

Teaching Replication in Quantitative Empirical Economics

Jan H. Höffler¹

Chair of Statistics, University of Göttingen, Germany

working paper for presentation at the

World Economics Association online conference:
“The economics curriculum: towards a radical reformation”

In empirical economics, a twofold lack of incentives leads to chronic problems with replicability: For authors of empirical studies providing replicable material is not awarded in as much as publishing new irreproducible studies is. Neither is authoring replication studies.

We offer a strategy to set incentives for replicability and replication. By integrating replication studies in the education of young scholars, we raise the awareness for the importance of replicability among the next generation of researchers and ensure that a big number of scientists get incentives to write replication studies: credit points and the prospect of publications at least of working papers already during their time as students.

By raising the number of researchers involved in replication and by providing an infrastructure for sharing their information, on the one hand we help to lower the amount of work researchers need to put into making their studies replicable. On the other hand, we facilitate the dissemination of insights derived from replication studies. This as a side effect imposes a significant threat of detection of irreproducible research, following the cases of recently introduced wiki projects for the revelation of plagiarism. In contrast to previous efforts like the report on the American Economic Review Data Availability Compliance Project, with our project we build the basis for the first replicable review paper on replicability as we give account of which studies were tested and which results were found in each case. After exploring several dozen studies published in highly ranked journals, we have not yet determined a single case where we see replicability is fully ensured. We identified two main problems: First, not all published results can be obtained from the replication material provided. Second, information about how the used data were obtained from the raw data is hardly ever sufficient.

For our investigation, we gave seminars at several faculties. We set up a wiki project for documenting the results of our replications as well as those found in the literature. In our database, we provide information about more than 1500 empirical studies, especially with regards to the availability of material for their replication.

We invite for discussion to develop standards for how to make research replicable and how to write replication studies. For this we provide information about existing projects that facilitate the sharing of material for empirical econometric research.

Keywords

replication, economic education, documentation, data archiving, statistical software

JEL Codes: A12, A20, C18, C80, Y80

¹ Platz der Göttinger Sieben 5, 37073 Göttingen, Germany

phone: ++49-(0)551-39-7406

fax: ++49-(0)551-39-7279

jhoeffl@uni-goettingen.de

Introduction

Replication refers to the duplication of published results (McCullough et al. 2006).

While in other sciences replicability is regarded as a fundamental principle for research and a prerequisite for the publication of results, in empirical economics it is still not treated as a top priority. Results are based on data and calculations that usually do not get published and that are not routinely controlled. Although there have already been warnings in the literature for decades about the dangers of the neglect of replication, the big picture has not changed much. It should be the standard, not the exception, however, only a small minority of journals in economics have introduced policies that should help to ensure the replicability of their published results. Even in the few cases of mandatory online archives for data and code used for calculations, replicability cannot be taken for granted at all (McCullough et al. 2006). McCullough and Vinod (2003) put it very clearly: “Research that is not replicable is not science and cannot be trusted either as a part of the profession’s accumulated body of knowledge or as a basis for policy.” This can be illustrated with the drastic case of social psychologist D. A. Stapel who, according to his home institution, the University of Tilburg, fabricated data for several dozen of his publications ([Stapel Investigation](#) and, e.g., Stapel et al. 2011a), which he and his supposedly innocent co-authors then had to retract (Retraction Watch 2013, and, e.g., Stapel et al. 2011b).

In empirical economics, a twofold lack of incentives leads to chronic problems with replicability: For authors of empirical studies the workload needed to make their material replicable is not awarded in the same way as publishing new irreproducible studies is. Neither is authoring replication studies.

The Replication Initiative

We offer a strategy to set incentives for replication and replicability. By integrating replication studies in the education of young scholars, we raise the awareness for the importance of replicability among the next generation of researchers and ensure that a big number of scientists get incentives to write replication studies: credit points for their studies and the prospect of publications at least of working papers already during their time as students (Wohlfarth 2012, Zakula 2012).

