A "Trojan Horse" in the peer-review process of fee-charging economic journals

Roberto Dell'Anno‡, Rocco Caferra*, Andrea Morone§

ABSTRACT

This paper aims to unmask the inadequacy of the review process of a sample of fee-charging journals in economics. We submitted a bait-manuscript to 104 academic economic journals to test whether there is a difference in the peer-review process between *Article Processing Charges (APC) journals* and *Traditional journals* which do not require a publication fee. The submitted bait-article was based on completely made-up data, with evident errors in terms of methodology, literature, reporting of results, and quality of language. Nevertheless, about half of the APC journals fell in the trap. Their editors accepted the article in the journals and required to pay the publication fee. We conclude that the *Traditional model* has a more effective incentive-mechanism in selecting articles, based on quality standards. Otherwise, we confirm that the so-called "Predatory Journals" – i.e. academic journals which accept papers without a quality check – exploit the APC scheme to increase their profits. They are also able to enter whitelists (e.g. Scopus, COPE). Accordingly, poor-quality articles published on APC journals shed the light on the weakness of methodologies based on a mechanical inclusion of academic journals in scientific database indexes, succeeding in being considered for bibliometric evaluations of research institutions or scholars' productivity.

JEL: *I20; A10.*
Keywords: *Predatory journal; Publication fees; Article Processing Charges; Peer-review.*

‡ Department of Economics and Statistics - CELPE, University of Salerno, Via Giovanni Paolo II, 132 – Fisciano (SA) 84084 – Italy. Email: rdellanno@unisa.it.

* Department of Economics, Management and Law, University of Bari A. Moro, Largo Abbazia Santa Scolastica, 53 - Bari 70124 - Italy. Email: rocco.caferra@gmail.com.

§ Department of Economics, Management and Law, University of Bari A. Moro, Largo Abbazia Santa Scolastica, 53 - Bari 70124 - Italy. Email: a.morone@gmail.com.

1 Introduction

Publishing articles on scientific journals is not only the foremost way for scholars to disseminate new ideas within the research community, but it also constitutes the source of scholar's bibliometric score.[1]

Quantitative analysis of research production (i.e. bibliometrics), aiming to combine a "number" to each scholar, is increasingly used to rank researchers and academic institutions. According to the slogan that "numbers don't lie" – because they seem a more "objective", comparable and unbiased measure of scholar's research productivity – we are observing a structural change in assessment criteria of the researchers' activity. Indeed, research institutions often encourage their affiliated to improve bibliometric performances because low ranking may penalize a department or a university reputation, and, especially for higher education, reputation affects students' and scholars' mobility as well as attractiveness in terms of sponsors and research funds. Although this (ab)use of bibliometric score are a well-known problem[2], more and more decision-makers use quantitative measures of research productivity to determine both the career paths of researchers and the success of academic institutions – e.g. the Italian national agency for the evaluation of universities and research institutes (ANVUR).

Unsurprisingly, academics have been largely converted to the "publish or perish" religion, and particularly for younger academics, this religion is biasing the scholars' behavior with a negative impact on the quality of research activity. Van Dalen and Henkens (2012) surveyed the high publication pressure perceived by researchers that negatively switched the interest of scientists from policy and knowing facts toward publication and citation within academic circles. According to Fanelli (2010), high publication pressure forces scientists to produce 'publishable' results at all costs and, in turn, fosters the ambition of researchers to become metric-wise.[3]

---

[1] Scholars' evaluations include production measures (such as the number of published papers on scientific journals) and impact measures (such as the Hirsch Index, which is the highest number of *n* papers receiving at least *n* citations). See Waltman (2015) for an extensive review and some useful recommendations in introducing citation impact indicators.

[2] See for instance the *Leiden Manifesto* (2015) and the *Dora Declaration* (2013), where it is claimed that quantitative measures have to be seen as an instrument and not as a goal of research activity. Both of them promote the idea that quantitative and qualitative measures are needed.

[3] However, the same author (Fanelli and Larivière 2016; Fanelli et al., 2017) pointed out as the prevalent belief that pressures to publish are causing scholars' misconducts (e.g. through "salami slicing"or duplicated publications) is likely false. In particular, Fanelli and Larivière (2016) estimated that "*The total number of papers published by researchers during their early career period (first fifteen years) has increased in recent decades, but so has their average number of co-authors. If we take the latter factor into account, by measuring productivity fractionally or by only counting papers published as the first author, we observe no increase in productivity throughout the century*.".

Our study provides evidence from the field that the combination of these incentive to increase bibliometric scores and the inclusion of open access journals which require an Article Processing Charges (APC) (also known as publication fee) to publish are conflicting with an effective decision process based on peer review to discriminate between valuable and useless research production.

In this sense, the issue of the paper is not whether bibliometric indicators can be considered useful to evaluate individual or institutional research – as they do provide a useful, even if not-exhaustive, assessment of research production – but if fee-charging journals can be considered as a trustworthy business model to prevent unethical and opportunistic conducts of publishers and scholars.

Two main (not alternative) business models to cover the cost of journal publications in scientific publishing exist the "*Traditional model*" which refers to subscription-based journals requiring the reader to pay to access to the journal and the "*Open Access model*" of publication, in which journals publish with open access – i.e., anyone has free online access to peer-reviewed research articles (Harnad, 2015) – but charge the author(s) of accepted articles of an APC.[4]

There are also "*hybrid business models"* in which the publisher requires a payment of a fee when submitting the manuscript (i.e. "submission fee")[5] or, to allow the authors to make their articles accessible to everyone on the web. However, differently from the open-access model with APC – traditional and hybrid models keep the editorial decision to accept manuscript separate from the payment of the fee.

We hypothesize that the existence of a publication fee is a sufficient condition to change the relationship among authors, publishers, and readers. The basic intuition is that while in the *Traditional model*, there is a stronger editor incentive in applying an effective peer review process to assess the quality of the submitted manuscript, this incentive disappears in the *APC journals*. It occurs because the APC model, transferring the role of funding the publications from users (i.e. scientific community) to the producers (authors) broke the incentive to look for the quality of the manuscripts because the publisher's

[4] There are two types of open access (OA) journals: Gold OA (when the publishing costs are covered by fees charged to the authors upon acceptance of their manuscripts for publication) and Green OA (when the publisher of a subscription journal allows the author to keep the non-commercial rights to his/her article so it can be posted in open internet archives). In this research, we analyze Gold OA journals. We thank an anonymous referee for this suggestion.

[5] In this hybrid category, we include journals that surcharge authors of accepted papers voluntarily opt to pay charges for color figures in print.

(increasing the revenue through the fee) and the author's aim (increasing his/her bibliometric score through an indexed journal article) is immediately achieved, without the "final users'" evaluation (i.e. readers).

The paper is organized as follows. Section 2 reviews the literature. Section 3 describes the Experimental design, econometric models, and reports the empirical outcomes. Section 4 concludes.

2 Literature

According to Osterloh and Frey (2015: 102), "*today rankings based on publications and citations dominate research governance. They serve as the basis for assessing the performance and impact of scholars, departments, and universities*."

Moed et al. (1985) stated that the proper use of bibliometric indicators can provide a "monitoring device" for a university. However, there might be some bias in the interpretation of the result. Indeed, years later, one of the co-authors suggested that the ranking of research institutions by bibliometric methods is an unsuitable tool for research performance evaluation (van Raan, 2005). Likewise, extensive research investigates how quantitative measures of research performances affect governance in academia and determine careers in universities.[6]

Widespread literature on the harmful consequences of bibliometric measures on the research profession exists. Weingart (2005, p. 118) defined them as "*theoretically unfounded, empirically crude and dependent on the data that we know to be imprecise*". Such types of metrics neither neutralize different customs of citing (e.g. article in biomedical research are cited six times more than the ones in mathematics) nor take into account the "*quantitative bias*" that could arise at the expense of the quality of research (e.g. the number of citations does not say everything about the quality of the paper that cited the considered one). Larivière and Gingran (2010) described this pattern recalling the "Matthew Effect": authors tend to read journals with a high impact factor and subsequently submit their articles in such journals, deflecting the attention to the quality of research they have previously read and simultaneously reducing the impact of a high-level publication in a low-prestige journal.

[6] An example might be found in Durante et al. (2011), which evaluated university performance based on familiarity by using the CIVR ("Steering Committee for the evaluation of research") score.

As Severin (2020) summarized, assessing the soundness and the originality of an article is crucial since only papers with some quality standards enter the scholarly record. In this way, low-quality papers cannot mislead policymakers and, recalling van Raan (2005), they do not potentially affect research evaluation.

However, also the review process has been in the spotlight. Smith (2006:178) criticized the common misperception that "*when something is peer-review is in some sense blessed*", shedding light on the bias of non-standardized procedures which results in the subjective evaluation of the reviewers. Seidl et al. (2005) empirically found that in some cases peer-review lacks impartiality, validity, and fairness, therefore an emerging body of research is testing the trustworthiness of the process in different journals. Langford and Guzdial (2015) showed the arbitrariness of the peer review process by an experiment conducted in 2014 by the organizers of the "Neural Information Processing Systems" conference. The organizers selected randomly the 10% of submitted manuscripts to the Conference and resubmitted them through the second round of review process. The result was that the conditional probability for an accepted submission to get rejected if examined by the second committee was equal to 60%! François (2015) considered this result as clear evidence for the randomness of the peer review, suggested some standardization methods.

The vulnerability of the review process has been also fostered by the exceeding growth rate of online journals[7] that can publish as much as they like, increasing the pressure and assigning papers to reviewers who are not experts in the area (Arns, 2014).

Our investigation is inspired by the aforementioned considerations and by the emergence of the selection of a trustworthy distinguishing method. This is of the utmost importance to reduce asymmetric information in scientific evaluation.

As discussed, judging the quality of a journal is not always a trivial task, this is particularly important for the so-called "Predatory Journals" which try to mask their low reputation level by asserting a consistent peer-review, but they only aim to earn money through the publication fee (i.e. Article Processing Charge). According to Beall (2016), they can be defined as journals (and publishers) that exploit the (gold) open-access model to profit from scholarly publishing dishonestly. In this sense they exploiting the open-access mechanism to maximized monetary gains, receiving funds directly from the authors who – in turn – aimed at improving the quantity of the research produced. Moreover,

[7] The annual number of articles indexed in the publisher Elsevier's Scopus database increased from around 1.2 million in 2000 to roughly 2.7 million in 2013. That is an increase of 113%, but some of this rise is simply due to articles from more journals being included in the later count (Arns 2014).

such journals pretend to use peer review for quality assurance, and spam academics with requests for submissions, to collect author payments (Bjork et al., 2019) and have an extremely high acceptance rate (Bjork 2019).

Shen and Bjork (2015: 1) stated that "*the problem of predatory open-access seems highly contained to just a few countries, where the academic evaluation practices strongly favor international publication, but without further quality checks*". On the same line, Xia et al. (2015) found that those who publish in these journals are young, inexperienced coming from developing countries. However, this phenomenon seems to be increasing also in other countries. Bagues et al. (2019) surveyed the impact that predatory journals have on the Italian academia, finding that some of these journals have managed to be included in citation indexes and some researchers took advantage of this to get a promotion, succeed especially when the committee (randomly selected) lack of expertise.

Different studies addressed their interest in evaluating the quality of their peer review. Baxt et al. (1998) used a fictitious manuscript to evaluate the peer-review performance founding that referees failed to check two-thirds of the major errors. Bohannon (2013) sent a bait-article to 304 journals, finding that more than half of the journals had accepted the paper, bringing into question the reliability of peer-reviewing. Sorokowski et al. (2017) criticized Bohannon's selection process because it did not include non-open-access journals, nor did it explicitly compare titles that did or did not have an impact factor. They re-designed the study to compare whitelist and blacklist journals, coming up with a similar result and highlighting a general tendency to capture some type of profit[8].

Following Bohannon (2013) approach, we aim to contribute to the current literature by providing a clear-cut normative approach to deal with predatory publishers. We have used the list of scientific journals employed by the Italian ANVUR to assess research production. Differently from the previous research, we have used the request of a publication fee as a distinctive factor, hence we have tested if the APC-funded model is a sufficient clue to infer the low reliability of journal peer-reviewing.

3 Experimental Design

3.1 Sample

[8] In their research, they created a profile of a fictitious scientist named Anna O. Szust and applied on her behalf to the editorial boards of 360 journals.

The ANVUR ranked 2714 international journals covering the Economic Area, namely Area 13[9], based on bibliometric indicators calculated until 2014. The sample selection consists of two steps. In the first step, we select the treatment group by screening the 2073 journals indexed in the scientific sub-categories of "Business"[10] or "Economics" by the Italian Research Quality Assessment exercise (2011-2014) (VQR, 2017) to pick those which require payment of APC.[11] We found that 52 journals match this condition, 30 of them are classified as "Business" and 22 in "Economics".

In the second step, we select the Control group by picking, from the VQR List, for each journal included in the Treatment group, a "*NoAPC*" (i.e., "Traditional") journal that: (1) belongs to the same area (i.e. "*Business*", "*Economics*"); (2) it is indexed in the same bibliometric database (ISI Web of Science and /or Scopus); (3) minimizes the difference between NoAPC- and APC-journal metrics (i.e. IF5, AIS, IPP, SJR, and the h-index). These bibliometric statistics of journals are computed by using a combination of the ISI Web of Science metrics (five-year impact factor - IF5; Article Influence Score AIS), Scopus metrics (Impact per Publication – IPP; SCImago Journal Rank - SJR) and h-index estimated by Google Scholar (53% of the sample) or by "Publish or Perish" software (47%) over the period 2010-2014. For non-ISI and non-Scopus journals, the metrics were imputed from the estimation of the correlation between IF5, AIS, IPP, SJR, and the h-index for the journal. These bibliometric data have been made public in the last release by ANVUR in February 2017.[12] Table 1 summarizes the bibliometric criteria used to define the units included in the Control Group.

[9], Area 13 is composed by 4 categories: Business (1217 journals), Economics (856 journals), General (3 journals), History (71 journals) and Statistics and Mathematical Methods (567 journals).

[10] Namely "Administration and Management" in the ANVUR list.

[11] This selection is both automatically and manually implemented. We create different loops in R to web scrape information from Journal Guide and DOAJ websites, such as publisher and APC, and to scan SCImago dataset, publishers' open access and pricing list, and Beall's list of predatory publishers. The residual unidentified reviews have been manually checked. For the sake of experimental design, we exclude Journals of the "Statistics" area.

