HESG Glasgow Wednesday 30th June 2004

Health Expenditure and Aggregate Income: A Time Series Investigation for Ireland 1960-2001

Patrick Moore
Department of Economics,
University College Cork, Ireland

DRAFT Please do not quote

Abstract

Despite the extensive documentation of a strong positive relationship between health expenditure and aggregate income, the precise nature of this relationship has been subject to controversy. Whether health care expenditures are stationary or nonstationary can have critical implications for researchers and policy makers desiring to model, explain and influence this important area of health economics. If expenditure on health care is nonstationary the implication is that government intervention to manage health care expenditures could possibly be prevented by the random nature of the series. From an econometric view, nonstationary HCE will make regressions on possible determinants difficult if not impossible to reliably implement. Dickey-Fuller (1979), Phillips-Perron (1988) and Kwitkowski, et al., (1992) methodologies are used to show that health care expenditures per capita and GDP per capita are non-stationary. The Perron (1997) methodology of endogenously determined breaks reports that the null of a unit root can be rejected in 99% of the cases for the health expenditure time series. The results obtained in this study so far suggest that it is unlikely for a long-term relationship to exist between health care expenditure and aggregate income for Ireland.

1. Introduction

This study conducts an empirical investigation into the relationship between health care expenditures (HCE) per capita and income, as measured by gross domestic product (GDP) per capita, for Ireland between 1960-2001. A strong positive correlation between HCE and income, GDP has been the consistent finding of research examining the determinants of aggregate health care expenditure (Hansen and King 1996). During the last decade economic researchers, such as Roberts (2000), have become increasingly aware of the difficulties that arise when unit roots are present in time series. To ignore these problems and proceed to estimate a regression containing nonstationary variables would omit important information about the underlying statistical and economic processes generating the data, if not leading to an entirely spurious regression. Therefore this research examines the possibility of nonstationarity in relation to a standard time series model of the macroeconomic demand for health care in Ireland in order to present a robust series for further analysis. While the literature in this area is extensive it is not conclusive due to the numerous methodological problems with the data, concrete conclusions have been difficult to form. These methodological problems are discussed in more detail in section 2.

Features of the health system are mentioned in the media virtually every week in Ireland. Expenditure on the health system in particular has attracted a growing amount of attention in recent years, this is perhaps due to the escalation in health spending, 241 percent for the period

1990 to 2001 (Wiley, 2001). At odds with this large increase is the decrease in public confidence in the care provided by the system, (Watson and Williams, 2001). Attempts have been made to gain a greater understanding of what exactly the government's money is funding and how said resources can be used more efficiently and effectively, for example the "Value for Money Audit" conducted by Deloitte and Touche (2001). The demand for health care is derived from the demand for health and is driven by a number of factors including demographics, public expectations, new treatments and technology. (Grossman, 1972) This demand places enormous pressure on the exchequer resources given that the Irish healthcare system is funded approximately 80 percent by the taxpayer (Wiley, 2001).

There is widespread agreement that income and health expenditure are closely linked, but the reliability of the precise mechanisms through which this relationship arises is far from clear (Newhouse, 1977; Culyer, 1988; Hansen and King, 1996). Currently the majority of research in this area is based on cross-country analysis typically involving 10 to 20 countries, despite the OECD's (2003) warning that due to the diverse nature of health expenditure measurement from country to country, vast inconsistencies exist in the data. Hansen and King (1998:380) have highlighted that "a deeper understanding of what drives HCE is more likely to be gained through careful country-by-country analysis".

One of the assumptions underlying traditional regression methods such as Ordinary Least Squares (OLS) is that of stationarity. A time series is said to be stationary when the data generating process is the same over time. If HCE or GDP is nonstationary then a regression of HCE on GDP is "spurious" and the results from estimating such a model are invalid. This would mean that the estimated income elasticity of HCE and the finding of a significant and positive relationship between the two variables may be meaningless. Accurate knowledge of the time series properties of HCE and GDP is a prerequisite to empirical testing. This study may verify that OLS regression analysis can be reliably used to examine the relationship between HCE and GDP. A greater understanding and a reliable model of the relationship between these two variables is required for a number of reasons.

Firstly rising health expenditure has been a cause of concern in most of the developed countries for several decades (OECD, 2001). This has increased pressure on taxes and social contributions due to the large proportion of this expenditure being publicly funded. Although Ireland currently has a young population relative to other developed countries there are fears that as ageing of the population accelerates HCE will rise (Jacobzone and Oxley, 2001). The large young population is also a concern in itself as those under the age of 15 along with individuals over the age of 65 are thought to be the most frequent users of health care (Hitiris, 1997).

Secondly Ireland has witnessed a decrease in the percentage of real GDP spent on health care from 7 percent in 1990 to 6.2 percent in 2001[1], this is in contrast to other developed countries such as USA or UK, where HCE as a percentage of GDP has steadily increased. The Irish figure for expenditure on health in 2001 of 6.2 percent of its real GDP compares poorly with an OECD average of 8.4 percent (OECD, 2003) and a United States figure of 13.9 percent. Given the large differences in funding, studies conducted across countries may not be relevant to the particular Irish situation.

Furthermore health care is not just shaped by income but a number of other determinants, such as, the structure of the health system (i.e. public/private mix) or the age structure of population,

[1] Real HCE data adjusted by price deflator based on nominal OECD series. Real GDP data adjusted by GDP deflator based on nominal OECD series.

each of which has a different level of importance for individual countries. Therefore a focused look at specific countries is required if any attempt is made to measure the correct share of national income to be allocated to health care. For any attempt at such measurement the relationship between the variables must be theoretically and empirically sound.