By raising the number of researchers involved in replication and by providing an infrastructure for sharing their information, on the one hand we help to lower the amount of work researchers need to put into making their studies replicable. On the other hand, we facilitate the dissemination of insights derived from replication studies. This as a side effect imposes a significant threat of punishment through embarrassment for irreproducible research, following the cases of recently introduced wiki projects for the revelation of plagiarism (<http://de.guttenplag.wikia.com>, <http://freyplag.wikia.com>). In contrast to previous efforts like the report on the American Economic Review Data Availability Compliance Project (Glandon 2010), with our project we provide the basis for the first replicable review paper on replicability as we give account of which studies were tested and which results were found in each case. After exploring several dozen studies published in highly ranked journals we have not yet determined a single case where we see replicability is fully ensured.

Our procedure is as follows: First we make a selection of studies for each course, depending on the knowledge of the students with regards to methodology, statistical software,

and previous knowledge in the fields of specialisation of the studies.² For most students it will be difficult enough to work with the same software as in the original study and to use the code published by the authors. However, some may find it interesting and possible to work with different software or to write their own code. The Journal of Applied Econometrics does not ask its authors to provide code, and it is a valid point that just working with the same code again bears the disadvantage that errors may be overlooked and often times those who replicate will not fully understand the details of the calculations and may not even notice this. As a first step it is however useful to work with code that is already there in order to see how such code is written by experienced professionals and in order to learn from this.

After selecting their topic according to their preferences, the students start with checking the material available for replication: Is the data set complete? Is the data sufficiently well described? Can it be tracked back to the raw data? It is important to do this at the very beginning of the course because some students may realise that they lack the material necessary for replication. If so they should carefully describe why this is the case, and this should be rewarded for their grade. Then they should select a new study for replication.

Students should write a short summary of the original paper and describe the data, always focusing on the replicability: Availability of raw data, completeness and quality of description of the data set. An overview of means and variances is helpful to get a better understanding of the data.

Looking at the programming code, students should show whether it is sufficient to replicate all published results, including graphs and results mentioned in the text but not in tables. They should describe whether it was clearly indicated which part of the code produced which results and whether the code is commented sufficiently.

For the interpretation, students should explain if results they may have found different from the publication still allow the conclusions drawn by the authors. They should analyse if the applied methods are useful in identifying the effects of relevance and if robustness checks are sufficient. They should draw their own conclusions if they are convinced by the findings of the original study. If possible they should investigate if they can confirm the results or call them into question using different data or methods.

Our main results so far can be summarised as such: Only few journals have a policy for all empirical quantitative studies they publish to archive data and code as well as instructions on how to use them. Even for the cases in which such policies are in place, they are usually not consequently enforced. And even for journals that have taken measures to enforce such policies we often could not obtain all the published results from the material provided (Appler 2012, Altinkaya 2012, Becker 2012, Cyrus 2012, Heidt 2012, Höfer 2013, Horstmann 2012, Mai 2012, Meyer 2012, Renner 2011, Richter 2012, Schneider 2009a, Schneider 2009b, ter Braak 2009, Weißer 2012, Winnige 2012³, Wohlfarth 2012, Zakula 2012). Amongst others, we were confronted with deviations in the number of observations from those published (Zhou 2011), deletion of identifying variables (Rempel 2011), unavailability of software extensions required to run the underlying code (Shaheen 2013), unavailability of datasets in the version used for the original study (Renner 2011). Some of the difficulties we encountered may in the end turn out to be rooted more in the limitations of our own abilities rather than in

² In the journals we looked at the software Stata was used by far most often. As this is also the software most of our students are familiar with, we nearly exclusively worked with this software so far. As previous studies have shown, different software can produce different results (McCullough and Vinod 2003). We thus would appreciate finding further partners for cooperation who could contribute expertise in using different software with their students. We consider it important to make replication possible for a large number of those interested. For this purpose we find it helpful to use free software such as R.

³ Unpublished working papers will be published upon acceptance of this manuscript.

the neglect of authors or editors to ensure replicability of their research. But not all of them. In one case an author even refused to provide code to us that by journal policy he and his co-authors would have been obliged to submit to the web archive.⁴

For journals that do not follow a data availability policy our experience is mixed: In some cases authors nicely provide help that allow replication to some extent (Enenkel 2011, Liebrand 2012). However, the transformation of raw data to the final dataset was not transparent (Enenkel 2011), and authors in these cases may use software for which code is often not saved in a way necessary for direct replication (Liebrand 2012).

