

Bianchi, Carluccio; Menegatti, Mario

Working Paper

On the potential pitfalls in estimating convergence by means of pooled and panel data

Quaderni di Dipartimento - EPMQ, No. 160

Provided in Cooperation with:

University of Pavia, Department of Economics and Quantitative Methods (EPMQ)

Suggested Citation: Bianchi, Carluccio; Menegatti, Mario (2004) : On the potential pitfalls in estimating convergence by means of pooled and panel data, Quaderni di Dipartimento - EPMQ, No. 160, Università degli Studi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi (EPMQ), Pavia

This Version is available at:

<https://hdl.handle.net/10419/87131>

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.

**On the potential pitfalls in estimating convergence
by means of pooled and panel data**

Carluccio Bianchi
(Università di Pavia)

Mario Menegatti
(Università di Parma)

160 (02-04)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Febbraio 2004

On the potential pitfalls in estimating convergence by means of pooled and panel data

by

Carluccio Bianchi

Dipartimento di Economia Politica e Metodi Quantitativi, Università di Pavia

Mario Menegatti

Dipartimento di Economia, Università di Parma

Abstract. We show that the use of pooled and panel data in estimating convergence across countries (or other territorial areas) may involve some pitfalls since this type of data cannot properly distinguish between actual convergence and the possibility of decreasing growth rates over time within each country (or geographical unit). We also show how the possible bias in the estimates can be eliminated.

Keywords: Convergence, growth, panel data, pooling

JEL Classification: O41, O47, O49

Acknowledgements: We wish to thank Gianni Amisano, Lorenzo Cappellari, Vincenzo Dall'Aglio, Francesco Daveri, Paolo Epifani, Andrea Lasagni, Matteo Manera, Rocco Mosconi, Massimiliano Serati and Gilberto Seravalli for their useful comments and suggestions on a previous version of the present paper. We are also particularly indebted to Eduardo Rossi for his great help in the Montecarlo simulations. The usual disclaimers obviously apply.

1. Introduction

One of the main implications of neoclassical growth theory is that in the long run per capita incomes of different countries (or other territorial areas) should converge either absolutely or conditionally on the peculiar features of the equilibrium steady states. The most commonly used test in checking for the existence of convergence (absolute or conditional) across economies concerns the relationship between the initial income level and the subsequent growth rate in a selected group of countries (or geographical units). If this relationship turns out to be negative, then it is argued that poor countries (areas) tend to grow faster than rich ones, so that convergence (and more precisely β -convergence) is claimed to exist.^{1,2}

Many empirical works studying convergence, starting from Barro (1991) and Mankiw, Romer and Weil (1992), perform the above described test using cross-countries data, where, for each economic system, the initial per capita GDP and the subsequent growth rate in a given period (usually some decades) are used. These cross-section analyses have been, subsequently, followed by other studies (e.g. Barro and Lee (1993), Islam (1995), Caselli et al. (1996), Evans (1997), Dowrick and Rogers (2002)) using pooling and/or panel estimation techniques, where for each country a higher number of observations referring to different points in time are considered. As well known, the common justification for the use of these techniques (see, for instance, Durlauf and Quah (1999) and Temple (1999)) is the possibility of increasing the number of observations used in the estimates and also, in the case of panel data analyses, the possibility of capturing the likely existence of economy-specific effects and even time-specific effects.

The aim of this work is to show the potential pitfalls connected with the use of pooled and panel data in testing for convergence across countries. In

¹ As we already mentioned, the empirical tests usually distinguish between absolute convergence (directly measuring the relationship between the growth rates and initial incomes) and conditional convergence (including variables capable of capturing the differences in the various countries steady states).

