WORKING   PAPERS
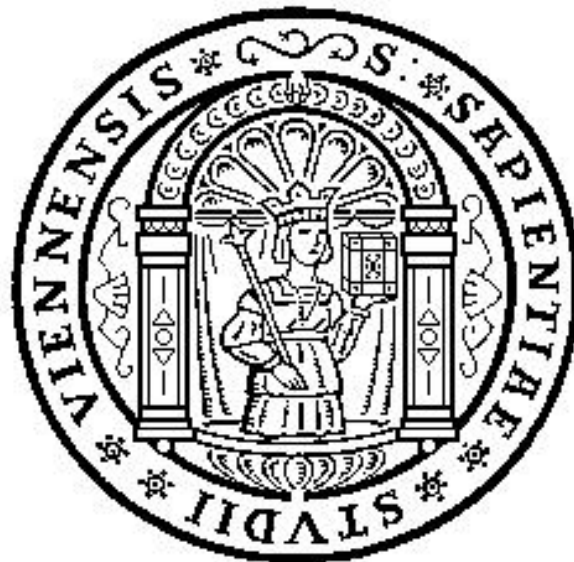
Adelina Gschwandtner
Michael A. Hauser

Modeling Profit Series: Nonstationarity and Long Memory

July 2005

Working Paper No: 0508

DEPARTMENT OF ECONOMICS

UNIVERSITY   OF   VIENNA

Modeling Profit Series: Nonstationarity and Long Memory

Adelina Gschwandtner[a] and Michael A. Hauser[b]

[a]University of Vienna, [b]Vienna University of Economics and Business Administration

This version July 2005.

Abstract:

The dynamic structure of profit rates for 156 US manufacturing companies is analyzed by means of fractional integration techniques as an alternative to the commonly used ARIMA models with respect to the "persistence of profits". The results show - despite the short lengths of the series - that 35.5% of the series have long range dependence and 54% are nonstationary. This is a confirmation of the of the strong challenge to the competitive environment hypothesis obtained by previous studies.

Address of authors:

[a]University of Vienna, Department of Economics, BWZ, Bruennerstr. 72, A-1210 Vienna, Austria. Tel: +43 1 4277 37480, Fax: +43 1 4277 37498, E-Mail: Adelina.Gschwandtner@univie.ac.at and [b]Vienna University of Economics and Business Administration, Department of Statistics and Mathematics, Augasse 2-6, A-1090 Vienna, Austria. Tel: +43 1 31336 4759, Fax: +43 1 31336 738, E-Mail: michael.hauser@wu-wien.ac.at

1 Introduction

Starting with the seminal contributions of Mueller(1977, 1986), much research has been undertaken in order to shed light on the empirical relevance of the competitive environment hypothesis. The main idea of this literature, which later became known as "the persistence of profit literature", is that in most markets firms may earn supernormal profits because markets are affected by external shocks, because of first mover advantages of dominating firms or because of barriers of existing entry/exit. Even though competition should reduce any over − or under − normal rents, the literature shows that it takes time for the competitive forces to work and therefore profits above the norm may exist over longer time periods. One of the main aims of this literature is to study how fast and to what extent abnormal profits are eroded. Some examples of this branch of research are given by Geroski and Jacquemin(1988), Kambhampati(1995), Goddard and Wilson(1999), McGahan and Porter(1999), Cable et al.(2001) and Glen et al.(2001), to mention just a few. Mueller(1990) analyzes profit persistence in 7 developed countries (Canada, West Germany, France, Japan, UK, US and Sweden) and concludes that in all these economies profits persist. One of the more recent contributions by Maruyama and Odagiri(2002) follows 376 Japanese firms which were previously analyzed for the period 1964-82 and finds that by adding 15 more years of data, the conclusion stays the same: profits still persist. Gschwandtner(2005) investigating 85 surviving US companies from 1950-1999 finds that competitive forces were not able to erode profits even after a period of 50 years.

In all these studies, autoregressive processes of low order (mostly of order one) have been the widely-used representations of the dynamics of profits and have become the modeling workhorse for evaluating the adequacy of the competitive environment hypothesis. Some few attempts have been made to extend this methodology. Cable et al.(2001) propose a structural time series analysis in order to describe the dynamics of company profits. Crespo Cuaresma and Gschwandtner(2005a) use a nonlinear approach for the same purpose and Crespo Cuaresma and Gschwandtner(2005b) consider time-varying persistence of profits.

