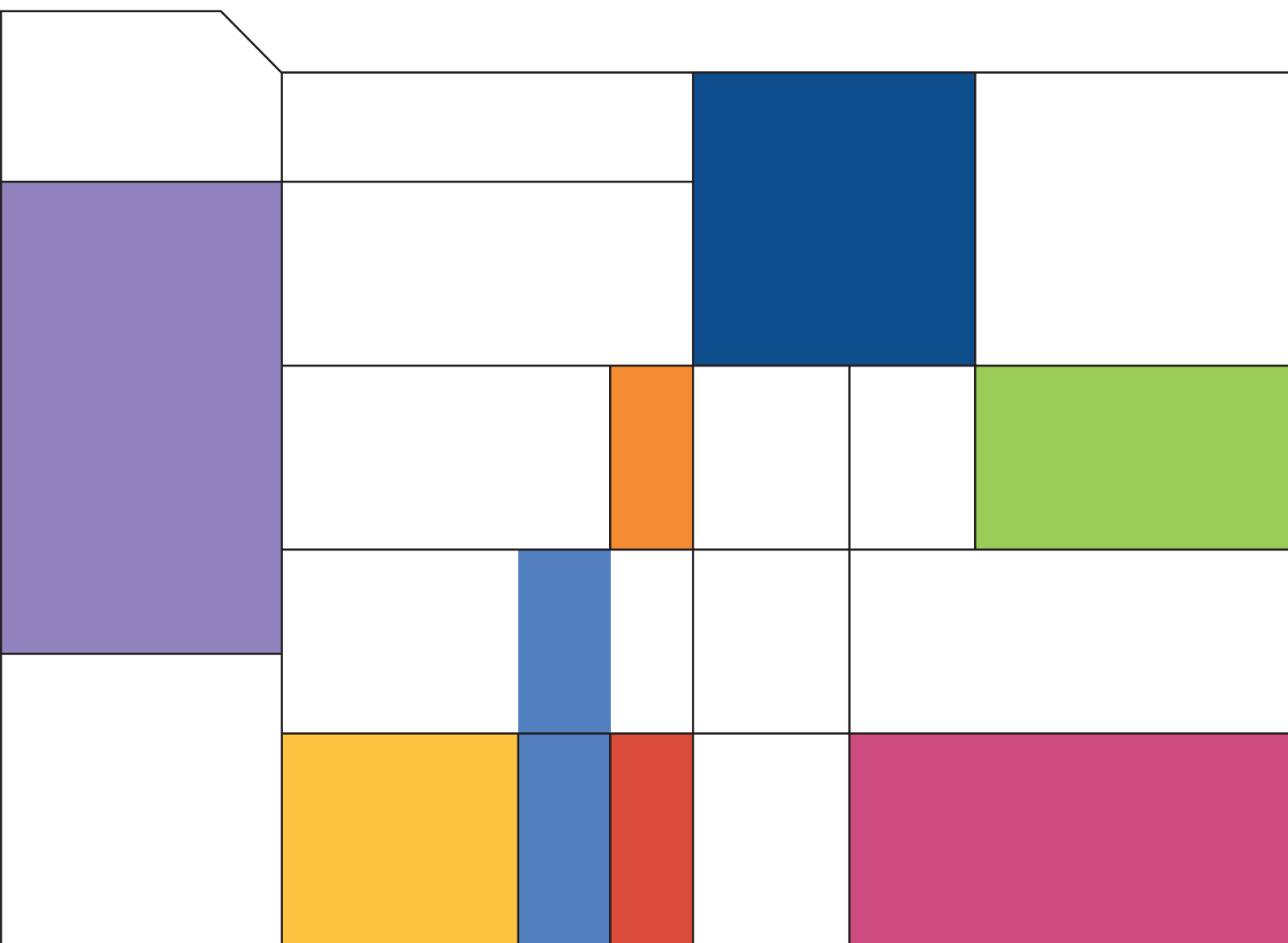




OECD Journal

Economic Studies



OECD Journal

Economic Studies

Volume 2016

The OECD Journal collection comprises four periodicals: *OECD Journal on Budgeting*, *OECD Journal: Journal of Business Cycle Measurement and Analysis*, *OECD Journal: Economic Studies* and *OECD Journal: Financial Market Trends*. Each periodical within *OECD Journal* retains its own ISSN, volume and issue numbering. The volume numbering of each journal corresponds to the calendar year. There will be one issue of *OECD Journal: Economic Studies* this year.

This work is published on the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

OECD Journal: Economic Studies, volume 2016
ISSN 1995-2848 (print)
ISSN 1995-2856 (online)
http://dx.doi.org/10.1787/eco_studies-v2016-1-en

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Note by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

Note by all the European Union member states of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Corrigenda to OECD publications may be found on line at: www.oecd.org/publishing/corrigenda.

© OECD 2017

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.

Foreword

OECD Journal: Economic Studies, an official publication of the Organisation for Economic Cooperation and Development (OECD) is aimed at a broad audience of policy-makers, academics and interested readers in the wider public. It publishes papers in different areas of economic policy analysis, generally based on empirical work with a cross-country dimension with the objective of contributing to the renewal of current approaches to economic issues and contributing to debate among economists and practitioners.

This volume addresses a rich set of topical issues through new analytical work. With the macroeconomic legacy from the recent crisis still apparent in many economies, a first group of articles considers the usefulness of early warning indicators and successful adjustment from crises with IMF programmes. A second group considers the outlook for long-term growth and the effectiveness of structural policies to boost it. A final set of articles, while all marshalling cross-country empirical evidence, reflect the diversity of topics covered in Economic Studies and include: an assessment of the effect of regional fragmentation on economic performance; the determinants of efficiency and contestability in banking systems; and an analysis of the role of selection in assessing student achievement.

All submitted papers are reviewed by the Board of Editors and, if deemed eligible, by two referees, one internal to the OECD and the other external. This guarantees that the papers that are accepted are consistent with both the standards that the OECD applies to its own work for the member countries and the requirements of academic excellence. Of course, the views expressed in the papers published remain the sole responsibility of the authors and should not be attributed to the OECD.

Articles are published online as they become available and assembled in an annual volume of articles released at the beginning of the following year.

Finally we would like to take this opportunity to thank Jean-Luc Schneider for his firm stewardship of the OECD Journal: Economic Studies over the years, his care and dedication will be greatly missed. We would also like to thank Catherine Brannigan Conway for her assistance in the preparation of this volume.

Editorial Board

Jean-Luc Schneider (Chairman), Deputy Director, OECD Economics Department, Paris

Sveinbjörn Blöndal, Head of Division, OECD Economics Department, Paris

Jorge Braga de Macedo, NOVA School of Business and Economics, Lisbon, NBER, Cambridge Mass. and CEPR, London; former president of the OECD Development Centre, Paris

Fabrizio Coricelli, Professor of Economics, Paris School of Economics, Université Paris 1 and Research Fellow, Centre for Economic Policy Research (CEPR), London

Chiara Criscuolo, OECD Directorate for Science, Technology and Industry, Paris

Anna D'Addio, OECD Directorate for Employment, Labour and Social Affairs, Paris

James K. Hammitt, Harvard University (Center for Risk Analysis), United States and Toulouse School of Economics (LERNA-INRA)

Bo Hansson, OECD Directorate for Education, Paris

Bo E. Honoré, Professor of Economics and Director of the Gregory C. Chow Econometric Research Program, Princeton University, US; Member of the Board of Trustees of the Danish National Research Foundation

Sébastien Jean, Centre d'études prospectives et d'informations internationales (CEPII), Paris, France. Policy associate and member of the Strategic Advisory Board, Leverhulme Centre for Research on Globalisation and Economic Policy (GEP), University of Nottingham

Francis Kramarz, Professeur École Polytechnique, Directeur, Centre de Recherche en Économie et Statistique (CREST) (Research Center for Economics and Statistics), Institut National de la Statistique et des Études Économiques (INSEE) (National Institute of Statistics and Economic Studies), Paris

Joaquim Oliveira Martins, Head of Division, OECD Public Governance and Territorial Development Directorate and Associate Professor, University Paris-Dauphine, Paris

Raed Safadi, Deputy Director, OECD Trade and Agriculture Directorate, Paris

Vincenzo Spezia, OECD Directorate for Science, Technology and Industry, Paris

David Turner, Head of Division, OECD Economics Department, Paris

Gert Wehinger, OECD Directorate for Financial and Enterprise Affairs, Paris

Catherine Brannigan Conway, OECD Economics Department, Paris

Acknowledgement

The Editorial Board would like to thank the following referees who kindly reviewed articles during the preparation of this volume.

Aida Caldera Sanchez, Senior Economist, Spanish Desk, Economics Department, OECD
Gilbert Cette, Director of the Microeconomic and Structural Analysis, Banque de France, Paris
Lewis Dijstra, Deputy Head of the Analysis Unit in the Directorate-General for Regional Policy of the European Commission, Brussels

Stuart Elliott, formerly of the OECD Directorate for Education and Skills, OECD, Paris
Miguel Ferreira, Banco BPI Chair in Finance at the Nova School of Business and Economics, Lisbon

Francesco Franco, Assistant Professor at the Nova School of Business and Economics, Lisbon
Pedro Gonzaga, Policy Analyst in the Directorate for Financial and Enterprise Affairs

Yvan Guillemette, Senior Economist in the Macroeconomic Analysis Division of the Economics Department, OECD, Paris

Eric A. Hanushek, Paul and Jean Hanna Senior Fellow at the Hoover Institution of Stanford University, Stanford, California

Edward Leamer, Chauncey J. Medberry Chair in Management/Professor in Economics & Statistics, UCLA, Los Angeles, California

Fabrice Murtin, Head of Section, Statistics Directorate, OECD, Paris

Mathilde Pak, Economist, Economics Department, OECD, Paris

Alexander Schiersch, Research Associate, Deutsches Institut für Wirtschaftsforschung, Berlin

Vincenzo Spiezia, Senior Economist, Directorate for Science, Technology and Innovation, OECD, Paris, OECD

David Turner, Head of Macroeconomic Analysis Division, Economics Department, OECD, Paris

Marco Vivarelli, Professor of Economics at the Università Cattolica del Sacro Cuore, and Director of the Institute of Economic Policy, Milan

Gert Wehinger, Principal Administrator, Directorate for Financial and Enterprise Affairs, OECD, Paris

Table of contents

Economic resilience: The usefulness of early warning indicators in OECD countries . . .	9
<i>by Mikkel Hermansen and Oliver Röhn</i>	
Fortune or fortitude? Determinants of successful adjustment with IMF programs . .	37
<i>by Martin Larch, Kristin Magnusson Bernard and Peter McQuade</i>	
Long-term growth and productivity projections in advanced countries.	71
<i>by Gilbert Cette, Rémy Lecat and Carole Ly-Marin CLM</i>	
The quantification of structural reforms in OECD countries: A new framework. . . .	91
<i>by Balázs Égert and Peter Gal</i>	
Administrative fragmentation and economic performance of OECD TL2 regions . . .	109
<i>by David Bartolini</i>	
National testing policies and educator based testing for accountability:	
The role of selection in student achievement	131
<i>by William C. Smith</i>	
Efficiency and contestability in emerging market banking systems.	151
<i>by Christian Daude and Julien Pascal</i>	

Economic resilience: The usefulness of early warning indicators in OECD countries

by
Mikkel Hermansen and Oliver Röhn*

The global financial crisis and the high associated costs have revived the academic and policy interest in “early warning indicators” of crises. This paper provides empirical evidence on the usefulness of a new set of vulnerability indicators, proposed in a companion paper (Röhn et al., 2015), in predicting severe recessions and crises in OECD countries. To evaluate the usefulness of the indicators the signalling approach is employed, which takes into account policy makers’ preferences between missing crises and false alarms. Our empirical evidence shows that the majority of indicators would have helped to predict severe recessions in OECD economies between 1970 and 2014. In the domestic areas, indicators that measure asset market imbalances (real house and equity prices, house price-to-income and house price-to-rent ratios), perform consistently well both in and out-of-sample. Domestic credit related variables appear particularly useful in signalling upcoming banking crises and in predicting the global financial crisis out-of-sample. Indicators of global risks consistently outperform domestic indicators in terms of their usefulness, highlighting the importance of taking international developments into account when assessing a country’s vulnerabilities. The good performance of the global indicators is however subject to a caveat: they are particularly suited to pick up recessions that affect a large number of countries simultaneously, such as the global financial crisis in 2008/09. The results are broadly robust to different definitions of costly events, different forecasting horizons and different time and country samples.

JEL classification: E32; E44; E51; F47

Keywords: Resilience, early warning indicators, vulnerabilities, imbalances, severe recessions, crises

* The authors are members of the Economics Department of the OECD. They would like to thank Aida Caldera Sanchez, Alain de Serres, Edward Leamer, Morten Rasmussen, Jean-Luc Schneider, and David Turner for helpful comments and Caroline Abettan for technical and editorial assistance. The views expressed in this paper are those of the authors and do not necessarily reflect those of the OECD and its member countries.

1. Introduction and main findings

The global financial crisis and the high costs associated with it have revived the academic and policy interest in “early warning indicators” of crises (e.g. Rose and Spiegel, 2011; Frankel and Saravelos, 2012; Alessi and Detken, 2011; Lo Duca and Peltonen, 2013 among many others). The early warning literature typically defines a crisis as the result of vulnerabilities and a trigger event. Despite recent methodological improvements of early warning models, predicting the timing of crises remains an extremely difficult task. Early warning models’ most important value-added therefore is to identify variables (or “early warning indicators”) that should be monitored to detect vulnerabilities, rather than to predict the exact start date of the next crisis. Vulnerability indicators can thus be a valuable input for monitoring economic risks, but should be complemented with other monitoring tools, including expert judgement.

The new wave of research has led many national and international institutions (e.g. European Commission and the International Monetary Fund) to develop their own set of vulnerability indicators and early warning models in the last few years. The OECD has also started to systematically monitor and publish indicators of potential macroeconomic and financial vulnerabilities in the *Economic Outlook* (EO) since the end of 2013 and more recently as part of country Economic Surveys.

The paper extends these OECD efforts by providing empirical evidence on the usefulness of a large set of vulnerability indicators in predicting costly events, measured as severe recessions. The paper draws on a set of vulnerability indicators proposed in a companion paper (Röhn et al., 2015). In Röhn et al. (2015) more than 70 vulnerability indicators are identified as particularly relevant for OECD countries based on a thorough review of the most recent evidence from the early warning literature and lessons learned from the global financial crisis. The indicators are grouped into six areas: i) financial sector imbalances, ii) non-financial sector imbalances, iii) asset market imbalances, iv) public sector imbalances, v) external sector imbalances, and vi) spillovers, contagion and global risks. The contribution of this paper is to assess the informational content of these vulnerability indicators for which sufficiently long time series exist by presenting both in-sample and out-of-sample results. The paper also complements earlier OECD research, which has focused on the impact of the size and composition of capital inflows on a country’s risk of suffering banking and currency crises or sudden stops (Furceri et al., 2011; Ahrend and Goujard, 2012), by considering a wider range of vulnerability areas and bad economic outcomes, including severe recessions, banking, currency and sovereign debt crises.

We focus on predicting severe recessions which is a novelty compared to most of the early warning literature, which has typically focused on particular types of economic crises, such as currency (e.g. Kaminsky et al., 1998) or banking crises (e.g. Demirgüç-Kunt and Detragiache, 2000) and more recently broader systemic financial events (e.g. Alessi and Detken, 2011; Lo Duca and Peltonen, 2013). This choice is motivated by two aspects: First, large drops in GDP per capita provide an efficient way to capture a range of costly economic

events and represents an outcome that policymakers are presumably most concerned to avoid. Second, severe recessions can be transparently defined and thus overcome the difficulty of identifying economic crises in an objective way.

The usefulness of the indicators is assessed on the basis of the signalling approach, one of the most commonly used early warning methodologies (e.g. Kaminsky et al., 1998; Borio and Lowe, 2002; Behn et al., 2013). The advantage of the signalling approach is that it can accommodate differences in data availability across countries and allows for the inclusion of a potentially larger number of vulnerability indicators than alternatives based on multivariate regression methods. This advantage is important because the aim of this study is to assess the predictive ability of each individual indicator rather than to devise a composite early warning indicator. According to the signalling approach, an indicator signals a vulnerable state of the economy if it crosses a threshold. Threshold levels are chosen so as to strike a balance between the risks of missing vulnerable states (so-called type I errors) and issuing many false alarms (so-called type II errors). In particular, a loss function is used to determine the optimal thresholds, which explicitly takes into account policymaker preferences between type I and type II errors. An indicator is labelled useful if its predictions result in a lower loss compared to a benchmark in which the indicator is ignored.

The main findings can be summarised as follows:

- The majority of vulnerability indicators appear to be useful early warning indicators for severe recessions when policymakers are strongly averse to missing severe recessions. Most indicators issue first warning signals on average more than 1.5 years before the onset of a severe recession, providing policymakers with a sufficiently long lead to react. However, the extent of the signalling power varies across indicators and the results are sensitive to the exact specification of policymakers' preferences between missing crises and false alarms.
- In the domestic areas, indicators that measure asset market imbalances (real house and equity prices, house price-to-income and house price-to-rent), perform consistently well both in and out-of-sample. Domestic credit related variables appear particularly useful out-of-sample and in signalling upcoming banking crises. The usefulness of indicators of external imbalances such as current account balances, official reserves and foreign currency exposure perform well in certain specifications. Fiscal imbalances are generally not found to be useful in signalling severe recessions and crises.
- Indicators of global risks consistently outperform domestic variables in terms of relative usefulness. In particular, measures of the global credit-to-GDP ratio (growth and gaps from a trend), a global equity price gap and a global house price gap perform particularly well both in-sample and out-of-sample. This highlights the importance of taking international developments into account when assessing a country's vulnerabilities. In an increasingly integrated world economy, vulnerabilities that build-up on the global level potentially transmit to countries around the world. The good performance of the global indicators is however subject to a caveat: as the indicators do not vary across countries they are particularly suited to pick up recessions that affect a large number of countries simultaneously, such as the global financial crisis in 2008/09.
- The results are broadly robust to different definitions of costly events (severe recessions versus defined economic crises, including banking, currency and sovereign debt crises), different forecasting horizons and different time and country samples.

The paper proceeds as follows. Section 2 discusses the data, including the choice of severe recessions as the dependent variable and the vulnerability indicators. Section 3 outlines the empirical approach and section 4 presents the results.

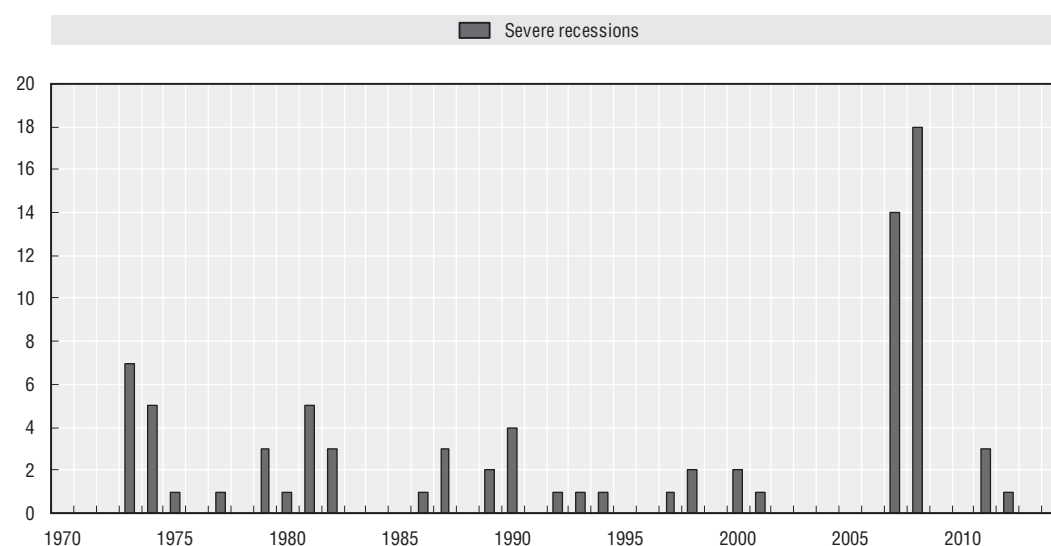
2. Data

2.1. Severe recessions as a measure of costly economic events

For the baseline evaluation, costly economic events are defined as severe recessions, putting the focus on tail events (left tail) as opposed to regular business cycle fluctuations. First, the Bry and Boschan (1971) algorithm is applied to identify peak and trough dates of business cycles in the level of GDP per capita over the period 1970-2014 for all 35 OECD countries. Next, severe recessions are defined as recessions with a fall in the level of GDP per capita from peak to trough above the median fall over the entire country-year sample, which is close to 3.5% of peak GDP per capita. The identified severe recessions are associated with an increase in the unemployment rate of on average 2.6 percentage points, confirming that the episodes come with a significant increase in idleness. Figure 1 displays the dates of the pre-severe recession GDP per capita peak and Table A1.1 in the Appendix shows the number of recessions and severe recessions for each country.

Figure 1. **Severe recession dates**

Number of countries



Note: The dates show the pre-recession peak in GDP per capita.

The use of severe recessions as the dependent variable constitutes a departure from most of the early warning literature, which has typically focused on particular types of economic crises, such as currency (e.g. Kaminsky et al., 1998) or banking crises (e.g. Demirgüç-Kunt and Detragiache, 2000) and more recently broader systemic financial events (e.g. Alessi and Detken, 2011; Lo Duca and Peltonen, 2013). Our focus on severe recessions is motivated by two aspects. First, we take a pragmatic approach with the aim to identify a set of indicators which can be used to make a broad and comprehensive assessment of potential vulnerabilities. For this purpose, large drops in GDP per capita provides an efficient way to capture a range of costly economic events and represents an outcome that policymakers are presumably most concerned to avoid.

Second, it is inherently difficult to define economic crises in an objective way. Crisis definitions often differ from one study to the next and it is not unusual for studies to disagree whether a particular episode constitutes a crisis and to differ on the exact timing of a particular crisis (e.g. Romer and Romer, 2015). Consequently, differences in crises definitions have led to differences in results. Moreover, most crisis indicators do not provide information on the relative severity of crises, but rather focus on the timing. In contrast, we focus on severe recessions which can be computed in an objective and transparent manner for large panels of countries over long time periods. The robustness section shows, nonetheless, that the results are broadly robust to different definitions of costly events including episodes of banking, currency and sovereign debt crises on the basis of the data collected by Babecký et al. (2014).

We pool severe recessions across countries, which is a standard approach in the recent early warning literature (e.g. Alessi and Detken, 2011; Lo Duca and Peltonen, 2013). Pooling across countries is necessary because each country only experiences a very limited number of severe recessions (on average around 2 to 3 over the time period considered). However, a larger sample size of course implies a more heterogeneous set of country experience and the model may be too simple to pick up these heterogeneities. As explained in more detail below, we partly account for these heterogeneities by allowing the *threshold values* to differ across countries. In addition, in the robustness section we also split the sample into high and low income OECD countries.

2.2. Vulnerability indicators

In this paper we rely on a set of vulnerability indicators that have been identified in a companion working paper (Röhn et al., 2015) based on a review of the most recent early warning literature and lessons learned from the global financial crisis. In Röhn et al. (2015) indicators are classified into five types of domestic vulnerabilities (or “imbalances”). These include i) financial sector imbalances, ii) non-financial sector imbalances, iii) asset market imbalances, iv) public sector imbalances and v) external sector imbalances. Besides domestic imbalances, economies are also vulnerable to shocks and crises originating in other countries through international spillovers and contagion through financial, trade and confidence channels. In addition, a set of global indicators common to all countries are thus also included.

Data availability across the indicators and across countries varies significantly. To capture a broad range of different severe recessions, we therefore only include vulnerability indicators for which data is available for at least 50% of the severe recessions identified over the period 1970-2014. Unfortunately, this excludes all financial sector imbalances indicators from the analysis as they are generally only available for a short time span of less than 10 years (see Röhn et al. [2015] for the precise data availability of the different indicators). To save space we do not report results for the public sector imbalance indicators, which have generally not been found to be useful for early warning in this study. The interested reader is referred to the working paper version of this article (Hermansen and Röhn, 2015). Table A1.2 in the Appendix provides details on the indicators included in the analysis and Table A1.3 reports pairwise correlations among the indicators. Indicators in the same imbalance area are generally positively correlated, with the highest correlations among credit and some real estate indicators. Across imbalance areas pairwise correlations are generally lower.

As is standard in the literature, we experiment with different transformations of the indicators. In particular, for several variables we also use deviations from a trend. The trends

have been calculated according to three different approaches: a) a slowly-adjusting one-sided HP-filter (gap1); b) a faster-adjusting one-sided HP-filter (gap2); and c) a 20 quarter (5 year) one-sided moving average (gap3).¹ All three trends are calculated recursively using only available data up to each point in time, ensuring only real time available information are utilised. In addition, several indicators are also expressed as growth rates: year-on-year growth rates (gr1); quarter-on-quarter growth rates (gr2); and cumulated growth rates over the preceding 6 quarters (4 years for annual series) (gr3). Table A1.2 in the Appendix lists the transformations employed for each indicator.

3. Empirical methodology: the signalling approach

The early warning literature has predominantly relied on two empirical approaches: signalling models (see e.g. Kaminsky et al. (1998) for an influential study) and multivariate logit/probit models (e.g. Demirgüç-Kunt and Detragiache, 1998).² The advantage of the signalling approach is that it accommodates differences in data availability across countries and allows for the inclusion of a potentially larger number of vulnerability indicators than the multivariate regression method. This is an important advantage, since the aim of this study is to assess the predictive ability of each individual indicator rather than to devise a composite early warning indicator. Hence we employ the signalling approach. A drawback of this method is that in its simplest form it ignores potential interactions among indicators and does not allow for standard statistical tests to assess the significance of the indicators.

The signalling approach is a non-parametric approach, which is based on the idea that a useful early warning indicator behaves differently in pre-crisis episodes compared to normal periods. A vulnerability indicator issues a warning signal of an upcoming crisis if the indicator exceeds a threshold, here defined by a percentile of an indicator's own distribution. Each indicator can then be evaluated according to the matrix below in which crisis occurrence and warning issuance are compared. A is the number of quarters across countries and time in which an indicator provides a correct signal, B is the number of quarters in which a wrong signal is issued, that is a signal was provided, but there was no crisis. C is the number of quarters the indicator does not issue a signal despite a crisis occurring. Finally, D is the number of quarters in which the indicator does not provide any warning signal, and rightly so because there was no crisis.

Evaluation matrix

	Crisis (within the following 8 quarters)	No crisis (within the following 8 quarters)
Signal issued	A	B
No signal issued	C	D

A signal is considered as correct if a crisis occurs within a fixed number of quarters after the signal is issued. In the baseline we set the number of quarters to 8, which is a commonly applied window in the literature (e.g. Kaminsky et al., 1998). In the robustness section we investigate whether the results are sensitive to the choice of this forecasting horizon. The first four quarters following the start of a severe recession are excluded from the evaluation sample since the behaviour of the vulnerability indicators is likely to be different during a severe recession compared to normal or pre-recession times (Bussiere and Fratzscher, 2006).

Ideally a threshold for each indicator should be chosen such that all observations fall into the A (a signal was issued and indeed there was a crisis) and D (a signal was not issued and indeed there was no crisis) cells. In reality, however, setting the threshold involves balancing two types of errors policy makers face. A high threshold would imply few crisis signals and a higher risk of missing a crisis (type I error). A low threshold on the other hand would increase the number of signals, but would also raise the number of false crisis signals (type II error). The optimal threshold is set by minimising the following loss function, which reflects the two types of errors as well as policymakers' relative preferences for either type. In particular we follow Sarlin (2013) and define the loss function as follows:³

$$\begin{aligned} L(x) &= \theta \cdot P \cdot T_1 + (1 - \theta) \cdot (1 - P) \cdot T_2 \\ &= \theta \cdot \frac{A+B}{A+B+C+D} \cdot \frac{C}{A+C} + (1 - \theta) \cdot \frac{B+D}{A+B+C+D} \cdot \frac{B}{B+D} \\ &= \theta \cdot \frac{C}{A+B+C+D} + (1 - \theta) \cdot \frac{B}{A+B+C+D}, \quad \theta \in [0, 1] \end{aligned}$$

where the threshold x determines the distribution between A-D and P and $(1-P)$ captures the unconditional probability of pre-crisis and normal periods, respectively. The parameter θ reflects the preferences of policymakers between type I (T_1) and type II (T_2) errors and is related to the relative costs of missing crises versus costs of false alarms (e.g. Demirgüç-Kunt and Detragiache, 2000). In case an alarm goes off, preventive action should be taken to avoid or at least reduce the cost of a possible crisis (see e.g. Caldera Sánchez et al. (2016) for a discussion of policy tools). If the alarm turns out to be false, costs are likely to be associated with the preventive action for example in terms of lower growth. T_1 , T_2 , and P can be estimated from the in-sample frequencies of A-D given a threshold x as shown in the second line of the equation above.⁴ With the experience of the global financial crisis, policymakers are likely to be more concerned about missing crises. Hence, below we focus on values of the preference parameter in the range of $\theta \in [0.5-0.9]$.

The threshold percentile is optimised over all countries, i.e. a common percentile is chosen which minimises the aggregate loss over all countries.⁵ However, to allow for the fact that the distributions of indicators are likely to differ substantially across countries, this optimal threshold percentile is applied to country-specific distributions of the indicators. Hence, the threshold values are allowed to differ across countries. For example, the 65th percentile might be the optimal threshold percentile for an indicator, i.e. the optimal threshold leaves 65% of the observations below the threshold for each country. In the case of the private credit gap indicator (the difference between the actual credit-to-GDP ratio and a trend) this corresponds to a threshold value of 4.4 percentage points of GDP for the United States above which a signal is issued, whereas the threshold value is 5.7 percentage points of GDP for Spain.

The performance of each indicator is assessed according to the following criteria:

- **Absolute Usefulness:** $U_a(x) = \min[\theta P, (1 - \theta)(1 - P) - L(x)]$. If the absolute usefulness is positive, there is a benefit in using the early warning indicator as it results in a lower loss ($L(x)$) for the policymaker than simply disregarding the indicator, which would result in a loss of $\min[\theta P, (1 - \theta)(1 - P)]$. This is because if the policymaker disregards the indicator, she has the choice between always or never signalling a crisis. If she always signals a crisis, so that $C = D = 0$, the resulting loss is $(1 - \theta)(1 - P)$. If she never signals a crisis, so that $A = B = 0$, the resulting loss is θP . The policymaker would choose the option which results in a lower loss given preferences θ and the unconditional crisis probability P .

Hence, the loss from disregarding the indicator is $\min[\theta P; (1-\theta)(1-P)]$. An indicator is then useful if it results in a lower loss than this benchmark loss. Moreover, it can be shown that for any given indicator threshold x the usefulness of the indicator is maximised at $\theta = (1-P)$. Since P is low in practice, this implies that for an indicator to achieve a high usefulness the policymaker needs to be significantly more concerned about the detection of crises than avoiding false alarms. Otherwise the policymaker could easily achieve a lower loss compared to the non-perfectly performing indicator by always assuming to be in a state of the high-frequency class, which is a normal period in our case.

- Relative Usefulness: $U_r(x) = \frac{U_a(x)}{\min[\theta P, (1-\theta)(1-P)]}$. Relative usefulness is the share of an indicators' absolute Usefulness U_a relative to the maximum possible Usefulness. A perfectly performing indicator would result in a value of the above loss function of zero so that the absolute usefulness of the measure would be $\min[\theta P; (1-\theta)(1-P)]$. Hence, U_r reports U_a as a percentage of the usefulness that a policymaker could achieve with a perfectly performing indicator. The relative usefulness is our preferred performance indicator as it allows for the comparison of models for policymakers with different preference parameters θ .

In addition to these two main criteria we also report several additional performance measures that have been frequently used in the literature:

- Adjusted noise-to-signal ratio (aNtS) = $[B/(B+D)]/[A/(A+C)]$, captures the ratio of the share of false alarms (noise) versus the share of correctly predicted crises (signal). A useful indicator is supposed to have an aNtS of less than 1. A value of 1 would result if an indicator provides purely random signals.
- Conditional probability = $A/(A+B)$. The probability of a crisis conditional on a signal being issued.
- Difference probability = $A/(A+B) - (A+C)/(A+B+C+D)$. The difference between the conditional probability and the unconditional probability of a crisis occurring. The larger the difference the better quality of the indicator. A negative difference probability implies that the indicator performs worse than an early warning system based on the simple unconditional probability of a crisis. In this case the indicator should not be applied.
- Average lead time (ALT). The average number of quarters prior to a crisis in which the first signal is issued.

In the results section below, the assessment of the indicators is mainly based on the (relative) usefulness criterion. This is because the usefulness measures allow to explicit accounting for policymaker preferences between missing crises and avoiding false alarms in contrast to the additional criteria such as the noise-to-signal ratio. To the extent that the other criteria provide additional insights we report them in some cases.

4. Results

4.1. Full-sample results

Table 1 shows the performance of each indicator according to the relative usefulness criterion together with the optimal threshold percentile for different values of the preference parameter θ . For a particular indicator, only the transformation that yields the highest relative usefulness is reported in the tables. The focus is on values of the preference parameter $\theta \in [0.7, 0.9]$, i.e. when policymakers have strong preferences for the detection of pre-severe recession episodes. For lower values of the preference parameter the indicators

Table 1. Full-sample performance of individual indicators

	Direction to be safe	$\theta = 0.7$			$\theta = 0.8$			$\theta = 0.9$		
		Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness
Non-financial sector imbalances										
Total private credit (% of GDP)	<				none	80	0.08	gap1	10	0.05
Private bank credit (% of GDP)	<	gr1	95	0.02	gr3	75	0.13	gap1	10	0.06
Household credit (% of GDP)	<				gap3	70	0.08			
Corporate credit (% of GDP)	<	gap1	95	0.01	gap2	90	0.05	gap1	25	0.09
Asset market imbalances										
Real house prices	<	gap3	95	0.06	gap3	85	0.15	gr3	40	0.09
House price-to-disposable income ratio	<	none	90	0.17	none	80	0.26	none	60	0.15
House price-to-rent ratio	<	none	90	0.08	none	80	0.18	gr3	40	0.12
Residential investment (% of GDP)	<	none	95	0.01	none	90	0.08	none	10	0.01
Real equity prices	<	gap1	90	0.04	gap1	85	0.12	gap3	45	0.07
External imbalances										
Current account balance (% of GDP)	>	none	5	0.01	none	10	0.07			
Foreign currency exposure index*	>									
Quantitative foreign currency exposure*	>				gr3	65	0.10	gr1	80	0.03
Foreign exchange reserves (% of GDP)	>	none	5	0.05	none	10	0.10			
Foreign reserves to M2*	>	none	15	0.05	none	15	0.14			
Real effective exchange rate (CPI)	<							gap3	5	0.01
Real effective exchange rate (ULC)	<							gr3	10	0.01
Export performance	>				gap3	10	0.03			
Spillovers, contagion and global risks										
Trade openness (% of GDP)	<				none	80	0.11			
Global private credit (% of GDP)	<				gr1	55	0.18	gr3	30	0.17
Global private bank credit (% of GDP)	<	gap1	80	0.10	gr3	70	0.31	gr3	60	0.22
VIX volatility index	<				gap2	65	0.25	gap2	55	0.24
Global real equity prices	<	gap1	90	0.23	gap3	75	0.30	gap3	50	0.27
Global real house prices	<	gr3	95	0.03	gap3	65	0.27	gap3	55	0.24

Note: The indicators are measured on a quarterly frequency, except for indicators marked with *, which are measured on an annual frequency. Relative usefulness measures the share of the usefulness of the indicator relative to a perfectly performing indicator (see section 3). Up to six different transformations have been tested for each indicator and only the best in terms of the relative usefulness criteria is reported. Gap1: deviation from a recursive, slowly-adjusting HP-filter with smoothing parameter $\lambda = 400000$ for quarterly series ($\lambda = 1600$ for annual series); gap2: deviation from a recursive, faster-adjusting HP-filter with smoothing parameter $\lambda = 26000$ for quarterly series ($\lambda = 100$ for annual series); gap3: deviations from a 20 quarter (5 year) moving average; gr1: year-on-year growth rates; gr2: quarter-on-quarter growth rates; gr3: cumulated growth rates over the preceding 6 quarters (4 years for annual series). Figures in bold indicate that the indicator is among the best 10 performing indicators in terms of the relative usefulness criterion for a given preference parameter θ . When indicators have a relative usefulness of zero the corresponding cells in the table are left blank.

Source: Authors' calculations.

are usually not useful and results are not reported here.⁶ The fact that the indicators only become useful for higher values of θ , as the reasoning above about the features of this loss function has shown, is a direct result of the low unconditional probability of pre-severe recession episodes in the sample around 10-20%. For $\theta \in [0.7, 0.9]$ most of the indicators are useful even if in several cases only marginally so. Table A1.4 in the Appendix also shows that most of the top performing indicators issue first warning signals on average more than 6 quarters before the onset of a severe recession, providing policymakers with a sufficiently long lead to react.

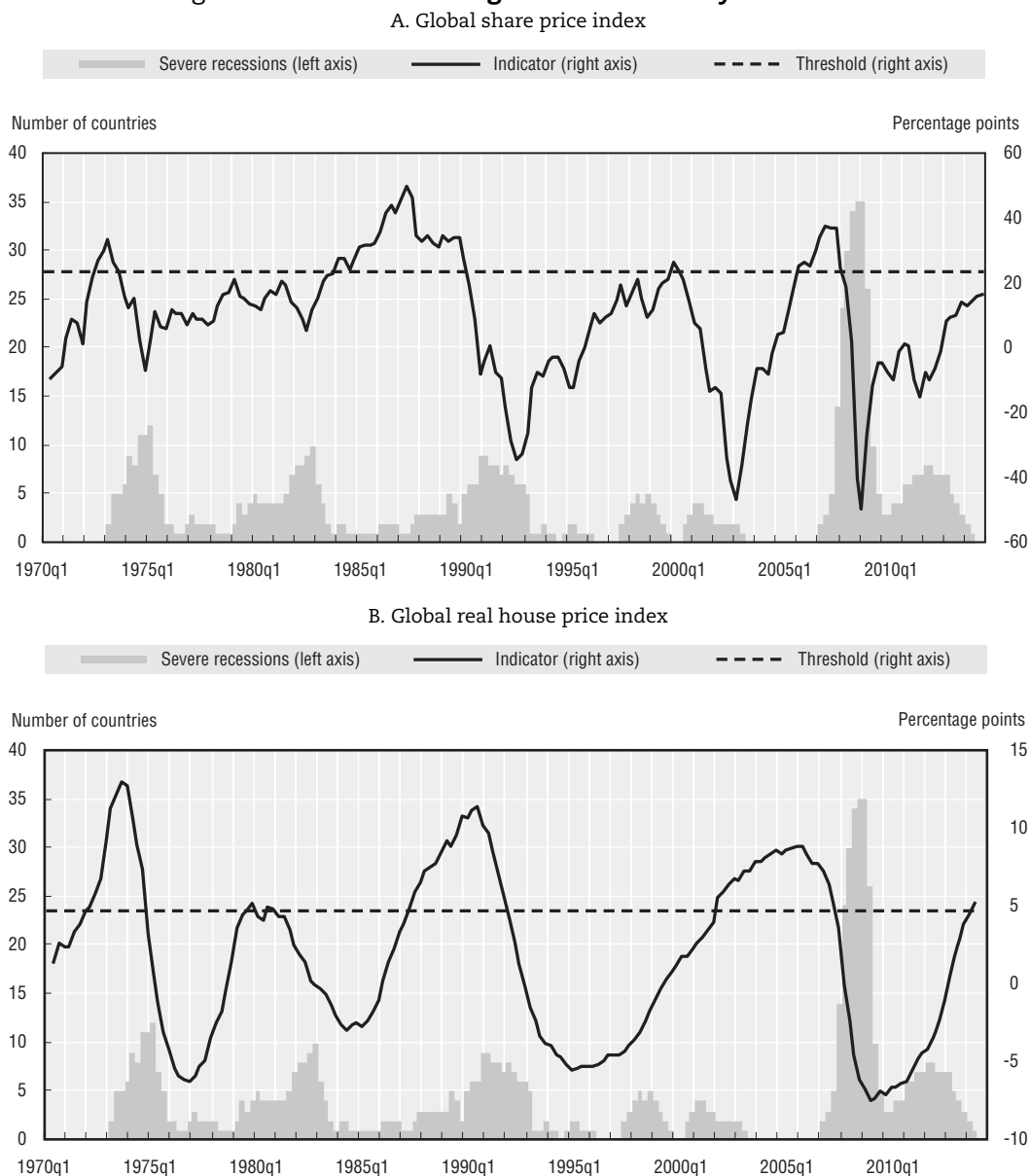
Global vulnerability indicators consistently outperform domestic indicators in terms of relative usefulness irrespective of the preference parameter θ . For $\theta \in [0.8, 0.9]$, the

(cumulative) growth of the global private bank credit-to-GDP ratio, a global equity price gap and a global house price gap perform particularly well. For example, growth of the global bank credit-to-GDP ratio is the best performing indicator for a value of the preference parameter of $\theta = 0.8$. The indicator reaches a relative usefulness of 31%, which means the indicator achieves 31% of the usefulness a policymaker would gain from an indicator that calls all severe recession episodes correctly and issues no false alarms. Table A1.4 provides further details on the performance of the indicators. The table shows that the global bank credit-to-GDP ratio indicator correctly calls 62% (1 minus type I error (0.38)) of the pre-severe recession episodes. Conditional on a signal being issued, the probability of a severe recession is 33%, which is 19 percentage points higher than the unconditional probability of a severe recession in the sample. Finally, the indicator issues a first warning signal on average 7 quarters ahead of the onset of a severe recession, providing policymakers with a sufficiently long lead to react.

The good performance of the global indicators is in line with the findings in the literature (e.g. Lo Duca and Peltonen, 2013; Babecký et al., 2014; Behn et al., 2013; Alessi and Detken, 2011) and highlights the importance to take international developments into account when assessing a country's vulnerabilities. In an increasingly integrated world economy, vulnerabilities that build-up on the global level potentially transmit to countries around the world. The good performance of the global indicators is however subject to a caveat: as the indicators do not vary across countries they are particularly suited to pick up recessions that affect a large number of countries simultaneously, such as the global financial crisis in 2008/09. Global indicators may be less suited to pick up vulnerabilities to more locally confined severe recessions. Figure 2 illustrates the performance of the two global risk indicators: global equity prices and house prices. It shows that good performance of the indicators is not solely driven by the global financial crisis. The indicators also pick up earlier episodes in which a range of countries experienced crisis simultaneously such as in the early 90s and mid-70s.⁷

Turning to the domestic vulnerability indicators, house price related indicators perform particularly well. The house price-to-disposable-income indicator is the best performing domestic indicator across different values of the preference parameter. In addition, the house price-to-rent ratio and real house price gap perform well. These results confirm findings in the literature that unsustainable real estate booms are often followed by costly economic downturns (e.g. Borio and Drehmann, 2009; Claessens et al., 2012).