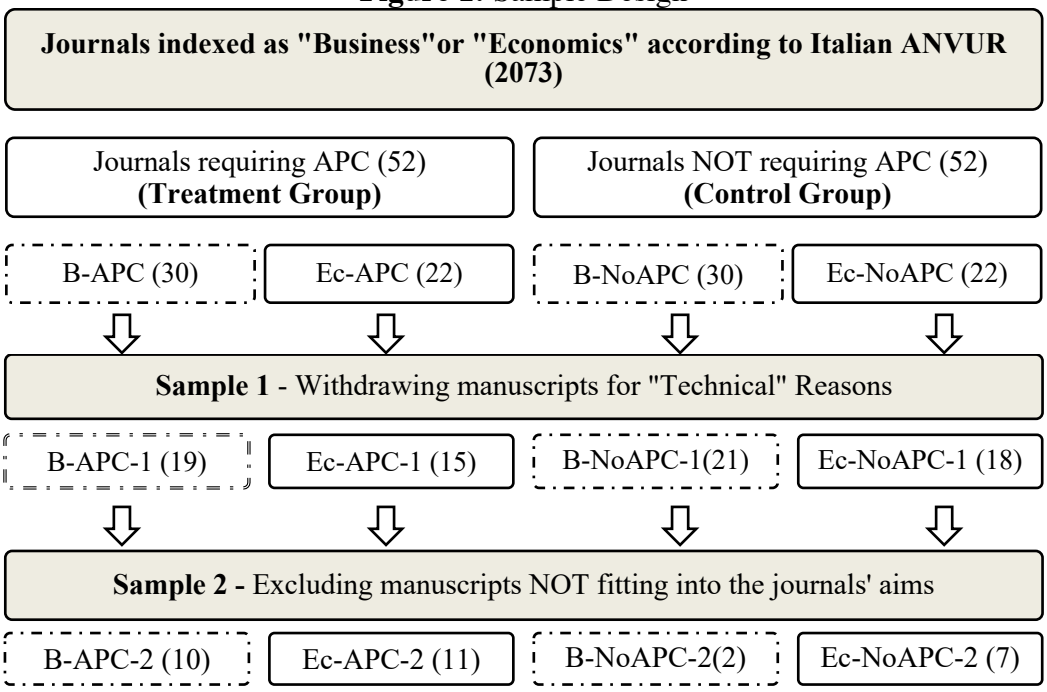
[12] Dataset and details on VQR exercise are available: http://www.anvur.it/attivita/vqr/vqr-2011-2014/gev/area-13-scienze-economiche-e-statistiche/

Table 1. Selecting Criteria for Treatment (APC) and Control (No APC) Groups

Variable	*Description*
ISI	Dummy for the presence in ISI WoS database (1=present, 0=not present)
Scopus	Dummy for the presence in Scopus database (1=present, 0=not present)
Repec	Dummy for the presence in Repec database (1=present, 0=not present)
Area	Scientific Area: Business (0) Economics (1)
IF5	5-year Impact Factor 2014. Source: ISI WoS.
AIS	Article Influence Score 2014. Source: ISI WoS.
IPP	Impact per Publication 2014. Source: Scopus.
SJR	SCImago Journal Rank 2014. Source: Scopus.
h-index	h-index over 2010-2014. Source: Google Scholar

Accordingly, we have identified 104 journals from the official database collected by the ANVUR for the last (second) Italian Research Quality Assessment exercise (2011-2014)[13].

Figure 1: Sample Design



Note: B = Business; Ec = Economics. Number of journals in parenthesis.

Once the theoretical sample has been defined, we identify the final dataset through additional screening processes which consist of excluding journals due to "technical" reasons. In particular, this sample (hereinafter Sample 1) is derived by dropping 31 journals for six types of impediments: (1) we exclude 15 journals because the submission process is managed by the same editorial office, therefore we submit the bait-manuscripts only to

[13] With regards to the ethical concerns of multiple submission, we adopted a well-established procedure in this strand of literature, see for instance Bohannon (2013), Martin and Martin (2016).

one journal for each editorial office; (2) for 7 cases, editor/editorial office sent back the manuscript because it was not following the format style guidelines of the journal[14]; (3) for 6 journals, submissions are no longer allowed; (4) for 1 case the submission was linked to the sending of personal documents of the submitter; (5) for 1 journal, the submission is possible only with editor invitation and (6) for 1 journal payment of the fee was required before submission. Accordingly, the empirical analysis in Sample 1 is based on 73 journals: 34 journals requiring APC and 39 journals not requiring APC (*No APC*).

A second test is based on a sample where we control for differences between the bait-manuscript's and the journals' topic (hereinafter Sample 2). Specifically, we drop from the Sample 1 those journals that, according to editors' emails sent to the "fake" submitters, reject the manuscript because it does not fit with the journal's aim. Sample 2 includes 30 journals: 21 of them requiring APC and 9 not requiring APC. Figure 1 summarizes the sample design.

Table 2 summarizes the Group's composition for Sample 1 and 2. We report p-values of tests checking if Treatment (APC) and Control (No APC) groups have equal means in terms of bibliometric scores. These tests confirm that there are no statistically significant differences between groups at a 5% level.

[14] Frequently, these requests of re-editing consisted in introducing new sections, inserting structured abstract, etc. If we had followed these editorial requests, we would have violated the hypothesis that all the Journals receive the same manuscript, therefore we decided to withdraw the manuscript from these journals.

Table 2. Descriptive statistics of Sample/Groups composition (Mean)

	T.S.ˣ	Sample 1			Sample 2		
Variable	***All***	***All***	***(Treat.) APC***	***(Contr.) No APC***	***All***	***(Treat.) APC***	***(Contr.) No APC***
Repec	**47.1%**	45.2%	35.3%	53.8%	53.3%	42.9%	77.8%
ISI	**5.77%**	5.5%	5.9%	5.1%	0.0%	0.0%	0.0%
Scopus	**26.0%**	27.4%	26.5%	28.2%	13.3%	19.0%	0.0%
IF5 *(Mean)*	**0.57**	0.65	0.71	0.59	0.54	0.62	0.34
*p-value**			0.42			0.097	
AIS *(Mean)*	***0.18***	*0.21*	0.24	0.19	*0.19*	0.22	0.12
*p-value**			0.28			0.097	
IPP *(Mean)*	***0.49***	*0.58*	0.57	0.59	*0.41*	0.48	0.26
*p-value**			0.88			0.14	
SJR *(Mean)*	***0.37***	*0.37*	0.35	0.40	*0.29*	0.31	0.25
*p-value**			0.37			0.42	
h-index *(Mean)*	***10.5***	*11.7*	12.41	11.13	*10.3*	11.4	7.78
*p-value**			0.50			0.15	
Age J. *(Mean)*	***15.2***	***16.1***	12.38	19.41	*9.86*	9.29	11.22
*p-value**			0.00			0.47	
Printed v. *(Mean)*	***0.63***	***0.62***	0.53	0.69	*0.67*	0.57	0.89
*p-value**			0.16			0.05	
DOAJ§	*26.9%*	*24.7%*	*50.0%*	*0.03%*	*23.3%*	*28.6%*	*11.1%*
COPE§	*79.8%*	*80.8%*	*61.8%*	*97.4%*	*66.6%*	*57.1%*	*88.9%*
Beall's List§	*17.3%*	*19.2%*	*38.2%*	*0.3%*	*33.3%*	*47.6%*	*0.0%*
# Journals	***104***	***73***	**34**	**39**	***30***	**21**	**9**

*Notes: ˣT.S.= Theoretical Sample; *p-value of two-sample t-test on equal means by assuming unequal variances (H_0);§ share of journals included in Directory of Open Access Journals (DOAJ); share of journals belong to Committee on Publication Ethics (COPE); share of journals whose publishers are included in the Beall's List.*

To check the equivalence between treatment and control groups to the presence of outliers, we also perform tests on medians. In particular Table 3 reports medians of group compositions and p-values of quantile regressions to test for the equality of medians between APC and No APC groups. With the exclusion of "*Age Journal*", all the tests fail to reject the null hypothesis that there is no difference between medians of treatment and control groups. Accordingly, Tables 2 and 3 validate the hypotheses that means and medians are not statistically different between APC and No APC groups.[15]

[15] We also test the hypothesis that two independent samples (APC and No APC) are from populations with the same distribution by using the Wilcoxon rank-sum test. These results - omitted for the sake of brevity - confirm the equivalence of distributions between samples.

Table 3. Descriptive statistics of Sample/Groups composition (Medians)

	T.S.×	Sample 1			Sample 2		
Variable	***All***	***All***	***(Treat.) APC***	***(Contr.) No APC***	***All***	***(Treat) APC***	***(Contr.) No APC***
Repec *(Median)*	**No**	No	No	Incl.	Incl.	No	Incl.
ISI *(Median)*	**No**	No	No	No	No	No	No
Scopus *(Median)*	**No**	No	No	No	No	No	No
IF5 *(Median)*	**0.43**	0.49	0.48	0.48	0.36	0.36	0.32
*p-value**			1.00			0.86	
AIS *(Median)*	***0.12***	*0.14*	0.15	0.14	*0.09*	0.08	0.12
*p-value**			0.50			0.79	
IPP *(Median)*	***0.32***	*0.34*	0.34	0.34	*0.22*	0.20	0.25
*p-value**			1.00			0.79	
SJR *(Median)*	***0.36***	*0.35*	0.35	0.35	*0.28*	0.30	0.25
*p-value**			0.93			0.60	
h-index *(Median)*	***9***	*10*	10	10	*7.5*	7	8
*p-value**			1.00			0.81	
Age J. *(Median)*	***11***	***14***	10	16	*9*	9	9
*p-value**			0.04			1.00	
Printed v. *(Median)*	***Yes***	***Yes***	No	No	*Yes*	Yes	Yes
*p-value**			1.00			1.00	
DOAJ[§] *(Median)*	*No*	*No*	*No*	*No*	*No*	*No*	*No*
COPE[§] *(Median)*	*Incl.*	*Incl.*	*Incl.*	*Incl.*	*Incl.*	*Incl.*	*Incl.*
Beall's List[§] *(Median)*	*No*	*No*	*No*	*No*	*No*	*No*	*No*
# Journals	***104***	***73***	**34**	**39**	***30***	**21**	**9**

Notes: See Table 2

3.2 The bait-manuscript

The bait-manuscript submitted to a sample of journals follows Bohannon's (2013) scheme. The manuscript had a credible layout for an economic journal article. It included 5 sections (Introduction, Literature, Methodology, Results and Discussion, Conclusions), and a data appendix. This fake manuscript has been written in Italian and automatically translated to English using Google Translator[16] (see the appendix 3 to read the submitted manuscript). In addition to the mistranslations we had also added an inconsistent and erroneous use of decimal separator, i.e. we use the comma to separate the integer part from the fractional part of numbers reported in main text and tables, while we use the point for the numbers reported in the appendixes. This kind of sloppiness is aimed to detect if there are differences in language checking between APC and not APC journals.

In terms of content, we submitted practically identical papers to two groups of journals grouped in two partially overlapping areas of economic research (i.e. Business and Economics). The only difference was the submitter's name. Indeed, we have used two fake identities, to make it easier to manage these multiple communications between the "fake

[16] https://translate.google.com/

author" and the editors. The name chosen was the translation of "*Misunderstood Genius"* in Welsh and Haitian Creole languages. For the affiliations, we have combined the generic "*National University of*" with the capital cities of the Welsh and Haitian institutions.[17] In particular, the submitter was the Assistant Prof. Camddeall Athrylith - Department of Economics, Management and Statistics, National University of Aberystwth, Wales for the sample of "*Administrative Sciences and Management*" journals, and the Assistant Prof. Jeni Konpreyansyon - Department of Economics, Management and Statistics, National University of Port-au Prince, Haiti for the "*Economics*" journals.

The title of bait-manuscript is: "Crime and Economic Growth. An empirical analysis for Germany". The abstract of the submitted papers is as follows: "*This paper examines the role of crime, enforcement, and taxation on economic growth. These effects are studied by a modified version of an endogenous growth model proposed by Loayza (1996). Econometric results are based on the German economy over the period 1992-2016. Empirical evidence confirms the theoretical model. We show that the relationship between crime and growth rate of GDP is negative in the long-run equilibrium.*"

The manuscript counted, at least, 71 errors that a skilled peer reviewer should identify and suggest the editor reject the manuscript. In particular, in addition to the slapdash quality of English, there were: 3 wrong first derivatives; 9 comments completely reversed compared to the empirical results reported in the table of the manuscript; 16 erroneous interpretations of statistical significance of reported t-students (on 45 estimated coefficients); incorrect interpretations of Durbin Watson statistics, Cointegration and Unit Roots, as a consequence of the wrong interpretation the transformations applied to the variables included in the model were inappropriate; the conclusions and policy implications were in contrast with the empirical outcomes; some data sources and code of variables were fictitious; some descriptive statistics were false (e.g., we report some means which were lower than the minimum values, or larger than maximum values of the variables). Some examples of these errors are shown in Figure 2.

[17] As Bohannon (2013: 62), we used authors and affiliations from developing countries because "*authors and institutions would arouse less suspicion if a curious editor were to find nothing about them on the Internet*."

Figure 2. Frames extrapolated from the bait-article

***Frame 2.a*: Examples of errors in analytical derivations (e.g., these two derivatives should be positive).**

(5) $$\eta = p\lambda(q-\lambda)$$

where, by imposing the constraint on the quality of the institutions $q > 2\lambda$, the partial derivatives will be obtained that guarantee the desired functional characteristics: $\frac{\partial \eta}{\partial \lambda} > 0$; $\frac{\partial \eta}{\partial p} > 0$; $\frac{\partial \eta}{\partial q} < 0$; $\frac{\partial^2 \eta}{\partial \lambda \partial q} > 0$; $\frac{\partial^2 \eta}{\partial p \partial q} < 0$; $\frac{\partial^2 \eta}{\partial \lambda^2} < 0$. Indicating with S the relative size of the legal economy compared to the total one

Frame 2.b: Examples of errors in the interpretation of empirical results

Table 1 - dependent Variable: growth rate of GDP per capita

	I	II	III	IV	V	VI	VII
Constant	-0,214 (-0,411)	-1,349** (-2,584)	-0,787** (-2,563)	-0,806 (-1,552)	-	-0,227 (-0,724)	0,242 (0,808)
Tax burden (τ_1)	-0,240 (-0,788)	-0,395 (-1,464)	-0,481 (-1,595)	-			
Tax burden^2 (τ_1)^2	-	-	0,225 (1,660)	-			
Pers. Inc.T. burden. (τ			-	-0,533* (-1,902			
Pers. Inc.T. burden.^2 (τ			-	-0,604* (-3,169)	(-5,200)	(-3,067)	(-2,191)
Criminal Economy (EC)	0,202*** (2,176)	0,158** (2,159)	0,090 (1,299)	0,212* (2,383)	0,218** (2,218)	0,206** (3,491)	0,197* (1,407)
Index of *enforcement* (λ)	-	-1,525** (-1,867)	-1,107*** (-2,105)	-0,878* (-1,693)	-0,489* (-1,691)	-0,666** (-1,700)	-
Control Var.							
Internat. Trade Openness	0,285*	0,360**	0,338*** 3,054)				
Tertiary Education	(0,098)	(1,945)	-				
workforce participation rate	0,304 (0,632)	-0,152 (0,707)	-	(0,714)	(0658)	-	-
Observations	24	24	24	24	24	24	24
R^2-adjustd	0,190	0,604	0,623	0,665	0,631	0,685	0,253
Durbin-Watson st.						0,498	0,357
LM Test serial-correl.[a]						0,878	0,693
BPG Test Heterosched.[b]	0,726	0,998	0,997	0,978	0,603	0,673	0,979
JB Test Normality[c]	0,648	0,385	0,667	0,778	0,744	0,560	0,652

This is the main result of the paper. In the Abstract, discussion and Conclusion we state that the "*relationship between crime and growth rate of GDP is negative*". At the contrary, the empirical result validates a positive correlation.