The nature of the income elasticity of health care expenditure, presents another area with conflicting opinions. Wagner's Law[2] is often mentioned in discussions on public sector growth, with Gemmell (1993) concluding that there is strong evidence in favour of Wagner's predictions. Cullis and Jones (1998) summarise the demand function arguments in support of Wagner's Law, that expenditures like health would exhibit elasticities greater than unity and as a result be classed as luxuries. However the size of the income elasticity may be incorrectly estimated due to the potential problem of non-stationarity in the data. Getzen (2000) suggests that the elasticity of health care expenditure at a national level is greater than 1 and at a micro level near 0 or negative making health care both luxury and necessity.

2. Literature

In recent decades health economists have devoted a lot of attention to the time series pattern of national HCE, and their findings have had an important impact on policy. The main hypothesis in this body of literature is that GDP per capita is the single most important determinant of the level of health care expenditure (Kanavos and Mossialos, 1996). The analysis of health care expenditure has been based on standard demand theory, typically focusing on the income elasticity of health care expenditure (HCE) estimated in functions linking per capita health care expenditure to per capita GDP. In particular, much of the early research on this subject found a strong positive correlation between HCE and GDP in developed countries, with income elasticity greater than one (Newhouse, 1977, 1987; Cullis and West, 1979). This suggests that health care is a 'luxury good" and the share of HCE in GDP will increase with per capita income. This is in stark contrast to what has occurred in Ireland particularly over the last decade, while the real amount of HCE has risen the proportion of GDP spent on health has fallen.

The seminal paper by Newhouse (1977) regressed HCE on GDP to identify the factors determining the quantity of health care services. His results showed an income elasticity greater than one and factors other than income, for example the price paid by the consumers, and the method of reimbursement were of marginal significance. The idea of health care as a luxury may seem counter intuitive, as care is often seen as meeting a 'need' rather than a 'want'. Despite this Newhouse's (1977) results concerning the income elasticity and high explanatory power of the relationship have been supported by a volume of literature using cross sectional data (Culyer, 1988; Gerdtham and Jonsson 1985; Gerdtham, 1987).

It is well known that the size of an estimating elasticity may be incorrectly estimated due to the potential problem of nonstaionarity in the data. Roberts (2000) highlights the problems of modeling the determinants of HCE, using time series, and warns against "drawing firm conclusions from aggregate level models in an area where theory provides little guidance." (Roberts, 2000:279). In particular she comments on Hitiris (1997) in which panel data for ten EU countries is used to analyse the determinants of aggregate health care expenditure. Roberts (2000) highlights the failure of Hitiris (1997) to deal with the time series properties of the data in particular the spurious regression problem which has grave consequences for the interpretation of his results. Furthermore she employs more recently developed techniques for unit roots in

[2] Long run tendancy for public expenditure to grow relative to national income. (Cullis and Jones, 1998)

panel data devised by Im *et al.* (1996) to show the possible errors involved in Hitris's (1997) exposition.

Five recent papers have focused on the stationarity and cointegration of HCE and GDP, dealing with the issues of time series analysis previously over looked. These papers show the recent trend of favouring panel based unit root test over univariate tests, the former allows for more powerful tests due to increased sample size and inclusion of heterogeneous cross-country information not available in single country tests. Despite being based on almost the same data[3] each paper has reached different conclusions. Firstly Hansen and King (1996) examines a time series for 20 OECD countries and carries out unit root tests and cointegration tests on each one. They do not reject the unit root hypothesis for either HCE or GDP, neither is the hypothesis of no cointegration between HCE, GDP and a range of other variables rejected. As a result they propose that panel data estimations of the relationship between HCE and GDP may be spurious. Akin to the previous paper Blomqvist and Carter (1997) use similar tests but in contrast they reject the cointegration hypothesis.

Roberts (2000) also preformed country by country and panel unit root tests finding clear evidence of unit roots *id est* nonstationarity, in the series. But she was unable to provide firm evidence regarding cointegration.

In contrast to the other recent papers McCoskey and Selden (1998) reject the null hypothesis of unit roots for both HCE and GDP. In testing national income and health expenditure McCoskey and Selden (1998) combine data on 20 OECD countries and use the panel data tests presented by Im *et al*. (1996) to test for unit roots. These same tests were used by Blomqvist and Carter (1997) and Roberts (2000), but they failed to reject the unit root hypothesis. Hansen and King (1998) who also rejected the unit root hypothesis, using exactly the same data, have highlighted a number of issues with the testing procedures employed by McCoskey and Selden (1998). They conclude that the different results for studies using Im *et al.* (1996) testing procedures and those using country specific ADF tests, lie in the restricitiveness of the null hypothesis of the former test. Given the varying results outlined in this paragraph, the low statistical power of the country-by-country tests is unlikely to be principally responsible for the difference between McCoskey and Selden (1998) and Hansen and King (1998).

The fifth recent paper, Gerdtham and Löthgren (2000), applies country specific and panel augmented Dickey-Fuller (ADF) tests. Also to avoid a possible multicollinearity problem as discussed by Hansen and King (1998), reverse hypothesis testing procedures developed by Kwiatkowski *et al.* (1992) are employed. The results obtained indicate that both HEC and GDP are nonstationary. Furthermore strong evidence is found supporting a long-run equilibrium cointegrating relationship between HEC and GDP.

[3] Exception being Roberts (2000), who uses a subsample of 10 EU countries.