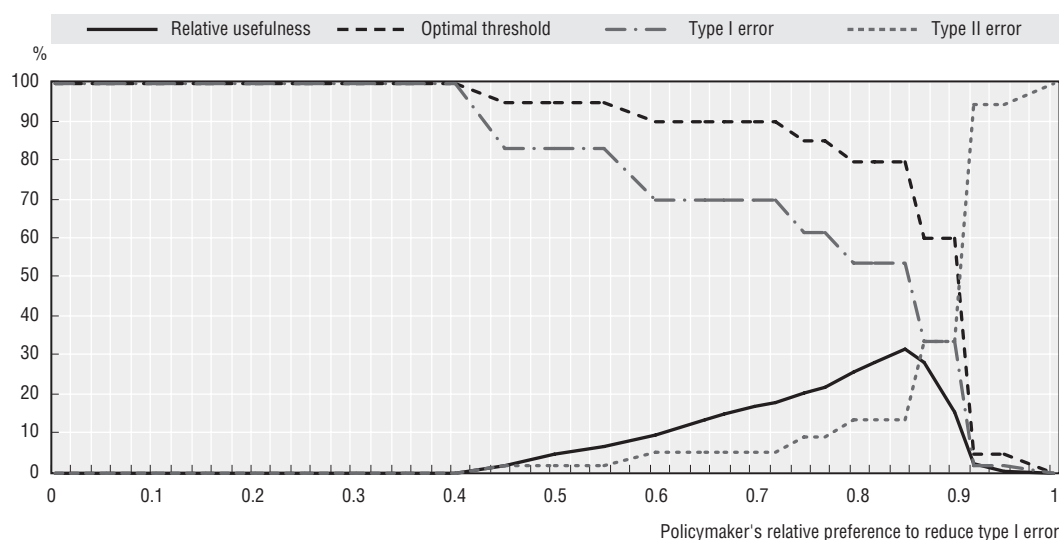
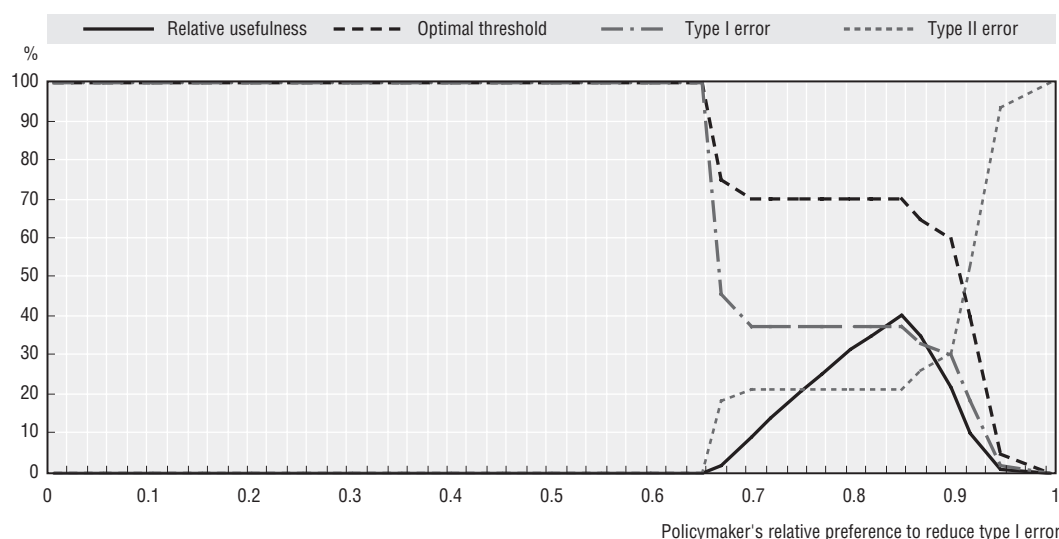
Domestic credit related variables also perform well, but are less robust to small variations in the preference parameter. In particular, growth in domestic bank credit-to-GDP ($\theta = 0.7$ and $\theta = 0.8$) and the gap of corporate credit-to-GDP are among the top-10 indicators in terms of the relative usefulness. While the literature has found credit variables to be among the most robust early warning indicators of financial crises (Borio and Lowe, 2002; Kaminsky and Reinhart, 1999; Reinhart and Rogoff, 2013; Schularick and Taylor, 2012; Taylor, 2012; Jordà et al., 2011), our results show that they are also relevant for the detection of severe recessions more broadly. Our results, however, also indicate that credit variables are less robust when the focus is more broadly on severe recessions in contrast to financial crises which have been the focus of the previous literature. We explore this issue further in the robustness section when we test whether the results change when we employ banking crises as our dependent variable. Finally, the results also suggest that official reserves (as a ratio of M2) are useful in predicting severe recessions (for $\theta = 0.8$).

Figure 2. **Performance of global vulnerability indicators**

Note: Grey areas represent the number of countries identified as being in a severe recession (from peak to trough). The global share price index and real house price index are both measured in deviation from a moving average (gap 3, see above). The optimal threshold is illustrated for $\theta = 0.8$. Global indicators are constructed as a weighted average across OECD countries (using GDP weights, see Table A1.2).

Source: Author's calculations.

These general observations notwithstanding, the findings also highlight the sensitivity of the results to small changes in the preference parameter θ . The relative usefulness and the ranking of best performing indicators (including the best transformation of an indicator in terms of relative usefulness) can vary with the value of θ . This sensitivity is further illustrated in Figures 3 and 4 which show the relationship between θ and the relative usefulness as well as the trade-off between type I and type II errors for two selected indicators: house price-to-disposable income ratio and global private bank credit-to-GDP. The figures show that the usefulness generally increases in θ up to a certain point after

Figure 3. **Usefulness of house price-to-disposable income ratio**Figure 4. **Usefulness of global private bank credit-to-GDP**

Note: The indicator is transformed into cumulated growth rates over the preceding 6 quarters.

which the usefulness drops quite sharply. After this point, because of the strong preference against missing crises, it is difficult for an indicator to beat the benchmark of always signalling a crisis (and therefore never missing a crisis). To reflect the fact that the results are sensitive to the exact value of θ but to keep the discussion tractable at the same time, we restrict presentation of the results to values of the preference parameter $\theta \in [0.8, 0.9]$ in the following. Given our focus on highly costly events, the choice of preference parameters reflecting strong preferences for their detection appears reasonable.

4.2. Out-of-sample results

As the ultimate objective of early warning indicators is to help predict future costly events, we also evaluate the performance of the indicators out-of-sample. In particular, we

want to evaluate whether the indicators provided useful information to detect the severe recessions associated with the global financial crisis and the European sovereign debt crisis. To this end, we split the sample into an “in-sample” estimation period and an “out-of-sample” evaluation period. The starting date of the in-sample estimation period depends on the data availability of the indicators. The in-sample estimation period ends in 2004Q4 to exclude the global financial crisis from the in-sample estimation period given our baseline forecasting horizon of 8 quarters. The out-of-sample evaluation period spans from 2005Q1 to 2012Q4. The out-of-sample evaluation period ends in 2012Q4 as we have only data to evaluate predictions until 2014Q4. Note that by splitting the sample in this way, almost half of the severe recession episodes are excluded from the estimation sample, highlighting the challenges involved in the out-of-sample exercise. In addition, the evaluation sample is heavily dominated by the global financial crisis. However, given the limited number of severe recessions it is difficult to devise an out-of-sample exercise in which a dominance of the global financial crisis would not occur.

The out-of-sample evaluation proceeds in the following steps. In a first step, the optimal threshold percentile for each indicator is calculated by minimising the loss function over the in-sample estimation period up to 2004Q4. These thresholds are then applied for each indicator in the evaluation sample. Given the threshold, the signals from the indicators are collected and type I, type II and the usefulness criterion are computed for values of the preference parameter $\theta \in [0.8, 0.9]$.⁸ It should be noted that if an indicator did not yield positive usefulness in the estimation sample, the indicator is excluded from the out-of-sample exercise by setting the threshold percentile to 100 (or 0).

The results are presented in Table 2. While for $\theta = 0.8$ only very few indicators achieve a positive relative usefulness, for $\theta = 0.9$ most of the indicators are useful and for several indicators the relative usefulness is even higher than in the full sample (see Table 1). Global indicators perform very well, confirming their superior full sample performance. For example, the global equity price gap indicator achieves a very high relative usefulness of 64%. It signals 84% of the pre-recession episodes correctly and only issues 17% false alarms. In addition, on average it issued a first warning signal 7.5 quarters ahead of the onset of the severe recessions (not reported). In addition, indicators of global credit and global real house prices also perform very well out-of-sample.

Turning to the domestic indicators, we find that credit and asset market indicators perform particularly well, similar to the full sample results. Domestic credit gap indicators perform very well and achieve a higher relative usefulness compared to the full sample results, highlighting the particular importance of unsustainable domestic credit booms in the global financial crises. From the asset market imbalance indicators the real equity price gap performs particularly well achieving a relative usefulness of 36%, substantially higher than for the full sample. The indicator also signals 90% of all pre-recession episodes correctly and started issuing warning signals on average more than 7.5 quarters before the onset of the financial crises (not reported). The gap of the residential investment-to-GDP ratio from trend also performs well out-of-sample and substantially better than in the full sample. The house price related indicators perform slightly worse out-of-sample compared to the full sample. In the case of the price-to-rent and price-to-disposable income ratios gaps from long-term trends now perform better than the levels of the ratios. External imbalance indicators do not perform very well out-of-sample and once again we do not find any role for the investigated fiscal imbalance indicators.

Table 2. **Out-of-sample performance of individual indicators**

	Direction to be safe	$\theta = 0.8$					$\theta = 0.9$				
		Transformation	Threshold	Relative usefulness	Type I error	Type II error	Transformation	Threshold	Relative usefulness	Type I error	Type II error
Non-financial sector imbalances											
Total private credit (% of GDP)	<						gap1	15	0.25	0.16	0.57
Private bank credit (% of GDP)	<	gr1	95	0.00	0.93	0.03	gap1	15	0.25	0.19	0.54
Household credit (% of GDP)	<						gap3	75	0.19	0.56	0.23
Corporate credit (% of GDP)	<						gap1	30	0.25	0.25	0.47
Asset market imbalances											
Real house prices	<	gr2	95	0	1	0.01	gap3	80	0.13	0.74	0.12
House price-to-disposable income ratio	<	gr2	95	0	0.99	0.02	gap3	85	0.11	0.78	0.08
House price-to-rent ratio	<	gap2	95	0	0.98	0.01	gr3	75	0.11	0.74	0.12
Residential investment (% of GDP)	<	none	85	0.06	0.85	0.04	gap3	65	0.21	0.57	0.17
Real equity prices	<						gap3	15	0.36	0.10	0.52
External imbalances											
Current account balance (% of GDP)	>	none	5	0	0.83	0.12	none	50	0	0.49	0.42
Foreign currency exposure index*	>						none	30	0	0.73	0.20
Quantitative foreign currency exposure*	>										
Foreign exchange reserves (% of GDP)	>	gr3	5	0.08	0.81	0.06	gap3	95	0.03	0.09	0.86
Foreign reserves to M2*	>	gr1	5	0	0.67	0.29					
Real effective exchange rate (CPI)	<						gap1	10	0.03	0.10	0.84
Real effective exchange rate (ULC)	<						gr1	15	0.05	0.12	0.79
Export performance	>						gap1	25	0.05	0.75	0.15
Spillovers, contagion and global risks											
Trade openness (% of GDP)	<						gr1	5	0.07	0.04	0.87
Global private credit (% of GDP)	<	gr3	95	0	0.97	0.1	gap1	25	0.42	0.12	0.45
Global private bank credit (% of GDP)	<	gap2	95	0	1	0.04	gr3	70	0.39	0.18	0.41
VIX volatility index	<						gap1	95	0	1	0.01
Global real equity prices	<						gap3	50	0.64	0.16	0.17
Global real house prices	<	gr3	75	0	0.99	0.01	gap3	55	0.36	0.47	0.16

Note: The in-sample dataset includes data until 2004q4 and the out-of-sample dataset spans from 2005q1 to 2012q4. See note to Table 1.
Source: Authors' calculations.

4.3. Robustness

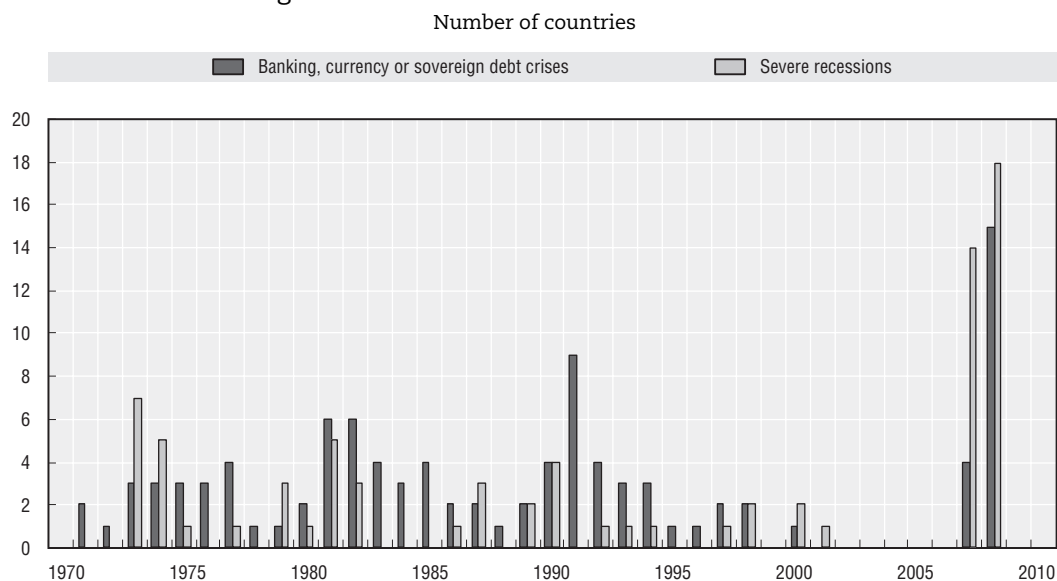
Alternative definitions of costly events

In the baseline we have been agnostic about the cause of a severe recession. This is in contrast to much of the early warning literature that has focused on particular types of crises, such as currency or banking crises (see Section 2). Therefore, we check whether the results hold when instead of using severe recessions as a definition of costly events, we use crises episodes instead.

Crises data are taken from Babecký et al. (2014). Their database covers banking, currency and sovereign debt crises among OECD countries over the period 1970Q1-2010Q4. Babecký et al. (2014) collect crisis dates from a range of studies, including the recent and well-known databases of Reinhart and Rogoff (2011) and Laeven and Valencia (2012). A crisis is identified if at least one study claims that a crisis occurred. The dataset has several advantages. First, annual crises dates have been cross-checked by country experts, who also converted them into quarterly frequency, to avoid relying overly on a specific crisis definition based on

changes in one or few variables. Second, its quarterly frequency allows a more precise assessment of the usefulness of early warning indicators. Figure 5 shows the dates of defined crises in comparison with the dating of severe recessions. The figure shows that crises and severe recession dates do not always coincide. Overall fewer severe recessions compared to crises are identified. On the other hand, during the global financial crisis the number of countries experiencing a severe recession is larger than the number of countries experiencing a (banking) crisis.

Figure 5. **Crises and severe recession dates**



Source: Babecký et al. (2014).

Table 3 reports the results for all types of crises included in the Babecký et al. (2014) database and for banking crises in particular. The results are qualitatively similar to the baseline. In particular global variables outperform domestic variables. In addition, the house price-to-disposable income ratio and the house price-to-rent ratio are among the best performing domestic variables. However, several differences also stand out. The official reserves-to-GDP ratio and an index of foreign currency exposure perform better in both crises samples compared to the baseline (for $\theta = 0.8$). In the “banking crises” sample, the domestic credit variables perform somewhat better than in the baseline, highlighting the particular role that credit booms play for banking crises.

Forecasting horizon

Next, we test the robustness of our results to the choice of the forecast horizon. In the baseline this horizon was set to 8 quarters. Table 4 shows that the results are very similar when the forecast horizon is set to 12 quarters for $\theta = 0.8$. Indeed relative usefulness of several of the indicators actually increases for this longer forecasting horizon. The indicators are less useful for $\theta = 0.9$. In contrast, for a forecasting horizon of 4 quarters the indicators become more useful for very strong preferences against missing crises. For $\theta = 0.9$ the top indicators are again similar to the baseline results even if the transformations of the best performing indicators sometimes differ compared to the baseline.

Table 3. **Robustness: economic crises**

	Direction to be safe	Baseline: severe recessions						Banking currency and sovereign debt crises						Banking crises					
		$\theta = 0.8$			$\theta = 0.9$			$\theta = 0.8$			$\theta = 0.9$			$\theta = 0.8$			$\theta = 0.9$		
		Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness
Non-financial sector imbalances																			
Total private credit (% of GDP)	<	none	80	0.08	gap1	10	0.05	none	85	0.11	all	10	0.02	none	85	0.17	gap3	55	0.20
Private bank credit (% of GDP)	<	gr3	75	0.13	gap1	10	0.06	gap3	85	0.12	gap1	10	0.01	gap3	90	0.14	gap3	55	0.20
Household credit (% of GDP)	<	gap3	70	0.08				none	75	0.06				none	75	0.08	none	20	0.14
Corporate credit (% of GDP)	<	gap2	90	0.05	gap1	25	0.09	gap3	80	0.09	gap1	5	0.02	gap3	90	0.10	gap3	40	0.09
Asset market imbalances																			
Real house prices	<	gap3	85	0.15	gr3	40	0.09	gap3	90	0.06	gap3	10	0.02	gap3	90	0.07	gap3	60	0.06
House price-to-disposable income ratio	<	none	80	0.26	none	60	0.15	none	80	0.18				none	85	0.19	none	80	0.21
House price-to-rent ratio	<	none	80	0.18	gr3	40	0.12	none	85	0.13	gap3	5	0.02	none	90	0.17	none	55	0.16
Residential investment (% of GDP)	<	none	90	0.08	none	10	0.01	none	60	0.08							gap2	5	0.03
Real equity prices	<	gap1	85	0.12	gap3	45	0.07	gap1	90	0.07	gap3	20	0.06	gap1	90	0.09	gap3	50	0.16
External imbalances																			
Current account balance (% of GDP)	>	none	10	0.07				none	25	0.09				none	10	0.03	none	50	0.06
Foreign currency exposure index*	>							none	15	0.13				none	15	0.13	none	100	0
Quantitative foreign currency exposure*	>	gr3	65	0.10	gr1	80	0.03	gap1	15	0.02	gr3	90	0				gap3	95	0.01
Foreign exchange reserves (% of GDP)	>	none	10	0.10				none	25	0.18				none	15	0.13	none	20	0.14
Foreign reserves to M2*	>	none	15	0.14				none	90	0.00				none	10	0.12	none	90	0.03
Real effective exchange rate (CPI)	<				gap3	5	0.01	gap3	35	0.01	gap3	20	0.02				gap3	35	0.11
Real effective exchange rate (ULC)	<				gr3	10	0.01	gap3	40	0.03	gr3	5	0.01				gap3	30	0.08
Export performance	>	gap3	10	0.03				gap3	15	0.04				gap3	5	0.02	gap3	75	0.06
Spillovers, contagion and global risks																			
Trade openness (% of GDP)	<	none	80	0.11				none	85	0.09				none	90	0.13	none	85	0.08
Global private credit (% of GDP)	<	gr1	55	0.18	gr3	30	0.17	none	75	0.13	gap3	25	0.10	none	75	0.13	gr1	55	0.27
Global private bank credit (% of GDP)	<	gr3	70	0.31	gr3	60	0.22	none	80	0.20	gap3	5	0.02	none	80	0.26	gap3	60	0.32
VIX volatility index	<	gap2	65	0.25	gap2	55	0.24	gap2	60	0.10	gap1	15	0.15	gap2	60	0.04	gap2	50	0.21
Global real equity prices	<	gap3	75	0.30	gap3	50	0.27	gap3	75	0.14	gap3	30	0.10	gap1	90	0.17	gap3	75	0.28
Global real house prices	<	gap3	65	0.27	gap3	55	0.24	gap3	60	0.10				gap3	80	0.08	gap3	70	0.17

Note: See note to Table 1.

Source: Authors' calculations.

Table 4. **Robustness: forecasting horizon**

	Direction to be safe	Baseline: 8 quarters						12 quarters						4 quarters					
		$\theta = 0.8$			$\theta = 0.9$			$\theta = 0.8$			$\theta = 0.9$			$\theta = 0.8$			$\theta = 0.9$		
		Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness
Non-financial sector imbalances																			
Total private credit (% of GDP)	<	none	80	0.08	gap1	10	0.05	none	75	0.12	gap1	5	0.02				gr3	75	0.10
Private bank credit (% of GDP)	<	gr3	75	0.13	gap1	10	0.06	gr3	60	0.21	gap1	10	0.02	gap3	95	0.02	gr3	75	0.16
Household credit (% of GDP)	<	gap3	70	0.08				gap3	60	0.13							gap3	70	0.12
Corporate credit (% of GDP)	<	gap2	90	0.05	gap1	25	0.09	gap1	25	0.13	gap1	10	0.02				gap1	70	0.13
Asset market imbalances																			
Real house prices	<	gap3	85	0.15	gr3	40	0.09	gr3	40	0.18	gr3	5	0.02	gap3	95	0.05	gap3	70	0.19
House price-to-disposable income ratio	<	none	80	0.26	none	60	0.15	none	65	0.29	gr3	10	0.05	none	95	0.15	none	80	0.29
House price-to-rent ratio	<	none	80	0.18	gr3	40	0.12	gr3	40	0.21	gr1	10	0.03	none	95	0.07	gap3	55	0.22
Residential investment (% of GDP)	<	none	90	0.08	none	10	0.01	gap3	60	0.10				none	95	0.01	none	85	0.11
Real equity prices	<	gap1	85	0.12	gap3	45	0.07	gr3	45	0.17	gap3	10	0.03				gap3	60	0.14
External imbalances																			
Current account balance (% of GDP)	>	none	10	0.07										none	5	0.01	none	25	0.16
Foreign currency exposure index*	>							none	95	0.01							none	80	0
Quantitative foreign currency exposure*	>	gr3	65	0.10	gr1	80	0.03	gr1	80	0.01				gr1	5	0.01	gr1	75	0.11
Foreign exchange reserves (% of GDP)	>	none	10	0.10				none	65	0.02				none	5	0.05	none	15	0.11
Foreign reserves to M2*	>	none	15	0.14										none	15	0.05	gap1	90	0.02
Real effective exchange rate (CPI)	<				gap3	5	0.01	gr3	15	0.05							gr1	45	0.06
Real effective exchange rate (ULC)	<				gr3	10	0.01	gr3	10	0.05							gap3	65	0.04
Export performance	>	gap3	10	0.03				gap1	80	0.03							gap3	25	0.03
Spillovers, contagion and global risks																			
Trade openness (% of GDP)	<	none	80	0.11				none	65	0.09							none	80	0.12
Global private credit (% of GDP)	<	gr1	55	0.18	gr3	30	0.17	gr1	55	0.22	gap3	15	0.08				gr3	55	0.23
Global private bank credit (% of GDP)	<	gr3	70	0.31	gr3	60	0.22	gr1	50	0.30	gr3	20	0.07	gap1	80	0.04	gap3	65	0.35
VIX volatility index	<	gap2	65	0.25	gap2	55	0.24	gap2	50	0.36	gap2	30	0.05	gr1	95	0.01	gap1	65	0.36
Global real equity prices	<	gap3	75	0.30	gap3	50	0.27	gap3	50	0.28	gap3	20	0.10	gap1	95	0.18	gap3	60	0.36
Global real house prices	<	gap3	65	0.27	gap3	55	0.24	gap3	55	0.36	gr3	40	0.11	gap3	95	0.04	gap3	55	0.25

Note: See note to Table 1.

Source: Authors' calculations.

Country sample

The baseline results are based on the set of OECD countries which are arguably a heterogeneous set of countries. Therefore, we investigate whether the results change when we consider more homogeneous sets of countries. In particular, we test whether the ranking of the best performing indicators changes between high-income and lower-income OECD countries. Table 5 shows that the results are broadly similar for low- and high-income OECD countries and also broadly comparable to the baseline results.⁹

Conclusions

The global financial crisis and the high associated costs associated with it have revived the academic and policy interest in “early warning indicators” of crises. The paper extends OECD efforts to monitor and detect early-on country risks, by providing empirical evidence on the usefulness of a new set of vulnerability indicators, proposed in a companion paper (Röhn et al., 2015), in predicting severe recessions and crises in OECD countries. To evaluate the usefulness of the indicators the signalling approach is employed, which takes into account policy makers’ preferences between missing crises and false alarms.

Our empirical evidence shows that the majority of indicators would have helped to predict severe recessions in the OECD economies between 1970 and 2014. In addition, most indicators issue first warning signals on average more than 1.5 years before the onset of a severe recession, providing policymakers with a sufficiently long lead to react. However, the extent of the signalling power varies across indicators and the results are sensitive to the exact specification of policymakers’ preferences between missing crises and false alarms.

In the domestic areas, indicators that measure asset market imbalances (real house and equity prices, house price-to-income and house price-to-rent), perform consistently well both in and out-of-sample. Domestic credit related variables appear particularly useful in signalling upcoming banking crises and in predicting the global financial crisis out-of-sample. Indicators of global risks consistently outperform domestic indicators in terms of their usefulness, highlighting the importance of taking international developments into account when assessing a country’s vulnerabilities. The good performance of the global indicators is however subject to a caveat: they are particularly suited to pick up recessions that affect a large number of countries simultaneously, such as the global financial crisis in 2008/09. The results are broadly robust to different definitions of costly events, different forecasting horizons and different time and country samples. The indicators identified as particular useful in this paper can be a valuable input for monitoring economic risks, but should be complemented with other monitoring tools, including expert judgement.

In this paper we have looked at vulnerability indicators in isolation. However, important interactions among indicators may be ignored. For example, some studies suggest that asset busts have larger repercussions on the wider economy if the preceding asset boom was credit financed (e.g. Jorda et al., 2015). Taking these interactions into account may improve forecasting accuracy (e.g. Keilis-Borok et al., 2000) and hence may also reduce the costs of policy responses associated with false alarms. We view work in this area as an important avenue for future research.

Table 5. **Robustness: country sample**

	Direction to be safe	Baseline						17 lowest income countries						18 highest income countries					
		$\theta = 0.8$			$\theta = 0.9$			$\theta = 0.8$			$\theta = 0.9$			$\theta = 0.8$			$\theta = 0.9$		
		Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness	Transformation	Threshold	Relative usefulness
Non-financial sector imbalances																			
Total private credit (% of GDP)	<	none	80	0.08	gap1	10	0.05	gap1	70	0.13	gap1	10	0.06	none	80	0.13	gr3	50	0.05
Private bank credit (% of GDP)	<	gr3	75	0.13	gap1	10	0.06	gr3	65	0.17	gap1	10	0.08	gr3	80	0.14	gr3	60	0.13
Household credit (% of GDP)	<	gap3	70	0.08				none	55	0.05	gap1	15	0.07	gap3	70	0.10	gap3	60	0.05
Corporate credit (% of GDP)	<	gap2	90	0.05	gap1	25	0.09	gap1	65	0.15	gap1	15	0.09	gap2	90	0.04	gap1	25	0.10
Asset market imbalances																			
Real house prices	<	gap3	85	0.15	gr3	40	0.09	gap3	70	0.21	gr2	5	0.01	gap3	85	0.17	gr3	40	0.17
House price-to-disposable income ratio	<	none	80	0.26	none	60	0.15	none	90	0.28	gr1	15	0.09	none	80	0.26	none	60	0.24
House price-to-rent ratio	<	none	80	0.18	gr3	40	0.12	gr1	50	0.22	gr1	25	0.08	none	80	0.19	gap3	40	0.21
Residential investment (% of GDP)	<	none	90	0.08	none	10	0.01	none	75	0.16	none	10	0.06	gap3	75	0.08			
Real equity prices	<	gap1	85	0.12	gap3	45	0.07	gap3	50	0.22	gap3	10	0.04	gap1	85	0.19	gap3	35	0.11
External imbalances																			
Current account balance (% of GDP)	>	none	10	0.07				none	30	0.21				none	5	0.02			
Foreign currency exposure index*	>							none	90	0.04	none	90	0.00	none	10	0.00			
Quantitative foreign currency exposure*	>	gr3	65	0.10	gr1	80	0.03	gr3	40	0.12	gr3	95	0.00	gr3	65	0.07	gr1	80	0.07
Foreign exchange reserves (% of GDP)	>	none	10	0.10				gr3	15	0.10	gap3	95	0.03	none	10	0.14	gap3	70	0.00
Foreign reserves to M2*	>	none	15	0.14				gap3	45	0.22	gr1	90	0.06	none	15	0.17			
Real effective exchange rate (CPI)	<				gap3	5	0.01	gap3	40	0.13							gap1	10	0.03
Real effective exchange rate (ULC)	<				gr3	10	0.01	gap3	45	0.18							gr3	10	0.03
Export performance	>	gap3	10	0.03				gap1	20	0.04	gap1	95	0.01	gap3	10	0.06	gap3	50	0.02
Spillovers, contagion and global risks																			
Trade openness (% of GDP)	<	none	80	0.11				none	65	0.09				none	80	0.19	none	65	0.03
Global private credit (% of GDP)	<	gr1	55	0.18	gr3	30	0.17	gr1	55	0.21	gr3	20	0.06	gr1	55	0.15	gr1	55	0.29
Global private bank credit (% of GDP)	<	gr3	70	0.31	gr3	60	0.22	gap1	70	0.31	gap1	5	0.04	gr3	70	0.32	gr3	65	0.39
VIX volatility index	<	gap2	65	0.25	gap2	55	0.24	gap2	65	0.26	gap1	20	0.17	gap2	65	0.25	gap2	55	0.35
Global real equity prices	<	gap3	75	0.30	gap3	50	0.27	gap3	60	0.38	gap3	45	0.19	gap1	90	0.29	gap3	55	0.33
Global real house prices	<	gap3	65	0.27	gap3	55	0.24	gap3	55	0.30	gap3	35	0.02	gap3	65	0.26	gap3	55	0.34

Note: See note to Table 1.

Source: Authors' calculations

Notes

1. For the choice of smoothing parameters we follow Alessi and Detken (2014) for quarterly series and use the conversion rule suggested by Ravn and Uhlig (2002) for annual series. In particular, for the slowly-adjusting HP-filter we use a smoothing parameter of $\lambda = 400\,000$ for quarterly series and $\lambda = 1\,600$ for annual series; for the faster-adjusting HP-filter we use a smoothing parameter of $\lambda = 26\,000$ for quarterly series and $\lambda = 100$ for annual series. See also Drehman et al. (2010) for a detailed analysis of the trend in credit-to-GDP for different smoothing parameters.
2. Focusing on business cycle forecasting, Keilis-Borok et al. (2000) use a pattern recognition algorithm developed for earthquake prediction to forecast recessions in the United States.
3. See Demirgüç-Kunt and Detragiache (2000) for a seminal contribution introducing loss functions to evaluate early warning indicators. Another commonly employed method for setting the optimal threshold is to minimise the (adjusted) noise to signal (aNtS) ratio (e.g. Kaminsky et al., 1998). The aNtS ratio is defined as the ratio of type II errors to one minus type I errors. This criterion has been shown to result in very high type I errors (i.e. large share of missed crises) (e.g. Berg and Pattillo, 1999) and is hence not considered here.
4. This formulation of the loss function differs from the one in Alessi and Detken (2011), which has also been widely used in the literature, by weighting the two types of errors by the unconditional probabilities P and $(1-P)$. The loss function proposed by Sarlin (2013) and employed in this paper is closer in spirit to the seminal contribution of Demirgüç-Kunt and Detragiache (2000). They show that the expected loss of a policymaker does not only depend on the costs of missing crises versus the costs of taking pre-emptive policy action when no crisis materialises, but also on the relative frequencies of the two events. Since crises are rare events (i.e. the unconditional probability of a crisis is low), the frequency of false alarms and hence of incurring unnecessary costs of pre-emptive policy action is potentially high. By accounting for the unconditional probabilities Sarlin's loss function has the advantage that differences in the frequency between crises (rare events) and normal periods (frequent events) are explicitly taken into account. The preference parameter θ can then be exclusively interpreted as the relative costs of missing crises versus costs of false alarms.
5. The optimisation procedure was run using a grid search over the percentile of an indicator's distribution range $[0,100]$ with incremental steps of 5.
6. The results for lower values of the preference parameter are available in the working paper version of this article.
7. This is corroborated by (unreported) results based on a sample that excludes the great recession (the "in-sample" results mentioned below). The global risk indicators are also in this shorter sample among the best performing early warning indicators.
8. We use the in-sample unconditional probability of a severe recession to compute the usefulness criterion out-of-sample as in Sarlin (2013).
9. We also tested if our results are robust when we exclude the 1970s from the sample. The reason is that financial liberalisation took off in the 1980s and hence countries became more intertwined and the type of shocks countries have faced may have changed. The results change very little when the 1970s are excluded from the sample and are therefore not reported here but are available in the working paper version of this article.

Bibliography

- Ahrend, R. and A. Goujard (2012), "International Capital Mobility and Financial Fragility – Part 1. Drivers of Systemic Banking Crises: The Role of Bank-Balance-Sheet Contagion and Financial Account Structure", *OECD Economics Department Working Papers*, No. 902, OECD Publishing, <http://dx.doi.org/10.1787/5kg3k8ksgglw-en>.
- Alessi, L. and C. Detken (2014), "Identifying excessive credit growth and leverage", *ECB Working Paper Series*, No. 1723, www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1723.pdf.
- Alessi, L. and C. Detken (2011), "Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity", *European Journal of Political Economy*, Vol. 27/3, pp. 520-533, <http://dx.doi.org/10.1016/j.ejpoleco.2011.01.003>.
- Babecký, J. et al. (2014), "Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators", *Journal of Financial Stability*, Vol. 15, pp. 1-17, <http://dx.doi.org/10.1016/j.jfs.2014.07.001>.

- Behn, M., C. Detken, T. Peltonen and W. Schudel (2013), "Setting Countercyclical Capital Buffers Based On Early Warning Models: Would it Work?", *ECB Working Paper Series*, No. 1604, www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1604.pdf.
- Benetrix, A., J. Shambaugh and P. Lane (2015), "International Currency Exposures, Valuation Effects and the Global Financial Crisis," *Journal of International Economics*, Vol. 96/1, pp. S98-S109, <http://dx.doi.org/10.1016/j.jinteco.2014.11.002>.
- Berg, A. and C. Pattillo (1999), "Predicting Currency Crises: The Indicators Approach and an Alternative", *Journal of International Money and Finance*, Vol. 18/4, pp. 561-586, [http://dx.doi.org/10.1016/S0261-5606\(99\)00024-8](http://dx.doi.org/10.1016/S0261-5606(99)00024-8).
- Borio, C. and Drehmann, M. (2009), "Assessing the Risk of Banking Crises – Revisited", *BIS Quarterly Review*, March 2009, pp. 29-46, www.bis.org/publ/qtrpdf/r_qt0903e.pdf.
- Borio, C. and P. Lowe (2002), "Assessing the Risk of Banking Crisis", *BIS Quarterly Review*, December 2002, pp. 43-52, www.bis.org/publ/qtrpdf/r_qt0212e.pdf.
- Bry, G. and C. Boschan (1971), *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs*, National Bureau of Economic Research, New York.
- Bussiere, M. and M. Fratzscher (2006), "Towards a New Early Warning System of Financial Crises", *Journal of International Money and Finance*, Vol. 25/6, pp. 953-973, <http://dx.doi.org/10.1016/j.jimonfin.2006.07.007>.
- Caldera Sánchez, A., M. Rasmussen and O. Röhn (2016), "Economic Resilience: What Role for Policies?" *Journal of International Commerce, Economics and Policy*, Vol. 07, No. 02.
- Claessens, S., M. Kose and M.M. Terrones (2012), "How Do Business and Financial Cycles Interact?", *Journal of International Economics*, Vol. 87, pp. 178-190.
- Demirgüç-Kunt, A. and E. Detragiache (2000), "Monitoring Banking Sector Fragility: A Multivariate Logit Approach", *The World Bank Economic Review*, Vol. 14, pp. 287-307.
- Demirgüç-Kunt, A. and E. Detragiache (1998), "The Determinants of Banking Crises in Developing and Developed Countries", *IMF Staff Papers*, Vol. 45/1, www.imf.org/external/Pubs/FT/staffp/1998/03-98/pdf/demirguc.pdf.
- Drehman, M. et al. (2010), "Countercyclical Capital Buffers: Exploring Options", *BIS Working Papers*, No. 317, www.bis.org/publ/work317.pdf.
- Frankel, J. and G. Saravelos (2012), "Can leading indicators assess country vulnerability? Evidence from the 2008-09 global financial crisis", *Journal of International Economics*, Vol. 87/2, pp.216-231, <http://dx.doi.org/10.1016/j.jinteco.2011.12.009>.
- Furceri, D., S. Guichard and E. Rusticelli (2011), "Episodes of Large Capital Inflows and the Likelihood of Banking and Currency Crises and Sudden Stops", *OECD Economics Department Working Papers*, No. 865, OECD Publishing.
- Hermansen, M. and O. Röhn (2015), "Economic Resilience: The Usefulness of Early Warning Indicators in OECD Countries", *OECD Economics Department Working Papers*, No. 1250, OECD Publishing, <http://dx.doi.org/10.1787/5jrxhgqfx3mv-en>.
- Jordà O., M. Schularick and A. Taylor (2011), "Financial Crises, Credit Booms, and External Imbalances: 140 Years of Lessons", *IMF Economic Review*, Vol. 59/2, pp. 340-378, <http://dx.doi.org/10.1057/imfer.2011.8>.
- Jordà, O., M. Schularick and A.M. Taylor (2015), "Leveraged Bubbles", *Journal of Monetary Economics*, Vol. 76(S), p. S1-S20, <http://dx.doi.org/10.1016/j.jmoneco.2015.08.005>.
- Kaminsky, G. and C. Reinhart (1999), "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems", *American Economic Review*, Vol. 89/3, pp. 473-500, <http://dx.doi.org/10.1257/aer.89.3.473>.
- Kaminsky, G., S. Lizondo and C. Reinhart (1998), "Leading Indicators of Currency Crises", *IMF Staff Papers*, Vol. 45, No. 1, www.imf.org/external/Pubs/FT/staffp/1998/03-98/pdf/kaminsky.pdf.
- Keilis-Borok, V., J. Stock, A. Soloviev and P. Mikhalev (2000), "Pre-recession Pattern of Six Economic Indicators in the USA", *Journal of Forecasting*, Vol. 19/1, pp. 65-80, [http://dx.doi.org/10.1002/\(SICI\)1099-131X\(200001\)19:1<65::AID-FOR730>3.0.CO;2-U](http://dx.doi.org/10.1002/(SICI)1099-131X(200001)19:1<65::AID-FOR730>3.0.CO;2-U).
- Laeven L. and F. Valencia (2012), "Systemic Banking Crises Database: An Update", *IMF Working Paper*, No. 12/163, www.palgrave-journals.com/imfer/journal/v61/n2/pdf/imfer201312a.pdf.
- Lo Duca, M. and T. Peltonen (2013), "Assessing Systemic Risks and Predicting Systemic Events", *Journal of Banking & Finance*, Vol. 37/7, pp. 2183-2195, <http://dx.doi.org/10.1016/j.jbankfin.2012.06.010>.

- Ravn, M. and H. Uhlig (2002), "On adjusting the HP-filter for the frequency of observations", *Review of Economics and Statistics*, Vol. 84/2, pp. 371-376, <http://dx.doi.org/10.1162/003465302317411604>.
- Reinhart, C. and K. Rogoff (2011), "From Financial Crash to Debt Crisis", *American Economic Review*, Vol. 101/5, pp. 1676-1706, <http://dx.doi.org/10.1257/aer.101.5.1676>.
- Reinhart, C. and K. Rogoff (2013), "Banking Crisis: An Equal Opportunity Menace", *Journal of Banking and Finance*, Vol. 37/11, pp. 4557-4573, <http://dx.doi.org/10.1016/j.jbankfin.2013.03.005>.
- Röhn, O., A. Caldera Sánchez, M. Hermansen and M. Rasmussen (2015), "Economic Resilience: A New Set of Vulnerability Indicators for OECD Countries", *OECD Economics Department Working Papers*, No. 1249, OECD Publishing, <http://dx.doi.org/10.1787/5jrxhgjw54r8-en>.
- Romer, C. and D. Romer (2015), "New Evidence on the Impact of Financial Crises in Advanced Countries", *NBER Working Paper Series*, No. 21021, <http://dx.doi.org/10.3386/w21021>.
- Rose, A. and M. Spiegel (2011), "Cross-country Causes and Consequences of the Crisis: An Update", Special Issue: Advances in International Macroeconomics: Lessons from the Crisis, *European Economic Review*, Vol. 55/3, pp. 309-324, <http://dx.doi.org/10.1016/j.eurocorev.2010.12.006>.
- Sarlin, P. (2013), "On Policymakers' Loss Functions and the Evaluation of Early Warning Systems", *Economics Letters*, Vol. 119, pp.1-7.
- Schularick, M. and A. Taylor (2012), "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008", *American Economic Review*, 102, No. 2, pp. 1029-61, <http://dx.doi.org/10.1257/aer.102.2.1029>.
- Taylor, A. (2012), "External Imbalances and Financial Crises", *NBER Working Paper Series*, No. 18606, <http://dx.doi.org/10.3386/w18606>.

Appendix

Table A1.1. **Crises and recessions by country**

	All recessions ¹	Severe recessions ¹	Crises ²	Banking crises ²	Currency crises ²	Sovereign debt crises ²
AUS	7	2	1	1	0	0
AUT	3	1	1	1	0	0
BEL	4	2	3	1	2	0
CAN	5	3	1	1	0	0
CHE	6	4	2	2	0	0
CHL	2	1	3	1	4	4
CZE	3	1	2	2	3	0
DEU	5	1	3	3	0	0
DNK	6	2	5	2	4	0
ESP	4	1	3	2	3	0
EST	3	1	2	2	1	0
FIN	5	3	4	1	4	0
FRA	5	1	2	2	0	0
GBR	5	3	4	4	1	0
GRC	7	5	5	2	3	1
HUN	3	1	5	2	0	4
IRL	4	1	4	2	2	1
ISL	6	4	6	3	4	0
ISR	6	2	2	1	4	0
ITA	4	2	3	1	4	0
JPN	5	2	1	1	0	0
KOR	3	3	3	3	1	0
LUX	5	4	2	1	1	0
LVA	2	1	3	2	1	0
MEX	5	4	3	2	3	2
NLD	5	4	1	1	0	0
NOR	3	2	4	2	6	0
NZL	5	4	4	2	2	0
POL	1	0	2	1	0	2
PRT	5	3	2	1	1	0
SVK	2	2	1	1	0	0
SVN	1	1	2	2	0	0
SWE	4	2	4	2	3	1
TUR	6	5	7	4	7	2
USA	4	3	2	2	0	0
Total	149	81	102	63	64	17

1. For the period 1970-2014, except for the following countries for which GDP series are not available for the full period: Chile (1986-), Czech Republic (1990-), Poland (1990-), Slovenia (1990-), Hungary (1991-), Slovak Republic (1993-), Estonia (1995-), and Latvia (1995-). Peak and trough dates in GDP per capita are identified using the Bry and Boschan (1971) algorithm. A severe recession is defined as a fall in GDP per capita from peak to trough exceeding the median fall, which is 3.4% of peak GDP per capita across the full country-year sample.
2. Over the period 1970-2010. A crisis can be a twin or triple crisis, which is why column 5-7 do not sum to column 4.

Source: Babecký et al. (2014) and authors' calculations.

Table A1.2. Dataset

Indicator	Description	Data Source	No. countries	Time coverage	Transformations
Non-financial sector imbalances					
Total private credit	Lending from all sectors (including foreign) to private non-financial sector in per cent of GDP.	BIS	27	1970Q4-2014Q2	Level, gap1-3, growth 1-3
Private bank credit	Lending from domestic bank sector to private non-financial sector in per cent of GDP.	BIS	27	1971Q2-2014Q2	Level, gap1-3, growth 1-3
Household credit	Lending from all sectors (including foreign) to households in per cent of GDP.	BIS	27	1994Q4-2014Q2	Level, gap1-3, growth 1-3
Corporate credit	Lending from all sectors (including foreign) to non-financial corporations in per cent of GDP.	BIS	27	1994Q4-2014Q2	Level, gap1-3, growth 1-3
Asset market imbalances					
Real house prices	Deflated by CPI. Indexed to 2010 = 100.	OECD	33	1988Q1-2014Q2	Gap1-3, growth 1-3
Price-to-disposable income ratio	Nominal house prices to nominal net household disposable income per capita. Indexed to 2010 = 100.	OECD	29	1995Q1-2014Q2	Level, gap1-3, growth 1-3
Price-to-rent ratio	Nominal house prices to rent prices. Indexed to 2010 = 100.	OECD	32	1991Q1-2014Q2	Level, gap1-3, growth 1-3
Residential investment as % of GDP	Gross fixed capital formation, housing, in per cent of GDP.	OECD	34	1970Q1-2014Q4	Level, gap1-3, growth 1-3
Real equity prices	Share price index deflated by CPI.	OECD	35	1981Q1-2014Q4	Gap1-3, growth 1-3
External imbalances					
Current account balance	In per cent of GDP.	OECD	35	1975Q1-2014Q4	Level
Foreign currency exposure index (fxagg)*	Index of the sensitivity of a country's portfolio to a uniform currency movement by which the domestic currency moves proportionally against all foreign currencies. Index between -1 (zero foreign-currency foreign assets and only foreign-currency foreign liabilities) and 1 (only foreign-currency foreign assets and only domestic-currency foreign liabilities).	Benetrix, Shambaugh and Lane (2015)	34	1990-2012	Level
Quantitative foreign currency exposure (netfx)*	Quantitative exposure to a uniform shift in the value of the domestic currency against all foreign currencies. It is calculated as the foreign currency exposure index (fxagg) multiplied by the sum of foreign assets and liabilities in % of GDP.	Benetrix, Shambaugh and Lane (2015)	34	1990-2012	Gap1-3, growth 1, 3
Foreign exchange reserves	In per cent of GDP.	IMF	35	1970Q1-2014Q4	Level, gap1-3, growth 1-3

Table A1.2. **Dataset** (cont.)

