution of the detected break point

Table 3 The performance of the model selection approach via AICc for structural break tests in semiparametric models (DGP3: two breaks)

	$\sigma^2 = 1$		$\sigma^2 = 1.5$		$\sigma^2 = 2$	
	2 breaks	$\hat{\tau}$	2 breaks	$\hat{\tau}$	2 breaks	$\hat{\tau}$
(T,N) = (30,20)	100.0	39.0	100.0	36.0	100.0	33.5
$\tau_1 = 10$		(9, 10)		(9,10)		(9, 10)
$\tau_2 = 20$		(19, 20)		(19,20)		(19, 20)
(T,N) = (60,10)	100.0	38.0	100.0	35.5	100.0	33.0
$\tau_1 = 20$		(19, 20)		(19,20)		(19, 20)
$\tau_2 = 40$		(39, 40)		(39,40)		(39, 40)

Note: Each entry from column 2–6 are the percentage of time that AICc selects the correct number of break (two breaks), or that they find the exact break points. The *numbers in parentheses* are the two sample observations that correspond to the 5th and 95th percentiles of the empirical distribution of the two detected break points

the other hand, since semiparametric approaches allow for flexible functional forms, the test is designed to detect a few significant structure breaks rather than many small (possible) breaks.

4 An application to OECD health expenditure

Rising health costs accompanied with aging populations have been the constant subjects of comments and discussions in popular media in most developed countries. An important issue in the health economics literature is the relationship between health care expenditure (HCE) and income (e.g., GDP). One approach to study this issue is via international comparisons of health expenditure.[5] The research on international comparisons of health expenditure has important policy implications. For example, the (U.S.) Technical Review Panel on the Medicare Trustees Reports (2004) reconfirmed the recommendations of the 2000 Technical Review Panel that the long-run growth rate of per capita HCE is one percentage point higher than the growth rate of per capita GDP.[6] This assumption is used in the annual Trustees Reports for long-range (e.g., 75 years) forecasts of future Medicare expenditure. The assumption necessarily implies an income elasticity of HCE equal to 1.333, if, as the trustees assumed, future GDP grows at its historical rate of 3%. Is this elasticity value high or low from the historical perspective? Despite the continuing research efforts, the empirical evidence on the size of the income elasticity of health expenditure still remains inconclusive. One reason is that the widely used log linear functional form is probably too restrictive in that it implies an income elasticity that is invariant across time and across countries in rather different development stages. If this homogeneity assumption on parameters does not hold, then the log linear model is misspecified, and it is likely to provide rather different estimates depending on which countries are included and which sample period is used in the analysis. In this section, we use a flexible varying coefficient model to examine the relationship between health expenditure and income. We apply the model selection procedure proposed in Sect. 2 to examine whether the health expenditure and income relationship has been subject to any form of structural changes.

4.1 Data and empirical model structures

We use annual observations on health expenditure for the 22 countries in the Organisation for Economic Co-operation and Development (OECD) for the period 1960–2002.[7] The data source is *OECD Health Data* (2006) (earlier editions of this database have been widely used). Excluding missing values, our sample includes 730 observations. Appendix A provides detailed information on data availability for each of the 22 countries considered in the article.

5 See Gerdtham and Jönsson (2000) for an up-to-date review of these studies. One reason why a large number of empirical studies use country level data is that neither a household/individual nor a state/province faces full health care costs.

6 Based on the quantitative analysis of a model with standard economic assumptions, Hall and Jones (2007) also projected that the optimal health share of spending in the U.S. "seems likely to exceed 30 percent by the middle of the century."

7 While data since 2002 are also available for most countries, year 2003 includes many more breaks in variable definitions than other years. It is difficult to conduct meaningful analysis allowing breaks within such a short period.

The dependent variable is per capita HCE. Following the literature (e.g., Hitiris and Posnett (1992) and Di Matteo (2005)), we include income and three non-income variables as explanatory variables. The three non-income variables are defined as follows: the proportion of HCE that is publicly funded (PUBL), the proportion of the population under the age 15 (POP15), and over the age 65 (POP65).[8] The local currency denominated HCE and GDP are converted by foreign exchange rates (XRATE) to U.S. dollars (later we will briefly discuss the results when these two variables are converted by purchasing power parity (PPP)). They are further transformed into real terms using the U.S. GDP price deflator (with 2000 as the base year). To be comparable in magnitude with other variables in the model, both HCE and GDP are expressed in hundreds before taking logarithms.

For ease of later presentation, we now define, at the risk of abusing notation, $X'_{it} = (X_{1,it}, X_{2,it}, X_{3,it}, X_{4,it}) = (\mathrm{GDP}_{it}, \mathrm{PUBL}_{it}, \mathrm{POP15}_{it}, \mathrm{POP65}_{it})$, where i indexes countries, as the vector of all explanatory variables (including state variables). Denote $X'_{-k,it} = (X_{1,it}, \ldots, X_{(k-1),it}, X_{(k+1),it}, \ldots, X_{4,it})$, omitting the kth element in X'_{it} for $k = 1, 2, 3, 4$. Similarly, denote the corresponding coefficient parameter vector $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ and $\beta_{-k} = (\beta_1, \ldots, \beta_{(k-1)}, \beta_{(k+1)}, \ldots, \beta_4)$. Using these notations, the explicit forms of the (pooled) linear and varying coefficient models are

$$y_{it} = \beta_0 + X'_{it}\beta + \varepsilon_{it}, \quad i = 1, 2, \ldots, N; \; t = 1, 2, \ldots, T, \tag{7}$$

and

$$y_{it} = g(X_{1,it}) + X'_{-1,it}\beta_{-1}(X_{1,it}) + \varepsilon_{it}, \tag{8}$$

where, given the panel nature of the data, we allow for a conditionally heteroskedastic error process of unknown form, namely, $E\left(\varepsilon^2_{it}|X_{it}\right) = \sigma^2(X_{it})$.[9] Note from model (8) that the state variable $X_{1,i,t}$ is $\mathrm{GDP}_{i,t}$. Here, the impact of GDP is twofold: it affects HCE directly through the varying intercept term $g(X_{1,it})$, and indirectly via coefficients of the other explanatory variables (namely, $\beta_{-1}(X_{1,it})$). We estimate a partially linear varying coefficient model (PLVC):

[8] According to the permanent income hypothesis, HCE in year t may also depend on income in previous years. For simplicity, here we do not include any lagged income variables. Nevertheless, POP65 is highly correlated with historical income. Its inclusion in the model can capture part of the lagged effect of income on HCE. Newhouse (1992); Okunade and Murthy (2002) and Di Matteo (2005) also pointed to the importance of technology in driving health care costs. Therefore, the income elasticity estimated from pooled regressions may also reflect the impact of new technology on health expenditure over time.