Table 1: Methodology of Studies Addressing Times Series Problems in HCE

Author	Tests for Stationarity	Tests for Cointegration	Income Elasticity	Stationarity	Cointegration
Hansen & King (1996)	ADF	Engle-Granger	Not calculated	2/3 found to be non-stationary No country had both series N-S	No cointegration
Blomqvist & Carter (1997)	Autocorrelations & first difference	Philips-Perron	Luxury	Both non-stationary	Cointegration
McCoskey & Selden (1998)	ADF, IPS	Not tested	Not calculated	Both Stationary	Not tested
Roberts (2000)	ADF, Philips-Perron	Engle-Granger Pesaran, Johansen	Not a luxury	Both non-stationary	No clear evidence of cointegration
Gerdtham & Löthgren (2000)	ADF, IPS, KPSS.	Shin test, KPSS	Not calculated	Both N-S	Cointegrated

From the literature in Table 1 it appears that the unit root question is relatively settled, but the question of whether a cointegrating relationship exists is what is dividing researchers. If the underlying series in an econometric specification contain unit roots, then traditional econometric analysis is invalid. However, if the series can be shown to have a cointegrating relationship between a nonstationary regressand and the nonstationary regressors, then the relationship can be modelled using an error correction mechanism allowing OLS to be safely employed.

This study brings together all the key concepts, theories and models in order to test empirically the stationarity of time series data on HCE and GDP in Ireland. To begin with Augmented Dickey-Fuller (ADF), Phillips-Perron (Philips and Perron 1988) and KPSS (Kwiatkowski *et al*., 1992) tests are applied to determine whether the series are stationary. The KPSS test is used similar to Gerdtham and Löthgren (2000), one reason for this is to avoid a possible multicollinearity problem as discussed by Hansen and King (1998). Given that time series involved are nonstationary based on the theoretical and empirical literature the time series are tested for cointegration. This is achieved using tests for cointegration similar to those used by Roberts (2000).

This study continues the same line of research as the five time series papers outlined but differs from previous studies in one principle way, it only uses data on one single country, Ireland, providing country specific analysis. McCoskey and Selden (1998:369) have highlighted that 'No single test is likely to be definitive in this rapidly-evolving area of econometric research" however the results of this study will help to mitigate concerns that time series analyses of national health care expenditure in Ireland are misspecified. Having established a reliable data set and model, this study will provide the answer to the question of how health expenditure in Ireland is affected by a percentage increase in GDP per capita.

3. DATA AND METHODOLOGY

3.1 Data Description

The Organisation for Economic Cooperation and Development (OECD) Health Data, 2003 is the source of all the variables. The OECD Health data give annual time series, spanning the years 1960 to 2001 for Ireland on total health care expenditure (HCE), Gross Domestic Product (GDP), population under 15 years of age (POP15) and the population 65 years or older, (POP65). From these sources a complete data set for Ireland was constructed covering the period 1960 to 2001. The figures in the GDP and Health Care Expenditure time series have been converted into Euro by applying the irrevocable conversion rate between the Punt and the EUR.[4] This conversion has been done for the complete time series, thus, the evolution over time of the historical national series is preserved.

To make useful comparisons over time, it is necessary to deflate nominal health expenditure by suitable price indices and to divide by the population to derive real spending per capita. The result is a ratio which is not affected by inflation, Real Health Care Expenditure per capita. Roberts (2000) criticises the use of GDP price indices for not being a good indicator of health care price inflation. Therefore the index used in this study for HCE is that provided by the OECD Health Data (2003), with a base year of 1995. The GDP deflator from the OECD Health Data (2003), has been used to construct the real GDP time series. Table 2 presents the basic definitions of the variables used in this study.

Table 2: Variable Definitions and Summary Statistics

Variable	Definition	Mean	Std D	Min	Max
HCE	Natural log of health care expenditure per capita in millions of Euro, expressed in 1995 prices.	6.38126	0.72554	4.96396	7.27046
GDP	Natural log of Gross Domestic Product per capita in millions of Euro, expressed in 1995 prices.	9.08339	0.45600	8.33830	10.04401
POP15	Natural log of proportion of the population under 15	-1.25858	0.12110	-1.54506	-1.16315
POP65	Natural log of proportion of the population over 65	-2.19847	0.02227	-2.24432	-2.16893

3.2 Model for Estimation

Inline with the literature reviewed the following model of health care expenditure will be estimated in logarithmic form for Ireland from 1960 to 2001:

$$HCE = \beta_0 + \beta_1 Y + \beta_2 POP65 + \beta_3 POP15 + u$$

HCE = per capita health expenditure
β_0 = intercept term
Y = GDP per capita
POP65 = proportion of population over 65
POP15 = proportion of population under 15
u = error term

[4] Established on January 1st 1999 for the EU-11 countries.

POP65 and POP15 are intended to represent the different health care needs of these high user groups. POP65 has showed a steady increase over the period, while POP15 peaked and now appears to be declining. Evidence that the elderly consume more health care per capita than people of a working age is a standard result showing as a positive coefficient on POP65 (Kleinman, 1974; Hitiris & Posnett, 1992; Hansen & King, 1996). POP15 however is more controversial, both Kleinman (1974) and Hansen and King (1996) reported a negative coefficient on POP15 and explained their results by highlighting that although young people may be high users of health care, they tend to use services with low per unit cost, for example vaccinations. The results obtained using this model may be sensitive to the inclusion of additional regressors, as outlined in section 1 there are numerous drivers for the demand for health care.