According to t-tests these coefficients are not statistically significant

They are significant at 10%

The second conclusion of the paper is that the "*. The econometric analysis showed a positive correlation between the two variables*"[enforcement and GDP growth]. At the contrary, the empirical result shows a not statistically significant coefficient at 5%,(or negative at 10%).

Tertiary education counts 22 observations (as reported in the appendix) therefore these regressions cannot include 24

Table 1 - dependent Variable: growth rate of GDP per capita

	I	II	III	IV	V	VI	VII
Constant	-0,214 (-0,411)	-1,349** (-2,584)	-0,787** (-2,563)	-0,806 (-1,552)	-	-0,227 (-0,724)	0,242 (0,808)
Tax burden (τ_1)	-0,240 (-0,788)	-0,395 (-1,464)	-0,481 (-1,595)	-			
Tax burden^2 $(\tau_1)^2$	-	-	0,225 (1,660)	-			
Pers. Inc.T. burden. (τ			-	-0,533* (-1,902			
Pers. Inc.T. burden.^2 (τ			-	-0,604* (-3,169)	(-5,200)	(-3,067)	(-2,191)
Criminal Economy (EC)	0,202*** (2,176)	0,158** (2,159)	0,090 (1,299)	0,212* (2,383)	0,218** (2,218)	0,206** (3,491)	0,197* (1,407)
Index of *enforcement* (λ)	-	-1,525** (-1,867)	-1,107*** (-2,105)	-0,878* (-1,693)	-0,489* (-1,691)	-0,666** (-1,700)	-
Control Var.							
Internat. Trade Openness	0,285* (1,939)	0,360** (2,673)	0,338*** (3,054)				
Tertiary Education	(0,098)	(1,945)	-				
workforce participation rate	0,304 (0,632)	-0,152 (0,707)	-	(0,714)	(0658)	-	-
Observations	24	24	24	24	24	24	24
R^2-adjustd	0,190	0,604	0,623	0,665	0,631	0,685	0,253
Durbin-Watson st.						0,498	0,357
LM Test serial-correl.[a]						0,878	0,693
BPG Test Heterosched.[b]	0,726	0,998	0,997	0,978	0,603	0,673	0,979
JB Test Normality[c]	0,648	0,385	0,667	0,778	0,744	0,560	0,652

According to t-tests these coefficients are not statistically significant

This is the main result of the paper. In the Abstract, discussion and Conclusion we state that the "*relationship between crime and growth rate of GDP is negative".* At the contrary, the empirical result validates a positive correlation.

They are significant at 10%

The second conclusion of the paper is that the ". *The econometric analysis showed a positive correlation between the two variables"*[enforcement and GDP growth]. At the contrary, the empirical result shows a not statistically significant coefficient at 5%,(or negative at 10%).

Tertiary education counts 22 observations (as reported in the appendix) therefore these regressions cannot include 24

Frame 2.c: Examples of errors in references

Banerjee A.V. - Duflo E., «Growth Theory through the Lens of Development Economics», in Durlauf S. - Aghion P. *Handbook of Economic Growth*, Holland: Elsevier Science, 2015, pp. 473–552.

Barro R.J. - Sala-i-Martin X., *Economic Growth*, New York: McGraw-Hill, 1995.

Dell'Anno R. - Morone A., «The criminal economy and underdevelopment», *Economic Systems*, vol. 58, n. 3, 2018, pp. 257-279.

Cimoli M. - Primi A. - Pugno M., «A Low-Growth Model: Illegality as a Structural Constraint», *Cepal Review*, vol. 88, 2006, pp. 85-102.

de Soto H., *The Other Path: The Invisible Revolutio* Publishers, 1989.

Caferra. «Consumption Growth and Crime: Some Evidence from Income Quintile Groups in Europe», *Applied Economics Letters*, vol. 13, 2017, pp. 529–532.

Engle R.F. - Granger C.W.J., «Co-integration and Error Correction: Representation, Estimation, and Testing», *Econometrica*, 55, 1987, pp. 251–276.

Friedman E. - Johnson S. - Kaufmann D. - Zoido-Lobaton P., «Dodging the grabbing hand: the determinants of unofficial activity in 69 countries», *Journal of Public Economics*, vol. 76, n. 3, 2000, pp. 459-493.

Johansen S. «Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models», *Econometrica*, 59, 1994, pp. 1551–1580.

Johansen S., *Likelihood-based Inference in Cointegrated Vector Autoregressive Models* Oxford: Oxford University Press, 1998.

Johnson S. - Kaufmann D. - McMill after communism», *Journ*

Jehle Jörg-Martin. Criminal Justice in Germany: Facts and Figures, Federal Ministry of Justice and Consumer Protection, Eight Edition 2017.

Some references and years are invented

This reference is the main data source and it does not exists. There is a 5th edition (published in 2013) and it ran until 2013, so we invented the data from 2013 to 2016.

Frame 2.d. Examples of errors in Database

	Sources	Means	MAX	MIN	# obs
growth rate of GDP per capita (%)	World Bank – "GDP per capita growth (annual %)". World Development Indicators Online. Washington, DC. Codice: NY.GDP.PCAP.ZK.GD	1.6	4.1	-1.0	24 ('92-'16)
Tax burden (%) (τ_1)	Eurostat - tax revenue (including social contributions) in % of GDP, [gov_10a_taxag]	39.0	43.7	31.1	24 ('92-'16)
Pers.Inc. Tax/GDP (%) (τ_2)	Eurostat - Tax on personal income in % of GDP [gov_11a_taxag]	13.5	15.8	9.3	24 ('92-'16)
Index of *enforcement* (λ)	Statistisches Bundesamt, (Destatis). Strafrechtliche Statistiken. (several years): Tab. 10.1 (1992, 1993); Tab. 3.1 (1994-2016) – numero persone denunciate per i quali l'Autorità giudiziaria ha iniziato l'azione penale / 100.000. La serie è trasformata in logaritmi.	5.6		5.7	24 ('92-'16)
Criminal economy/GDP (%)	Jehle (2017). Reddito criminale prodotto in % of GDP (Tab. 1, pag. 11).	38.3	37.1	26.9	24 ('92-'16)
International Trade Openness (%)	World Bank – "Merchandise trade as a share of GDP". World Development Indicators Online. Washington, DC. Codice: IMP.EXP.TR	37.4	47.8	[illegible]	24 ('92-'16)
Tertiary Eduction (%)	World Bank – "School enrollment, tertiary (% gross). World Development Indicators Online. Washington, DC. Codice: SE.TER.ENRR (manca il valore per il 1998)	40.8	67.1	24.8	22 ('92-'16)
Workforce participation rate (%)	World Bank – "Labor participation rate, total (% of total population ages 15+). World Development Indicators Online. Washington, DC. Codice: SL.TTF.CA.ZS	48.9	40.4	49.9	24 ('92-'16

This code does not exists. The right code is: NY.GDP.PCAP.KD.ZG

This variable and code do not exist in Eurostat database

This variable does not exist. The text is not in English, but German and Italian

MEAN<MIN

MEAN>MAX

These variables and codes do not exist.

3.3 Results

In January 2019, each of two alter-ego of Assistant Prof. *Misunderstood Genius* sent his bogus manuscript to 104 journals, due to "technical reasons" the original sample size reduces to 73 journals by defining Sample 1. We disentangle the journal's evaluation process in three steps:

- the first step in which editor and/or reviewers evaluate whether the manuscript sufficiently suits the journal's aim. In favorable cases, these journals are included in Sample 2.
- The second round in which editors and/or referees evaluate the scientific significance of the submitted manuscript.
- The third round in which editors and/or reviewers who had required revisions, evaluate the same paper without any further revision. Indeed, after few weeks from revision request, *Misunderstood Genius* re-submitted the original version without any changes but, in his cover letter for the editor and reviewers, he stated: *"Dear Editor, please find attached the revised manuscript. Sincerely [author's name]"*.

Figure 3 shows the outcomes of this analysis of Sample 1 and Sample 2.[18]

[18] Appendix 1 reports outcomes separating them in scientific areas. These results confirm the general finding in Figures 3 and 4. Appendix 2 lists journals and their most relevant statistics for this empirical exercise.

Figure 3: Results based on Sample 1 - (APC Vs NoAPC)

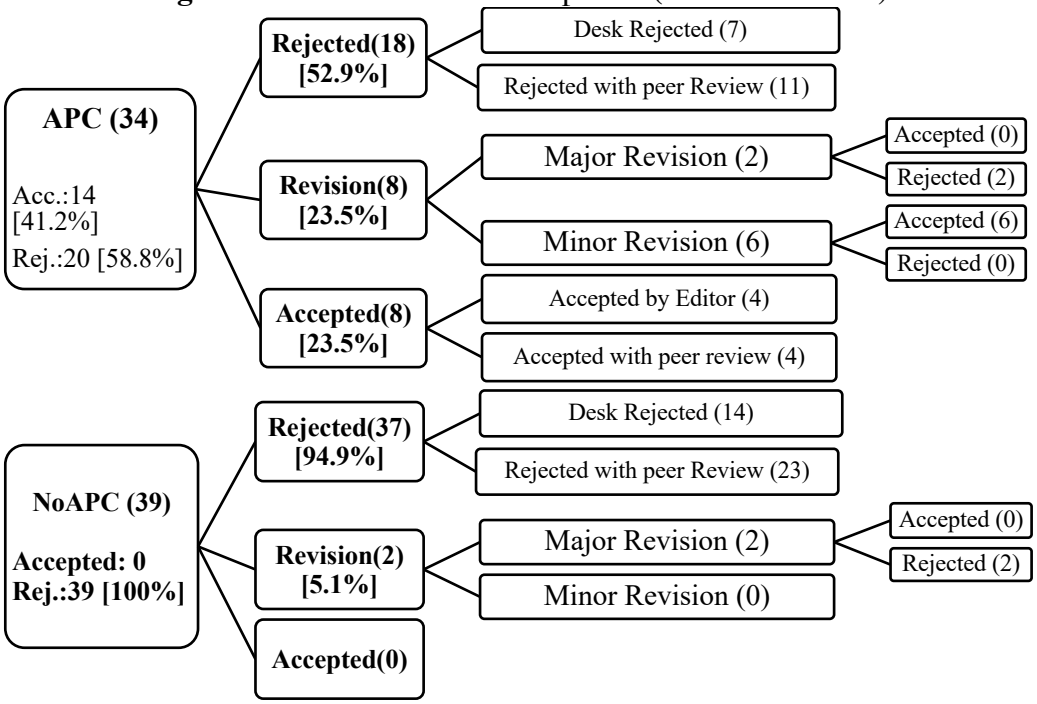


Figure 4: Results based on Sample 2 (APC *vs*. NoAPC)

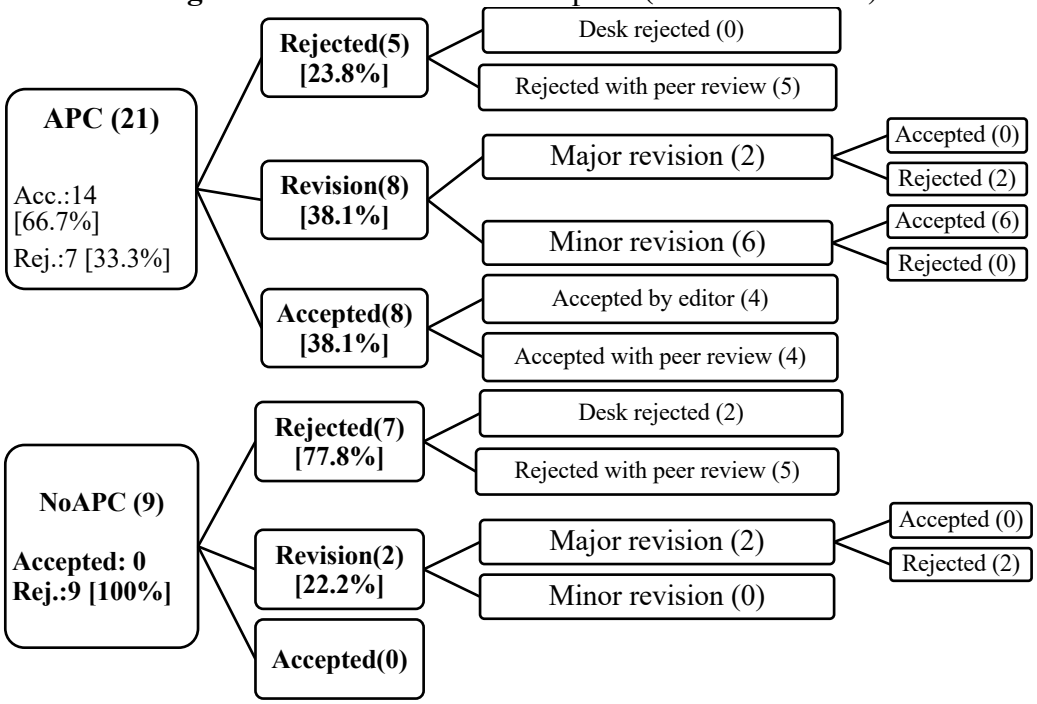
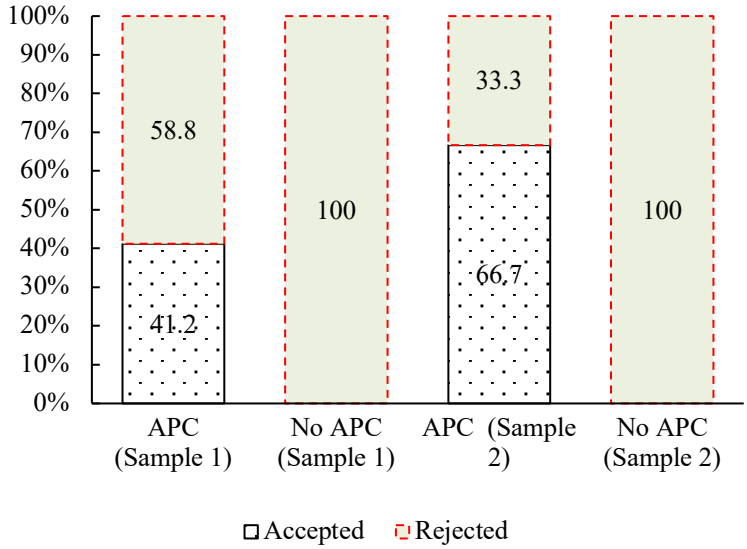


Figure 3 and Table 4 highlight as our quasi-natural experiment validates the hypothesis that publication fees (APC) are a sufficient condition to deeply change the relationship among authors and editors. For the *Traditional model (No APC)*, an effective peer review process to assess the quality of the submitted manuscript exists, this incentive disappears for *APC* journals. In particular, APC journals exhibit a significantly higher acceptance

rate (41.2% in Sample 1 and 66.7% in Sample 2) than No APC (0% in both Samples), marking a clear line between treatment and control groups (see Figure 5).