Indicator	Description	Data Source	No. countries	Time coverage	Transformations
Foreign exchange reserves *	In per cent of money supply M2.	World Bank	31	1970-2013	Gap1-3, growth 1,3
Real effective exchange rate	Competitiveness indicator. Relative consumer prices (CPI), overall weights based on exports of goods.	OECD	35	1970Q1-2014Q4	Gap1-3, growth 1-3
Real effective exchange rate	Competitiveness indicator. Relative unit labour costs (ULC) for overall economy and overall weights based on exports of goods.	OECD	35	1970Q1-2014Q4	Gap1-3, growth 1-3
Export performance	Exports of goods and services relative to export market for goods and services.	OECD	35	1975Q1-2014Q4	Gap1-3, growth 1-3
Spillovers, contagion and global risk					
Trade openness	Sum of exports and imports in per cent of GDP.	OECD	35	1970Q1-2014Q4	Level
Global total private credit (% of GDP)	Weighted average of total private credit-to-GDP ratios across countries for each quarter. Weights defined by nominal GDP at Purchasing Power Parity (PPP).	BIS	27	1970Q4-2014Q4	Level, gap1-3, growth 1-3
Global private bank credit (% of GDP)	Weighted average of private bank credit-to-GDP ratios across countries for each quarter. Weights defined by nominal GDP at PPP.	BIS	27	1971Q2-2014Q2	Level, gap1-3, growth 1-3
Global real equity prices	Weighted average of country share price indexes for each quarter. Weights defined by nominal GDP at PPP.	OECD	35	1981Q1-2014Q4	Gap1-3, growth 1-3
Global real house prices	Weighted average of country real house price indexes for each quarter. Weights defined by nominal GDP at PPP.	OECD	33	1988Q1-2014Q2	Gap1-3, growth 1-3
VIX	Implied volatility of the S&P 500 index over the next 30 days. Calculated using a range of options on the S&P 500 index. The VIX is quoted in percentage points and can roughly be interpreted as the expected movement in the S&P 500 index over the next 30-day period.	Datastream	-	1990Q2-2015Q1	Level, gap1-3, growth 1-3

Note: The indicators are measured on a quarterly frequency, except for indicators marked with *, which are measured on an annual frequency. Up to six different transformations have been tested for each indicator and only the best in terms of the relative usefulness criteria is reported. Gap1: deviation from a recursive, slowly-adjusting HP-filter with smoothing parameter $\lambda = 400000$ for quarterly series ($\lambda = 1600$ for annual series); gap2: deviation from a recursive, faster-adjusting HP-filter with smoothing parameter $\lambda = 26000$ for quarterly series ($\lambda = 100$ for annual series); gap3: deviations from a 20 quarter (5 year) moving average; gr1: year-on-year growth rates; gr2: quarter-on-quarter growth rates; gr3: cumulated growth rates over the preceding 6 quarters (4 years for annual series).

Table A1.3. **Correlations among vulnerability indicators**

	Total private credit	Private bank credit	Household credit	Corporate credit	Real house prices	Price-to-disposable income	Price-to-rent	Residential investment	Real equity prices	Current account balance	Foreign reserves	Real effective exchange rate (CPI)	Real effective exchange rate (ULC)	Export performance	Trade openness	Global private credit	Global private bank credit	VIX	Global share price index	Global real house price index
Total private credit	1.00																			
Private bank credit	0.76	1.00																		
Household credit	0.78	0.82	1.00																	
Corporate credit	0.92	0.52	0.47	1.00																
Real house prices	0.42	0.49	0.34	0.24	1.00															
Price-to-disposable income	0.16	0.27	0.09	0.08	0.80	1.00														
Price-to-rent	0.29	0.39	0.23	0.16	0.92	0.86	1.00													
Residential investment	-0.35	-0.27	-0.22	-0.23	-0.37	-0.09	-0.30	1.00												
Real equity prices	0.46	0.44	0.42	0.33	0.47	0.17	0.34	-0.17	1.00											
Current account balance	0.41	0.30	0.31	0.38	0.00	-0.01	0.05	-0.20	-0.14	1.00										
Foreign reserves	0.01	0.11	0.05	-0.04	0.11	0.00	0.04	-0.09	0.03	0.09	1.00									
Real effective exchange rate (CPI)	0.12	0.14	0.13	0.02	0.11	0.10	0.09	0.01	0.14	0.02	-0.18	1.00								
Real effective exchange rate (ULC)	0.15	0.16	0.16	0.02	0.31	0.29	0.31	-0.06	0.23	-0.04	-0.13	0.76	1.00							
Export performance	0.22	0.22	0.18	0.13	-0.14	-0.14	-0.01	-0.11	0.03	0.18	-0.02	0.01	-0.13	1.00						
Trade openness	0.51	0.20	0.18	0.64	0.10	-0.10	-0.01	-0.22	0.21	0.24	0.21	-0.07	-0.12	-0.06	1.00					
Global private credit	0.41	0.36	0.40	0.28	0.47	0.01	0.22	-0.40	0.53	0.12	0.13	-0.03	0.09	0.01	0.27	1.00				
Global private bank credit	0.47	0.41	0.45	0.34	0.50	0.04	0.25	-0.41	0.57	0.13	0.16	0.01	0.10	-0.03	0.32	0.94	1.00			
VIX	0.08	0.07	0.04	0.06	0.03	-0.01	0.02	-0.04	-0.06	-0.02	0.02	-0.02	0.00	0.00	0.02	0.14	0.16	1.00		
Global share price index	0.39	0.35	0.36	0.26	0.49	-0.01	0.25	-0.32	0.54	0.04	0.16	-0.04	-0.05	-0.07	0.36	0.89	0.88	-0.11	1.00	
Global real house price index	0.48	0.45	0.47	0.30	0.56	0.11	0.34	-0.37	0.52	0.03	0.14	0.06	0.08	-0.14	0.31	0.85	0.83	-0.21	0.85	1.00

Note: Pairwise correlations across all countries and time periods for untransformed indicators. The table only include indicators on quarterly frequency.

Source: Author's calculations.

Table A1.4. Detailed full-sample performance of individual indicators

	Transformation	No. crises episodes	No. observations	Direction to be safe	Thresh. percentile	Rel. usefulness	Usefulness	aNTS	Type I error	Type II error	Cond. prob.	Diff. prob.	ALT
$\theta = 0.8$													
Global private bank credit (% of GDP)	gr3	65	3849	<	70	0.31	0.04	0.35	0.38	0.22	0.33	0.19	7.0
Global real equity prices	gap3	81	4488	<	75	0.30	0.04	0.35	0.43	0.20	0.35	0.19	6.7
Global real house prices	gap3	67	3670	<	65	0.27	0.03	0.43	0.35	0.28	0.30	0.14	7.0
Price-to-disposable income ratio	none	51	3140	<	80	0.26	0.03	0.29	0.54	0.13	0.36	0.22	5.9
VIX volatility index	gap2	55	3289	<	65	0.25	0.03	0.42	0.35	0.27	0.29	0.14	6.6
Price-to-rent ratio	none	65	3526	<	80	0.18	0.02	0.38	0.63	0.14	0.33	0.17	6.0
Global private credit (% of GDP)	gr1	67	3933	<	55	0.18	0.02	0.53	0.30	0.37	0.25	0.10	7.0
Real house prices	gap3	67	3670	<	85	0.15	0.02	0.39	0.69	0.12	0.32	0.17	5.8
Foreign reserves to M2 *	none	72	1072	>	15	0.14	0.02	0.48	0.72	0.13	0.34	0.14	1.3
Private bank credit (% of GDP)	gr3	65	3849	<	75	0.13	0.02	0.49	0.56	0.21	0.26	0.11	6.4
$\theta = 0.9$													
Global real equity prices	gap3	81	4488	<	50	0.27	0.02	0.51	0.19	0.42	0.27	0.11	7.0
Global real house prices	gap3	67	3670	<	55	0.24	0.02	0.48	0.23	0.37	0.27	0.12	6.9
VIX volatility index	gap2	55	3289	<	55	0.24	0.02	0.49	0.25	0.37	0.26	0.11	6.7
Global private bank credit (% of GDP)	gr3	65	3849	<	60	0.22	0.02	0.44	0.30	0.31	0.29	0.14	6.9
Global private credit (% of GDP)	gr3	66	3881	<	30	0.17	0.01	0.73	0.12	0.64	0.19	0.05	7.3
Price-to-disposable income ratio	none	51	3140	<	60	0.15	0.01	0.51	0.34	0.34	0.25	0.10	6.7
Price-to-rent ratio	gr3	61	3339	<	40	0.12	0.01	0.70	0.19	0.57	0.21	0.05	6.9
Corporate credit (% of GDP)	gap1	52	3118	<	25	0.09	0.01	0.82	0.12	0.72	0.18	0.03	7.4
Real house prices	gr3	63	3510	<	40	0.09	0.01	0.72	0.21	0.57	0.20	0.05	6.8
Real equity prices	gap3	81	4488	<	45	0.07	0.01	0.69	0.24	0.52	0.21	0.06	7.1

Note: The indicators are measured on a quarterly frequency, except for indicators marked with *, which are measured on an annual frequency. Relative usefulness measures the share of the usefulness of the indicator relative to a perfectly performing indicator (see section 3). Up to six different transformations have been tested for each indicator and only the best in terms of the relative usefulness criteria is reported. Gap1: deviation from a recursive, slowly-adjusting HP-filter with smoothing parameter $\lambda = 400000$ for quarterly series ($\lambda = 1600$ for annual series); gap2: deviation from a recursive, faster-adjusting HP-filter with smoothing parameter $\lambda = 26000$ for quarterly series ($\lambda = 100$ for annual series); gap3: deviations from a 20 quarter (5 year) moving average; gr1: year-on-year growth rates; gr2: quarter-on-quarter growth rates; gr3: cumulated growth rates over the preceding 6 quarters (4 years for annual series). aNTS stands for adjusted noise-to-signal ratio. ALT stands for average lead time and is measured as the number of quarters unless the indicator is indicated with a * in which case it is measured as the number of years. For the calculation of the statistics see section 3 in the main text.

Source: Authors' calculations.

Fortune or fortitude? Determinants of successful adjustment with IMF programs

by

Martin Larch, Kristin Magnusson Bernard and Peter McQuade*

Full adjustment programs in the wake of crisis episodes exact a major toll on a country's economy, yet not all are blessed with success. We identify adjustment needs by a country's decision to approach the IMF for official assistance. We then investigate the factors conducive to successful exit from official assistance during more than 170 adjustment episodes by means of a panel regression framework. In contrast to the existing literature, we do not use absolute benchmarks. We define success as a resumption of real GDP growth and a reduction of government debt compared to the pre-program period. Our econometric results suggest stringent policy action do play a role for the probability of success. At the same time, we also find that successful exit also very much depends on supportive external conditions and, linked to that, the degree of openness of an economy.

JEL classification: E61, F33, G01, H81

Keywords: Fiscal adjustment, financial crises, IMF lending

* Martin Larch and Peter McQuade work with the European Commission, Directorate-General for Economic and Financial Affairs, Kristin Magnusson Bernard with the European Central Bank. Correspondence: martin.larch@ec.europa.eu

Disclaimer: The views expressed in this paper are those of the authors and should not be attributed to the European Commission or the European Central Bank.

Acknowledgements: We are grateful to Uwe Boewer, Servaas Deroose, Zeno Enders, Juergen von Hagen, Paul van den Noord and Peter Pontuch and two anonymous referees for helpful comments and Salvador Barrios for access to his data set. Valuable suggestions by the participants of the 16th Banca d'Italia Public Finances Workshop in Perugia and the 2014 fiscal conference of the ECB are also gratefully acknowledged. Balint Tatar, Orhan Chiali and Jacek Szelozyński provided excellent statistical assistance.

1. Introduction

Defining the right adjustment strategy for getting back to a sustainable path after a crisis episode has become a topic of a heated debate. The global economic downturn that started in the summer of 2007 has been the most economically costly since the Great Depression. From its original epicenter in the U.S. financial sector, it quickly spread around the world, leaving plummeting growth and soaring unemployment in its wake. A large number of emerging market economies requested financial assistance from the IMF and regional financing sources as market-based financing at affordable terms disappeared.

A second wave of the crisis erupted in 2010 as concerns about debt sustainability in the euro area came to the forefront, exacerbated by fears of contagion in a monetary union and the associated detrimental knock-on effects on financial stability. By the end of 2013 four of the then seventeen euro-area countries were in full economic adjustment programs,¹ an unprecedented situation among advanced economies.

IMF support programs, including those combined with assistance from regional lenders such as the European Financial Stability Facility (EFSF) and the European Stability Mechanism (ESM), typically come with a series of policy conditions which in many cases require incisive adjustment measures. While political determination and faithful implementation play a crucial role, the final outcome of an adjustment program depends on an array of factors including those that are not necessarily under the control of policy makers.

In this paper we take a fresh look at what determines successful exit from IMF assistance programs. We do this by examining a large sample of IMF supported programs inceptioned during the years 1993-2010. We introduce a measure of performance, based on real GDP growth and government debt, which differs from the ones generally used in the literature, in that it relies on a relative not absolute threshold. It takes into account the performance of the country prior to the program. We believe that a performance-based criterion is superior, as it does not tar all countries with the same brush. A successful exit is likely to look different for catching up countries as opposed to more mature economies. Moreover, many existing definitions of success focus on the extent to which program conditionality is met. But then, compliance with program conditions may not be sufficient to ensure success if external factors are not propitious or, if propitious, do not spill over to the domestic economy.

The results derived from our empirical analysis largely confirm the findings in the literature with some interesting and important nuances. Among the variables that can be affected by policy choices, faster fiscal adjustment and a lower initial deficit contribute significantly and positively to a successful adjustment episode. Decisive financial sector repair is also conducive to successful adjustment as lack of credit to the private sector significantly lowers the chances of success, while a systemic banking crisis per se need not be detrimental. The role of exchange rate flexibility is less clear cut, possibly due to different off-setting effects. The probability of successful adjustment is considerably higher if the global growth situation is favorable and the economy is open enough to benefit from the

external conditions. Finally, while our primary aim is not to discuss the optimal design of IMF programs, our results suggest that more official financing (which conversely is typically coupled with a more gradual fiscal adjustment) does not significantly contribute to success, while more stringent conditionality and especially in the structural area does appear to exert a significantly positive effect. Our findings turn out to be robust across a wide range of specifications and for controlled variations in the sample.

The rest of this paper is organized as follows. Section 2 provides additional background and motivation to our work. Section 3 presents stylized macroeconomic facts for the countries in our sample, and discusses how to construct an indicator of successful adjustment. Section 4 presents results from regression analysis of factors that increase the probability of successful adjustments, including alternative definitions thereof. It also discusses implications for the recently completed programs in Europe. Section 5 concludes.

2. Motivation and literature

Formulating an adjustment strategy in the wake of a crisis episode requires a thorough understanding of a country's particular characteristics and the particular environment in which it operates: idiosyncrasies matter. Euro area countries, for instance, exhibit features that are very specific; most notably they are part of a monetary union where monetary policy is delegated to the ECB; the official financing made available through the recent EU-IMF supported programs for euro area countries was much larger than the historical average; and some of the EU-IMF-supported programs involved countries with very high government debt-to-GDP ratios. Nevertheless, learning some stylized facts from past crisis episodes and trying to build on past experiences is a valid strategy to improve policies going forward and to possibly assess the odds of success of on-going or recently completed adjustment programs.

To be useful, any analysis of factors conducive to successful adjustment after a crisis episode requires a systematic benchmark against which that success is evaluated. A first difficulty arises with how to identify a need for adjustment. Banking crises often, but not always, carry such economic costs that adjustment is needed. The same holds for both the speed and magnitude of fiscal adjustment, which might be very gradual if market or official financing continues to flow and debt levels are manageable.

This paper makes the assumption that a country requesting an IMF program is indicative of a significant adjustment need, as well as a certain urgency to advance the process to a considerable degree within a given time period. We consider the political costs of asking for official assistance to be sufficiently high that countries would not request it if adjustment needs were manageable otherwise. We acknowledge that the adjustment needs for countries under IMF programs vary substantially, including the actual need for financing, and will aim to control for this in our estimations.

This choice of sample could potentially bias the results, as countries availing themselves of official assistance might share certain characteristics likely to influence the results and ideally warranting the use of a control group not receiving financial assistance (Przeworski and Vreeland, 2000; Ghosh et al., 2002; Hardoy, 2003; Hutchison, 2004; Atoyan and Conway, 2005; Barro and Lee, 2005; Bas and Stone, 2011). The economic situation might be worse, which would reduce the likelihood of a successful adjustment, but the reform commitment might also increase with the external policy scrutiny. However, we still believe that this is the most appropriate sufficiently large sample available. Even within our potentially biased sample there are successful and unsuccessful adjustment episodes and

exploring the factors that discriminate between success and failure conditional on a country asking for official assistance is still useful.

A second issue regards how to formulate a benchmark against which the success of an adjustment episode is assessed. Some earlier literature finds that benchmark by looking for instance at process-related or accounting elements such as whether a program was completed without interruption or the share of conditions met during a program (Ivanova et al., 2001). Other studies focus on the stated aims of IMF financial assistance programs, such as providing short-run macroeconomic stabilization including stemming capital outflows. They typically assess the impact of IMF programs on economic growth (Dreher, 2006) or the current account (Ghosh et al., 2002).

Our first contribution is to construct an indicator of successful adjustment that improves on existing ones. Instead of using absolute thresholds for all countries regardless of their economic characteristics as IMF (2012) our definition is formulated relative to countries' pre-crisis levels of growth and debt, as they were sufficiently favorable for the country to finance itself on the capital markets. We believe that this approach is warranted given the highly varying country characteristics in our sample.

In a second step, we try to identify factors that predict successful adjustment according to the indicator by means of regression analysis with a larger sample and a more thorough delineation of factors that could be affected by certain policies – the 'hard work' – vis-à-vis those outside the countries' control such as global growth or risk appetite – 'help from abroad'. Our set of possible factors conducive to successful economic adjustment draws on previous literature, regardless of whether the adjustment took place under the aegis of an IMF-supported program or not. Barrios and Langedijk (2010) find that large current account deficits can significantly impair the ability of countries to achieve successful fiscal consolidations, but that absence of nominal exchange rate adjustment need not be a major impediment. The negative effects on growth and fiscal sustainability from banking crises, especially if they are preceded by a credit boom and followed by a credit crunch, were already studied by Calvo et al. (2006) and Cerra and Saxena (2008) but naturally came to the forefront during the last crisis given its origins in the U.S. financial sector (IMF, 2009; Laeven and Valencia, 2012; and Abiad et al., 2011).

Our empirical strategy of regressing economic and policy variables on a binomial indicator of successful macroeconomic adjustment relies on an approach that is well-established in the economic literature. In particular, there is a link with the relatively rich literature on successful fiscal consolidation which developed since the mid-1990s (see for instance Alesina and Perotti, 1995; Alesina and Ardagna, 1998; von Hagen et al., 1998; von Hagen et al., 2001; von Hagen and Strauch, 2001; Guichard et al., 2007; Larch and Turrini, 2011).

Most euro-area financial assistance programs have been completed fairly recently. Drawing firm conclusions about their success would therefore entail relying excessively on projections, which might be subject to bias. They are therefore not included in our sample, but we discuss their future challenges by assessing their probability of success using the estimated coefficients from our regression.

3. Features of adjustment spells and measures of success

Macroeconomic situation in countries receiving financial assistance.

In this section we conduct a descriptive analysis of the sample and provide an overview of trends in key macroeconomic variables prior to, during and following economic

adjustment programs. The sample covers 176 completed or expired IMF General Resource Account (GRA) supported programs extended to a total of 59 countries from 1993 and onwards.²

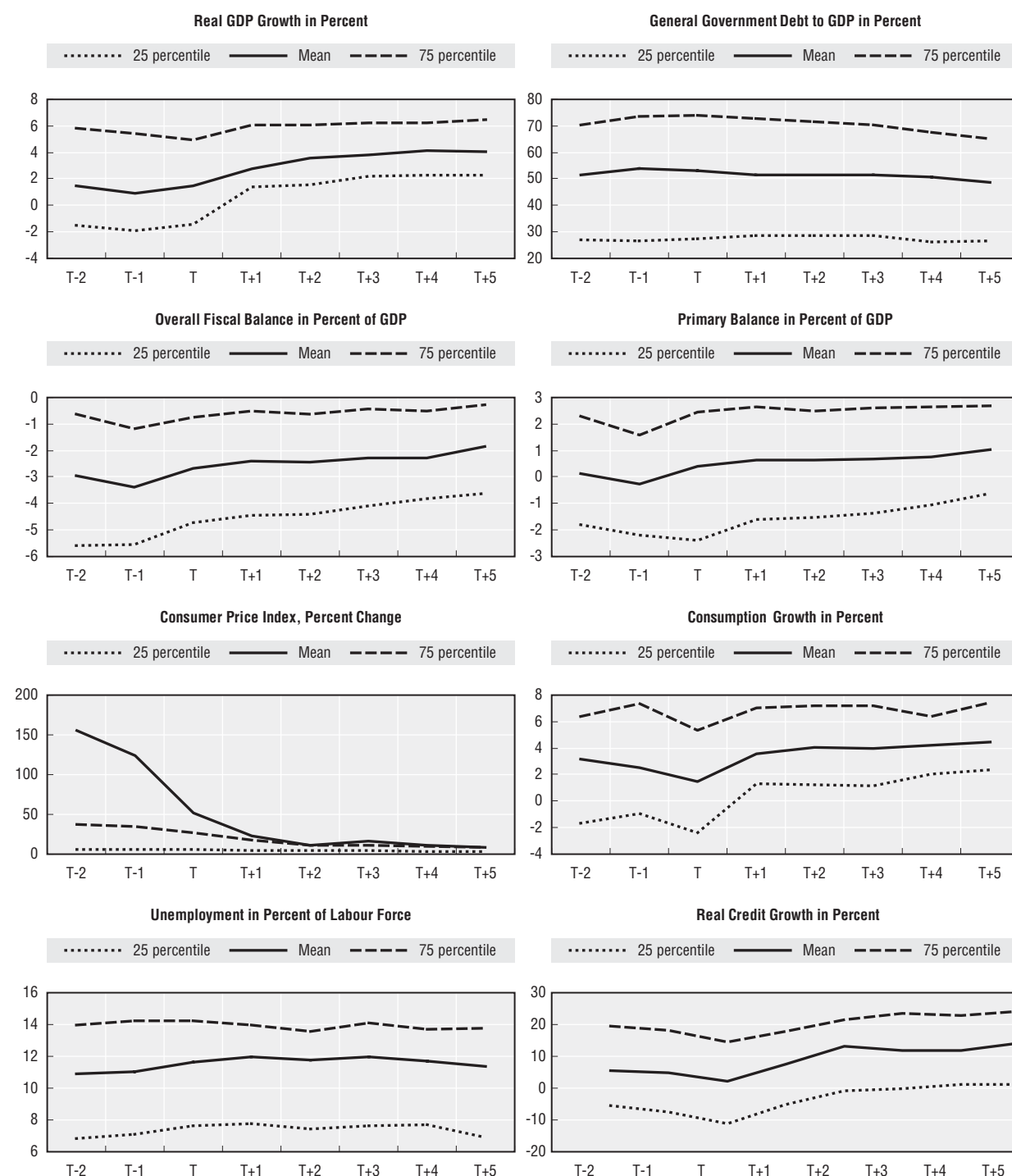
Among the 176 programs in the sample 83 percent were Stand-by-Agreements (SBA) and the remaining ones Extended Funding Facilities (EFF). On average a country had three programs with the IMF, however there are countries which completed up to seven GRA-supported programs. The average duration of an SBA amounted to 1.7 years with a maximum duration of three years, while EFFs were longer on average with a mean of 2.9 years and a maximum of four years. The average size of a GRA-supported program reached 234 percent of the country's quota with a standard deviation of 443 percent. Countries which were considered to meet the IMF's Exceptional Access Criteria³ in some cases reached above 3000 percent of the country quota. Nearly 30 percent of programs did not disburse any of the official available financing as countries could meet their financing needs through other sources. Countries drew on average 48 percent of the funds approved under the program, with the figure rising to 68 percent when non-disbursing programs are excluded.

A key decision for our analysis is to identify the horizon over which adjustment is assessed. In order to assess trends in key macroeconomic variables over the relevant horizon we define pre-program, program, and post-program periods annually.⁴ The year of program start is denoted with T , the pre-program period includes years $T-2$ and $T-1$, the program period comprises the interval $[T, T+2]$ and the post-program period refers to $[T+3, T+5]$. The appropriateness of the definition of the pre-program and the program period could be questioned as programs can start at any time in the year. We will later try to control for this in our econometric analysis by varying the starting point depending on the date of program inception. Moreover, policy measures tend to impact key macroeconomic aggregates with different lags. Decisively implemented measures aimed at improving the overall fiscal balance translate rather rapidly into fiscal headline figures, while structural reforms might impact real GDP growth with a delay of several years. Improvements in unemployment figures typically come only late in the recovery phase. Furthermore, using unemployment figures as an indicator for a revival of the labor market comes with disadvantages as a shrinking labor force (e.g. due to workers dropping out of the formal labor market) and falling employment might have offsetting effects on unemployment. However, unemployment remains the only labor market indicator available for a sufficiently large sample.

Another difficulty results from the fact that programs tend to have different durations, depending e.g. on the perceived time required for the needed adjustment. Average program duration within the sample is approximately 1.9 years with a standard deviation of 0.9 years. Therefore, from this perspective the definition of the program period as the time elapsed between T and $T+2$ seems appropriate, while bearing in mind that plenty of programs were shorter than two years. Some programs were immediately followed by a successor agreement as further adjustment was needed, and the improvement achieved in the post-program period might not be directly attributable to the original program. A further issue arises when defining the end of an adjustment program as programs can also end at any time in the year.

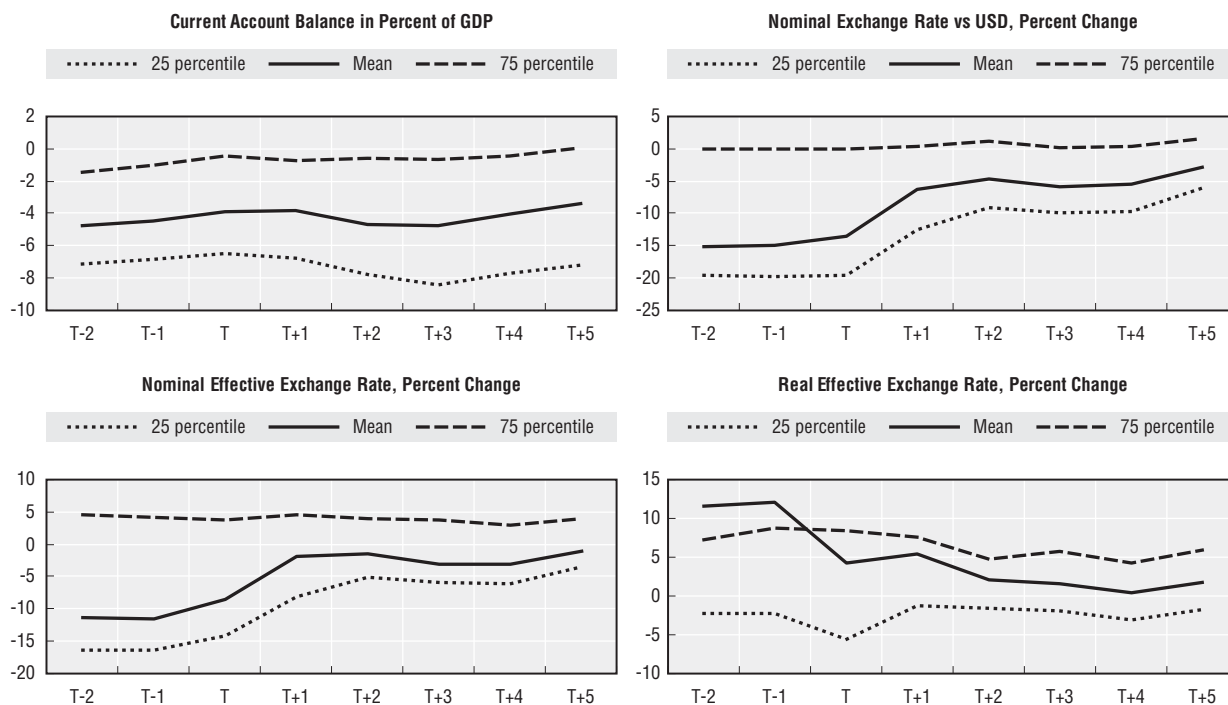
Bearing in mind the above caveats, the main objective of the descriptive analysis is not to draw firm conclusions about program episodes. Rather we try to get a preliminary feel for the data ahead of the more rigorous regression analysis. On average key macroeconomic indicators improve during an IMF supported economic adjustment program (Figure 1).

Figure 1. **Key macroeconomic variables in countries under an IMF economic adjustment program**



Countries seeking financial assistance recorded a slowdown of real GDP growth in the pre-program period, though average growth was still positive. A significant part of the countries suffered a recession as shown by the 25 percentile trend line. Fiscal indicators exhibit a similar pattern: both the overall fiscal balance and primary balance deteriorated

Figure 1. **Key macroeconomic variables in countries under an IMF economic adjustment program** (cont.)



Note: T denotes the year of program start.

Source: IMF World Economic Outlook Database, IMF reports, IMF International Financial Statistics, Darvas (2012), World Bank and other sources. For a detailed description of the data sources see Appendix; authors' calculations. Real credit growth, real effective exchange rate and nominal exchange rate vs. the USD are year-on-year percentage changes. The real effective exchange rate, the nominal exchange rate vs the USD and the nominal effective exchange rate are measured as the foreign currency price of one unit of domestic currency (indirect quotation). The mean of the change in the consumer price index is influenced by countries suffering from hyperinflation of above 1 000 percent per year.

pre-program and general government debt increased. The current account recorded substantial deficits during the run-up to the program but already improved just prior to the program which could be attributable to the fact that capital often flows out from countries suffering from macroeconomic disturbances. The interpretation of current account developments is, however, not straightforward. In many countries, a high current account deficit prior to the adjustment episode might be one of the vulnerabilities – particularly if it is financed by a large share of short-term portfolio flows – prompting it to seek financial assistance when financing dries up, and a reduction of it then suggests a more sustainable position going forward. However, in a fast-growing and capital-poor country standard economic theory recommends running a current account deficit to finance investment needs, and a widening current account deficit might then mean that external funds to finance such projects is again forthcoming post-crisis. Inflation was typically high before program start, but also influenced by outliers. Therefore, when assessing trends in inflation, it is more appropriate to take the 25 and 75 percentile lines as reference values. Real consumption growth was on a deteriorating path in the pre-program period. Unemployment was constantly growing prior to program start and real domestic credit growth – measured as the percentage growth in real credit advanced to the private sector – was rather sluggish and on a declining path. The currency typically underwent a marked depreciation against the US dollar during the first year of a program. Competitiveness – measured as a depreciation of the real exchange rate – typically worsened before a program and held steady thereafter.

In the year of the program start, real GDP growth slightly improved with economic growth returning gradually in the years afterwards. Countries that suffered from a deep recession in the pre-program period experienced an even stronger rebound as suggested by the 25 percentile line. Overall fiscal and primary balances also improved significantly in the start year and the adjustment continued under the program, while general government debt entered a declining path. When the general government debt-to-GDP ratio ranged above 50 percent at program start the average decline was more significant.⁵ The trend in the current account shows a rather mixed picture, while inflation returned to modest levels and was clearly on a declining path. Unemployment kept growing until T+1, in line with expectations of labor markets reacting with a lag to a rebound of the economy, but did not decline significantly later on and hence stabilized at a slightly higher level compared to pre-program levels. Inflation moderated after the program was put in place, furthermore real consumption growth accelerated and returned to a higher path compared with pre-program. Real credit growth rebounded during the program period and remained at a significantly higher level than prior to program start. Both the real effective exchange rate and the nominal exchange rate versus the US dollar and currencies of the given countries' trading partners stabilized gradually under the program.

Overall, on average key macroeconomic indicators appear to improve measurably during the program period and continue to evolve favorably after its end. Real GDP growth accelerated, the general government debt-to-GDP ratio declined, both the overall fiscal and primary balance improved, real consumption growth rebounded and the currency stabilized.

How to measure successful adjustment

This section sets the basis for our econometric analysis aimed at identifying factors which influence the outcome of an adjustment program. Prior to a more formal regression analysis we need a benchmark against which to evaluate success in order to identify factors enhancing the probability of successful exit from a macroeconomic adjustment program.

There are at least three conceivable ways of defining success: i) an "accounting" definition according to which a program is considered a success if compliance with the policy conditions agreed under the program is high; ii) a "market-based" definition that looks at whether market access is regained during or at the end of the program; and iii) a more "macroeconomic" definition which is centered on the economic performance of the country during the years following the completion of the program. The accounting definition, or variants of it, is the most common in the literature. Ivanova et al. (2001) are an early example; they use the share of policy conditions met by a program country or whether a program was interrupted or not. More recently, in the European context Pisani-Ferry et al. (2013) also assess program implementation on the basis of whether and how many key program conditions were met.

Ideally, the three definitions listed above should be complementary in the sense that success should manifest itself via a consistent compliance with policy conditions, combined with a return to market financing and a favorable macroeconomic performance after the end of the program. In practice, however, things are likely to be more complex. Specifically, not all policy conditions may be equally relevant and market confidence has shown to be volatile and not always forward-looking. Moreover, while compliance with policy conditions may provide an indication of the determination of policy makers, it is no guarantee of success in a broader sense if external economic conditions are not conducive. For these reasons, we concentrate on the sustainability dimension and gauge the success of an economic

adjustment program as a resumption of economic activity and improvements in the general government debt position.

In particular, we compare average real GDP growth in the five years after program start with average growth in the five years prior to program start to judge whether real GDP growth returned to the country's own benchmark level following the start of the adjustment program. In addition, we investigate whether general government debt entered a declining path within five years following program start. This differs from IMF (2012) and Barrios and Langedijk (2010) which both use absolute thresholds regardless of a country's own growth performance and ability to finance a certain debt level from market sources prior to the crisis. We believe this approach is warranted because a favorable growth and debt performance differ markedly within our sample due to the large variation in country characteristics, as well as differing market perceptions over time. Hence, those programs are deemed to be successful which recorded both favorable economic growth and declining public debt according to the following criteria:

1. Post-adjustment real GDP growth rate to reach 3/4 of pre-crisis one.
 - If average real GDP growth in $[T-5, T-1]$ was ≤ 3 percent then growth has to be above 2.25 percent to succeed.
 - If average growth in $[T-5, T-1]$ was between 3 percent and 6 percent then the average growth in $[T+1, T+5]$ has to reach at least 3/4 of the growth in $[T-5, T-1]$.
 - If average growth in $[T-5, T-1]$ was higher than 6 percent, then above 4.5 percent average growth in $[T+1, T+5]$ is necessary to succeed.
2. General Government Debt to GDP ratio to decline by 5 percent.
 - If average general government debt to GDP in $[T+1, T+5]$ was below 25 percent, then the trends in public debt are considered to be irrelevant from outcome perspective.
 - If average general government debt to GDP in $[T+1, T+5]$ was above 25 percent, then if general government debt peaked between T and $T+5$ and declined by at least 5 percent compared to the peak value, then the program is deemed to be successful as regards the evolution of general government debt.
 - The country does not default on its debt in $[T+1, T+3]$.

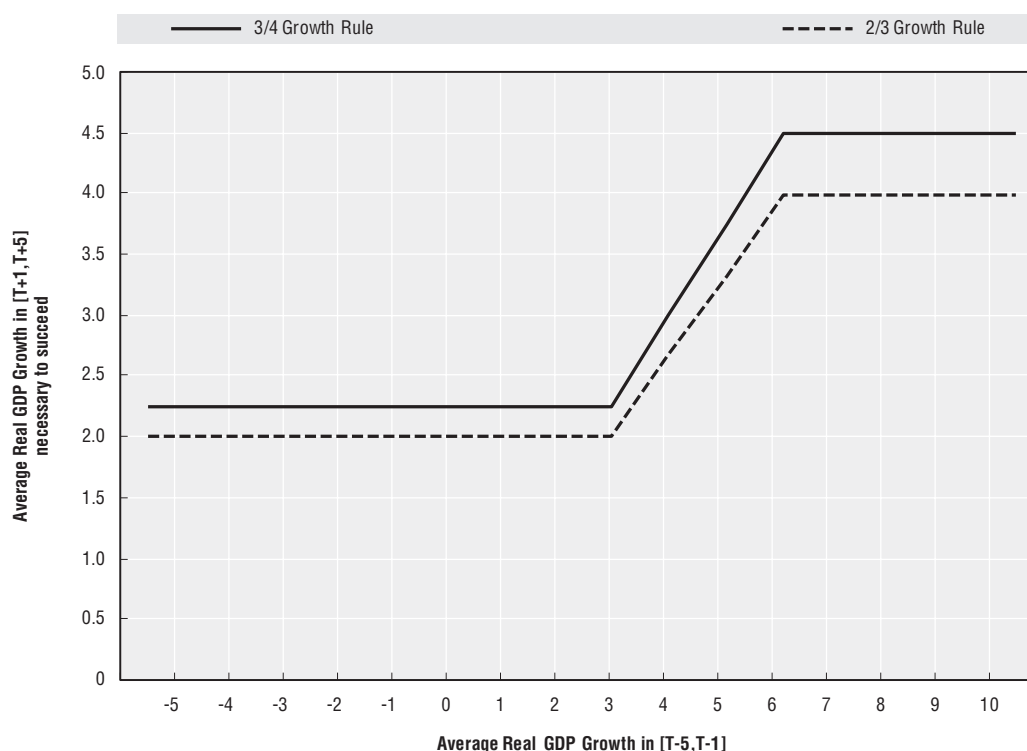
Growth criterion

Our rationale for choosing simple five-year averages for real GDP growth as opposed to e.g. measures of potential output is mainly to increase robustness. Lack of sufficient data for many countries precludes the use of e.g. a production function methodology for estimating potential growth. A more readily available approach such as the Hodrick-Prescott (HP) filter on the other hand suffers from the well-known end point problem (Mise et al., 2005).⁶ Another valid criticism is that for measuring the performance of an economy over the long-term, five years might be short at first sight. Here again we are trying to strike a balance between the availability of GDP data on the one hand, and, on the other hand, that policy measures implemented in the post-program period or afterwards could affect economic growth.

Our decision to use a three-pronged definition of success is motivated by both technical and economic concerns. Several countries in our sample were suffering from negative average real GDP growth prior to the program; and even for countries suffering from low, but still positive average growth, maintaining it cannot be considered meeting the key objective of putting the economy on a dynamic and sustainable growth path. Consequently, for the

low-growth cases (below 3 percent average growth before program start) at least an average growth of 2.25 percent in $[T+1, T+5]$ (3/4 of 3 percent) is required to succeed. A similar sustainability argument can be made for choosing an absolute cut-off point for countries recording high growth before program start, i.e. above 6 percent. The pre-crisis growth rates for these countries might simply have been symptoms of overheating, and as such it should not be considered a failure if they are not reached after the program. Therefore, for the high-growth cases the adjustment episode is deemed to be successful, if average real GDP growth reaches above 4.5 percent in $[T+1, T+5]$, which is again 3/4 of 6 percent.⁷ A graphic illustration of the growth criterion is provided in Figure 2.

Figure 2. **Growth criterion for success**

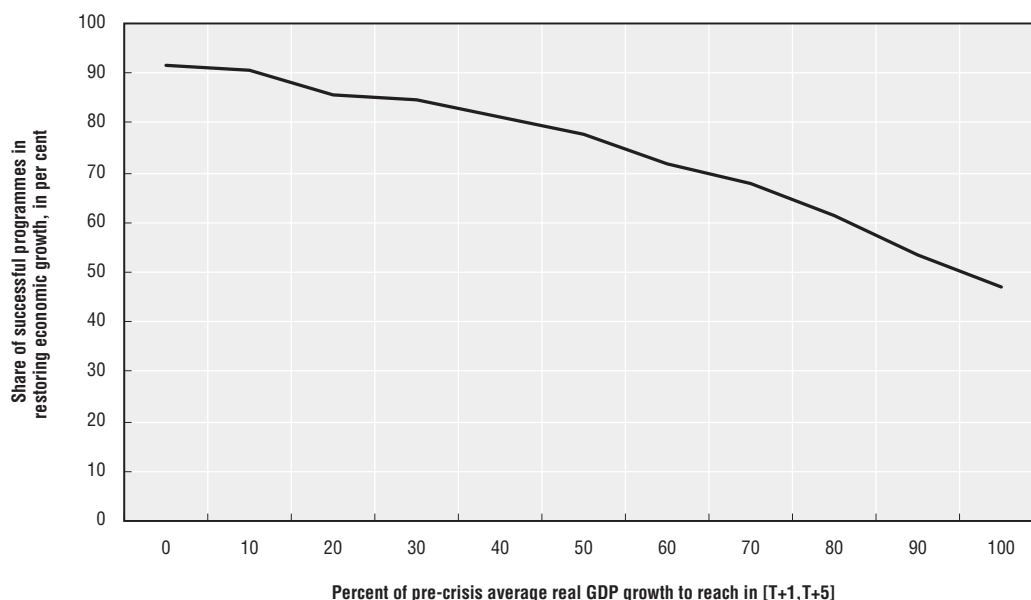


Note: T denotes the year of program start.

Source: Authors' calculations.

On the basis of our criterion for real GDP growth nearly 65 percent of adjustment episodes in our sample can be classified as successes. When we lower the bar so that only 2/3 of the average growth prior to the program is required to be reached after program start, the success rate increases only marginally to 68 percent. We therefore consider our definition to be fairly robust with respect to the cut-off points. To address outstanding concerns regarding the effect of borderline cases, we will later rerun the regressions using the modified criterion for growth to see whether results from the regression still hold. We also calculated the number of programs successful in restoring growth when modifying the required percentage of average real GDP growth in $[T-5, T-1]$ necessary to be reached in $[T+1, T+5]$ from 0 to 100 percent in 10 percent steps (Figure 3). From the figure it is evident that the lower bound for the indicator would be achieving about 50 percent of the pre-crisis average growth otherwise a far too high share of the adjustment episodes is judged as success. In addition, there is no sudden change in the slope of the line shown in the figure.

Figure 3. **Share of successful programs in restoring growth as a function of the percent of average pre-crisis real GDP growth to be reached in [T+1,T+5]**



Note: T denotes the year of program start.

Source: Authors' calculations.

Debt criterion

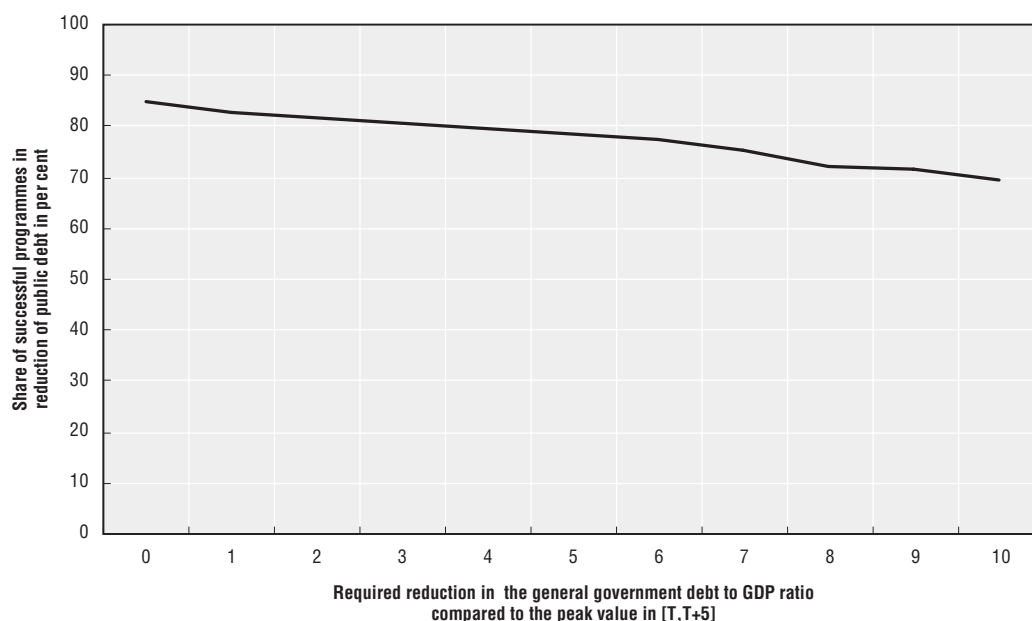
The aim of the debt criterion is to ensure a return to sustainable public finances, thus enabling continued market financing. We take a peaking of the general government debt-to-GDP ratio and a reversal in the trend to be the key signal of sustainability. The literature and policy prescriptions for how much debt needs to be reduced to be considered sustainable are inconclusive (Reinhart and Rogoff, 2010, Herndon et al., 2013, as well as e.g. the Maastricht criterion for EU countries). We take an agnostic view and consider an attempt at restoring fiscal sustainability as successful if it lowers the general government debt-to-GDP ratio by at least five percent in [T+1,T+5] compared with the peak value it reaches in [T,T+5]. Requiring a reduction of the debt-to-GDP ratio by an absolute five percentage points independently from the level of the debt-to-GDP ratio would privilege countries with high debt-to-GDP ratios “consolidating” through growing nominal GDP (denominator effect). A reduction in the level of general government debt instead affects the debt-to-GDP ratio in the same way independently from the level of the debt-to-GDP ratio.

A reduction in the level of general government debt can, however, also be achieved by a restructuring of the public debt, disorderly or planned. To make sure that only the latter are considered successes, a criterion with respect to disorderly default on outstanding debt had to be included in our definition. Accordingly, a default in [T+1,T+3] on debt would be considered a failure. The period [T+1,T+3] was chosen particularly with the aim to exclude those cases where a restructuring or debt release was part of the IMF-supported adjustment program, which was also cross-checked with program documents.⁸

This definition of the fiscal sustainability leads to a success rate of over 78 percent, i.e. substantially above the success rate found for the 3/4 economic growth criterion. Again, we checked for robustness of the debt criterion by shortening the time available for a reduction of the debt-to-GDP ratio to four years instead of five years and prolonged the period in which

no default or debt restructuring may occur by one year. This modified criterion lowered the success rate by about three percentage points, and as such the criterion can be argued robust from this perspective. The debt criterion was also tested for robustness with respect to the percent value by which the debt-to-GDP ratio is required to decline compared to the peak value. The figure below suggests that the debt criterion is fairly robust regarding the variation in the percent threshold as there is again no sudden change in the slope of the curve.