3.3 Testing for Stationarity

Firstly to establish if the time series are stationary three tests are undertaken; The Augmented Dickey Fuller (ADF) test, the Phillips-Perron test and the Kwitkowski, Philips, Shin and Schmidt (KPSS) test.[5] The KPSS test is used in response to criticism from McCoskery and Selden (1998) about the low power of the Augmented Dickey Fuller test and Hansen and King's (1998) comments on the possibility of multicollinearity in the ADF regressions if time trends are included. The KPSS test, (Kwitkowski, *et al.*, 1992) has a null of trend stationarity which may be a means of eliminating the possible multicollinearity problem, providing a more powerful unit root test. These tests for the presence of unit roots are outlined in more detail in following subsections.

Given that the time series involved are non-stationary for the model to constitute a valid long run relationship it must be a cointegrating regression. Various combinations of three methodologies have been used to test for cointegration in the literature. Firstly using the Engle-Granger (1987) two step procedure with Augmented Dickey Fuller (ADF). Secondly the Johansen (1991) procedure and thirdly a test proposed by Pesaran *et al.* (1996) to test for long run relationships when the orders of the underlying regressors are unknown.[6] The first two tests for cointegration have been used extensively in this particular area and other studies using time series analysis, (Engle and Granger, 1991), (see Table 1). The Pesaran *et al.* (1996) test is relatively new, it has been already employed by Roberts (2000) and provides an alternative to the more established methods of Engle-Granger and Johansen.

While the Engle and Granger methodology has come under criticism (Enders 1995) it would appear to suit this study given the model specified in section 3.2 and the sample size. However one problem that is relevant is the test's reliance on a two step estimator. The first step being the generation of the error series, the second step uses these errors to estimate a regression.[7] Therefore any error introduced in step one may be carried into step two.

A priori one would expect a cointegrating relationship between HCE and GDP. Both POP15, and POP65 are bounded variables and therefore theoretically cannot have a long term effect on HCE.

[5] The null Hypothesis is that the series are stationary, I(0), around the level or trend

[6] Pesaran *et al*. (1996) test involves a bounds testing procedure which can be implemented whether the underlying regressors are I(0), I(1) or mutually cointegrated.

[7] In this case the regression estimated is: $\Delta\hat{e}_t = a\hat{e}_{t-1} + \varepsilon_t$

3.3.1 Augmented Dickey Fuller Test

The Dickey Fuller (DF) and Augmented Dickey Fuller (ADF) tests are the most widely used tests for the presence of unit roots in Econometrics. The ADF is used when y_t follows an AR(p) process, then the error term will be autocorrelated to compensate for the misspecification of y_t.[8] The null hypothesis is that of a stocastic trend, nonstationarity, against the alternative of a deterministic trend, stationarity. The tests involve running the following equation and can be altered accordingly for τ_{μ}, constant and no trend and τ, no constant and no trend:

$$\Delta y_t = a + b_t t + \gamma y_{t-1} + \sum_{j=1}^{p-1} \phi_j \Delta y_{t-j} + \varepsilon_t$$

Where $\Delta y_t \equiv y_t - y_{t-1}$ and t indicates time and H_0: $\gamma = 0$

The t-statistic for γ does not have the familiar mean-zero t-distribution under the null hypothesis but tables of critical vales have been constructed, (Dickey and Fuller, 1987). Given that the mean of y is not zero and has a non zero drift[9] for HCE and GDP the ADF test is used including both a constant and a trend. However to ensure that the assumptions made are correct the sequential testing strategy suggested in Perron (1988) is implemented.

The ADF is the standard test used for unit roots, despite much grumbling about its low power in the literature, it has been used in almost all of the major papers in the area, (see Table 2). Much of the question about this test relates to its use in cross country analysis, it is employed here given that this study is country specific and the test is widely acceptable. The period examined in this study covers the years 1960 to 2001 and is based on the latest OECD Health Data (OECD, 2003). Thus an increased sample size is used relative to previous studies. This increased sample size may also improve the power of unit root testing procedures.

3.3.2 Phillips-Perron Test

The Phillips-Perron (1988) test while similar to the ADF test makes fewer stringent assumptions about the distribution of the errors. There is no requirement for the disturbance term to be serially uncorrelated or homogeneous. Enders (1995) briefly outlines the test as follows, considering the two regressions:

$$y_t = a_0^* + a_1^* y_{t-1} + \mu_t$$

$$y_t = \tilde{a}_0 + \tilde{a}_1 y_{t-1} + \tilde{a}_2(t - T/2) + \mu_t$$

T = number of disturbance terms and $E(\mu_t) = 0$

Phillips and Perron test uses this regression to derive a test statistic that can be used to test the hypotheses about the coefficients a_i^* and $\tilde{a}_I$ under the null hypothesis that the data are generated by $y_t = y_{t-1} + \mu_t$. The critical values for the Phillips-Perron test are the same as those used for Dickey-Fuller tests.

[8] Autocorrelated errors will invalidate the use of the DF distributions, which are based on the assumption that the error term u_t is white noise.

[9] From exploratory data analysis.