Figure 5. Acceptance/Rejection rates comparison between APC and No APC.



3.4 Statistical Analysis

In this section, we analyze by statistical methods if the previous results provide sufficient evidence to conclude that the difference in acceptance rate between APC and No APC is statistically significant and does not depend on insufficient sample size and unbalanced sample.

The validity of this analysis has three potential drawbacks: (1) the experimental design is not based on a random sampling method. On one hand, this approach allows us to have more comparable units between treatment and control groups of journals, on the other, it may reduce the meaning of power tests on the minimum sample size that assumes random sampling method. (2) The effective sample size is limited (73 or 30 observations). (3) If from a theoretical viewpoint the evidence that no control group journal has accepted the bait manuscript intuitively supports the key hypothesis of this research, from a statistical viewpoint, it generates a problem of perfect prediction in outcome model.[19]

[19] Specifically, given that our outcome variable is binary when a journal accepts the manuscript, it separates the control and treatment group perfectly. Consequently, we cannot estimate by a Probit/Logistic regression the statistical significance of APC on acceptance rate, given that we observe acceptance only in the case of APC journals.

Taking into account the previous caveats, in the following, we will show as all the applied tests converge on the same results that journals with APC accept fake-manuscript significantly more than "traditional" journals.

As a first preliminary step, we assess the validity of our findings, by conducting a binomial power test to verify the minimum sample size for each group attainable through our data given a significative level of 5% and power of 90% (Table 4).

Table 4 Power test on the Minimum sample size.

	Sample 1				Sample 2			
	N.		Acceptance r.		N.		Acceptance r.	
APC (Treatment Group)	34		0.4118		21		0.6667	
No APC (Control Group)	39		0.0000		9		0.0000	
significative level (α)	0.05				0.05			
power (1-β)	0.80	0.90	0.80	0.90	0.80	0.90	0.80	0.90
Alternative	One-sided		Two-sides		One-sided		Two-sides	
Minimum N for each Group	11	15	14	19	6	7	7	9

Bearing in mind the first caveat on the lack of random assignment in the quasi-experimental design method, Table 4 indicates that both the samples have a sufficient size.

In a second step, we perform a test on the difference between acceptance rates between control and treatment groups. Table 5 reports statistical tests on the hypothesis that the acceptance rates are equal in two samples (APC Vs No APC).

Table 5. Tests on the Difference between Acceptance rates

	Obs.	Sample 1	Obs.	Sample 2
APC (Treatment Group)	34	0.4118 (0.086)	21	0.6667 (0.105)
No APC (Control Group)	39	0.0000 (0.000)	9	0.0000 (0.000)
H_0: APC = No APC (p-value)		0.000		0.000
H_1: APC ≠ No APC (p-value)		1.000		1.000
H_1: APC < No APC (p-value)		1.000		1.000
t-stat on difference between groups		-4.806		-6.325
Welch's degrees of freedom		33		20

Note: Standard Error in parenthesis. Unequal variances between groups are assumed.

According to Table 5, we cannot reject the hypothesis of a statistically significant difference in the acceptance rates between the two groups.

The third step of our analysis is aimed to apply more advanced techniques to compare the two groups making use of matching estimators (see Stuart, 2010). The rationale of this extension is that, if there are covariates related to treatment assignments that affect the potential outcomes then the t-tests on the difference between means of two groups may be misleading because the missing data on the potential outcomes are informative.

Accordingly, we look for covariates that are related both to the potential outcomes and the treatment.

In particular, we control for three (not bibliometric) variables[20]: the location of the publisher - because it may have an impact on the type of reviewers a journal can attract (Shen and Bjork, 2015) (Table 6); the age of the journals - because it may be correlated with journal reputation and status (Table 7) and whether the journals have a paper as well as an online version or publish online-only (Table 7).

Table 6: Distribution of the location of the publisher (Sample 1)

Country Group	No APC	APC	Rejected	Accepted	Total
Western Europe 1 (Gbr)	18 (24.7%)	8 (11.0%)	26 (35.6%)	0 (0.0%)	26 (35.6%)
Western Europe 2 (Deu; Che)	13 (17.8%)	5 (6.8%)	18 (24.7%)	0 (0.0%)	18 (24.7%)
Western Europe 3 (Ita; Nld)	2 (2.7%)	2 (2.7%)	4 (5.5%)	0 (0.0%)	4 (5.5%)
North America (Can; Usa)	5 (6.8%)	7 (9.6%)	6 (8.2%)	6 (8.2%)	12 (16.4%)
Eastern Europe (Rom; Ukr) & Western Asia (Tur)	0 (0.0%)	4 (5.5%)	2 (2.7%)	2 (2.7%)	4 (5.5%)
Africa (Nga)	0 (0.0%)	3 (4.1%)	2 (2.7%)	1 (1.4%)	3 (4.1%)
Southern Asia (Ind)	0 (0.0%)	2 (2.7%)	0 (0.0%)	2 (2.7%)	2 (2.7%)
Eastern Asia (Chn)	0 (0.0%)	2 (2.7%)	0 (0.0%)	2 (2.7%)	2 (2.7%)
South-Eastern (Sgp)	1 (1.4%)	0 (0.0%)	1 (1.4%)	0 (0.0%)	1 (1.4%)
Pacific (Nzl)	0 (0.0%)	1 (1.4%)	0 (0.0%)	1 (1.4%)	1 (1.4%)
Total	**39**	**34**	**59**	**14**	**73**

Table 6 provides evidence that, although publishers requiring publication fees are for the most part located in more economically developed countries (82.2% of 34 APC journals)[21], the predatory publishers - with the relevant exclusion of Canada[22] - are predominantly located in more economically developed countries (57% of 14 accepted manuscripts).

[20] We thank an anonymous referee for this suggestion.

[21] United Kingdom (26), Germany (16), Canada (7), USA (5) other Western European countries (6).

[22] Specifically, six APC journals published by the Canadian Center of Science and Education accepted the bait manuscript, followed by China (2), India (2), New Zealand (1), Nigeria (1), Romania (1), Ukraine (1).

Table 7: Distribution of Age of Journal (Samples 1, 2)

Journal	3-10	11-20	21-30	31-40	40-54	Total
APC	18 (24.7%)	13 (17.8%)	2 (2.7%)	1 (1.4%)	0 (0.0%)	34 (46.6%)
No APC	10 (13.7%)	14 (19.2%)	7 (9.6%)	5 (6.8%)	3 (4.1%)	39 (53.4%)
Accepted	9 (12.3%)	5 (6.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	14 (19.2%)
Rejected	19 (26.0%)	22 (30.1%)	9 (12.3%)	6 (8.2%)	3 (4.1%)	59 (80.8%)
Printed	17 (23.3%)	15 (20.5%)	8 (11.0%)	4 (5.5%)	1 (1.4%)	45 (61.6%)
Only online	11 (15.1%)	12 (16.4%)	1 (1.4%)	2 (2.7%)	2 (2.7%)	28 (38.4%)
Total Samp.1	**28 (38.4%)**	**27 (37.0%)**	**9 (12.3%)**	**6 (8.2%)**	**3 (4.1%)**	**73 (100%)**
APC	15 (50.0%)	6 (20.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	21 (70.0%)
No APC	5 (16.7%)	2 (6.7%)	1 (3.3%)	1 (3.3%)	0 (0.0%)	9 (30.0%)
Accepted	9 (30.0%)	5 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	14 (46.7%)
Rejected	11 (36.7%)	3 (10.0%)	1 (3.3%)	1 (3.3%)	1 (3.3%)	17 (56.7%)
Printed	13 (43.3%)	5 (16.7%)	1 (3.3%)	1 (3.3%)	1 (3.3%)	21 (70.0%)
Only online	7 (23.3%)	3 (10.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	10 (33.3%)
Total Samp.2	**20 (66.7%)**	**8 (26.7%)**	**1 (3.3%)**	**1 (3.3%)**	**0 (0.0%)**	**30 (100%)**

Table 7 shows that APC journals usually have a lower age than traditional journals. The positively skewed distribution is also observed if we only consider predatory journals (i.e., the mean age of journals which accepted the bait-manuscript is 9.86 years old, while the oldest predatory journal is 16 years old). Similarly, right-skewed distribution emerges for printed versus only online journal issues,

The fourth step of our analysis consists of applying matching estimators. These estimators are based on the idea of comparing the outcomes of journals that are as similar as possible with the sole exception of their treatment status.

By using a potential-outcomes framework, we define as y_0 the observed editor's decision for submission to "No APC" Journal (i.e., control group) and y_1 the potential editorial decision for the same journal if it were in the treatment group. With this framework, the difference between the observed outcome y_0 and the imputed potential outcome y_1 is an estimate of the individual-level treatment effect.

In the following, we report both the mean of the difference between y_1 and y_0 (Average Treatment Effect - ATE) and the mean of the difference between y_1 and y_0 among the journals that require APC (i.e. Average Treatment Effect on the Treated - ATET).[23] We apply treatment-effect estimators based on Nearest-neighbor matching (NNM) and the propensity-score matching (PSM).

The NNM determines the "nearest" journal by using a weighted function of the covariates for each submission. Specifically, we apply a 1:1 nearest neighbor matching where we select for each APC journal, the No APC journal with the smallest distance measured by

[23] The estimates of ATET require a less restrictive assumption of "*strong ignorability*" than to estimate the ATE (see Wooldridge 2010). Although ATET and ATE estimates converge on the same result, we consider ATET findings more reliable for our model.

the Mahalanobis method. This type of matching is nonparametric because no explicit functional form for either the outcome model or the treatment model is specified. However, the cost of this flexibility is that we need a larger sample size to get to the true value than an estimator that imposes a functional form. Given our small dataset, we suggest considering NNM findings cautiously.

The second matching approach is PSM. It determines the "nearest" by combining all the covariate information into estimated treatment probabilities - known as propensity scores - and uses this single continuous covariate as the matching variable.

Table 8 reports the output of these analyses based on sample 1.

Table 8. Nearest-Neighbor and Propensity Score Matching Est. (Sample 1)

	Nearest-Neighbor M.		**Propensity Score M.**	
	ATE	ATET	ATE	ATET
APC Vs No APC (coeff.)	0.213	0.412	0.219	0.412
(p-value)	(0.031)	(0.000)	(0.000)	(0.000)
Treated obs (Raw/ Matched)	34/73	34/34	34/73	34/34
Control obs (Raw/ Matched)	39/73	39/34	39/73	39/34
Number obs (Raw/ Matched)	73/146	73/68	73/146	73/68

Note: For NNM we require exact matching on the binary variables of scientific area ("Business", "Economics")

Following Rubin's (2008) recommendation, we examine whether the treatment model balanced the covariates with standardized differences of means and variance ratios. Tables 9 and 10 report estimated margins based on Probit regression.[24]

Table 9. Nearest-Neighbor Matching Diagnostic (Sample 1)

Outcome model	Margins	Raw data*		ATE*		ATET*	
y = Rejected/Accepted	dy/dx (p-val)	Std Diff Mean	Variance ratio	Std Diff Mean	Variance ratio	Std Diff Mean	Variance ratio
Repec	-.213 (.000)	-.375	.923	-.027	.994	.000	1.00
Scopus	-.143 (.004)	-.038	.965	-.031	.968	.000	1.00
SJR	-.634 (.002)	-.212	.789	-.088	.756	-.033	1.57
h-index	.016 (.007)	.161	2.14	.112	1.38	.231	2.56
Age Journal	-.033 (.000)	-.781	.277	-.564	.374	-.348	1.29
D_Pub_4(Can;Usa)	.329 (.000)	.206	1.47	.079	1.18	.238	1.58
D_Pub_5(Rom; Ukr; Tur)	.229 (.106)	.509	.000	.338	.000	.509	.000
D_Pub_6(Nga)	.400 (.000)	.433	.000	.290	.000	.433	.000
Pseudo R^2	.535						
Chi^2(p-value)	17.9 (.021)						
Obs	73						

[24] We estimate several alternative specifications of outcome and treatment model, including non-linear and polynomials in the covariates. However, given those specifications methods require larger datasets that we have, we lean towards specifications that are more parsimonious rely on having enough data.

Note: [*]A perfectly balanced covariate has a standardized difference of zero and variance ratio of one.

Table 10. Propensity-Score Matching Diagnostic (Sample 1)

Treatment model	Margins	Raw data		ATE		ATET	
y = No APC/APC	dy/dx (p-val)	Std Diff Mean	Vari-ance ratio	Std Diff Mean	Vari-ance ratio	Std Diff Mean	Vari-ance ratio
Scopus	-.159 (.074)	-.038	.965	-.171	.769	.211	1.34
SJR	-.445 (.011)	-.212	.789	-.179	.765	-.145	4.25
h-index	.011 (.056)	.161	2.14	-.312	2.44	.017	11.2
COPE	-.259(.102)	-.972	9.48	.129	.848	.294	.948
Age Journal	-.014 (.002)	-.781	.277	-.197	.587	-.217	1.14
Printed issues	-.200 (.033)	-.334	1.17	.029	.980	-.334	1.05
D_Pub_1(Gbr)	-.393 (.014)	.206	1.47	.140	1.08	-.482	1.10
D_Pub_2(Deu; Che)	-.451(.004)	.509	.000	-.176	.748	-.441	1.56
D_Pub_3(Ita; Nld)	-.454 (.047)	.433	.000	-.155	.597	.033	.441
D_Pub_4(Can; Usa)	-.202 (.269)	.433	.000	.175	1.57	.206	5.73
Pseudo R^2	.411						
Chi^2(p-value)	27.0 (.003)						
Obs	73						

See note of table 9

The output of Tables 9 and 10 indicate that NNM and PSM improve the level of balance of covariates with the exclusion of the distributions of the h-index and dummy for journals that are located in North America). In general, these diagnostic tests show that covariates are balanced in both raw and matched data.

In conclusion, all our statistical tests suggest that we cannot reject the hypothesis of a statistically significant difference in the acceptance rates between APC and No APC journals (Tables 5 and 8). This implies that there is sufficient evidence to state that a significant share of APC journals in the economic area (41% in Sample 1 and 67% in Sample 2) accept papers without making a proper and accurate peer review.

4 Conclusion

This research aims to analyze how the APC business model, independently from the Open Access nature of journals, biases the review process extraordinarily rising the acceptance rate.

To carry out this research we submitted a bait-manuscript to 73 (out of 104) academic economic journals to test if there was a difference in the peer-review process between a Treatment Group (34 "*APC-charging journals*) and a Control Group *(39 Traditional journals* i.e. that do not require payment of publication fee).