Figure 4. **Share of successful programs in debt reduction as a function of the required reduction of the debt-to-GDP ratio compared with its peak value in $[T, T+5]$**



Note: T denotes the year of program start.

Source: Authors' calculations.

Results

Finally, taking the two criteria together, we find that 103 programs are deemed to be successful which results in an overall success rate of 58.5 percent. This shows that countries which succeeded according to the growth criterion typically succeeded in achieving debt reduction as well. Table 1 provides an overview on the number of successes when using modified criteria. The table shows again that modifications of our definition lead to very modest changes in the success ratio.

Table 1. **Success rate of IMF-supported economic adjustment programs**

	General Government Debt to decline by 5% until T+5 and no default in $[T+1, T+3]$	General Government Debt to decline by 5% until T+4 and no default in $[T+1, T+4]$
Post-adjustment real GDP growth rate to reach 3/4 of pre-crisis one	58.5%	58.0%
Post-adjustment real GDP growth rate to reach 2/3 of pre-crisis one	60.8%	60.2%

Note: T denotes the year of program start.

We also assessed the joint distribution with respect to both criteria, i.e. the variation in the percent of the pre-crisis average real GDP growth to be reached in $[T+1, T+5]$ and the variation in the percent value by which the debt-to-GDP ratio is required to decline compared with its peak value in $[T, T+5]$.

Clearly, the debt criterion in our definition of success is not orthogonal to the growth criterion. A reduction of the debt-to-GDP ratio may result from consolidation efforts but also from an increase in the denominator, that is, economic growth. At the same time, the two criteria are not perfectly correlated and add information to the definition. We later control for the correlation across the two criteria by running regressions with a definition of success that relies on economic growth only.

A number of programs in our sample – 26 in total – are back-to-back, that is, the formal completion of one arrangement is followed by a new one. In 10 cases the predecessor programs are unsuccessful, based on our definition, 16 are successful. While such cases make the interpretation of success or failure a bit thornier (the recovery phase of the predecessor program overlaps with the implementation phase of the successor arrangement), we did not exclude them from our sample as they may still include useful information; this is certainly true for unsuccessful back-to-back programs. Hence, successor arrangements are not grouped to form one program; rather they are treated as separate programs. To assess the impact of back-to-back cases on our regression analysis we control for them when doing robustness checks.

4. Regression results

In this section we investigate factors associated with successful adjustment episodes as defined in the previous section. We first translate the success indicator using the growth and debt criterion into a binary variable, which is then used as the dependent variable in a probit regression. We assess the importance of a range of explanatory variables found to matter in the literature on fiscal and macroeconomic adjustment referenced in Section 2. Broadly

Table 2. **Explanatory variables used in the regression analysis**

Fiscal Balance in T	Fiscal balance in the year of program start measured as general government net lending in percent of GDP
Fiscal Balance adjustment	The change in general government net lending in percent of GDP under the program, percentage points
Real GDP Growth in T	Year-on-year percentage change
Primary Balance in T	General government net lending in percent of GDP excluding interest expenditures
Primary Balance adjustment	Change in primary balance in percent of GDP during the program, percentage points
Public Debt in T	General government debt to GDP at program start
World GDP growth	Average year-on-year percentage change between T+1 and T+5
Banking Crises	Dummy variable taking the value 1 if a banking crisis was ongoing in the year of program start using the definition in Laeven and Valencia (2012)
Credit Crunch	Dummy variable taking value 1 if real credit advanced to private sector recorded negative growth in at least two years between T and T+2
Exchange Rate Regime in T	IMF classification taking values from 1 to 15; a higher value indicates a more flexible exchange rate regime.
Openness Indicator	Measured as exports plus imports divided by GDP
VIX in T	Chicago Board of Exchange S&P 500 Implied Volatility Index capturing the risk appetite of the market and taking high values in times of turbulence and crisis
Nominal Effective Exchange Rate; change	Adjustment in nominal effective exchange rate under the program. An increase in the NEER is equivalent to an appreciation
Current Account Balance	In percent of GDP
Change in Current Account Balance	Percentage point change in the current account balance under the program

Note: T denotes the year of program start.

speaking, these can be divided into those that can be influenced directly or indirectly by policy action during the period of adjustment – the ‘hard work’ part – and those outside the control of the country in question – the external conditions. The following list provides an overview of the explanatory variables, where T denotes the year of program start. More details about the variables used and the respective data sources are provided in the Appendix.

For the estimation we use a panel probit with a population-averaged (PA) generalized estimating equation (GEE) approach with exchangeable correlation matrix based on Liang and Zeger (1986). Notice that our panel is not a ‘real’ panel in the sense that each program constitutes one observation and years in which no program was put in place were excluded from the sample. Nevertheless, since many countries had more than one program our model goes beyond a pure cross-section analysis. A summary of the estimations is provided in Table 3.1.

Table 3.1. **Factors conducive to successful adjustment; results from probit regressions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Primary Balance in T	0.077* (0.039)	0.092** (0.040)	0.093** (0.038)	0.091*** (0.033)	0.099*** (0.031)	0.090*** (0.029)	0.082** (0.033)	0.114*** (0.038)	0.118*** (0.039)
Primary Balance Adjustment	0.087** (0.039)	0.073** (0.037)	0.075* (0.039)	0.068* (0.038)	0.077** (0.039)	0.085** (0.040)	0.090** (0.036)	0.128*** (0.040)	0.121*** (0.043)
Public Debt in T	-0.004 (0.003)	-0.006 (0.004)	-0.007* (0.004)	-0.007* (0.004)	-0.006 (0.004)	-0.008* (0.004)	-0.008** (0.004)	-0.007** (0.003)	-0.008** (0.003)
Real GDP Growth in T	0.056*** (0.019)	0.059*** (0.022)	0.073** (0.030)	0.064** (0.032)	0.081*** (0.029)	0.096*** (0.026)	0.093*** (0.027)	0.094*** (0.031)	0.092*** (0.030)
World GDP Growth [T+1,T+5]		1.388*** (0.256)	1.489*** (0.285)	1.573*** (0.309)	1.673*** (0.309)	1.907*** (0.286)	2.115*** (0.287)	2.171*** (0.311)	2.272*** (0.349)
Banking Crisis Dummy			0.783** (0.318)	0.868*** (0.333)	0.816** (0.342)	0.898*** (0.305)	1.042*** (0.285)	1.240*** (0.384)	1.254*** (0.383)
Credit Crunch Dummy				-0.497** (0.233)	-0.531** (0.231)	-0.536** (0.231)	-0.497** (0.244)	-0.530** (0.263)	-0.548** (0.270)
Exchange Rate Regime					0.067** (0.031)	0.085*** (0.031)	0.064* (0.033)	0.092*** (0.032)	0.088*** (0.033)
Openness Indicator						0.008** (0.003)	0.009*** (0.003)	0.008** (0.004)	0.007* (0.004)
CBOE S&P 500 Volatility Index							-0.057*** (0.020)	-0.044** (0.020)	-0.049** (0.021)
NEER Adjustment								0.022** (0.010)	0.023** (0.010)
Current Account Balance in % of GDP in T									-0.004 (0.015)
Adjustment in Current Account Balance									0.022 (0.021)
No. of Observations	176	176	176	176	176	176	176	176	176

Note: Standard errors in parentheses: p < 0.01, ** p < 0.05, * p < 0.1. For variable definitions see Table 2. T denotes the year of program start.

The first and most parsimonious specifications includes only basic macro-fiscal variables (column (1) and (2) in Table 3.1). Our hypothesis is that the degree of initial fiscal vulnerabilities matter for the success of the adjustment. In particular, we include the primary budget balance and government debt at program start in percent of GDP as well as the improvement of the primary budget balance ratio over the program horizon. We estimated the model using the primary instead of the headline balance and the adjustment thereof because the measure excluding interest expenditure provides a better measure of the de facto fiscal adjustment.⁹ In addition, countries agreeing to a GRA-supported program typically cover some of their financial needs by drawing on the credit provided by the program. This is usually extended at somewhat concessional terms and

has a bearing on the interest bill which needs to be kept in mind when extrapolating the results to discuss adjustment episodes in general. We find both the primary balance at program start and its adjustment to contribute to successful adjustment in a positive and highly significant manner, which is in line with what theory would suggest as well as with the results in the earlier literature on successful fiscal consolidation referred to above: the lower the initial fiscal imbalance and the larger the fiscal adjustment the higher the odds for success (e.g. Larch and Turrini, 2011).

It is important to note that our results do not indicate whether the speed of adjustment was in any way optimal, in the sense of striking the right balance between its impact on economic activity and progress towards sustainable public finances. We refer here to the debate about the size of fiscal multipliers triggered by Blanchard and Leigh (2013) according to which fiscal consolidation could be self-defeating in view of larger than expected fiscal multipliers. Based on their findings, Blanchard and Leigh have argued in favor of a less-now-and-more-later type of fiscal adjustment. Evidently, this does not clash with our finding about the role of fiscal consolidation for successful macroeconomic adjustment: Blanchard and Leigh focus on the short term, while our analysis looks at the medium-term horizon when possible negative effects of fiscal consolidation are likely to be outweighed by positive effects. There is a growing body of empirical literature showing that fiscal consolidation triggers a medium-term adjustment process, typically via the labor market, which has a positive impact on a country's competitiveness and, ultimately, its aggregate level of economic activity; see Alesina and Perotti (1995) for an early contribution to this field, as well as were Hernandez de Cos and Moral-Benito (2014) and Lamo et al. (2014) for recent and specific evidence of this labor market or competitiveness channel.

There is also a vivid discussion on the level at which, if at all, public debt impacts negatively on growth (Reinhart and Rogoff, 2010; Herndon et al., 2013). In our framework, we both require a positive association with growth and material reduction of debt to consider the adjustment process a success, and can, therefore, only include the initial level of debt among our explanatory variables to avoid spurious correlations. We find only a small and not significant negative effect of the debt-to-GDP ratio at program start in our parsimonious specification.

Real GDP growth at program start was positively associated with successful adjustment episodes, which simply and intuitively suggests that adjustment episodes where growth had already returned at program inception had a higher probability of success.

The recent crisis shed light on the role of the financial system during adjustment episodes, with emerging conclusions that banking crises are typically associated with slower and more protracted recoveries. Therefore, we would expect that a banking crisis affects the outcome of an adjustment episode in a negative way. However, and in contrast to existing research, in our regressions the corresponding dummy variable is estimated to exert a positive effect (column (3)). A closer look at the banking crises identified by Laeven and Valencia (2012) however show that while the effects on GDP growth and public debt are typically negative, they vary widely. One possible explanation is that the gravity of initial conditions amplifies political awareness: the worse the starting position, the easier it is for policy makers to table and agree on necessary measures. Such a link between initial conditions and political determination is corroborated by studies on fiscal adjustment (Barrios et al., 2010; Larch and Turrini, 2011). Moreover, banking crisis may weigh on the adjustment only in instances when they lead to a protracted lack of credit that the recovery is hurt (Abiad et al., 2011; Calvo et al., 2006).

We therefore included a credit crunch dummy taking on the value one if real credit growth is flat or negative for two years during the period $[T, T+2]$. The results in column (4) show that a credit crunch has a significantly negative influence on the outcome which is in line with a priori expectations. This again suggests that hard work, i.e. sufficient repair of the financial sector after banking crisis, pays off in terms of a higher likelihood of a successful adjustment. While we do not measure financial sector repair directly, empirical evidence shows that an impaired banking sector significantly weighs on a country's macroeconomic performance (e.g. Nkusu, 2011). Therefore, more active and effective policies aimed at restoring credit growth are conducive to a successful adjustment.

Turning to external variables, we find that external demand (average real world GDP growth for the period $[T+1, T+5]$) had a strong positive association with success (column (2)). This is in line with expectations as increasing demand for export goods is certainly supportive for economic growth and may also help cushion negative effects of a decline in domestic demand on the back of fiscal retrenchment. To underpin this hypothesis from an econometric perspective, openness was added to the model as well with results reflected in column (6). The positive coefficient suggests that the chance of a successful exit from an IMF adjustment program increases with the degree of openness, again as expected.

A more flexible exchange rate regime is typically found to be helpful for economic adjustment through its favorable price effects on export goods, which is also confirmed by our model in column (5). However, currency flexibility, or more specifically depreciation, may also have unfavorable effects. A high degree of pass-through to import prices may partly offset the gain resulting from relatively cheaper export goods relative to the rest of the world.¹⁰ If debt is denominated in foreign currency, a depreciation of the home currency leads to higher debt servicing costs.¹¹ We included the percentage change of the nominal effective exchange rate (NEER) to investigate the relative importance of these channels, and the results in column (8) show us that its effect is significantly positive, suggesting that an appreciation of the home currency is positively associated with successful adjustment. This could suggest that the negative effect of a currency depreciation on the debt servicing costs in the short-run outweighs the possible benefits which would arise from a gain in competitive advantage, but could also be a spurious correlation in the sense that both the resumption of growth and the appreciation of the currency stem from enhanced confidence in the sovereign more generally.

International investors' risk appetite is also likely to matter for successful adjustment, as it could result in different perceptions of countries' creditworthiness over time regardless of their economic fundamentals. We proxy risk appetite with the VIX index as in the existing literature. According to the results reported in column (7) a higher implied volatility – or conversely, lower risk tolerance – is significantly negatively associated with successful adjustment.

A large current account deficit is often seen as an important vulnerability that could be incompatible with successful adjustment, although the expected effects are not clear-cut as discussed earlier.¹² We included the initial position and the adjustment carried out in the current account under the program in the regression, but did not obtain significant results.

Most of our explanatory variables are measured at the onset of the financial assistance program (period T) and should not be affected by subsequent events. At the same time, some of our explanatory variables could potentially be endogenous to an adjustment program. In particular, a forceful implementation of a program could generate confidence effects which

then may spill over to the primary budget balance. To address this possibility we instrument the observed change of the primary budget balance over the change in public debt, the deficit and the current account balance in the years leading up to the crisis.¹³ The current account balance is included because in many cases, even if first incurred by the private sector, in the event of a crisis foreign debt is ‘unloaded’ on the government sector. We implement the instrumental variables using a probit model with clustered errors, where the cross-sectional identifier is specified as the country to allow for intragroup correlation. The results are reported in Table 3.2.

Table 3.2. **Factors conducive to successful adjustment; results from instrumental variables probit regressions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Primary Balance in T	0.155 (0.110)	0.193*** (0.064)	0.181*** (0.065)	0.240*** (0.050)	0.256*** (0.036)	0.258*** (0.040)	0.246*** (0.036)	0.254*** (0.039)	0.269*** (0.057)
Primary Balance Adjustment	0.280 (0.240)	0.296*** (0.113)	0.227* (0.134)	0.323*** (0.094)	0.310*** (0.103)	0.369*** (0.080)	0.355*** (0.093)	0.363*** (0.091)	0.417*** (0.059)
Public Debt in T	-0.005 (0.008)	-0.006 (0.005)	-0.006 (0.005)	-0.008* (0.004)	-0.006 (0.005)	-0.009** (0.004)	-0.010** (0.004)	-0.010** (0.004)	-0.013*** (0.004)
Real GDP Growth in T	0.079 (0.055)	0.084* (0.044)	0.133*** (0.045)	0.090** (0.042)	0.123*** (0.044)	0.106** (0.050)	0.099** (0.046)	0.098** (0.046)	0.069 (0.055)
World GDP Growth [T+1,T+5]		1.354** (0.563)	1.898*** (0.531)	2.071*** (0.677)	2.400*** (0.695)	2.288** (0.890)	2.485*** (0.913)	2.513*** (0.891)	1.796 (1.300)
Banking Crisis Dummy			1.045*** (0.380)	1.281*** (0.485)	1.280** (0.500)	1.200** (0.500)	1.267** (0.501)	1.358*** (0.499)	1.170** (0.569)
Credit Crunch Dummy				-0.964** (0.390)	-1.061*** (0.368)	-0.902** (0.376)	-0.955** (0.400)	-0.930** (0.383)	-0.656 (0.434)
Exchange Rate Regime					0.086** (0.039)	0.094** (0.042)	0.097** (0.041)	0.108*** (0.042)	0.078 (0.050)
Openness Indicator						0.006** (0.003)	0.008** (0.003)	0.007** (0.003)	0.004 (0.004)
CBOE S&P 500 Volatility Index							-0.025 (0.030)	-0.021 (0.031)	-0.005 (0.035)
NEER Adjustment								0.014 (0.012)	0.018 (0.011)
Current Account Balance in % of GDP in T									-0.011 (0.023)
Adjustment in Current Account Balance									-0.003 (0.052)
No. of Observations	136	136	136	136	136	136	136	136	136

Note: Standard errors in parentheses: p < 0.01, ** p < 0.05, * p < 0.1. For variable definitions see Table 2. T denotes the year of program start.

Reassuringly, there are only two variables, which, while preserving the expected algebraic sign, loose statistical significance in the instrumental variable approach: the indicator gauging global risk appetite and the nominal effective exchange rate. For the rest, the results largely confirm the findings obtained from the simple probit regressions, especially as regards the role played by the starting conditions – both macro and fiscal –, the size of the fiscal adjustment, the external environment and the degree of openness.

The estimated coefficients in Table 3.1 are useful to assess the sign and the statistical significance of the explanatory variables in our model. They do not gauge the marginal impact on the probability of success. Because of the nonlinearity of the underlying probit model, the marginal effect also depends on the values of all explanatory variables. Table 4 shows the average partial effect (APE) and the partial effect at the average (PEA) of each explanatory variable. The impact of average world GDP growth after program start is the

highest while the marginal effect of openness and public debt are the lowest. However, this picture is distorted as the standard deviation differs substantially across explanatory variables. In order to improve comparability, the marginal impact was recalculated by adding one standard deviation to the particular explanatory variable and keeping all other variables at their sample average. The last column in Table 4 shows the respective results. The relative impact of average world GDP growth is still the strongest on the probability of success, nevertheless the differences in the magnitude of the effects are much smaller across variables.¹⁴ These findings underpin the conclusion that the success of an adjustment program is significantly influenced by ‘external help’, that is, the external economic environment prevailing during the program. Yet policy measures that impact on public finances and the financial sector are also important.

Table 4. **Partial effects on the probability of success**

Variable	Mean	Std. Dev.	Min	Max	Average partial effect	Partial effect at the average	Average impact of $dx = \sigma_x$	Impact of $dx = \sigma_x$ at the average
Public Debt in T	53.273	37.407	5.123	270.183	-0.002	-0.003	-0.065	-0.103
Real GDP Growth	1.517	5.429	-16.228	15.600	0.023	0.034	0.117	0.161
Primary Balance in T	0.386	4.724	-19.543	17.423	0.027	0.042	0.123	0.168
Primary Balance Adjustment	0.195	3.639	-17.379	17.573	0.030	0.047	0.107	0.149
World GDP Growth	3.702	0.444	2.909	4.789	0.518	0.791	0.207	0.253
Banking Crisis	0.216	0.413	0.000	1.000	0.265	0.356	0.117	0.161
Credit Crunch	0.386	0.488	0.000	1.000	-0.129	-0.196	-0.062	-0.099
Exchange Rate Regime	7.830	4.439	1.000	15.000	0.022	0.034	0.095	0.134
Openness Indicator	81.952	39.025	14.679	260.632	0.002	0.003	0.069	0.100
VIX - S&P 500 Volatility Index	20.429	6.599	12.372	32.583	-0.010	-0.016	-0.070	-0.111
NEER Change	-1.759	30.580	-87.977	329.152	0.005	0.008	0.150	0.199

Note: For variable definitions see Table 2. T denotes the year of program start. The average partial effect and the partial effect at the average for the binary variables, banking crisis and credit crunch, was obtained by calculating the difference in the predicted probability when altering the value of the particular variable from zero to one. ‘ σ_x ’ is defined as the standard deviation of variable ‘x’. The probability of success across all programmes in our sample based on average values of the explanatory variables in our baseline specification is 66.4 percent.

Extensions

One factor of interest is whether the degree of imbalances experienced by a country affects the results. We proxy the degree of imbalances with whether the IMF program was disbursing or not, with the hypothesis that smaller adjustment needs meant that the country could continue to finance itself on the market and did not need to draw on official financing. We therefore added a dummy variable reflecting whether disbursements occurred under the program.¹⁵ Results suggest that this does not play a role as the coefficient is slightly negative and does not substantially differ from zero (results see Table A.2.1 in the Appendix). The share of the funds drawn under the program relative to the amount available and the size of the program relative to the country’s economy were also found to be insignificant. The results did not change significantly when the instrumental variables approach was applied (see Table A.2.2 in the Appendix).

We also investigated which types of economic reforms are most conducive to successful adjustment. To ensure a common definition, we identify these reforms by way of the conditionality agreed under the program. Broadly speaking there exist two different subgroups of conditions, quantitative performance criteria (QPC) and structural conditions.¹⁶ QPCs are quantifiable and measurable criteria while structural conditions are often non-quantifiable criteria and consist of policy measures aimed at implementing structural

reforms (e.g. to the labor or product markets) and in general correcting disruptions to the supply side of the economy. First, we assess whether the total amount of criteria per se matter for the chance of success. Results show that the probability of a successful exit is increasing with the overall number of criteria and conditions (see Table A.3.1 in the Appendix). We also included separately the total number of the QPCs and the total number of structural conditions into the model. We found that the impact of both the QPCs and the structural conditions is positive but only the coefficient of the structural conditions is significant. Five years might appear to be a rather short period of time for the assessment of the impact of structural reforms, but we tried to strike a balance between allowing a sufficient time frame for the full effects to be felt while not letting so much time pass that we in fact mainly picked up the effects of other factors. Finally, there are three different types of structural conditions, so called prior actions (formerly: conditions for completion of the review), structural performance criteria and structural benchmarks. We estimated the regression including the QPCs and three different types of structural conditions separately. Results show that only prior actions and structural performance criteria have a statistically significant positive impact on the outcome. The importance of structural reforms is confirmed by the two-stage instrumental variables approach which addresses the possible endogeneity of the fiscal adjustment (see Table A.3.2 in the Appendix). In fact, only structural performance criteria, which typically refer to growth-enhancing policy measures for the medium to long run, preserve their statistical significance.

In addition, to assess the possible impact of the prevailing political situation in a country implementing an IMF-supported adjustment program we tested a number of variables from the World Bank Political Institutions data base. We focused attention on indicators capturing or gauging the political clout of government such as the Herfindahl index, the vote share of the government parties, and the margin of majority. Our prior was that a stronger government ensures a higher degree of political ownership, a stronger implementation and finally a higher probability of success. This prior was not confirmed by our estimates: either the political indicators were not significant and/or had the wrong sign.

Finally, we also tested the possible role played by the degree of economic flexibility. Our prior was that more flexible economies should find it easier to meet our definition of success when going through a crisis. We used the well-known Fraser index of economic freedom as a general proxy for economic flexibility with rather mixed results.¹⁷ The estimated coefficient turns out to be statistically significant only occasionally but is not robust.

Robustness checks

We performed a number of sensitivity checks and found that our results hold up.¹⁸ First, we reran the baseline regression with the modified criteria for success outlined in Section 3. Using the modified criterion for public debt, i.e. changing the relevant horizon for the reduction to occur, and maintaining the original criterion for economic growth did not substantially change the coefficients. When applying both the modified criterion for economic growth, i.e. lowering the bar to 2/3 of average growth, and public debt reduction, we again obtain similar results. Lastly, when using the modified criterion for economic growth and the original criterion for public debt reduction we obtain the same results and all variables are significant.

We also tested the baseline regression using only the 3/4 growth criterion as the dependent variable. The results mainly hold up with the exception that the impact of the adjustment in the nominal effective exchange rate variable is less than one third if compared

with the baseline regression and also turns insignificant which again appears to support the hypothesis on the link between exchange rate flexibility and debt reduction posited earlier.

We re-classified the 15-notch variable of exchange rate flexibility into a binary one (fixed vs. non-fixed exchange rate regime) and found that a fixed exchange rate regime has a negative impact on success which is in line with findings from the baseline model. Throughout our regression analysis we assumed that the public debt-to-GDP ratio enters the regression equation as a linear variable. In order to capture possible non-linear effects linked to a high debt-to-GDP ratio we used squared values and found that the results remain unchanged. We also included a dummy variable for highly indebted countries at program start instead of the debt-to-GDP ratio. The coefficient turned out to be negative, yet it failed to be significant even when varying the cut-off point in the definition of a “highly-indebted country”.

We applied gradual cut-offs in terms of income levels to our sample and found that our results held up, although some variables became insignificant when the sample size shrunk by more than a third. Our results also held up to the exclusion of countries in early stages of post-communism transition and to the exclusion of a specific region as well (e.g. Asia, Latin-America and the Caribbean, non-EU Europe, Europe).

A number of financial programs which only lasted for a rather short period and were succeeded by the next agreement in the following year are included in the sample, which makes it difficult to determine the length of the adjustment episode. We excluded back-to-back programs where the predecessor program was unsuccessful. However, this did not have a significant impact on the regression results in our baseline specification.

We also changed the reference year as a significant part of the adjustment in the primary balance was carried out in the year of the program start. We estimated the baseline model using T-1 (the year before program start) as the reference year for the initial condition with respect to the primary balance and the adjustment carried out, and found no significant difference in the coefficients compared with the baseline model, which again indicates that exiting the program with sound public finances positively contributes to program success. We also redefined the reference year for the program end contingent on in which month in the year the program ended and found that results remain similar.

Comparison with market perceptions of success

A common view of a successful adjustment following a financial assistance program is that the country can again fully finance itself from the markets, without the aegis of an IMF-supported program. To compare this notion of success with our indicator requiring improvement in the growth and debt situation, we define market-based success as a country not requesting another IMF program within a certain time frame.¹⁹ Another advantage of this definition is that market participants should take spill-overs between different government securities into account when making their investment decisions, and therefore this measure of success ought to internalize the cross-country effects of adjustments in different countries to a greater extent than our growth-debt indicator. By contrast, a clear downside of this definition pertains to the question of how sustainable market rates are in the medium to long term. Concretely, market access may be regained at the price of higher yields.

We find that 36 percent (32 percent) of countries did not request a follow-on program within a two (three) year horizon. This means that the success rate for a market-based

indicator is about half of the one using the growth and debt situation, which could be taken as evidence that contagion concerns are prevalent for market participants when assessing the creditworthiness of a sovereign exiting an adjustment episode. When this success rate is used as the dependent variable instead, most of the estimated coefficients still have the expected signs but their magnitude and significance change somewhat compared to the growth-debt criterion. The market-based indicator gives relatively higher weight to the countries' own economic characteristics, i.e. growth at program start, fiscal adjustment and openness. Of the external variables, only the VIX remains significant while trading partner demand (which was found to be a key explanatory variable for the previous indicator) becomes insignificant.

What do our results imply for recently completed programs in europe?

In this section we use the estimated coefficients to investigate implications for the adjustment programs in Europe which alongside contributions from European lenders also involved IMF assistance. Specifically, we calculate the probability of success for the programs for Ireland launched in 2010, Portugal in 2011, Cyprus in 2013, Romania in 2011 and 2013 and the second adjustment program for Greece of 2012. We do not cover the current Greek program of 2015 because it does not involve a financial contribution of the IMF. We use available values at end-October 2013, including in particular the European Commission autumn forecast.

Naturally, the results need to be treated with substantial caution as the findings of our panel regressions are very likely to miss specific factors conducive to the success or failure of a particular program. For instance, compared to the “average” country in the sample, the European programs (with the exception of Romania) faced a more challenging fiscal and debt situation and lack of exchange rate flexibility. On the other hand, they were outliers in terms of institutional quality, GDP per capita and financial depth, all of which *ex ante* could be expected to facilitate adjustment. Moreover, as mentioned above, the adjustment strategy in the euro area programs aimed to address other issues such as a need to prevent contagion and preserve financial stability, and our indicator cannot assess whether these objectives were met or not.

Our exercise relies to a large extent on projections, which for some variables especially exchange rate developments are surrounded by a high degree of uncertainty. At the same time, estimating the probability of success with past forecast data offers the possibility to compare the predictions of our model with outturns. For programs that started after 2012 we assume no change in the NEER throughout the program and beyond. This assumption seems justified as the NEER does not impact the probability of a success substantially.

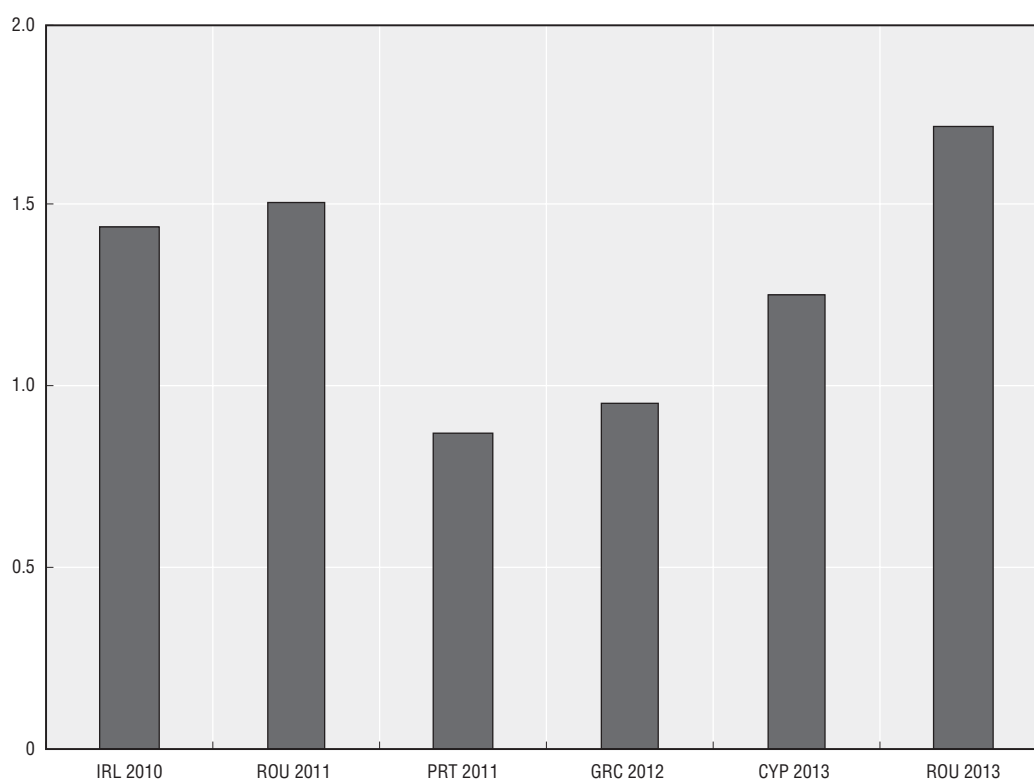
Before turning to the results of our ‘out-of-sample’ assessment, it may be worth stressing that our success indicator may be particularly stringent for countries such as Portugal and Greece which had accumulated a whole array of structural problems, as opposed to Ireland, where problems were instead concentrated in the financial sector. A recovery of growth and a reduction of debt is arguably more difficult to achieve for the former.

From the chart below, we can see that based on information available in autumn 2013 the European programs in all cases reached the average probability of success in the sample and exceeded it substantially for the majority of countries. Ireland benefitted to a large extent from its openness, Romania from its relatively comfortable fiscal situation at the onset of the program, and Cyprus both from its relatively open economy but also from

entering its program at a time of much more robust global growth and lower risk aversion compared the other euro area programs. The second program in Greece had a success probability almost exactly in line with the average in the sample, which must be considered a major achievement given its very challenging fiscal situation at the onset of the crisis and relatively low degree of economic openness. The fact that Greece eventually applied for a third program financed by the ESM, does not invalidate this relatively sanguine assessment. The renewed recourse to financial assistance was to a large extent determined by political factors. Portugal's success probability was marginally lower, because the country's growth performance was expected to turn out, and effectively did turn out, to be underwhelming. The country did not benefit from the improving external environment owing to its comparatively low degree of openness of its economy.

Taken together, these findings again underline the importance of taking both domestic policy achievements and the external environment into account when judging a country's adjustment process. Notably, if the latter is taken into consideration, then the progress under the European programs was on average substantially higher than sometimes argued.

Figure 5. Probability of successful economic adjustment, based on data available at end-October 2013



Note: Conditional probability over unconditional probability. Results obtained using the estimated coefficients in the baseline regression, including structural reforms. IRL – Ireland, ROU – Romania, PRT – Portugal, GRC – Greece, CYP – Cyprus.
Source: Authors' calculations.

5. Concluding remarks

We reviewed more than 170 adjustment episodes, identified by the need to approach the IMF for official assistance, with the aim of pinning down factors that help countries resume positive growth and reduce government debt levels. We found that decisive policy action,

especially faster fiscal adjustment, progress on financial sector repair and structural reforms, contribute significantly and positively to a successful adjustment episode. Regarding the importance of a functioning financial system, we note that a banking crisis per se need not be detrimental for successful adjustment i) as long as it allows for the continued extension of credit to the private sector and ii) because it strengthens political determination to take decisive actions.

We find that initial vulnerabilities in the form of high fiscal imbalances lower the chances of successful adjustment. The effect of exchange rate flexibility is not clear-cut. While flexible exchange rates are found to increase the likelihood of success we also find some evidence indicating that appreciation episodes are conducive to a successful adjustment, which contrasts with the often-made claim that exchange rate depreciation and an export-led recovery are prerequisites for success. Turning to external factors, more trade openness helps, and the probability of successful adjustment is considerably higher if global growth is favorable. Our results suggest that more official financing (which conversely is typically coupled with less fiscal adjustment) does not significantly contribute to success, while more stringent conditionality especially in the structural area appears to exert a significantly positive effect.

Our results have important implications for adjustment processes in the euro area. End 2013, global conditions were forecast to (and actually did) improve compared to the conditions present during the early days of the crisis, that is, when the programs for Greece, Portugal, Ireland and Cyprus were put in place. The very adverse conditions under which these countries undertook their initial adjustment therefore needs to be strongly acknowledged when assessing their progress in restoring growth and debt sustainability. It also means that while more support from external demand materialised for the programs starting later such as Cyprus, it is still imperative to continue fiscal consolidation, and financial sector and structural reforms. Finally, program countries will benefit more from improving external conditions if they are sufficiently open or manage to open up their economy.

Notes

1. The 2012-13 financial sector assistance programme for Spain had a narrower objective and is not considered as an official crisis program for the remainder of this paper. Financing arrangements which give access to official financing but are not monitored by means of *ex-post* conditionality, such as the IMF's Flexible Credit Line, are not included in the analysis as eligibility criteria for such instruments require lack of outright macroeconomic adjustment needs.
2. A complete list of programs included in the sample is provided in the Appendix. There were more than 200 GRA-supported programs put in place since 1993; our data sample shrank somewhat due to unavailability of data for some variables of interest for certain countries. The selection of GRA-supported programs means that low-income-economies with no access to international financial markets were excluded from the sample. Programs that started in 2011 or afterwards are excluded as we would have to partly rely on projected variables which could bias the results. Programs not completed until end-August 2013 were excluded on corresponding grounds.
3. [www.imf.org/external/pubs/ft/sd/index.asp?decision=14064-\(08/18\)](http://www.imf.org/external/pubs/ft/sd/index.asp?decision=14064-(08/18)).
4. We follow the same approach as in IMF (2012).
5. In approximately 40 per cent of the sample the General Government Debt to GDP ratio ranged above 50 per cent at program start and declined on average from above 85 per cent to below 75 per cent between T and T+5.
6. In addition, several successor states of the former Soviet Union or of the former Socialist Federal Republic of Yugoslavia did not even exist before the program was agreed and for some countries the real GDP series is far too volatile.

7. Moreover, a continuous measure would have clashed with the relatively large share of countries transitioning from planned to market economies during the 1990s, as they quite frequently experienced years of extremely high or low growth. Another minor problem, partly addressed already above, arises from the fact that for 13 programmes real GDP growth is available only for a slightly shorter period than five years before programme start. However, since all these countries suffered from negative average real GDP growth rates during the year, it seems plausible to allocate them to the lowest growth criterion bar. The sole exception is Bosnia and Herzegovina which recorded immense economic growth after the end of the civil war; therefore we will assume that average prior to the program was above 6 per cent, leading to a growth criterion of 4.5 per cent on average after program start.
8. In our sample there are 19 programs where a default or a debt restructuring occurred during the period $[T, T+3]$. In seven of those cases debt restructurings were part of program design. It is also worth noting that for all but a few of the remaining cases, the country also did not meet the growth criterion, and as a result only a handful of programs were considered failures on the base of a default alone.
9. The improvement of the primary balance actually serves as a proxy of fiscal adjustment as neither comparable estimates of the structural budget balance nor narrative estimates of discretionary fiscal policy are available for the large set of countries used. While potential output can easily be calculated with for instance an HP-filter, budgetary sensitivities are not readily available beyond OECD members. Assuming that government revenues have an overall elasticity equal to one, a possible guesstimate of the budgetary sensitivity is the government expenditure-to-GDP ratio which is available for most countries. However, this approach involves serious simplifications which may not be warranted, especially in cases where countries lose access to market funding.
10. The extent of the gain in competitiveness as a result of currency devaluation depends largely on the share of import goods necessary for the production of export goods and the added value in the export sector.
11. A possible criticism is that foreign currency denominated debt was mostly issued in USD and the nominal effective exchange rate is the exchange rate vis-à-vis a basket in which the weight of the USD might be small. Yet, under no arbitrage assumptions a depreciation of the home currency vis-à-vis other currencies would result in a depreciation versus the USD as well if keeping the relative price of the currencies of the trading partners constant versus the USD. Therefore on average the NEER should also reflect exchange rate movements versus the USD.
12. An improving current account balance also signals capital outflows, while an economic recovery in emerging markets usually goes along with capital inflows. Exchange rate effects on the interest bill might further obscure clear results.
13. Specifically, we use the change in public debt in the five years leading up to the crisis, while the deficit and the current account balance are measured as the values observed two years prior to the crisis episode.
14. The effect of binary variables needs to be interpreted with caution when assessing the variation by one standard deviation.
15. The number of non-disbursing adjustment programs is 50.
16. The IMF streamlined the number of quantitative performance criteria in 2002 leading to an overall reduction in the number of quantitative performance criteria. Therefore, the QPC series suffers from a structural break which may also result in lack of significance.
17. www.freetheworld.com/release.html.
18. Estimation results are available from the authors upon request.
19. An alternative (but more resource-consuming) option would have been to look at the conditions, especially currency, yields, coupons and maturity, at which a country could issue government bonds.
20. This does not turn out to be an issue as all programs started in 2011 or afterwards are excluded from the sample.
21. The nominal values were calculated from the credit advanced to private sector expressed as percentage share nominal GDP by multiplying with the nominal GDP.

References

- Abiad, A., G. Dell'Ariccia and B. Li (2011), "Creditless Recoveries", *IMF Working Paper* 11/58.
- Alesina, A. and R. Perotti (1995), "Fiscal Expansions and Fiscal Adjustments in OECD Countries", *NBER Working Paper* No. 5214 (Cambridge, Massachusetts, National Bureau of Economic Research).
- Alesina, A. and S. Ardagna (1998), "Tales of Fiscal Adjustment", *Economic Policy*, Vol. 13, No. 27, pp. 489-585.
- Atoyan, R. and P. Conway (2005), "Evaluating the Impact of IMF Programs: A Comparison of Matching and Instrumental-Variable Estimation", *The Review of International Organizations*, Vol. 1, No. 2, pp. 99-124.
- Barrios, S. and S. Langedijk (2010), "Fiscal consolidation with external imbalances", in *External Imbalances and Public Finances in the EU*, *European Economy Occasional Paper* 66.
- Barrios, S., S. Langedijk and L. Pench (2010), "EU fiscal consolidation after the financial crisis. Lessons from past experiences", *European Economy Economic Paper* 418.
- Barro, R. and J.W. Lee (2005), "IMF Programs: Who is Chosen and What are the Effects?", *Journal of Monetary Economics*, Vol. 52, No. 7, pp. 1245-69.
- Bas, M. and R. Stone (2011), "If Life Sends you Lemons: Adverse Selection and Growth under IMF Programs", www.rochester.edu/College/PSC/stone/working_papers/IMFGrowth3.pdf.
- Blanchard, Oliver and D. Leigh (2013), "Growth Forecast Errors and Fiscal Multipliers", *IMF Working Paper* 13/1.
- Calvo, G., A. Izquierdo and E. Talvi (2006), "Sudden Stops and Phoenix Miracles in Emerging Markets", *American Economic Review Papers and Proceedings*, Vol. 96, No. 2, pp. 405-10.
- Cerra, V. and S.C. Saxena (2008), "Growth Dynamics: the Myth of Economic Recovery", *American Economic Review*, Vol. 98, No.1, pp. 439-57.
- de Cos, P.H. and E. Moral-Benito (2014), "The Role of Public Wages in Fiscal Consolidation Processes", *Working Paper* presented at the ECFIN workshop on: Government wage bill: Determinants, interactions and effects.
- Darvas, Z. (2012), "Real effective exchange rates for 178 countries: a new database", *Bruegel Working Paper* 2012/06.
- Dreher, A. (2006), "IMF and Economic Growth: The Effects of Programs, Loans and Compliance with Conditionality", *World Development*, Vol. 34, No. 95, pp. 769-88.
- Ghosh, A., T. Lane, M. Schulze-Gattas, A. Bulir, J. Hamann, A. Mourmouras (2002), "IMF Supported Programs in Capital Account Crises", *IMF Occasional Paper* 210.
- Guichard, S., M. Kennedy, E. Wurzel and C. André (2007), "What Promotes Fiscal Consolidation: OECD Country Experiences", *OECD Economics Department Working Papers* No. 553.
- von Hagen, J., R. Perotti and R. Strauch (1998), "Sustainable Public Finances", London: CEPR.
- von Hagen J., A. Hughes-Hallett and R. Strauch (2001), "Budgetary Consolidation in EMU", *European Economy, Economic Papers* 148.
- von Hagen, J. and R. Strauch (2001), "Fiscal Consolidations: Quality, Economic Conditions, and Success", *Public Choice* Vol. 109, pp. 327-46.
- von Hagen J., A. Hughes-Hallett and R. Strauch (2002), "Budgetary Consolidation in Europe: Quality, Economic Conditions, and Persistence", *Journal of the Japanese and International Economies*, Vol. 16, pp. 512-535.
- Hardoy, I. (2003), "Effect of IMF Programmes on Growth: A Reappraisal Using the Method of Matching", Paper presented at the European Economic Association, Stockholm, August 20-24, 2003.
- Herndon, T., M. Ash and Pollin (2013), "Does public debt consistently stifle economic growth? A Critique of Reinhart and Rogoff", *Cambridge Journal of Economics*, Vol. 38, No. 2, pp. 257-279.
- Hutchison, M. (2004), "Selection Bias and Output Costs in IMF Programs", *EPRU Working Paper* 04-15, www.econ.ku.dk/epru/files/wp/wp-04-15.pdf.
- International Monetary Fund (2003), "Lessons from the Crisis in Argentina", prepared by the Policy Development and Review Department, in consultation with the other Departments.
- International Monetary Fund (2009), "From Recession to Recovery: How Soon and How Strong?", *World Economic Outlook*, Ch.3, pp. 97-132.

- International Monetary Fund (2012), "Background Paper 3: Outcomes of Fund-Supported Programs", in *2011 Review of Conditionality*.
- Ivanova, A., W. Mayer, A. Mourmouras and G. Anayiotos 2001, "What Determines the Success or Failure of Fund-Supported Programs?", *IMF Working Paper* Nov/01.
- Laeven, L. and F. Valencia (2012), "Systemic Banking Crisis Database: An Update", *IMF Working Paper* 12/63.
- Lamo, A., E. Moral-Benito and J. Perez (2014), "Austerity through public employment in bad times: exploiting the crowding-out and the competitiveness channels?", Working Paper presented at the ECFIN workshop on: Government wage bill: Determinants, interactions and effects.
- Larch, M. and A. Turrini (2011), "Received Wisdom and Beyond: Lessons from Fiscal Consolidation in the EU", *National Institute Economic Review*, Vol. 217, No. 1, pp. R1-R18.
- Liang, K.-Y. and S. Zeger (1986), "Longitudinal data analysis using generalized linear models", *Biometrika*, Vol. 73, No. 1, pp. 13-22.
- Mise, E., Kim, T.-H. and P. Newbold (2005), "On suboptimality of the Hodrick-Prescott Filter at time series endpoints", *Journal of Macroeconomics*, Vol. 27, 2005, pp. 53-67.
- Nkusu, M. (2011), "Nonperforming Loans and Macrofinancial Vulnerabilities in Advanced Economies", *IMF Working Paper* No. 11/161.
- Pisani-Ferry, J., A. Sapir and G. Wolff (2013), "EU-IMF assistance to euro area countries: an early assessment", *Bruegel Blue Print Series*, Vol. 19.
- Przeworski, A. and J.R. Vreeland (2000), "The effect of IMF programs on economic growth", *Journal of Development Economics*, Vol. 62, pp. 385-421.
- Reinhart, C. and K.S. Rogoff (2010), "Growth in a Time of Debt", *American Economic Review Papers and Proceedings*, Vol. 100, pp. 573-578.
- The World Bank (2002), "Regaining Fiscal Sustainability and Enhancing Effectiveness in Croatia".