² It is worthwhile emphasizing that this paper is not concerned with the issue of the most appropriate measure of economic convergence and thus with the debate around the so-called Galton's fallacy stimulated by Quah (1993). Indeed the aim of the work is to show that, within the

particular, it can be claimed that the use of a technique simultaneously considering observations referring to different countries in the same period of time and observations referring to the same country in different periods of time may imply that the traditional convergence coefficient will capture not only the possible tendency for poor countries to grow faster than rich ones but also the possible tendency for each single country to grow at a decreasing rate over time. Although these two phenomena are both implied by neoclassical theory in the transition phase towards the equilibrium steady state, it is certainly possible that in reality they occur separately, so that, for instance, there may be no actual convergence across countries, while at the same time the growth rate within each country may decrease over time. This problem implies that the usual tests about convergence performed using pooling and panel techniques may be misleading, since they could indicate the presence of convergence even in cases where there is none.³

The paper proceeds as follows. Section two illustrates a numerical example where the GDPs of a constructed group of countries diverge while pooling and panel estimates indicate instead the existence of convergence. Section three suggests a possible technique to compute an unbiased test for convergence in a data set containing many observations for each country. Section four concludes.

2. A simple example

The simplest econometric test for convergence across countries is a cross-section regression. A common implementation of this technique follows the

traditional β -convergence approach, the use of pooling and panel techniques may produce misleading results.

³ It is worthwhile emphasizing that pooling and panel techniques are used to test the conclusions of a wide variety of models and theories, besides the convergence issue. The possible shortcomings studied in this paper only refer to the specific problem examined.

suggestion proposed by Mankiw, Romer and Weil (1992), according to which the following equation is estimated:⁴

$$\log\left(\frac{y_{jm}}{y_{jh}}\right) = \mathbf{a} + \mathbf{b} \log(y_{jh}) + \mathbf{g} \log(s_j) + \mathbf{d} \log(n_j + d + g) + \mathbf{e}_j \quad [1]$$

with $m > h$, $j = 1, \dots, N$

where y_{jh} and y_{jm} are respectively the levels of per capita income in country j at the initial time h and at the final time m , s_j is the average saving rate of country j in the time period from h to m , n_j is average population growth in the same period, d is the average depreciation rate, g is the average rate of technological progress and \mathbf{e}_j is a random error term. The depreciation rate and the rate of technological progress are assumed to be equal across countries (as usually done in the literature). The usual conclusion of econometric analyses is that if the β coefficient is negative then the sample data support the convergence conjecture.

In a large part of the convergence literature cross-section estimates are substituted or accompanied by a pooling regression. In this case equation [1] is estimated considering N countries and T periods, so that, for each country, we have a total number of T observations for any variable. The advantage of pooled estimates, with respect to cross-section ones, is thus given by the fact that the total number of observations is increased to $N \times T$.

The same data set can be also used to perform a panel data analysis whereby either country-specific fixed effects or both country-specific and time-specific fixed effects are introduced.⁵ In the first case the equation to be studied becomes:

$$\log\left(\frac{y_{jm}}{y_{jh}}\right) = \mathbf{a}_j + \mathbf{b} \log(y_{jh}) + \mathbf{g} \log(s_j) + \mathbf{d} \log(n_j + d + g) + \mathbf{e}_j \quad [2]$$

⁴ It should be noticed that Mankiw, Romer and Weil's original estimates concerned a cross-section of countries in a single period of time and that they also considered a model including human capital. In this paper we neglect human capital in order to simplify the exposition, while maintaining the validity of the conclusions we reach.

⁵ For the interpretation of this model see Islam (1995) and Durlauf and Quah (1999).

with $m > h$, $j = 1, \dots, N$

while in the second case the equation to be estimated is:

$$\log\left(\frac{y_{jm}}{y_{jh}}\right) = \mathbf{a}_j + \mathbf{h}_i + \mathbf{b} \log(y_{jh}) + \mathbf{g} \log(s_j) + \mathbf{d} \log(n_j + d + g) + \mathbf{e}_j \quad [3]$$

with $m > h$, $j = 1, \dots, N$; $i = 1, \dots, T$ ⁶.

Again the usual claim is that there is convergence if β is negative.

A panel-data technique exhibits two advantages with respect to a cross-section regression: first, it increases the number of the observations; second, it makes it possible to consider the potential existence of technological heterogeneity between countries.