We had the support of our home university's Teaching Centre and could draw on its resources for research oriented teaching to bachelor students since we won their award for research oriented teaching and learning three times. It should not be underestimated that teaching bachelor students to try to replicate well published empirical econometric studies requires intensive supervision and a lot of patience. Students need to be instructed carefully in order to avoid as much frustration as possible, even though some is inevitable: Bachelor students will usually not be able to understand every detail of such studies, and they should know that this is neither expected nor necessary. In order to investigate transparency it is often sufficient to show that not all published results can be obtained with the archived replication material – which is not exactly a result a typical student feels enthusiastic about. Students need to be guided such that they do not make unrealistic plans on how to extend on existing studies that can only lead to disappointment. And it is not helpful to the discipline if students who do not manage to replicate empirical studies start sending angry emails to the authors, particularly in cases in which the “lack of replicability” lies in the lack of economic education of the student. In our experience replication, especially when done in groups that interact well, can benefit students by allowing them to better understand how the body of knowledge evolves through empirical research. Tendencies to take every empirical result presented to them as a “proof” of anything will not persist. On the other hand, replication can help to show that it is too simple to say that “There are three kinds of lies: Lies, damned lies, and statistics” (Twain 1906). With our concept and the material we had compiled we could convince professors at several other faculties for cooperation: At the University of Toronto, the University of Bonn, at our own university's Faculty of Social Sciences, and at the Graduate Institute, Geneva. A replication seminar was introduced for the MAGKS inter-university graduate program comprised of six German universities - Philipps University Marburg, RWTH Aachen, Justus-Liebig-University Gießen, Georg-August-University Göttingen, the University of Kassel and the University of Siegen. We give a PhD seminar at Nanjing Agricultural University. The exchanges turned out to be very fruitful and we gained many insights that we could not have gotten otherwise. For documenting the results of our replications as well as those found in the literature we set up a wiki project. In our database we provide information about more than 1400 empirical studies, especially with regards to the availability of material for their replication.⁵ We want to contribute to the identification of studies the scientific community regards as especially important for replication as is already practised by the 3ie project for the replication of impact evaluations in development economics (<http://www.3ieimpact.org/en/evaluation/impact-evaluation-replication-programme>).

⁴ As an exception we do not name the respective study here as we do not have written proof of the authors' refusal.

⁵ A smaller version of this wiki will be made accessible to all members of the conference upon acceptance of this manuscript: <http://replications.wiki.gwdg.de>

Those who cooperate with us in teaching replication get access to the full wiki which we plan to make available to the public by October 2013.

Finally, we provide instructional videos on replication.⁶ Our teaching resources can be freely used by any interested institutions or individuals such that everyone can participate in the improvement of replicability in empirical econometrics.

Journals investigated

We so far focus on six journals that provide data of empirical studies in online archives. Five journals of the American Economic Association that all follow a similar data availability policy: The American Economic Review (Bernanke 2004)⁷ and the American Economic Journals (Macroeconomics, Microeconomics, Applied Economics and Economic Policy). Furthermore, the Journal of Political Economy that adopted the American Economic Review's data availability policy.⁸ It is unclear to us why the Papers and Proceedings of the American Economic Review should be held to a lesser scientific standard and are exempted from the data availability policy of the American Economic Review. The Biometrical Journal is the only journal we are aware of that has a Reproducible Research Editor who checks the replicability of published results at least from the material provided, even though that journal does not require its authors to contribute to its data archive.⁹ Students, even PhD students at prestigious faculties, to our experience are not usually aware of the fact that reviewers do not routinely check the results of the studies they referee. Another problem to us seems even more widespread - and more difficult to solve: Even for the few journals that ask their authors to provide information about how the final data that get archived were obtained from the raw data, no or by far insufficient details are supplied. Material for replication submitted voluntarily by authors of studies published in the American Economic Review Papers and Proceedings did not fulfil the requirements of replicability from raw data (Sailer 2011, Cyrus 2012). There seems to be a lack of standards for issues like how to document data cleaning and the merging of different data sets. Furthermore, many institutions that provide data frequently change datasets and do not archive each version of them. It seems promising to us that the project [Data Cite](#) introduces the [digital object identifier](#) system also for datasets, and we hope this will become an established standard.