[9] $\varepsilon_{.,t}$ may also be serially correlated. Therefore, we also conducted some simulations to evaluate the performance of the GLR test in DGPs allowing ε_t to be autocorrelated. We find that the test performance is not sensitive to the presence of autocorrelation. Also note that to estimate a literal "world income elasticity" from model (8), a weighted (say, by each country's population) least squares regression may be more appropriate. Following most of the literature, here we use simple pooled regression, which could be justified by interpreting each country's data as providing a separate draw of country experience unrelated to size of a country.

Table 4 Linear regression results (data converted by XRATE)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2	AICc
Panel (a): Full sample: 1960–2002							
Estimate	−3.200**	1.321**	−0.235**	−0.227**	0.231**	0.958	−2.420
Std error	(0.363)	(0.018)	(0.041)	(0.068)	(0.046)		
Panel (b): Sub-sample I: 1960–1974							
Estimate	−3.935**	1.397**	−0.012	−0.292	0.050	0.944	−2.220
Std error	(0.780)	(0.033)	(0.048)	(0.155)	(0.069)		
Panel (c): Sub-sample II: 1975–1991							
Estimate	−3.089**	1.204**	−0.329**	0.057	0.254**	0.957	−3.149
Std error	(0.395)	(0.015)	(0.036)	(0.084)	(0.058)		
Panel (d): Sub-sample III: 1992–2002							
Estimate	−2.513**	1.124**	−0.700**	0.368**	0.480**	0.918	−3.218
Std error	(0.467)	(0.021)	(0.045)	(0.106)	(0.103)		

Notes: 1. The table reports the estimation results of the linear model (1). The dependent variable is HCE (foreign exchange rate-adjusted). The regressors include, in order, a constant, GDP, PUBL, POP15, and POP65.
2. The *numbers in parentheses* are heteroskedasticity-consistent standard errors of the estimates. The symbol ** denotes significant at the 1% level.

$$y_{it} = \beta_0 + X'_{it}\beta(X_{1,it}) + \varepsilon_{it}. \tag{9}$$

4.2 Linear specifications

Panel (a) of Table 4 presents the parameter estimates of model (7) and their heteroskedasticity-consistent standard errors obtained from the full sample.[10] The income elasticity is 1.321 and the estimate is statistically significant at any conventional level. This result implies that health care is a luxury good. The magnitude of the estimate falls in the range reported in the literature (see Di Matteo (2003) for brief surveys of the literature). Although smaller in magnitude, the effects of the other three non-income variables, PUBL, POP15, and POP65, on health expenditure are all statistically significant and have signs consistent with the literature. For example, the elasticity of POP65 is 0.231, meaning that older people tend to use more health care than the working age population, ceteris paribus. Countries with high fractions of public finance tend to have lower per capita health expenditure (with an elasticity of −0.235). However, we need to interpret these results with caution. As shown shortly,

[10] We assume that all variables are stationary. Although the debate on whether HCE and GDP contain unit roots is not over, recent empirical evidence appears to suggest that both series, aggregated at the national level, can be characterized as stationary if structural breaks are allowed in the tests (Jewell et al. 2003; Carrion-i-Silvestre 2005). Furthermore, in conducting the key model specification tests, we rely on the bootstrap method. This should alleviate the bias in the statistical inference if HCE and GDP are actually nonstationary.