In the present paper we propose to analyze the dynamics of profits for 156 US manufacturing companies which managed to survive from 1950-1999, by means of fractional integration. There are several macroeconomic and financial studies revealing the long memory properties of economic time

series. Gil-Alana and Robinson(1997) analyze an extended set of Nelson and Plosser US macro data by means of fractionally-based tests proposed by Robinson(1994). Other examples are Diebold and Rudebusch(1989), Baillie and Bollerslev(1994) and Gil-Alana and Robinson(2000).

The present paper proposes fractional integration as an alternative and plausible way of modeling profit series as opposed to the conventional autoregressive and unit root approach. Both a variety of different degrees of slow converence to a market profit rate, and a diversity of different degrees of nonstationarity will become possible. The results will go in line with the main results of the profit persistence literature.

The paper contributes by being the first study to apply autoregressive fractionally integrated moving average processes of profit series. To our knowledge the series used is the longest time series available on profits. The present methodology is a new way to explain the dynamics of profits which allows for a smooth transition from stationarity to nonstationarity. Unit root behavior is well known to be a problem in the profit persistence literature. At the same time the methodology extends the simple autoregressive approach by allowing a variety of degrees of convergence to a specific profit market rate. Therefore a new indicator of profit persistence is constructed. We analyse the small sample properties of our sample and show this methodology can successfully be applied in small samples.

The structure of the paper is as follows. Section two deals with methodological issues: the concept of "persistence of profits", a comparison of traditional autoregressive moving average models and fractional integration and a summary of the available estimation techniques for nonstationary fractionally integrated series. Section three gives the design of the empirical study, describes the data and presents the results. A short section concludes.

2   Methodological issues

2.1   "Persistence of profits"

In case of deviations the notion of an equilibrium generally requires forces that establish a new equilibrium again. In a stochastic setting the deviation from the equilibrium values x_t^*, $y_t = x_t - x_t^*$,

are then viewed as weakly stationary with a mean of zero. So that without further stochastic shocks, a convergence to the mean will be observed.

The "persistence of profits" literature questions this postulate to hold in general by reporting evidence that the means of individual profit rates are different from zero for a large percentage of firms, or that many profit series exhibit nonstationarity. Previous studies might have suffered from the low power of unit root tests against stationary autoregressive alternatives in short series, and so might have chosen the null hypothesis of nonstationarity too often. We will reconsider the classification in stationary and nonstationary series by using a model which allows for a smooth transition from stationarity to nonstationarity, fractional integration. Long memory turns out as a natural extension of conventional autoregressive model by aggregation. Granger(1980), for example, shows that the sum of an infinite number of independent AR(1) processes with beta distributed coefficients on the interval (0,1) exhibits fractional integration. Similar properties do hold also for models with dependent processes. The data investigated below belong to the largest 500 US firms in 1950, which survied until 1999. Each of the firms may be assumed to produce a bunch of different products each of them obeying to its own profit process. So the aggregate profit rate series may well exhibt long range dependence.

2.2   Fractional integration

Autoregressive moving average, ARMA, models provide a rich class for modeling short run effects. Their long-run properties, however, vary radically with the modulus of the characteristic roots of the AR polynomial. For example in case of an AR(1) in the form $(1 - \alpha L)y_t = \epsilon_t$, $|\alpha| < 1$, ϵ_t (Gaussian) white noise, y_t is covariance stationary, while for $\alpha = 1$, it is nonstationary though not explosive, and for $|\alpha| > 1$ it is explosive. The null asymptotic distribution for $H_0 : \alpha = 1$ against $H_A : \alpha \neq 1$ is nonstandard and Pitman efficiency is lacking, reflecting this nonsmooth behavior. An alternative to model long run behavior are models for fractional integration.

We start with a stationary process y_t which is fractionally integrated, FI(d), $-1/2 < d < 1/2$.

$$(1 - L)^d (y_t - \mu) = \epsilon_t$$

The difference operator may be defined (more generally for $d \in \mathbb{R}$) via the hypergeometric function

4

$$F(-d, 1, 1; L)$$

$$(1 - L)^d = \Pi_d(L) = \sum_{j=0}^{\infty} \pi_{d,j} L^j = \sum_{j=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(d-j+1)\,\Gamma(j+1)} (-1)^j L^j$$

The Wold representation of y_t has the form

$$y_t = \mu + \Psi(L)\,\epsilon_t$$

with $\Psi(L) = \Pi_{(-d)}(L)$.