The leading journal with a regular section focusing on replication of published studies is the Journal of Applied Econometrics (Pesaran 2003). As this journal typically publishes technically more demanding studies and it does not require its authors to archive the code they used to obtain their results we considered their material as insufficient at least for our bachelor students. The replications published however inspired us and we regarded that journal's data archive as such a valuable resource that we included most of the studies published in our wiki dataset.

In our wiki, we give an overview about journal policies on replication. To this date, only a minority of journals have introduced mandatory online archives for data and code used for quantitative empirical studies (Huschka and Wagner 2012). To our knowledge no journal has found a convincing strategy to achieve transparency of data cleaning and to deal with other issues concerning the manipulation of raw data. Very few journals regularly publish replication studies, most prominently the Journal of Applied Econometrics. We do not know of any other systematic collection of such replication studies that is accessible to the public. This project requires the collective work of the community since a large amount of replication

⁶ A first video is already available in German – a presentation with an introduction to replication in general and to our project (Kneib 2012).

⁷ <http://www.aeaweb.org/aer/data.php>

⁸ <http://www.press.uchicago.edu/journals/jpe/datapolicy.html?journal=jpe>

⁹ <http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291521-4036/homepage/EditorialBoard.html>

work is impossible to find in acceptable time for any single group given that the results are often just mentioned as asides of published studies. Especially when it comes to reproductions, i.e., empirical work on the same question as in a previous study but with different data or methodology, specific expertise is needed for each subfield of economics in order to assess the results in the context of the existing literature. For the above described reasons we invite to further discussion about how policies should be designed in order to ensure replicability.

Related Research

In order to develop standards for how to make research replicable and how to write replication studies we provide information about existing projects that facilitate the sharing of material for empirical econometric research like the Harvard dataverse (Crosas 2011) or the [runmycode](#) page for sharing code and data and enabling readers to run code with a cloud technology (Stodden et al. 2012).

We document projects from related disciplines that focus on replication and from which economists can learn, like the replication project in psychology that collects information about replication studies (Spellman 2012, <http://psychfiledrawer.org>). Especially with regards to teaching the psychologists' perspective is very helpful to us (Frank and Saxe 2012).

Conclusion and Outlook

Much remains to be done to ensure full replicability of quantitative empirical research. Those journals that have data availability policies need to enforce them, and such policies need to become a universal standard in our discipline. Analogously, standards are needed for how to proceed for replication. In order to lower the amount of work every journal has with establishing such standards there should be a common effort to find them. And the results of replication studies need to be documented such that it becomes easy to find and compare them. The initiative may be better taken away from journals and their editors as they may be subject to conflicts of interest (Laband and Piette 1994). So crowd-sourcing that enables everyone to comment and make a contribution to us seems an approach that avoids such conflicts.

In our further research we plan to investigate how the degree of replicability of published studies influences the citations they get. To us it seems plausible that datasets of well replicable research will be used by other scientists, too, who will then cite the original work. If this is the case, we could show that there is another incentive to make research replicable, both for authors and journals.

References:

Altinkaya, Hatice 2012, 'Replication of Meghir and Palme (2005): Educational Reform, Ability, and Family Background', unpublished working paper.

Appler, Felix 2012, 'Replication of Magruder (2010): Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa', unpublished working paper.

Becker, Daniel 2012, 'Replication of Cipollone and Rosolia (2007) «Social Interactions in High School: Lessons from an Earthquake»', unpublished working paper.