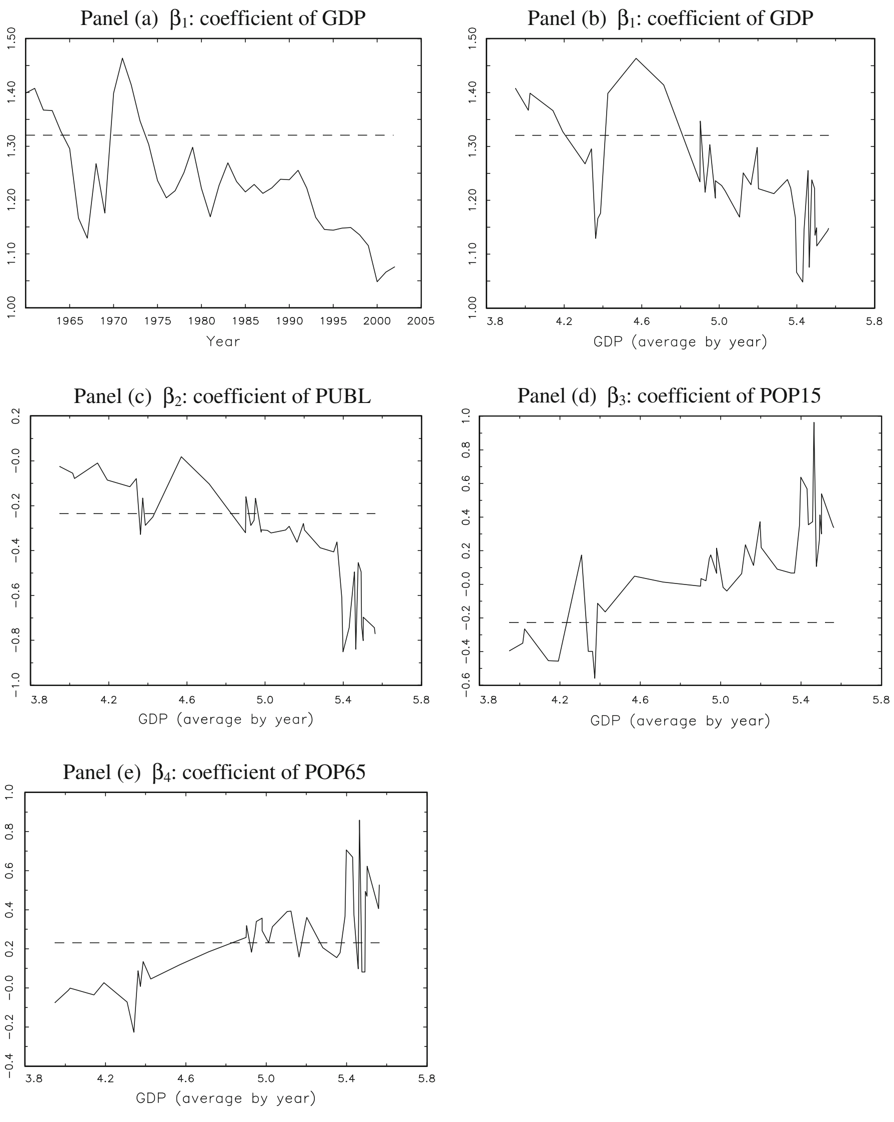

Fig. 1 The elasticity estimates from linear model (1) (data converted by XRATE) *Note*: The *dashed line* is estimated from the full sample. The *solid line* represents recursive estimates

the simple linear model does not provide a good description of data since it ignores the problem of parameter heterogeneity.

To provide some intuitive evidence on the variations in the elasticity estimates, we re-estimate model (8) on an annual basis. Panel (a) of Fig. 1 plots the recursive cross-section income elasticity estimates against the calendar years, which reveals an unstable relationship between HCE and GDP. The downward trend is obvious in both the 1960s and the 1990s. Substantial variations in parameters across time are

Table 5 Structural breaks tests (data converted by XRATE)

Number of breaks	Linear model		PLVC model	
	Break date	Minimized AICc	Break date	Minimized AICc
0 break	N/A	−2.420	N/A	−2.740
1 break	1974	−2.783	1974	−3.002
2 breaks	1974, 1991	−2.854	1970, 1990	−3.090

the indications of structural breaks. To formally address this issue, we use a search procedure similar to Bai and Perron (1998) structural break test to study whether the data can be better characterized by linear model (1) with structural breaks. To be comparable with the later results using semiparametric techniques, here we allow for a maximum of two structural breaks in the testing procedure. A trimming of 25% of the total of 43 years (i.e., 11 years) of observations is applied to both ends of the timely ordered panel data.[11] The left panel of Table 5 shows that the model with two breaks achieves the minimum AICc, verifying the existence of regime shifts in health expenditure as reported in the recent studies that used similar data. The first break occurred in 1974, 1 year after the first oil price shock. The second break occurred in 1991 and largely coincided with the beginning of the long economic expansion in most developed countries led by the U.S. through most of the 1990s.

Panels (b), (c), and (d) of Table 4 summarize the regression results for three sub-sample periods identified by the above structural break test. The income elasticity decreased from 1.397 in the first sub-sample period to 1.21 in the second sub-sample period (1975 through 1991), and further down to 1.124 during the last period of 1992–2002. In contrast, elasticities of PUBL and POP65 increased (in absolute values) during the sample period. Finally, the impact of POP15 changed from a negative value to a positive and significant one, in line with the pattern shown in Fig. 1.

4.3 Semiparametric specifications

Merely finding that income elasticity varies over time (Panel (a) of Fig. 1) is not very useful for its relevance in contemporary policy analysis. Therefore, we also plot in Panel (b) the same income elasticity estimates against the annual cross-nation average GDP. It can be seen that Panel (b) shares a similar pattern with Panel (a), suggesting that the GDP level plays an important role in observed annual variations in HCE's income elasticity estimates. Panels (c), (d), and (e) plot the elasticity estimates of PUBL, POP15, and POP65 against the average GDP, respectively. They also suggest that the impact of these non-income variables on HCE may have changed at different economic development levels. Note that while a trend was evident in all three graphs, the signs of the elasticity estimates for the two age variables changed during the sample period.

[11] As our data is an unbalanced panel, such a trimming approach is not the same as trimming 25% of total observations.

Table 6 The model specification tests(data converted by XRATE)

Test statistic	p-Value	Bootstrap empirical distribution				AICc	R^2
T_n		99%	95%	90%	80%	(H_o/H_A)	(H_o/H_A)
Panel (a)							
$H_o: y_{i,t} = \beta_0 + X'_{i,t}\beta(\text{GDP}_{i,t}) + \varepsilon_{i,t}$ vs. $H_A: y_{i,t} = g(\text{GDP}_{i,t}) + X'_{-1,i,t}\beta_{-1}(\text{GDP}_{i,t}) + \varepsilon_{i,t}$							
−0.204	0.990	−0.137	−0.145	−0.150	−0.155	−2.740/ − 2.506	0.971/0.963
Panel (b)							
$H_o: y_{i,t} = \beta_0 + X'_{i,t}\beta + \varepsilon_{i,t}$ vs. $H_A: y_{i,t} = \beta_0 + X'_{i,t}\beta(\text{GDP}_{i,t}) + \varepsilon_{i,t}$							
0.455	0.000	0.069	0.056	0.050	0.044	−2.420/ − 2.740	0.958/0.971
Panel (c)							
$H_o: y_{i,t} = \beta_0 + X_{1,i,t}\beta_1 + X'_{-1,i,t}\beta_{-1}(\text{GDP}_{i,t}) + \varepsilon_{i,t}$ vs. $H_A: y_{i,t} = \beta_0 + X'_{i,t}\beta(\text{GDP}_{i,t}) + \varepsilon_{i,t}$							
0.025	0.000	0.018	0.014	0.011	0.009	−2.723/ − 2.740	0.970/0.971
Panel (d)							
$H_o: y_{i,t} = \beta_0 + X_{2,i,t}\beta_2 + X'_{-2,i,t}\beta_{-2}(\text{GDP}_{i,t}) + \varepsilon_{i,t}$ vs. $H_A: y_{i,t} = \beta_0 + X'_{i,t}\beta(\text{GDP}_{i,t}) + \varepsilon_{i,t}$							
0.159	0.000	0.025	0.019	0.017	0.014	−2.608/ − 2.740	0.966/0.971
Panel (e)							
$H_o: y_{i,t} = \beta_0 + X_{3,i,t}\beta_3 + X'_{-3,i,t}\beta_{-3}(\text{GDP}_{i,t}) + \varepsilon_{i,t}$ vs. $H_A: y_{i,t} = \beta_0 + X'_{i,t}\beta(\text{GDP}_{i,t}) + \varepsilon_{i,t}$							
0.151	0.000	0.027	0.022	0.019	0.016	−2.611/−2.740	0.966/0.971
Panel (f)							
$H_o: y_{i,t} = \beta_0 + X_{4,i,t}\beta_4 + X'_{-4,i,t}\beta_{-4}(\text{GDP}_{i,t}) + \varepsilon_{i,t}$ vs. $H_A: y_{i,t} = \beta_0 + X'_{i,t}\beta(\text{GDP}_{i,t}) + \varepsilon_{i,t}$							
0.065	0.000	0.025	0.020	0.017	0.014	−2.692/ − 2.740	0.969/0.971