For $-1/2 < d < 0$ the autocorrelation function is negative, $\rho(\tau) < 0$, $\tau \neq 0$, with $\sum_{-\infty}^{\infty} \rho(\tau) = 0$. Correspondingly, the spectrum of y_t at frequency zero is null. For $0 < d < 1/2$ the autocorrelation function is positive, $\rho(\tau) < 0$, $\tau \neq 0$, with $\sum_{-\infty}^{\infty} \rho(\tau) = \infty$, induced by its slow hyperbolic decay (long memory). Correspondingly, the spectrum of y_t at frequency zero is infinite. For $d = 0$ we are in the class of conventional ARMA (stationary and invertible).

The integrated process of y_t, Y_t, is defined by $Y_t - Y_{t-1} = y_t$, or

$$Y_t = Y_0 + \sum_{\tau=1}^{t} y_\tau$$

with initial values $Y_t = 0$, $t \leq 0$, given. Replacing y_t by its Wold representation we obtain

$$Y_t = Y_0 + \mu\,t + \widetilde{\Pi}_{(-1-d),t}(L)\,\epsilon_t$$

where the elements $\tilde{\pi}_{(-1-d),t,j}$ of $\widetilde{\Pi}_{(-1-d),t}(L)$ converge to the elements $\pi_{(-1-d),j}$ of $\Pi_{(-1-d)}(L)$ as $t \to \infty$. So in case of $-1/2 < d < 0$, i.e. $1/2 < (1+d) < 1$, the process Y_t may be called drift-reverting, which seems to be more appropriate than "mean-reverting".

For $\mu = 0$ the variance of the sample mean of realizations of a fractionally integrated process of length n depends on n and d.

$$Var(\overline{y}_n) = O(n^{2d-1})$$

for $d > -1/2$. $\overline{y}_n = \sum_{t=1}^{n} y_t/n$. Due to the large proportion of long cycles in the series - as perceptible in the frequency representation of ARFIMA processes below - long range dependence enforces a slower convergence of the sample mean than in the short memory case. The increasing memory in the series slows down its convergence in terms of its variance until no convergence at $d = 1/2$. Further, for the nonstationary region $1/2 < d < 1$ the variance of the sample mean is smaller than that of an I(1) model, increasing as d approaches 1, while the unconditional mean does

not exist.

Another interpretation of the fractional integration parameter d is found by inspection of the (pseudo) spectral density $f(\lambda)$ - see below Sec. 2.4 - close to frequency zero. It holds

$$f(\lambda) = O(\lambda^{-2d})$$

The d-value measures the slope of the (pseudo) spectral close to the origin.

2.3 Estimation of nonstationary fractional integration

The estimation of stationary fractionally integrated models have been discussed extensively in the literature. Hauser(1997), and Hauser(1999) respectively, summarize semiparametric, and maximum likelihood approaches and discuss their small sample properties.

There are three different approaches to cope with stationarity and nonstationarity when modeling long memory:

(i) time domain approximate maximum likelihood,

(ii) fractional differentiation and testing the appropriate hypothesis,

(iii) frequency domain estimation with special tapers to eliminate trends.

(i) Beran(1995) developed an approximate time domain maximum likelihood estimator for ARFIMA-$(p, d, 0)$ models with $d > -1/2$. This allows the simultaneous estimation of the AR coefficients (stationarity of the AR polynomial assumed) and the fractional integration parameter for $d > 1/2$ without previously taking differences of integer order. Beran, Bhansali and Ocker(1998) further show that the model selection criteria AIC (Akaike's information criterion), SIC (Schwartz information criterion) among others, have the same properties as under conventional ARMA models. However, Monte Carlo simulations in Beran, Bhansali and Ocker(1998), table 2, indicate that the estimated d values, when the mean of the data is unknown and has to be estimated, might exhibit a serious negative bias in small samples.

(ii) Gil-Alana and Robinson(1997), e.g., have performed a sequence of tests developed by Robinson(1994) along a grid of d values to classify historic macro economic series of the US according

to their nonstationarity behavior. Their first step is to choose a hypothesis $H_0 : d = d_0$, then they fractionally difference the series, i.e. $(1 - L)^{d_0} y_t$, and perform a LM-type test based on periodogram regressions. In the latter also AR(p) components of a fixed order are allowed. The test statistics turn out to be asymptotically standard normal distributed under regularity conditions. In order to choose the order of the AR component several values for p for each candidate of d_0 are considered. p is chosen to minimize the absolute value of the test statistics.