3.3.3 KPSS Test

Kwiatkowski *et al.* (1992) presented a test with stationarity as the null hypothesis. It has subsequently been used by Gerdtham and Löthgren (2000) and offers a recent alternative to the more traditional ADF. This test is based on the decomposition of the series into the sum of a deterministic trend, a random walk and a stationary error component as:

$$y_t = \delta_t + r_t + \varepsilon_t$$

where r_t is a random walk: $r_t = r_{t-1} + u_t$
and u_t are individually and identically distributed $(0,\sigma_u^2)$.
The initial value r_0 is treated as fixed and serves the role of an intercept in the model.
The KPSS test is a one sided LM-test for the stationarity hypothesis Ho: $\sigma_u^2 = 0$, that the unit root process is fixed at the initial value r_0 and hence that y_t is trend stationary. The LM statistic for each series is:

$$\eta = \frac{\frac{1}{T}\sum_{t=1}^{T} S_{it}^2}{s_i^2(l)}$$

Where $S_{it} = \sum^{T}_{j=1}\varepsilon_{ij}$ and s_i^2 is the consistent Newey and West (1987) estimate of the long-run variance of the disturbance terms ε_{ij} defined as:

$$\sigma_i^2 = \lim_{T\to\infty} T^{-1}E(S_{iT}^2)$$

Two KPSS tests are run in this study, the first with a null hypothesis that the series are stationary, I(0), around a level (η_μ) and the second with a null hypothesis that the series are stationary, I(0), around a deterministic linear trend (η_τ). Give the small sample size involved in this study the KPSS test may not give accurate results, (Kwiatkowski *et al*, 1992)

3.4 Testing for Cointegration

In their article Engle-Granger (1987) set out a two step procedure to test for the presence of a cointegrating relationship between two time series, y_t and x_t. Firstly the variables are tested for their order of integration this is usually done using an ADF test. If the two variables are integrated of order one then the long-run equilibrium relationship can be estimated in the form: $y_t = \hat{\alpha}_0 + \hat{\alpha}_1 x_t + e_t$. If the series of estimated residuals, $\{\hat{e}_t\}$ from this regression are found to be stationary then the series are cointegrated of order (1, 1). In order to establish stationarity the residuals are tested using the ADF as outlined previously in section 3.3.1.

3.5 Perron Structural Break Methodology

Perron (1997) developed a methodology of endogenously determined breaks, building on previous work in this area. The models are estimated over all possible break dates in the data set, without making any prior assumptions about the break date. The date chosen is the one which maximises the probability of rejection of the unit root hypothesis. Perron describes his method as taking the extreme view as regards the choice of break date. "The choice of break points is effectively made to be perfectly correlated with the data. This case is instructive to study because if one can still reject the unit root hypothesis under such a scenario it must be the case that it would be rejected under a less stringent assumption." (Perron, 1997:356).

Two models are estimated, the first allowing only a change in the intercept.
Model 1:

$$y_t = \mu + \theta DU_t + \beta T + \delta DUM_t + \alpha y_{t-1} + \sum c_i \Delta y_{t-i} + \varepsilon_t$$

H_0: $\alpha = 1$ (a unit root)
$DU_t = 1$ for $t > T_b$
$DUM_t = 1$ for $t = T_b + 1$
$T_b =$ the endogenously determined time of the break.

The methodology chooses the break point at the minimum value of the t statistic, t_α. The truncation lag parameter k is also treated as unknown, with the truncation chosen using a general to specific procedure rather than relying on methods based on the information criterions like the Bayesian used to determine the lags in the ADF test in this study. This is done to avoid size distortions and/or power loss.
Model 2:

$$y_t = \mu + \theta DU_t + \beta T + \gamma DT_t + \delta DUM_t + \alpha y_{t-1} + \sum c_i \Delta y_{t-i} + \varepsilon_t$$

$DT_t = 1$ for $t > T_b$

4. Discussion of Results

4.1 Exploratory Data Analysis

A priori one would expect health expenditure and income to be trended, graphs show an obvious upward trend in both time series. Also visually it does not appear that the mean, variance and autocovariances of the individual series are time invariant, while this points towards nonstationarity it is by no means conclusive. Nonstationarity is only due in part to the deterministic trend, it is likely that the variables may possess stochastic trends as well, (Roberts, 2000). This will be confirmed, when the regression is run, (See Table 3), by a high R^2 and an extremely low Durbin Watson statistic and supported visually by correlograms. It can be seen that while the real amount of health care expenditure is increasing, reasonably steadily, the percentage of Real GDP attributed to health has been steadily declining over the last 20 years of the period. This fact has also been highlighted by Wiley (2001).

POP65 has steady increased over the period, while POP15 peaked and now appears to be declining. The increase in the elderly population should have a positive influence on HCE as the population in POP65 are considered to consume more health care per capita than people of a working age (Kleinman, 1974; Hitiris & Posnett, 1992; Hansen & King, 1996). The effect of the decline in POP15 since the mid eighties is more controversial. Following on from Kleinman (1974) and Hansen and King (1996) POP15 may have a negative affect on HCE.

When an Ordinary Least Squares (OLS) regression is run on the data the results observed are presented in Table 3. The high R^2 obtained may be due to the sustained upward trends, which are evident in both HCE per capita and GDP per capita, rather than any real relationship between the two variables. Also an R^2 greater than the Durbin Watson statistic is a helpful rule of thumb to suspect that the estimated regression is spurious (Granger and Newbold, 1974).