- Bernanke, Ben 2004, 'Editorial Statement', *American Economic Review*, 94 (1), 404.
- Coelho, Philip R. P., Frederick De Worken-Eley III, James E. McClure 2005, 'Decline in Critical Commentary, 1963-2004', *Econ Journal Watch*, 2(2), 355-61.
- Coelho, Philip R. P., James E. McClure 2006, 'Why Has Critical Commentary Been Curtailed at Top Economics Journals? A Reply to Robert Whaples' *Econ Journal Watch*, 3(2), 283-91.
- Crosas, Mercè 2011, 'The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data', *D-Lib Magazine*, 17(1/2).
- Cyrus, Philipp 2012, 'Replication of Field, Jayachandran, and Pande (2010): Do Traditional Institutions Constrain Female Entrepreneurship? A Field Experiment on Business Training in India', unpublished working paper.
- Enenkel, Kathrin 2011, 'Replication of Gould and Klor (2010): Does terrorism work?', unpublished working paper.
- Frank, Michael C., Rebecca Saxe 2012, 'Teaching Replication', *Perspectives on Psychological Science*, 7(6), 600-4.
- Glandon, Philip 2011, 'Appendix to the Report of the Editor: Report on the American Economic Review Data Availability Compliance Project', *American Economic Review: Papers & Proceedings*, 101(3), 695-9.
- Heidt, Lea 2012, 'Replication of Kuhn, Kooreman, Soetevent, and Kapteyn (2011) «The Effects of Lottery Prizes on Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery»', unpublished working paper.
- Höfer, Susanna 2012, 'Replication of Oltmans Ananat (2011) «The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality»', unpublished working paper.
- Horstmann, Elaine 2012, 'Replication of Lucas (2010) «Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka»', unpublished working paper.
- Huschka, Denis, Gert G. Wagner 2012, 'Data accessibility is not sufficient for making replication studies a matter of course', RatSWD Working Paper Series, No. 194.
- Kneib, Thomas 2012, 'Replication in Empirical Economics', presentation (in German) at the Pluralist Event in addition to the Annual Conference of the German Economic Association, organized by the committee Real World Economics: <http://www.real-world-economics.de/index.php?id=108>
- Laband, David N., Michael J. Piette. 1994, 'Favoritism versus search for good papers: Empirical evidence regarding the behavior of journal editors' *Journal of Political Economy*, 102(1), 194-203.
- Liebrand, Andreas Tonio 2012, 'Replicating: The Value Premium and Time-Varying Volatility', unpublished working paper.
- Mai, Linda 2012, 'Replication of Imai and Takarabe (2011) «Bank Integration and Transmission of Financial Shocks: Evidence from Japan»', unpublished working paper.

McCullough, B.D., K. A. McGeary, T. D. Harrison 2006, 'Lessons from the JMCB Archive', *Journal of Money Credit and Banking*, 38(4), 1093-107.

McCullough, B.D., Hrishikesh D. Vinod 2003, 'Verifying the Solution from a Nonlinear Solver: A Case Study', *American Economic Review*, 93(3), 873-92.

Meyer, Ann-Kathrin 2012, 'Replication of Pettersson-Lidbom (2010) «Dynamic Commitment and the Soft Budget Constraint: An Empirical Test»', unpublished working paper.

Pesaran, Hashem 2003, 'Introducing a replication section', *Journal of Applied Econometrics*, 18(1), 111.

Rempel, Mark 2011, 'Replication of Ferraz and Finan (2011) «Electoral Accountability and Corruption: Evidence from Audits of Local Governments»', unpublished working paper.

Renner, Maren 2011, 'Replication of Luttmer and Singhal (2011): Culture, Context, and the Taste for Redistribution', unpublished working paper.

Retraction Watch 2013, 'Retraction 46 arrives for Diederik Stapel', <http://retractionwatch.wordpress.com/2013/02/05/retraction-46-arrives-for-diederik-stapel>

Richter, Markus 2011, 'Replication of Alesina and Fuchs-Schündeln (2007) «Good-Bye Lenin (or Not?): The Effect of Communism on People's Preferences»', unpublished working paper.

Sailer, Arne 2011, 'Replication of Bergin, Feenstra, and Hanson (2009) «Offshoring and Volatility: Evidence from Mexico's Maquiladora Industry»', unpublished working paper.

Schneider, Jan-Peter 2009a, 'Replication of Post, van den Assem, Baltussen, and Thaler (2007): Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show', unpublished working paper.

Schneider, Jan-Peter 2009b, 'Replication of Grove and Wu (2007) «The Search for Economics Talent: Doctoral Completion and Research Productivity»', unpublished working paper.

Shaheen, Sondos 2013, 'Replication of Esteban, Mayoral and Ray (2011) «Ethnicity and Conflict: An Empirical Study»', unpublished working paper.