(iii) Adopting pseudo spectral densities, i.e. spectral densities for nonstationary processes, using the Whittle objective function and allowing AR parametrization the approach by Velasco and Robinson(2000) avoids Beran's problem, and is likely to be more powerful than the sequential testing procedure of Robinson(1994). (The latter also holds for the semiparametric Geweke-Porter-Hudak approach as proposed by Hurvich and Ray(1995).) It will be applied to our data set and is described in more detail below.

2.4 Whittle pseudo-maximum likelihood estimation for nonstationary times series

For our purpose, we modify slightly the presentation of Velasco and Robinson(2000), short VR, and restrict it to $-1/2 < d < 3/2$ and to $\mu = 0$ (no drift) for nonstationary models ($1/2 \leq d < 3/2$). The pseudo spectral density of an autoregressive fractionally integrated moving average, ARFIMA, process is

$$f(\lambda|\theta, \sigma^2) = |1 - e^{-i\lambda}|^{-2d} \frac{|\beta_q(e^{-i\lambda})|}{|\alpha_p(e^{-i\lambda})|} \frac{\sigma^2}{2\pi}$$

with $\lambda \in (0, \pi]$ and $d > -1/2$. The AR and the MA polynomials $\alpha_p(L)$ and $\beta_q(L)$ of order p and q are assumed to have roots outside the unit circle, and to have no roots in common. For $d > 1/2$ the pseudo spectral density is interpreted as the limit of the expected sample periodogram, as in the stationary framework. For $-1/2 < d < 1/2$ the pseudo spectral density reduces to the commonly used spectral density.

The parameter vector θ collects all parameters except the error variance, σ^2.

Instead of the original data we use the tapered series. The periodogram for tapered data is given by

$$I(\lambda_j) = \frac{1}{n \sum_t h_t^2} |\sum_{t=1}^{n} h_t y_t \exp(-i t \lambda_j)|^2$$

7

at the frequencies $\lambda_j = 2\pi j/n$, and taper $\{h_t\}_{t=1}^n$. The purpose of tapering is the elimination of trends and down weighting of nonstationarity. For example the Zhurbenko taper of order 2 (VR p.1231) has weights, $h_t / \sum h_t^2$, forming a triangle. It can cope with nonstationarity of type $d < 3/2$, of type $d < 2$ if μ is known to be null, or of a linear trend. The expense is a loss of ordinates. Only every second ordinate would be included in the discrete frequency domain objective function below, since this taper destroys the orthogonality of the sine and cosine transforms at close frequencies. The full cosine bell taper showing properties of order 1 and 3, is suitable for $-1/2 < d < 1/2$ with $\mu \in \mathbb{R}$, and $d < 3/2$ if $\mu = 0$ making only the exclusion of the first periodogram ordinate necessary. The well known Tukey-Hanning taper, e.g., is of order 1 and so apt for $-1/2 < d < 1/2$ with all ordinates included.

Actually, two tapers will be employed: The Tukey-Hanning taper for stationary models - its competing performance is illustrated in Hauser(1999) - and the full cosine bell taper, FCB. The Tukey-Hanning taper is given by

$$
h_t = \begin{cases}
\frac{1}{2}[1 - \cos\{\pi(t - \frac{1}{2})/l\}] & t = 1, \ldots, l \\
1 & t = l+1, \ldots, n-l \\
\frac{1}{2}[1 - \cos\{\pi(n - t + \frac{1}{2})/l\}] & t = n-l+1, \ldots, n
\end{cases}.
$$

with $2\,l = 4\sqrt{n}$ the number of data points altered. The full cosine bell taper is given by

$$
h_t = \frac{1}{2}(1 - \cos\frac{2\pi t}{n})
$$

For estimation the discrete frequency-domain Whittle function for the tapered data is used as objective function.

$$
\mathcal{L}(\theta, \sigma^2) = -\sum_{j=k}^m \log f(\lambda_j | \theta, \sigma^2) - \frac{1}{2\pi} \sum_{j=k}^m \frac{I(\lambda_j)}{f(\lambda_j | \theta, \sigma^2)}
$$

where m is the largest integer in $(n-1)/2$. The actual estimation is performed using the reduced function with respect to σ^2. For the Tukey-Hanning taper all ordinates are included, i.e. $k = 1$, while for the FCB the first is excluded, so $k = 2$. Given the true model, $\theta = \theta_0$, the properties of the Whittle estimates based on the FCB tapered data are asymptotically normal

$$
\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, 4\pi\,\Phi_{FCB}\,\Sigma_0^{-1})
$$

with $\Phi_{FCB} = 35/18$ and Σ_0 the analogue to the information matrix at θ_0, cp. VR.