Note: The GLR test statistic T_n is based on Cai et al. (2000a). The number of bootstrap replications is 500

Based upon the above graphic evidence, we now start the semiparametric analysis by first testing whether the more interesting PLVC model (9) is sufficient to capture the underlying nonlinearity of the data relative to the basic varying coefficient model (8). The GLR test results are reported in Panel (a) of Table 6.[12] We fail to reject the null model (PLVC) at any conventional significance level. The AICc of the PLVC model is −2.740, smaller than that of the alternative model (8), also suggesting that the PLVC model is a better choice for the data. Therefore, we use model (9) as our benchmark semiparametric model. Panel (b) presents the linearity test result where the null model is the linear specification (7) and the PLVC is the alternative model. The linearity is clearly rejected for the data. Note that this rejection is consistent with the ranking of the two competing models based on their AICc values in Table 5 (−2.420 vs. −2.740).

Another type of model specification test we conduct is whether the elasticity of one of the four explanatory variables is constant in the PLVC model (in addition to the constant intercept term β_0). Panels (c) through (f) summarize the results for testing the null hypothesis $\beta_k(\text{GDP}_{i,t}) = \beta_k$, $k = 1, 2, 3$, and 4, respectively. In all four cases, we reject the null hypothesis. The impacts of both income and non-income variables on health expenditure, therefore, are not constant everywhere but vary with the income level, confirming the simple rolling-sample OLS estimates presented in Fig. 1.

Figure 2 plots the smooth elasticity estimates from model (9) for GDP, PUBL, POP15, and POP65 based on the full sample. To provide a sense of how accurately these elasticities are estimated, the density of the data is plotted at the bottom of each graph with each observation represented by a symbol $\times$. We also provide pointwise bounds for the estimates. The bounds are formed by the 5th and the 95th percentiles of their distributions. For comparison, the elasticity estimate from the linear model (7) (which is invariant to different income levels) is also plotted as a dashed line. Focusing on Panel (a), we observe that the income elasticity ($\beta_1(\text{GDP}_{i,t})$) is relatively flat when (log) GDP is between 3.5 and 5.5 (approximately US\$ 3,300–24,500). The estimate based on the linear model (1.321) is mostly within the 90% bounds of semiparametric estimates for GDP in this range. However, when income level is either very low or very high, the income elasticity of health expenditure is negatively related to income levels. Table 7 tabulates the mean values and the 10th, 50th (median) and 90th percentiles of the PLVC model estimates of income elasticity (recall there are 730 of them). The mean and median estimates are close to the linear estimator. However, the lower 10th percentile is 1.139 at the higher income end, which is statistically smaller than the linear estimator of 1.321. Associated with lower income levels are much higher elasticity estimates. For example, at the point of GDP equal to 3.038 (US\$ 2,087) (second lowest GDP in the sample), the income elasticity is estimated to be the highest (2.463).

[12] In setting the smoothing parameter h, we followed Li et al. (2002) and used $c_0 = 1$. We also examined various values of c_0 ranging from 0.5 to 2 and found that our basic results hold. We also used the data-driven method to find optimal c_0 (hence h) by minimizing AICc over c_0 for models (2), (6) or their restrictive versions. The value of c_0 found this way falls in the range of (0.5, 1.7) in most cases considered throughout the paper. For a few cases, the data-driven method concludes with c_0 smaller than 0.5, leading to under-smoothed functions of coefficient estimates (hence even more nonlinear than presented here). Nevertheless, the basic conclusions drawn here still hold for those few cases.