Appendix

Data sources and definitions

1. **IMF GRA-supported adjustment programs start date and end date:** Source: IMF Monitoring of Fund Arrangements, IMF Annual Reports.
2. **Real GDP growth:** Annual percentage change. Source: IMF World Economic Outlook Database; for a handful observations the data set was complemented from IMF country reports.
3. **Public Debt:** Gross debt of general government or of the public sector broadest coverage available, in percent of GDP. Source: IMF World Economic Outlook Database, IMF Historical Public Debt Database Fall 2012 Vintage, IMF Monitoring of Fund Arrangements, IMF country reports, OECD Economic Outlook, World Bank, Eurostat, Republic of Croatia Ministry of Finance, Ministry of Finance Romania.
4. **Fiscal Balance:** of general government or the public sector broadest coverage available, in percent of GDP. The basic source is the IMF World Economic Outlook Database and missing data was complemented from other sources (see below). Source: IMF World Economic Outlook Database, IMF country reports, IMF (2003), OECD Economic Outlook, World Bank (2002), Peru Reserve Bank, Republic of Turkey Prime Ministry Undersecretariat of Treasury, AMECO, Eurostat.
5. **Primary Balance:** of general government or of the public sector broadest coverage available, in percent of GDP. The basic source is the IMF World Economic Outlook Database and missing data was complemented from other sources (see below). Source: IMF World Economic Outlook Database, IMF country reports, IMF (2003), OECD Economic Outlook, World Bank (2002), Peru Reserve Bank, Republic of Turkey Prime Ministry Undersecretariat of Treasury, AMECO, Eurostat, Bosnia and Herzegovina National Bank Annual Reports, Inter-American Development Bank, Central Bank of Indonesia.
6. **Cyclically Adjusted Balance:** Fiscal balance of general government or of the public sector broadest coverage available, adjusted for cyclical components, in percent of GDP. The output gap was estimated with the Hodrick-Prescott filter using $\lambda = 100$.
7. **Cyclically Adjusted Primary Balance:** Primary balance of general government or of the public sector broadest coverage available, adjusted for cyclical components, in percent of GDP. The cyclically adjusted primary balance was obtained by subtracting interest expenditure to the cyclically adjusted balance.
8. **Adjustment in Net Lending, Primary Balance, Cyclically Adjusted Balance, Cyclically Adjusted Primary Balance:** Difference in the value of the variable, respectively, between the value in the year of the program end and in the year of the program start. In case the program ended between the first and eighth month of the relevant year, the value from the year before the year of the program end is taken as the value for the year of the program end. For cases in which the adjustment program lasted only one year, that is the difference between the start year and end year of the program is one, no such adjustment is undertaken.

9. **Real world GDP growth:** Annual percentage change of the real world GDP growth, five year arithmetic average between T+1 and T+5. Source: IMF World Economic Outlook Database.
10. **Banking Crisis:** Dummy variable takes value 1 if a banking crisis was on-going in the year of the programme start, otherwise 0. For banking crises started in 2008 or afterwards no end date is provided in the source data base and these crises are marked as still on-going. Therefore, we assumed that a banking crisis has not ended yet if it started in 2008 or afterwards.²⁰ Source: Laeven and Valencia (2012).
11. **Credit Crunch:** Dummy variable takes value 1 if real credit advanced to private sector recorded negative growth in at least two years between T and T+2, otherwise 0. The annual percentage change in nominal credit advanced to private sector was corrected for inflationary effects by dividing through the relative change in the GDP deflator.²¹ Source: World Bank, IMF country reports, National Bank of Romania (credit advanced to private sector); IMF World Economic Outlook Database, IMF country reports (nominal GDP and GDP deflator).
12. **Exchange Rate Regimes:** Exchange Rate Regime Reinhart and Rogoff Classification, annual fine classification. Source: Carmen M. Reinhart Author Website, www.carmenreinhardt.com.
13. **Openness indicator:** Exports plus imports divided by GDP at current prices. Source: Penn World Table 7.1.
14. **VIX:** Chicago Board of Exchange S&P 500 Implied Volatility Index; annual, calculated as the average of the daily “last price” of the index. Source: Bloomberg.
15. **Nominal Effective Exchange Rate (NEER) change:** Percentage change in the nominal effective exchange rate: values in the year of program end and program start were compared. In case the program ended between the first and eighth month of the relevant year, the value from the year before the year of the program end is taken as the value for the year of the program end. For cases in which the adjustment program lasted only one year, that is the difference between the start year and end year of the program is one, no such adjustment is undertaken. Source: Darvas (2012).
16. **Current Account Balance:** in percent of GDP in USD. Source: IMF World Economic Outlook Database; for a handful observations the data set was complemented from IMF country reports.
17. **Program Conditionality:** Source: IMF Monitoring of Fund Arrangements.
18. **Program Size, Country Quota and Disbursement:** Source: IMF Monitoring of Fund Arrangements, IMF Annual Reports, IMF Financial Data Query Tool, IMF Lending Arrangements.
19. **Consumer Price Index:** Source: IMF World Economic Outlook.
20. **Real Effective Exchange Rate:** Source: Darvas (2012).
21. **Nominal Exchange Rate:** Source: IMF International Financial Statistics; complemented from IMF World Economic Outlook by calculating the ratio of GDP measured in USD and home currency.

Table A.1.1. **Sample of IMF assistance programmes**

Country name	ISO	Year
Algeria	DZA	1994, 1995
Angola	AGO	2009
Argentina	ARG	1992, 1996, 1998, 2000, 2003
Armenia	ARM	1995, 2009
Azerbaijan	AZE	1995, 1996
Belarus	BLR	1995, 2009
Bolivia	BOL	2003
Bosnia and Herzegovina	BIH	1998, 2002, 2009
Brazil	BRA	1998, 2001, 2002
Bulgaria	BGR	1994, 1996, 1997, 1998, 2002, 2004
Colombia	COL	1999, 2003, 2005
Costa Rica	CRI	1993, 1995, 2009
Croatia	HRV	1994, 1997, 2001, 2003, 2004
Czech Republic	CZE	1993
Dominica	DMA	2002
Dominican Republic	DOM	1993, 2003, 2005, 2009
Ecuador	ECU	1994, 2000, 2003
El Salvador	SLV	1993, 1995, 1997, 1998, 2009, 2010
Estonia	EST	1993, 1995, 1996, 1997, 2000
FYR Macedonia	MKD	1995, 2000, 2003, 2005
Gabon	GAB	1994, 1995, 2000, 2004, 2007
Georgia	GEO	1995, 2008
Greece	GRC	2010
Guatemala	GTM	2002, 2003, 2009
Honduras	HND	2008, 2010
Hungary	HUN	1993, 1996, 2008
Iceland	ISL	2008
Indonesia	IDN	1997, 1998, 2000
Jamaica	JAM	1992, 2010
Jordan	JOR	1994, 1996, 1999, 2002
Kazakhstan	KAZ	1994, 1995, 1996, 1999
Korea	KOR	1997
Latvia	LVA	1993, 1995, 1996, 1997, 1999, 2001, 2008
Lesotho	LSO	1994, 1995, 1996
Lithuania	LTU	1993, 1994, 2000, 2001
Maldives	MDV	2009
Mexico	MEX	1995, 1999
Moldova	MDA	1993, 1995, 1996, 2010
Mongolia	MNG	2009
Pakistan	PAK	1993, 1994, 1995, 1997, 2000, 2008
Panama	PAN	1995, 1997, 2000
Papua New Guinea	PNG	1995, 2000
Paraguay	PRY	2003, 2006
Peru	PER	1993, 1996, 1999, 2001, 2002, 2004, 2007
Philippines	PHL	1994, 1998
Poland	POL	1993, 1994
Republic of Congo	COG	1994
Romania	ROU	1994, 1997, 1999, 2001, 2004, 2009
Russia	RUS	1995, 1996, 1999
Serbia	SRB	2001, 2002, 2009
Seychelles	SYC	2008
Slovak Republic	SVK	1994
Sri Lanka	LKA	2001, 2009
Thailand	THA	1997

Table A.1.1. **Sample of IMF assistance programmes (cont.)**

Country name	ISO	Year
Turkey	TUR	1994, 1999, 2002, 2005
Ukraine	UKR	1995, 1996, 1997, 1998, 2004, 2008, 2010
Uruguay	URY	1996, 1997, 1999, 2000, 2002, 2005
Venezuela	VEN	1996
Vietnam	VNM	1993

Source: IMF Monitoring of Fund Arrangements, IMF Annual Reports.

Table A.2.1. **Extensions of baseline model; disbursement, program size and draw down under program**

	(1)	(2)	(3)	(4)
Primary Balance in T	0.112*** (0.036)	0.118*** (0.036)	0.119*** (0.037)	0.117*** (0.036)
Primary Balance Adjustment	0.126*** (0.042)	0.129*** (0.043)	0.130*** (0.043)	0.131*** (0.043)
Public Debt in T	-0.007* (0.004)	-0.007** (0.003)	-0.007** (0.004)	-0.009** (0.004)
Real GDP Growth in T	0.092*** (0.027)	0.096*** (0.027)	0.098*** (0.027)	0.099*** (0.027)
World GDP Growth [T+1,T+5]	2.165*** (0.422)	2.210*** (0.428)	2.229*** (0.429)	2.174*** (0.422)
Banking Crisis Dummy	1.245*** (0.358)	1.204*** (0.370)	1.195*** (0.367)	1.246*** (0.364)
Credit Crunch Dummy	-0.526** (0.259)	-0.552** (0.261)	-0.560** (0.263)	-0.536** (0.261)
Exchange Rate Regime	0.096*** (0.036)	0.097*** (0.035)	0.098*** (0.035)	0.087** (0.035)
Openness Indicator	0.008** (0.004)	0.008** (0.004)	0.008** (0.004)	0.008** (0.004)
CBOE S&P 500 Volatility Index	-0.043** (0.020)	-0.048** (0.021)	-0.050** (0.021)	-0.048** (0.020)
NEER adjustment	0.022*** (0.008)	0.023*** (0.008)	0.024*** (0.008)	0.021*** (0.008)
Disbursement	-0.123 (0.318)			
Programme Size to IMF Quota		0.019 (0.035)		
Amount Drawn to IMF Quota			0.042 (0.045)	
Amount Drawn to Programme Size				0.387 (0.355)
No. of Observations	176	176	176	176

Note: Standard errors in parentheses; *** p < 0.01, ** p < 0.05, * p < 0.1. For variable definitions see Table 2. T denotes the year of program start. Dependent variable: program success indicator (binary) using baseline definition from Section 2.

Table A.2.2. **Extensions of baseline model; instrumental variables approach; disbursement, program size and draw down under program**

	(1)	(2)	(3)	(4)
Primary Balance in T	0.268*** (0.041)	0.257*** (0.041)	0.255*** (0.040)	0.255*** (0.037)
Primary Balance adjustment	0.375*** (0.091)	0.364*** (0.090)	0.361*** (0.091)	0.365*** (0.091)
Public Debt in T	-0.008 (0.005)	-0.010** (0.004)	-0.010** (0.004)	-0.010** (0.005)
Real GDP Growth in T	0.092* (0.051)	0.098** (0.046)	0.099** (0.045)	0.097** (0.046)
World GDP Growth [T+1,T+5]	2.534*** (0.980)	2.513*** (0.887)	2.520*** (0.878)	2.514*** (0.892)
Banking Crisis Dummy	1.408*** (0.540)	1.331*** (0.503)	1.347*** (0.496)	1.363*** (0.500)
Credit Crunch Dummy	-0.946** (0.393)	-0.929** (0.383)	-0.925** (0.381)	-0.936** (0.390)
Exchange Rate Regime	0.118** (0.051)	0.110*** (0.042)	0.110*** (0.041)	0.109** (0.043)
Openness Indicator	0.006* (0.003)	0.007** (0.003)	0.007** (0.003)	0.007** (0.003)
CBOE S&P 500 Volatility Index	-0.014 (0.029)	-0.023 (0.032)	-0.023 (0.033)	-0.020 (0.031)
NEER adjustment	0.013 (0.012)	0.016 (0.010)	0.016 (0.010)	0.013 (0.012)
Disbursement	-0.439 (0.340)			
Programme Size to IMF Quota		0.017 (0.027)		
Amount Drawn to IMF Quota			0.024 (0.035)	
Amount Drawn to Programme Size				-0.055 (0.380)
No. of Observations	136	136	136	136

Note: Standard errors in parentheses; *** p < 0.01, ** p < 0.05, * p < 0.1. For variable definitions see Table 2. T denotes the year of program start. Dependent variable: program success indicator (binary) using baseline definition from Section 2.

Table A.3.1. **Extensions to baseline model; program conditionality**

	(1)	(2)	(3)
Primary Balance in T	0.109*** (0.037)	0.109*** (0.037)	0.108*** (0.038)
Primary Balance Adjustment	0.122*** (0.042)	0.121*** (0.043)	0.130*** (0.042)
Public Debt in T	-0.007** (0.004)	-0.007** (0.004)	-0.007** (0.004)
Real GDP Growth in T	0.095*** (0.027)	0.095*** (0.027)	0.094*** (0.027)
World GDP Growth [T+1,T+5]	2.220*** (0.436)	2.217*** (0.437)	2.301*** (0.467)
Banking Crisis Dummy	1.208*** (0.367)	1.206*** (0.366)	1.208*** (0.380)
Credit Crunch Dummy	-0.577** (0.265)	-0.571** (0.266)	-0.553** (0.270)
Exchange Rate Regime	0.094** (0.037)	0.095** (0.037)	0.109*** (0.038)
Openness Indicator	0.009** (0.004)	0.009** (0.004)	0.009** (0.004)
CBOE S&P 500 Volatility Index	-0.057*** (0.021)	-0.059** (0.023)	-0.062** (0.024)
NEER Adjustment	0.025*** (0.008)	0.024*** (0.008)	0.026*** (0.009)
All Conditions	0.014** (0.006)		
Structural Conditions		0.015** (0.006)	
Quantitative Performance Criteria		0.009 (0.031)	-0.009 (0.034)
Prior Action/Necessary for Compliance			0.023* (0.013)
Structural Performance Criteria			0.074* (0.041)
Structural Benchmarks			0.001 (0.011)
No. of Observations	176	176	176

Note: Standard errors in parentheses; *** p < 0.01, ** p < 0.05, * p < 0.1. For variable definitions see Table 2. T denotes the year of program start. Dependent variable: program success indicator (binary) using baseline definition from Section 2.

Table A.3.2. **Extensions to baseline model; instrumental variables approach; program conditionality**

	(1)	(2)	(3)
Primary Balance in T	0.242*** (0.040)	0.235*** (0.041)	0.237*** (0.053)
Primary Balance adjustment	0.377*** (0.075)	0.362*** (0.085)	0.351** (0.139)
Public Debt in T	-0.011*** (0.004)	-0.011** (0.004)	-0.010** (0.005)
Real GDP Growth in T	0.092* (0.050)	0.094* (0.048)	0.095* (0.051)
World GDP Growth [T+1,T+5]	2.370** (0.966)	2.433** (0.950)	2.846*** (0.911)
Banking Crisis Dummy	1.259*** (0.452)	1.273*** (0.446)	1.359*** (0.457)
Credit Crunch Dummy	-0.823** (0.361)	-0.817** (0.367)	-0.975** (0.395)
Exchange Rate Regime	0.117** (0.046)	0.123*** (0.046)	0.142*** (0.046)
Openness Indicator	0.009** (0.004)	0.009** (0.004)	0.009* (0.005)
CBOE S&P 500 Volatility Index	-0.032 (0.035)	-0.042 (0.039)	-0.045 (0.044)
NEER adjustment	0.018* (0.011)	0.017 (0.011)	0.016 (0.011)
All Conditions	0.011 (0.008)		
Structural Conditions		0.015 (0.010)	
Quantitative Performance Criteria		-0.012 (0.026)	-0.039 (0.036)
Prior Action/Necessary for Compliance			0.025 (0.018)
Structural Performance Criteria			0.077** (0.037)
Structural Benchmarks			-0.004 (0.017)
Observations	136	136	136

Note: Standard errors in parentheses; *** p < 0.01, ** p < 0.05, * p < 0.1. For variable definitions see Table 2. T denotes the year of program start. Dependent variable: program success indicator (binary) using baseline definition from Section 2.

Long-term growth and productivity projections in advanced countries

by

Gilbert Cette, Rémy Lecat and Carole Ly-Marin*

In this period of high uncertainty about future economic growth, we have developed a growth projection tool for 13 advanced countries and the euro area at the 2100 horizon. This high uncertainty is reflected in the debate on the possibility of a “secular stagnation”, fueled by the short-lived Information and Communication Technology (ICT) shock and the current low productivity and GDP growth in advanced countries. Our projection tool allows for the modelling of technology shocks, for different speeds of regulation and education convergence, with endogenous capital growth and TFP convergence processes.

We illustrate the benefits of this tool through four growth scenarios, crossing the cases of a new technology shock or secular stagnation with those of regulation and education convergence or of absence of reforms. Over the period 2015-2100, the secular stagnation scenario assumes yearly TFP growth of 0.6% in the US, leading to a 1.5% GDP growth trend. The technology shock scenario assumes that the third technological revolution will, in the US, provide similar TFP gains to electricity during the second industrial revolution, leading to a 1.4% TFP trend, to which we add a TFP growth wave peaking in 2040, and thus to an average GDP growth rate of 3%. In non-US countries, GDP growth will depend on the implementation of regulation reforms, the increase in education and on the distance to the country-specific convergence target, namely the US, as well. Over the period 2015-2060, for the euro area, Japan and the United Kingdom, benefits from regulation and education convergence would amount to a 0.1 to 0.4 pp yearly growth rate depending on the initial degree both of rigidity and the TFP distance to the US.

JEL classification: O11, O33, O43, O47, O57

Keywords: Growth, productivity, long-term projections, structural reforms, innovation, education

* Gilbert Cette (gilbert.cette@banque-france.fr), Rémy Lecat (remy.lecat@banque-france.fr) and Carole Ly-Marin (carole.ly-marin@banque-france.fr) are from the Banque de France and Gilbert Cette and Carole Ly-Marin also work at the Aix-Marseille School of Economics (AMSE). The views expressed in this paper do not necessarily reflect those of the Banque de France or the Eurosystem. The authors would like to thank Roxanne Tabouret for her precious research assistance.

1. Introduction

In developed countries, future long-term growth topics have received growing attention over the last few years in the context of the Great Recession and of lower average growth than at any time since WWII. The current situation thus logically raises the question: *are developed countries going to suffer from a long period of low growth or are they once again going to benefit from a wave of higher growth associated with a new technology revolution?* The prospect of a long period of low growth is alarming, not only because it would reduce purchasing power and improvements in living standards, but also because it would make it harder to face what Gordon (2012, 2013) names the “six headwinds”. These “six headwinds”, which are already in action and could potentially play a larger role in the future, are: i) a reversal of the demographic dividend; ii) a plateau in educational attainment; iii) rising income and wealth inequalities; iv) globalization; v) energy and environmental risks; and vi) the twin household and government deficits. These “six headwinds” could contribute to low growth and in turn, low growth would make them more damaging.

To use an expression coined by Hansen (1939), and applied to the current situation by Summers (2014, 2015) and Eichengreen (2015) among others, are we facing the risk of “secular stagnation”?¹ Such a “secular stagnation” appears to have both demand and supply side origins.

Concerning demand, it appears to be the consequence of a structural disequilibrium between saving and investment. A large saving glut could have numerous causes: for example, a growing share of emerging countries (such as China) with high saving rates in the world economy, a growing share of large companies with high saving rates in developed countries (such as the GAFA: Google, Amazon, Facebook and Apple), the increase of income inequalities as the saving rate increases with income, etc. Other causes may be, among others, declining investment opportunities or decreasing investment prices. From this disequilibrium, the natural real interest rate could turn negative and in the context of low inflation and of the zero lower bound limit to interest rates, monetary policy would not be able to lower the saving rate. With regard to the supply side, productivity has slowed down in all advanced countries since the beginning of the 2000s, before the current crisis.² This fact is now well documented by several recent studies (see for example for the US, Gordon, 2012, 2013, 2014, 2015, or Byrne, Oliner and Sichel, 2013, and for all advanced countries, Crafts and O'Rourke, 2013, or Bergeaud, Cetté and Lecat, 2016a). These countries already suffered from productivity slowdowns at the end of the 1960s, during the 1970s, towards the end of the 1980s, and during the 1990s, except for the US where, from the mid-1990s, we observe an acceleration due to faster improvement in information and communication technologies (ICT hereafter) productive performances (Jorgenson, 2001, was the first of numerous papers to stress this last point). Productivity growth is now very low, even close to nil in some countries, and, for some authors such as Gordon (2012, 2013, 2014, 2015), this situation could be the future of long-term productivity.

This study deals with the topic of future long-term growth, and is associated with a user-friendly software tool that allows the construction of different scenarios.³ Previous studies have suggested future long-term scenarios associated with different types of issues. Duval and Maisonneuve (2009) proposed a long-term scenario at the 2060 horizon for the world divided into ten areas. The main questions treated in this study were the crucial role of education and the growth impact of reforms in non-US areas. Fouré, Benassy-Quéré and Fontagné (2013) developed scenarios for 147 countries at the 2050 horizon. Key issues included the role of energy, female participation and the international circulation of savings financing capital growth.

The originality of our study is twofold. First, it is a supply side approach that attempts to infer productivity growth from technological innovations in the US, to which other countries converge as it is considered to be the technological frontier; education and regulation being the main drivers of this productivity catch-up process. Second, as mentioned previously, the analysis is supplemented by a user-friendly software that enables the construction of different scenarios related to productivity growth from technology shocks at the frontier (the US) and the catch-up process of other countries from education and regulation hypotheses. As it is a supply side approach that gives a large role to technology waves, the horizon is very long and scenarios can be built with an annual step from the current period to 2100, the end of the century. Mainly using the Bergeaud, Cetté and Lecat (2016a) database,⁴ we consider 13 developed countries and the euro area, which together represent more than half of current world GDP. Specific scenarios can be developed for each of them.⁵

Concerning possible settings, users operating the study's associated software can build diverse productivity scenarios for the US, the technological frontier. Given exogenous employment growth and an endogenization of capital growth, US GDP growth is determined by the total factor productivity (TFP) scenario. For other countries, productivity scenarios depend on the US's, but also on the country's specific catch-up process, itself subject to hypotheses concerning education and regulation. As for the US, with the same exogenous employment growth and endogenization of capital growth, non-US country GDP growth scenarios are associated with these productivity ones.

The remainder of the paper is organized as follows: Section 2 presents the model and Section 3 the data. Sections 4 and 5 detail the endogenization of capital and the productivity catch-up process respectively. Section 6 comments some scenarios built with the software associated with the model. Section 7 concludes.

2. The model

The model is a supply one based, at the national level, on the following usual two-factor (capital and labor) Cobb-Douglas production function, with constant returns to scale:

$$Q_{i,t} = TFP_{i,t} \cdot K_{i,t-1}^{\alpha} \cdot (N_{i,t} \cdot H_{i,t})^{1-\alpha} \quad (1)$$

where i and t variable indexes indicating for which country i and which year t the variable is considered. Q is the volume of GDP, TFP the total factor productivity, K the volume of capital installed at the end of the year, N employment, i.e. the number of workers, and H the average number of hours worked per year and per worker. α is the elasticity of output Q to capital K and, as in other studies (see for example Bergeaud et al., 2016a), we assume constant elasticity over time for all countries and the calibration: $\alpha = 0.3$.

Total factor productivity takes into account the contribution of education to the quality of labor input through the average years of schooling in the working age population. This

contribution is calibrated with a 5% return on years of schooling, estimated in Bergeaud, Clette and Lecat (2016b) on the same database. This return lies within the range of estimates of “Macro-Mincer” equations such as Soto (2002), Cohen and Soto (2007) and Barro and Lee (2010).

This relation (1) can be written in growth rate terms:

$$(1') \quad \Delta q_{i,t} = \Delta \text{tfp}_{i,t} + \alpha \cdot \Delta k_{i,t-1} + (1 - \alpha) \cdot (\Delta n_{i,t} + \Delta h_{i,t})$$

or

$$(1'') \quad \Delta q_{i,t} = \Delta \text{tfp}_{i,t} + \alpha \cdot (\Delta k_{i,t-1} - \Delta n_{i,t} - \Delta h_{i,t}) + (\Delta n_{i,t} + \Delta h_{i,t})$$

where x corresponds to the logarithm of the variable X ($x = \log(X)$), and Δx is the usual approximation for the growth rate of X . ($k_{i,t-1} - n_{i,t} - h_{i,t}$) is called capital intensity.

To build a future long-term scenario, for each country i , employment N and working hours H are exogenous. The quantification of volume of capital K and of the TFP comes from specific assumptions and relations.

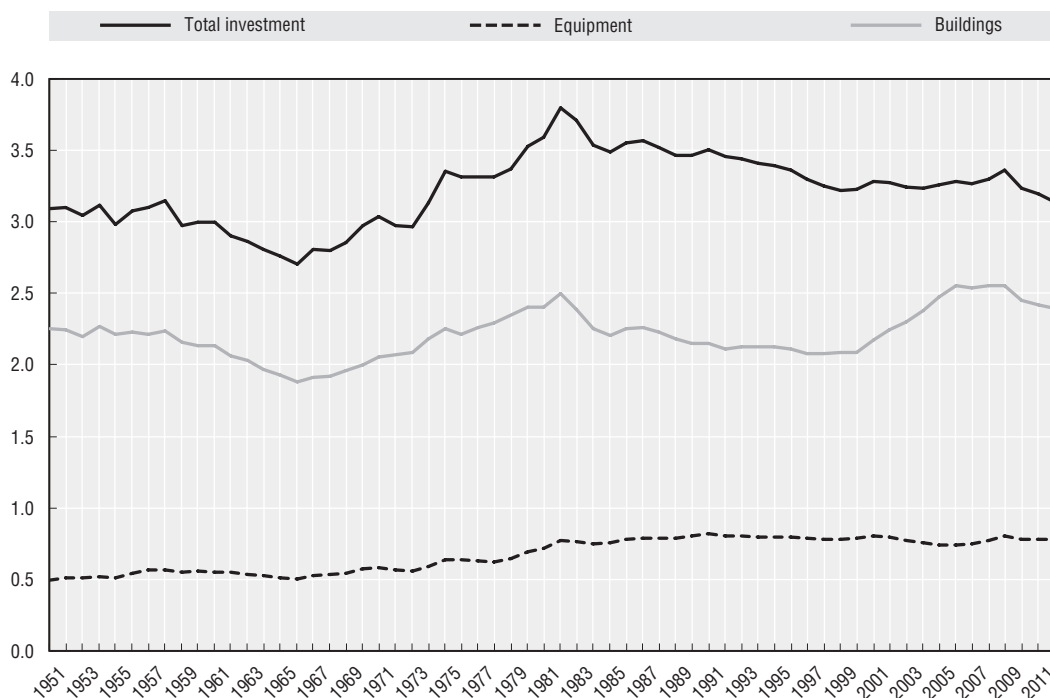
Concerning capital, we assume that in the long term, at the potential path, the capital coefficient (ratio of capital divided by GDP) remains constant in nominal terms (Clette, Kocoglu and Mairesse, 2005):

$$\Delta p q_{i,t} + \Delta q_{i,t} = \Delta p k_{i,t} + \Delta k_{i,t-1} \quad (2)$$

where P_Q is the GDP price ($p q = \log(P_Q)$) and P_K the investment in the fixed productive capital price ($p k = \log(P_K)$).

As in Clette, Kocoglu and Mairesse (2005), we observe in nominal terms over the last few decades in the US notable stability in the capital coefficient (Chart 1). The stability assumption thus seems reasonable.

Figure 1. **Capital coefficient, at current prices, in the US**
(Ratio of capital stock to GDP in current prices)



From relation (2), we obtain the relation (2)', which is used to build long-term capital scenarios:

$$(2') \quad \Delta k_{i,t-1} = \Delta p q_{i,t} + \Delta q_{i,t} - \Delta p k_{i,t}$$

To build long-term capital evolution scenarios using relation (2'), we need a corresponding scenario of the relative investment price ($p_k - p_q$) evolution. We assume that this relative price evolution depends, through an error correction model (ECM), on TFP evolution. The underlying theory is that, in the US, quality improvements in investment in terms of productive performance are at least partly incorporated into the measurement of investment prices in national accounts through hedonic or matching methods. This is mainly done for ICT since this investment benefits more than others from performance improvements (for a summary on these aspects, see Byrne, Oliner and Sichel, 2013, and Byrne, Fernald and Reinsdorf, 2016). So, the investment performance gains impact both from investment prices and TFP. Taking all this into consideration, the relation estimated for the US is the following:

$$\Delta p_{US,t} = a_0 + a_1 \cdot \Delta p_{US,t-1} + a_2 \cdot \Delta t f p_{US,t} + a_3 \cdot \Delta t f p_{US,t-1} + a_4 \cdot (p_{US,t-1} - a_5 \cdot t f p_{US,t-1}) + u_t$$

where $p_{US,t}$ corresponds to the relative investment price in the US (in logarithm):

$$(4) \quad p_{US,t} = p k_{US,t} - p q_{US,t}.$$

Increases in $t f p_{US,t}$ lower the relative investment price $p_{US,t}$ as the quality-adjusted price of investment decreases.

Relation (3) is estimated for the US. To build long-term relative investment price scenarios, we first build a long-term TFP scenario for the US and then use the estimated parameters of relation (3) to build the US relative price scenario. For the other countries i , we assume (as proposed by Schreyer, 2000, and numerous papers thereafter) that the relative price is the same as in the US:

$$(4') \quad p_{i,t} = p_{US,t}$$

The long-term TFP scenarios are exogenous for the US. For the other countries i , we assume a catch-up process represented through an error correction model (ECM) in which intervene the level of education of the working age population,⁶ regulation of the labor and product markets and, in the short run, changes in the capacity utilization rate, in the employment rate and in average working time. The following catch-up relation (5) is estimated:

$$\begin{aligned} \Delta t f p_{i,t} = & \alpha_1 \cdot [t f p_{US,t-1} - t f p_{i,t-1}] + \alpha_1 \cdot \alpha_2 \cdot R_{i,t-1} \cdot [t f p_{US,t-1} - t f p_{i,t-1}] + \alpha_3 \cdot \Delta C U R_{i,t} + \alpha_4 \cdot \Delta h_{i,t} \\ & + \alpha_5 \cdot \Delta E R_{i,t} + \gamma_i + \delta_t + u_{i,t} \end{aligned}$$

where R is the level of anticompetitive regulation (product market x labor market regulation), CUR the capacity utilization rate, ER the employment rate (here employment over population), and the γ_i and δ_t country and year fixed effects.

In relation (5), the long-term TFP catch-up process to the US level depends on anticompetitive regulation on product and labor markets (more regulation lower the catch-up process, the expected sign of α_2 being negative). In the short run, TFP growth also depends on changes in the capacity utilization rate of production factors (with a positive impact, the expected sign of α_3 being positive) and on changes in the employment rate and hours (with a negative impact, the expected sign of α_4 and α_5 being negative). These three short term impacts are estimated in other studies (see Bourlès and Cetté, 2007, or Aghion et al., 2009, for surveys and estimates). The employment rate and working hours have decreasing

returns as less productive workers are recruited as the employment rate increases and because of tiredness effects with regard to working hours. Country and time dummies capture all the unobserved factors that may impact the convergence target: regulation that is not measured in our indicators (such as tariffs or banking regulations), the quality of management, corruption, etc. Year fixed effects capture the cycle common to all countries in our sample but also common movement in convergence targets across all these countries. It may appear, for example, that the ICT technology shock impacts the TFP level in the United States because of the share of ICT-producing industries in the US economy, which is lower in most countries of our sample, leading to a common downward shock in the convergence target for all non-US countries. We assume that the US remains the technology leader over the whole period and that the convergence process continues as during our estimation period (1994-2010), which corresponds to the first ICT wave.

Relation (5) is estimated over our country panel. Non-US country TFP scenarios can be built using its estimated parameters and exogenous scenarios concerning US TFP.

3. Data

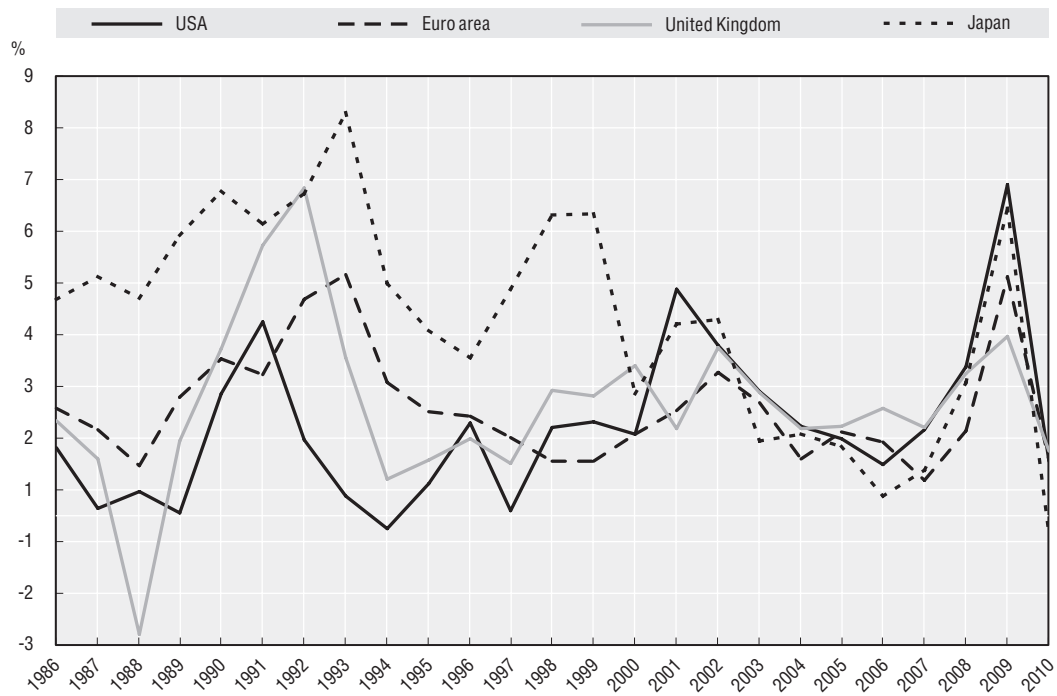
These models require the use of national accounts data (GDP, equipment and building investment, price of output and of investment in fixed productive capital, employment), of degrees of production factor utilization (capital utilization rates, hours worked per employee), population, education and regulation data.

For national accounts data (GDP, investment, employment, working hours, GDP and investment prices) and for education, population and production factor utilization data, we essentially use the dataset from Bergeaud et al. (2015, 2016a and 2016b) which gathers data for 13 OECD countries and the euro area over the period 1890-2013. These countries have been chosen for the large share of world GDP they represent: the G7 (the United States, Japan, Germany, France, the United Kingdom, Italy and Canada), four other euro area countries (Spain, the Netherlands, Belgium and Finland) and two other countries (Australia and Sweden). In addition, a euro area has been reconstituted, aggregating Germany, France, Italy, Spain, the Netherlands, Belgium, Portugal and Finland (93.2% of the euro area GDP in 2010). A detailed description of the construction of this dataset is given in Bergeaud et al. (2016a and 2016b).

The capital indicator is constructed using the perpetual inventory method (PIM) applied to each of the two components (equipment K_E and buildings K_B) with the corresponding investment data (I_E and I_B). The yearly depreciation rates used to build the capital series using the PIM are 10.0% for equipment and 2.5% for buildings following Cette et al. (2009) and are assumed to be constant across time and space. Finally, the damage that occurred during the World Wars, earthquakes in Japan and the civil war in Spain are, as much as available information will allow, taken into account in building the capital series. Chart 2 presents the yearly growth rate of capital intensity (the total capital stock over hours worked) over the estimation period; this displays a counter-cyclical pattern, as the capital stock tends to be more inert than hours worked.

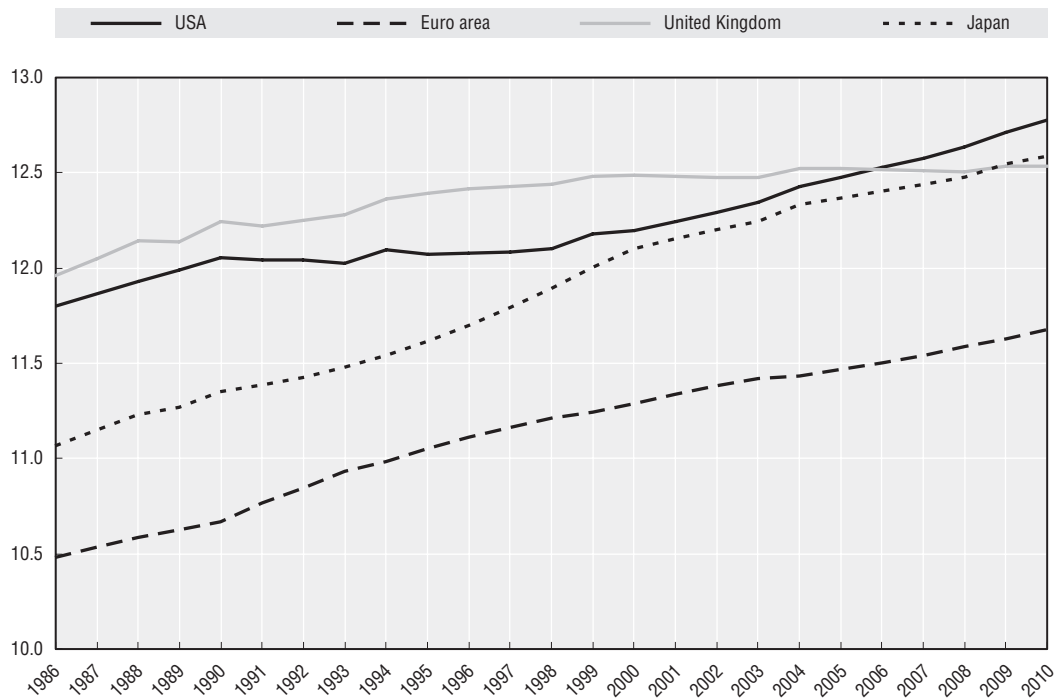
For education, we also use data collected by Bergeaud et al. (2016b). Educational attainment is defined by the average time spent studying in the total population over the age of 15 or 25⁷ using Van Leeuwen and Van Leeuwen-Li (2014). As shown in Chart 3, educational attainment rose continuously over the estimation period, with particularly large gains for the euro area and Japan.

Figure 2. **Capital intensity yearly growth rate (%)**



Source: Bergeaud et al. (2016b).

Figure 3. **Average years of schooling in the population aged over 15 (in years)**



Source: Van Leeuwen and Van Leeuwen-Li (2014).

For population and employment rate projections up to 2060, we use those of the OECD (2015) and allow for parameterization afterwards. We therefore assume exogeneity of both these parameters, which may be endogenous to growth itself or regulation. Age structure could be a major determinant of long-term employment rate projections but is not concerned in cyclical developments. Lastly, the impact of regulation should be captured through our convergence equation.

We use OECD regulation indicators for employment protection legislation (EPL – OECD, 2013) and product market regulation (PMR – Koske et al., 2015). For employment protection legislation, we use permanent employment legislation as it is most representative of the stock of employment contracts. For product market regulation, we use the composite indicator encompassing state control, barriers to entrepreneurship and barriers to trade and investment. Both these indicators range between 0 (least regulated) and 6 (most regulated). Our regulatory indicator is the product of this EPL and PMR, assuming a non-linear relationship and a possible offsetting of one regulation by the other.⁸

4. Capital stock dynamics

Capital stock dynamics are assumed to depend on the stability of the capital stock to GDP ratio in nominal terms (see Section 2). Hence, to project the capital stock, we need to determine the trajectory of the relative investment price. As seen in equation (3), relative investment prices can be related to TFP through a dynamic error-correction model framework (ECM), as innovation tends to lead to decreasing investment prices, especially when quality adjustment is taken into account, compared to output prices.⁹ It is estimated on US data, not on a country-to-country basis, as US investment prices incorporate quality-adjusted information more efficiently than other countries, especially for recent periods.

Estimates are presented over different periods (from 1930 to 1990) in Table 1. We use the one-step estimate of the ECM, with the coefficients in first difference representing the short-term relationship. Changes in relative investment price are fairly inert, as shown by the large and significant coefficient of $\Delta \ln(p_{t-1})$ over all the periods. The short-term relationship with TFP is complex as the coefficient and the lagged coefficient are of opposite signs. In our preferred estimates (1970-2013), the overall coefficient of TFP is negative but small. The long-term relationship corresponds to the level variables ($\ln(p_{t-1})$ and $\ln(tfp_{t-1})$), the coefficient for $\ln(p_{t-1})$ being the error-correction term. It is negative in all of the columns, but is only significant for the 1970-2013 period. The coefficient of $\ln(tfp_{t-1})$ is, from equation (3), the product of α_4 , the error-correction term, and α_5 , the long-term relationship between the relative investment price and total factor productivity. Hence, α_5 is significant and equals 0.73, which implies that technological progress, as reflected by a 1% increase in TFP, will lead to a long-term decrease of 0.73 percentage point in relative investment prices. This estimate is consistent with Fisher's (2006) reaction function for investment-specific technology shocks. 1970-2013 is our preferred specification as the period is short enough to have a stable relationship among the variables¹⁰ and recent enough to have quality-adjusted data, but also because the period is long enough to estimate the coefficients precisely. Residual stability tests and cointegration tests are also more satisfactory on that period.¹¹ Overall, the sign of the coefficients does not vary considerably from one estimate to the other but their significance and magnitude can display large variations. The results are robust to the exclusion of the crisis period from 2008 onwards.

Table 1. **Relative investment price estimates (equation (3)) on US data**

Estimations for 2013 from... Dependent variable	1930 $\Delta \ln(p_t)$	1950 $\Delta \ln(p_t)$	1970 $\Delta \ln(p_t)$	1990 $\Delta \ln(p_t)$
$\Delta \ln(p_{t-1})$	0.463**** (3.62)	0.581**** (5.57)	0.697**** (5.22)	0.737**** (6.28)
$\Delta \ln(tf p_t)$	0.0641 (0.61)	-0.149 (-1.41)	-0.268** (-2.09)	-0.267 (-1.69)
$\Delta \ln(tf p_{t-1})$	-0.0753 (-0.53)	0.206 (1.66)	0.209 (1.13)	0.410*** (2.89)
$\ln(p_{t-1})$	-0.0166 (-0.84)	-0.0211 (-0.81)	-0.110** (-2.15)	-0.151 (-1.52)
$\ln(tf p_{t-1})$	-0.00341 (-0.34)	-0.00873 (-0.57)	-0.0803** (-2.12)	-0.0804 (-1.12)
N	84	64	44	24
adj. R^2	0.172	0.402	0.546	0.680

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

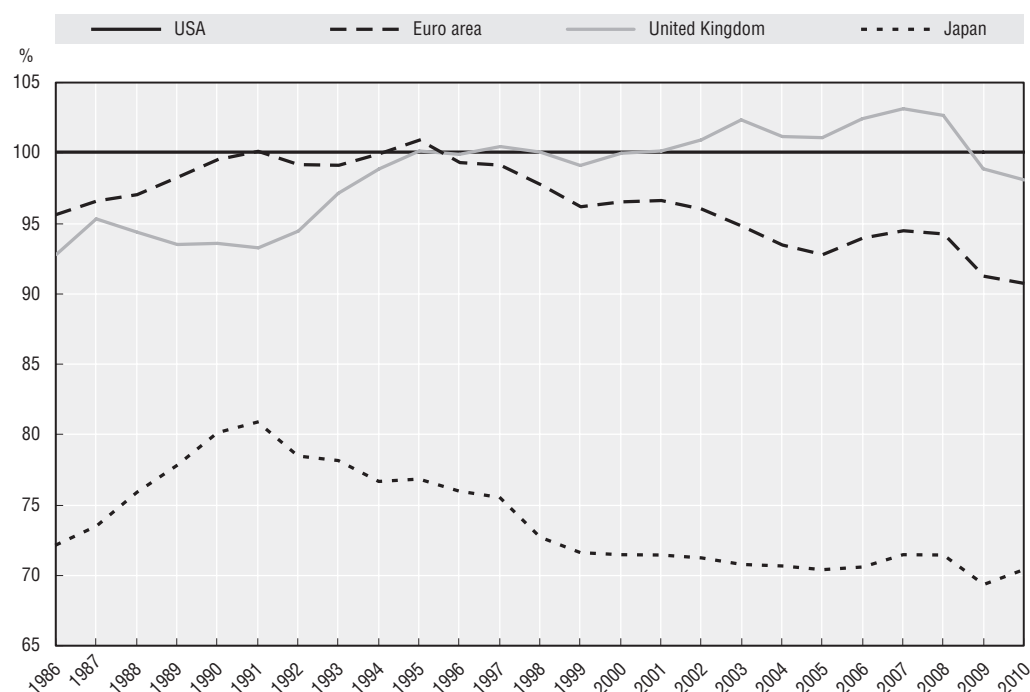
Note: One-step error-correction model estimated on US data. First, we test the variables' integration order with two different unit root tests: the Augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test. As both time series are integrated of order 1, we model a cointegrated process and test its fitting results with Engle's Lagrange multiplier test, detecting for the presence of autoregressive conditional heteroskedasticity, and the Breusch-Godfrey and Durbin's alternative tests, both assessing for serial correlation in the disturbance and find that they are reliable from 1950 onwards (all tests available on request).