The submitted bait-article was based on completely invented data, with evident errors in terms of methodology, literature, reporting of results, quality of language, policy im-

plications, and some fake references. Nevertheless, 41.2% (Sample 1) and 66.7% (Sample 2) of APC journals fell into the trap: the editors accepted the manuscript and, required us to pay the fee to proceed to publish the article in their journals.

Accordingly, the peer review process instead of working as a mechanism aimed to recognize quality, preventing plagiarism, expanding knowledge and promoting innovative research[25] becomes an empty word recalled by predatory journals to catch, on the one hand, inexperienced (or unscrupulous) scholars and, on the other hand, to fulfill the standard requirements to be indexed on some of the most relevant databases of peer-reviewed publications (i.e. Scopus). In particular, we have found that two journals indexed by Scopus accepted the bait-article, while no journals included in the ISI - Web of Science fell into the trap. For that reason, we agree with Wang and Waltman's (2016) result that Web of Science performs better than Scopus in terms of accuracy.[26] Furthermore, this research has also revealed that predatory behavior conducts may occur also for journals included in Directory of Open Access Journals (1 case); journals which belong to Committee on Publication Ethics (7 cases) and for publishers that are not included in the Beall's List (6 journals).

We find that the APC business model has shown its inappropriateness to preserve the standard of scientific publishing because the conflict of interests between journals and authors to publish poor research studies disappears. Accordingly, due to the great consequences of bibliometric scores in determining career paths of researchers and performances of academic institutions, we conclude that relying exclusively on the editors' and authors' ethics, is not an effective method to preserve opportunistic behaviors.

We conclude that the *Traditional model* has a more effective incentive-mechanism to select based on quality standards. This model can adequately solve the publishers' commercial goal from the scientific community's aim to distinguish between good and poor quality of scientific studies. This is made possible because traditional journals, through effectual peer-review, tend to publish only accurate articles because they attract more journal citations, these quotations increase Impact Factors (IFs) and, in turn, increase revenue from subscriptions and attentiveness for sponsors. Journals with higher IF also increase the prestige of members of the editorial board and, in turn, the quality of the peer-review process still increases in the long term.

[25] See Gans and Shepherd (1994), Seidl et al. (2005), Smith (2016), Ross-Hellauer (2017) for concerns in the peer-review process.

[26] Indeed the peer review process is one of the required conditions to be included in the Scopus and ISI - Web of Science databases.

This study has two main policy implications. The first one deals with the issue of the efficient use of public resources in academic research. Indeed, payment of APC is usually sponsored by the research department, and for useless articles, it constitutes a clear waste of research funds (for instance Assistant Prof. *Misunderstood Genius* had to pay about 4,900 US dollars as APC for the 14 accepted manuscripts). The second implication deals with the effect that these fee-charging journals have on scholars and institutions bibliometric scores. Bibliometric statistics are often used in a comparative way (e.g. the Italian national agency for the evaluation of universities and research institutes sets minimum standards of research production for candidates to the National Scientific Qualification System based on median values of the overall research production of Italian academic community) therefore, publications on predatory journals may distort informative sets for decision making in universities and academic recruitment. Accordingly, additional checks may be required, since a merely quantitative approach is not sufficient in establishing the researcher's performances as well as in defining the quality of a scientific journal.

From a normative viewpoint, a potential solution to disentangle the relationship between APC business model and predatory conduct might be making the review process of published articles public - but keeping the identities of reviewers anonymous.[27]

[27] For instance, this is the practice followed by the APC Journal "Economics: The Open-Access, Open-Assessment E-Journal".

References

[1] Bagues, M. F., Sylos-Labini, M., Zinovyeva, N., (2019). A Walk on the Wild Side: 'Predatory' Journals and Information Asymmetries in Scientific Evaluations. Research Policy 48, no. 2: 462-477.

[2] Baxt W.G., Waeckerle J.F., Berlin J.A., Callaham M.L. (1998). Who Reviews the Reviewers? Feasibility of Using a Fictious Manuscript to Evaluate Peer Reviewer Performance. Annals of Emergency Medicine 32(3): 310-317.

[3] Beall J. (2016). Essential Information about Predatory Publishers and Journals. International Higher Education, 86: 2-3.

[4] Björk, B. C. (2019). Acceptance rates of scholarly peer-reviewed journals: A literature survey. El Profesional de la Información. 28(4), e280407.

[5] Björk, BC, Kanto-Karvonen, S Harviainen JT (2019). How Frequently are Articles in Predatory Open Access Journals Cited. arXiv preprint arXiv:1912.10228

[6] Bladek M (2014) DORA: San Francisco declaration on research assessment (May 2013). Coll Res Libr News 75:191–196. http://crln.acrl.org/content/75/4/191.full

[7] Bohannon J. (2013). Who's Afraid of Peer Review? Science, 342 (6154): 60-65.

[8] Durante, R., Labartino, G., Perotti, R. (2011). Academic Dynasties: Decentralization and Familism in the Italian Academia. Available at NBER: http://www.nber.org/papers/w17572

[9] Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US states data. PLoS ONE, 5(4), e10271. doi: 10.1371/journal.pone.0010271

[10] Fanelli, D., Larivière, V. (2016). Researchers' Individual Publication Rate Has Not Increased in a Century. PloS one, 11(3), e0149504. https://doi.org/10.1371/journal.pone.0149504.

[11] Fanelli, D., Costas,R., Ioannidis, J. P. A. (2017). Meta-assessment of bias in science. Proc. Natl. Acad. Sci. U.S.A. 114, 3714-3719.

[12] François, O (2015). Arbitrariness of peer review: A Bayesian analysis of the NIPS experiment. arXiv:1507.06411.2015.

[13] Gans J.J., Shepherd G. B. (1994). How Are the Mighty Fallen: Rejected Classic Articles by Leading Economists. The Journal of Economic Perspectives, 8(1): 165-179.

[14] Grösche, P., Schmidt, U., Seidl, C. (2005). The Performance of Peer Review and a Beauty Contest of Referee Processes of Economics Journals. Estudios de Economía Aplicada. 23: 505-551.

[15] Harnad, S., (2015). Open access: what, where, when, how and why. In, Holbrook, J. Britt and Mitcham, Carl (eds.) Ethics, Science, Technology, and Engineering: An International Resource. Farmington Hills MI, US. Macmillan Reference.

[16] Haustein, S., Larivière, V. (2014). The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects. À paraître dans (Eds.) I. Welpe, J. Wollersheim, S. Ringelhan & M. Osterloh, Incentives and performance: Governance of knowledge-intensive organizations, Cham: Springer International Publishing.

[17] Hicks, D, Wouters, P, Waltman, L, Rijcke, S De, & Rafols, I (2015). Bibliometrics: the Leiden Manifesto for research metrics. Nature, nature.com.

[18] Langford, J, Guzdial, M, (2015). The arbitrariness of reviews, and advice for school administrators. Communications of the ACM 58 (4):12-13.

[19] Moed, H., Burger, W., Frankfort, J., van Raan, Ton. (1985). The use of bibliometric data for the measurement of university research performance. Research Policy. 14(3). 131-149. 10.1016/0048-7333(85)90012-5.

[20] Martin A., Martin T. (2016). A not-so-harmless experiment in predatory open access publishing. Learned Publishing, 29(4), 301-305.

[21] Osterloh M., Frey B. S., (2015). Ranking Games. Evaluation Review 39(1): 102-129

[22] van Raan, T. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. Scientometrics 62: 133-143. 10.1007/s11192-005-0008-6.

[23] Ross-Hellauer T. (2017). What is open peer review? A systematic review. F1000Research, 6, 588. https://doi.org/10.12688/f1000research.11369.2

[24] Rubin, D. B. 2008. For objective causal inference, design trumps analysis. Annals of Applied Statistics 2: 808–840.

[25] Seidl C., Schmidt U., Grösche P. (2005). The performance of peer review and a beauty contest of referee processes of economics journals. Estudios de Economía Aplicada, 23: 249-286.

[26] Severin, A., Strinzel, M., Egger, M., Domingo, M., Barros, T., (2020). Who reviews for predatory journals? A study on reviewer characteristics bioRxiv 2020.03.09.983155.

[27] Shen, C., Björk, B. C. (2015). Predatory'open access: a longitudinal study of article volumes and market characteristics. BMC medicine, 13(1), 230.

[28] Smith R. (2006). Peer review: a flawed process at the heart of science and journals. Journal of Royal Society of Medicine 99(4): 178–182.

[29] Solomon D.J., Björk B. (2012). A Study of Open Access Journals Using Article Processing Charges. Journal of the American Society for Information Science and Technology 63(8):1485–1495,

[30] Sorokowski Piotr, Kulczycki Emanuel, Sorokowska Agnieszka, Pisansk Katarzyna (2017). Predatory journals recruit fake editor, Nature News, Mar 22.

[31] Starbuck W. H. (2005). How Much Better Are the Most-Prestigious Journals? The Statistics of Academic Publication. Organization Science 16(2):180-200.

[32] Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical science: a review journal of the Institute of Mathematical Statistics, 25(1), 1.

[33] van Dalen, Hendrik P., Henkens, Kene, (2012). Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey. CentER Discussion Paper Series No2012003. Available at SSRN: https://ssrn.com/abstract=1983205

[34] van Noorden, R. (2020). Hundreds of scientists have peer-reviewed for predatory journals. Nature.

[35] Wallace F.H., Perri T. J., (2018). Economists behaving badly: publications in predatory journals Scientometrics 115(2):749-766.

[36] Waltman, L. (2016). A review of the literature on citation impact indicators, Journal of Informetrics, 10(2): 365-391

[37] Weingart, P. Impact of bibliometrics upon the science system: Inadvertent consequences? Scientometrics 62(1): 117–131.

[38] Wang, Q., Waltman, L., (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus, Journal of Informetrics, 10(2): 347-364.

[39] Wooldridge, J. M. 2010. Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge, MA: MIT Press.

[40] Xia, J., Harmon, J., Connolly, K., Donnelly, R., Anderson, M., Howard, H., (2014). Who publishes in "predatory" journals?. Journal of the Association for Information Science and Technology. 66. 10.1002/asi.23265.