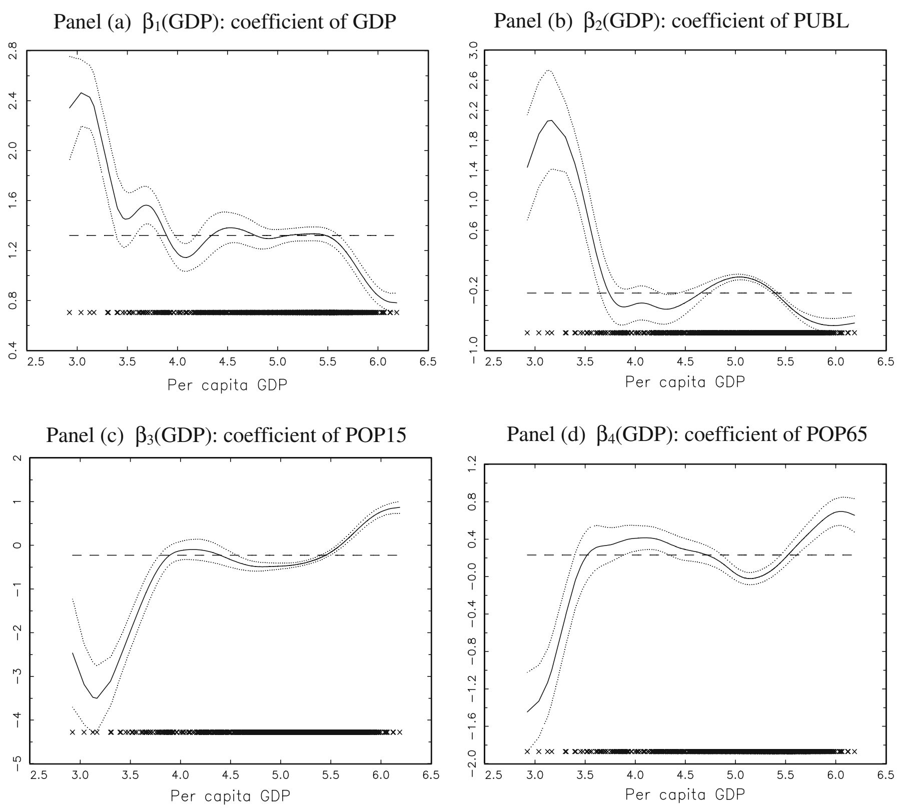

Fig. 2 Full sample elasticity estimates from PLVC model (6) and linear model (1) (data converted by XRATE). *Note*: The *solid line* represents smooth coefficient estimates. The *dotted lines* are the 5th and 95th percentiles of their distributions. The level *dashed line* represents the estimate from the linear model. Each symbol × represents a data point

The semiparametric estimates for PUBL, POP15, and POP65 are plotted in Panels (b), (c), and (d) of Fig. 2. Similar to the pattern of income elasticity, the elasticities of the three variables are all flat and move around their linear model counterparts for intermediate income levels. However, for extreme values of income, we find a negative correlation between elasticity of PUBL and income level and a positive correlation between elasticities of the two age variables and income level. While a higher percentage of HCE that is publicly funded is in general associated with lower per capita GDP, the relation is reversed at very low levels of income. The graphs also illustrate that, when a society has a relatively high proportion of young population, its share of income spent on health care is typically lower, *ceteris paribus*. However, when per capita GDP attained US$ 27,000, such young societies tend to have higher per capita health expenditure. This result is consistent with the finding based on the sub-sample period of 1992–2002 using linear regressions, where the elasticity estimate of POP15 is positive and significant (Panel (d) of Table 4).

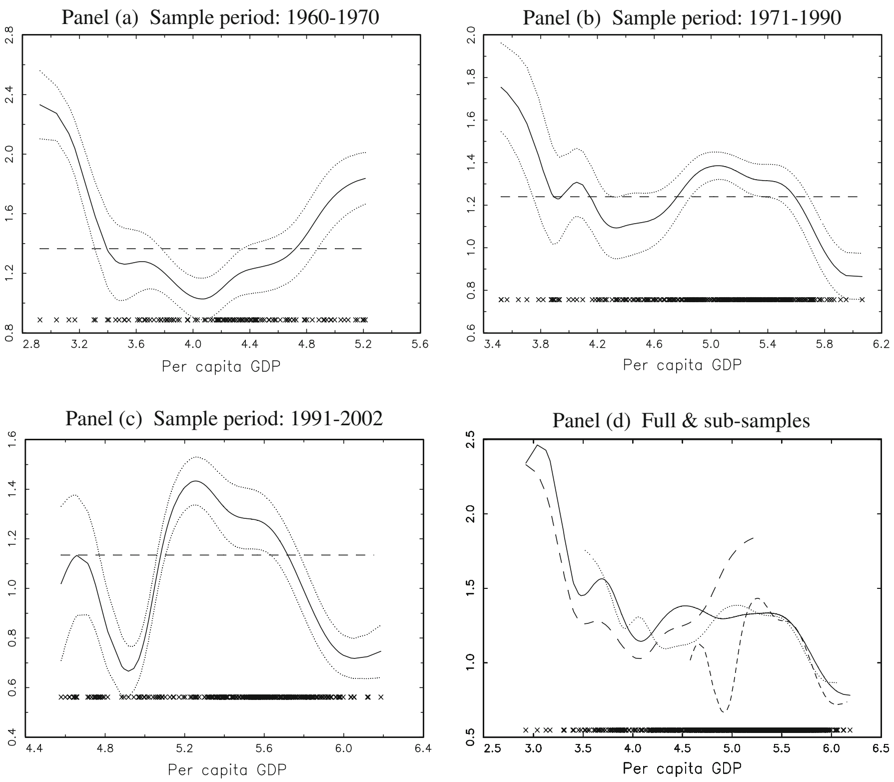

Fig. 3 Sub-sample income elasticity estimates β_1(GDP) from the PLVC model (6) (data converted by XRATE). *Notes*: 1. In Panels (a), (b), and (c), the *solid line* represents smooth coefficient estimates. The *dotted lines* are the 5th and 95th percentiles of their distributions. The level *dashed line* represents the estimate from the linear model. 2. In Panel (d), the *solid line* is estimated based on the full sample. The *dashed*, *dotted and short dashed lines* represent estimates based on the sub-sample periods 1960–1970, 1971–1990, and 1991–2002, respectively. The corresponding number of observations in each sub-sample is 114, 357, and 259, respectively. 3. Each symbol × represents a data point

Paralleling the analysis within the linear framework, we now test whether there still exist structural breaks in the data after allowing for parameters to change smoothly at different levels of income according to model (9). The results are reported in the right panel of Table 5. Based on the model selection method via AICc, we find that the PLVC model (9) with two breaks is preferred to those with 0 or 1 break. The two breaks occurred in 1970 and 1990, respectively. Both preceded the dates identified by the above linear method (1974 and 1991).[13] It can also be seen from Table 5 that the PLVC model (9) with two breaks achieves minimum AICc among all linear and semiparametric specifications, with or without structural breaks.

[13] 1970 is also the year when the number of cross sections in the sample jumped from an average of 9.8 over the 1960-1969 period to 16. To rule out the possibility that the identified break in 1970 is caused by this sudden change in the member countries, we conduct the structural break test with data limited to the ten countries that are in the 1969 cross section. We still find that 1970 is a structural break point.