3 The empirical study

3.1 Data specifications

The database contains yearly data on profits for 156 surviving US manufacturing companies for the period 1950-1999. The sample corresponds to those firms among the largest 500 US manufacturing companies (in terms of sales) as of 1950 for which a complete time series on profits spanning the period 1950-1999 exists. The main data source was Compustat. Moody's Industrial Manual was used for missing points and Global Vantage for the most recent years. The Compustat variable name corresponding to the proxy for income is "Income before extraordinary items" and it represents the income of a company after all expenses, including special items, income taxes and minority interests but before provisions for common and/or preferred dividends. Total assets include current assets plus net property, plant and equipment plus other noncurrent assets. For more information of the construction on the database see Gschwandtner(2005).

Profit is defined as net income divided by total assets and throughout the study the profit rate of company i at time t, π_{it}, is defined as the relative deviation from an economy-wide measure[1] at time t, Π_t^*. Suppose Π_{it} denotes the profit of company i at period t, and Π_t^* the economy-wide measure in period t, then π_{it} is defined as $(\Pi_{it} - \Pi_t^*)/\Pi_t^*$. The normalization by the economy wide measure has two advantages. First it removes the impact of the business cycles or other shocks common to all companies, and second it improves upon the normalization by the sample mean usually used in the literature. The profits of the sample studied might not be abnormal with respect to the own sample average but might be well above (or below) the economy measure.

3.2 Model selection and testing

Given 156 yearly profit rate series of relatively short lengths with only 50 observations, we are trying to obtain as reliable estimates and tests for the fractional integration parameter as possible. So we assume that in case of nonstationarity the differenced series has a mean of null excluding long run drifts in profit rates, which is plausible and does not make the implementation of more volatile estimators of higher order necessary.

Low order ARFIMA(p, d, q) models with p, q up to 2 are estimated by means of the Whittle objective

9

function using FCB tapered data. The Schwartz criterion is used to choose the most appropriate model among them for each series. The results for the corrected AIC, AICC, are also given for comparison.

$$SIC(p, q) = -2 \log \mathcal{L}(\theta, \sigma^2) + \log(n)\,(p + q + 2)$$

$$AICC(p, q) = -2 \log \mathcal{L}(\theta, \sigma^2) + 2\,\frac{(p + q + 2)\,n}{n - p - q - 3}$$

with $\theta = \theta(p, d, q)$ and $\sigma^2 = \sigma^2(p, d, q)$. The d estimates thus obtained are then improved. First the series are fractionally differenced with the estimated d-values - using the partial autocorrelation function - and the chosen models are reestimated using the Tukey-Hanning taper. The final d-estimate is then the sum of both d estimates. (A similar 2-step procedure is proposed in VR.)

In order to test for the alternative of integer integrated models, i.e. $d = 0$ or $d = 1$, we perform two likelihood ratio tests, one with $H_0 : d = 0$, and the other with $H_0 : d = 1$. In the first case we restrict d in the chosen model to be zero, and compare both the restricted and the unrestricted estimated model using the Whittle function together with the Tukey-Hanning taper. In the second case, we integer difference the series once, and proceed as if $d = 0$. Both LR-tests are asymptotically χ^2 distributed with one degree of freedom. Due to the short length of our series we use their small sample distributions as given in table 1. Thereby normal errors are assumed and the number of replications is 1000. For illustration the critical values for the FCB taper are also calculated.

Table 1: $(1 - \alpha)$-Quantiles of the small sample distribution of LR-tests, $n = 50$. Comparison of the Tukey-Hanning and the Full Cosinus Bell taper. The small sample $(1 - \alpha)$-probabilities given the asymptotic 0.90, 0.95 and 0.99-quantiles according to the χ_1^2 (2.71, 3.84, 6.63) are printed in brackets.