Appendix 1 – Results by Sample and by Area

Figure A.1.a: Results Sample 1 - Administrative and Management

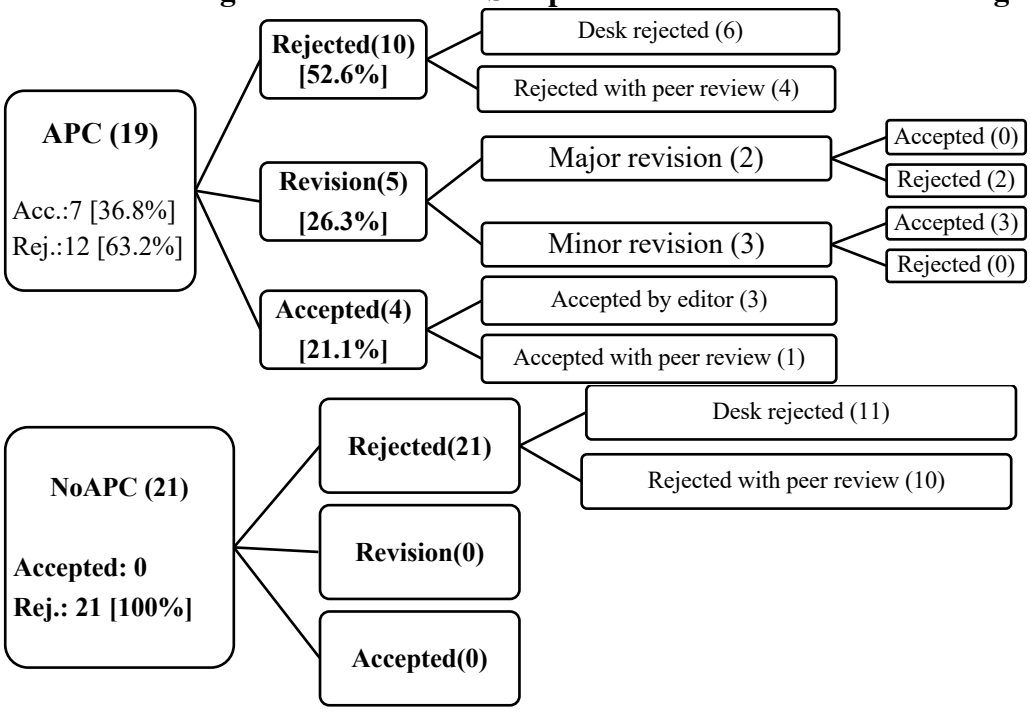


Figure A.1.b: Results Sample 1 - Economics

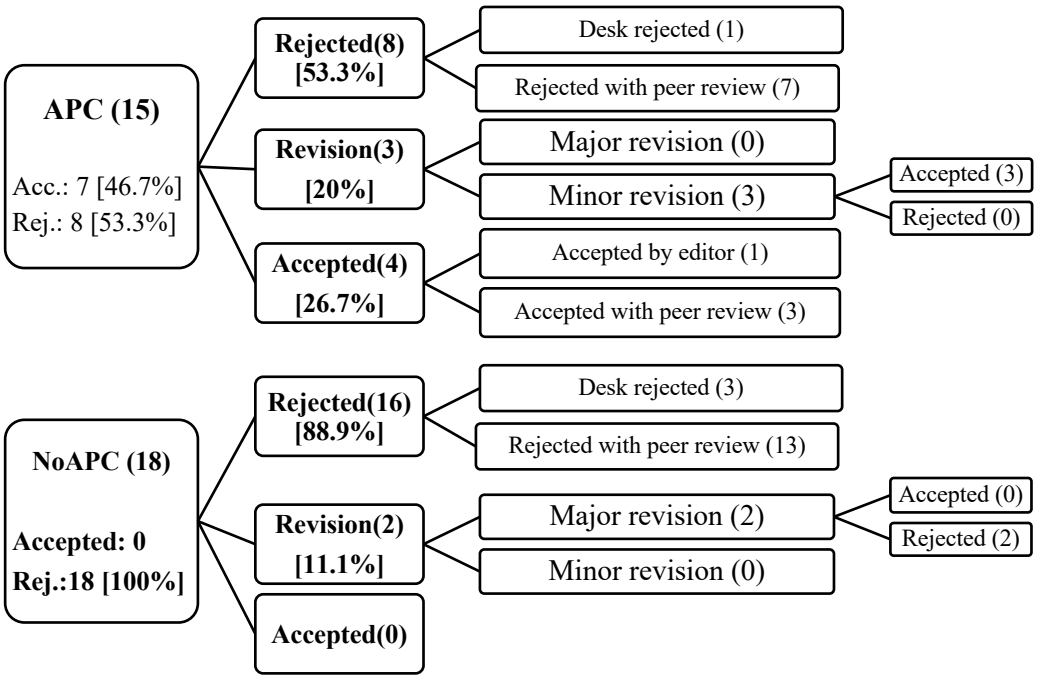


Figure A.2.a: Results Sample 2 - Business

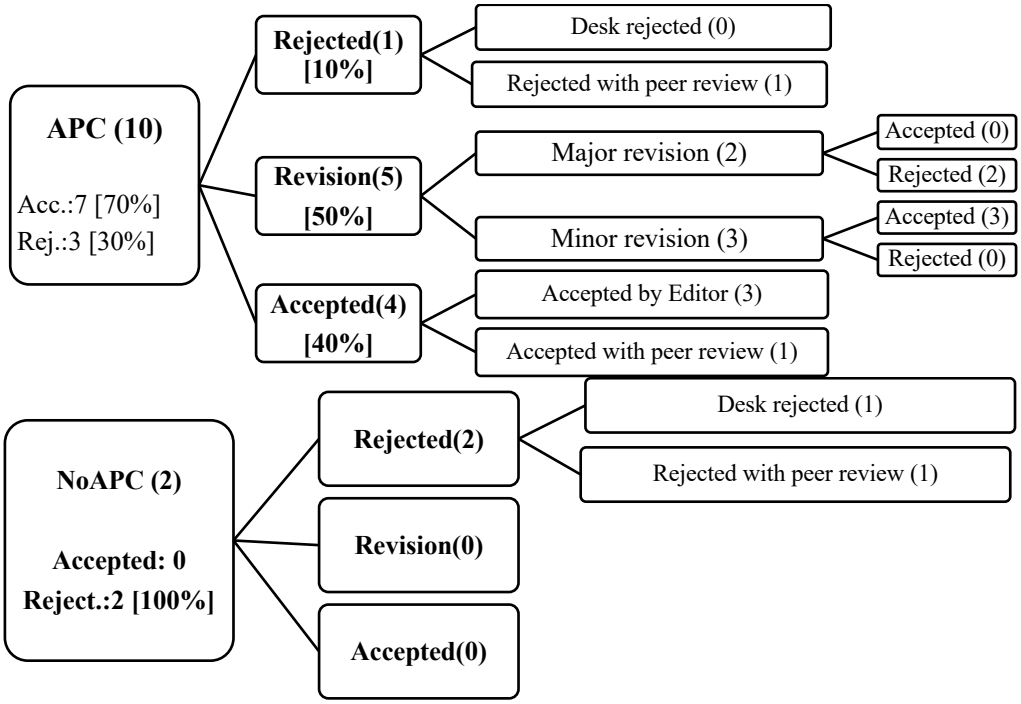
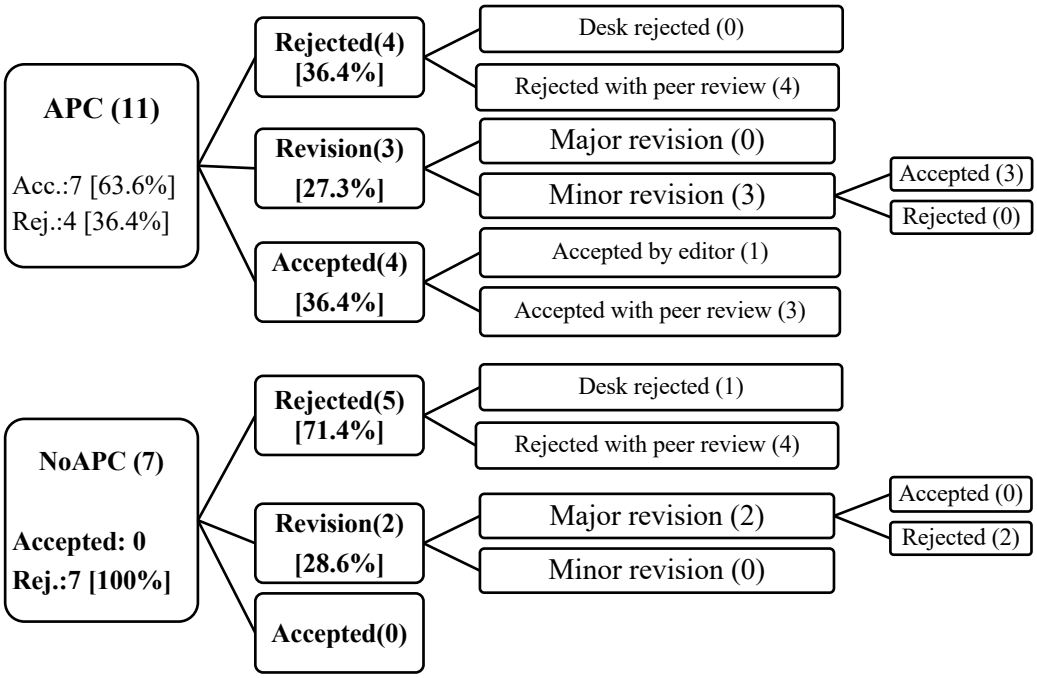


Figure A.2.b: Results Sample 2 - Economics



Appendix 3 – - The Bait-Article

********** **Bait Manuscript**********

Crime and Economic Growth. An empirical analysis for Germany.

For "Administrative and Management" Journals:

Camddeall Athrylith
Department of Economics, Management and Statistics
National University of Aberystwth, Wales
c.athrylith@gmail.com

For "Economic" Journals:

Jeni Konpreyansyon
Department of Economics, Management and Statistics
National University of Port-au Prince, Haiti
j.konpreyansyon@gmail.com

Abstract
This paper examines the role of crime, enforcement and taxation on the economic growth. These effects are studied by modified version of an endogenous growth model proposed by Loayza (1996). Econometric results are based on the German economy over the period 1992-2016. Empirical evidence confirms the theoretical model. We show that the relationship between crime and growth rate of GDP is negative in the long run equilibrium.

Keywords: Crime; Economic Growth; Enforcement.

Acknowledgements
The author is grateful to participants at the AIB-TD 2018 Annual Meeting and ABM 2017 Annual Meeting for their valuable comments and suggestions. In addition, authors are grateful for the financial support received from the Ministry for Education, Research and Competitiveness of [*Wales or Haiti*] (project grants 104781 - Ref. EDU2018-270475-C10-02).

1. - Introduction

The relationship between economic growth and crime (EC) is a topic of great relevance in the public debate. This work aims to analyze the effects that policies to combat crime and taxation could have on the rate of long-term GDP growth when an illegal sector operates in the economic system.

The theoretical reference approach is that of the neoclassical theory of endogenous growth. In particular we propose a model inspired by Loayza (1996) adapted to be suitable for an empirical application. The à la Loayza model (1996) has two characteristics that justified its preference with respect to other formulations proposed in the literature. The first is to have a simple basic structure (1 sector, 1 asset, 1 consumer-producer agent, constant returns, etc.) and flexible that allows to focus the attention on other characteristics of the analyzed phenomenon (eg the system sanctions, the relationship with the C, etc.) usually over-simplified to facilitate analytical tractability. The second advantage of the model proposed by Loayza (1996) consists in the ability to identify, for an economic system in which legal and illegal sector coexist, conditions of steady state equilibrium suitable for use for empirical analysis.

The analytical approach will offer only a partial explanation of the complex link between economic growth and criminal economy.

The relationship between economic growth and EC is a source of heated debate among scholars not so much on the "if" the two sectors (legal and illegal) interact, but on the type and sign of the interactions.

In this sense, the inclusion of EC among the factors influencing growth is supported by a literature, empirical theory, consolidated (eg de Soto, 1989; Johnson, Kaufmann, Mcmillan and Woodruff, 2000; Friedman, Johnson , Kaufmann and Zoido-Lobaton, 2000, Carillo and Pugno, 2004, Banerjee and Duflo, 2005, Cimoli, Primi and Pugno. On the contrary, it is more controversial to state, if the EC has a pro- or anti-cyclical function, if the effects are independent of the degree of economic development, if relations of (Granger) causality exist from or to the EC, etc.[1]

In summary, the contribution of this work can be summarized in the attempt to formalize, and then empirically test for Germany, a model of endogenous growth where both formal institutions (eg penal and administrative sanctions system, quality of institutions, sector efficiency) public, etc.) informal (Crime) play a role in determining long-term growth.

The work is organized as follows: the second paragraph presents the theoretical model that will be calibrated for the German economy (third paragraph); the fourth paragraph empirically tests whether the hypotheses of the model are confirmed with reference to Germany in the period 1992-2016. The fifth paragraph summarizes the main results of the analysis and offers some general conclusions. Two appendices provide more details regarding the construction of the database and analysis of propaedeutic cointegration to empirical analysis.

2. - A model of endogenous growth with Criminal Sector

[1] With reference only to recent works dealing with these issues, we should mention: Dell'Anno and Morone (2018), Caferra (2017).

In this section we will present the growth model used as a reference for the empirical analysis of the effects of some economic policy variables on the GDP growth rate. This objective is pursued by adapting the endogenous growth model proposed by Loayza (1996). The representation of the proposed economy makes it possible to examine how crime reacts to variations, ceteris paribus, enforcement, the quality of institutions, the productivity of the public sector and, through these, the implications that they have on the rate of GDP growth per capita. long-term.

The model is constructed by hypothesizing a production technology that is a function of the supply of congestible public goods. The economy consists of agents who offer a measure of capital on the market that includes both physical capital and human capital. The production technology exhibits constant returns of scale and produces a single good used, indifferently, for consumption and for investment. Following the approach of Barro and Sala-i-Martin (1992), it is assumed that the rate of return on capital depends on the ratio between the amount of public goods (G) and total production (Y_i). On the basis of these hypotheses, the product of the i-th agent is given by:

(1) $$Y_i = A\left(\frac{G}{Y}\right)^{\alpha} k_i \qquad \text{with} \quad 0 < \alpha < 1 \text{ e } A > 0$$

where: A is the parameter that measures productivity (exogenous); k_i the capital endowment of the i-th agent; α is the elasticity of production to variations in the ratio (G/Y).

There are two sectors in the economy:

- the legal sector $\left(y_i^L\right)$, where agents pay a proportional tax on income with a rate τ and whose revenue is used by the government to finance the provision of productive public services (G), enforcement activities and other non-productive expenses.

(2) $$y_i^L = (1-\tau) A\left(\frac{G}{Y}\right)^{\alpha} k_i, \qquad \text{with} \quad 0 < \tau < 1$$

- the criminal or illegal sector $\left(y_i^I\right)$, where the agents renounce a share π^e of their income as a sanction and whose amount is used by the government to finance exclusively the enforcement activity. It is hypothesized that the criminal discounts the penalty for the probable probation of assessment for which the sanctioning rate incorporates the taxpayer's subjective perception of being subjected to control.

The operators of the illegal market will have access only to a share (δ) of public services offered by the public sector (G) whose use is permitted without affecting the status of illegality of economic activity (non-excludable goods) [2]. In symbols:

(3) $$y_i^I = \left(1-\pi^e\right) A\left(\frac{\delta G}{Y}\right)^{\alpha} k_i, \qquad \text{with } 0 < \pi < 1 \text{ and } 0 < \delta < 1$$

[2] For example, public spending to finance services for the judicial protection of contracts, the social security and social security system for workers (formal), public subsidies to businesses, etc.

The provision of public services is financed through taxation on legal production:

(4) $$G = \eta(q, \lambda, p)(\tau, Y^L), \qquad \text{with } 0 < \eta \leq 1$$

where $\eta(\cdot)$ is the share of revenue destined to the financing of rival productive public services and $1-\eta(\cdot)$ is the part destined to the financing of unproductive expenses, including the fight against illegality. It is hypothesized that $\eta(\cdot)$ it is positively correlated with the quality of institutions (q), this is motivated by the fact that better institutions make it possible to administer the bureaucracy more effectively and efficiently $\left[(\partial\eta/\partial q) > 0\right]$.

Particular attention was paid to the formulation of the variables that determine the activity against crime (π^e). Compared to Loayza (1996), where an increase in enforcement was possible only by reducing public resources destined to purchase "productive" public goods and services, in the model proposed here, it is hypothesized that the increase in control activity (π^e) produces an increase in revenue greater than the cost of the service (unproductive). From this it follows that the greater revenue collected can also be used for public expenses that are propaedeutic to growth. In particular, if enforcement (p) or criminal proceedings (λ) become more frequent and / or administrative penalties (σ) or penalties (f) are harder, crime is reduced $\left[\left(\partial y^I / \partial \pi^e\right) < 0\right]$. Furthermore, it is assumed that the amount of resources necessary to improve the enforcement system is decreasing to improve the quality of the institutions and that the yield, in terms of share of revenue allocated to public production expenditure, of the prosecution in criminal cases (λ) has decreasing returns will scale. In formal terms the share of revenue allocated to the financing of productive public spending (G) is given by:

(5) $$\eta = p\lambda(q - \lambda)$$

where, by imposing the constraint on the quality of the institutions $q > 2\lambda$, the partial derivatives will be obtained that guarantee the desired functional characteristics: $\frac{\partial\eta}{\partial\lambda} > 0$; $\frac{\partial\eta}{\partial p} > 0$; $\frac{\partial\eta}{\partial q} < 0$; $\frac{\partial^2\eta}{\partial\lambda\partial q} > 0$; $\frac{\partial^2\eta}{\partial p\partial q} < 0$; $\frac{\partial^2\eta}{\partial\lambda^2} < 0$. Indicating with S the relative size of the legal economy compared to the total one $\left(Y = Y^L + Y^I\right)$:

(6) $$S \equiv \frac{Y^I}{Y}$$

From (4) and (6) we can derive the relationship between the provision of public services and total production:

(7) $$\frac{G}{Y} = \eta(q, \lambda, p)\left(\tau(1 - S)\right)$$

Due to the simultaneous presence of congestion phenomena in the public service and due to the non-contribution of Illegal Economy to the financing of productive public goods, we will, ceteris paribus, reduce the productivity of all agents as the size of the illegal sector increases.

A fundamental role in the model proposed in this work is carried out by the expected sanctioning rate (π^e). It determines whether or not the agent participates in the illegal sector and its formalization constitutes the main element of differentiation with respect to Loayza (1996). In modeling the system of sanctions it is assumed that the agent, in deciding whether to operate in the illegal sector, subjectively assesses the direct pecuniary consequences - administrative sanction (σ) - as well as those of a criminal nature - probability of reporting to the judicial authority (λ) and associated costs (f). The functional form of the expected sanctioning rate (π^e), is explicit as a linear relation both of the level of diffusion of illegality and of the probability of assessment. These factors have a proportional effect on the two components that determine the amount of the penalty (expected). The first, linked to the direct monetary cost (pecuniary administrative sanction proportional to the tax evaded: $\sigma\tau$) and, second, connected to the possible denunciation to the judicial authority resulting from the evasive act ascertained. In economic terms, it can be interpreted as the share of production necessary to finance the tax payer's defensive activity (or even as a shadow price of the psychological cost of having to support a judicial proceeding). Assuming, for simplicity, a fixed cost for the defensive activity or psychological cost (f), this will be multiplied by the probability that the enforcement will follow a complaint from the judicial authority (λ). In formal terms:

$$\pi^e = (\sigma\tau + \lambda f)\, pS \tag{8}$$

In summary, the expected sanctioning rate π^e will be a function: (1) of the probability of assessment $\left[\left(\partial\pi^e/\partial p\right) > 0\right]$; (2) the administrative sanctioning tax rate evaded from criminal proceeds $\left[\left(\partial\pi^e/\partial\sigma\right) > 0\right]$; (3) the effectiveness of the penal sanction system, measured as the probability that a finding results in a complaint to the judicial authority $\left[\left(\partial\pi^e/\partial\lambda\right) < 0\right]$; (4) the cost of defensive activity in the jurisdictional seat (or psychological cost of being subject to court proceedings) $\left[\left(\partial\pi^e/\partial f\right) > 0\right]$; (5) the size of the illegal sector $\left[\left(\partial\pi/\partial S\right) > 0\right]$; (6) of the tax rate $\left[\left(\partial\pi^e/\partial\tau\right) > 0\right]$.