Based on the structural break test results, we divide the data into three sub-samples, 1960–1970, 1971–1990, and 1991–2002. We re-estimate the PLVC model with each sub-sample and plot the estimated income elasticity $\hat{\beta}_1(GDP_{i,t})$ in Panels (a), (b), and (c) of Fig. 3, respectively. From the panels, we observe some significant differences in the elasticity estimates between the three sub-samples. The variation of $\hat{\beta}_1(.)$ is similar to a U-shaped curve in the 1960s (Panel (a)). In contrast, Panel (b) shows that in the 1971–1990 sample, the highest-income economies tend to have lower income elasticity of HCE than the lowest-income economies. The variation of $\hat{\beta}_1(.)$ in the latest period shows a more complicated picture.

In Panel (d), we plot together the estimates from the full sample and the three sub-samples, where it is now clear that most variations are around GDP = 5 and are from the estimates based on the first and the third sub-samples. However, it should be pointed out here that for the same level of GDP, the sub-sample estimates may show more variations than the full sample estimates, but they also show wider confidence intervals. This is because the number of data points used in the sub-samples is smaller than in the full sample.

Table 7 tabulates the mean and three percentiles of $\hat{\beta}_1(.)$ for each sub-sample. Although there were significant variations of the estimates within each sub-sample the income elasticity decreased on average from the first to the third sub-sample. For log GDP = 5.08 (which is the full sample average), the corresponding income elasticity estimate is 1.78 in the first sub-sample. The estimate is 1.38 in the second and only 1.15 in the third sub-sample. Furthermore, 25% of the estimates fall below the unit value in the 1991–2002 sub-sample. This occurred in both very high-income economies and relatively low-income ones. For high GDP (larger than 5.2), there is a clear negative relation between income elasticity and GDP. Nevertheless, the downward trend appears to have reversed at the highest levels of GDP (recall Panel (c) of Fig. 3).

From the lower three panels of Table 7, we can also see that the elasticity estimates for PUBL, POP15, and POP65 vary with GDP within each sub-sample, and vary across samples for the same level of GDP.

To end this subsection, we compare how the varying coefficient model in-sample fits the data relative to the linear model. The predictions are made using parameter estimates from the full sample.[14] For the 12 countries included in the year 1960 sample, the root mean squared errors (RMSE) for the PLVC model is 0.233, a reduction of 25% from that of the linear model (0.312). The largest gain comes from predicting two outliers, health expenditure of Spain and the U.S. Both models predicted better for the 22 countries in the year 2002 sample. However, there is still a 15% reduction in prediction errors when the PLVC model is used. In predicting the U.S. outlier, the PLVC model has an error of $1,118, much smaller than the $1,731 error from the linear model (the actual real U.S. health expenditure was $5,110 in 2002). We further plot the actual and predicted time series of HCE for the U.S., which shows that the difference in prediction accuracy between the two models is significant[15].

[14] The figure that plots the actual HCE and the predicted values from the two competing models (7) and (9) for the years 1960 and 2002 is available from authors upon request.

[15] This figure is available from authors upon request as well.

Table 7 Semiparametric estimation results (data converted by XRATE)

Coefficient	Mean	10th percentile	Median	90th percentile
Panel (a) Full sample: 1960–2002				
$\hat{\beta}_1$(GDP)	1.288	1.139	1.318	1.371
$\hat{\beta}_2$(GDP)	−0.263	−0.593	−0.287	−0.030
$\hat{\beta}_3$(GDP)	−0.250	−0.485	−0.291	0.268
$\hat{\beta}_4$(GDP)	0.220	−0.010	0.228	0.454
Panel (b) Sub-sample I: 1960–1970				
$\hat{\beta}_1$(GDP)	1.324	1.071	1.238	1.721
$\hat{\beta}_2$(GDP)	0.013	−0.268	−0.140	0.696
$\hat{\beta}_3$(GDP)	−0.857	−1.835	−0.552	−0.389
$\hat{\beta}_4$(GDP)	−0.170	−0.574	−0.113	0.151
Panel (c) Sub-sample II: 1971–1990				
$\hat{\beta}_1$(GDP)	1.278	1.112	1.314	1.380
$\hat{\beta}_2$(GDP)	−0.294	−0.519	−0.261	−0.146
$\hat{\beta}_3$(GDP)	−0.059	−0.304	−0.131	0.453
$\hat{\beta}_4$(GDP)	0.296	0.117	0.224	0.507
Panel (d) Sub-sample III: 1991–2002				
$\hat{\beta}_1$(GDP)	1.142	0.784	1.221	1.351
$\hat{\beta}_2$(GDP)	−0.543	−0.764	−0.620	−0.287
$\hat{\beta}_3$(GDP)	0.165	−0.282	0.086	0.734
$\hat{\beta}_4$(GDP)	0.314	−0.070	0.307	0.677

4.4 Empirical results using PPP conversions

In this section, we briefly report model specifications and estimations for health expenditure using PPP-converted HCE and GDP.[16] We first conduct the structural break test with the linear model (7). AICc is minimized for the model with two breaks. The two identified break dates are 1974 and 1991, the same as in Table 5 using exchange rate-converted data. The full-sample income elasticity estimate is 1.603, which is higher than its counterpart using XRATE-converted data (1.321). All three sub-sample income elasticity estimates are also generally higher than those in Table 4 for the corresponding periods.

We also conduct a battery of semiparametric model specification tests. The GLR test results are very similar to the findings in Table 6 that use XRATE-converted data. We fail to reject, at the relatively conservative 10% level, the PLVC model (9) when

[16] To save space, the results in this part are not reported in Tables. However, they are available from authors upon request.

it is tested against the basic varying coefficient model (8). As before, the linear model (7) is clearly rejected. Therefore, are the four restrictive PLVC models in which the coefficient of one of the four explanatory variables is constant. All four elasticities are, therefore, subject to change over different income levels. Comparing the AICc values of H_o and H_A models, we find that the above GLR test results are fully supported by the model selection approach via the AICc criterion.