In order to show the possible shortcomings of the approaches we have just described we examine a simple numerical example founded on the Solow-Swan growth model. We consider, in particular, a closed economy where firms produce a single good Y using the Cobb-Douglas technology:

$$Y_t = K_t^q (A_t L_t)^{1-q} u_t \quad [4]$$

where Y_t is output, K_t capital, L_t labour, A_t the level of technology and u_t is a random shock.⁷ Since, as usual, constant returns to scale are assumed to prevail, we have the following expression for GDP per head y_t :

⁶ In the actual estimates, however, only T-1 time dummies are introduced, in order to avoid the regressors coefficient matrix to be singular.

⁷ If we let Y_{nt} stand for the so-called natural or potential output, that is the trend level of output that would prevail in the Solow-Swan model in the absence of shocks, then the subsequent output gap between actual and potential output would be $\frac{Y_t - Y_{nt}}{Y_{nt}}$, usually approximated by

$\log \frac{Y_t}{Y_{nt}} = \log u_t$. In line with the actual experience of developed countries we constrain the shock

to produce an output gap not greater, in absolute value, than 3%. Since the shock $\log u_t = \tilde{u}_t$ is

$$y_t = \frac{Y_t}{L_t} = \left(\frac{K_t}{L_t} \right)^q A_t^{1-q} u_t = k_t^q A_t^{1-q} u_t \quad [5]$$

We assume furthermore that capital depreciates through time at the constant rate d , while population grows at the constant rate n and technology improves at the constant rate g so that $A_{t+1} = A_t(1+g)$. We finally assume that saving is a constant fraction s of output. Under these assumptions capital accumulation is described by the usual dynamic equation⁸:

$$Dk_t = sy_t - (d+n)k_t \quad \text{with } s \in (0,1), \quad [6]$$

where $Dk_t = k_{t+1} - k_t$.

Using this structure, we generate a data set containing the output levels for eight economies for a time period of 100 years, estimating the values of y_t and k_t for each year and for every economy. We then perform a Montecarlo experiment simulating 1000 times the possible growth paths of our constructed economies. The whole data set is obtained by supposing $s = 0.5$ and $n + d = 0.08$, while $g = 0.02$. The level of the initial capital is set equal to 10 in the poorest country and is then increased by one unit in passing from an economic system to the next one. The value of coefficient q in the production function is set equal to 0.6 in the poorest country and is then increased by 0.01 in passing from an economic system to the next one. The differences in the production function and in the initial level of capital are the unique elements of heterogeneity between countries.

normally distributed, i.e. $\tilde{u}_t \sim N(0, \mathbf{s}^2)$, then σ^2 will be such that, for instance,

$$\Pr\left(-0.03 \leq \frac{\tilde{u}_t}{\mathbf{s}} \leq 0.03\right) = 0.95.$$

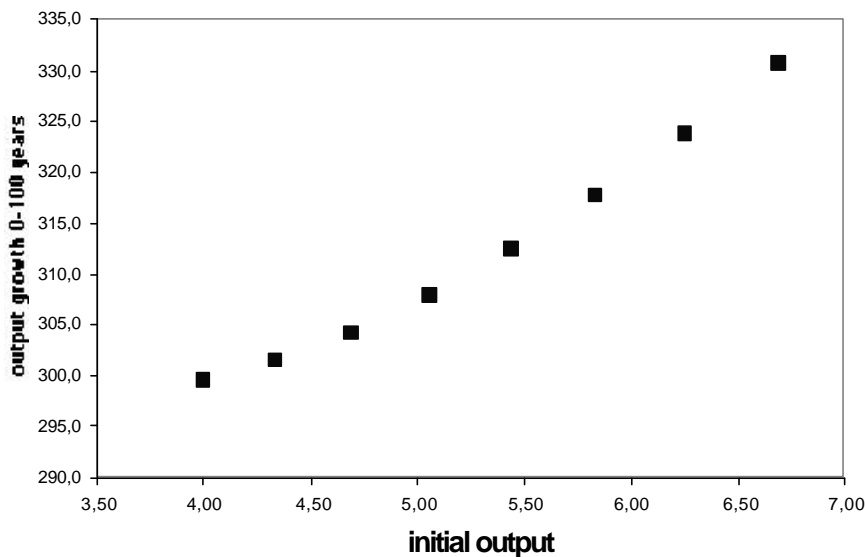
⁸ Actually the dynamic equation [6] holds exactly only in continuous time. For the sake of simplicity, however, we shall use it as a discrete time approximation to compute the data of our example.

