

Hartwig, Jochen; Sturm, Jan-Egbert

Working Paper

An outlier-robust extreme bounds analysis of the determinants of health-care expenditure growth

KOF Working Papers, No. 307

Provided in Cooperation with:

KOF Swiss Economic Institute, ETH Zurich

Suggested Citation: Hartwig, Jochen; Sturm, Jan-Egbert (2012) : An outlier-robust extreme bounds analysis of the determinants of health-care expenditure growth, KOF Working Papers, No. 307, ETH Zurich, KOF Swiss Economic Institute, Zurich, <https://doi.org/10.3929/ethz-a-007337182>

This Version is available at:

<https://hdl.handle.net/10419/80829>

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.

KOF Working Papers

An outlier-robust extreme bounds analysis of the determinants
of health-care expenditure growth

Jochen Hartwig,^a and Jan-Egbert Sturm^{a, b}

^a KOF Swiss Economic Institute, ETH Zurich, Switzerland

^b CESifo Munich, Germany

KOF

ETH Zurich
KOF Swiss Economic Institute
WEH D 4
Weinbergstrasse 35
8092 Zurich
Switzerland

Phone +41 44 632 42 39
Fax +41 44 632 12 18
www.kof.ethz.ch
kof@kof.ethz.ch

An outlier-robust extreme bounds analysis of the determinants of health-care expenditure growth

Jochen Hartwig^{*,a} and Jan-Egbert Sturm^{a,b}

^a *KOF Swiss Economic Institute, ETH Zurich, Switzerland*

^b *CESifo Munich, Germany*

This version: June 2012

Abstract

Hartwig (2008) has presented empirical evidence that the difference between real wage growth and productivity growth at the macroeconomic level is a robust explanatory variable for deflated health-care expenditure growth in OECD countries. In this paper, we test whether this finding is robust to the inclusion of additional covariates, applying different versions of Extreme Bounds Analysis (EBA) to data for 33 OECD countries over the period 1970-2010. As far as it is statistically feasible, all macroeconomic and institutional determinants of health-care expenditure growth that have been suggested in the literature are included in the EBA. Furthermore, we analyse to what extent outliers in the data influence the results using an outlier-robust MM estimator. Our results confirm Hartwig's earlier finding. A number of additional both covariate- and outlier-robust determinants are also identified.

JEL classification: C12; C23; I10

Keywords: Health-care expenditure; 'unbalanced growth'; Extreme Bounds Analysis; MM estimator, OECD panel

* Address of corresponding author: KOF Swiss Economic Institute, ETH Zurich, Weinbergstrasse 35, 8092 Zurich, Switzerland. Phone: +41 44 632 7331. Fax: +41 44 632 1218. E-mail: hartwig [at] kof.ethz.ch

1. Introduction

While the number of studies on the determinants of health-care expenditure (HCE) is substantial (see Gerdtham and Jönsson, 2000),² the number of determinants uncovered by this literature is not. Back in 2003, an OECD working paper concluded from a literature review that “[w]hile *income* (generally proxied by GDP *per capita*) is the main driving force in all studies” (Docteur and Oxley, 2003, p. 73), there was little consensus on the income elasticity of HCE. Neither was there any consensus on the significance of the ageing of the population, nor of labour market variables, life-style factors, technological change or institutional arrangements in national health systems for the rise in HCE. The various studies usually test different sets of explanatory variables, highlighting the problem of model uncertainty in this field of research. When the same variables are tested, they are sometimes found to be statistically significant, sometimes not. For certain variables, not even the sign on the coefficient is stable across studies.³

Answering Gerdtham and Jönsson’s (2000) call for strengthening the theoretical basis for the macroeconomic analysis of health expenditure, Hartwig (2008) suggested revisiting Baumol’s (1967) model of unbalanced growth to this end. In a nutshell, Baumol’s model is a two-sectoral neoclassical growth model that has a sector with high productivity growth, which carries a low weight in the economy, and a sector with low productivity growth, which carries a high weight. Baumol assumes that the high-productivity-growth sector, which he calls the ‘progressive sector’, sets the rate at which nominal wages grow for both sectors: they grow in line with productivity in the progressive sector. Given the weights of the two sectors, however, the average productivity growth rate will resemble the productivity growth rate of the ‘non progressive’ sector. Consequently, the excess of average wage growth over average productivity growth in the economy approximately shows how much unit costs in the low-productivity-growth sector rise and how much revenues consequently have to be raised there. So if the health sector is representative of the low-productivity-growth sector, the excess of nominal wage growth over productivity growth should drive health-care expenditure growth. Hartwig (2008) confirms this empirically using data from a panel of 19 OECD countries over the time period 1971-2003.

The shift of expenditures into (services) sectors with low productivity growth like health care and education – also called ‘stagnant’ sectors – is known as ‘Baumol’s cost disease’. It is a nominal story. Prices do not (have to) rise in the progressive sector, while they (have to) rise in the stagnant sector in order to keep real wages in line with stagnant productivity. Therefore, the share of the stagnant sector in total expenditure rises over time. This is the

² Hartwig (2008) offers an overview.

³ See the annex in Docteur and Oxley (2003).

‘cost disease’. In Baumol’s model, the stagnant sector gains weight for purely nominal reasons. In ‘real’ (deflated) terms, Baumol assumes the proportion of the two sectors to stay constant. For this reason, Hartwig’s (2008) preferred specification for testing Baumol’s model was regressing nominal per-capita health-care expenditure growth on the difference between nominal wage growth and productivity growth. However, as basically the entire empirical literature on the determinants of health-care expenditure (growth) models deflated HCE,⁴ Hartwig (2008) also deflates both HCE and wages by the GDP deflator. He finds the difference between real wage growth and productivity growth to be a statistically and economically significant explanatory variable for real per-capita HCE growth – a finding that is robust to including per-capita GDP growth (‘the main driving force’ of HCE in earlier studies) as an explanatory variable.

This paper investigates whether this finding is robust to the inclusion of other covariates. More than 70 such covariates have been suggested in the literature. Here we use a statistical technique known as Extreme Bounds Analysis (EBA) that is specifically designed to investigate how robust results are to changing the set of covariates. EBA has originally been applied to the field of economic growth (see Levine and Renelt, 1992, Sala-i-Martin, 1997, Sturm and de Haan, 2005), where – much like in the field of health-care expenditure – model uncertainty is very high, and a large number of potential determinants has been suggested by the literature. EBA has since spread to fields of research other than economic growth like political economy (see Dreher et al., 2009a, 2009b, Gassebner et al., 2012, Moser and Sturm, 2011 and Sturm et al., 2005) or environmental economics (see Gassebner et al., 2011). As far as we know, there is no application of EBA to the field of health economics so far.

As EBA utilises the classical least squares estimator, the presence of outliers in the data can strongly distort the results and lead to unreliable conclusions. To take care of this problem, we check our OLS results against results generated by an outlier-robust MM estimator: a sophisticated method to estimate parameters that are insensitive to the inclusion of outliers.

The paper is structured as follows. The next section discusses our dataset. Section 3 explains the methodologies of Extreme Bound Analysis and the MM estimator used, and section 4 presents the results. Section 5 concludes.

2. Data

For inclusion in our Extreme Bounds Analysis, we aim at complete coverage of health-care expenditure determinants that have been suggested by the literature. Hartwig (2008) divides this literature into two stages. In the first stage, the close correlation between HCE and GDP

⁴ Recently, though, Bates and Santerre (2012) tested the robustness of the results obtained by Hartwig (2008) using *nominal* US state level data. Including five additional covariates, they confirm the significance of the ‘cost disease’ for explaining the rise in nominal HCE.

was established, the income elasticity of HCE was investigated and covariates were tested for their significance. Gerdtham et al. (1998) delivers a comprehensive list of potential determinants from this first stage. Besides GDP, socio-demographic and technological factors as well as a long list of institutional variables (mostly dummy variables) pertaining to specifics of the national health systems are included in the empirical model. For example, one of the institutional dummy variables takes the value of one for countries (and years) with fee-for-services as the dominant means of remuneration in primary care and zero otherwise. We reproduce and update these variables to the extent possible.

The data source for most of the non-dummy variables is the OECD Health database, which also contains economic, socio-demographic and even technological data (as long as they are health-related).⁵ For the institutional dummy variables, however, we rely on the information in Gerdtham et al. (1998) on how to construct them. Gerdtham et al.'s dataset covers 24 OECD countries and the time period 1970-1991. We carry forward these time series with information from Christiansen et al. (2006). They use almost the same set of explanatory variables as Gerdtham et al. (1998) and give information on the institutional characteristics of health systems for the period 1980-2003. Christiansen et al. (2006) investigate European Union (EU) instead of OECD countries, however. This means that for the OECD countries that are also EU members, we can ideally establish time series for the institutional dummy variables that cover the period 1970-2003. For the non-EU OECD countries, however, the series end in 1991.

Finally, we used information from Paris et al. (2010) to further update our data on the institutional setting. Paris et al. (2010) do not report time series data; they describe the state of the national health systems for 29 OECD countries in 2008/09. This gives us data points for the institutional dummy variables for 2008/09. Furthermore, we assume that if the value we derive from Paris et al. for 2008/09 – 0 or 1 – is the same as the value for 2003 we get from Christiansen et al. (2006) or the value for 1991 we get from Gerdtham et al. (1998) for the non-EU OECD countries, then there has been no change in the meantime, and we close the gaps in the time series with the respective values. If the values are not the same, however, we conclude that there has been a change in the institutional setting at some unknown point in time, and we take the values for the in-between years as missing.