Table 3: OLS Regression Results, HCE Dependent Variable

	Coefficient	Standard Error	t-statistic	p-value
Constant	-25.78997	3.150297	-8.18652	0.0000000
GDP	2.67239	0.174326	15.32982	0.0000000
POP15	4.93299	0.737411	6.68961	0.0000000
POP65	-6.14602	1.558606	-4.11651	0.0001996

R^2	0.963143
Durbin Watson Statistic	0.567638

In order to further the preliminary investigation into the stationarity of both time series, the corresponding correleograms have been plotted. In both of these correlograms the autocorrelation function (ACF), starts at a very high value and gradually decreases, this type of pattern is another general indication that a series is nonstationary. Upon visually analysing the correlograms of the first difference of both series it would appear that the series are both integrated of order one, I(1).[10]

Therefore *a priori*, inline with the literature, it is expected that HCE and GDP are non-stationary, integrated of order one. However graphical plots are only a guide, statistical test must be used to confirm if the series are stationarity. The results of these tests are reported in section 4.2.

4.2 Unit Root Tests

Table 4 presents three versions of the ADF test, with all the variables in natural logs: $\tau_{\tau,}$ constant and trend; $\tau_{\mu,}$ constant and no trend; τ, no constant and no trend. Because the data are trended, the purpose of the unit root test is to determine whether the series is consistent with an I(1) process with a stochastic trend, or if it is consistent with an I(0) process, i.e. it is stationary, with a deterministic trend. The null hypothesis is that the series are integrated of order one, I(1). The number of lags has been selected using the Bayesian Information Criterion (BIC). The results show in Table 4 that that for most of the variables the null hypothesis of at least one unit root cannot be rejected, this means that the series have a unit root, i.e. they are non-stationary.

[10] The correlograms for all series, and their first difference are plotted in the Appendix, A3 along with graphs of the partial accorrelation functions (PACF).

Table 4: Augmented Dickey-Fuller Tests

	HCE	GDP	
τ_τ constant & trend	-1.1023	0.8713	
Lags	0	0	
τ_μ constant, no trend	-1.9027	2.3854	
Lags	0	1	
τ no constant no trend	3.4392^+	3.3886^+	
Lags	0	1	
Critical Values	**1%**	**5%**	**10%**
τ_τ	-4.15	-3.50	-3.18
τ_μ	-3.58	-2.93	-2.60
τ	-2.62	-1.95	-1.61

Notes:

1. $^+$ rejects the null hypothesis at the 1% level.
2. All series in natural logs.

The results in Table 4 show that the null hypothesis of a unit root cannot be rejected in the case of all series. These results are backed up by the Phillips-Perron results shown in Table 5.

Table 5: Phillips-Perron Tests

t-Statistic			Significance Level	Critical Value
HCE	-1.75907		1%	-3.593
GDP	1.86468		5%	-2.932
POP15	5.25419	*	10%	-2.604
POP65	-1.39908			
DHCE	-5.70414	*		
DGDP	-4.05223	*		
DPOP15	-1.51671			
DPOP65	-2.88141	*		

Notes:

1. * significant at the 5% level.
2. All tests calculated with an intercept and 4 lags in the error process.

However in the Augmented Dickey Fuller and Phillips-Perron tests, the unit root is the null hypothesis to be tested, and the way in which classical hypothesis testing is carried out ensures that the null hypothesis is accepted unless there is strong evidence against it. This point has been highlighted extensively in the literature, McCoskery and Selden (1998) directly comment on the low power of Dickey Fuller tests. Kwiatkowski, *et al.* (1992) present an alternative unit root test in which the null hypothesis of stationarity is tested against the alternative of a unit root, the KPSS test.

Table 6: KPSS Tests

	Lag	HCE	GDP		Significance Level	Critical Value
η_μ	0	3.48728*	3.97773*	η_μ	1%	0.739
	1	1.81295*	2.08548*		2.5%	0.574
	2	1.24886*	1.44515*		5%	0.463
	3	0.96682*	1.12450*		10%	0.347
	4	0.79861*	0.93267*			
η_τ	0	0.91773*	0.58991*	η_τ	1%	0.216
	1	0.47325*	0.31822*		2.5%	0.176
	2	0.32445*	0.22894*		5%	0.146
	3	0.25078*	0.18574*		10%	0.119
	4	0.20762*	0.16107*			

Notes:
1. * shows those values that are significant at the 5% level.
2. ** shows those values that are significant at the 10% level.
3. The null hypothesis is I(0) around level (η_μ) and I(0) around trend (η_τ).
4. Critical values taken from Kwiatkowski *et al.*, (1992:Table 1).

Table 6 shows two versions of the KPSS test. The null hypothesis is that the series are stationary, I(0), around a level (η_μ), or I(0) around a deterministic linear trend (η_τ). The hypothesis of level stationarity is rejected, this result is not very surprising in light of the obvious deterministic trends present in the series. Therefore a test with a null hypothesis of stationarity around a deterministic linear trend is performed. The null hypothesis of I(0) is rejected in HCE, GDP and POP15 at the 5% level, making these results consistent with the presence of a unit root. The null hypothesis of stationarity about a level is not rejected in all cases for Pop65 variable. These are similar findings to Hansen and King (1996) and Gerdtham and Löthgren (2000), and suggest that POP65 may be stationary, while there is evidence of the presence of unit roots in HCE, GDP and POP15 series.

On first differencing HCE and testing for the presence of a unit root using ADF none was found. This result was confirmed by the KPSS test, which showed an acceptance of a null of stationarity, suggesting that HCE is integrated of order one, I(1). The null of a unit root is also rejected for the first difference of GDP, while the KPSS test rejects the null of stationarity. This inconsistency between the two tests may be due to the small sample size used. (Kwiatkowski *et al.,* 1992). The Phillips-Perron test in Table 5, for both the GDP series and its first difference, suggest that GDP is integrated of order one I(1) and also confirms the result for HCE.