Taking into account the neoclassical hypothesis of perfect mobility of productive factors between the legal and illegal sectors, we will have that in equilibrium the rate of return between the two sectors must necessarily be equal (eq. 9). From this condition we obtain the share of Illegal Economy compared to the official GDP of equilibrium. By restricting the solution of the model to an internal solution, that is, where illegal and legal economics coexist, substituting and simplifying the condition of equality of the sectoral yield rates (2) and (3), we obtain:

$$\left(1-\pi^e\left(S,\sigma,\tau,p,\lambda,f\right)\right)\delta^{\alpha} = (1-\tau) \tag{9}$$

Substituting the (8) in (9) and expressing as a function of the relative dimension of the EC we have:

(10) $$S = \frac{\delta^{\alpha} - \tau + 1}{\delta^{\alpha} p(\sigma\tau + \lambda f)}$$

The internal solution is guaranteed by imposing the following restrictions on the parameters:

(11) $$S > 0 \Rightarrow \tau > 1 - \delta^{\alpha}$$

(12) $$S < 1 \Rightarrow \delta^{\alpha}\left[1 - p(\sigma\tau + \lambda f)\right] + \tau < 1$$

From (12) we can derive the constraint on the parameter $\sigma(\tau, p, \lambda, \delta, \alpha, f)$:

(13) $$\sigma > \frac{\delta^{\alpha}(1 - p\lambda f) + \tau - 1}{\tau\delta^{\alpha} p}$$

From (13), to ensure that the administrative penalty is greater than the tax evaded, we obtain the constraint on $\tau(p, \lambda, \delta, \alpha, f)$:

(14) $$\tau > \frac{1 + \delta^{\alpha}(p\lambda f - 1)}{1 + \delta^{\alpha} p}$$

From (14), to guarantee a positive tax rate, we can determine the lower limit of the probability of assessment $p(\lambda, \delta, \alpha, f)$:

(15) $$p > \frac{\delta^{\alpha} - 1}{\delta^{\alpha}\lambda f}$$

Considering that $p < 1$, from (15) we derive that the lower limit to the enforcement parameter $\lambda(\delta, \alpha, f)$:

(16) $$\lambda > \frac{\delta^{\alpha} - 1}{f\delta^{\alpha}}$$

This constraint is always verified by values of $\lambda > 0$, since the right-hand side of the inequality is always negative since both the public sector productivity compared to the private sector (α) and the share of public services that can not be excluded to the illegal economic operators (δ) included in the unit range and the cost of defense in the judicial process is always positive $(f > 0)$.

From (10), and respecting the constraints on the parameters that guarantee internal solution to the illegal sector (11 - 15), we can see how variations in the variables affect the EC dimension:

Hp.1: When the probability of assessment increases then the EC is reduced $\left[(\partial S/\partial p) < 0\right]$

- *Hp.2:* When the rate of administrative sanctions increases then the EC is reduced $\left[(\partial S/\partial \sigma) < 0\right]$;
- *Hp.3:* When the tax rate increases then the EC increases $\left[(\partial S/\partial \tau) > 0\right]$;
- *Hp.4:* When the criminal relevance of the crimes increases then the EC is reduced $\left[(\partial S/\partial \lambda) < 0\right]$;
- *Hp.5:* When the cost of criminal proceedings (legal protection, court costs, psychological cost) increases

then the EC is reduced $\left[\left(\partial S/\partial f\right)<0\right]$;

Hp.6: When the productivity of the public sector, relative to the private sector, increases then the EC increases $\left[\left(\partial S/\partial \alpha\right)<0\right]$;

- *Hp.7:* When the share of public services that cannot be excluded by criminals (possibility of free riding) increases then the EC increases $\left[\left(\partial S/\partial \delta\right)>0\right]$.

Substituting the (7) in (3) we have that under equilibrium conditions between the EC and the legal economy the net return on capital (r) is given by:

$$r=\left[A\left(1-\tau\right)\tau^{\alpha}\right]\left\{\eta\left(q,\lambda,p\right)\left[1-S\left(\sigma,\tau,p,\lambda,\delta,\alpha,f\right)\right]\right\}^{\alpha} \tag{17}$$

The first addend $\left[A\left(1-\tau\right)\tau^{\alpha}\right]$ corresponds to the case in which there is no EC. In this case, the rate of return on capital is a function that is first increasing and then decreasing with respect to the tax rate τ.

As far as the best consumer choice is concerned, Loayza (1996) uses a utility function that has a constant intertemporal substitution elasticity (EIS).

$$Max\ U=\int_{0}^{\infty}\frac{c_i^{1-\theta}-1}{1-\theta}\ e^{-\rho t}dt \tag{18}$$

Where ϱ is the (constant) rate of inter-temporal preference of consumption and whose EIS is given by reciprocal aversion related to risk (θ).

The dynamic budget constraint indicates that the variation in the representative consumer's wealth is equal to the difference between income from capital net of tax and consumption $\left[c_i\left(t\right)\right]$.

$$\dot{k}_i\left(t\right)=y_i\left(t\right)-c_i\left(t\right)=rk_i\left(t\right)-c_i\left(t\right) \tag{19}$$

The Hamiltonian of the problem is:

$$H=\frac{c_i^{1-\theta}-1}{1-\theta}e^{-\rho t}+\mu\left[rk_i-c_i\right] \tag{20}$$

he conditions for maximization are: $H_c=0\rightarrow e^{-\rho t}c_i^{\theta}=\mu$; $H_k=-\dot{\mu}\rightarrow r=-\frac{\dot{\mu}}{\mu}$ and the condition of transversality, which requires the consumer not to borrow without limits, is $\mu_T k_T=0$. From the solution of the optimal problem, we have the rate of consumption growth that is constant and equal to γ:

$$\frac{\dot{c}_i\left(t\right)}{c_i\left(t\right)}=\gamma=\frac{1}{\theta}\left[r-\rho\right] \tag{21}$$

In a steady state, the solution of the model implies that the rates of capital growth, of both legal $\left(Y^L\right)$ and illegal $\left(Y^I\right)$ aggregate production $\left(Y\right)$, are constant and equal to the growth rate of consumption $\left(\gamma\right)$. We can therefore write that:

(22) $$\gamma = \frac{1}{\theta}\left\{\left[A(1-\tau)\tau^{\alpha}\right]\left\{\eta(q,\lambda,p)\left[1-S(\sigma,\tau,p,\lambda,\delta,\alpha,f)\right]\right\}^{\alpha}-\rho\right\}$$

Which explicitly becomes: (23) $$\gamma = \frac{1}{\theta}\left\{\left[A(1-\tau)\tau^{\alpha}\right]\left\{p\left(\lambda q+\lambda^{2}\right)\left[1-\frac{\delta^{\alpha}-\tau+1}{\delta^{\alpha}p(\sigma\tau+\lambda f)}\right]\right\}^{\alpha}-\rho\right\}$$

3. An empirical verification for the German economy

In this section we propose an empirical verification of the hypotheses used for the construction of the theoretical model with reference to the German economy of the last thirty years. In general, the empirical investigations that analyze the interactions between legal and illegal economics meet two fundamental problems to which this analysis does not subtract.

The first derives from the inevitable inclusion in the regression of variables with obvious measurement errors (for example, estimates of illegality, enforcement indices). The second limitation derives from the fact that, if one wanted to use an exhaustive model of interactions between legal and illegal economics, the number of variables to be included would be disproportionate to the (small) sample size conditioned by the limited quantity / quality / extent of available data. A comprehensive representation of the phenomenon would therefore result in a reduction in degrees of freedom that would further jeopardize the reliability of the estimates. Ultimately, the presence of measurement errors, the necessary omission of some relevant variables and the reduced sample size require us to evaluate the results of the empirical analysis with caution.

The empirical verification that is proposed is aimed at essentially testing whether, as hypothesized in the theoretical model, the growth rate of real German per capita GDP is influenced by changes in the tax burden, the degree of enforcement and the size of the criminal sector.

3.1 Database

The database is built with reference to the Germans and covers a period that goes, for the variables without missing values, from 1992 to 2016. The presence of missing values and the use of the series in raw differences, reduces the sample of the estimated regressions depending on the model specification. Referring to Appendix 1 for details of data sources and their definitions, in this section the meaning of the variables used will be explained, highlighting the main problems related to the construction of the sample.

The dependent variable is represented by the growth rate of real GDP per capita, published in the World Bank database. The tax pressure index was used as a proxy for the proportional tax rate.

For the EC, the estimates of Jehle (2017) of the ratio of income produced by crime as a percentage to GDP were used. Although this variable does not perfectly coincide with the size of the EC, it is a reliable measure of the same. Statistics on the degree of enforcement are difficult to find. For Germany, in fact, official data are not available on the number of checks or on the results of the same, and the institutions in charge publish, in a structured way and with homogeneous methods, information related to their inspection activity. In this work, attempts have been made to circumvent the problem by creating a proxy for criminal enforcement based on the judicial statistics published by Eurostat. It is calculated as the ratio between the number of subjects reported

by the judicial authority for criminal offenses per 100,000 inhabitants. The hypothesis that underlies the use of this indicator is that an increase in the number of subjects reported should indicate a greater effectiveness of enforcement.

To take into account the problems related to the omission of relevant variables, three control variables were included, aimed at explaining some of the determinants of growth that the theoretical model neglects by simplification. In particular we include an index of the economy's opening up to international trade, an indicator of the quality of human capital and the rate of participation in the labor force. In order to check the robustness of the results, alternative specifications of the regression model were used.

3.2 Econometric methodology and results of empirical analysis

From the work of Nelson and Plosser (1982) the attention to the presence of stochastic trends in the historical series has increased. In the case of static regression with OLS estimators that use variables in levels, the presence of integrated processes of the same order can lead to a spurious regression, with the consequence of supporting the existence of relationships between variables even when these are stochastically independent . A solution to prevent this problem is to differentiate the variables so as to obtain stationarity. Engle and Granger (1987) show, however, that not necessarily static regressions at levels with integrated variables are spurious regressions. In fact, there is the possibility that a linear combination of processes I (1) turns out to be a stationary process, I (0) (i.e. cointegration). In this case, the cointegration relationship is interpreted as the long-term static equilibrium relationship. Two approaches are adopted in the literature to verify the presence of cointegration: the residual two-step procedure to test the null hypothesis of absence of cointegration (Engle and Granger, 1987; Phillips and Ouliaris, 1990) and the regression approach of reduced rank based on estimates of the multivariate system due to Johansen (1997, 1999). The small sample size limits, once again, the reliability of these tests. In particular, Johansen's systemic approach is totally inapplicable to our sample.

Appendix 2 shows the results of the univariate tests for the integration order and the Engle-Granger and Phillips-Ouliaris tests for the presence of cointegration among the variables included in the model. The univariate analysis shows that all the variables are I (1) while from the cointegration tests we reject the hypothesis of cointegration between the variables of the model. Based on these results, the most appropriate approach to avoid the risk of spurious regression is the transformation into raw differences and the subsequent estimation with OLS estimators.

Having avoided the problem of spurious regression, an econometric model is estimated in order to verify whether the effect of changes in the tax burden, the level of enforcement and the criminal economy on the real growth rate per capita confirm, for the German economy, the model forecasts.

The unavailability of data with adequate extension has not made possible an empirical verification of the other parameters of the theoretical model (for example, quality institutions, probabilities of assessment, administrative sanctions, efficiency of the public sector, public expenditure that can not be excluded).

The specified models are deliberately thrifty in terms of the variables included in the specification due to the small sample size. In order to reconcile the effects of multicollinearity and endogeneity between tax pressure

and the proxy of the criminal economy in the evaluation of output, it is considered preferable to use the ratio between direct direct taxes on GDP (models IV to VII). A summary of the most significant regressions is shown in Table 1.

Table 1 - dependent Variable: growth rate of GDP per capita

	I	II	III	IV	V	VI	VII
Constant	-0,214 (-0,411)	-1,349** (-2,584)	-0,787** (-2,563)	-0,806 (-1,552)	-	-0,227 (-0,724)	0,242 (0,808)
Tax burden (τ_1)	-0,240 (-0,788)	-0,395 (-1,464)	-0,481* (-1,595)	-	-	-	-
Tax burden^2 $(\tau_1)^2$	-	-	0,225* (1,660)	-	-	-	-
Pers. Inc.Tax burden. (τ_2)	-	-	-	-0,533** (-1,902)	-0,611** (-2,022)	-0,641*** (-2,404)	-0,536 (-1,442)
Pers. Inc.Tax burden.^2 $(\tau_2)^2$	-	-	-	-0,604*** (-3,169)	-0,802*** (-5,200)	-0,654*** (-3,067)	-0,613*** (-2,191)
Criminal Economy (EC)	0,202*** (2,176)	0,158** (2,159)	0,090 (1,299)	0,212* (2,383)	0,218** (2,218)	0,206** (3,491)	0,197* (1,407)
Index of *enforcement* (λ)	-	-1,525** (-1,867)	-1,107*** (-2,105)	-0,878* (-1,693)	-0,489* (-1,691)	-0,666** (-1,700)	-
Control Var.							
Internat. Trade Openness	0,285* (1,938)	0,360** (2,673)	0,338*** (3,054)	0,466*** (3,805)	0,470*** (3,860)	0,457*** (3,997)	0,248*** (2,251)
Tertiary Education	0,022 (0,098)	0,339* (1,945)	-	0,224 (1,301)	0,000 (0,002)	-	-
workforce participation rate	0,304 (0,632)	-0,152 (0,707)	-	0,198 (0,714)	0,238 (0658)	-	-
Observations	24	24	24	24	24	24	24
R^2-adjustd	0,190	0,604	0,623	0,665	0,631	0,685	0,253
Durbin-Watson st.	0,657	0,557	0,357	0,480	0,557	0,498	0,357
LM Test serial-correl.[a]	0,950	0,717	0,432	0,803	0,989	0,878	0,693
BPG Test Hetero-sched.[b]	0,726	0,998	0,997	0,978	0,603	0,673	0,979
JB Test Normality[c]	0,648	0,385	0,667	0,778	0,744	0,560	0,652

*Note: ***, **, * indicate statistical significance at 1% , 5% and 10%.*
All variables are in first differences. The t-statistics are calculated using standard errors robust to heteroskedasticity (White, 1980).
[a] Durbin Watson statistic for serial autocorrelation in residuals. The statistic DW shows that there is no positive serial correlation.
[b]Breusch-Godfrey LM Serial correlation test; [c]BPG: Breusch-Pagan-GodfreyTest for heteroskedasticity;
[d] Jarque-Bera test for normality of residuals. For the three tests (a, b, c) the p-values are reported where the null hypotheses are respectively: absence of serial correlation up to the 2nd order, absence of heteroskedasticity and absence of asymmetry and kurtosis in the residuals.