⁵ We mostly used the 2011 version of the OECD Health database, which is the first version that is only available online. We noticed however, that the list of variables available online is not as comprehensive as the list available in earlier CD ROM versions of the database. So whenever we could not find necessary data online, we referred to the 2010 CD ROM version of the OECD Health Database. Our dataset covers the following 33 countries: Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, the United Kingdom and the United States.

After this digression into the construction of the institutional dummy variables, let us return to the determinants of health-care expenditure that have been suggested in the literature. As said, Hartwig (2008) identified a second stage of the literature that started in the mid-1990s. During this stage, the focus shifted towards the time series properties of HCE and GDP; (panel) unit root and cointegration tests as well as structural breaks tests were performed. To find new explanatory variables for HCE was not the main interest of researchers during this stage.⁶ Nevertheless, for our purpose we need an overview over the determinants of HCE that have been suggested during the second stage. We draw on Martín et al. (2011) for this purpose. They performed a systematic search for literature on the determinants of HCE in medical databases and principal health economics journals over the period 1998-2007. Their review of 20 studies yields the determinants that have been suggested in the literature since Gerdtham et al. (1998) and prior to Hartwig (2008). Some of the studies reviewed by Martín et al. (2011) are micro-level studies which focus on the question whether rising HCE with age is caused by aging as such or by ‘proximity to death’. These studies typically analyse micro datasets from health insurance companies to compare ex post the health-care costs for survivors with costs for those who have died. As our focus is on the macroeconomic level, we leave aside those studies reviewed by Martín et al. (2011) which focus on the micro-level.⁷

Table 1 lists all macro-level explanatory variables for HCE that have been suggested in the literature we reviewed. In the top-down dimension, the table has seven sections. Section 1 gives the dependent variable: real per-capita HCE (HCEPC). Note that, since our aim is to test the robustness of the results in Hartwig (2008), but also in order to avoid statistical problems involved in modelling non-stationary data, we convert all level variables into growth rates.⁸ This means for instance that, although HCEPC is listed in the table, we will model the growth rate (log difference) of HCEPC. Variables that are shares – for instance GDP shares or population shares – will enter our model in first differences, in other words, we will use the change in the shares as explanatory variables.

<Insert Table 1>

Section 2 in the top-down dimension of Table 1 lists the two potential determinants of per-capita HCE growth we are mainly interested in. Per-capita GDP (growth) has long been

⁶ As a matter of fact, Potrafke (2010) argues that Hartwig (2008) has opened up a ‘third stage’ of the literature in which the interest in new – and hopefully theory-based – explanatory variables for HCE takes centre stage.

⁷ See also Christansen et al. (2006, pp. 2-7) for the distinction between micro- and macro-level studies.

⁸ Note that Gertham and Jönsson (2000, p. 48), in their contribution to the *Handbook of Health Economics*, conclude: “It may also be important to replicate Gerdtham et al. (1998) with extended data sets and also with respect to growth rates of health expenditure as in Barros (1998), and new methods ...”. This epitomises our approach.

acknowledged to be an important driver of per-capita HCE (growth) – Docteur and Oxley (2003) call it “the main driving force in all studies”. The second variable is our ‘Baumol variable’, as will be explained shortly. We want to check whether the explanatory power of these two variables is robust to the inclusion of other covariates.⁹

Table 1 does not list Hartwig’s (2008) ‘Baumol variable’ – the difference between real wage growth and productivity growth –, but another variable: the change in income distribution. Why? – In our preparatory data analysis, we calculated the correlations between all explanatory variables. We did this because the main weakness of the method of Extreme Bounds Analysis is that it cannot decently cope with multicollinearity. Two highly correlated variables should therefore ideally not enter both the EBA. We found a very high correlation ($R = 0.83$) between Hartwig’s (2008) ‘Baumol variable’ and the change in income distribution – an explanatory variable that was suggested by Karatzas (2000), who regresses the level of real per-capita health-care expenditure on, among other variables, the income distribution as expressed by the ratio of nominal wages to nominal GDP. Karatzas (2000, p. 1087, fn. 10) claims that “the income distribution variable constitutes a key predictor since it affects the value of the income elasticity of the per capita real health care spending”. Although he does not link his income distribution variable to Baumol’s model of unbalanced growth in any way, it is clear that the change in income distribution as defined by Karatzas measures the same thing as the ‘Baumol variable’ constructed by Hartwig: Whenever real wages grow more (less) than productivity, the wage share in GDP rises (declines). Since we should not include both the ‘Baumol variable’ and the change in the income distribution in the EBA because of their high correlation, we need to choose one of them. We opt for the change in the income distribution because we have more observations for this variable than for the difference between real wage growth and productivity growth (924 vs. 785). Note, however, that this choice does not affect any of our results in a meaningful way.

Section 3 in the top-down dimension of Table 1 lists socio-demographic factors that have been suggested as explanatory variables for HCE. These are dominated by population shares. We choose not to include all these population shares in the EBA for the following reasons. First, having too many population variables will put too large a weight on them in the EBA results. If many Z vector variables are population variables, then a large share of the regressions will consist of combinations of population variables. That is creating an imbalance. Secondly, variables that are substantially overlapping, like POP6574 and POP6584, are bound to be highly correlated generating multicollinearity problems and thereby reducing the likelihood that any of these variables will turn out to be significant. So

⁹ In the EBA jargon, variables in the ‘M vector’ are included in all regressions. All the other explanatory variables, which will only be used in a sub-set of regressions, are called ‘Z vector variables’ (see section 3 for details).

we decided to include only three population variables: the share of the population 4 years and under (covering potentially higher than average HCE for children), the share of the population between 65 and 79 years (covering potentially rising health-care costs after retirement) and the share of the population over 80 years (covering potential extra costs for long-term care).

The remaining sections in the top-down dimension of Table 1 list institutional, price, and technological variables that have been suggested as determinants for HCE, and finally the variables needed to construct some of the explanatory variables. The shaded areas of Table 1 contain variables that we either modify vis-à-vis the literature (light grey) or drop (dark grey). The column headed ‘Remarks’ gives some further details. Also, this column indicates whether the number of observations for a variable is low. We exclude variables with less than 350 observations from our preferred model, which leaves us with 38 explanatory variables.¹⁰ Table 2 provides descriptive statistics for all variables.

<Insert Table 2>

3. Methodology

To examine the sensitivity of the individual variables on per capita HCE growth, we apply (variants of) extreme bounds analysis, as suggested by Leamer (1985) and Levine and Renelt (1992). This approach has been widely used in the economic growth literature. The central difficulty in this research – which also applies to the research topic of the present paper – is that several different models may all seem reasonable given the data but yield different conclusions about the parameters of interest. Equations of the following general form are estimated:

$$(1) \quad Y = \alpha M + \beta F + \gamma Z + u,$$

where Y is the dependent variable; M is a vector of ‘standard’ explanatory variables; F is the variable of interest; Z is a vector of up to three possible additional explanatory variables, which the literature suggests may be related to the dependent variable; and u is an error term. The extreme bounds test for variable F states that if the lower extreme bound for β – the lowest value for β minus two standard deviations – is negative, and the upper extreme bound for β – the highest value for β plus two standard deviations – is positive, the variable F is not robustly related to Y .

As argued by Temple (2000), it is rare in empirical research that we can say with certainty that one model dominates all other possibilities in all dimensions. In these circumstances, it

¹⁰ We include the variables with less than 350 observations in an alternative model for which the results are available on request. The main conclusions are not affected by this.

makes sense to provide information about how sensitive the findings are to alternative modeling choices. Extreme bounds analysis (EBA) provides a relatively simple means of doing exactly this. Still, the approach has been criticized in the literature. Sala-i-Martin (1997) argues that the test applied poses too rigid a threshold in most cases. Assuming that the distribution of β has at least some positive and some negative support, the estimated coefficient changes signs if enough different specifications are considered. We therefore report not just the lowest and highest coefficient estimates, but also the percentage of the regressions in which the coefficient of the variable F is significantly different from zero at the 10 percent level. Moreover, instead of analyzing just the extreme bounds of the estimates of the coefficient of a particular variable, we follow Sala-i-Martin's (1997) suggestion to analyze the entire distribution. Following this suggestion, we not only report the unweighted parameter estimate of β , but also the unweighted cumulative distribution function (CDF(0)), that is, the fraction of the cumulative distribution function lying on one side of zero.¹¹

Whereas EBA or any of its alternatives can deal with model uncertainty, i.e. whether results are robust to the selection of covariates, it does not take care of the inclusion of so-called outlying, or unusual, observations. Although empirically oriented economists often test the residuals of their regressions for heteroskedasticity and structural change, they hardly ever test for unusual observations. Still, it is quite likely that many of the data sets we use contain unusual observations. Some authors therefore suggest using so-called robust estimation techniques (see, for example, Temple, 1998 or Sturm and de Haan, 2005).¹² Robust estimators can be thought of as trying to seek out the most coherent part of the data, the part best approximated by the model being estimated. These estimators will not be led astray by outliers.