While it has been established that HCE and GDP are both integrated of order one, it is less clear whether POP15 and POP65 are stationary. However as the population variables are bounded they cannot theoretically have a long run effect on HCE and as a result are omitted from the cointegration analysis.

4.3 Cointegration Results

The Engle-Granger test was conducted to test for a co-integrating relationship between health care expenditure and income per capita. The first step in the Engle-Granger methodology was to establish the order of integration of each series. This has already been done using the Augmented Dickey Fuller (ADF) and Phillips-Perron tests. Both series were found to be integrated of order one, I(1). These deviations were not found to be stationary therefore there is not a co-integrating relationship between the health expenditure per capita and income per capita. In order to ensure that autocorrelation did not interfere with this result the residuals were also tested using the ADF.

4.4 Structural Break

However an examination of the plot of residuals reveals evidence that there is a possible structural change between 1974 and 1990. It can be no coincidence that these dates coincide with revisions in the reporting of health care expenditure in Ireland. None of the literature reviewed experienced similar data problems. However there is a large volume of general unit root literature to draw on. Following Perron (1997), it is well known that a structural break can be mistaken for nonstationarity in time series. Perron (1997) highlights that failure to account for an existing structural break will lead to substantial loss of power to reject the null hypothesis of a unit root in the ADF and Phillips-Perron tests. Having used the Perron methodology of endogenously determined breaks, it is clear from the results that the structural break in the HCE is affecting the unit root testing procedures as the null of a unit root can be reject in 99 percent of the cases. This result is consistent with the HCE series being stationary around a breaking mean and trend function. A longer time period would perhaps yield more information.

5 Conclusions

5.1 Main Conclusions

This study has tested time series for health expenditure and income for the presence of unit roots, in order to avoid the problem of spurious regression when examining the relationship between these two variables. Unit roots are useful for indicating whether the finite sample data used exhibits stationary or non-stationary attributes as oppose to making definitive statements like 'real HCE is non-stationary'. Whether health care expenditures are stationary or nonstationary can have critical implications for researchers and policy makers desiring to model, explain and influence this important area of health economics. A greater understanding of the relationship between HCE and GDP would aid policy formation, in particular it could guide the creation of the next health strategy for Ireland. It would also help to answer the question of whether health care is a luxury or necessity. From an econometric view, nonstationary HCE will make regressions on possible determinants difficult if not impossible to reliably implement. From an international perspective this study has highlighted some of the potential problems of examining individual countries rather than examining cross-country panel data.

The principal finding of this study so far, is that the variables in a standard model of aggregate health care expenditure for Ireland were collectively non-stationary in levels for the finite period studied. This implies that one of the key assumptions underlining OLS is violated, making the usual statistical tests of significance fraught with uncertainty. However a possible structural break in the HCE series places a question over the result for health expenditure. The Perron methodology of endogenously determined breaks demonstrated the effect of the structural break on the unit root testing procedure. The possibility of a structural break has not been considered in most of the current literature. The results obtained in this study suggest that it is unlikely for a long-term, cointegrating, relationship between health care expenditure and aggregate income to exist in Ireland.

5.2 Further Research

Given the findings in this study there is considerable scope for future theoretical and empirical work on a model of health care expenditure for Ireland. Further research must investigate the robustness of these results and attempt to alleviate the problems encountered.

The empirical studies previously conducted in this area, outlined in this study in section 2, have been slow to investigate new variables which influence health expenditure. Now that Ireland is in a slow down period after huge growth in GDP over the last decade, one possible variable which may influence health expenditure in the future is the size of the budget deficit. If a government has a high budget deficit it is likely to put a constraint on health expenditure, with the opposite also applying. Another variable which may be important to include is proximity to death, recently Seshamani and Gray (2004) have produced evidence to suggest that age and proximity to death to have significant effects on health expenditure.

Although the literature reviewed in section 2 emphasises the relationship between GDP and health expenditure, Gross National Product (GNP) may be more suitable for the Irish situation. Given the difference between Ireland's GDP and GNP the latter variable may be a more accurate representation of Irish aggregate income.

Re-examining the possibility of structural breaks in the time series could further develop the question of unit roots and employing the methods used in other areas of economics to solve this estimation problem.

In conclusion while this study has not solved the methodological problems of a time series examination of the relationship between income and health expenditure it has laid the foundations for further research by highlighting the potential problems associated with such an analysis. These problems are not unique to health expenditure but span all areas where time series are used. Until they are solved the debate will continue on the income elasticity of health expenditure and on how exactly aggregate income influences health expenditure in the Irish economy.

ACKNOWLEDGEMENTS
This paper benefited from suggestions and comments made by the Health Economics Group at University College Cork.

REFERENCES

Blomqvist, A. and Carter, R. (1997). Is Health Care Really a Luxury? *Journal of Health Economics,* Vol. 16, p207-229.

Cullis, J., and West, P. (1979). *The Economics of Health; An Introduction.* Robertson, Oxford.

Cullis, J and Jones, P. (1998). *Public Finance and Public Choice*. Oxford University Press, New York.

Culyer, A.J., (1988). Health Expenditures in Canada: Myth and Reality; Past and Future. *Canadian Tax Paper No. 82*, Canadian Tax Foundation, Toronto.

Culyer, A.J., (1990). Cost Containment in Europe. In OECD (1990) *Health Care Systems in Transition*, p29-40. OECD Paris.

Deloitte and Touche (2001). *Value for Money Audit*, Department of Health and Children, Dublin.