5. Catch-up dynamics

Although the TFP dynamics for the United States are exogenously determined by the technology scenario, the TFP dynamics for other countries are determined by a convergence process towards the technological frontier, defined as the US level. We thus estimate convergence equation (5), whose results are reported in Table 2. In Chart 4, we see that, for the euro area, Japan and the UK, catching up with the US level took place until the mid-1990s, when the ICT shock led to a relative acceleration of US TFP. The euro area and the UK reached the US level, while others, such as Japan, remain far from it, which illustrates that countries may converge to different target TFP levels.

Our baseline specification includes convergence to the frontier, interaction of regulation and the frontier, all control variables, country fixed effects and year fixed effects from 1994 onwards. Other specifications include different periods, the regulation gap with the US rather than the level of regulation, different time periods, different returns on education and estimates without control variables.

Convergence to the frontier appears fast, especially in the period of the ICT technology shock that benefited the US more than most of the other countries in our panel: ignoring the interaction between the distance to the technological frontier and the level of regulation, about 14% of the gap is closed in one year for most specifications. However, this speed of convergence is reduced by the impact of labor and product regulation, whose average level is 7.35 over the sample. The coefficient of the interaction of regulation and the frontier is negative and significant, as expected. The speed of convergence is slower for countries with stringent labor and product regulations. Hence, the average speed of convergence is halved at the average regulation level, at 7% of the TFP gap on average.¹² Moreover, countries converge to different levels relative to the US, as estimates include country fixed effects and year fixed effects. Another specification choice could have been to introduce the level of regulation directly into the estimates so that the level as well as the speed of convergence depends on regulation. However, this is fully captured by our fixed effects, which encompass all other

Figure 4. **Total factor productivity as a % of the US level**

Source: Bergeaud et al. (2016b).

Table 2. **Catch-up estimates (equation (5))**

	(1) Baseline	(2) Regulations gap	(3) 1986-2010	(4) 1990-2010	(5) Return on education gap = 7% $\Delta \ln(tfp_t)$	(6) Without control variables $\Delta \ln(tfp_t)$
Dependent variable	$\Delta \ln(tfp_t)$	$\Delta \ln(tfp_t)$	$\Delta \ln(tfp_t)$	$\Delta \ln(tfp_t)$	$\Delta \ln(tfp_t)$	$\Delta \ln(tfp_t)$
Frontier	0.138**** (3.87)	0.139**** (4.00)	0.0839*** (3.31)	0.102**** (3.46)	0.140**** (3.97)	0.167**** (4.32)
Regulation · Frontier	-0.00920** (-2.54)		-0.00453** (2.58)	-0.00483** (-2.02)	-0.00921*** (-2.91)	-0.0111**** (-2.99)
Regulation gap · Frontier		-0.00980*** (-2.65)				
ΔCUR	0.245**** (4.16)	0.246**** (4.18)	0.340**** (7.94)	0.314**** (7.25)	0.244**** (4.05)	
$\Delta \ln(\text{hours})$	-0.237** (-2.40)	-0.238** (-2.41)	-0.290*** (-3.19)	-0.273*** (-3.00)	-0.244** (-2.48)	
$\Delta \text{Employment rate}$	-0.255** (-2.02)	-0.257** (-2.06)	-0.0399 (-0.31)	0.0423 (0.30)	-0.265** (-2.07)	
Constant	0.0129*** (3.30)	0.0130*** (3.31)	0.0128**** (3.47)	0.00967 (1.51)	0.0114*** (3.02)	0.0167**** (3.48)
Observations	192	192	288	252	192	192
Adjusted R^2	0.710	0.711	0.579	0.594	0.712	0.632

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: 1994-2010 estimates apart from columns (3) and (4). Estimates include country and year fixed effects. Robust t-stats are reported.

unobservable factors not captured in our indicators (for example banking regulation). The use of the regulation gap with the US level does not alter our estimates of the impact of regulation, but it is not our preferred specification, as part of the impact of changes in US regulation should be captured in the US TFP level, which is included in the frontier variable.

The control variables yield the expected results. Capacity utilization rates capture the country-specific cycle, which is not encompassed in the year dummies which capture the whole country sample cycle. Due to the imperfect measurement of factor utilization, which biases our TFP measure (Cette et al., 2015), this control is useful to correct for all cyclical effects. It is positive and significant in all specifications and its coefficient is fairly stable. The coefficients on the variations of hours worked per employee and of employment rates are negative, due to fatigue effects of longer hours and the recruitment of less productive workers as employment rates increase. The coefficients are negative, significant and stable across specifications, apart from the employment rate, which is not significant over 1986-2010 and 1990-2010. The magnitude of the coefficients is smaller than in Bourlès and Cette (2007), but their estimates are on productivity per employee and not TFP.¹³

Our preferred estimate is the 1994-2010 one (column (1)). It yields the highest R^2 compared to other time periods, encompasses all control variables, uses a return on education more in line with literature and, as previously explained, the regulation gap leads to complex interpretations. Regulation measurement methodologies are also more robust and stabilized from the mid-1990s.

We may be concerned by a potential reverse causality between changes in TFP and regulation: in times of crisis, structural reforms may be easier. However, the literature on the political economy of structural reforms is not unanimous on this point; second, this reverse causality would bias our coefficient upward, which remains negative and significant; finally, regulations are lagged.

6. Scenarios

The outlook for productivity growth over the next few decades is subject to large uncertainties (6.1.). For this reason, very contrasting scenarios are built with our software (6.2.).

6.1. Large uncertainties

The future of productivity growth over the next few decades could differ at the technological frontier (the United States) compared to other developed countries.¹⁴ It depends dramatically on improvements in ICT -or other technologies- performance and on diffusion. For this reason, the future of productivity growth is highly uncertain, and optimistic or pessimistic scenarios are both realistic. Gordon (2012, 2013, 2014, 2015), for example, is very skeptical about the extent of the current technology wave and argues that these contributions from ICT will be very small. According to him, labor productivity growth could stay on average close to 1¼% per year, which corresponds to what was observed over the sub-period 1975-95, from the first oil shock to the ICT productivity growth wave. Byrne, Oliner and Sichel (2013) propose two steady-state scenarios. In the first lower-bound one, the improvement in ICT performance, measured by ICT relative price growth, could remain on the same path as that observed on average during the 2004-12 sub-period. It translates into a slowdown in ICT performance improvement compared with the 1995-2004 sub-period but also, in a slightly less pronounced way, compared with the 1974-95 sub-period. But it also means an acceleration in comparison with what we observe in more recent years. In the second upper-bound scenario, the improvement in ICT performance could follow an intermediate path relative to that observed in the 1974-95 long sub-period and that observed in the favorable 1995-2004 sub-period. The contribution of ICT to future productivity growth (both from ICT capital deepening and from TFP in ICT producing sectors) differs between these two scenarios. It ranges from 1.8% to 2.5% for

average annual labor productivity growth in the non-farm business sector. We could imagine more pessimistic scenarios than the lower-bound one (like what Gordon proposes) but also more optimistic scenarios than the upper-bound one.

A more optimistic scenario than the Byrne et al. (2013) upper-bound one could yield different results. The following ones are those primarily discussed in the literature: the first one stems from the fact that significant improvements could happen in the near future in the semiconductor industry (see ITRS, 2013a, 2013b). The next operational one could be the 3D chip; it will allow fast improvements in ICT performance for many years. This development would increase the future contribution of ICT to productivity growth, which would engender a second ICT productivity growth wave benefiting the US. Secondly, large productivity gains could be generated by the extended use of available chips and ICT capacities in several new areas. This development was identified in 2005 by the ITRS as the “*More than Moore*” process. Pratt (2015) gives some emphasis to the potential improvements in robotics and Mokyr et al. (2015) to the gains in research efficiency. It also justifies Brynjolfsson and McAfee’s (2014) optimistic approach, which nonetheless stresses the need for appropriate institutions to reap the benefits from these potential gains. Rapid progress or even revolutions could take place in other domains, such as biotech, energy, agriculture... and fuel a “third technological revolution” scenario.

In the future, non-US advanced countries’ productivity growth could also be positively influenced by a possible catch-up with the higher ICT diffusion level observed in the US. Numerous studies have suggested reasons for the gap in ICT diffusion in non-US advanced countries compared to the leading country, including the level of post-secondary education among the working age population as well as labor and product market rigidities. For example, efficient use of ICT requires a higher degree of skilled labor than the use of other technologies. The required firm reorganization for effective ICT adoption can be constrained by strict labor market regulations. Moreover, low levels of competitive pressure, resulting from product market regulation, can reduce the incentive to efficiently exploit production techniques. Numerous empirical analyses have confirmed the importance of these factors.¹⁵ Among others, Cetto and Lopez (2012) show, through an econometric approach, that the US enjoys the highest level of ICT diffusion because of a higher level of post-secondary education among the working age population and less restrictive product and labor market regulations. This means that the future of productivity will depend, in all advanced countries, on technological progress, on the “*More than Moore process*”, but also on institutional changes through structural reforms to reap the full benefits from these advances. Concerning non-US countries, these institutional changes could potentially play a large role. Indeed, they would help to speed up the convergence to the US level of ICT productive capital diffusion.

6.2. Two types of growth scenarios

Two very different types of growth scenarios are proposed with the software (see Box for projection tool user manual) as to illustrate how much possible future outlooks differ. In the first one, named “Secular Stagnation”, US TFP growth stays indefinitely at the low level observed both before (1974-1990) and after (2005-14) the ICT productivity growth wave associated with the ICT third industrial revolution. In the second, called “technology shock”, the US enjoys another large productivity growth wave associated with a third industrial revolution based on ICT or other technologies. The underlying TFP growth is then assumed to go back to the level observed during the 1990-2005 sub-period. We add to

this trend growth a technology shock over four decades equivalent to the TFP contribution of electricity during the second industrial revolution in the 20th century. For each of these two US growth scenarios, two sub-scenarios have been built concerning non-US countries. In the first one, education and regulation stay at their level at the beginning of the period, and TFP catch-up does not accelerate over the period. In the second, education and regulations converge to US levels over the 2015-30 sub-period, which allows an acceleration of TFP convergence to the US level. For the UK, we did not include a specific treatment with regard to Brexit, as this shock is hard to calibrate, but it creates an additional downward risk on all the scenarios presented above.

“Secular stagnation” scenarios

In these scenarios, US TFP growth is stabilized at 0.6% per year, which corresponds to the level observed both before (1974-1990) and after (2005-14) the ICT productivity growth wave associated with the ICT third industrial revolution, meaning that the US does not enjoy a new technology shock. We use OECD employment projections until 2060 and after that the contribution of employment stays around 0.5 pp per year. The hours worked per worker decrease by -0.1% per year, which corresponds to the evolution observed during the most recent historical sub-period 2005-14.¹⁶ The average length of education stays stabilized at 12.8 years.

Under these assumptions, over the period 2015-2100, US GDP growth averages 1.5% (see Chart 5), with a contribution of 0.6pp from TFP, 0.5pp from capital intensity and 0.4pp from hours worked (itself decomposed in a contribution of 0.5pp from employment and -0.1pp from hours worked per worker).

Concerning non-US countries, two scenarios are proposed. In the first one, without reforms, length of education and regulation remain stabilized at their 2014 levels. In the second, length of education and regulation converge to the US level (if their initial 2014 levels are above US ones, they remain stable) over the 2015-35 sub-period.¹⁷ In both cases, hours worked per worker growth remain on their 2005-14 trend until 2060, when they are stabilized, and employment growth corresponds to OECD projections.

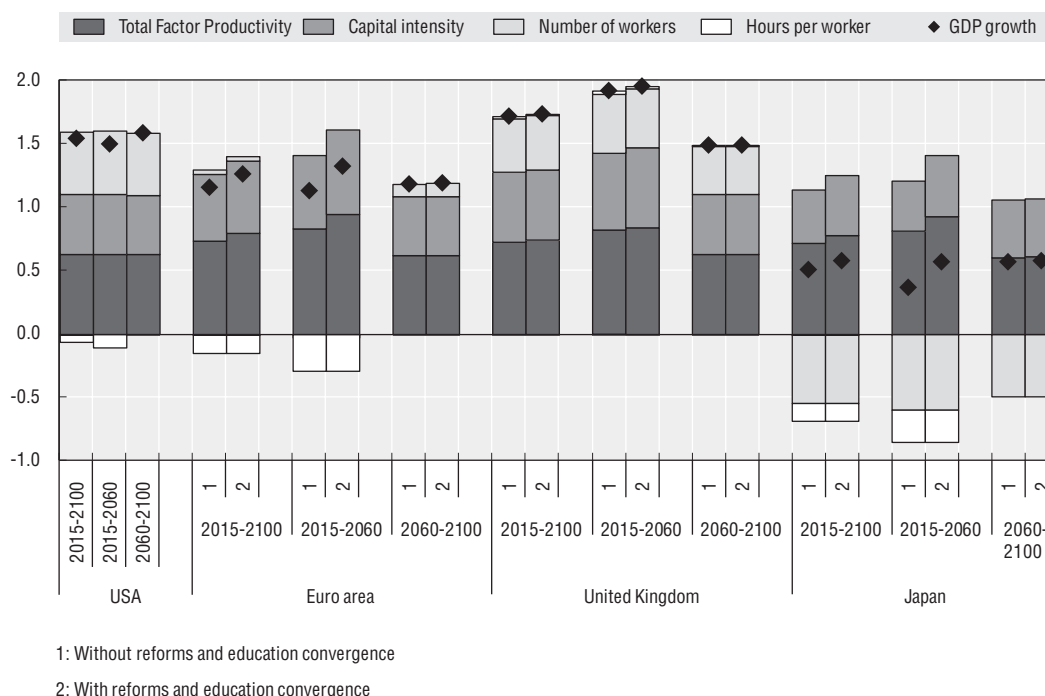
These two growth scenarios differ slightly on average for non-US countries over the 2015-2100 sub-period as reforms act on the speed of convergence, which is low as a large share of the convergence to the target was achieved by 2015 for most advanced countries. Reforms speed up convergence for the lagging countries over the first period 2015-60, with TFP gains higher by 0.2pp in the euro area and 0.1pp in Japan, but less in the UK. These higher TFP gains boost the contribution of capital intensity by 0.1pp in the three areas.

In both reform hypotheses, this secular stagnation scenario corresponds to low future GDP and productivity growth in developed countries. It means that the different headwinds identified by Gordon (2012, 1013) and mentioned in the introduction will be challenging to face. This scenario would be alarming, leading to risks of social and possibly political instability.

“Technology shock” scenarios

In the “technology shock” scenario, the US enjoys another large productivity growth wave mainly associated with the ICT third industrial revolution. This new wave could result from “More than Moore” advances (functional diversification of semiconductor-based devices) or upcoming positive technology shock (3D chips in particular) as emphasized by

Figure 5. **Average yearly GDP growth and contributions under the secular stagnation by hypothesis**



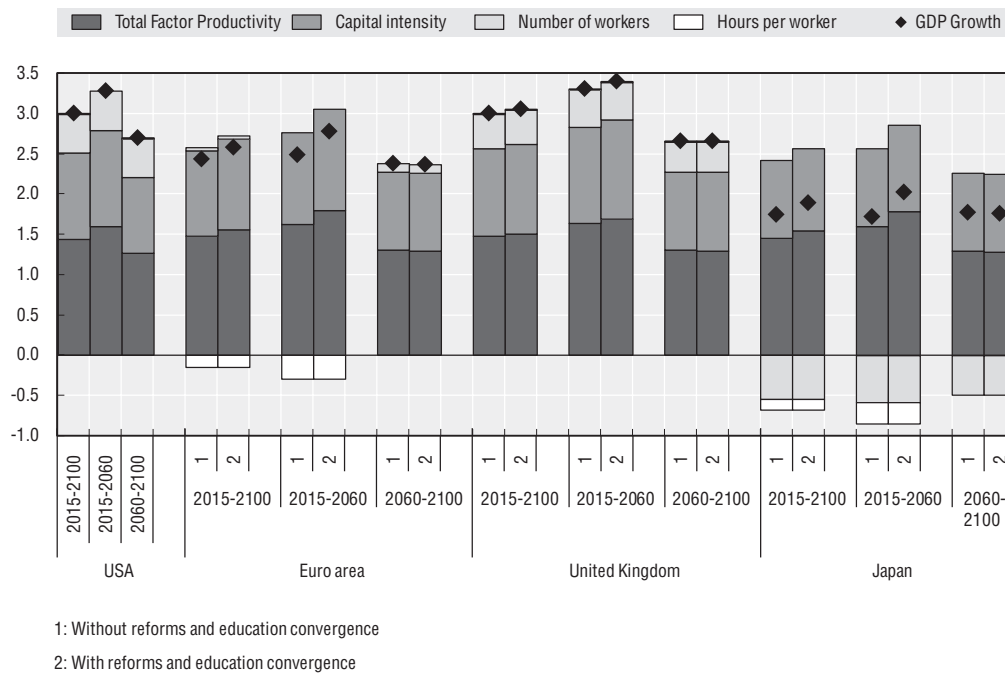
Note: This chart presents average yearly GDP growth under the secular stagnation hypothesis, with and without reforms (convergence of regulation and average years of education in the working-age population to the US level). GDP growth is decomposed into the contribution of TFP, capital intensity, the number of employees (labor) and hours worked per employee (hours), as presented in equation (1st).

the International Technology Roadmap for Semiconductors (ITRS, 2013b). It could stem from advances in machine learning (Brynjolfsson and McAfee, 2014) or robotics (Pratt, 2015). It could also come from other domains such as biotech, energy, agriculture, etc. Mokyr et al. (2015) emphasize that technological progress is hard to predict, but that professionalization in R&D activity, developments in new research tools and the overall increase in the number of world researchers support an optimistic view.

In this scenario, TFP growth is assumed to return to the pace observed during the 1990-2005 sub-period, from which we remove the contribution from the ICT technology shock¹⁸ in the US over that period (yielding a 1.3% per year TFP trend, which is also the 1995-2014 average growth rate). This contribution is removed as a technology shock is added over this trend as a technology shock which is assumed to take place over four decades, with overall TFP gains associated with the ICT third industrial revolution equivalent to those observed for electricity during the second industrial revolution in the 20th century, as estimated in Bergeaud, Cetté and Lecat (2016b). The overall technology shock hence amounts to a 27% increase in TFP over these 40 years.¹⁹ Hours worked per employee are assumed to be stable. In this scenario, the US average annual GDP growth is 1.8pp higher than in the secular stagnation scenario during 2015-60 and 1.2pp higher during 2060-2100.

In this scenario (see Chart 6), the yearly US GDP growth rate accelerates to 3% over the 2015-2100 period, with a 1.4pp contribution from TFP, 1.1pp from capital intensity and 0.5pp from employment. Thus, the contributions from TFP and capital intensity are more than double those in the secular stagnation scenario.

Figure 6. **Average yearly GDP growth and contributions under the technology shock hypothesis**



Note: This chart presents average yearly GDP growth under the technology shock hypothesis, with and without reforms (convergence of regulation and average years of education in the working-age population to the US level). GDP growth is decomposed into the contribution of TFP, capital intensity, the number of employees (labor) and hours worked per employee (hours), as presented in equation (1").

Here again, two sub-scenarios have been built for non-US countries. In the first one, education and regulation stay at their level at the beginning of the period, and the TFP catch-up process takes place at its current pace. In the second, education and regulations converge to US levels over the 2015-35 sub-period, which allows an acceleration of TFP convergence to the US level.

This acceleration takes place mostly during the 2015-60 period, with the TFP contribution 0.2pp higher in the euro area and Japan, and the capital intensity contribution 0.2pp higher in the euro area and 0.1pp higher in Japan during this sub-period. The gains are more limited for the United Kingdom (0.1pp for TFP) as regulation levels are already close to US levels. Over the whole 2015-2100 period, the GDP growth rate would come out at a yearly average of 2.5% in the euro area, 3.0% in the United Kingdom and 1.9% in Japan. This pace would correspond to the 1974-90 trend growth rate for the euro area, and the 1950-74 rate for the United Kingdom, but would remain much lower than any 20th century trend growth rate for Japan, as the overall contribution of labor would be negative.

In the first scenario (no change in education levels and no reforms), average annual GDP growth is higher than in the secular stagnation, during 2015-2060, by 1.3pp in the euro area, 1.1pp in the UK and 1.4pp in Japan. These gaps are 1.2pp in the euro area, 1.1pp in the UK and 1pp in Japan during 2060-2100.

In these technology shock scenarios, there is no doubt that the different headwinds set out by Gordon (2012, 2013) and mentioned in the introduction would be easy to face, in the US and also in non-US countries, if there were a convergence of education and regulation to US levels.

Box 1. The projection tool: A user-friendly software

The software is a user-friendly projection tool that allows the user to compute different growth scenarios for 13 developed countries based on inputted parameters concerning the technological frontier's growth path and the follower countries' convergence process. This software is freely available at www.longtermproductivity.com.

In an Excel format, the file contains 16 sheets that can be separated into 3 categories:

- The hypothesis sheet where parameters are entered concerning the projections for each country with regard to TFP and the follower countries' convergence, relative prices, average hours worked per year and per worker, employment, labor and product market regulation, and education.
- The individual countries' sheets where the user can observe visual output of his inputted parameters in the hypothesis sheet (GDP annual growth rate, TFP annual growth rate and level, capital annual growth rate, total employment growth rate, average hours worked per year and per worker annual growth rate, relative price annual growth rate, annual product and labor market regulation indexes, and annual level of education in years).
- The TFP sheet recapitulating all of the countries' TFP growth rates in data format and charts.

To program a long-term growth scenario, the user has to input assumptions into the hypothesis page.

The user is first invited to enter his assumptions about the technological frontier, the United States, with regard to TFP, relative capital prices, average hours worked per year and per worker, education and employment.

For the leading country, the user can implement (or not) a technology shock and control for its magnitude and duration, in addition to a TFP trend. There are four parameters: the trend's yearly growth rate, the wave length in years, its peak year, and its height over the trend.

For the trend's yearly growth rate, the user can either choose to directly input a percentage by clicking on "manual entry", or he can compute a past trend by clicking on "past trend" and enter into the input box a start year and an end year. If the user simulates a technology shock, he inputs its duration in years in the box labelled "wave length", in which year it reaches its peak and by what percentage in the boxes labelled respectively "wave peak year" and "wave peak height over trend". If the user does not wish to have a technology shock, he simply inputs 0% in the "wave peak height over the trend" box.

For the relative capital price, the user can either choose to manually enter a yearly growth rate, which can be directly inputted or computed on a past trend, or choose the paper's modeled trend.

Concerning average hours worked per year per worker, the user is asked to enter three parameters: the yearly growth rate, the target number of hours and its catch-up start date. For the yearly growth rate, the user can either enter it manually or compute a past trend in addition to a minimum and a maximum number of hours worked per year per worker. If the user wishes to make assumptions about working hours' legislation, he can input a target number of hours that the country will converge towards starting from the year inputted in "catch-up start date". If not, he simply manually inputs 0% in the "yearly growth rate" and the country's previous number of hours in the "target number of hours", which is indicated in the adjacent box.

Finally, the user is asked to make assumptions about the employment rate from 2060 (when the OECD's projections end). He can either input a manual entry or compute a past trend.

Box 1. The projection tool: A user-friendly software (cont.)

The user can now make assumptions about the follower countries' convergence process to the technological frontier. All countries have the same parameters related to assumptions about TFP, relative capital price, average hours worked per year and per worker, education, market regulation, and employment.

For TFP, the user can either choose to manually enter the country's catch-up parameters or he can choose the paper's modeled trend. If he chooses the "manual entry", he will control the trend's yearly growth rate (manual entry or past trend), the date at which the country starts converging to the leader's targeted TFP level, the targeted TFP level, and the speed of convergence.

The relative capital price can either be manually entered, by actually inputting a yearly growth rate or computing a past trend, or can follow the paper's modeled trend.

Average hours worked per year per worker are set through three parameters: the yearly growth rate, the target number of hours and its catch-up start date.

For the yearly growth rate, the user can either make a manual entry or compute a past trend in addition to a minimum and a maximum number of hours worked per year per worker.

If the user wishes to make assumptions about working hours legislation, he can input a target number of hours to which the country will converge, starting at the "catch-up start date". If not, he simply manually inputs 0% in the "yearly growth rate" and the country's previous number of hours in the "target number of hours", which is indicated in the adjacent box.

The user can implement up to two market regulation reforms on both the labor market and the product market. To do so, he first clicks on "regulations 1" and enters the catch-up start and end year into the new inputted index. For a second reform the step is repeated. If the user does not wish to introduce market reforms, he simply re-enters the country's previous index (indicated in the output box).

Finally, the user is asked to make assumptions about the employment rate from 2060 (when the OECD's projections end). He can either input a manual entry or compute a past trend.

At the end of the hypothesis sheet, in the parameters section, the user can see an ON/OFF switch for the HP function. We advise you to only turn it ON once you have entered all the assumptions as the Hodrick-Prescott filter is rather time consuming when computing data.

7. Concluding remarks

The analyses proposed in this paper illustrate how future growth in advanced countries is driven by technology improvements at the technological frontier (the US) and the catch-up process to this frontier in terms of productivity in other countries. However, their respective paces are very uncertain in the future. In the US, it depends on new technological changes, in particular in the ICT industry, and on the extension of the use of available ICT capacities in several areas (the "*more than Moore*" process). In other countries, it also depends on the institutional economic environment and on the capacity to implement large-scale and ambitious structural reforms.

The user-friendly software presented in the paper was used to build different scenarios illustrating these uncertainties at a very long horizon, from the current period to 2100, the end of the century. These scenarios are deliberately extremely contrasting, and show how wide the range of possible future growth is for the developed countries. In the lower

scenario, improvements in purchasing power and living standards would be very small, making it challenging to face the “headwinds” talked about by Gordon (2012, 2013). Social and political stability could be seriously threatened in such a low growth scenario. By contrast, in the upper scenario, gains in purchasing power and living standards would be great, and Gordon’s “headwinds” would be easily overcome.

There is no consensus in the literature on the future of growth drivers and, for this reason it seems difficult to attribute probabilities to these diverse scenarios. We assume that the future growth of developed countries will probably settle between the extreme scenarios described in the paper, but the literature is not yet very conclusive in helping identify exactly where. But the software can be used to build more precise scenarios for those who have a specific opinion on what may happen regarding growth drivers over the next decades.

This software is of course based on a set of hypotheses and simplifications. It does not yet allow for the introduction of different demographic scenarios or for an explicit account to be taken of the issue of environmentally sustainable growth. It is the first step in a long research program, the next two steps being the inclusion of both of these aspects. Bearing this in mind, it is still a practical tool that enables us to represent different possible growth paths, and to characterize the associated issues in developed countries.

Notes

1. The causes of secular stagnation risk were for Hansen (1939) a demographic slowdown but also a disequilibrium between saving and investment.
2. Productivity has accelerated in Spain since 2008, the beginning of the current crisis, for very specific reasons (see Bergeaud, Cetto and Lecat, 2016a).
3. Available at www.longtermproductivity.com.
4. This database can be accessed at www.longtermproductivity.com.
5. These countries are the G7 (the United States, Japan, Germany, France, the United Kingdom, Italy and Canada), four other euro area countries (Spain, the Netherlands, Belgium and Finland) and two other advanced countries (Australia and Sweden).
6. Total factor productivity is adjusted for education, using the estimated return on years of schooling of 5% from Bergeaud, Cetto and Lecat (2016b): see equation (1'). The adjustment term for education, when extracted from tfp , is hence $-0.05 \times (EDU_{US,t-1} - EDU_{i,t-1})$ with EDU , the average years of schooling in the working-age population.
7. The calculation starts with primary school and does not include kindergarten or any other type of education received before the age of 6.
8. If one body of regulation is at its maximum (6) and the other at the minimum (0), the overall regulation index will be 0. In practice, both types of regulation tend to be correlated across countries.
9. The relationship between investment-specific technology shocks and the relative price of investment may be disrupted by several factors, such as mark-up changes or variation in the use of intermediate inputs and factors of production in technology-producing sectors (see Fisher, 2009). That is why we expect this relationship to hold more strongly in the long run than in the short run.
10. Benati (2014) shows over a long time horizon that relative investment prices and TFP are not cointegrated when structural breaks are not allowed; when introduced, long-horizon covariation is either nil or negative.
11. Cointegration is not rejected only for the regression from 1970 onwards (at the 5% threshold), while it is rejected for all the other regressions (results are presented with a 20-year interval, but regressions at a 10-year interval were also tested).
12. As our specification is original, these results are not easily comparable with those in previous studies. Nevertheless, the sign and amplitude of the interaction term impact seem consistent with those in previous studies, which have also introduced interaction between the product of labor and product

market regulation indicators and the distance to the TFP frontier indicator. See for example Aghion et al. (2009).

13. And the employment rate corresponds here to the ratio of employment to population and not to the ratio of employment to working age population, as in Bourlès and Cette (2007).
14. See Cette (2014, 2015) on these aspects.
15. See Aghion et al. (2009), Guerrieri et al. (2011) and Cette and Lopez (2012), who use country-level panel data, as well as Cette, Lopez and Mairesse (2016), who employ sectoral-level panel data.
16. From 2060 onwards, hours worked per employee are stabilized.
17. This reform effort is not exceptional: such an increase of the average length of education and such a decrease in the regulation index were observed in the 1990s and the 2000s in these countries.
18. As estimated in Bergeaud, Cette and Lecat (2016b).
19. This figure corresponds to the overall gains from electricity, from which we remove the gains from ICT already accounted for up to 2015.

References

- Aghion, P., P. Askenazy, R. Bourlès, G. Cette and N. Dromel (2009), "Education, Market Rigidities and Growth", *Economics Letters* 102 (1), pp. 62-65.
- Benati, L. (2014), "Do TFP and the relative price of investment share a common I(1) component?", *Journal of Economic Dynamics and Control*, Vol. 45, August, pp. 239-261.
- Bergeaud, A., G. Cette and R. Lecat (2015): "GDP per capita in advanced countries over the 20th century", *Working Papers* No. 549, Banque de France.
- Bergeaud, A., G. Cette and R. Lecat (2016a), "Productivity trends from 1890 to 2012 in advanced countries", *The Review of Income and Wealth*, Vol. 62, issue 3, pp. 420-444.
- Bergeaud, A., G. Cette and R. Lecat (2016b), "The role of production factor quality and technology diffusion in 20th century productivity growth", *Working Papers* No. 588, Banque de France, forthcoming in *Cliometrica*.
- Barro, R. and J. Lee (2010), "A new data set of educational attainment in the world, 1950-2010", NBER, *Working Papers* No. 15902, National Bureau of Economic Research, Inc.
- Bourlès, R. and G. Cette (2007), "Trends in 'structural' productivity levels in the major industrialized countries", *Economics Letters*, Elsevier, Vol. 95(1), pages 151-156, April.
- Brynjolfsson, E. and A. McAfee (2014), "The second machine age – Work, progress, and prosperity in a time of brilliant technologies", Kindle edition.
- Byrne, D., J. Fernald and M. Reinsdorf (2016), "Does the United States have a Productivity Slowdown or a Measurement Problem?", NBER, *Working Papers Series* No. 2016-3, March.
- Byrne, D., M. Oliner and D. Sichel (2013), "Is the information technology revolution over?", *International Productivity Monitor*, No. 25, Spring, pp. 20-36.
- Cette, G. (2014), "Does ICT remain a powerful engine of growth", AFSE Presidential Address, *Revue d'Economie Politique*, 124 (4), July-August, pp. 473-492.
- Cette, G. (2015), "Which role for ICTs as a productivity driver over the last years and the next future?", *Digiworld Economic Journal*, Communications & Strategies, No. 100, 4th quarter, pp. 65-83.
- Cette, G., N. Dromel, R. Lecat and A.-C. Paret (2015), "Production Factor Returns: The Role of Factor Utilization", *The Review of Economics and Statistics*, MIT Press, Vol. 97(1), pp. 134-143, March.
- Cette, G., Y. Kocoglu and J. Mairesse (2005), "ICT and potential output growth", *Economics Letters*, Vol. 87, Issue 2, pp. 231-234.
- Cette, G., Y. Kocoglu and J. Mairesse (2009), "Productivity growth and levels in France, Japan, the United Kingdom and the United States in the Twentieth Century", NBER *Working Papers*, No. 15577.
- Cette, G. and J. Lopez (2012), "ICT demand behaviour: An international comparison", *Economics of Innovation and New Technology*, Taylor and Francis Journals, Vol. 21(4), June, pp. 397-410.
- Cette, G., J. Lopez and J. Mairesse (2016), "Upstream product market regulations, ICT, R&D and productivity", forthcoming in *The Review of Income and Wealth*.

- Cohen, D. and M. Soto (2007), "Growth and Human Capital: Good Data, Good Results", *Journal of Economic Growth*, 12(1):51-76.
- Crafts, N. and K. O'Rourke (2013), "Twentieth Century Growth", *Discussion Paper Series*, CEPR, No. 9633, September.
- Duval, R. and C. de la Maisonneuve (2010), "Long-run growth scenarios for the world economy", *Journal of Policy Modeling*, Elsevier, Vol. 32(1), pp. 64-80, January.
- Eichengreen, B. (2015), "Secular Stagnation: The Long View", *American Economic Review, Papers and Proceedings*, 105(5), pp. 66-70.
- Fisher, J. (2006), "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks," *Journal of Political Economy*, 114, pp. 413-451.
- Fisher, J. (2009), "Comment on Letting Different Views about Business Cycles Compete", *NBER Chapters*, 24, 457-474.
- Fouré, J., A. Benassy-Queré and L. Fontagné (2013), "Modelling the world economy at the 2050 horizon", *Economics of Transition*, Wiley-Blackwell, 21(4), pp. 617-654.
- Gordon, R.J. (2012), "Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds", *NBER, Working Papers* 18315, National Bureau of Economic Research, Inc.
- Gordon, R.J. (2013), "US Productivity Growth: The Slowdown Has Returned after a Temporary Revival", *International Productivity Monitor*, Centre for the Study of Living Standards, Vol. 25, Spring, pp. 13-19.
- Gordon, R.J. (2014), "The Demise of U.S. Economic Growth: Restatement, Rebuttal, and Reflections", *NBER, Working Papers* 19895, February.
- Gordon, R.J. (2015), "Secular Stagnation: A Supply-Side View", *American Economic Review, Papers and Proceedings*, 105(5), pp. 54-59.
- Guerrieri, P., M. Luciani and V. Meliciani (2011), "The determinants of investment in information and communication technologies", *Economics of Innovation and New Technology* 20 (4), pp. 387-403.
- Hansen, A. (1939), "Economic Progress and Declining Population Growth", *American Economic Review*, Vol. 29, No.1, pp. 1-39.
- ITRS (2013a), "International Technology Roadmap for Semiconductors", 2012 edition update.
- ITRS (2013b), "International Technology Roadmap for Semiconductors", 2013 edition.
- Jorgenson, D. (2001), "Information technology and the US economy", *The American Economic Review*, Vol. 91, March, No. 1.
- OECD (2013), "Protecting jobs, enhancing flexibility, A new look at employment protection legislation", Chapter 2 of the 2013 Edition of the OECD Employment Outlook.
- Koske, I., I. Wanner, R. Bitetti and O. Barbiero (2015), "The 2013 update of the OECD product market regulation indicators: Policy insights for OECD and non-OECD countries", *OECD Economics Department Working Papers*, 1200/2015.
- Maddison, A. (2001), "The World Economy: A Millennium Perspective", Development Centre Studies, Development Centre of the Organization for Economic Co-operation and Development.
- Mokyr, J., C. Vickers and N.L. Ziebarth (2015), "The history of technological anxiety and the future of economic growth: Is this time different?", *Journal of Economic Perspective*, Vol. 29, No. 3, Summer, pp. 31-50.
- Pratt, G.A. (2015), "Is a Cambrian explosion coming from robotics?", *Journal of Economic Perspective*, Vol. 29, No. 3, Summer, pp. 51-60.
- Schreyer, P. (2000), "The Contribution of Information and Communication Technology to Output Growth: A Study on the G7 Countries", *OECD Science, Technology and Industry Working Papers* 2000/2, OECD.
- Soto, M. (2002), "Rediscovering Education in Growth Regressions", *OECD Development Centre Working Papers* 202, OECD Publishing.
- Summers, L. (2014), "U.S. Economic Prospects: Secular Stagnation, Hysteresis, and the Zero Lower Bound", *Business Economics*, Vol. 49, No. 2, pp. 65-74.
- Summers, L. (2015), "Demand Side Secular stagnation", *American Economic Review, Papers and Proceedings*, 105(5), pp. 60-65.
- Van Leeuwen, B. and J. van Leeuwen-Li (2014), "Education since 1820", In: J.L. van Zanden et al. (eds.), "How Was Life?: Global Well-Being since 1820", OECD Publishing, pp. 88-101.

The quantification of structural reforms in OECD countries: A new framework

by

Balázs Égert and Peter Gal*

This document describes and discusses a new supply side framework that quantifies the impact of structural reforms on per capita income in OECD countries. It obtains the overall macroeconomic reform impacts by aggregating over the effects on physical capital, employment and productivity through a production function. On the basis of reforms defined as observed changes in policies, the paper finds that product market regulation has the largest overall single policy impact five years after the reforms. But the combined impact of all labour market policies is considerably larger than that of product market regulation. The paper also shows that policy impacts can differ at different horizons. The overall long-term effects on GDP per capita of policies transiting through capital deepening can be considerably larger than the 5- to 10-year impacts. By contrast, the long-term impact of policies coming only via the employment rate channel materialises at shorter horizon.

JEL classification: D24, E17, E22, E24, J08

Keywords: Structural reforms, simulation, per capita impact

* Balázs Égert (balazs.egert@oecd.org) and Peter Gal (peter.gal@oecd.org) are from the OECD Economics Department. The paper benefitted from useful comments and suggestions from Andrea Bassanini, Alain de Serres, Sean Dougherty, Falilou Fall, Andrea Garnero, Alexander Hijzen, Catherine L. Mann, Jean-Luc Schneider, participants at the OECD's 2016 Spring Working Party 1 meeting and two anonymous referees. The views expressed in this paper are those of the authors and do not necessarily reflect those of the OECD and its member countries.

1. Introduction

This document describes and discusses a new simulation framework that quantifies the impact of structural reforms on per capita income. The overall medium- to long-term supply side impact is derived through components of the production function: multi-factor productivity, capital intensity and the employment rate. It builds on and extends the previous frameworks developed in the Economics Department (ECO), which include Bouis and Duval (2011); Barnes et al. (2013); and the long-term scenario model elaborated in Johansson et al. (2013) (see Figures A1.2-A1.4 in Égert and Gal, 2016). The framework used in ECO for the G-20 and occasional country desk quantifications is based on Bouis and Duval (2011). In what follows, this will be referred to as the old framework.

In developing a new empirical framework for the quantification of the medium- to long-term impact of reforms, a number of objectives have been established at the outset (OECD, 2015). One is to extend the set of policy channels included in the framework to significantly broaden the range of quantifiable reforms. A second objective is to update the framework to cover the post-crisis period. A third objective is to improve the framework's internal consistency with respect to the country and time coverage, the empirical specification and econometric method used to estimate policy impacts. Fourth, country-specific policy effects should be better taken into account. A final objective is a better integration of emerging market economies.

This document is a first step towards addressing the first three objectives. First, relative to the old framework, the number of policy variables and channels through which they influence GDP per capita is increased. Second, the new framework is based on relationships estimated over a period including the immediate post-crisis years (1985-2011), although it ends in 2011 due to the availability of some policy indicators. Third, internal consistency is increased by a considerable extent: new estimates for the three supply-side channels are based on a similar sample of countries and time span. The sample size is almost identical for the individual supply-side channels.¹ The estimates are also based on a unified estimation framework (dynamic OLS in a panel setting).

This document could be viewed as work in progress. The estimated relationships will be re-assessed, and new and more detailed policy variables will be added as need arises and new data become available. The estimation approach will be refined. Ongoing work looking into ways how to incorporate country-specific policy influences (Égert and Gal, 2016). The results presented in this document are based on *average* policy effects obtained on an OECD sample. Country-specific effects can be identified in a panel setting by conditioning the impact of individual policies on the stance of other policies or via policy interactions. This will allow for the incorporation of a potentially large set of additional policy areas including institutions and policy areas with limited time-series availability (e.g. subcomponents of the PMR indicator, housing market regulations or the rule of law indicator). Work on integrating EMEs into the quantification framework will also be completed in the next phase. This prompts two important questions: i) an increased cross-country heterogeneity in the

estimations, and ii) a more detailed look at the role of the quality of institutions, which may determine the relation between policies and outcomes. Finally, the estimates based on macroeconomic time series will be cross-checked with those obtained from using sector- and firm-level data to better control for potential endogeneity.²

This document is structured in three parts. Part 1 compares the main features of the old and the new frameworks. In particular, it discusses how the new framework improves on the old framework by i) incorporating more policy channels, ii) enhancing the internal consistency of the policy impacts, and iii) using updated estimates for them. Part 2 presents in detail how these new coefficient estimates are obtained. Part 3 finally sets out the new simulation framework and shows the impact of policy changes on per capita income and its supply-side components. It differs from typical OECD studies analysing *specific* policy impacts in a number of ways. First, the simulation framework considers the impact of a large number of policies by controlling for a wide range of other policies (including them in the regressions at the same time). Second, most projects typically look at the isolated relationship between policies on the one hand productivity, investment or labour market outcomes on the other hand. The framework is used to aggregate these effects into an overall impact on per capita income.

2. A consistent framework to quantify the impact of structural reforms

We follow previous OECD papers by assessing the impact of structural reforms on per capita income based on a production function approach (Barnes et al., 2013; Bouis and Duval, 2011; Johansson et al., 2013). In this paper, the influence of policies on GDP is assessed through their influence on its supply side components: multi-factor productivity (MFP), capital intensity and employment. Within this framework, the relationship between policies and these components is obtained from a range of cross-country reduced-form panel regressions. The overall impact on GDP per capita is obtained by aggregating the policy effects of the various channels.³

The framework presented in this paper seeks to improve on a number of dimensions compared to previous OECD studies. First, a considerably larger number of policy determinants are analysed for MFP and employment and policy determinants of capital deepening are introduced. Second, internal consistency is improved in three ways. To start with, supply-side channels are used in a consistent manner: different levels of disaggregation of the supply side components are not mixed across policy areas (e.g. employment for some policies, the labour force participation and unemployment rate for others). In addition, econometric estimates are obtained using the very same up-dated dataset (SPIDER dataset) for as many countries as possible (25 for the employment rate and above 30 for MFP and capital deepening) and estimation technique. Finally, changes in policy measures and the horizons at which their impact is measured are standardised. The last major improvement relates to the updating of the coefficient estimates. We make use of econometric estimates covering the post-crisis period (mostly until 2011-13) for all three supply-side channels (MFP, capital deepening and the employment rate).⁴

3. Empirical estimates used in the simulation framework

The baseline estimates used in the simulation framework are taken from Égert (2016a, 2016b) and Gal and Theising (2015). They are summarised hereafter.⁵

3.1. Multi-factor productivity

We calculate aggregate MFP as the residual of output once capital and labour are accounted for. Human capital is included in our measure of MFP: MFP series, which exclude human capital (output minus human and physical capital and labour) are implausible: a decline over decades in countries close to the frontier, and the USA being far from the frontier (see details in Égert, 2016a).

The policy determinants of MFP are taken from Égert (2016a) and the main results are summarised in Table 1 (columns 1 to 3). These results suggest that a strong negative relationship can be identified between *product market regulations*, captured by the overall ETCR indicator (and its subcomponents barriers to entry and public ownership) and MFP mostly if only country but no year fixed effects are used.⁶ If both country and time fixed effects are included in the regressions, only the coefficient estimate on public ownership is found to be statistically significant, overall ETCR and barriers to entry have large standard errors for the full sample. When using a subsample, given by data available on general spending on basic research (rather than business spending on R&D funded by industry used for the large sample), the overall ETCR indicator and its subcomponents become again negative and statistically significant.

None of the *labour market regulation* indicators⁷ are statistically significant when added one by one to the baseline regression.⁸ But ALMP becomes statistically significant if used in logs. The positive sign on ALMP may indicate that more spending on ALMP helps labour reallocation towards more productive uses by reducing skill mismatches (Adalet-McGowan and Andrews, 2015). Adding year fixed effects to the regressions switches off the innovation intensity variable. In the simulations, estimates obtained using country fixed effects only will be used.

3.2. Capital deepening

The policy determinants of capital deepening are taken from Égert (2016b). The results are summarised in Table 1 (columns 4 and 5). In the baseline specification, K/Y is regressed on the user cost of capital (decomposed into relative investment prices, the real interest rate and corporate taxes) and product market regulation. Labour market policies are added to the regressions in a later stage. The results show that the corporate taxes-to-GDP ratio and the relative investment price variable bear the expected negative sign and are statistically significant. But the real interest rate is found to be only weakly related to the capital stock. To be fully consistent with theory, we keep all three components of the user cost of capital in the specifications augmented by structural policy indicators.

Product market regulation, measured by the ETCR indicator, shows a fairly robust negative relationship with the capital stock series. It is robust to alternative (smaller) country samples and time coverage (excluding the crisis). Finally, the results do not change ostensibly if the regressions are carried out on a reduced common sample covering all labour market policy indicators as well (Table 1).

The employment protection legislation (EPL) indicator has a strong and quantitatively important negative relationship to the capital stock. Its coefficient estimate is precisely estimated for both the level and log-linear specifications. The estimated elasticities indicate that a one-step increase in EPL is associated with a decline of about 0.2 per cent in the capital stock-to-output ratio.⁹ It should be noted that the negative relation between the ETCR indicator and the capital stock remains statistically significant and of the same magnitude (Table 1).