	Taper		
	Tukey-Hanning	FCB	FCB
$1 - \alpha$	$H_0 : d = 0$	$H_0 : d = 0$	$H_0 : d = 1$
0.90	3.040 (0.876)	3.908 (0.810)	4.837 (0.783)
0.95	4.432 (0.930)	5.307 (0.894)	6.575 (0.857)
0.99	7.105 (0.985)	8.637 (0.974)	12.464 (0.951)

It turns out that the simulated distribution for the LR-test using the Tukey-Hanning taper exhibits only minor deviations from the asymptotic χ_1^2, contrary to the case of the FCB and $H_0 : d = 1$.

3.3 Application to the data

The application of the model selection criterion SIC and the 2-step estimation procedure outlined in Section 2 yields the following distribution of the estimated d-values for the 156 series as depicted in Figure 1. Table 2 gives some simple descriptive indicators.

Figure 1: Histogram of the d-estimates. Models chosen by SIC, 2-step estimation.
*** INCLUDE ABOUT HERE ***

Table 2: Descriptive statistics for the estimated d-values. Models chosen by SIC, 2-step estimation. 156 series.

Mean	0.516	Median	0.517
Standard Deviation	0.400	Variance	0.160
Skewness	-0.210	Kurtosis	3.953
Minimum	-0.922	Maximum	1.764

Three estimates are located to the very left of the histogram and two to the very right. After testing (see below) they are assigned either to the integer values $d = 0$ or $d = 1$. So, we interpret their occurrence as the result of overparametrized models. The overall shape is almost symmetric with a mean of 0.516. The hypothesis of a normal distribution could be rejected at a 5% but not at a 1% level using the Jaques-Bera test.

The next step is to test for the integer values, $H_0 : d = 0$ and $H_0 : d = 1$. At a significance level of 5% four (six for the AICC) categories of various degrees of integration are distinguished. Table 3 gives the results.

Table 3: Decisions for the 156 profit rate series based on the LR-tests $H_0 : d = 0$ and $H_0 : d = 1$. Numbers of series falling into each category.

	$d < 0$	$d = 0$	$0 < d < 0.5$	$0.5 \leq d < 1$	$d = 1$	$d > 1$	total	no decision
SIC	0	37	33	21	61	0	152	4
AICC	1	45	24	15	55	1	141	15

Remarkably 4 series cannot be assigned by means of the LR-tests when relying on the SIC. The reason seems to be the large variance in the parameter estimates of the chosen models and so a flat likelihood function, which might be due to model inadequacy and/or to the short length of the series. The AICC rule yields considerably more "no decision" cases than the SIC. In all those models the number of parameters is larger than in the SIC model, and seem to be overparametrized. Since the overall pattern is essentially the same, we will refer below only to the SIC results.

Ignoring the no decision cases, the estimated and tested d values are all between zero and one, categorized as stationary ($d = 0$ and $0 < d < 1/2$) and nonstationary ($1/2 \leq d < 1$ and $d = 1$). Fractionally integrated models are chosen as a good approximation for 35.5% of the series with 21.7% stationary and 13.8% nonstationary. This is remarkable in the light of the shortness of the series and the conservative character of the testing procedure for $d = 0$ and $d = 1$.

Neither significant negative nor long memory beyond $d = 1$ is observed, and 46% (54%) of the series are stationary (nonstationary). This is comparable to the stationarity analysis from the "Persistence of Profit" studies. Kambhampati(1995), using the Dickey-Fuller test for I(1) could not reject nonstationarity of profits in only 13 out of 42 cases for Indian industry-level data. Goddard and Wilson(1999) employing data for 335 U.K. firms over the period 1972-91 likewise report nonstationarity in approximately 75% of firms in the sample. Crespo Cuaresma and Gschwandtner(2005a) find for the same data set but using the normalization by the sample mean in more than 32% of the cases that the null of a unit root cannot be rejected at 5% significance level using individual Dickey Fuller tests.

Further we find that the type of normalization of the data is not essential for our conclusions, as the results for the unnormalized profit rates and the models chosen with the SIC are very similar. (Tables are available on request.)

4   Summary

Notwithstanding the small number of observations, we find evidence that 35.5% of the series under study can be well approximated by fractionally integrated processes showing various degrees of long range dependence ($0 < d < 1$). Only 46% of the series are found stationary and so 54% nonstationary. This is remarkable since the introduction of the long memory parameter d allows for various degrees of stationarity/nonstationarity, and so for a larger class of stationary (also nonstationary) low order models. The frequent occurence of nonstationarity strongly contradict the competitive environment hypothesis which implies that the erosion of profits happens in a mean reverting process.