Spellman, Barbara A. 2012, 'Introduction to the Special Section on Research Practices', *Perspectives on Psychological Science*, 7(6), 655-6.

Stapel, Diederik A., und Siegwart Lindenberg 2011a, 'Coping with Chaos: How Disordered Contexts Promote Stereotyping and Discrimination', *Science*, 332 (6026), 251-253.

Stapel, Diederik A., und Siegwart Lindenberg 2011b, 'Retraction of Stapel and Lindenberg, *Science* 332 (6026) 251-253', *Science* 334 (6060), 1202.

Stapel Investigation: Joint Tilburg/Groningen/Amsterdam investigation of the publications by Mr. Stapel, <https://www.commissielevelt.nl>

Stodden, Victoria, Christophe Hurlin, Christophe Prignon 2012, 'RunMyCode.Org: A Novel Dissemination and Collaboration Platform for Executing Published Computational Results', SSRN Working Paper: <http://dx.doi.org/10.2139/ssrn.2147710>

ter Braak, Daniel 2009, 'Replication of Brender and Drazen (2008): How Do Budget Deficits and Economic Growth Affect Reelection Prospects? Evidence from a Large Panel of Countries', unpublished working paper.

Twain, Mark (1906), 'Mark Twain's Own Autobiography: The Chapters from the North American Review', ed. Michael J. Kiskis, Wisconsin Studies in Autobiography, University of Wisconsin Press.

Weißer, Christoph 2012, 'Replication of Obstfeld, Shambaugh, and Taylor (2010): Financial Stability, the Trilemma, and International Reserves', unpublished working paper.

Whaples, Robert 2006, 'The Costs of Critical Commentary in Economics Journals', *Econ Journal Watch* 3(2), 275-282.

Winnige, Stefan 2012, 'Replication of Thornton (2008): The Demand for, and Impact of, Learning HIV Status', unpublished working paper.

Wohlfarth, Paul 2012, 'Replication in the narrow sense of Banzhaf/Walsh (2008)', Replication Working Paper No. 2. http://zfs.uni-goettingen.de/uploads/media/Replication_WP_02_2012.pdf

Zhou, Zhili 2011, 'Replication of Duflo, Dupas, and Kremer (2011) «Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation on Kenya»', unpublished working paper.

Zakula, Björn 2012, 'Narrow Replication of Ashcraft (2005): Are Banks Really Special?', Replication Working Paper No. 1. http://zfs.uni-goettingen.de/uploads/media/Replication_WP_01_2012_01.pdf

The research for this project was [funded by the Institute for New Economic Thinking \(INET\)](#).

Appendix: Reading List and exam guidelines of the 2013 Course “Advanced Econometrics” taught together with Lore Vandewalle at the Graduate Institute, Geneva.

Reading List “Advanced Econometrics”
2013, Graduate Institute, Geneva

Prof Lore Vandewalle and Jan H. Höffler

PAPER 1: Matching used in robustness check (Diff in diff is the main methodology)

Harrison, A., and J. Scorse (2010), Multinationals and Anti-sweatshop Activism, *American Economic Review*, 100(1): 247-73.

OR

Fowlie, Holland and Mansur (2012), What Do Emissions Markets Deliver and to Whom? Evidence from Southern California’s NOx Trading Program, *American Economic Review*, 102(2): 965-993.

PAPER 2: Regression Discontinuity Designs used in robustness check (IV is the main methodology)

Clots-Figueras, I. (2012), Are Female Leaders Good for Education? Evidence from India, *American Economic Journal: Applied Economics*, 4(1): 212-44.

IMPORTANT: check if the data can be obtained for free! If not, another paper will be allocated (just forward me the email of the author saying that the data cannot be given for free)

OR

Clark, D. (2007), The Performance and Competitive Effects of School Autonomy, *Journal of Political Economy*, 117(4): 745-783.

PAPER 3: Duration Models (and probit)

Iyer, R. and M. Puri (2012), Understanding Bank Runs: The Importance of Depositor-Bank Relationships and Networks, *American Economic Review*, 102(4): 1414-45.