In the usual presentation of outliers it is stressed that one or more observations may be measured with a substantial degree of error. As Swartz and Welsch (1986, p. 171) put it: "OLS and many other commonly used maximum likelihood techniques have an unbounded influence function; any small subset of the data can have an arbitrarily large influence on their coefficient estimates. In a world of fat-tailed or asymmetric error distributions, data

¹¹ Sala-i-Martin (1997) proposes using the (integrated) likelihood to construct a weighted CDF(0). However, the varying number of observations in the regressions due to missing observations in some of the variables poses a problem. Sturm and de Haan (2001) show that this goodness of fit measure may not be a good indicator of the probability that a model is the true model, and the weights constructed in this way are not equivariant to linear transformations in the dependent variable. Hence, changing scales result in rather different outcomes and conclusions. We thus restrict our attention to the unweighted version. Furthermore, for technical reasons – in particular our unbalanced panel setup – we are unable to use extensions of this approach, like Bayesian Averaging of Classical Estimates (BACE), as introduced by Sala-i-Martin et al. (2004), or Bayesian Model Averaging (BMA).

¹² Robustness in this case is defined in terms of the observations included in the regression. Of course, model uncertainty and the role of outliers may be related as outliers can have consequences for the choice of variables.

errors and imperfectly specified models, it is just those data in which we have the least faith that often exert the most influence on the OLS estimates.”

Following Barnett and Lewis (1994, p. 316), we define an outlier as an observation ‘lying outside’ the typical relationship between the dependent and explanatory variables revealed by the remaining data. For instance, point A in Figure 1(a) is clearly an outlier. Outliers in the dependent variable – i.e. in the y-direction – often possess large positive or large negative residuals, which are easy to detect by plotting the residuals. Observations may be outlying for several reasons. The most obvious one involves problems with the quality of the data. Outliers in the explanatory variables may be more problematic than outliers in the dependent variable. As Figure 1(b) shows, an unusual observation in the x-direction (B) can actually tilt the OLS regression line. In such a case we call the outlier a (bad) leverage point. Note that looking at the OLS residuals cannot discover bad leverage points. If a leverage point tilts the regression line, deleting the points with the largest OLS residuals implies that some ‘good’ points would be deleted instead of the ‘bad’ leverage point.

<Insert Figure 1>

Basically, there are two ways to deal with outliers: regressions diagnostics and robust estimation. Diagnostics are certain statistics mostly computed from the OLS regression estimates with the purpose of pinpointing outliers and leverage points. Often the unusual observations are then removed or corrected after which an OLS analysis on the remaining observations follows. When there is only one unusual observation, some of these methods work quite well. However, single-case diagnostics are well known to be inadequate in the presence of multiple outliers or leverage points (Temple, 2000).

Take for instance Figure 1c. Deleting either of the two outliers will have little effect on the regression outcome and will therefore not be spotted by the single-case diagnostics. The potential effect of one outlying observation is clearly masked by the presence of the other. Testing for groups of observations to be influential might solve this masking effect problem. However, a serious problem in the multiple observation case is how to determine the size of the subset of jointly influential observations. Suppose we are interested in detecting all subsets of size $m=2,3,\dots$ of observations that are considered to be jointly outliers and/or high-leverage. A sequential method might be useful, but where to stop? In the multiple observation case the number of possible subsets for which each diagnostic measure of interest can be computed is: $n!/[m!(n-m)!]$, where n is number of observations. For $m=5$ and $n=50$ this already results in over 2 million diagnostics. Therefore we prefer so-called robust regression techniques that employ estimators that are not strongly affected by (groups of) outliers.

Over the last few years, several robust-to-outliers methods have been proposed in the statistical literature. High break-down point estimators, like Least Median of Squares, Least

Trimmed Squares, or so-called S-estimators are able to resist a contamination of up-to 50 percent of outliers. However, these estimators are known for their low efficiency at a Gaussian error distribution. To cope with this, Yohai (1987) introduced MM-estimators that combine high-breakdown point and a high efficiency (Verardi and Croux, 2008).

Instead of squaring the residuals in the minimization process as done with OLS, within the class of S- and MM-estimators each residual undergoes a transformation dampening the influence of large residuals. For the MM-estimator, this normalizing scale is robustly determined in a first step using a so-called S-estimator that has excellent robustness properties. We use the algorithm as implemented by Verardi and Croux (2008). It starts by randomly picking N subsets of k observations, where k is the number of regression parameters to estimate.¹³ For each subset, the equation that fits all points perfectly is obtained yielding a trial solution of an outlier robust S-estimator (using the Tukey Biweight loss function). On the basis of the residuals, a scale estimate is obtained for each subset. An approximation for the scale estimate to be used in the final MM-estimation is then derived from the subset that leads to the smallest scale. Following Maronna et al. (2006), we chose a breakdown point of 50 percent and an efficiency relative to the Gaussian case of 70 percent, which requires setting the Tukey Biweight parameter to 1.547 for the first step S-estimator and to 2.697 in the second step.¹⁴ As far as inference is concerned, standard errors robust to heteroskedasticity are computed according to the formulas available in the literature (see e.g. Croux et al., 2008).

4. Results

In a first step we regress the growth rate of deflated per-capita health-care expenditure (dlhcepc) only on our two ‘M vector’ variables per-capita real GDP (dlgdppc) growth and the ‘Baumol variable’ change in income distribution (did). The first four columns of Table 3 show the results for the OLS estimations without fixed effects, with fixed country effects, fixed year effects and two-way fixed effects respectively. The remaining columns report results for the MM estimations, which identify a varying number of outliers depending on whether fixed effects are included or not and, if yes, which kind of fixed effects. As the table shows, both explanatory variables are always significant at the 1% level. The estimated coefficients suggest that a one percentage point increase in per-capita GDP growth raises the growth rate of per-capita HCE by 0.5-0.7 percentage points. Against the background of the old debate about whether health care is a luxury or a necessity, this finding is in line with “recent studies using pooled time-series cross-section data and a wider range of explanatory

¹³ The number N of sub-samples to generate is chosen to guarantee that at least one subset without outliers is selected with high probability (see Salibián-Barrera and Yohai (2006) on how to achieve this).

¹⁴ Increasing Gaussian efficiency would lead to a higher estimation bias making comparison between the OLS and MM-regression based results more difficult.

variables (which) suggest elasticities near or less than one” (Docteur and Oxley, 2003, p. 73). In other words, our results lend no support to the view that health care is a luxury.

The coefficients on the income distribution variable suggests that a one percentage point increase in the wage share raises the growth rate of per-capita HCE by 1.2-1.6 percentage points. To allow for a comparison with the results in Hartwig (2008), we replaced the variable ‘did’ with the ‘Baumol variable’ used in that paper – the difference between real wage growth and productivity growth. The coefficients we found on that variable were also thoroughly significant at the 1% level with values around 0.5. This is somewhat lower than the coefficient values found in Hartwig (2008), which were in the range of 0.8-0.9.¹⁵

<Insert Table 3>

To check whether our findings are robust to the inclusion of additional explanatory variables, we opt for Sala-i-Martin’s version of Extreme Bounds Analysis. As before, we apply both the OLS and the MM estimators with and without fixed effects. By including up to three additional variables from the ‘Z vector’, we estimate 66’711 OLS regressions. Each ‘Z vector’ variable is included in 7’176 of them; the two ‘M vector’ variables are of course always included. The MM estimation is very time consuming. Therefore, we have restricted the maximum number of additional ‘Z vector’ variables to two and thereby produced 666 combinations in which we tested the ‘M vector’ variables.

The left-hand-side of Table 4 shows the results for the OLS estimation with both fixed country and year effects.¹⁶ The first column gives the average of the estimated β -coefficients for that particular variable. The next two columns report the lowest and highest estimated coefficients. Column (4) gives the percentage of the regressions in which the coefficient on the variable is significantly different from zero at the 10 percent level. The final column reports the results of the cumulative distribution function (CDF) test, in other words, it shows the percentage of the cumulative distribution function lying on one side of zero. CDF(0) indicates the larger of these areas under the density function either above or below zero, so it will always lie between 50% and 100%. The variables in Table 4 are ordered based on their CDF(0).

¹⁵ A coefficient of 0.6 was found for the sub-period 1982-92. There are a number of reasons why the coefficients could differ, for instance the different sample of countries (33 OECD countries in this study vs. 19 in Hartwig, 2008) and the different estimation period (1970-2010 vs. 1971-2003). Also, Hartwig (2008) used random, not fixed effects and the growth rate of the GDP deflator as covariate.

¹⁶ With respect to our main concern – to find out whether GDP growth and the ‘Baumol variable’ are robust explanatory variables for HCE growth – the estimations with fixed country effects, fixed year effects, and no fixed effects yield the same result as the estimation with two-way fixed effects. The two-way fixed effects estimation identifies a slightly larger set of robust ‘Z vector’ variables than the other three approaches. All results are available from the authors upon request.