Table 1. Initial output and output growth in the eight countries

Country	Initial output (= y_0)	Output growth after 100 years (= $\log(y_{100}/y_0)$)
Country 1	3.98	299.9
Country 2	4.32	301.9
Country 3	4.67	304.8
Country 4	5.03	308.3
Country 5	5.41	312.9
Country 6	5.81	318.1
Country 7	6.23	324.2
Country 8	6.67	331.1

The structure proposed in the example is such that the larger the initial capital in a given economic system the higher the growth rate in the period considered. We have thus a positive relation between initial output and subsequent growth, implying a phenomenon of *divergence* in per capita GDPs across countries. This phenomenon is illustrated in table 1 and figure 1.⁹

Figure 1. Divergence across countries per capita GDPs



⁹ It is worthwhile noticing that the values of the output growth rates after 100 years reported in table 1 and figure 1 are the averages of the 1000 simulations for each economy.

We now use the data generated in the just described experiment to estimate equations [1], [2] and [3]¹⁰, subdividing the whole period of 100 years into four sub-periods of 25 years. We estimate the three equations for every simulated growth path and we repeat the estimations 1000 times. The average estimate¹¹ of coefficient β implied by a pooling equation is -0.273 , with a standard error in the simulations of 0.003 . When a panel approach is used, the value of the same coefficient is -0.294 (with a standard error of 0.003) if equation [2] is estimated, while it is -0.494 (with a standard error of 0.070) if equation [3] is estimated instead. In all three cases the average values of the β coefficient are strongly significant, even though with the wrong sign¹². Thus we reach the conclusion that, while the generated data should exhibit, by construction, a divergence in per capita incomes, the pooling and panel data estimates give a strong indication in favour of convergence.

3. An unbiased convergence test

As anticipated in the introduction, the reason for the possible failure of pooling and panel data techniques in distinguishing between convergence and divergence is related to the fact that these models end up with estimating two different simultaneous phenomena. Indeed, the sign of the β coefficient in equations [1], [2] and [3] indicates at the same time:

- whether initially poor countries grow faster than initially rich ones (the convergence issue);
- whether the growth rate within each country is decreasing over time.

These two phenomena, although both caused by the assumption of diminishing returns to physical capital, can occur either together or separately.

¹⁰ In our example the variables s , n , d and g are assumed to be constant. For this reason the estimates are performed using the initial income as the unique regressor, since the values of s , n , d and g are captured by the constants of the regressions.

¹¹ For simplicity, the estimate is performed using the OLS technique.

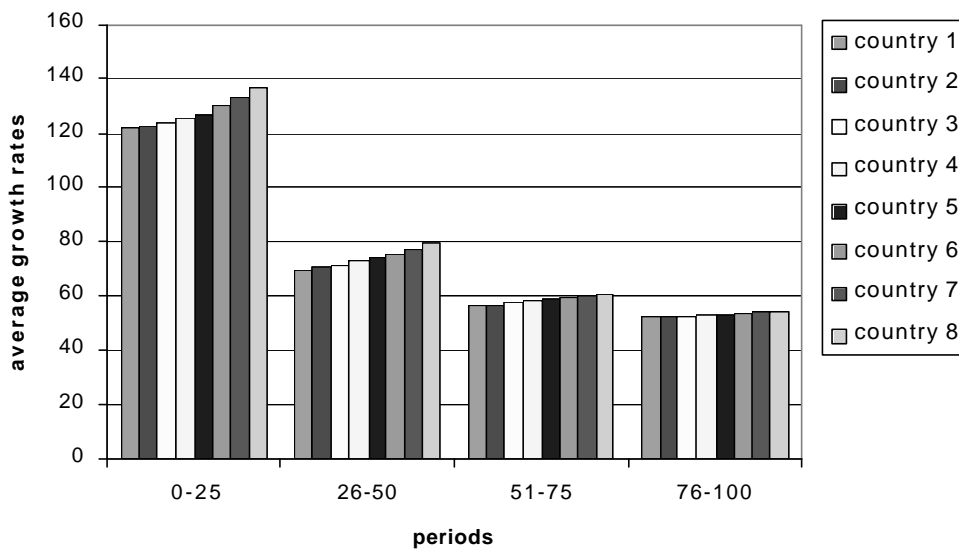
¹² It should be added that the β coefficients are also highly significant in every single regression run, with average t-values around -13 , -18.5 and -41 respectively for equations [1], [2] and [3].