OR

Costa, D. and M. Kahn (2007), Surviving Andersonville: The Benefits of Social Networks in POW Camps, *American Economic Review*, 97(4): 1467-87.

PAPER 4: Randomized Controlled Trials

Dupas, P. and J. Robinson (forthcoming), Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya, *American Economics Journal: Applied Economics*.

OR

Angrist, J., D. Lang and P. Oreopoulos (2009), Incentives and Services for College Achievement: Evidence from a Randomized Trial, *American Economic Journal: Applied Economics*, 1(1): 1-28.

PAPER 5: Instrumental Variables

Edmonds, E., N. Pavcnik and P. Topalova (2010), Trade Adjustment and Human Capital Investments: Evidence from Indian Tariff Reform, *American Economic Journal: Applied Economics*, 2(4): 42-75.

OR

Bosch, M. and M. Manacorda (2010), Minimum Wages and Earnings Inequality in Urban Mexico, *American Economic Journal: Applied Economics*, 2(4): 128-149.

Exam Advanced Econometrics: REPLICATION

Follow the instructions carefully!

Timeline exam

February 25: Email the composition of groups

You have to form groups of 4 students by February 25 and send an email to lore.vandewalle@iheid.ch with the composition of each of the groups. Each group will get one of the following econometric tools allocated (randomly):

- Matching
- RDD
- Duration models
- RCT
- Instrumental variables

For each of the tools we selected two highly published papers, out of which you can choose the one to be replicated. You can switch tools between groups based on mutual agreement. Do inform me in case you do so.

March 6: One page feedback on whether the paper is suitable for replication

You should make sure that the paper of your choice can be replicated:

- Is raw data available?
- Is the data set complete?
- Is the data sufficiently well described?

If you realised that one of the two papers cannot be replicated, you should carefully describe why this is the case (this will be rewarded for your exam!). If none of the papers can be replicated, we will send you another one.

May 22: Deadline seminar paper

One seminar paper has to be handed in per group. Explain in detail how you proceeded for the replication, and which conclusions you draw. Below are some suggestions about how to proceed:

- 1 Short summary of the original paper
- 2 Data
 - Is raw data available?
 - Is the data set complete?
 - Is the data sufficiently well described?
 - Make an overview of means and variances
- 3 Stata code
 - Are all results replicable?
 - Clearly indicate which part of the code produced which results?
 - Are the codes commented sufficiently?
- 4 Interpretation
 - If the results you find are different, does it change the interpretation?
 - Is the applied method useful in identifying the effects of relevance?
 - Are the described robustness checks conducted?
 - Are the conclusions drawn in the paper convincing?
- 5 Can you confirm the results or call them into question using different data or methods?

Examples on how to write a replication study can be found at:

- “Instructions for authors” on journal websites, e.g. by the Journal of Applied Econometrics
- Examples in the literature, e.g. the Albouy versus Acemoglu, Johnson and Robinson debate (syllabus). Other examples:
 - Fisher, A. C., W. M. Hanemann, M. J. Roberts, and W. Schlenker (2012), “The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment.” *American Economic Review*, 102(7): 3749-60.
 - Mueller, H. (2012), “Growth Dynamics: The Myth of Economic Recovery: Comment.” *American Economic Review*, 102(7): 3774-77.
 - Goettingen replication working papers
- Further information on replication and its methodology: <http://www.3ieimpact.org/en/evaluation/impact-evaluation-replication-programme>

May 29: Exam (evening)

Each group has to discuss the replication of another group (random allocation).

On the evening of the exam, each paper will be allocated 35 minutes: 15 minutes to present the replication, 5 to 10 minutes for the discussion by another group and 10 to 15 minutes for a class discussion. In between the third and the fourth presentation, we will take a 10 minutes break.

A note on the evaluation

It is hard to say in advance which papers will be easy to replicate (if you tried something that did not work out, do write it down in your paper.). Some papers will allow extensions, for others it will be impossible. All of these papers are highly published, so not finding any problem or new extension does not necessarily imply you did a bad job! The aim of this exercise is to provide a good replication: what matters is being as complete as possible in judging the paper.

The 6 points of the exam will be allocated as follows:

- Written paper: 4 points
- Clarity presentation: 1 point
- Discussion: 1 point