The first thing to note from Table 4 is that our baseline model is robust to the inclusion of additional explanatory variables. The estimated cumulative distribution function for the coefficients on the ‘Baumol variable’ (change in income distribution) completely lies on the right-hand side of zero; for real per-capita GDP growth this is almost the case. The underlying estimated coefficients are statistically significant in 99.9 and 99.6 percent of the cases, respectively. The average coefficients are 0.72 for per-capita GDP growth and 1.37 for the change in the wage share.

Secondly, only a few of the additional variables we have tested change their sign infrequently and can thus be counted as robust explanatory variables as well. If we apply the criterion that $CDF(0) > 0.95$,¹⁷ these variables are: (1) the dummy variable for countries with fee-for-services as the dominant means of remuneration in primary care (ffsa), (2) the growth in acute beds per 1’000 inhabitants (dlbedsi), (3) the change in the rate of unemployment (dunemp), (4) the growth in the number of patients undergoing renal dialysis per 100’000 population, (5) the change in the female participation ratio (dfpr), (6) the growth in per capita real expenditure on health administration (dlta) and (7) the growth in tobacco consumption in grams per capita 15+ (dltobc).

While most of the average coefficient signs on these seven variables are in line with prior expectations, this does not hold for two of them. Growth in the number of patients undergoing renal dialysis should raise health care expenditure, yet the sign on this variable is robustly negative. Also, an increase in the female participation ratio has been theorised to contribute to raising HCE because it leads to replacing unpaid care at home – mostly carried out by women – by for-pay services. Still, we find robustly negative coefficients on this variable. We will discuss in an instant whether our findings for these two variables – as well as for the other robust variables – are due to outliers in the data.

Before that, let us have a look at the statistical significance of the coefficients. The growth in acute beds per 1’000 inhabitants (dlbedsi) is a very robust explanatory variable, but it is significant in only 6.8% of the regressions. The other six robust variables are significant in between 26% (dltobc) and 60% of the regressions (dfpr, dlta). Other variables that are often significant, but not robust, are the change in the share of public health expenditure in total expenditure (dpuhes) and the change in the share of inpatient expenditure in total expenditure (dtexmc). Interestingly, the average coefficients on these two variables are negative.

Finally, before turning to the outlier-robust MM estimation, let us emphasize that most of the variables we tested are not robust, like for instance the age-structure variables and all institutional variables except the dummy variable for countries with fee-for-services as the

¹⁷ Sala-i-Martin applies the criterion $CDF(0) > 0.9$, but we consider this to be too low given the one-sidedness of the test.

dominant means of remuneration in primary care (ffsa). This does not come as a surprise, however, since it has already been pointed out in the introduction that the literature has not been very successful in uncovering robust explanatory variables for HCE (growth) apart from GDP (growth). For a review of this literature, again see the annex in Docteur and Oxley (2003).

<Insert Table 4>

The right-hand-side of Table 4 reports the results of the outlier-robust MM estimations. Again, we show results for the estimation with two-way fixed effects, but the main conclusions are not sensitive to the choice of fixed effects. We register that even after the removal of outliers, per-capita GDP growth and the ‘Baumol variable’ remain robust explanatory variables for per-capita HCE growth. The sign on the change in income distribution (did) is again positive in 100 percent of the regressions, the sign on GDP growth in 97.2 percent. However, compared to the OLS estimation the average coefficient value on GDP growth drops from 0.72 to 0.41, which means that health is even less of a luxury than was previously thought. Also, the percentage of the regressions in which per-capita GDP growth is significant drops from 99.6 to 80.1 percent. This is still a high value, but the significance of GDP growth is clearly somewhat sensitive to the exclusion of outliers. This is not the case for the ‘Baumol variable’, though, which remains significant in 98.2 percent of the regressions. The average coefficient value remains virtually unchanged vis-à-vis the OLS estimates.

On the other hand, only two of the seven ‘Z vector’ variables found to be robust in the OLS estimation ‘survive’ the switch to outlier-robust MM estimation. These are: (1) the dummy variable for countries with fee-for-services as the dominant means of remuneration in primary care (ffsa) and (2) the growth in per capita real expenditure on health administration (dlta). However, the MM estimation discloses three other robust ‘Z vector’ variables: (3) the lagged change in the government share (dgsh1), (4) the growth rate of the number of land traffic fatalities (dlaccident) and (5) the growth rate of gross expenditure on R&D (dlgerd).

The change in the share of inpatient expenditure in total expenditure (dtexmc) remains often significant, though not robust. The same is now true for the change in the insurance coverage of the population (dcovero). The average sign for the former variable remains negative; for the latter variable it is positive (as expected). The negative sign on ‘dtexmc’, which is in contrast to earlier findings (see Docteur and Oxley 2003, p. 74), might reflect a rearrangement visible, for instance, in Swiss data. Reforms to hospital remuneration have made inpatient treatment financially less attractive for Swiss hospitals. This has contributed to a surge in expenditures for outpatient treatment by hospitals in Switzerland over the past 20 years – these have been the fastest growing expenditure component in the Swiss health

system by a wide margin. If hospitals prefer outpatient treatment to inpatient treatment for financial reasons, the switch to outpatient treatment must be expected to boost total HCE growth instead of dampening it.

With respect to the signs of the five variables emerging as robust from the MM estimation, the positive sign on ‘ffsa’ conforms to prior expectations. Fee-for-service remuneration in primary care excites practitioners to offer a lot of treatments (including unnecessary ones), which raises costs. Also, that rising real expenditure on health administration (dlta) raises total HCE does not come as a surprise. The government share has been introduced into the literature as an explanatory variable by Hitiris, who sees it as a proxy for the level of development of a country, arguing that “rich countries experiencing a high rate of growth of income and public revenue would be expected to direct significant resources to public spending, including health care” (Hitiris, 1997, p. 3). He finds a positive and significant impact of the government share on the level of per-capita health-care expenditure. This is confirmed by Roberts (2000), as well as our results.¹⁸

R&D expenditure has been proposed as a proxy for technological progress – including progress in medical technologies – by Okunade and Murthy (2002). They expect a significantly positive sign on this variable – and find one. This is confirmed by our outlier-robust estimation. The only sign that seems odd at first sight is the negative sign on the growth rate of the number of land traffic fatalities. But this is probably what has happened: the death rate has dropped resulting in higher expenditure on those who are (now) severely injured.

As a final step in our empirical analysis, we include all variables that emerge as robust from either the OLS or the MM estimations into the baseline model in order to investigate, which of them are statistically significant. The right-hand-side of Table 5 reports the results of the outlier-robust MM estimations. Columns (5) and (6) show that *all* variables that emerged as robust explanatory variables for HCE growth either from the OLS EBA regressions (column 5) or from the MM EBA regressions (column 6) are statistically significant at least at the 10 percent level, while the two baseline variables remain highly significant with fairly stable coefficients across the three specifications.

This picture changes somewhat when the extended baseline models are estimated with OLS (see the left-hand-side of Table 5). Although the two baseline variables remain highly significant, their coefficient values vary more across the specifications. When the seven covariates that emerged as robust from the OLS EBA regressions are included, the coefficient

¹⁸ Note that we lagged the change in the government share (dgsh) – as well as the growth rate of per-capita real insurance premiums (dlins) – because of concerns over reverse causality. The un-lagged version of dgsh is also robust.

on the ‘Baumol variable’ jumps to a value above 2, and the coefficient on per capita GDP growth becomes greater than 1, indicating that health care might be a luxury after all. The most striking difference between the two parts of Table 5 is, however, that only four out of ten covariates are significant when the extended baseline models are estimated with OLS rather than with the outlier-robust MM estimator. These four covariates are: (1) the growth in acute beds per 1’000 inhabitants (dlbedsi), (2) the change in the rate of unemployment (dunemp), (3) the growth in the number of patients undergoing renal dialysis per 100’000 population – notably with a *positive* sign – and (4) the growth in per capita real expenditure on health administration (dlta). Obviously, the statistical significance of the other six covariates is impaired by outliers in our sample of data.

<Insert Table 5>

5. Conclusion

‘What drives health-care expenditure?’ was the question asked by Hartwig (2008). Many scholars had asked that question before, and many different drivers have been proposed by a flourishing literature. However, so far there has not been a systematic investigation into which of these proposed drivers are robust explanatory variables for health-care expenditure or its growth. This paper aims at closing this gap in research, applying the methodologies of Extreme Bounds Analysis and outlier-robust MM estimation to OECD data. We aimed at including all macroeconomic and institutional determinants of health-care expenditure (growth) that have been suggested by the literature in our EBA, however multicollinearity among some explanatory variables and lack of observations for others prompted us to modify or exclude a number of variables from our preferred EBA specification. Nevertheless, the latter contains 38 potential determinants.

Our results confirm the broad picture emerging from the literature that there are not many robust drivers of HCE growth. However, we were able to identify ten possible candidates. Four of them: the growth in acute beds per 1’000 inhabitants, the change in the rate of unemployment, the growth in the number of patients undergoing renal dialysis per 100’000 population and the growth in per capita real expenditure on health administration emerged as significant in both OLS and outlier-robust MM regressions. Also, we are able to confirm the long-standing insight originating from Newhouse (1977) that GDP (or income) drives health-care expenditure. In line with more recent studies, our results predominantly suggest that health-care is not a luxury at the macroeconomic level. What seems to be a new insight is that the impact of GDP growth on HCE growth is strengthened by outliers in the data. In our outlier-robust EBA estimation, GDP growth turns out insignificant in roughly 20 percent of the regressions.