In the numerical example considered in section two, in particular, we have no convergence (as shown in table 1 and in figure 1), but at the same time we have a growth rate decreasing over time in each economic system (as shown in table 2 and in figure 2). The negative values of β obtained estimating equations [1], [2] and [3] are thus caused by this second phenomenon, which is sufficiently strong to overcome the divergence in growth rates occurring across countries.

Table 2. Average initial output and average growth rate in each sub-period

Period	Average initial output	Average growth
0-25 years	5.27	127.4
25-50 years	19.00	73.7
50-75 years	39.99	58.5
75-100 years	72.09	53.0

Figure 2. The evolution of growth rates over time



Since the possible bias in the convergence test is due to the simultaneous presence in each country of an increasing output level and a decreasing output growth, this bias can be easily eliminated by estimating the value of the β coefficient in a model where the relevant variables to be used in the regressions are, for each sub-period, the deviations from the average between countries. Indeed, by using deviations from sub-period averages, the model compares the

countries growth rates and initial incomes net of the effects of these variables evolution over time. Thus the equation to be estimated becomes:

$$\log\left(\frac{y_{jm}}{y_{jh}}\right) - av\left(\log\left(\frac{y_{jm}}{y_{jh}}\right)\right) = \mathbf{b}[\log(y_{jh}) - av(\log(y_{jh}))] + \mathbf{g}[\log(s_j) - av(\log(s_j))] + \mathbf{d}[\log(n_j + d + g) - av(\log(n_j + d + g))] + \mathbf{e}_j \quad \text{with } m > h \quad [7]$$

where $av(z_t)$ is the average of variable z_t between countries at time t . It should be noticed, moreover, that equation [7] is formally equivalent to estimating a model where time-specific, but no country-specific, fixed effects are introduced, i.e. to estimating equation [3] with $\mathbf{a}_j=0$ "j.

When equation [7] is estimated, using the data of the numerical example described in section two, we get an average value for coefficient β of 0.098, with a standard error in the simulations of 0.007¹³. The estimate of equation [7], then, enables to show the actual divergence occurring across countries in a way that traditional models [1], [2] and [3] previously examined cannot capture.

It is important to notice, however, that the model underlying equation [7] exhibits some differences with respect to traditional pooling and panel techniques. As we pointed out in section two, pooling and panel regressions were introduced in convergence analysis in order to increase the number of observations used in the estimates, and, in the case of panel-data analyses, also to take into account technological differences between countries. It is obvious that in estimating equation [7] the first goal is achieved but not the second one. This fact does not imply, however, that the possible existence of significant technological differences between countries cannot be properly considered in convergence analysis. If necessary, equation [7] can be corrected either by modelling the country-specific technology as a function of other observable variables or by introducing some proxies for it. Technological differences between countries can therefore be reintroduced in convergence analyses by substituting the country-specific effects with an explicit formulation capable of capturing them.

¹³ The average t-value of the β coefficient in the 1000 regressions run is 3.9.

4. Conclusions

This paper showed the possible pitfalls connected with the use of pooled and panel data in estimating convergence. Using the data set generated by a simple numerical example, founded on the neoclassical Solow-Swan model, pooling and panel techniques indicated the existence of convergence across countries while per capita incomes actually diverged.

We claim that this result is due to the fact that pooling and panel techniques are incapable of distinguishing between the possible tendency for poor countries to grow faster than rich ones and the possible tendency for each country to grow at a decreasing rate over time.

We finally showed the existence of a possible solution to the shortcomings described, consisting in the estimate of a model where each variable to be used in the regressions is the difference between the actual observation in any country at any point in time and the average between countries of the observations in the same period of time.