We also examine $\hat{\beta}_1(.)$, $\hat{\beta}_2(.)$, $\hat{\beta}_3(.)$, and $\hat{\beta}_4(.)$ over 730 different GDP values (full sample). The income elasticity estimate $\left(\hat{\beta}_1\right)$ increases from 1.08 to 1.84 as log GDP increases from 3.87 to 5.05, which is opposite to what we observed in Panel (a) of Fig. 2 for the same range of GDP. Nevertheless, at higher income levels, the estimate decreases with GDP, the same pattern as seen earlier. The mean and median values of $\hat{\beta}_1$ are 1.524 and 1.481, respectively. Both are lower than the above OLS estimate from the linear model, but higher than the estimates from PLVC regressions using XRATE-converted data. However, the PPP estimate of income elasticity carries much less variation (1.011–1.853) than the alternative one in Panel (a) of Fig. 2 (0.781–2.463). This is not surprising given that the sample observations of HCE and GDP are more homogeneous when converted by PPP than by exchange rates. Nevertheless, comparing with those in Fig. 2, we find that variations in the elasticities of PUBL, POP15, and POP65 are similar using both measures of income and health expenditure. This is as expected since these three variables are not subject to the influence of conversion.

Applying the same procedure as described earlier, we conduct the semiparametric structural break test with the PLVC model. Two breaks are found. The first break occurred in 1970. The second break occurred in 1983, which is different than the finding from the linear model (dated 1991). It is also quite far from the year 1990, the break date identified also using the semiparametric techniques but based on the other measure of data. According to the new break dates, we re-estimate the PLVC model with three sub-samples. For most levels of GDP in each sub-sample, the linear estimator of income elasticity falls outside the 90% bounds of the semiparametric estimates, meaning that the income elasticity is not constant. It varies most in the second regime (1971–1983) with its 10th percentile as low as 0.785 and 90th percentile as high as 2.220. We also examine the full sample estimate as well as the sub-sample estimates. Comparing to the corresponding results using the alternative measures of data (Fig. 3), we find more variations between the estimates from different samples. Nevertheless, we observe many of the outlying estimates also have large standard errors due to the sparseness of data in the local areas.

5 Further discussions

Projections of HCE involve enormous uncertainty due to dramatic innovations in medical technology and procedures (supply side), in addition to possibly smoother changes in consumers' preferences for health care (demand side) (OECD 2006). Contributing factors also include various reforms in health care delivery systems, financing, etc., which may induce changes in both supply and demand. Nevertheless, historical

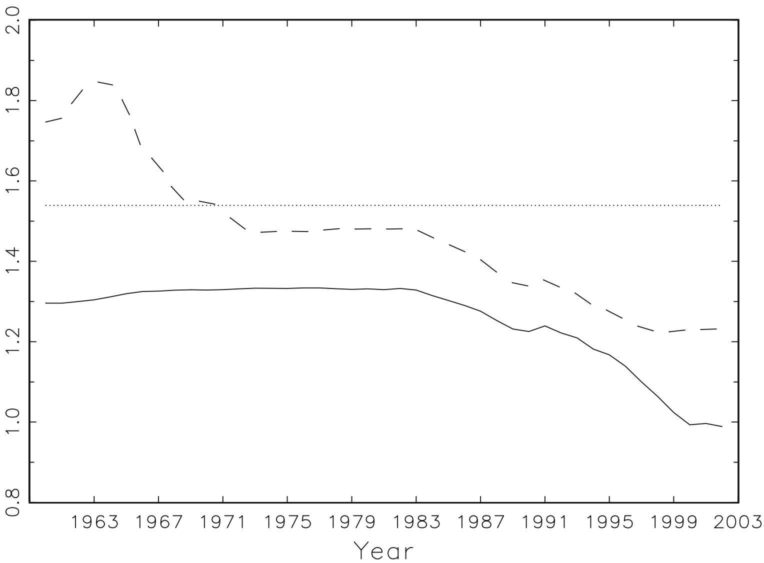

Fig. 4 Income elasticity estimates for U.S. *Note*: The *dotted line* is the estimate from the linear model (1) using the U.S. data alone. The *solid line* represents varying estimates for the U.S. from the semiparametric PLSC model (6) using full sample OECD data (foreign exchange rate-converted). The *dashed line* is based on the PPP-converted data

evidence, properly assembled, may still shed some light on this complicated process. As a demonstration, this section provides a brief discussion related to U.S. health expenditure.

We mentioned earlier that for the long-range projection purpose, (U.S.) Medicare Trustees Reports assume income elasticity of HCE to be 1.333. How well does this assumption hold from the international and historical perspectives? In Fig. 4, we plot the income elasticity estimate from the linear model based on U.S. data alone (which is a constant), and the semiparametric estimates for the U.S. based on the full sample of OECD data using two different conversion factors. The simple linear estimate is 1.539, higher than the assumed 1.333. Using PPP-converted OECD data, the semiparametric estimate is larger than 1.333 for most of the 43 years. Nevertheless, it starts to decrease in the early 1980s and appears to stabilize around 1.23, which is fairly close to 1.333, in the late 1990s. Using the exchange rate-converted data, the semiparametric estimate of income elasticity falls to around unit in the last few years. However, it is much higher in the first half of the sample (1960–1985), ranging from 1.30 to 1.33. Also, if the reduction in income elasticity during the 1990s is primarily attributable to an increasing share of individuals who participate in managed care relationships (HMOs), then once the market penetration of managed care stabilizes, the linkage between income and health expenditure may rebound (Rettenmaier and Wang 2006). Overall, the evidence presented in Fig. 4, compounded by the aging population, seems to suggest that the long-run growth rate of HCE assumed by the Review Panel is reasonable despite the recent slowdown in the U.S. and some other developed countries.

6 Conclusions

The issue of parameter instability in statistical models has attracted much attention from both theoreticians and applied researchers. Exploring recent advances in the theory of change-point estimation in linearly dependent processes and in the nonparametric statistics, we proposed and evaluated the possibility of using the standard model selection approach to test for structural breaks in semiparametric varying coefficient models. Monte Carlo evidence shows that the proposed procedure can be a useful tool for applied researchers.

We reexamined the relationship between HCE and income for 22 OECD countries. We found that the relationship was highly nonlinear and has been subject to both types of structural changes: smooth changes and regime shifts. Therefore, neither a semiparametric model only allowing for smooth changes in parameters nor a linear model allowing for structural breaks may be sufficient to capture the underlying nonlinearity in the relationship. This result highlights the difficulty in modeling the seemingly straightforward relationship between HCE and income.