The most important aim of our study, however, was to check the robustness of Hartwig's (2008) assertion that Baumol's (1967) model of 'unbalanced growth' offers a theoretical framework for explaining the rise in health spending. Note that we do not say: We wanted to check whether HCE was driven by Baumol's 'cost disease'. One simply cannot do that using deflated data. Hartwig (2008) uses nominal data in a first step and shows that nominal wage growth in excess of productivity growth drives nominal HCE growth. This is what Baumol's model predicts. The small (low-weight) manufacturing sector sets the rate of nominal wage growth for the overall economy. The large (heavy-weight) 'stagnant' sector with low (or no) productivity growth raises prices in line with rising nominal wage costs. Therefore, expenditures are channelled into the stagnant sector, the weight of which in nominal GDP becomes ever larger. This is the 'cost disease': it is about nominal flows. Baumol's model does not predict that the share of the stagnant sector in deflated data rises. On the contrary: his model assumes that the relation of the two sectors in real terms stays constant.

Still, most of the empirical literature on the determinants of health-care expenditure operates with deflated data, so in a second step Hartwig (2008) deflated nominal wages and HCE by the GDP deflator and re-estimated his model. He found that real wage growth in excess of productivity growth still contributes significantly to explaining real HCE growth, and the present paper confirms that this finding survives Extreme Bounds Analysis and outlier robust estimation. How do we have to interpret this finding?

Obviously, the difference between real wage growth and productivity growth drives the wage share. If real wages grow exactly at the same rate as productivity at the macroeconomic level, the wage share stays constant. If they grow more, the wage share rises and vice versa. Baumol's model of unbalanced growth implicitly assumes a constant wage share. In the progressive sector, nominal wages grow at the same rate as productivity and the price level doesn't change, so real wage growth and productivity growth are equal. In the stagnant sector productivity growth is zero, and prices rise to the same extent as wages. So here also real wage growth and productivity growth are equal. Note that this constellation is sufficient to cause the 'cost disease'. What the empirical results in Hartwig (2008) and the present paper show is that the 'cost disease' is so to speak aggravated if wage earners manage to increase the wage share. Conversely, it also means that if the wage share drops this puts a break on the growth of health-care expenditure.

Barros (1998), in his analysis of the determinants of health expenditure growth came across an empirical regularity he could not explain. He writes: "We detect, nevertheless, a clear slowdown in health care expenditures (as GDP share) growth in the decade 1980-1990 relative to the average evolution in the two previous decades. The determinants of this slowdown are not identified by our model and the issue clearly calls for further research" (Barros, 1999, p. 540). The years since 1980 have witnessed a marked decline in the wage

share in most OECD countries. This decline – in conjunction with Baumol’s model of unbalanced growth – offers a ready explanation for the deceleration in HCE growth that has puzzled Barros.

Acknowledgements

We wish to thank Terkel Christiansen, Mickael Bech, Jørgen Lauridsen and Pascal Nielsen for sharing their data with us. Arlette Frener’s support in processing our data is gratefully acknowledged.

References

- Barnett, V., Lewis, T., 1994. *Outliers in Statistical Data*. Wiley, New York.
- Barros, P.P., 1998. The black box of health care expenditure growth determinants, *Health Economics* 7 (6), 533–44.
- Bates, L.J., Santerre, R.E., 2012. Does the U.S. health care sector suffer from Baumol’s cost disease? Evidence from the 50 states. Mimeo.
- Baumol, W.J., 1967. Macroeconomics of unbalanced growth: the anatomy of urban crisis. *American Economic Review* 57 (3), 415–26.
- Christiansen, T., Bech, M., Lauridsen, J., Nielsen, P., 2006. Demographic changes and aggregate health-care expenditure in Europe. ENEPRI Research Report No. 32.
- Crivelli, L., Filippini, M., Mosca, I., 2006. Federalism and regional health care expenditures: an empirical analysis for the Swiss cantons. *Health Economics* 15 (5), 535–541.
- Croux, C., Dhaene, G., Hoorelbeke, D., 2008. Robust Standard Errors for Robust Estimators. Mimeo. [www.econ.kuleuven.be.christophe.croux/public](http://www.econ.kuleuven.be/christophe.croux/public).
- Di Matteo, L., 2005. The macro determinants of health expenditure in the United States and Canada: assessing the impact of income, age distribution and time. *Health Policy* 71 (1), 23–42.
- Docteur, E., Oxley, H., 2003. Health-care systems: lessons from the reform experience. OECD Health Working Papers No. 9.
- Dreher, A., Sturm, J.-E., Vreeland, J.R., 2009a. Development Aid and International Politics: does membership on the UN Security Council influence World Bank decisions? *Journal of Development Economics* 88 (1), 1–18.
- Dreher, A., Sturm, J.-E., Vreeland, J.R., 2009b. Global horse trading: IMF loans for votes in the United Nations Security Council. *European Economic Review* 53 (7), 742–757.
- Gassebner, M., Lamla, M.J., Sturm, J.-E., 2011. Determinants of pollution: what do we really know? *Oxford Economic Papers* 63 (3), 568–595.

- Gassebner, M., Lamla, M.J., Vreeland, J., 2012. Extreme bounds of democracy, *Journal of Conflict Resolution*, forthcoming.
- Gerdtham, U.-G., Jönsson, B., 2000. International comparisons of health expenditure: theory, data and econometric analysis. In: Culyer, A.J., Newhouse, J.P. (Eds.), *Handbook of Health Economics*, vol. 1A. Elsevier, Amsterdam.
- Gerdtham, U.-G., Jönsson, B., MacFarlan, M., Oxley, H., 1998. The determinants of health expenditure in the OECD countries: a pooled data analysis. In: Zweifel, P. (Ed): *Health, the Medical Profession, and Regulation*, Kluwer Academic Publishers, Dordrecht.
- Giannoni, M., Hitiris, T., 2002. The regional impact of health care expenditure: the case of Italy. *Applied Economics* 34 (14), 1829–1836.
- Hartwig, J., 2008. What drives health care expenditure? – Baumol’s model of ‘unbalanced growth’ revisited. *Journal of Health Economics* 27 (3), 603–623.
- Hitiris, T., 1997. Health care expenditure and integration in the countries of the European Union. *Applied Economics* 29 (1), 1–6.
- Karatzas, G., 2000. On the determination of the US aggregate health care expenditure. *Applied Economics* 32 (9), 1085–99.
- Koenig, L., Siegel, J.M., Dobson, A., Hearle, K., Ho, S., Rudowitz, R., 2003. Drivers of healthcare expenditures associated with physician services. *The American Journal of Managed Care* 9 (Special Issue 1), SP34–SP42.
- Leamer, E.E., 1985. Sensitivity analyses would help. *American Economic Review* (75) 3, 308–313.
- Levine, R., Renelt, D., 1992. A sensitivity analysis of cross-country growth regressions. *American Economic Review* 82 (4), 942–963.
- Martín, J.J.M., del Amo González, M.P.L., García, M.D.C, 2011. Review of the literature on the determinants of healthcare expenditure. *Applied Economics* 43 (1), 19–46.
- Maronna, R.A., Martin, D.R., Yohai, V.J., 2006. *Robust Statistics. Theory and Methods*, Wiley, Chichester.
- Mosca, I., 2007. Decentralization as a determinant of health care expenditure: empirical analysis for OECD countries. *Applied Economics Letters* 14 (7), 511–515.
- Moser, C., Sturm, J.-E., 2011. Explaining IMF lending decisions after the Cold War. *Review of International Organizations* 6 (3), 307–340.
- Newhouse, J.P., 1977. Medical-care expenditure: a cross-national survey. *Journal of Human Resources* 12 (1), 115–125.

- Okunade, A.A., Murthy, V.N.R., 2002. Technology as a ‘major driver’ of health care costs: a cointegration analysis of the Newhouse conjecture. *Journal of Health Economics* 21 (1), 147–159.
- Paris, V., Devaux, M., Wei, L., 2010. Health systems institutional characteristics. A survey of 29 OECD countries. *OECD Health Working Papers* No. 50.
- Potrafke, N., 2010. The growth of public health expenditures in OECD countries: Do government ideology and electoral motives matter? *Journal of Health Economics* 29 (6), 797–810.
- Roberts, J., 2000. Spurious regression problems in the determinants of health care expenditure: a comment on Hitiris (1997). *Applied Economics Letters* 7 (5), 276–283.
- Sala-i-Martin, X., 1997. I just ran two million regressions. *American Economic Review* 87 (2), 178–183.
- Sala-i-Martin, X., Doppelhofer, G., Miller, R.I., 2004. Determinants of long-term growth – a Bayesian Averaging of Classical Estimates (BACE) Approach. *American Economic Review* 94 (4), 813–835.
- Salibian-Barrera, M., Yohai, V., 2006. A fast algorithm for S-regression estimates. *Journal of Computational and Graphical Statistics* 15 (2), 414–427.
- Swartz, S., Welsch, R.E., 1986. Applications of Bounded-influence and diagnostic methods in energy modeling. In: Belsley, D.A., Kuh, E. (Eds.), *Model Reliability*. MIT Press, Cambridge MA.
- Sturm, J.-E., de Haan, J., 2001. How robust is Sala-i-Martin’s robustness analysis, University of Groningen. Mimeo.
- Sturm, J.-E., Berger, H., de Haan, J., 2005. Which variables explain decisions on IMF credit? An Extreme Bounds Analysis. *Economics & Politics* 17 (2), 177–213.
- Sturm, J.-E., de Haan, J., 2005. Determinants of long-term growth: new results applying robust estimation and extreme bounds analysis. *Empirical Economics* 30 (3), 597–617.
- Temple, J., 1998. Robustness tests of the augmented Solow model. *Journal of Applied Econometrics* 13 (4), 361–375.
- Temple, J., 2000. Growth regressions and what textbooks don’t tell you. *Bulletin of Economic Research* 52 (3), 181–205.
- Verardi, V., Croux, C., 2008. Robust Regression in Stata. KU Leuven, Working Paper KBI 0823.
- Yohai, V. 1987. High breakdown-point and high efficiency estimates for regression. *The Annals of Statistics* 15, 642–665.