Table 1. **MFP and investment: long-term coefficient estimates used in the simulations**

Explanatory variables	Dependent variable				
	Multi-factor productivity			log(capital stock/output)	
	(1)	(2)	(3)	(4)	(5)
c	9.863**	10.492**	10.244**	0.956**	1.261**
Product market regulation					
ETCR aggregate	-0.037**		-0.047**	-0.035**	-0.025**
ETCR public ownership		-0.027**			
Labour market policies					
log ALMP			0.029**		
EPL					-0.152**
Intermediate outcomes					
trade openness (size adjusted)	0.007**	0.008**	0.006**		
business exp. On R&D by industry	0.071**	0.059**	0.047**		
elements of the user cost of capita					
log relative investment prices				-0.377**	-0.608**
long-term real interest rate				0.002	0.004*
corporate taxes/GDP				-0.024**	-0.026**
error correction term	-0.033**	-0.043**	-0.052**	-0.026**	-0.022**
Adjusted R-squared	0.952	0.959	0.964	0.919	0.947
Country/year fixed effects	Yes/No	Yes/Yes	Yes/No	Yes/Yes	Yes/Yes
No. of observations/countries	755/34	755/34	570/32	705/32	600/31
Years	1985-2011	1985-2012	1985-2013	1985-2014	1985-2015

Note: Estimates based on using the dynamic OLS (DOLS) estimator with one lag and one lead. Standard errors are heteroscedasticity robust. ** and * denote significance at the 5% and 10% level, respectively. The MFP regressions include human capital and output gap and the capital deepening regression output gap as control variables.

3.3. Employment rate

Policy determinants of the employment rate are taken from Gal and Theising (2015), who build on previous work from the OECD Reassessed Jobs Strategy (Bassanini and Duval, 2006). The impact of policies on the aggregate employment rate is derived from estimation results obtained for four demographic groups (youth, prime age women and man; and the elderly). Such a breakdown makes it possible to capture more policies than regressions focusing directly on the overall employment rate or a breakdown by skill levels. Estimation results, summarised in Table 2, show that among tax-benefit and activation policies, lower unemployment benefit replacement rates, increased spending on active labour market policies (ALMP), lower labour tax wedges all tend to boost employment. These findings broadly confirm existing results but based on an updated dataset and on a somewhat different methodology.¹⁰ Regarding wage-setting institutions, the excess coverage of wage agreements, its interaction with the tax wedge and the level of minimum wage tend to affect employment rates negatively. Product market regulation captured by the ETCR indicator correlates negatively with the aggregate employment rate. Finally, while the EPL indicator has no clear robust relationship with the aggregate employment rate (Table A4.3), it does have a heterogeneous impact across various segments of the population. This makes aggregate effects potentially dependent on the composition of the working age population by skills and demographic groups.

Indeed, some of the other policy effects also show significant and intuitive heterogeneities across segments of the population

- Higher unemployment benefit replacement rates affect most negatively the employment of the elderly and the low educated;

- ALMP spending has positive effects for each segments of the population, mostly so for the youth;
- Minimum wage impacts are stronger for the youth than for prime age men and the elderly.¹¹ The minimum wage impacts are even stronger for prime wage women.¹²
- Stricter employment protection legislation tends to decrease female employment rates, although this result is not robust to the exclusion of family policies. Effects on men are ambiguous. It also has opposing effects on the low educated (lowering employment) and on the highly educated (increasing employment) (Table A4.4). The explanation can be that stricter EPL may hold back less the hiring of highly qualified workers (they are more likely to provide a better match for the firm) and impacts less their firing (which may be more costly). These opposing effects across groups of the population help to explain why it is difficult to find robust aggregate effects;
- Raising the legal retirement age increases labour force participation for the elderly;
- More spending on in-kind family benefits such as childcare and longer maternity leaves increase employment rates of the working-age female population.

Table 2. **Employment rate – long-term coefficient estimates**

Dependent variable: employment rate				
Explanatory variables	Youth	Prime age women	Prime age men	Elderly
	(1)	(2)	(3)	(4)
Tax-benefit and activation policies				
UE benefit replacement rate	-0.183**	-0.204**	-0.147**	-0.343**
ALMP spending on unemployed, as % of GDP/capita (HP-trend)	0.147**	0.092**	0.047**	0.063**
Tax wedge (single, no ch.)	-0.866**			
Tax wedge (couple, 2 ch.)		0.004	-0.274**	-0.260**
Wage setting institutions				
Excess coverage	0.072	-0.171**	0.025	0.105
Excess coverage * Tax wedge (single, no ch.)	-3.627**			
Excess coverage * Tax wedge (couple, 2 ch.)		-0.938**	0.079	0.623*
Minimum wage (%median)	-0.311**	-0.421**	0.043	-0.093
Labour and product market regulations				
EPL regular contracts	1.599	-2.746*	-0.569	1.710
ETC regulation	1.032	-1.533**	0.232	0.630
Dummy for the presence of minimum wage (%median)	7.128	13.516**	-1.823	2.452
Policies primarily affecting women				
Family benefits in cash (% of GDP)		-0.967		
Family benefits in kind (% of GDP)		4.698**		
Number of weeks of maternity leave		0.265**		
Pension system – primarily affecting the elderly				
Legal age for pensions (total)				0.851**
error correction term	-0.303**	-0.145**	-0.294**	-0.160**
Adjusted R-squared	0.978	0.960	0.907	0.977
Country/year fixed effects	Yes/Yes	Yes/Yes	Yes/Yes	Yes/Yes
No. of observations/countries	422/25	420/25	420/25	422/25
Years	1987-2010	1987-2010	1987-2010	1987-2010

Note: Estimates based on using the dynamic OLS (DOLS) estimator with one lag and one lead. Heteroskedasticity and autocorrelation robust (Newey-West) standard errors are shown in parentheses. See more details and Gal and Theising (2015). ** and * denote significance at the 5% and 10% level, respectively.

4. The new simulation framework

This section illustrates how the new quantification framework. Box 1 describes how policy impacts on the three supply-side channels can be aggregated to total per capita income effects.

Box 1. Calculating total policy impacts on per capita income

Theoretical considerations

In the new framework, similarly to previous frameworks, structural policies affect per capita income through the supply side components. The appropriate aggregation across the components is straightforward in a standard neo-classical model with a Cobb-Douglas aggregate production of the following form.

$$Y = K^\alpha (hL)^{1-\alpha}, \quad 0 < \alpha < 1 \quad (1)$$

with h denoting labour-augmenting (Harrod-neutral) technological progress. Note that the empirical construction of the MFP measure that is used for the estimations relies on the formulation in equation (1).¹³ However, under the assumption of constant returns to scale, equation (1) can be rewritten in the following way:

$$Y = \text{MFP}(K^\alpha L^{1-\alpha}) \quad (2)$$

where there is a very close link between multi-factor productivity (MFP) and h : $\text{MFP} = h^{1-\alpha}$. Introducing per capita measures and after some rearrangements, per capita income can be expressed as a function of MFP, the capital-output ratio (K/Y) and the employment rate (L/N_{wa}):

$$\ln\left(\frac{Y}{N_{pop}}\right) = \frac{1}{1-\alpha} \ln(\text{MFP}) + \frac{\alpha}{1-\alpha} \ln\left(\frac{K}{Y}\right) + \ln\left(\frac{L}{N_{wa}}\right) + \ln\left(\frac{N_{wa}}{N_{pop}}\right) \quad (3)$$

where N_{pop} and N_{wa} stand for total population and working age population, respectively.

The advantage of this formulation is that in a standard setting, all components are separable and independent from each other. Specifically, the capital-output ratio does not depend on either productivity or employment, neither is the employment rate influenced by productivity or capital.¹⁴

For simulating the effects of changes in policies, the above equation will be used in growth rates:

$$\Delta \ln\left(\frac{Y}{N_{pop}}\right) = \frac{1}{1-\alpha} \Delta \ln(\text{MFP}) + \frac{\alpha}{1-\alpha} \Delta \ln\left(\frac{K}{Y}\right) + \Delta \ln\left(\frac{L}{N_{wa}}\right) + \Delta \ln\left(\frac{N_{wa}}{N_p}\right) \quad (4)$$

where Δ captures differences over time, which can be interpreted as percentage changes. As mentioned above, MFP in our empirical framework uses the Harrod-neutral specification. Hence (4) can be rewritten as follows:

$$\Delta \ln\left(\frac{Y}{N_{pop}}\right) = \Delta \ln(h) + \frac{\alpha}{1-\alpha} \Delta \ln\left(\frac{K}{Y}\right) + \Delta \ln\left(\frac{L}{N_{wa}}\right) + \Delta \ln\left(\frac{N_{wa}}{N_p}\right) \quad (5)$$

Similarly to the calculation of MFP a standard value for capital elasticity is set in the simulations ($\alpha = 0.33$). The last term capturing the share of working age population will be assumed to be unchanged over the simulation horizon. Alternatively, demographic projections by the United Nations could be used over the projection horizon (long-term scenarios project of the OECD, Johansson et al., 2013).

Box 1. Calculating total policy impacts on per capita income (cont.)**Practical considerations**

MFP and capital deepening are measured in logarithms, while the employment rate is measured in percentage points (between 0 and 100). The simulation framework requires that the reform impacts are expressed in log-points for each supply side component. Percentage point changes in the employment rate are thus transformed into log-points by dividing the changes in the employment rate by the latest observed employment rate for the working age population L/N_{wa} (which was 67% in 2013, averaged across all countries in the sample):

$$\Delta \ln \left(\frac{L}{N_{wa}} \right) = \frac{\Delta \left(\frac{L}{N_{wa}} \right)}{\frac{L}{N_{wa}}}$$

Another issue about aggregation is how to obtain the aggregate employment effect from the demographic and skill groups of the population. Policy effects for these groups are aggregated using the groups' weight in the working age population. For the illustrative simulations presented in this paper, the population structure of the average OECD country is used in the latest available year (Table A6.2).

4.1. Choosing the size of policy changes for illustrative purposes

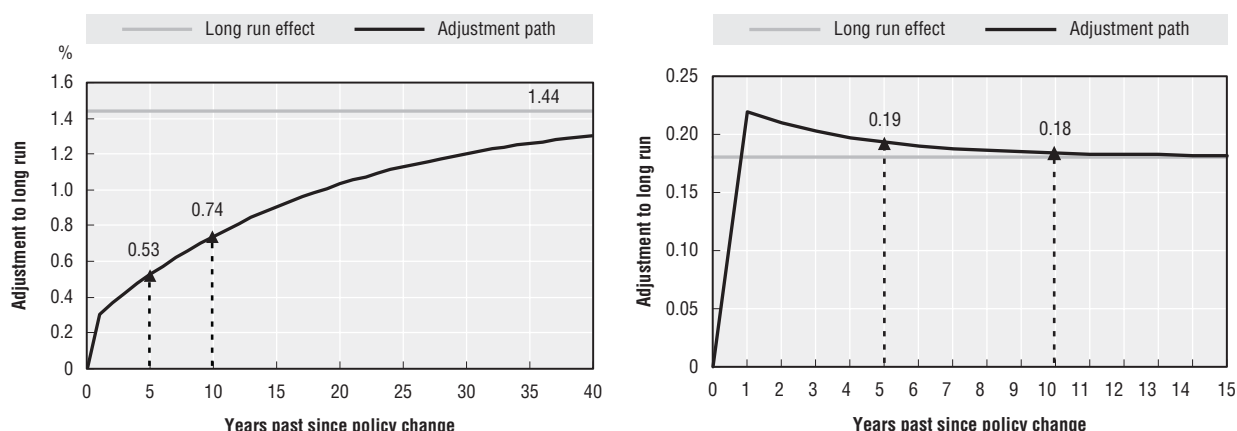
One needs to determine the magnitude of changes in the structural policy indicators to quantify the impact of structural policies on per capita income. Ideally, in each case, one should use details on planned policy changes and translate them into the policy indicators used for the estimations. In practice, this is not always possible. Details are not always sufficient and if they are, it is not always easy to map them into our policy indicators.

For such cases and for illustrative purposes, we define a reform measure for each policy, which is based on policy changes observed in the past. More specifically, reforms are determined as the average improvements in the policy indicators in two-year windows. Only those consecutive years are used during which policy indicators moved into the reform direction in both years. Appendix 5 in Égert and Gal (2016) discusses alternative reform scenarios. It is important to stress that the impact of reforms is linear: if reform intensity (the change in the policy indicator) doubles, the impact on per capita income will also double.

4.2. Obtaining policy effects over different horizons

Policy effects identified by the coefficient estimates are the long-run effects (Tables 1 and 2). This implies full convergence, which can take a very long time, depending on the estimated speed of adjustment parameters. For MFP and capital deepening, the estimated error correction terms are estimated in the range of -0.03 to -0.05, and for employment, they are around -0.2 (Tables A4.1A4.2 in Appendix 4 in Égert and Gal, 2016). A speed of adjustment coefficient of $\rho = -0.05$ implies that 90% of the convergence occurs after about 45 years.¹⁵ This long term convergence can be speeded up if the short term effect is large and points to the same direction. Such an initial “boost” is found for the ETCR impact on MFP, for instance (see Figure 1). Policymakers are typically interested in policy impacts at shorter time horizons. Therefore, policy impacts are also calculated and presented for 5 and 10 years after the reforms took place.

The adjustment path is calculated using the estimated error correction term and the contemporaneous, short run policy effect from the error correction model. The adjustment

Figure 1. **Adjustment path towards the long-run impact: two examples**

Note: ETCR stands for Energy, Transport and Communication Regulation. The size of the reforms are measured by the average changes of the indicators to a more favourable direction, observed over two consecutive years, across all OECD countries in the sample and years (1985-2013). The estimates for the long- and short run effects of the policies and the convergence parameters are taken from column 3 of Table 1 and Table A4.1 for MFP and from Table A4.3 for the employment rate. For each year t following the policy change, the adjustment path of the policy effect is $[1 - (1 + \rho)^{t-1}](\beta x - \gamma x) + \gamma x$, where ρ , β and γ are the estimates for the speed of adjustment, the long run and the short run impact of the policy, and x is the size of the reform shock.

path uses estimates of the short-run effects and the speed of adjustment terms from Tables A4.1-A4.2 in Appendix 4 in Égert and Gal (2016). The overall impact will therefore depend on the adjustment path to the long-run equilibrium and the number of years considered after the policy changes. To compute policy impacts over the desired horizon, the convergence path for each policy is evaluated for each supply-side channel. These effects are then aggregated across policies and supply side components, in accordance with equation (5) in Box 1.

The policy effects are linear with respect to changing the size of the reform measure, but they are non-linear depending on the time horizon over which they are evaluated. Put differently, the impact of reforms is twice as much if the reform shock is twice as large. However, the impact 5 years after reforms occurred can be different from the half of the impact predicted for 10 years after the reforms were introduced. As described above, the effect is determined by the shape of the adjustment path to the total long term effect. Figure 1 shows these adjustment paths and the resulting non-linear impacts in terms of different horizons for two different policies. The adjustment path for an ETCR impact on MFP is slow and decelerating. Adjustment of an ALMP impact on the employment rate is characterised by an initial overshooting and then a gradual return to the long-run equilibrium.

4.3. Simulation results

This section presents the impact of policies on per capita income levels. The overall impact mostly reflects average effects across countries, as country-specific effects play a much smaller role in the new framework at the current stage.

4.3.1. Average country effects in the new framework

Overall impact on per capita income 5 years after the policy changes took place (Figure 2 and Table 3):

- Product market regulation measured by the ETCR indicator has the largest overall direct policy impact: 0.7%. This is because ETCR influences per capita income through all three supply-side channels (productivity, capital deepening and the employment rate). The

impact through MFP is 0.5. This is comparable with the impact of other policies. For instance, the impact of increased ALMP spending, a reduction in tax wedge and in the minimum wage or in the number of maternity leave weeks ranges from 0.36% to 0.49% (Table 3, column 2). The influence of trade openness is similar in magnitude of that of ETCR. However, it requires further empirical work to pin down through which policies openness impact on MFP and thus on per capita income.

- Some policies have only a minor impact on per capita income levels. For instance, raising the legal retirement age, increasing business spending R&D funded by industry and cutting excess coverage add about 0.1% to per capita income when using past positive policy changes over two consecutive years (Table 4, column 2). The small impact of R&D may seem puzzling. There is some evidence that less developed OECD countries may benefit more from technology diffusion and adoption through the trade channel and that innovation intensity (more spending on R&D) may be more important for more advanced OECD countries (Égert, 2016a).
- The overall impact of all labour market policies is considerably larger than the overall effect of ETCR and the remaining policies (corporate taxes and R&D spending). Nevertheless, adding up the effect of all labour market policies implies that reforms are carried out simultaneously in all quantifiable policy areas, which is rather implausible in practice.
- Table 3 (and Tables A6.3 to A6.4 containing detailed impacts on longer horizons) can indeed provide precious help for policy makers for the elaboration of comprehensive structural reform packages. Depending on the ease with which reforms can be implemented, policies could be picked to reach policy objectives in terms of overall impact on per capita income.

Comparing the impact of policies across different horizons: 5 and 10 years after policy changes and the long-run effect (Table 4):

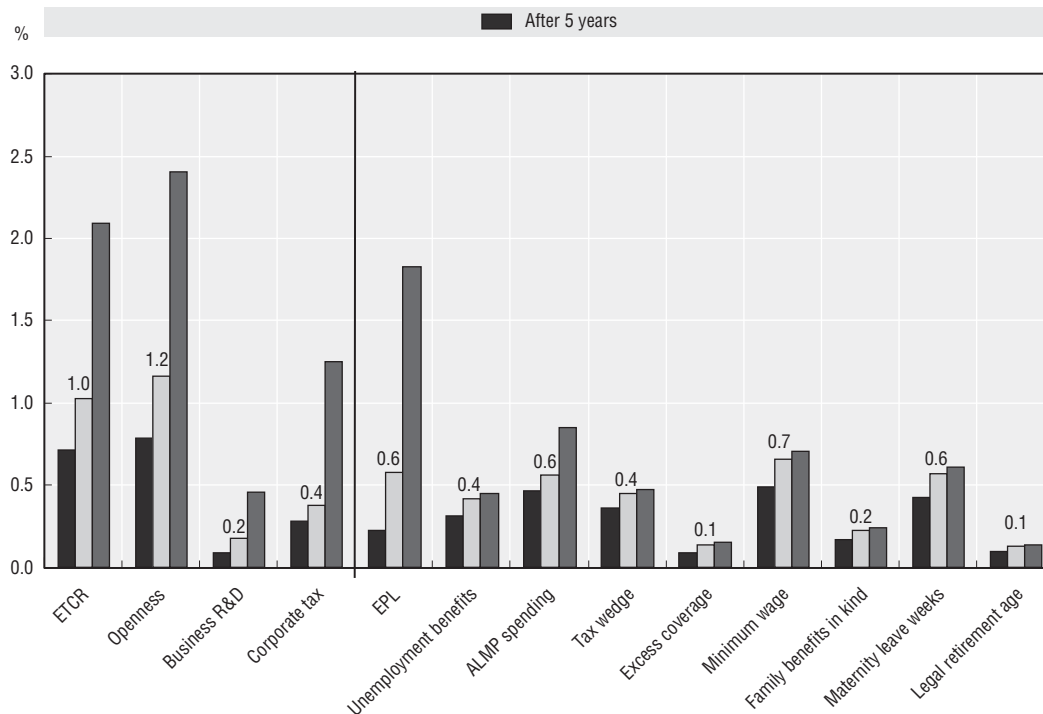
- For some policies, the overall long-term effects on GDP per capita can be considerably larger than the 5- to 10-year impacts. These policies include ETCR, EPL and ALMP spending.
- The total impact of other policies, mostly labour market policies transiting only via the employment rate channel, materialises at shorter horizon. Hence, the impact at different horizons – reported in Table 6 – are similar in magnitude.

Impact of policies on the separate supply-side channels (Figure 3):¹⁶

- Different policies have different impacts on the separate supply-side channels. For instance, corporate taxes have a much larger impact on investment compared to product market regulations (ETCR). The impact of legal retirement age on the employment rate is roughly 5 times smaller than those found for minimum wages, maternity leave weeks or ALMP spending (Figure 4).
- The impact of EPL reform on the employment rate is subject to large uncertainties. The effect is very small and negative if estimates from aggregate employment equations are used, reflecting the mechanism that in the short run, less strict EPL can induce more firings. However, when the overall impact is obtained by aggregating the impacts across separate segments of the population – either by educational attainment levels or by demographic groups – positive effects can be identified, up to half a percentage point. Uncertainties about the *average employment* impact of EPL have also been emphasised in earlier findings in the literature (Boeri et al., 2015).

- Some policy areas (family benefits, pension age) can be assessed and included in the framework only by demographic groups. Hence as a benchmark in future quantification exercises, the effects found when aggregating across demographic groups could be used.

Figure 2. **The impact of reforms on GDP per capita 5 and 10 years after the reforms**



Note: Typically observed reforms are measured here by the average of all beneficial two-year policy changes that were observed over two consecutive years in the sample.

4.3.2. Country-specific effects in the new framework

The new simulation framework allows for country heterogeneities in three different ways:

- Heterogeneous effects through different reform intensity (see Table A5.1 in Appendix 5 in Égert and Gal, 2016).
- Heterogeneous effects through the interaction of wage setting institutions. The effect of labour tax wedge is found to depend on the level of excess coverage of wage bargaining (Gal and Theising, 2015, and de Serres, Hijzen and Murtin, 2014). The payoffs from lowering the average labour tax wedge can be substantially higher for countries with higher excess coverage. The gains in the employment rate from a cut in the labour tax wedge (a reduction corresponding to the average observed over two consecutive years) are by 0.4 percentage point larger for a country with a high excess coverage than for one where it is low (Figure 4).
- Heterogeneous effects through differences in the age, gender and skill composition of the population. In principle, they can be another source of country-specificity in the framework, as the labour market results show heterogeneities across these groups. However, these effects are rather small in practice. For instance, using the lowest and the highest shares of the most responsive groups to policies (i.e. youth or low educated)

Table 3. **The impact of reforms on GDP per capita and its supply side components 5 years after the reforms**

Structural policy areas	Size of a typically observed reform	Total effect on GDP per capita	Impact on supply side components		
			MFP	K/Y	L/N
		<i>In percent</i>	<i>In percent</i>	<i>In percentage points</i>	
Product market regulation		0.72			
ETCR (0-6, 6 is strictest)	-0.31	0.72	0.53	0.07	0.10
Intermediate policy channels mainly affecting productivity		0.87			
Openness	4.01	0.79	0.79		
R&D (business exp.)	0.10	0.09	0.09		
Investment specific policies		0.28			
Corporate tax	-0.98	0.28		0.57	
Labour market policies		2.88			
Labour market regulations		0.22			
EPL (regular contr., 0-6, 6 is strictest)	-0.30	0.22		0.24	0.07
Tax-benefit and activation policies		1.39			
Unemployment benefits	-1.42	0.31			0.21
ALMP spending	3.18	0.46	0.09		0.25
Tax wedge	-2.28	0.36			0.24
Tax wedge (single)	-1.39	0.25			0.17
Wage setting institutions		0.58			
Excess coverage	-1.89	0.09			0.06
Minimum wage	-2.48	0.49			0.32
Labour market policies for specific demographic groups		0.69			
Family benefits in kind	0.11	0.17			0.11
Maternity leave weeks	4.83	0.42			0.28
Legal retirement age	0.57	0.10			0.06

Note: Typically observed reforms are measured here by the average of all beneficial two-year policy changes that were observed over two consecutive years in the sample. The total GDP/capita effect uses the employment effects from the aggregation across demographic groups, and it uses the formula in eq. (5) to aggregate across supply side components.

makes only a marginal difference (0.1 to 0.2 percentage point) for the final impact on aggregate employment (these results are not reported here). Hence in practice, differences in the relative size of these demographic groups, as well as in the skill composition of countries have only limited effects on per capita income. This serves as one more motivation to expand the framework with more non-linear effects and interactions between policy areas.

5. Conclusion

This document describes and discusses a new simulation framework that quantifies the impact of structural reforms on per capita income. Compared to earlier attempts, the new framework developed in this paper broadens the range of quantifiable reforms, updates the underlying empirical relationships, covers the post-crisis period and improves the framework's internal consistency. The document presents the new coefficient estimates on the three main supply-side components (MFP, capital and employment). The paper is a step in a gradual, ongoing process to continuously improve and update the quantification of the effect of structural reforms on per capita income levels. Further work is needed to better account for country-specific effects and to extend the analysis to emerging market economies. Last but not least, the extent to which the macroeconomic estimates are consistent with results obtained on the basis of sector- and firm-level data will be verified in future work.

Table 4. **The impact of reforms on GDP per capita at different horizons**

Structural policy areas	Size of a typically observed reform	Total effect on GDP per capita		
		After 5 years %	After 10 years %	After full convergence (long run) %
Product market regulation		0.72	1.02	2.09
ETCR (0-6, 6 is strictest)	-0.31	0.72	1.02	2.09
Intermediate policy channels mainly affecting productivity		0.87	1.34	2.86
Openness (perc. of GDP)	4.01	0.79	1.17	2.40
R&D (business exp.) (perc. of GDP)	0.10	0.09	0.17	0.46
Investment specific policies		0.28	0.38	1.25
Corporate tax (as perc. of GDP)	-0.98	0.28	0.38	1.25
Labour market policies		2.88	4.03	5.78
Labour market regulations		0.22	0.57	1.83
EPL (regular contr., 0-6, 6 is strictest)	-0.30	0.22	0.57	1.83
Tax-benefit and activation policies		1.39	1.74	2.10
Unemployment benefits (perc. of earnings)	-1.42	0.31	0.42	0.45
ALMP spending (per unemployed, as perc. of GDP/capita)	3.18	0.46	0.57	0.85
Tax wedge (perc.points)	-2.28	0.36	0.45	0.47
Tax wedge (single) (perc.points)	-1.39	0.25	0.31	0.32
Wage setting institutions		0.58	0.79	0.86
Excess coverage (perc.points)	-1.89	0.09	0.14	0.15
Minimum wage (perc. of median)	-2.48	0.49	0.66	0.70
Labour market policies for specific demographic groups		0.69	0.92	0.99
Family benefits in kind (perc. of GDP)	0.11	0.17	0.23	0.24
Maternity leave weeks	4.83	0.42	0.57	0.61
Legal retirement age	0.57	0.10	0.13	0.14

Note: Typically observed reforms are measured here by the average of all beneficial two-year policy changes that were observed over two consecutive years in the sample. The total GDP/capita effect uses the employment effects from the aggregation across demographic groups, and it uses the formula in eq. (5) to aggregate across supply side components.

Using typical past reforms as a basis for simulation, the new framework provides a number of results for the main policy variables. First, five years after the reforms, product market regulation has the largest overall single policy impact. However, the combined impact of all labour market policies is considerably larger than that of product market regulation and the remaining policies combined (corporate taxes and R&D spending). Some specific policies have only a minor impact on per capita income levels. They include the legal retirement age and business spending on R&D.

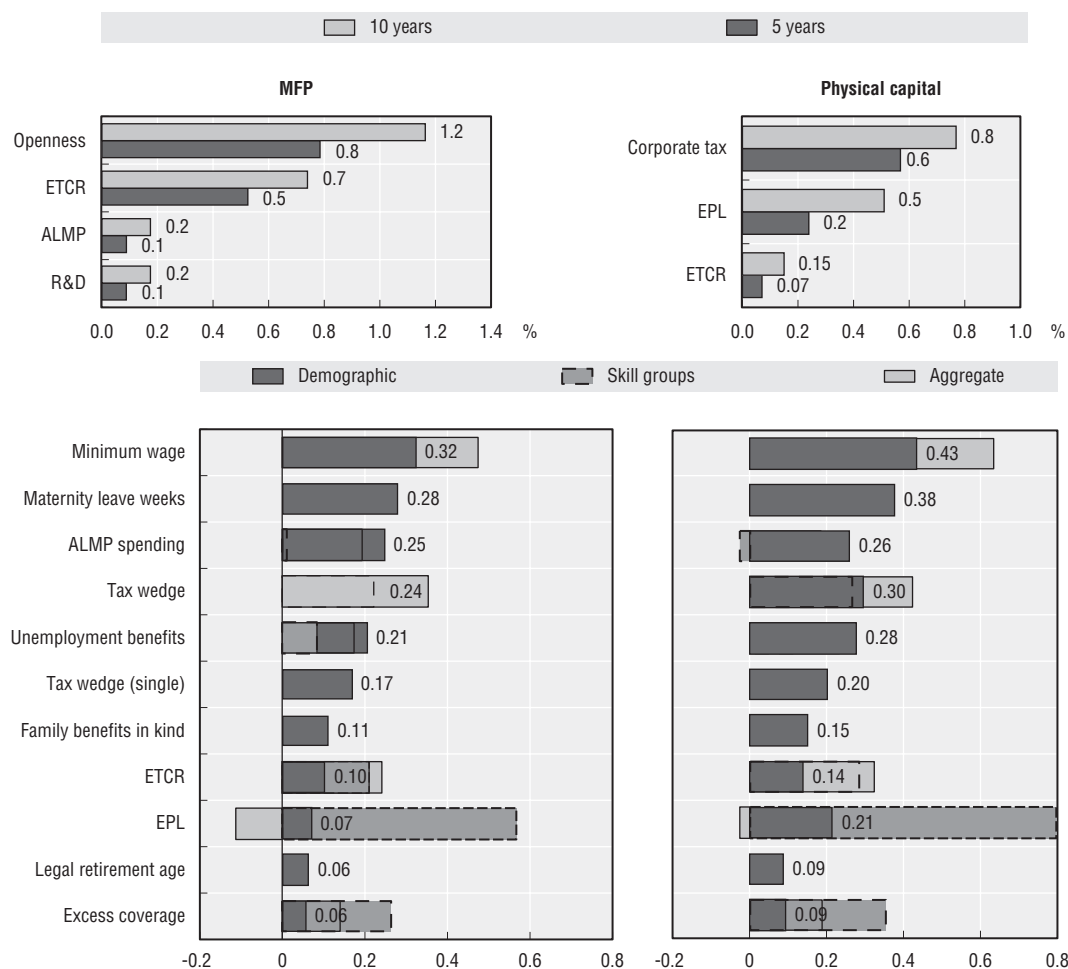
Second, policy impacts can differ at different horizons. For some policies, the overall long-term effects on GDP per capita can be considerably larger than the 5- to 10-year impacts. This is particularly the case of policies that influence GDP per capita through capital deepening (product market regulation, employment protection legislation and spending on active labour market policies). The long-term impact of other policies, mostly labour market policies transiting only via the employment rate channel, materialises at shorter horizon.

Third, the new framework shows the determinants of policies through the separate supply-side channels:

Fourth, different policies have different impacts on the separate supply-side channels. For instance, corporate taxes have a much larger impact on investment compared to product market regulations. The impact of legal retirement age on the employment rate is

Figure 3. **Effects of improving structural policy areas: details**
by supply side components after 5 and 10 years

Predicted effects of typically observed reforms* in each policy area

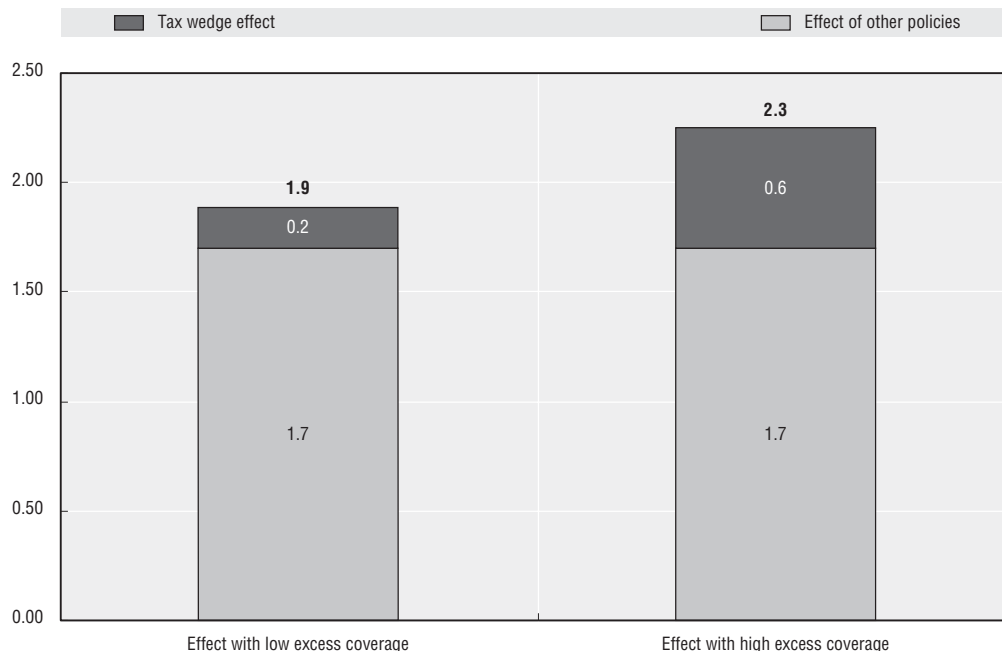


roughly 5 times smaller than those found for minimum wages, maternity leave weeks or ALMP spending.

Finally, there is no robustly significant relationship between employment protection legislation reform on the aggregate employment rate. Some policy areas (family benefits, pension age) can be assessed and included in the framework only by demographic groups. Hence as a benchmark in future quantification exercises, the effects found when aggregating across demographic groups could be used.

Figure 4. Employment effects of tax wedge reforms in a low vs. high excess coverage country

Predicted effects of typically observed reforms* in each policy area



Note: Typically observed reforms are measured as the average improvements in the policy indicators over all two year windows that show improvements in both periods (see Table 5, column 2). The low and high excess coverage country corresponds to the 25th and 75th percentile of the excess coverage distribution in 2011 (Poland and Greece, respectively).

Notes

1. For instance, policy effects on labour market outcomes are analysed for specific policies (independently of other possibly relevant policies) on substantially different country samples. The effects of unemployment benefits, tax wedge and active labour market policies are taken from Bassanini and Duval (2006). The effect of childcare spending reported in Jaumotte (2003) is used. The first study covers 20 countries and the period 1982 to 2003, and uses OLS and SUR for estimation purposes. The second paper looks at 17 countries and 1985 to 1999. It employs 2-stage least squares to estimate policy effects. For a more detailed comparison, see Tables A1.1 and A1.2.
2. Macroeconomic estimates are well- suited to quantify the overall macroeconomic effects of policies. But policy effects may be better identified in sector- or firm-level datasets. However, computing overall macroeconomic policy effects using sector- and firm-level estimates require specific assumptions: most of them obtain partial equilibrium results and results that are differential in nature (identifying from differences between two groups) and not aggregate economy-wide effects.
3. Appendix 1 in Égert and Gal (2016) provides a detailed comparison of the old and new frameworks.
4. A more detailed comparison can be found in the working paper version of this paper (Égert and Gal, 2016).
5. It should be noted that a number of sensitivity checks are carried out in Égert (2016a, 2016b) and Gal and Theising (2015). They confirm that the results summarised hereafter are fairly robust to alternative specifications regarding time and country coverage, different controls and estimation methods.
6. The MFP and ETCR series have common trends captured by year fixed effects. These trends are strongly correlated with each other. The correlation between the time fixed effects of MFP and the demeaned overall ETCR series is 0.72 (the series are also demeaned in the regressions including country fixed effects). When we compare the time fixed effects in the MFP and ETCR series, the correlation is 0.77. This is not surprising as time fixed effects explain about 89% of the variation of the demeaned overall ETCR series. When decomposing the overall ETCR indicator into i) barriers to

entry and ii) public ownership, public ownership survives the inclusion of year fixed effects. This variable could potentially be used for the purpose of quantification (at the expense of covering fewer policy areas).

7. Three labour market regulation indicators are used: EPL for permanent contracts. ALMP spending and the unemployment benefit replacement ratio.
8. The finding that EPL is statistically not significant stands somewhat in contrast with the literature using sector- and firm-level data relying on difference-in-difference approaches. For instance, Bassanini et al. (2009) finds for a set of 16 OECD countries from 1982 to 2003 that country-level EPL is associated with lower MFP growth in sectors with higher layoff rates. Rincon-Aznar and Siebert (2012) show the negative relation to hold for manufacturing sectors but not for the services sectors (Table 5, on a somewhat different sample: 11 OECD countries from 1984 to 2007). Using firm-level data for the USA, Autor et al. (2007) report mixed evidence on the negative relation between employment protection (good faith exception) and the level of MFP: the coefficient estimates are negative but only one coefficient in two is precisely estimated (Table 8, panel 1, first row). Dougherty et al. (2011) show that state-level employment regulation lowers MFP levels in Indian firms operating in more-labour intensive industries. However, our results cannot be fully compared with these results. A difference-in-difference approach identifies the extent to which employment regulation affects differently sectors or industries with different characteristics. But it does not inform us whether employment protection is good or bad, neither the size of the impact for overall MFP. It would require additional assumptions regarding the impacts on the presumably less affected industries or firm groups (control group). In the macro estimates, the coefficient on EPL is also negative and statistically significant if it is identified both along the within and between dimensions, i.e. when EPL's cross country variation is also taken into account in regressions not including country fixed effects (Egert, 2016a).
9. This negative relationship is robust to alternative country coverages (for narrower samples composed of more developed OECD countries) and to the definition of the capital stock (real capital stock, capital stock/output, capital stock/workers). This result needs qualification. The effect of EPL on investment is not clear-cut in the existing body of research. The literature reports no evidence that labour market regulation has any impact of investment at the macroeconomic level and for several OECD countries (Kerdrain et al, 2010). There is mixed evidence on the relation between capital stock and labour market regulation at the firm level. There is evidence for European firms that more stringent EPL reduces investment per worker and capital per worker (Cingano et al., 2010). By contrast, for US firms, research suggests higher firing costs (wrongful discharge exceptions) are linked to higher capital stock and capital-to-labour ratios. But the effect becomes negative when state-specific trends are used. A rise in capital may be related to a correction of an earlier downturn and that the introduction of more stringent firing regulations followed a rise of the capital-to-labour ratio (Autor et al. 2007). For Italian firms, estimation results show that the introduction of unjust-dismissal costs raises the capital-to-labour ratio in firms with less than 15 employees, compared to larger firms (Cingano et al., 2015).
10. The statistically significant relation between these policies and the overall employment rate (Table A4.2) are robust to various sensitivity checks, including different estimation methodologies, control variables and a time period covering only the pre-financial crisis period (Table 10-12 and D1-D3 in Gal and Theising, 2015). Nevertheless, jackknifing the sample, i.e. dropping one country at a time from the sample, shows some sensitivity to the country coverage.
11. The magnitude of the estimated impact (-0.3 for the youth) seems consistent with studies showing elasticities of -0.1 to -0.2 (see recent survey by Neumark, 2015 and OECD, 2015b). This is because we use the Kaitz index (median to minimum wage), which in our sample averages at 50% (Table 5, Gal and Theising, 2015). Hence 1 percentage point (pp) increase in it translates into a 2 pp increase, on average, for the minimum wage level. Therefore coefficients obtained when using the level of the minimum wage should be multiplied by two to make them comparable with our coefficients.
12. Our coefficient estimates for prime-age women are larger than those reported in the literature using similar datasets (Addison and Ozturk, 2012). The differences may be due to different model specification and data coverage. Therefore, care should be taken when using these estimates for quantification.
13. MFP used for the estimations is calculated as follows:

$$\ln(MFP_t) = \ln(Y_t) / (1 - \alpha) - \ln(L_t) - \ln(CLF_t) - \alpha / (1 - \alpha) \ln(K_t),$$

where CLF adjusts labour input for people working but not living in the country or those working abroad for domestic companies $\alpha = 0.33$, the standard value in the literature and fixed across countries and over time for ensuring comparability in a simple manner.

14. Considering capital intensity, when r is the real interest rate, the capital-output ratio in equilibrium is given by $\frac{K}{Y} = \frac{\alpha}{r}$. In a more elaborate setting, the real interest rate can be replaced by the user cost of capital, which includes the relative price of investment goods and corporate taxes as further determinants. In addition, excessive regulation can introduce frictions that suppress capital accumulation – a mechanism that can be captured by product and labour market regulation indicators.
As for the employment rate, both labour supply and labour demand determinants enter as policy channels in equilibrium (hence no need to include wages or productivity on top of them).
15. $\ln(1 - 0.9) / \ln(1 - 0.05) \approx 45$ years. The half-life, i.e. the time over which half of the convergence to the new long-run equilibrium happens can be calculated as $\ln(1 - 0.5) / \ln(1 - 0.05) \approx 13.5$ years.
16. In addition to the baseline results based on demographic groups, the predicted impacts for the employment rate and the core set of policies are shown for two alternative approaches: i) results obtained for the overall employment rate; and ii) results obtained for skill groups.

References

- Adalet-McGowan, M. and D. Andrews (2015), “Labour Market Mismatch and Labour Productivity: Evidence from PIAAC Data”, *OECD Economics Department Working Papers* No. 1209.
- Addison, J.T. and O.D. Ozturk (2012), “Minimum Wages, Labor Market Institutions and Female Employment: A Cross-Country Analysis”, *Industrial and Labor Relations Review*, 65(4), 779-809.
- Autor, D.H., W.R. Kerr and A.D. Kugler (2007), “Does employment protection reduce productivity? Evidence from US States”, *Economic Journal*, 117(521), 189-217.
- Barnes, S., R. Bouis, P. Briard, S. Dougherty and M. Eris (2013), “The GDP impact of reform: A simple simulation framework”, *OECD Economics Department Working Papers* No. 834.
- Bassanini, A. and R. Duval (2006), “Employment Patterns in OECD countries: Reassessing the Role of Policies and Institutions”, *OECD Social, Employment and Migration Working Papers* No. 35.
- Bassanini, A., L. Nunziata and D. Venn (2009), “Job Protection Legislation and Productivity Growth in OECD Countries”, *Economic Policy*, Vol. 24(58), pp. 249-402.
- Bouis, R. and R. Duval (2011), “Raising potential growth after the crisis”, *OECD Economics Department Working Papers*, No. 835.
- Bouis R., R. Duval and F. Murtin (2011), “The policy and institutional drivers of economic growth across OECD and non-OECD economies: New evidence from growth regressions”, *OECD Economics Department Working Papers*, No. 843.
- Cette, G., J. Lopez and J. Mairesse (2013), “Upstream product market regulations, ICT, R&D and productivity”, *NBER Working Papers*, No. 19488.
- Cette, G., J. Lopez and J. Mairesse (2013), “Product and labour market regulations, production prices, wages and productivity”, *NBER Working Papers*, No. 20563.
- Cingano, F., M. Leonardi, J. Messina and G. Pica (2010), “The effect of employment protection legislation and financial market imperfections on investment: Evidence from a firm-level panel of EU countries”, *Economic Policy*, 25(61), 117-63.
- Cingano, F., M. Leonardi, J. Messina and G. Pica (2015), “Employment protection legislation, capital investment and access to credit: Evidence from Italy”, *Economic Journal*, <http://10.1111/ecoj.12212>.
- Dougherty, S., V.C. Frisancho Robles and K. Krishna (2011), “Employment protection legislation and plant-level productivity in India”, *NBER Working Paper Series*, No. 17693.
- Égert, B. (2016a), “Regulation, institutions and productivity: new macroeconomic evidence from OECD countries”, *OECD Economics Department Working Paper* (forthcoming).
- Égert, B. (2016b), “Regulation, institutions and aggregate investment: New evidence from OECD countries”, *OECD Economics Department Working Paper* (forthcoming).
- Égert, B. and P. Gal (2016), “The quantification of structural reforms: Introducing country-specific policy effects”, *OECD Economics Department Working Paper*, No. 1354.
- Engle, R.F. and C.W.J. Granger (1987), “Co-integration and error correction: representation, estimation and testing”, *Econometrica*, Vol. 55(2), pp. 251-276.

- Gal, P. and A. Theising (2015), "The macroeconomic impact of policies on labor market outcomes in OECD countries: A reassessment", *OECD Economics Department Working Paper* No. 1271.
- Hendry, D.F. (1986), "Econometric modelling with cointegrated variables: An overview", *Oxford Bulletin of Economics and Statistics*, Vol. 48(3), pp. 201-212.
- Jaumotte, F. (2003), "Labour Force Participation of Women: Empirical Evidence on the Role of Policy and Other Determinants in OECD Countries", *OECD Economic Studies*, No. 37, pp. 51-108.
- Johansson, A., Y. Guillemette, F. Murtin, D. Turner, G. Nicoletti, C. de la Maisonneuve, P. Bagnoli, G. Bousquet and F. Spinelli (2013), "Long-term growth scenarios", *OECD Economics Department Working Papers*, No. 1000, OECD Publishing.
- Neumark, D. (2015) "The Effects of Minimum Wages on Employment", *Federal Reserve Bank of San Francisco Economic Letter*, 2015-37.
- Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica*, Vol. 49(6), pp. 1417-1426.
- OECD (2015a), "Quantification of structural reforms", *ECO/CPE/WP1/2015(5)*.
- OECD (2015b), *OECD Employment Outlook*, OECD Publishing, Paris.
- Phillips, P.C.B. and B.E. Hansen (1990), "Statistical inference in instrumental variables regression with I(1) processes", *The Review of Economic Studies*, Vol. 57(1), 99-125.
- Rincon-Aznar, A. and W.S. Siebert (2012), "Employment production, productivity, wages and jobs in Europe", *INDICSER Discussion Paper* No. 36.
- De Serres, A., A. Hijzen and F. Murtin (2014), "Unemployment and the Coverage Extension of Collective Wage Agreements", *European Economic Review*, Vol. 71, pp. 52-66.
- Stock, J. and M.W. Watson (1993), "A simple estimator of co-integrating vectors in higher order integrated systems", *Econometrica*, Vol. 61(4), pp. 783-820.