Our results have implications for health policy study. First, as the income elasticity is not constant but varies with income levels, the share of a nation's resources allocated to heath industry is likely to change in the course of economic development. Second, in forming their long-term forecasts of HCE, currently less-developed countries could and should study the experiences of more developed countries. However, caution must be taken since even the estimated nonlinear relationship may change significantly over time.

Acknowledgments The authors thank two referees for valuable comments that greatly improved the paper, and Amy Hopson and Brian Alg for editorial assistance.

Appendix A: Sample period of OECD health expenditure data

Country	Data availability
Australia	1960, 1963, 1966, 1969, 1971–2002
Austria	1960–2002
Belgium	1995–2002
Canada	1960, 1965, 1970–2002
Denmark	1971–2002
Finland	1960–2002
France	1960, 1965, 1970, 1975, 1980, 1985, 1990–2002
Germany	1970–1990, 1992–2002
Greece	1970, 1980, 1987–2002
Iceland	1960–2002
Ireland	1960–2002
Italy	1988–2002
Japan	1960–2002
Netherlands	1972–2002

Appendix A: continued

Country	Data availability
New Zealand	1970–2002
Norway	1960–2002
Portugal	1970–2002
Spain	1960–2002
Sweden	1970–2002
Switzerland	1985–2002
United Kingdom	1960–2002
United States	1960–2002

Source: OECD *Health Data* (2006)

References

Ahmad I, Leelahanon S, Li Q (2005) Efficient estimation of a semiparametric partially linear varying coefficient model. Ann Stat 33:258–283

Andrews DWK, Lee I, Ploberger W (1996) Optimal change point test for normal linear regression. J Econom 70:9–38

Bai J, Perron P (1998) Estimating and testing linear models with multiple structural changes. Econometrica 66:47–78

Bai J, Perron P (2006) Multiple structural change models: a simulation analysis. In: Corbea D, Durlauf S, Hansen BE (eds) Econometric theory and practice: frontiers of analysis and applied research. Cambridge University Press, Cambridge pp 212–237

Cai Z, Kuan CM(2005) Nonparametric modeling for conditional capital asset pricing models (Unpublished manuscript)

Cai Z, Fan J, Yao Q (2000a) Functional-coefficient regression models for nonlinear time series. J Am Stat Assoc 95:941–956

Cai Z, Fan J, Li R (2000b) Efficient estimation and inferences for varying-coefficient models. J Am Stat Assoc 95:888–902

Carrion-i-Silvestre JL (2005) Health care expenditure and GDP: are they broken stationary? J Health Econ 24:839–854

Chou SY, Liu JT, Huang CJ (2004) Health insurance and savings over the life cycle—A semiparametric smooth coefficient estimation. J Appl Econom 19:295–322

Di Matteo L (2003) The income elasticity of health care spending: a comparison of parametric and nonparametric approaches. Eur J Health Econ 4:20–29

Di Matteo L (2005) macro determinants of health expenditure in U.S. and Canada: assessing the impact of income, age distribution and time. Health Policy 71:23–42

Fan J, Huang T (2005) Profile likelihood inferences on semi-parametric varying-coefficient partially linear models. Bernoulli 11:1031–1057

Gerdtham UG, Jönsson B (2000) International comparisons of health expenditure: theory, data and econometric analysis. In: Culyer AJ, Newhouse JP (eds) Handbook of health economics, vol 1A, Chap 1. Elsevier Sciences, Amsterdam pp 11–53

Hall RE, Jones CI (2007) The value of life and the rise in health spending. Q J Econ 122:39–72

Hastie TJ, Tibshirani RJ (1999) Generalized additive models. Chapman & Hall/CRC, London

Hitiris T, Posnett J (1992) The determinants and effects of health expenditure in developed countries. J Health Econ 11:173–181

Hurvich CM, Simonoff JS, Tsai CL (1998) Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. J R Stat Soc Ser B 60:271–293

Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. Biometrika 76:297–307

Jansen DW, Li Q, Wang Z, Yang J (2008) Fiscal policy and asset markets: a semiparametric analysis. J Econom 147:141–150

Jewell T, Lee J, Tieslau M, Strazicich MC (2003) Stationarity of health expenditures and GDP: evidence from panel root tests with heterogeneous structural breaks. J Health Econ 22:313–323

Li Q, Huang C, Li D, Fu T (2002) Semiparametric smooth coefficient models. J Bus Econ Stat 20:412–422

Li Q, Wang S (1998) A simple consistent bootstrap test for a parametric regression function. J Econom 87:145–165

Liu RY (1988) Bootstrap procedures under some non i.i.d. models. Ann Stat 16:1696–1708

Liu J, Wu S, Zidek JV (1997) On segmented multivariate regressions. Stat Sin 7:497–525

Narayan PK (2006) Examining structural breaks and growth rates in international health expenditures. J Health Econ 25:877–890

Newhouse JP (1992) Medical care costs: how much welfare loss? J Econ Perspect 6:3–21

OECD (2006) Projecting OECD health and long-term care expenditures: What are the main drivers? OECD Economics Department working papers, No. 477, OECD Publishing, Paris

Okunade AA, Murthy VNR (2002) Technology as a 'major driver' of health care costs: a cointegration analysis of the Newhouse conjecture. J Health Econ 21:147–159

Perron P (2006) Dealing with structural breaks. In: Patterson K, Mills TC (eds) Palgrave handbook of econometrics, vol 1: econometric theory. Palgrave Macmillan, New York pp 278–352

Pesaran MH, Timmermann A (2007) Selection of estimation window in the presence of breaks. J Econom 137:134–161

Rettenmaier AJ, Wang Z (2006) Persistence in medicare reimbursements and personal medical accounts. J Health Econ 25:39–57

Robinson PM (1988) Root-n-consistent semiparametric regression. Econometrica 56:931–954

Technical Review Panel on the Medicare Trustees Reports. Review of assumptions and methods of the Medicare Trustees' Financial Projections. December 2004, Washington DC

Ullah A (1985) Specification analysis of econometric models. J Quant Econ 1:187–209

Wang Z (2006) The joint determination of the number and the type of structural changes. Econ Lett 93(2):222–227

Yao YC (1988) Estimating the number of change-points via Schwarz' criterion. Stat Prob Lett 6:181–189