Tables

Table 1: Macro-level explanatory variables for HCE suggested in the literature

Measure	Definition	Suggested by	Data Source	Remarks
<i>Dependent variable</i>				
HCEPC	Total expenditure on health, per capita, national currency units (NCU) at 2000 GDP prices		OECD Health Data 2011	
<i>M vector variables</i>				
GDPPC	Gross Domestic product, per capita, NCU at 2000 GDP prices	passim	OECD Health Data 2011	
DID	Change in income distribution	Karatzas (2000), Hartwig (2008)	ID = COMP/GDPN	
<i>Socio-demographic factors</i>				
ACCIDENT	Land traffic accidents, deaths per 100000 population	Koenig et al. (2003)	OECD Health Data 2011	
ALCC	Alcohol intake, litres per capita 15+	Gerdtham et al. (1998)	OECD Health Data 2011	
DP	Population density (Population per square kilometre)	Crivelli et al. (2006)	DP = POP/SURFACE	
FPR	Female participation ratio, % of active population	Gerdtham et al. (1998)	OECD Health Data 2010	
GSH	Public expenditure as percentage of GDP	Roberts (2000)	GSH = G/GDPN	
HCEI	Total expenditure on health, per capita in US\$ PPP at the start of the decade	Barros (1998)		Barro models average growth rates over three decades and includes for each decade the expenditure level of the first year of the decade. With annual growth rates, we cannot mimic this.
HCEI2	Total expenditure on health, per capita in US\$ PPP at the start of the decade squared	Barros (1998)	HCEI2 = HCEI*HCEI	Barro models average growth rates over three decades and includes for each decade the expenditure level of the first year of the decade. With annual growth rates, we cannot mimic this.
LE65F	Life expectancy at age 65 for females	Christiansen et al. (2006)	OECD Health Data 2011	Because the growth in LE65F and LE65M is highly correlated, we replaced the two variables by their average LE65.
LE65M	Life expectancy at age 65 for males	Christiansen et al. (2006)	OECD Health Data 2011	Because the growth in LE65F and LE65M is highly correlated, we replaced the two variables by their average LE65.
MORT	Mortality rate (Potential years of life lost per 100000 population 0-69)	Crivelli et al. (2006)	OECD Health Data 2011	
OBESI	Obese population, measured, % of population	Koenig et al. (2003)	OECD Health Data 2011	Less than 350 observations
PO	Dummy variable, one for countries with below-average per-capita GDP	Crivelli et al. (2006)	Built based on GDPPPP	
POP80	Share of population 80 years and over (%)	Mosca (2007)	OECD Health Data 2011	
POP75	Share of population 75 years and over (%)	Gerdtham et al. (1998)	EUROSTAT	Not included to avoid too many population variables
POP65	Share of population 65 years and over (%)	Christiansen et al. (2006)	OECD Health Data 2011	Not included to avoid too many population variables
POP6584	Share of population between 65 and 84 years (%)	Di Matteo (2005)	EUROSTAT	Not included to avoid too many population variables
POP6574	Share of population between 65 and 74 years (%)	Di Matteo (2005)	EUROSTAT	Replaced by POP6579 (using POP65 and POP80)
POP4564	Share of population between 45 and 64 years (%)	Di Matteo (2005)	EUROSTAT	Not included to avoid too many population variables
POP2544	Share of population between 25 and 44 years (%)	Di Matteo (2005)	EUROSTAT	Not included to avoid too many population variables
POP2044	Share of population between 18 and 44 years (%)	Di Matteo (2005)	EUROSTAT	Not included to avoid too many population variables
POP19	Share of population 18 years and under (%)	Mosca (2007)	EUROSTAT	Not included to avoid too many population variables

POP04	Share of population 4 years and under (%)	Gerdtham et al. (1998)	EUROSTAT
UNEMP	Unemployment rate (% ratio to labour force)	Christiansen et al. (2006)	OECD Health Data 2010
TOBC	Tobacco consumption, grams per capita 15+	Gerdtham et al. (1998)	OECD Health Data 2011

Institutional factors

BUDCEILA	Dummy variable, one for countries with direct budgetary controls in the ambulatory sector, zero otherwise	Gerdtham et al. (1998)		Construction not documented in the paper. Variable could not be built.
CAPITA	Dummy variable, one for countries with capitation as the dominant means of remuneration in primary care, zero otherwise	Gerdtham et al. (1998), Christiansen et al. (2006)	Gerdtham et al. (1998), Christiansen et al. (2006), Paris et al. (2010)	
CASEHO	Dummy variable, one for countries with fee-for-service or payment by bed days in in-patient care, zero otherwise	Christiansen et al. (2006)	Christiansen et al. (2006), Paris et al. (2010)	
CEILHO	Dummy variable, one for countries with global prospective budgets of hospital spending, i.e. countries having hospital spending either on budget or established levels which should not be exceeded, zero otherwise	Christiansen et al. (2006)	Christiansen et al. (2006)	Less than 350 observations
COPAY	The share of total expenditure on health paid for out of pocket (%)	Gerdtham et al. (1998)	OECD Health Data 2010	Dropped due to high (negative) correlation with PUES, which has more observations (952 vs. 620).
COPAYGP	Dummy variable for range of copayments for general practitioner visits, one for all, two for some groups of patients, three for no	Christiansen et al. (2006)	Christiansen et al. (2006)	Highly correlated with COPAYGP. Merged to COPAYDUM: One for countries with some copayment for either GP or HO, zero otherwise
COPAYHO	Dummy variable for range of copayments for hospital stays, one for all, two for some groups of patients, three for no	Christiansen et al. (2006)	Christiansen et al. (2006)	Highly correlated with COPAYHO. Merged to COPAYDUM: One for countries with some copayment for either GP or HO, zero otherwise
COVERO	Insurance coverage of the population (%)	Gerdtham et al. (1998)	OECD Health Data 2011	
FFSA	Dummy variable, one for countries with fee-for-services as the dominant means of remuneration in primary care, zero otherwise	Gerdtham et al. (1998)	Gerdtham et al. (1998), Paris et al. (2010)	
FREEHO	Dummy variable, one for countries with free choice of hospitals, two for countries with limited or no free choice	Christiansen et al. (2006)	Christiansen et al. (2006), Paris et al. (2010)	Highly correlated with FREEGP and FREESP. Merged to FREE: One for countries with free choice of either HO or GP or SP
FREEGP	Dummy variable, one for countries with free choice of general practitioner, two for countries with limited or no free choice	Christiansen et al. (2006)	Christiansen et al. (2006), Paris et al. (2010)	Highly correlated with FREEHO and FREESP. Merged to FREE: One for countries with free choice of either HO or GP or SP
FREESP	Dummy variable, one for countries with free choice of specialist, two for countries with limited or no free	Christiansen et al. (2006)	Christiansen et al. (2006), Paris et al. (2010)	Highly correlated with FREEHO and FREEGP. Merged to FREE: One for countries with free choice of either HO or GP or SP
GATEKEEP	Dummy variable, one for countries with physicians as (compulsory) gatekeepers, zero otherwise	Gerdtham et al. (1998), Christiansen et al. (2006)	Gerdtham et al. (1998), Christiansen et al. (2006), Paris et al. (2010)	
GLOBALHO	Dummy variable, one for countries which remunerate their hospitals mainly by global budget, zero otherwise	Christiansen et al. (2006)	Christiansen et al. (2006), Paris et al. (2010)	
HCSYS	Dummy variable, one for countries with public integrated systems, two for countries with public contract (reimbursement) system, three for countries with a mixed model and four for countries in transition	Gerdtham et al. (1998), Christiansen et al. (2006)	Gerdtham et al. (1998), Christiansen et al. (2006)	Split into 4 variables, two of which (former dummies 3 and 4) are dropped because they contain too many zeros. Former dummy 1: HCSYSPI. Former dummy 2: HCSYSPC
INS	Per capita real insurance premiums	Karatzas (2000)	INS = (SSS+PINS)/POP/GDPDEF*100	