5. References

- Baltagi B. H. (2001), *Econometric Analysis of Panel Data*, John Wiley & Sons Ltd., Chichester, UK.
- Barro R. J. (1991), "Economic Growth in a Cross Section of Countries", *Quarterly Journal of Economics*, **106**, 407-433.
- Barro R. J. and Lee J.-W. (1994), "Sources of Economic Growth", *Carnegie-Rochester Conference Series on Public Policy*, **40**, 1-46.
- Caselli F., Esquivel G. and Lefort F. (1996), "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics", *Journal of Economic Growth*, **1**, 363-389.

- Dowrick S. and Rogers M. (2002), “Classical and Technological Convergence: Beyond the Solow-Swan Growth Model”, *Oxford Economic Papers*, **54**, 369-385.
- Durlauf S. N. and Quah D. (1999), “The New Empirics of Economic Growth”, in J. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, North-Holland, Amsterdam, 235-308.
- Evans, P. (1998), “Using Panel Data to Evaluate Growth Theories”, *International Economic Review*, **39**, 295-306.
- Hsiao C. (1986), *Analysis of Panel Data*, Cambridge University Press, Cambridge, UK.
- Islam, N. (1995), “Growth Empirics: A Panel Data Approach”, *Quarterly Journal of Economics*, **110**, 1127-1140.
- Islam, N. (2003), “What Have we Learnt from the Convergence Debate?”, *Journal of Economic Surveys*, **17**, 309-362.
- Mankiw N. G., Romer D. and Weil D. N. (1992), “A Contribution to the Empirics of Economic Growth”, *Quarterly Journal of Economics*, **107**, 407-437.
- Quah D. (1993), “Galton’s Fallacy and Tests of Convergence Hypothesis”, *Scandinavian Journal of Economics*, **95**, 427-443.
- Solow R. M. (1956), “A Contribution to the Theory of Economic Growth”, *Quarterly Journal of Economics*, **70**, 65-94.
- Swan T. W. (1956), “Economic Growth and Capital Accumulation”, *Economic Record*, **32**, 334-361.
- Temple J. (1999), “The New Growth Evidence”, *Journal of Economic Literature*, **37**, 112-156.

**Dipartimento di economia politica e metodi quantitativi
Università degli studi di Pavia**

List of the lately published Technical Reports
(available at the web site: "<http://economia.unipv.it/Eco-Pol/quaderni.htm>").

Quaderni di Dipartimento

#	Date	Author(s)	Title
144	06-02	I.Epifani A.Lijoi	Exponential functionals and means of neutral to the right priors
145	06-02	I.Pruenster P.Bertoletti	A note on third-degree price discrimination and output
146	12-02	P.Berti L.Fratelli P.Rigo	Limit theorems for predictive sequences of random variables
147	01-03	A. Lijoi I. Pruenster	Practicable alternatives to the Dirichlet process
148	02-03	A. Lijoi Pruenster S.G.Walker	Extending Doob's consistency theorem to nonparametric densities
149	02-03	V.Leucari G.Consonni	Compatible Priors for Causal Bayesian Networks
150	02-03	L. Di Scala L. La Rocca G.Consonni	A Bayesian Hierarchical Model for the Evaluation of a Web Site
151	02-03	G.Ascari	Staggered Prices and Trend Inflation: Some Nuisances
152	02-03	G.Ascari P.Gagnepain	How inefficient are football clubs? An evaluation of the Spanish arms race
153	03-03	P.Dellaportas C.Tarantola	Categorical data squashing by combining factor levels
154	05-03	A. Lijoi Pruenster	A note on the problem of heaps
155	09-03	P.Berti P.Rigo	Finitely additive uniform limit theorems
156	09-03	P.Giudici C.Tarantola	Web Mining pattern discovery
157	09-03	M.A.Maggi U.Magnani M.Menegatti	On the Relationships between Absolute Prudence and Absolute Risk Aversion
158	10-03	P.Bertoletti	Uniform Pricing and Social Welfare
159	01-04	G.Ascari N.Rankin	Perpetual Youth and Endogenous Labour Supply: A Problem and a Possible Solution