On the basis of the results obtained, we note that the coefficient that links the tax burden to the growth rate of GDP (models I, II and III) is statistically not different from zero. On the other hand, the quadratic link between GDP growth and direct tax pressure is statistically significant. The concave downward relationship between these two variables confirms the non-linear nature between economic growth and the rate of taxation envisaged by the steady state equilibrium relation (equation 22).

The link between GDP and EC growth rate is negative. This result supports the hypothesis used by Loayza (1996) for the construction of the model. In particular, the author estimates, for the Latin American countries, that an increase of 1 percentage point of the EC compared to the GDP - ceteris paribus - corresponds to a

reduction in the real legal GDP per capita of 3 percentage points. According to our estimates for Germany, over the period 1986-2016, an EC increase of 1 percentage point reduces the growth of real GDP per capita by around 0.2%.

Based on the assumptions of the theoretical model, the negative correlation between the EC dimension and economic growth is motivated by the assumption that legal production also depends on public services subject to congestion to which the criminal sector, while not contributing to financing, can access for the part of non-excludable services. The hypothesis of "common" public goods and services (non-excludable and rivals) used in the model therefore implies that a greater EC, limiting the ability of the policy maker to finance those public goods and services necessary for the development of the country, produces a reduction of the long-term growth rate.

The enforcement indicator of justice is statistically significant (at a level of significance of 5%). This result therefore supports the hypothesis of the proposed theoretical model according to which greater enforcement increases the rate of GDP growth. Unfortunately, the lack of data on the other, and more relevant, enforcement parameters (probabilities of assessment, level of administrative and penal sanctions) and quality of institutions do not allow us to empirically test the reliability of the importance of enforcement on economic growth as hypothesized in the theoretical model.

In conclusion, the empirical analysis presented, although conditioned by the poor quality of some data series, corroborates the predictions of the theoretical model with reference to long-run relationships between GDP growth rate, EC and direct tax pressure and enforcement.

4. - Conclusions

This paper analyzes the relationship between economic growth and the criminal economy in a context where a significant role is played by the quality of the institutions (understood as factors relevant to the effectiveness of the police system of sanctions and law enforcement). An adaptation of the endogenous growth model proposed by Loayza (1996) has been proposed. It was complemented by a more realistic formalization of the system of sanctions structured and calibrated with reference to the German economy.

The econometric analysis was carried out with data on the German economy in the period 1992-2016. The results of this survey should however be evaluated with caution due to the small sample size and the non-direct observability characteristic of some of the variables included in the regression (EC, enforcement variables, institutional quality), with the consequence that these data inevitably present measurement errors. Having said this, the results of the empirical analysis show that:

- the link between the overall tax burden and the GDP growth rate is statistically weak, while the correlation with direct tax pressure is statistically significant. In line with the theoretical model there is a quadratic (concave) link between per capita GDP growth and an average direct tax rate.
- In relation to the effects of more effective enforcement, the effect of greater contrast to crime on the rate of growth has been verified empirically. The econometric analysis showed a positive correlation

between the two variables. From a regulatory point of view, this analysis suggests that an increase in the fight against crime has positive effects on the growth of the long-term (legal) economy.

- The relationship between GDP and EC growth rate is estimated with a negative sign, confirming the hypotheses of the theoretical model. The inverse link between the EC dimension and economic growth derived, in the theoretical model, from the existence of non-excludable and rival public goods and services for which, the major EC, reducing the ability to finance the policy maker of those public goods and services preparatory to economic development, limited the steady-state growth rate. The estimated empirical model shows that for each increment of a percentage point of the EC there is a reduction of about 0.2 percentage points of the rate of growth of real GDP per capita.

References

Andrews D. - Zivot. E., «Further evidence on the Great Crash, the oil price shock, and the unit-root hypothesis», *Journal of Business and Economic Statistics* , vol. 10, 1992, pp. 251–270.

Banerjee A.V. - Duflo E., «Growth Theory through the Lens of Development Economics», in Durlauf S. - Aghion P. *Handbook of Economic Growth*, Holland: Elsevier Science, 2015, pp. 473–552.

Barro R.J. - Sala-i-Martin X., *Economic Growth*, New York: McGraw-Hill, 1995.

Dell'Anno R. - Morone A., «The criminal economy and underdevelopment», *Economic Systems*, vol. 58, n. 3, 2018, pp. 257-279.

Cimoli M. - Primi A. - Pugno M., «A Low-Growth Model: Illegality as a Structural Constraint», *Cepal Review*, vol. 88, 2006, pp. 85-102.

de Soto H., *The Other Path: The Invisible Revolution in the Third Worlds*, New York: Harper and Row Publishers, 1989.

Caferra. «Consumption Growth and Crime: Some Evidence from Income Quintile Groups in Europe», *Applied Economics Letters*, vol. 13, 2017, pp. 529–532.

Engle R.F. - Granger C.W.J., «Co-integration and Error Correction: Representation, Estimation, and Testing», *Econometrica*, 55, 1987, pp. 251–276.

Friedman E. - Johnson S. - Kaufmann D. - Zoido-Lobaton P., «Dodging the grabbing hand: the determinants of unofficial activity in 69 countries», *Journal of Public Economics*, vol. 76, n. 3, 2000, pp. 459-493.

Johansen S. «Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models», *Econometrica*, 59, 1994, pp. 1551–1580.

Johansen S., *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press, 1998.

Johnson S. - Kaufmann D. - McMillan J. - Woodruff C., «Why do firms hide? Bribes and unofficial activity after communism», *Journal of Public Economics*, vol. 76, n. 3, 2000, pp. 495-520.

Jehle Jörg-Martin. Criminal Justice in Germany: Facts and Figures, Federal Ministry of Justice and Consumer Protection, Eight Edition 2017.

Kwiatkowski D. - Phillips P.C.B. - Schmidt P. - Shin Y., «Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root», *Journal of Econometrics*, vol. 54, 1992, pp. 159-178.

Loayza N.V., «The economics of the illegal sector: a simple model and some empirical evidence from Latin America», *Carnegie-Rochester Conference Series on Public Policy*, vol. 45, 1996, pp. 129-162.

MacKinnon J.G., «Numerical Distribution Functions for Unit Root and Cointegration Tests», *Journal of Applied Economics*, vol. 11, 1996, pp. 601-618.

Nelson C.R. - Plosser C.I., «Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications», *Journal of Monetary Economics*, vol. 10, 1982, pp. 139-162.

Ng S. - Perron P., «Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power», *Econometrica*, vol. 69, 2001, pp. 1519-1554.

Perron P. - Ng S., «Useful Modifications to Some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties», *Review of Economic Studies*, vol. 63, 1996, pp. 435-463.

Phillips P.C.B. - Ouliaris S., «Asymptotic Properties of Residual Based Tests for Cointegration», *Review of Economic Studies*, vol. 58, n. 1, 1990, pp. 165-193.

White H., «A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity», *Econometrica*, vol. 48, 1980, pp. 817–838.

Appendix 1: Database

Table 2 : Database

	Sources	Means	MAX	MIN	# obs
growth rate of GDP per capita (%)	World Bank – "GDP per capita growth (annual %)". World Development Indicators Online. Washington, DC. Codice: NY.GDP.PCAP.ZK.GD	1.6	4.1	-1.0	24 ('92-'16)
Tax burden (%) (τ_1)	Eurostat - tax revenue (including social contributions) in % of GDP, [gov_10a_taxag]	39.0	43.7	31.1	24 ('92-'16)
Pers.Inc. Tax/GDP (%) (τ_2)	Eurostat - Tax on personal income in % of GDP [gov_11a_taxag]	13.5	15.8	9.3	24 ('92-'16)
Index of *enforcement* (λ)	Statistisches Bundesamt, (Destatis). Strafrechtliche Statistiken. (several years): Tab. 10.1 (1992, 1993); Tab. 3.1 (1994-2016) – numero persone denunciate per i quali l'Autorità giudiziaria ha iniziato l'azione penale / 100.000. La serie è trasformata in logaritmi.	5.6	7.7	5.7	24 ('92-'16)
Criminal economy/GDP (%)	Jehle (2017). Reddito criminale prodotto in % of GDP (Tab. 1, pag. 11)	38.3	37.1	26.9	24 ('92-'16)
International Trade Openness (%)	World Bank – "Merchandise trade as a share of GDP". World Development Indicators Online. Washington, DC. Codice: IMP.EXP.TR	37.4	47.8	29.0	24 ('92-'16)
Tertiary Eduction (%)	World Bank – "School enrollment, tertiary (% gross). World Development Indicators Online. Washington, DC. Codice: SE.TER.ENRR (manca il valore per il 1998)	40.8	67.1	24.8	22 ('92-'16)
Workforce participation rate (%)	World Bank – "Labor participation rate, total (% of total population ages 15+). World Development Indicators Online. Washington, DC. Codice: SL.TTF.CA.ZS	48.9	40.4	49.9	24 ('92-'16

Appendix 2: Unit Root Test and Cointegration Analysis

Table 3 shows some of the most used tests to establish the presence of unit root in the univariate series. In the literature it is known that the presence of structural breaks, the reduced sample size, the multiplicity of trends, could compromise the reliability of classical tests such as the Augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP) test. To support the choice of the differentiation order, the tests proposed by Andrews and Zivot (1992) - ZA are also performed: suggested in case of presence of structural break; Kwiatkowski, Phillips, Schmidt and Shin (1992) - KPSS: which, unlike the previous ones, is a stationary and not a unit root test, so it is useful to verify the robustness of previous outputs; the 4 tests proposed by Ng and Perron (2001) - NG-P: representing a more efficient version of the Phillips-Perron test modified by Perron and Ng (1996).

In the event of divergent results regarding the integration order, the integration order was established on the basis of what was suggested by the majority of the tests with reference to the 5% threshold value.

Table 3 - ADF, PP, ZA; KPSS, NG-P Unit Root Tests

Variabiles	**Specific. Test**	**Transf. Series**	**ADF p-v.♣**	**PP p-v.♣**	**ZA p-v.**	**KPSS (c. v. 5%)***	**NG-P (critical value 5%)°**			
							MZa	**MZt**	**MSB**	**MPT**
						Int.=0.463	Int.= -8.1	Int.= -1.98	Int.= 0.233	Int.= 3.17
						I e T= 0.146	I e T=-17.3	I e T=-2.9	I e T= 0.168	I e T= 5.48
growth rate of GDP per capita	Intercetta	Livelli	0.078	0.041	0.302	0.739	-8.96	-2.96	0.231	2.92
		Differn.	0.070	0.051	0.053	0.058	-8.91	-2.91	0.237	2.75
Tax burden	Trend e intercetta	Livelli	0.357	0.557	0.065	0.147	-4.28	-1.43	0.335	21.01
		Differn.	0.016	0.003	0.447	0.054	-12.54	-2.45	0.200	7.57
Pers.Inc. Tax/GDP	Trend e intercetta	Livelli	0.395	0.332	0.217	1.998	-3.30	-1.27	0.384	27.26
		Differn.	0.069	0.001	0.035	0.077	-12.69	-2.46	0.195	7.58
Index of enforcement	Trend e intercetta	Livelli	0.041	0.033	--	0.024	-1.46	-0.78	0.532	53.95
		Differn.	0.019	0.019		0.022	-5.71	-1.69	0.296	15.97
Criminal economy/GDP	Intercetta	Livelli	0.119	0.071	0.053	0.510	-8.42	-2.05	0.244	2.91
		Differn.	0.003	0.090	0.262	0.201	-10.23	-2.26	0.220	2.42
International Trade Openness	Intercetta	Livelli	0.809	0.782	0.003	4.972	-1.74	-0.65	0.376	10.374
		Differn.	0.110	0.090	0.114	0.322	-11.77	-1.42	0.206	2.095
Tertiary Eduction	Trend e intercetta	Livelli	0.239	0.511	0.095	7.609	-0.16	-0.13	0.806	129.41
		Differn.	0.109	0.020	0.159	0.117	-11.35	-2.29	0.202	8.49
Workforce participation rate	Intercetta	Livelli	0.006	0.007	0.000	0.009	-9.57	-2.15	0.227	3.14
		Differn.	0.009	0.004	0.039	0.008	-12.47	-2.49	0.200	1.99

Note: ♣MacKinnon (1996); °Ng-Perron (2001, Table 1); *Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1).
Hypothesis null: the series has unit root for ADF, PP, ZA, NG-P, null hypothesis: the series is stationary for KPSS.
For ADF and PP tests inclusive of 1 delay excluding the enforcement variable that for the reduced size of the series (lag = 0). For ZA the maximum number of included lag is 1. For the enforcement index the limited sample size does not allow the software (Eviews 9.2) to estimate the p-value. For KPSS and NG-P the automatic criteria are used to establish the number of delays and the non-parametric method of estimating the spectrum, respectively, Akaike info Criterion and AR-spectral OLS.

To test the presence of cointegration, the OLS estimates are shown in table 4 with the variables in the levels of the models specified in table 1. The results of the residual based tests proposed by Engle and Granger (1987, EG) and Phillips and Ouliaris are reported (1990, PO). They are unit root tests applied to the residues obtained from the static OLS estimate of non-stationary variables (for the EG test the non-stationary test is the ADF, for the PO test it is the PP). The time series will be cointegrated if the residuals of the static OLS regression are stationary. The result reported in table 2 is robust to changes in trend specification (intercept, linear trend and intercept).

Table 4: Cointegration Tests (Engle-Granger e Phillips-Ouliaris)

Models	**I**	**I**	**II**	**II**	**IV**	**IV**	**V**	**V**	**VI**	**VI**
Observations	24	24	24	24	24	24	24	24	24	24
Trend specification	I	I,T	I	I,T	I	I,T	I	I,T	I	I,T
EG Test (tau-stat)	0.03	0.06	0.02	0.06	0.05	0.03	0.01	0.05	0.02	0.02
PO Test (tau-stat)	0.05	0.04	0.01	0.05	0.03	0.05	0.02	0.05	0.08	0.04

Note: I: Intercept; T: linear trend. For EG and PO, the p-values associated with the τ-statistics of the tests are reported.

Table 4 shows that models estimated with OLS in levels always have stationary residuals. Therefore, it is not possible to reject the null hypothesis of absence of cointegration and the use of the raw differences used for the estimates shown in table 1 is therefore appropriate.

It should be noted that the conclusions reached on the basis of unitary root tests and, above all, from the cointegration analysis must be evaluated with caution due to the small sample size.

************************End Bait Manuscript************************