MIXEDGP	Dummy variable, one for countries with a mix of capitation and fee-for-services as the dominant means of remuneration in primary care, zero otherwise	Christiansen et al. (2006)	Christiansen et al. (2006), Paris et al. (2010)	
NHSD	Dummy variable, one for countries with a decentralised National Health System, zero otherwise	Mosca (2007)	Mosca (2007)	Less than 350 observations
OVERBILL	Dummy variable, one for countries with high potential for free setting of medical care prices and overbilling, 0.5 for countries with some potential and zero otherwise	Gerdtham et al. (1998)	Gerdtham et al. (1998)	Dropped because it is constant over time meaning it drops out of the fixed effects regressions.
PUBREIMB	Dummy variable, one for countries with public reimbursement as the dominant means of remuneration in inpatient care, zero otherwise	Gerdtham et al. (1998)	Gerdtham et al. (1998)	Dropped because of high correlation with HCSYSPC
PUHES	Public health expenditure as a share of total health expenditure	Christiansen et al. (2006)	OECD Health Data 2010	
REIMBMOD	Dummy variable, one for countries with direct payment by patient before reimbursement by insurer (reimbursement model), zero otherwise	Gerdtham et al. (1998)		Construction not documented in the paper. Variable could not be built.
SHIC	Dummy variable, one for countries with a centralised Social Health Insurance System, zero otherwise	Mosca (2007)	Mosca (2007)	Less than 350 observations
SHID	Dummy variable, one for countries with a decentralised Social Health Insurance System, zero otherwise	Mosca (2007)	Mosca (2007)	Less than 350 observations
TA	Per capita real expenditure on health administration	Karatzas (2000)	OECD Health Data 2010	
TEXMC	The share of inpatient expenditure in total health expenditure (%)	Gerdtham et al. (1998)	OECD Health Data 2010	
WS	Dummy variable, one for countries with wage and salary as the dominant means of remuneration in primary care, zero otherwise	Gerdtham et al. (1998), Christiansen et al. (2006)	Gerdtham et al. (1998), Christiansen et al. (2006), Paris et al. (2010)	

Prices

HPI	Price index for total expenditure on health (2000=100)	Karatzas (2000)	OECD Health Data 2010	
-----	--	-----------------	-----------------------	--

Technological and capacity factors

BEDSFPSH	Ratio of beds in for-profit hospitals to acute care beds	Koenig et al. (2003), Gerdtham et al. (1998)	BEDSFPSH = BEDSFP/BEDS	Less than 350 observations
BEDSH	Acute care beds per general hospital	Giannoni/Hitiris (2002)	BEDSH = BEDS/HOSP	
BEDSI	Acute care beds per 1000 inhabitants	Christiansen et al. (2006)	BEDSI = BEDS/POP	
CERV	Cervical cancer: percentage of females aged 50-69 screened, survey or programme data	Koenig et al. (2003)	OECD Health Data 2010	Less than 350 observations
DLGERD	Gross expenditure on R&D, compound annual growth rate (constant prices)	Okunade/Murthy (2002)	OECD Main Science and Technology Indicators (online access July, 11, 2011)	
DOCTCA	The stock of practicing physicians per 1000 population	Gerdtham et al. (1998)	DOCTCA = PHYS/POP	
DOCTCA*FFSA	The product of DOCTCA and FFSA	Gerdtham et al. (1998)	= DOCTCA*FFSA	Dropped because it is not feasible to include interacted variables in EBA.

HRD	Total expenditure on health R&D, national currency units (NCU) at 2000 GDP prices	Okunade/Murthy (2002)	OECD Health Data 2010	Less than 350 observations
MAMMO	Mammography: percentage of females aged 50-69 screened, survey or programme data	Koenig et al. (2003)	OECD Health Data 2011	Less than 350 observations
NURCA	Number of actively employed nurses per 1000 population	Karatzas (2000)	NURCA = NUR/POP	
OUTPSH	Share of total surgeries performed on an outpatient basis	Koenig et al. (2003)	OUTPSH = SPOUT/TSP	Less than 350 observations
PERSH	Number of healthcare and non-healthcare personnel per hospital	Giannoni/Hitiris (2002)	PERSH = HOEMP/HOSP	Less than 350 observations
PHYSH	Physicians per 100 hospital beds	Christiansen et al. (2006)	PHYSH = PHYSHO/BEDS*100	Less than 350 observations
RAT	The ratio of specialists to general practitioners	Karatzas (2000)	RAT = (PHYS – GP)/GP	Less than 350 observations
REND	Patients undergoing renal dialysis, rate per 100000 population	Gerdtham et al. (1998)	OECD Health Data 2011	
TOMSCA	Tomographic scanners per million population	Christiansen et al. (2006), Koenig et al. (2003)	TOMSCA = CT+MRI+PET	Less than 350 observations

Variables needed to build explanatory variables

BEDS	Number of acute care beds		OECD Health Data 2011
BEDSFP	Number of beds in for-profit hospitals		OECD Health Data 2011
COMP	Compensation of employees, million of NCU		OECD Health Data 2010
COMPR	Compensation of employees, million of NCU at 2000 GDP prices		OECD Health Data 2010
CT	Number of Computed Tomography Scanners		OECD Health Data 2010
EMP	Total employment, thousands of persons		OECD Health Data 2010
G	General government's total outlays, mln. of NCU		OECD Health Data 2010
GDPDEF	GDP deflator, price index (2000=100)		OECD Health Data 2011
GDPN	Nominal GDP, million of NCU		OECD Health Data 2010
GDPPPP	GDP per capita in US\$ PPP		OECD Health Data 2011
GP	Number of general practitioners		OECD Health Data 2010
HOEMP	Total hospital employment, number of persons		OECD Health Data 2011
HOSP	Number of general hospitals		OECD Health Data 2011
MRI	Number of Magnetic Resonance Imaging units		OECD Health Data 2010
NUR	Number of practicing nurses		OECD Health Data 2011
PET	Number of PET Scanners		OECD Health Data 2010
PHYS	Number of practising physicians		OECD Health Data 2011
PHYSHO	Number of physicians employed in hospitals		OECD Health Data 2011
PINS	Private insurance, million of NCU		OECD Health Data 2010
POP	Total population, thousands of persons		OECD Health Data 2011
SPOUT	Outpatient surgical procedures ("day cases")		OECD Health Data 2010
SSS	Social security schemes, million of NCU		OECD Health Data 2010
SURFACE	Surface of country, square kilometres		CIA World Factbook

TSP	Total surgical procedures, number	OECD Health Data 2010
WSEMP	Number of wage and salaried employees, thousands of persons	OECD Health Data 2010

Notes: Variables in light grey are included in a modified form. Those in dark grey have not be included in the analysis.

Table 2: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
dlhcepc	1028	4.01	5.06	-28.33	30.60
dlgdppc	1218	2.23	3.02	-14.88	12.26
did	1055	0.02	1.35	-6.60	10.68
dlaccident	1101	-2.90	12.64	-99.49	90.15
dlalcc	1176	-0.10	5.65	-38.30	30.01
ddp	1272	0.81	2.79	-67.83	48.07
dfpr	901	0.29	0.79	-4.00	11.00
dgsh1	733	0.10	2.03	-11.90	15.51
dlle65	1192	0.78	1.29	-4.85	7.87
dmort	1114	-0.14	0.26	-2.62	1.77
dobese	67	0.30	0.63	-1.50	1.80
po	924	0.55	0.50	0.00	1.00
dpop80	1360	0.06	0.08	-0.90	0.40
dpop6579	1360	0.07	0.16	-0.50	1.40
dpop04	966	-0.07	0.16	-0.76	0.54
dunemp	981	0.08	1.01	-4.20	5.80
dltoabc	800	-1.57	6.02	-54.81	53.54
capita	1066	0.17	0.38	0.00	1.00
caseho	688	0.14	0.35	0.00	1.00
ceelho	289	0.72	0.45	0.00	1.00
copaydum	1428	0.17	0.38	0.00	1.50
dcovero	1053	0.37	2.23	-5.80	36.40
ffsa	1042	0.39	0.49	0.00	1.00
free	1428	0.39	0.49	0.00	1.00
gatekeep	1083	0.54	0.50	0.00	1.00
globalho	689	0.43	0.49	0.00	1.00
hcsyspi	1428	0.25	0.44	0.00	1.00
hcsyspc	1428	0.27	0.44	0.00	1.00
dlins1	503	4.42	9.45	-79.57	66.09
mixedgp	1075	0.24	0.43	0.00	1.00
nhsd	209	0.26	0.44	0.00	1.00
dpuhes	952	0.07	2.25	-16.10	15.10
shic	220	0.20	0.40	0.00	1.00
shid	220	0.25	0.43	0.00	1.00
dlta	544	4.53	21.19	-150.41	248.49
dtexmc	730	-0.25	2.64	-22.00	17.40
ws	1095	0.15	0.35	0.00	1.00
dlhpi	416	6.98	9.29	-32.85	73.90

dbedsfpsh	231	0.23	1.04	-6.30	3.95
dlbedsh	397	0.05	4.83	-30.11	27.93
dlbedsi	639	-1.69	3.51	-42.08	11.08
dcerv	150	0.46	2.37	-7.10	16.00
dlgerd	668	5.32	7.12	-24.37	64.34
dldoctca	638	2.02	3.64	-28.79	17.53
dlhrd	268	5.91	32.42	-302.04	210.41
dmammo	158	1.12	4.16	-17.40	25.70
dlnurca	370	1.34	3.61	-21.66	18.71
doutpsh	201	1.16	1.84	-5.12	8.46
dlpersh	172	2.59	4.40	-17.35	32.41
dlphysh	202	3.05	4.46	-19.35	25.38
drat	238	0.55	14.36	-128.13	132.71
dlrend	602	6.21	10.24	-51.85	88.85
dltomsca	127	6.55	7.47	-3.10	37.10

Note: Variables in light grey have less than 350 observations.