Dicky D., and Fuller, W. (1979). Distribution of the estimators for Autoregressive Time Series with a Unit Root. *Journal of American Statistical Association*, Vol. 74, p427-437.

Enders, W., (1995) *Applied Economic Time Series*. John Wiley and Sons, New York.

Engel, R. and Granger, C. (1987). Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica,* Vol 55, p251-276.

Engel, R. and Granger, C. (1991). *Long-Run Economic Relationships: Readings in Cointegration*. Oxford University Press, Oxford and New York.

Gerdtham, U., Anderson, F, Sogaard, J and Jönsson, B. (1988). Econometric Analysis of Health Care Expenditures: A Cross-section Study of OECD Countries. Centre for Medical Technology Asseeement, University of Linkoping, Sweden.

Gerdtham, U. and Jönsson, B. (1991). Price and Quality in International Comparisons of Health Care Expenditure. *Applied Economics,* Vol. 23, p1519-1528.

Gerdtham, U. and Jönsson, B. (2001). International Comparisons of Health Expenditure. *Handbook of Health Economics,* Culyer, A. and Newhouse, J., (eds.) Vol. 1(A), Elsvier.

Gerdtham, U. and Löthgren, M. (2000). On Stationarity and the Cointegration of International Health Expenditure and GDP. *Journal of Health Economics*, Vol. 19(4), p461-475.

Getzen, T. (2000) Health Care is an Individual Necessity Luxury and a National Luxury; Applying Multi-level Decision Models to the Analysis of Health Care Expenditure. *Journal of Health Economics*, Vol. 19(2), p259-270.

Granger, C. W., and Newbold P., (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, Vol. 2, p111-120.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, 80(2), p223-255.

Hansen, P., and King A., (1996). The Determinants of Health Care Expenditure: a Cointegration Approach. *Journal of Health Economics,* Vol. 15, p127-137.

Hansen, P., and King A., (1998). Health Care Expenditure and GDP: Panel Data Unit Root Test Results – Comment. *Journal of Health Economics,* Vol. 17, p377-381.

Hitiris, T., (1997). Health Care Expenditure and Integration in the Countries of the European Union. *Applied Economics*, Vol. 29, p1-6.

Hitiris, T. and Posnett, J. (1992). The Determinants and Effects of Health Expenditure in Developed Countries. *Journal of Health Economics*, Vol. 11, p173-181.

Im, K. S., Pesaran, M. H., Shin, Y., (1996). Testing for Roots in Heterogeneous Panels. *DAE Working Paper No. 9526*, University of Cambridge.

Jacobzone, S., Oxley, H., (2001). Healthcare Expenditure: A Future in Question. OECD Observer, 9th December 2001, p15, OECD, Paris.

Johansen, S., (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregression Models. *Econometrica,* Vol. 59,p 1551-1580.

Kanavos, P. and Mossialos, E., (1996). The Methodology of International Comparisons of Health Care Expenditures: Any Lessons for Health Policy? *Discussion Paper No. 3*, LSE Health, The London School of Economics and Political Science.

Kleiman, E., (1974). The Determinants of National Outlay on Health. In Perlman, M., ed. The Economics of Health and Medical Care. Macmillan, London.

Kwitkowski, Philips, Shin, Schmidt, (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. *Journal of Econometrics*, Vol. 54, p159-178.

MacFarlan ,M. and Oxley, H., (1994). *Health Care Reform Controlling Expenditure and Increasing Efficiency.* OECD, Paris.

McCoskery, S. and Selden, T., (1998), Health Care Expenditure and GDP: Panel Data Unit Root Test Results. *Journal of Health Economics*, Vol. 17, p369-376.

Newhouse, J.,P. and Phelps, (1974). Price and Income Elasticities for Medical Care Services. In M. Perlman (ed.) *The Economics of Health and Medical Care,* MacMillan, London.

Newhouse, J., (1977). Medical Care Expenditure; A Cross-National Survey. *Journal of Human Resources,* Vol. 12, p115-125.

Newhouse, J., (1987). Cross-National Differences in Health Spending; What do they Mean? *Journal of Health Economics*, Vol. 6, p159-162.

OECD (2001). *Health at a Glance*, OECD, Paris.

OECD (2003). *Health Data 2002 CD-ROM,* OECD, Paris.

Perron, P. (1988). Trends and Random Walks in Macroeconomic Time Series, *Journal of Economic Dynamics and Control*, Vol. 12, p297-332.

Perron, P. (1997). Further Evidence on Breaking Trend functions in Macroeconomic Variables. *Journal of Econometrics*, Vol. 80, p355-385.

Phillips, P. and Perron, P. (1988). Testing for a Unit Root in Times Series Regression. *Biometrica*, Vol. 75, p335-346.

Roberts, J., (2000). Spurious Regression Problems in the Determinants of Health Care Expenditure: a Comment on Hitris. *Applied Economics Letters,* Vol. 7(5), p279-283.

Seshamani, M. and Gray, A. (2004) Ageing and Health-care Expenditure: the Red Herring Argument Revisited. *Health Economics,* Vol 13, p303-314.

Watson, D. and Williams J. (2001). *Perceptions of Quality of Care in The Public and Private Sectors in Ireland,* ESRI, Dublin.

Wiley, M., (1998). Health Expenditure Trends in Ireland: Past, Present and Future. In *The Irish Health System in the 21st Century,* Oak Tree Press, Dublin.

Wiley, M., (2001). Reform and Renewal of the Irish Health Care System: Policy and Practice, *Budget Perspectives: Proceedings of a Conference held on 9th October 2001*, ESRI, Dublin.