References

Baillie, R.T. and Bollerslev,T., 1994, Cointegration, fractional cointegration and exchange rate dynamics, Journal of Finance, 49, 737-745.

Beran, Jan, 1995, Maximum Likelihood estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models, Journal of the Royal Statistical Society, Series B, 57(4), 659-672.

Beran, Jan, Bhansali, R.J. and Ocker, D., 1998, On Unified Model Selection for Stationary and Nonstationary Short- and Long-Memory Autoregressive Processes, Biometrika, 85, 921-934.

Cable, J., Jackson, R.H.G. and Rhys, H., 2001, Profit Cycles: The Dynamics of Corporate Earnings Revisited, Mimeo, School of Management and Business, University of Wales.

Crespo Cuaresma, J. and Gschwandtner, A., 2005a, The competitive environment hypothesis revisited: Nonlinearity, nonstationarity and profit persistence, Applied Economics, forthcoming.

Crespo Cuaresma, J. and Gschwandtner, A., 2005b, Tracing the dynamics of competition: Evidence from company profits, Working Paper Nr.0504, Department of Economics, University of Vienna.

Diebold, F.X. and Rudebusch, G.D., 1989, Long run memory and persistence in aggregate output, Journal of Monetary Economics, 2, 189-209.

Geroski, P.A. and Jacquemin, A., 1988, The Persistence of Profits: A European Comparison, The Economic Journal, 98, 375-389.

Gil-Alana, L.A. and Robinson, Peter M., 1997, Testing of unit root and other nonstationary hypotheses in macroeconomic time series, Journal of Econometrics, 80, 241-268.

Gil-Alana, L.A. and Robinson, Peter M., 2000, Testing of Seasonal Fractional Integration in UK and Japanese Consumption and Income, Discussion Paper No. EM/00/402, The Suntory Centre, London School of Economics and Political Science.

Glen, J., Lee, K. and Singh, A., 2001, Persistence of Profitability and Competition in Emerging markets, Economics Letters, 72, 247-253.

Goddard, J.A. and Wilson, J.O.S., 1999, The Persistence of Profit: A New Empirical Interpretation, International Journal of Industrial Organization, 17, 663-687.

Granger, C.W.J., 1980, Long memory relationships and the aggregation of dynamic models, Journal of Econometrics, 14, 227-238.

Gschwandtner, A., 2005, Profit persistence in the "very" long run: evidence from survivors and exiters, Applied Economics, 37(7), 793-806.

Hauser, M.A., 1997, Semiparametric and nonparametric testing for lon memory: a Monte Carlo study, Empirical Economics, 22, 247-271.

Hauser, M.A., 1999, Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo simulation, Journal of Statistical Inference and Planning, 80, 229-255.

Hurvich, C.M. and Ray, B.K., 1995, Estimation of the memory parameter for nonstationary or non-invertible fractionally integrated processes, Journal of Time Series Analysis, 16(1), 17-42.

Kambhampati, U.S., 1995, The Persistence of Profit Differentials in Indian Industry, Applied Economics, 27, 353-361.

McGahan, A.M. and Porter, M.E., 1999, The Persistence Of Shocks To Profitability, The Review of Economics and Statistics, 81(1), 143-153.

Mueller, D.C., 1977, The Persistence of Profits Above the Norm, Economica, 44, 369-380.

Mueller, D.C., 1986, Profits in the Long Run, Cambridge University Press, Cambridge.

Mueller, D.C.(ed.), 1990, Dynamics of Company Profits: An International Comparison, Cambridge University Press, Cambridge.

Maruyama, N. and Odagiri, H., 2002, Does the 'persistence of profits' persist?: a study of company profits in Japan, 1964-97, International Journal of Industrial Organization, 20, 1513-1533.

Robinson, Peter M., 1994, Efficient tests of nonstationary hypotheses, Journal of the American Statistical Association, 89, 1420-1437.

Velasco, Carlos and Robinson, Peter M., 2000, Whittle Pseudo-Maximum Likelihood Estimation for Nonstationary Time Series, Journal of the American Statistical Association, 95(452), 1229-1243.

Footnotes:

[1] The economy wide measure is the median of the profit of a sample consisting of more than 175000 observations and more than 15500 US manufacturing companies.

Figure1: Histogram of the d—estimates. Models chosen by SIC, 2—step estimation.