Table 3: Baseline regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Ordinary Least Squares				MM-estimations			
	Only constant	Fixed country	Fixed year	Fixed both	Only constant	Fixed country	Fixed year	Fixed both
dlgdppc	0.538*** (11.07)	0.464*** (8.934)	0.712*** (12.34)	0.619*** (9.492)	0.397*** (6.384)	0.389*** (4.710)	0.540*** (5.619)	0.527*** (6.196)
did	1.508*** (14.88)	1.493*** (14.73)	1.275*** (11.70)	1.276*** (11.66)	1.548*** (10.24)	1.556*** (7.284)	1.412*** (8.493)	1.541*** (7.816)
Observations	924	924	924	924	924	924	924	924
R-squared	0.245	0.284	0.327	0.358				
p-val. vs.none		0.03	0.00	0.00				

Notes: t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: EBA results

	OLS regressions					MM regressions				
	Avg.	Min.	Max.	%sign.	CDF(0)	Avg.	Min.	Max.	%sign.	CDF(0)
dlgdppc	0.72	-0.53	2.47	99.55	99.87	0.41	-0.29	0.80	80.93	97.15
did	1.37	0.63	2.51	99.87	100.00	1.36	0.08	2.36	98.20	100.00
ffsa	2.69	0.82	11.85	44.44	100.00	2.09	-2.08	8.03	27.78	97.22
dlbedsi	0.05	-0.01	0.16	6.80	99.03	0.08	-0.04	0.40	33.33	91.67
dunemp	0.38	-0.03	0.82	44.44	98.77	0.14	-0.18	0.45	2.78	80.56
dlrend	-0.04	-0.12	0.01	45.68	96.30	0.03	-0.03	0.05	50.00	91.67
dfpr	-0.52	-2.15	0.08	60.19	96.12	-0.02	-0.92	0.39	11.11	58.33
dlta	0.02	-0.07	0.06	60.49	95.06	0.06	0.02	0.09	97.22	100.00
dltohc	0.04	-0.02	0.07	25.93	95.06	0.00	-0.04	0.08	13.89	66.67
dgsh1	0.10	-0.05	0.29	5.83	94.17	0.26	0.13	0.38	88.89	100.00
dlle65	0.23	-0.28	0.57	42.68	93.90	0.17	-0.24	0.73	2.78	91.67
dtexmc	-0.25	-0.36	0.14	86.42	93.83	-0.15	-0.30	0.08	72.22	94.44
ws	-1.08	-2.97	3.13	23.75	93.75	-0.24	-2.68	1.42	13.89	80.56
po	0.44	-0.23	2.55	2.47	92.59	0.72	-0.35	2.28	19.44	88.89
dlnurca	-0.07	-0.22	0.05	0.00	92.59	-0.05	-0.31	0.10	22.22	86.11
dpop80	-1.92	-5.32	3.61	1.23	87.65	0.40	-3.60	4.72	0.00	61.11
dldoctca	-0.05	-0.78	0.11	11.65	85.44	0.04	-0.13	0.16	19.44	86.11
dlins1	0.01	-0.01	0.04	0.00	85.37	0.00	-0.06	0.03	5.56	58.33
gatekeep	0.49	-1.96	3.43	0.00	85.19	0.16	-1.23	1.96	2.78	66.67
dpop6579	0.46	-4.91	2.64	2.47	82.72	0.07	-2.00	1.65	2.78	66.67
dpuhes	-0.12	-0.48	0.43	74.07	80.25	0.14	-0.14	0.74	30.56	86.11
mixedgp	-1.32	-7.09	1.74	16.05	79.01	-0.46	-2.20	4.30	16.67	91.67
hcsyspc	0.23	-1.17	2.29	2.47	79.01	-0.09	-1.60	0.96	2.78	63.89
globalho	0.37	-2.08	2.38	0.00	75.31	-0.29	-2.38	1.15	2.78	83.33
capita	-0.31	-5.80	21.83	23.16	75.08	-0.24	-1.65	4.07	5.56	88.89
dpop04	0.74	-3.14	9.17	6.17	74.07	-0.60	-3.39	5.10	8.33	80.56
hcsyspi	-0.38	-2.33	1.61	3.70	69.14	-0.27	-2.33	0.35	11.11	75.00
dcovero	0.08	-0.16	1.44	10.68	66.99	0.12	-0.61	0.27	80.56	88.89
free	0.04	-2.87	2.05	1.23	61.73	0.06	-2.52	0.84	5.56	69.44
dmort	0.00	0.00	0.00	3.70	60.49	0.00	-0.01	0.00	8.33	72.22
caseho	0.15	-2.95	4.12	0.26	60.32	-0.15	-3.23	3.00	11.11	80.56
dlalcc	0.00	-0.12	0.28	33.01	60.19	0.00	-0.18	0.10	11.11	63.89
dlaccident	0.00	-0.04	0.10	2.91	57.28	-0.04	-0.08	0.02	91.67	97.22
dlgerd	0.00	-0.08	0.06	0.00	57.28	0.08	-0.01	0.12	77.78	97.22
dlhpi	-0.03	-0.43	0.18	4.82	56.63	0.04	-0.20	0.27	33.33	72.22
ddp	0.10	-0.55	2.59	1.94	52.43	0.02	-0.25	0.76	0.00	69.44
copaydum	-0.02	-2.44	1.35	0.00	51.46	-0.10	-2.40	0.61	2.78	58.33
dlbedsh	0.02	-0.07	0.24	8.74	50.49	0.02	-0.07	0.17	52.78	50.00

Notes: Each cell contains information on the estimated β -coefficients. The columns “Avg.”, “Min.” and “Max” report the average, minimum and maximum β -coefficients, respectively. The column “%sign.” reports the percentage of cases in which the estimated coefficient estimate is significant at the 10 percent level. The column “CDF(0)” reports the percentage of the cumulative distribution function lying on one side of zero. The OLS-results are based upon a total of 66’711 regressions. The MM-results are based upon 666 regressions. Values above 95% are highlighted in grey.

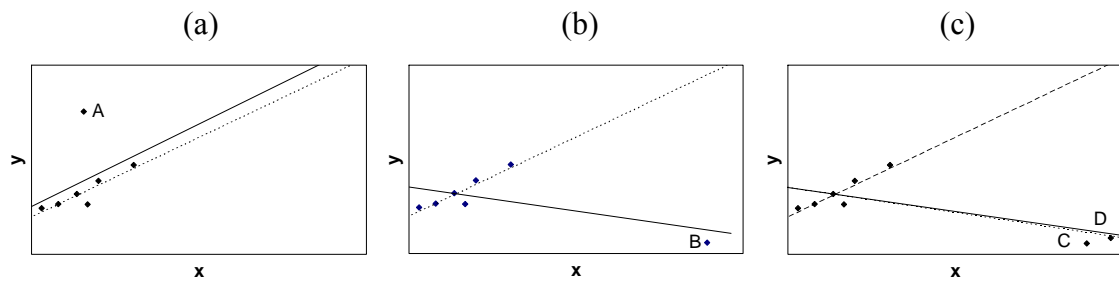
Table 5: Extended regression results

	(1)	(2)	(3)	(4)	(5)	(6)
	Ordinary Least Squares			MM-estimations		
	Baseline	Robust OLS	Robust MM	Baseline	Robust OLS	Robust MM
dlgdppc	0.619*** (9.492)	1.065*** (5.432)	0.685*** (4.796)	0.527*** (6.196)	0.854*** (19.55)	0.625*** (4.326)
did	1.276*** (11.66)	2.058*** (6.156)	1.713*** (7.140)	1.541*** (7.816)	1.382*** (25.68)	1.696*** (9.585)
ffsa		0.620 (0.497)	4.212 (1.350)		0.831*** (5.387)	2.776** (2.522)
dlbedsi		0.142** (2.100)			0.278*** (32.49)	
dunemp		1.279*** (3.202)			1.152*** (15.45)	
dlrend		0.0536** (2.049)			0.0421*** (19.25)	
dfpr		0.391 (0.711)			-0.736*** (-12.40)	
dlta		0.0412 (1.659)	0.0450*** (6.656)		0.0240*** (4.162)	0.0543*** (12.39)
dltoabc		0.00362 (0.0926)			0.0292** (2.248)	
dgsh1			-0.0578 (-0.521)			0.202*** (3.858)
dlaccident			0.00979 (0.677)			-0.0715*** (-7.164)
dlgerd			-0.0164 (-0.358)			-0.0522* (-1.945)
Observations	924	136	259	924	136	259
R-squared	0.358	0.614	0.521			

Notes: t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figures

Figure 1: Outlying observations and bad leverage points



Notes: The solid lines represent the OLS estimates including the unusual observation(s). The dotted lines represent the OLS estimates without the unusual observations A, B, or C. The dashed line represents the OLS estimate without observations C and D.