Administrative fragmentation and economic performance of OECD TL2 regions

by
David Bartolini*

The present work investigates the relationship between administrative fragmentation and regional per capita GDP growth rate, using a panel of OECD TL2 regions in the period 1996-2011. According to the fiscal decentralisation literature, fragmentation should enhance growth as local governments can implement policies that better match citizens' needs, thus providing services and public goods in a more efficient way. The presence of many local governments, however, may result in overlapping functions, (dis)economies of scale, and uncoordinated policies.

The empirical analysis shows that regions with high municipal fragmentation display lower growth, thus supporting the view that costs prevail on benefits. The introduction of regional territorial characteristics – namely, the share of population living in rural areas – provides a different picture. The negative impact of fragmentation decreases with the share of population living in rural areas. Indeed, in extremely “rural” regions the effect turns mildly positive. This is because the costs and benefits of fragmentation have a different weight in urban and rural regions. The key insight is the different distribution of the population over the territory: more concentrated in urban than in rural regions. This implies that, for a given level of municipal fragmentation overlapping of function is more severe in urban regions (where people are likely to commute over municipal boundaries) than in rural area. In the same vein, for the same level of municipal fragmentation access to the local government is more difficult in rural areas (where people is sparsely located within municipal boundaries) than in urban areas.

The policy implications of the analysis are twofold. Firstly, reducing municipal fragmentation may have a heterogeneous impact within a country, thus raising concern for one-size-fits-all policies of municipal agglomeration. For instance, the principle guiding municipal amalgamation should not be the average municipal size at the country level, but it should be weighted for the rural/urban characteristics of each region. Secondly, the analysis suggests that the increasing size of metropolitan areas should be accompanied by a consistent amalgamation of local administrations, otherwise representing an obstacle to the full realisation of agglomeration economies.

JEL classification: R11, R50

Keywords: Regional development, administrative fragmentation, local governments

* I would like to thank for very useful comments, Luiz de Mello, Joaquim Oliveira Martins, Enrique Garcilazo, Raffaele Trapasso, Alexander Lembcke, Abel Schumann, Isabelle Chatry, as well as two anonymous referees which provided detailed written comments. The ideas and opinions expressed in this paper are sole responsibility of the author and do not reflect the position of the OECD or any of its member countries.

Introduction

Good governance is an important ingredient of countries economic development. Along with the “traditional” factors of production – labour, capital and technology – governance has received increasing attention as driver of economic growth (Williamson, 1975; North, 1990; Acemoglu and Robinson, 2006). An important component of governance is a well-functioning public administration which provides policies and regulations that are essential for a sound business environment. In this respect, the territorial structure of the public administration – how many local governments? what functions? – is particularly relevant for regional economic growth (OECD/KIPF, 2012). The present work deals with these questions by considering the partition of regional population into local governments. To this end an indicator of municipal fragmentation is constructed for 250 OECD TL2 regions,¹ and its relationship with the per capita regional GDP growth is tested empirically.

Administrative fragmentation is measured by the number of municipalities per 100 000 inhabitants (OECD, 2016). This indicator captures the size of local governments with respect to the population served, thus neutralising any bias introduced by regional size.² Municipal fragmentation is evaluated on the basis of its impact on the annual growth rate of regional per capita GDP, in the period 1996-2011.

The relevance of administrative fragmentation for regional economic growth rests on the trade-off between proximity to people’s needs and externalities, as defined by the theorem of fiscal decentralisation (Oates, 1999). Accordingly, higher fragmentation would result in municipalities that can better tailor policies to citizens’ preferences, but also suffer from the presence of externalities due to the risk of overlapping policies, (dis)economies of scale, and transaction costs for coordinating policies.

Previous empirical studies have mainly focused on metropolitan areas, investigating the role of fragmentation in the provision of public goods for the whole metropolitan area (for instance, transport and housing policies). Indeed, one of the main problems with municipal fragmentation is the need to co-ordinate policies that have an impact beyond the local government’s boundaries (policy spillovers). Nelson and Foster (1999), for instance, find that municipal fragmentation reduces income per capita in US large metropolitan areas; OECD research provide evidence of a negative relationship between municipal fragmentation and productivity of large metropolitan areas (OECD, 2015).

The impact of municipal fragmentation, however, may differ according to the type of territory considered. Stansel (2005) finds a positive relationship between administrative fragmentation and per capita income growth when including metropolitan areas of all sizes in the analysis. This is probably due to the different weight that the costs of fragmentation have in less densely populated areas. In particular, when the population is scattered over a vast territory there is less commuting over municipal boundaries, thus reducing the problem of policy spillovers and the need of (costly) co-operation. Therefore, the benefits of fragmentation may overcome the costs in rural areas, while for the same level of fragmentation the contrary may happen in densely populated regions.

The distinction between urban and rural regions is the main contribution of this paper. To this end, an indicator of population concentration within the regional boundaries is constructed, by taking into account the share of regional population living in rural areas; it is referred to as the rural index.³ This indicator is a continuous indicator ranging from 0 (no rural population) to 100 (only rural population), which allows to test the hypothesis that the negative impact of fragmentation on economic growth increases with the degree of urbanisation (decreases with the rural index).

The empirical investigation is conducted on a cross-section of 250 TL2 OECD regions, covering 23 OECD countries. Results show that municipal fragmentation has a negative impact on the annual growth rate of real per-capita GDP, averaged over the period 1996-2011. GDP growth is also negatively affected by the share of population living in rural areas, that is, rural regions tend to grow less than urban ones. However, the interaction between fragmentation and the rural index confirms the hypothesis of a negative impact in urban regions. In particular, the lower the rural index the higher the negative impact of fragmentation.

The economic intuition for this result rests on the different weight that costs and benefits of fragmentation have in urban and rural regions due to the different distribution of the population over the territory: less concentrated in rural than in urban regions. This implies that, for a given level of municipal fragmentation, overlapping of functions is more severe in urban areas, where people are more likely to commute over municipal boundaries, than in rural areas. By contrast, the benefit of fragmentation in terms of better match with citizens' preferences and better (and easier) access to local politicians tends to be more beneficial for people living in rural areas – it is also important in terms of “voice” of rural community towards the higher levels of government (Hirschman, 1970). To summarise, the empirical analysis confirms that a given level of municipal fragmentation has a different impact on the economic growth of urban and rural regions. The drawbacks of fragmentation that hinder economic growth in urban regions are counterbalanced by the benefit of proximity to citizens in rural regions.

The policy implications of the analysis are twofold. Firstly, reducing municipal fragmentation may have a heterogeneous impact within a country, thus raising concern for one-size-fits-all policies of municipal agglomeration. For instance, the principle guiding municipal amalgamation should not be the average municipal size at the country level, but it should be weighted for the rural/urban characteristics of each region. Secondly, the analysis suggests that the increasing size of metropolitan areas should be accompanied by a consistent amalgamation of local administrations, otherwise representing an obstacle to the full realisation of agglomeration economies.

The rest of the paper is organised as follows. The next section provides a review of the literature on municipal fragmentation and economic growth. Then, a brief discussion of the measures of fragmentation and a snapshot of current administrative structures in OECD countries are provided. The paper continues with the description of the data and the econometric model. Empirical results are then presented, along with a discussion of policy implications. A final section concludes with a summary of the main results and suggestions for further work.

Related Literature

A consequence of municipal fragmentation is the increased number of local government in a given territory which may result in increasing competition. It is possible to gather most of the literature according to the value attached to competition. Two competing theories emerge: a polycentric, and a centrist model (Nestor and Foster, 1999). The polycentric view tends to stress the positive impact of competition between local governments for the efficient provision of public services, and the possibility to benchmark the behaviour of local administrators with their neighbours – namely, electoral yardstick competition (Bartolini and Santolini, 2012); furthermore, a fragmented territorial administration offers a wider choice of service and tax bundles to firms and residents (Tiebout, 1956). The public choice literature can also be considered polycentric, given the importance of “taming” the Leviathan through the proliferation of competing administrative bodies (Brennan and Buchanan, 1980). By contrast, the centrist view stresses the drawbacks of competition in terms of “race to the bottom”, the emergence of dis-economies of scale in the provision of local public goods, and the presence of policy spillovers that would require costly co-operation among local governments (Hamilton et al., 2004). To sum up, the actual impact of administrative fragmentation on regional economic growth depends on the size of these contrasting effects, and thus remains a question to be answered by the empirical analysis.

The empirical literature has mainly focused on US metropolitan areas.⁴ Some works back the centrist view, providing evidence of a drag on economic growth. For instance, both Nelson and Foster (1999) and Paytas (2001) show that administrative fragmentation reduces income per capita in US metropolitan areas. Ahrend et al. (2014) show that municipal fragmentation reduces the wage premium of large metropolitan areas. There are also works backing the polycentric view. Randall and Gronberg (1990) show that fragmentation increases local spending to attract firms and people, thus contributing to economic growth. Stansel (2005) covers 314 US metropolitan areas, focusing on the number of municipalities and counties per 100 000 residents, finding a positive impact of administrative fragmentation and economic growth.

The apparent contrast between these set of empirical works can partly be attributed to differences in the indicator of fragmentation and the territory considered (metropolitan areas, counties, regions, etc.). As regards the type of indicator, Grassmueck and Shield (2010) criticise the use of the number of municipalities as it overlooks differences in administrative powers of municipalities. To overcome this shortcoming they consider an index which takes into account the spending authority of local administration – Hirschman Herfindahl concentration index. Interestingly, this indicator is positively related to economic growth in metropolitan areas, while the number of municipalities negatively affects economic growth. This highlights the importance of the way in which fragmentation is measured. The basic assumption is that local governments with large spending powers have greater influence to affect the economic system than local administrations with less power. The other important caveat is the geographical scope of the analysis. Most of the literature focuses on Metropolitan areas (mainly in the USA). Stansel (2005) notes that the discrepancy with Nelson and Foster (1999) could depend on the sample used. While Nelson and Foster (1999) focus on large metropolitan areas, Stansel (2005) considers metropolitan areas of all sizes, thus suggesting that the relationship between fragmentation and growth is negative in large metro areas, while turning positive in smaller ones.

This brief and selected overview shows that the literature fails to take into account territorial characteristics that might be determinant for the actual impact of fragmentation. The present work takes a step in this direction, by combining measures of administrative fragmentation (absolute and population weighted) with territorial characteristics (rural index), which capture the distribution of the population over the whole territory of each country (OECD 2009b; OECD 2012).⁵

Finally, it is worth mentioning that municipalities can also differ in terms of their “capacity” to devise and implement policies. One of the potential problems of fragmentation is the reduction in the size of municipalities, thus increasing the chances of capacity gaps: limited financial capacity, limited range of skills of public administrators, etc. OECD research has focused on the problems of governance gaps, implicitly suggesting that administrative fragmentation may hinder economic growth if the gaps are not adequately filled (Charbit, 2011; Charbit and Michalun, 2009).

Measuring fragmentation and territorial characteristics

This section provides a discussion of the two crucial elements of the analysis: (1) the measure of municipal fragmentation; and (2) the distribution of the regional population over the territory.

Administrative fragmentation has been studied mainly in relation to US metropolitan areas. Three main measures of fragmentation are proposed: the absolute number of local councils in a given territory; the number of local councils weighted for the population; and indexes based on the functions and powers of local councils. These indicators differ in terms of information content and international comparability. The simplest way to account for administrative fragmentation is to consider the absolute number of municipalities (Cutler and Glaeser, 1997). This indicator, however, fails to take into account regional population differences, thus suffering from the fact that most populous regions tend to have more local governments than scarcely populated regions (Dolan, 1990).

The second type of indicators is weighted by population size. In the present work, the municipal fragmentation index consists of the number of municipalities within a region divided by the regional population, and multiplied by 100 000. Given the number of local councils the indicator decreases with the population, implicitly assuming that the number of municipalities increases with population size. Thus accounting for the fact that bigger regions tend to have more municipalities. This indicator is also used in other OECD works (OECD, 2015; OECD, 2016).

Both the weighted and the population-weighted indicators, however, treat all local councils in the same way. They do not take into account the administrative powers and functions delegated to the local level of government. A solution would be to create an indicator which takes into account the degree of decentralisation in each country – e.g. a fiscal dispersion measure (Dolan, 1990). The basic assumption is that local governments with large spending powers have greater influence to affect the economic system than local administrations with less power. This indicator, however, requires a considerable amount of information about the functions actually allocated to the local level, and raises issues about data homogeneity across countries given differences in administrative and legal frameworks.

To summarise, although the analysis would benefit from the introduction of a more refined index of (functional) fragmentation, the focus on the absolute and relative measure

of fragmentation assures a higher degree of homogeneity of the dataset and allows a better international comparison.

In the present work, the urban/rural characteristics of OECD TL2 regions are captured by the rural index. This index is calculated starting from the classification of “local communities” in urban and rural according to the OECD regional typology framework, and then aggregating at the regional level in order to obtain the share of the population living in rural areas. Therefore, regions are not classified as rural or urban, but each region has its own degree of “rurality”, thus avoiding the problem of selecting similar regions into different categories. Recently the European Commission uses the concept of “population grid” in order to define territorial characteristics. The rural index could be improved by using this methodology which provides a smaller and more uniform unit of analysis.

This indicator is correlated to regional population density, but provides a finer measure of the distribution of population within regions. For instance, population density cannot distinguish between two regions with the same population size and surface areas, but where population is concentrated in one city or evenly spread over the territory. The rural index would be lower for the region in which the population is concentrated in few cities.

Finally, the choice of the OECD territorial level 2 might raise concerns about the representation of a vast territory by a single rural index. We acknowledge that TL2 regions may be constituted by both rural and urban areas, and in this respect TL3 regions present more homogeneous territories.⁶ The choice however fell on the TL2 level for two main reasons: firstly, the administrative border of TL3 regions risk to cut across metropolitan areas, which may span over municipalities belonging to different TL3 regions; secondly, the dataset on TL3 regions only includes European countries and Japan, thus reducing the possibility to generalise the results to all OECD countries.

Sub-national government structure in OECD countries

Most OECD countries present a multilevel government structure. Public administrative functions are shared among several territorial tiers (e.g., regional government, provincial government, municipal government). The number of governmental tiers within a country represent the vertical fragmentation of the public administration, which accounts for the number of government overlapping over given territory. At each government tier, however, there are more than one government body. The way in which the territory is divided at a given government tier represents the horizontal fragmentation of the public administration.

Table 1 provides an overview of both dimensions of administrative fragmentation in OECD countries. The number of municipalities (and their average size) varies a lot among OECD countries: from 67 municipalities in New Zealand, to more than 36 thousands in France. Also the vertical dimension varies from countries with only one tier like Slovenia, Estonia and Portugal, where only the municipal level is present, to countries with three sub-national tiers of government like France, Italy, and the USA.

All countries have a municipal level of government,⁷ thus justifying the use of this territorial level in our analysis. Considering horizontal fragmentation at an intermediate level of government would sensibly reduce the sample size, and would reduce international comparability, as the administrative functions and the political role of the intermediate level varies across countries much more than the municipal level does.

Table 1 provides only an indication of absolute municipal fragmentation, which appears to be closely linked to the size of countries. When considering the number of

Table 1. **Subnational government structure in OECD countries, 2012**

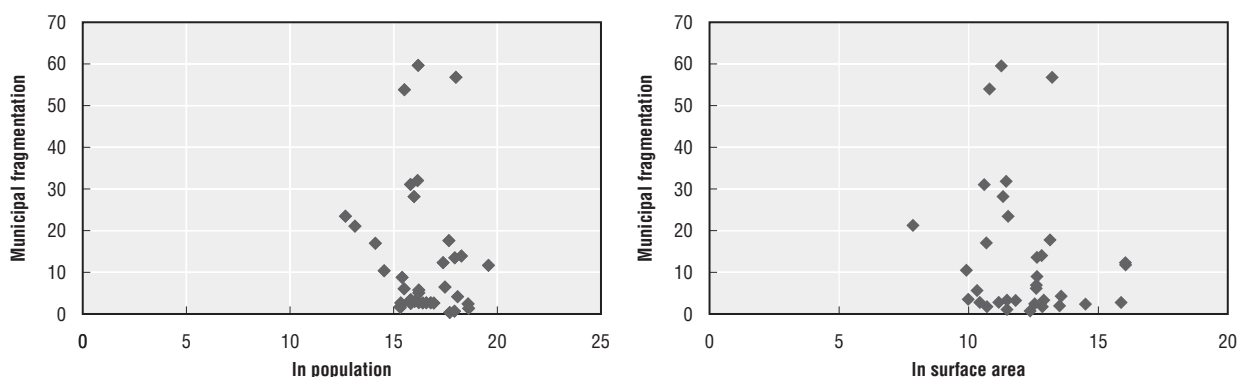
2012-13	Municipal level	TL3 (intermediate or State/Regional level)	TL2 (State/Regional level)
Federations and quasi-federations			
Australia	565		8
Austria	2 354		9
Belgium	589	10	6
Canada	4 147		13
Germany	11 327	295	16
Mexico	2 457		32
Spain	8 116	52	17
Switzerland	2 408	26	
United States	35 879	3 031	50
Unitary States			
Chile	345		15
Czech Republic	6 253	14	
Denmark	98		5
Estonia	226		
Finland	320		
France	36 700	101	27
Greece	325	13	
Hungary	3 177	19	
Iceland	74		
Ireland	114		
Israel	254		
Italy	8 092	110	20
Japan	1 719	47	
Korea	227	17	
Luxembourg	106		
Netherlands	408	12	
New Zealand	67	11	
Norway	428	18	
Poland	2 479	380	16
Portugal	308		2
Slovak Republic	2 927	8	
Slovenia	211		
Sweden	290	20	
Turkey	2 950	81	
United Kingdom	406	28	3

Source: OECD 2013, "Sub-national governments in OECD countries: Key data" (brochure), OECD, Paris.

municipalities weighted for country's population however a different picture emerges. Figure 1 plots municipal fragmentation, constructed as the number of municipalities per 100 000 inhabitants in each country,⁸ against countries' population and surface area – both expressed in logarithms. It appears that the larger municipal fragmentation is present in countries of average population and surface area. The three countries with the highest degree of fragmentation are the Slovak Republic (54), France (56.7), and the Czech Republic (59.5). Although France and the Czech Republic differ in terms of population size and surface area, they are both at the top of the distribution of municipal fragmentation.

The econometric model and the data

Administrative fragmentation affects the "closeness" of the public administration to citizens. This can be an advantage as local policy makers can better tailor policies to

Figure 1. **Municipal fragmentation does not depend on the size of countries**

Note: The sample consists of OECD member countries (see Table 1 for a complete list).

Source: Elaboration on OECD (2013) Subnational governments in OECD countries: Key data.

citizens' needs, thus contributing to regional growth. At the same time the proliferation of local administrations determine a cost in terms of reduced economies of scale, overlapping of functions, and co-ordination problems. For instance, the construction of transport infrastructure may be delayed by the need of consulting several stakeholders, or it may result in overlapping of projects. Therefore the impact of fragmentation on economic growth depends on which of the two effects prevail.

The econometric model captures this trade-off by including an interaction term between the level of fragmentation and the rural index, which rests on the hypothesis that the costs of fragmentation are higher than the benefits in urban areas while the contrary applies in rural areas. If this is the case, the coefficient associated with the interaction term should have a positive sign.

The empirical analysis is conducted on a cross-section of 250 TL2 regions, representing 23 OECD countries, for the period 1996-2011. Data are obtained from the OECD Regional Statistics Database. The econometric equation stems from a reduced-form model which directly accounts for the effect of fragmentation on economic growth. The empirical analysis is conducted considering a linear econometric model where the institutional variables are interacted with the rural index. In particular, we expect a negative sign of the coefficient associated with municipal fragmentation, β_2 , and a positive sign of the coefficient associated with the interaction term, β_3 .

The empirical analysis is conducted using the following econometric specification,

$$g_i = \text{cons} + \beta_1 \text{rur}_i + \beta_2 \text{frag}_i + \beta_3 \text{rurfrag}_i + \beta_4 X_i + \varepsilon_i$$

where g_i is the annual average growth rate of *per capita* regional gross product in region i , cons is a constant term, rur is the rural index for region i , and frag accounts for municipal fragmentation in region i . Finally, X_i is a vector of control variables, including population density, the initial level of regional *per capita* GDP, an indicator of patents registration, education attainment of working age population, and a dummy variable for the political system (unitary or federal).

The growth of the regional GDP per capita is calculated as the average annual growth of *per capita* GDP at constant PPP for the years 1996-2011.⁹ The rural index is constructed taking the ratio of the population living in *rural areas* with respect to the total population of the TL2 region. Rural areas are defined as settlements/municipalities where the density of

the population is lower than 150 inhabitants per square kilometre (OECD 2011). The average value of the rural index in our sample is 0.43, indicating that in the “average” region 43% of the population lives in rural areas. The indicator of municipal fragmentation is constructed as the ratio between the number of municipalities and the population in each TL2 region. Data about number of municipalities are obtained from countries’ Census statistics.¹⁰

Several economic and demographic indicators that are important determinant of economic performance are taken into account as control variables. Among them of particular importance are human capital and the political system.¹¹ Human capital is one of the main determinants of labour productivity, contributing to GDP growth (Barro and Sala-i-Martin, 2004). The attainment level of education of the workforce is used as proxy for human capital; this indicator considers the share of the labour force with primary, secondary and tertiary education, respectively.¹² A dummy variable accounting for the political system – Federal or Unitary – is considered, as local governments in federal countries may have more autonomy than in unitary ones. Also included as control variables are innovation intensity, which should positively affect economic growth (although the impact may occur with some lag), the initial level of regional GDP per capita, which accounts for economic convergence, and population density, that should increase the regional growth rate because of agglomeration benefits in metropolitan areas.

The model is estimated using an OLS (ordinary least square) estimator and robust standard errors. The choice of using a cross-section rather than a panel is driven by the lack of time series data on municipal fragmentation. However, many of the variables of interest are time invariant or rarely changing through time, so that not much information could be added by a panel. This is the case for the fragmentation index and the rural index. The results of the empirical analysis are driven by the comparison between regions rather than the evolution of fragmentation within each region. The analysis is conducted also with country dummies to capture regional differences that are given by countries’ specific characteristics. For instance, two regions in different countries may exhibit a different growth path because of country specific macroeconomic policies.

The correlation matrix between the explanatory variables shows that there might be a potential problem of multicollinearity among the education variables. In particular, Table 2 shows a strong correlation between primary and secondary education (0.85). For this reason, the variable referring to secondary education is excluded from the estimated equation.

Table 2. **Correlation matrix**

	Rural index	Fragmentation	Federal	Population density	g1996	Patents	Primary education	Secondary education	Tertiary education
Rural index	1								
Fragmentation	0.1527	1							
Federal	0.1705	-0.0576	1						
Population density	-0.4351	-0.1237	0.0213	1					
g1996	-0.1118	-0.1085	0.4828	0.2956	1				
Patents	-0.1248	-0.1943	0.3091	0.0357	0.4718	1			
Primary education	-0.185	0.0519	-0.345	0.0166	-0.1272	-0.2655	1		
Secondary education	0.2885	0.0227	0.2135	-0.1487	-0.0877	0.0805	-0.8473	1	
Tertiary education	-0.1891	-0.1407	0.2185	0.2341	0.3851	0.309	-0.2079	-0.3423	1

Interestingly, there is not much correlation between the municipal fragmentation variable and the variable representing population density (-0.12). As expected there is some correlation between the rural index and the indicator of population density (-0.43), but it is not very large, thus we run the analysis both with and without controlling for population density. Furthermore, post estimation statistics for detecting multicollinearity within the regressors provide a negative result; the variance inflation factor is below the critical level for all the regressors used in the analysis (see appendix A3).

Given that the analysis is a cross-section, it is hard to control for endogeneity that might arise from region-specific effects that are not included among the explanatory variables. In addition the impossibility to control for capital formation, i.e., a standard determinant of GDP growth, may lead to a problem of missing variables.

Another caveat of our approach is that regions tend to differ in terms of administrative frameworks. The subnational governance structure of Picardie differs substantially from California or Australia's South Wales. The impact of fragmentation can be influenced by the subnational administrative framework. To take into account this source of heterogeneity, countries fixed effect are considered in some of the specifications of the model, as the administrative framework is the same for regions within the same country.

Given these caveats the empirical analysis provides a first insight into the relationship between administrative fragmentation and population density, providing evidence of a strong correlation if not causation.

Estimation results

The first column of Table 3 presents the results of the estimation of a model without territorial characteristics, i.e., without the rural component. The coefficient associated with municipal fragmentation is negative but not statistically significant as the negative impact of fragmentation in urban regions is counterbalanced by the positive effect in rural regions.

The inclusion of the rural index confirms the hypothesis of a differentiated impact of fragmentation. The second column of Table 3 shows that the coefficients associated with municipal fragmentation and the interaction term are statistically significant at the 1% level. The sign of the coefficients is as expected, with the impact of fragmentation being negative but decreasing with the rural index (because of the positive coefficient of the interaction term). The negative and statistically significant coefficient associated with the rural index, confirms that rural region tend to grow less than urban ones. Similar results are obtained if we substitute the dependent variable with the value added (VA) per worker, a measure of labour productivity. This shows that labour productivity is likely to be the main channel through which administrative fragmentation impacts on regional GDP per capita growth.¹³

Control variables have the expected sign: the initial value of GDP is negative and significant supporting the idea of regional convergence; education attainment of the labour force has a positive and significant impact on regional performance; and population density has a positive impact, although it loses significance once the rural index is introduced.¹⁴

The empirical analysis is replicated with the introduction of country dummies to capture the influence of countries specific characteristics on regional growth. Column 3 shows that the main results are still valid, in particular the coefficient of the interaction term is of similar magnitude and still significant at the 1% level. The coefficients associated with the control variables lose statistical significance because the country dummy tend to

Table 3. **Estimation results**

Dependent variable: annual average growth of per capita GDP	(1) g_total (1996-2011)	(2) g_total (1996-2011)	(3) g_total (1996-2011)	(4) g_pre crisis (1996-2007)	(5) g_post crisis (2008-11)
Rural index		-0.00821*** (0.00286)	-0.00680** (0.00270)	-0.00669** (0.00323)	-0.00710 (0.00453)
Municipal fragmentation	-1.39e-05 (2.24e-05)	-0.000112*** (3.73e-05)	-4.34e-05* (2.23e-05)	-2.65e-05 (2.89e-05)	-8.99e-05** (3.86e-05)
Fragmentation squared					
Interaction: rural fragmentation		0.000181*** (4.82e-05)	0.000143*** (2.53e-05)	7.74e-05** (3.33e-05)	0.000322*** (4.94e-05)
Federal constitution	0.00165 (0.00213)	0.00177 (0.00215)	0.00379 (0.00614)	-0.00487 (0.00698)	0.0276*** (0.0101)
g1996	-5.71e-07*** (1.88e-07)	-5.68e-07*** (1.85e-07)	-5.58e-09 (7.40e-08)	-3.66e-08 (9.47e-08)	7.97e-08 (1.12e-07)
Patent intensity	-8.49e-06 (1.01e-05)	-1.06e-05 (9.90e-06)	-9.29e-06* (5.37e-06)	-7.47e-06 (6.11e-06)	-1.43e-05 (9.03e-06)
Primary education	-0.000263*** (5.84e-05)	-0.000279*** (5.81e-05)	2.45e-05 (0.000158)	2.01e-05 (0.000167)	3.67e-05 (0.000238)
Tertiary education	0.000235*** (8.34e-05)	0.000195** (8.50e-05)	0.000230 (0.000221)	0.000232 (0.000262)	0.000224 (0.000300)
Population density	1.89e-06* (1.07e-06)	1.17e-06 (1.12e-06)	-3.91e-07 (8.60e-07)	-1.66e-07 (9.83e-07)	-1.01e-06 (9.22e-07)
Constant	0.0319*** (0.00368)	0.0374*** (0.00402)	0.0129 (0.0143)	0.0240 (0.0161)	-0.0175 (0.0203)
Country dummies	No	No	Yes	Yes	Yes
Observations	250	250	250	250	250
R-squared	0.301	0.332	0.738	0.624	0.807

Robust standard errors in parentheses.

*** p < 0.01, ** p < 0.05, * p < 0.1

capture most of the cross-regional variation. For instance, most of the variation in the level of education attainment is cross country not within the same country.

The estimated effect of municipal fragmentation on regional performance is therefore given by the direct effect captured by β_2 , and the coefficient of the interaction term β_3 . According to the estimation in column (3), where country dummies are taken into account, the marginal effect of fragmentation on GDP growth is given by the following

$$\frac{\partial g}{\partial frag} = -(0.000043) + (0.000143)rur$$

which is negative for small values of the rural index. If all the regional population lives in urban areas, the rural index is equal to zero and the marginal impact of fragmentation is -0.000043. The impact turns positive for regions in which more than 30% of the population lives in rural areas, which corresponds to 30% of the regions in the sample and 76% of the population. The magnitude of the marginal effect implies that a 10% reduction in the index of fragmentation produces a 0.8% increase in the annual per capita growth rate.

The empirical results are robust to the introduction of country dummies, to the measure of GDP in terms of labour productivity (GVA per worker), and to an alternative measure of fragmentation: the absolute number of municipalities. The latter is an important robustness check, because our analysis seems not to be affected by the volatility observed in previous works. For instance, Dolan (1990) reports that the impact of

fragmentation drastically changes when considering absolute fragmentation instead of per capita fragmentation. This is another indication of the importance of accounting for the distribution of the population within the territory in order to identify the impact of administrative fragmentation. Moreover, the robustness of our results with respect to changes in the indicator used to measure fragmentation can be taken as a further proof of the importance of including a territorial indicator (e.g., rural index, population density, population concentration, etc.) in the analysis. Finally, the inclusion among the dependent variables of the regional employment at place of work does not change the main results of the analysis, as well as the exclusion of patent intensity and population density. Estimation results are available upon request to the author.

The analysis suggests that the reduction of fragmentation is beneficial only in urban regions. This is due to the higher geographical concentration of population in urban areas, so that economic interactions are likely to spill over municipal administrative borders. For instance, workers living in urban areas are likely to commute across municipal borders, which may constitute an obstacle in case of an uncoordinated transportation policy. By contrast, in rural regions the population is spread over a vast territory, interactions are weaker and workers are less likely to cross administrative borders. In those areas, political representation and access to politicians is more a concern than the coordination of policies.

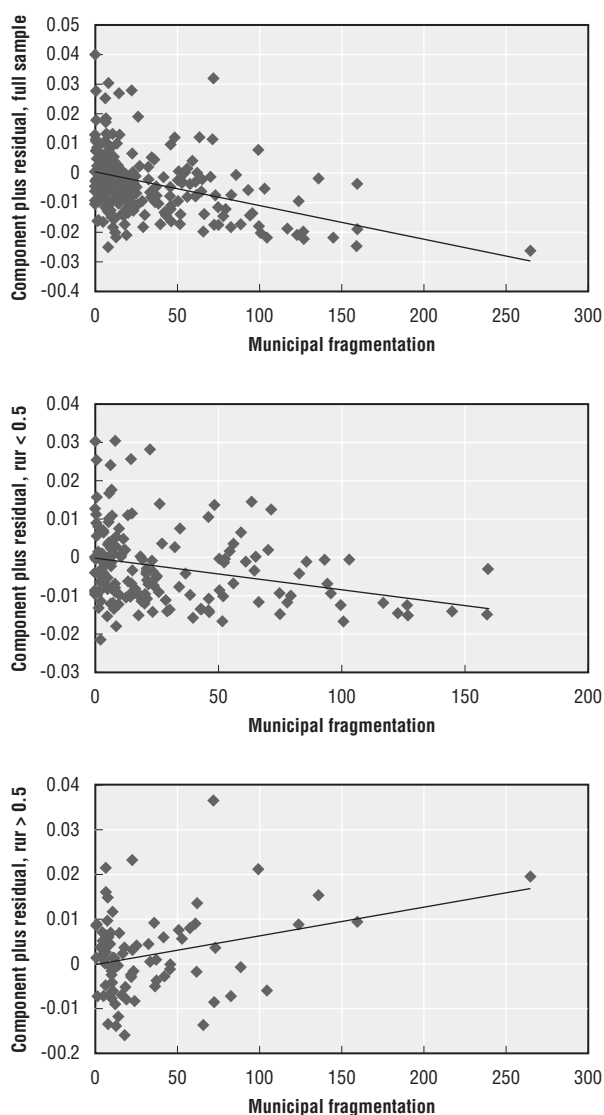
A graphical representation of the empirical results is obtained by running a post regression diagnosis. Figure 2 shows three different component-plus-residual plots of fragmentation for the whole sample, a sub-sample of urban and rural regions, respectively.

The three plots in Figure 2 show that the negative impact of municipal fragmentation is actually driven by the regions with a rural index lower than 0.5 (166 regions). The sub-sample of regions with a rural index greater than 0.5 (84 regions) shows actually a positive relationship, although the relationship is really driven by a few regions.

Figure 4 plots the average impact of horizontal fragmentation (computed using the estimated marginal impact) on TL2 regions with respect to their population density. As population density increases the impact becomes negative. It is worth noting that this relationship is not obtained if we consider only population size, meaning that the estimated impact of the administrative reform does not depend on the population size of a region but on the concentration (density) of the population in rural areas (rural index). In the appendix, the same relationship is computed and presented separately for each country in our sample.

The empirical analysis shows that in few countries a reduction of municipal fragmentation is likely to have a homogeneous effect, either positive or negative (Figure 4). In Belgium, and Luxembourg the positive effect is driven by the high density of population over the whole territory, which in geographical terms (especially for Luxembourg) is quite small. For the UK, the result is most probably driven by the partition of the territory in large TL2 statistical units, so that most of the population lives in the urban centres within these regions. In most of the countries, however, the effect is estimated to be positive in some regions and null or negative in others.

The empirical analysis covers a 15 year period, from 1996 to 2011. The most relevant event occurring in this period is the financial crisis which hit most advanced economies in 2007 and the subsequent “great recession”. In order to account for this change in the economic environment, the analysis is replicated on two sub-periods: the pre-crisis period, from 1996 to 2007; and the post-crisis period 2008-11. The results of the analysis performed

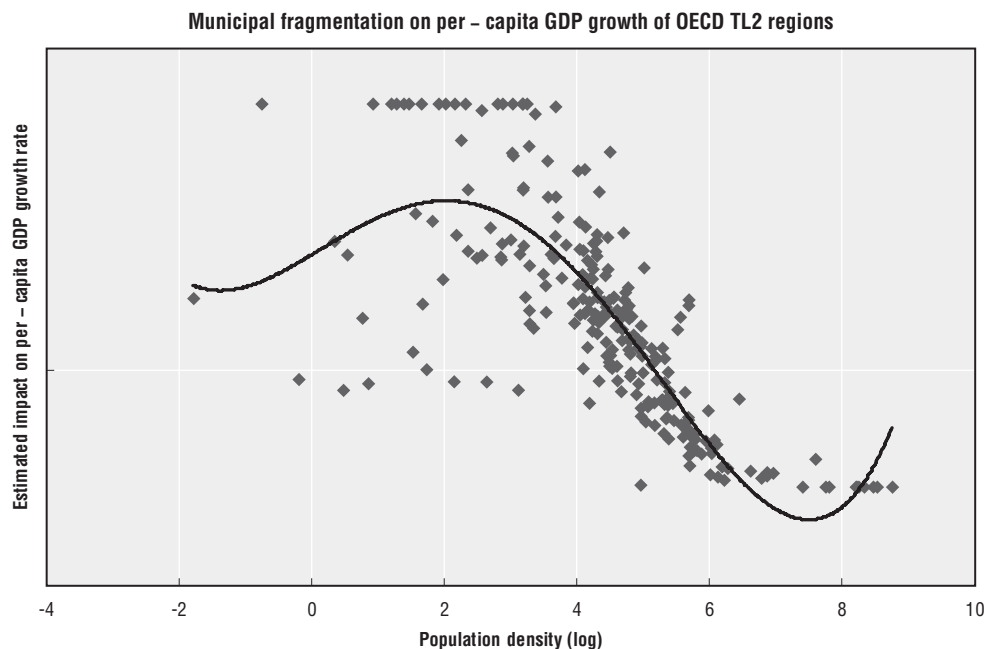
Figure 2. **Impact of fragmentation on per capita GDP – post regression diagnostic**

Source: Author's calculations.

on these two sub-samples are consistent with the analysis conducted over the whole period, the rural index and the indicator of administrative fragmentation display a negative impact on GDP growth, and the interaction term is positive. The difference is on the magnitude of the estimated coefficient and its statistical significance. It appears that the negative impact of the rural index is statistically significant only in the pre-crisis period, while the negative impact of fragmentation is statistically significant only in the post-crisis period. Interestingly, fragmentation seems to emerge as a significant problem only during the crisis. The large magnitude and statistical significance of the interaction term in the post-crisis period reinforces the message that municipal fragmentation is a problem in mainly urban regions and seems to be amplified in periods of economic recession or stagnation.

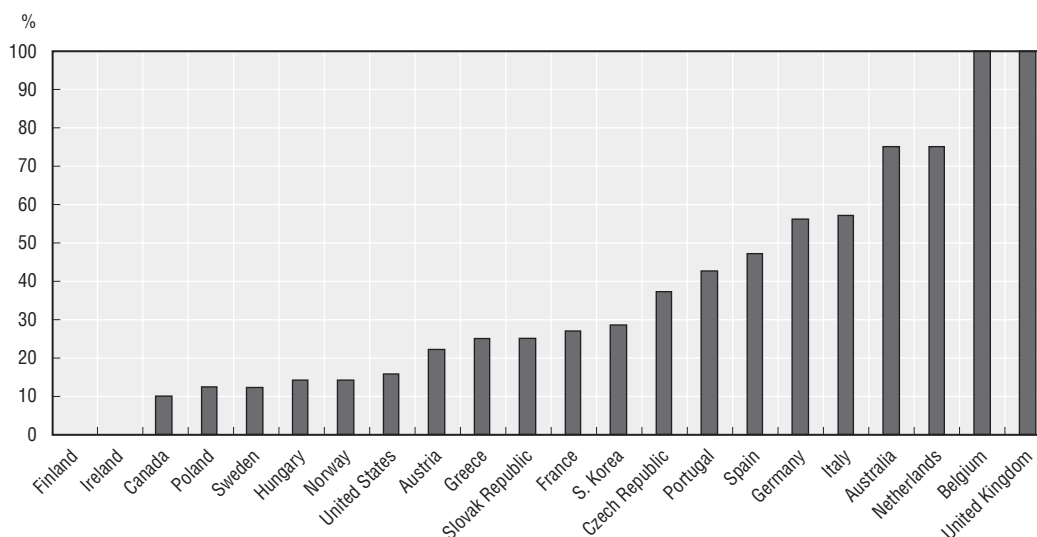
To summarise, the empirical analysis supports the hypothesis that municipal fragmentation hinders regional economic growth in urban areas, with an increasing impact given by the inverse of the rural index. Results are robust to different specifications of the

Figure 3. **Estimated effect of institutional fragmentation on annual per-capita GDP growth, in TL2 regions (with country effects)**



Source: Author's calculations.

Figure 4. **Share of TL2 regions within each country gaining from a reduction of municipal fragmentation**



Note: The definition of gaining and losing regions is based on the estimated coefficient associated with municipal fragmentation.

Source: Author's calculations.

model (inclusion of country dummies) and the use of different indicators (VA per worker). The fact that results are robust to the inclusion of fragmentation both in absolute term and relative to population highlights the importance of including a territorial dimension in the analysis. In terms of policy, the analysis provides empirical support to a place-based approach to territorial reforms. For instance, a policy aiming at reducing the number of

municipalities should focus more on urban than rural areas. In other words, it might be efficient in terms of economic development to reduce the number of municipalities in urban areas, while keeping – even small – municipalities in rural regions.

Concluding remarks

The analysis provides empirical evidence for a differentiated impact of municipal fragmentation in urban *vis à vis* rural regions. In particular, fragmentation hinders GDP per capita growth the more urban is the region. In fact, in regions where most of the population lives in rural areas municipal fragmentation appears to have no impact or enhancing growth. For OECD TL2 regions with more than 30% of people living in rural areas, fragmentation is actually positively correlated to economic growth.

The results of the empirical analysis suggest that reforms of the administrative structure of a country should take into account regional territorial differences, advocating for a place-based approach to institutional reform. The reduction of municipal fragmentation may lead to better economic performance only in regions where most of the population lives in urban areas (e.g., regions that contain metropolitan areas).

The policy implications of the analysis are twofold. Firstly, reducing municipal fragmentation may have a heterogeneous impact within the country, thus raising concern for one-size-fits-all policies of municipal agglomeration. For instance, the principle guiding municipal amalgamation should not be the average municipal size at the country level, but it should be weighted for the rural/urban characteristics of each region. Secondly, the analysis suggests that the agglomeration process should be complemented by a consistent amalgamation local administrations, otherwise representing an obstacle to the full realisation of agglomeration economies.

The present work represents a first step towards a better understanding of the relationship between subnational administration and economic performance. Further analysis should be conducted in order to account for the actual functions that are managed at the local level and the role played by vertical fragmentation (i.e., the presence of a multitier government structure) on the economic performance of regions.

Notes

1. The OECD provides a territory typology which includes two subnational levels: Territorial Level 2 (TL2) and Territorial Level 3 (TL3). The territory of each country is partitioned in TL2 regions – which corresponds to the territory of regions in Italy and France and to the territory of the States in the USA. Then each TL2 region is further partitioned in TL3 entities – e.g., departments in France and counties in the USA.
2. Larger regions, in terms of population, have also a larger absolute number of local governments. Nevertheless, the robustness of the empirical analysis is tested using the absolute value of fragmentation – i.e., not weighted by the regional population.
3. This measure is correlated to population density, but avoids the risk of having low densely populated regions although most of the population concentrates in few areas.
4. This literature review is based on Wolman et al. (2011).
5. To our knowledge the only work that controls for the territorial characteristics of the jurisdiction is Dolan (1990), which considers the impact of municipal fragmentation on the cost of government (a composite indicator which accounts for several dimensions of costs in the public administration). Dolan (1990) considers the ratio between agricultural land and commercially and industrially developed land as a measure of the territorial characteristics. He finds, however, no direct impact of its rural index on government cost.

6. The TL3 territory is the way in which TL2 territories are further divided. It corresponds to the territory of *province* in Italy, *departments* in France, and counties in the USA. The territorial level does not necessarily correspond to an administrative tier. For instance, in the Netherlands there is no administrative body corresponding to the TL2 territory, as below the central government there is only the provincial administration (corresponding to the TL3 definition) and the municipal level.
7. For Canada the municipal level corresponds to “census subdivisions”; Indian reserves and settlements as well as special purpose entities are excluded. For the United States the municipal level comprises only general purpose entities. For Australia the municipal level corresponds to “local government areas”, which includes a variety of entities defined in each of the six States; unincorporated areas are not included.
8. This is not the indicator used in the empirical analysis, which is constructed at the regional level.
9. For the TL2 regions of Norway and Korea, the average annual growth is calculated for the period 2003-07. For the TL2 regions of Turkey the average annual growth is calculated for the period 2004-07.
10. Data on the number of counties in US are from the National Association of Counties, Research Division, 440 First Street, NW, Washington, DC 20001. (202) 393-6226 (as published in the webpage www.charlestoncounty.org/stats/bystate.htm). For the US, sub-county general purpose administrative units are used as “municipalities”.
11. Neither physical capital nor investments in physical capital are available at the regional level. In the model presented this element is partly captured by the level of GDP of the region which is likely to be correlated with the amount of physical capital. Thus the level of GDP can be a proxy for the level of physical capital in the absence of such regional indicator.
12. Two Canadian TL2 regions, namely Yukon and Northern Territories, have been excluded from the analysis because of lack of data on education attainment.
13. The work is conducted using GDP per capita because data are available for more years and more regions than data on value added per worker.
14. The same results are obtained when removing population density from the regression, in particular the coefficient associated with municipal fragmentation is still negative but not statistically significant.

References

- Acemoglu, D. and J.A. Robinson (2006), *Economic Origins of Dictatorship and Democracy*, Cambridge University Press.
- Akai, N. and M. Sakata (2002), “Fiscal Decentralization Contributes to Economic Growth: Evidence From State-level Cross-section Data for the United States”, *Journal of Urban Economics*, 53, pp. 93-108.
- Ahrend, R., E. Farchy, I. Kaplanis and A. Lembcke (2014), “What Makes Cities More Productive? Evidence on the Role of Urban Governance from Five OECD Countries”, *OECD Regional Development Working Papers*, No. 2014/05, OECD Publishing, Paris, [10.1787/5f2432cf2d8p-en](https://doi.org/10.1787/5f2432cf2d8p-en)
- Barro, R. and X. Sala-i-Martin (2004), *Economic Growth*, MIT Press, second edition.
- Brennan, D. and J. Buchanan (1980), *The power to tax: analytical foundations of a fiscal constitution*. Cambridge University Press, Cambridge.
- Charbit, C. (2011), “Governance of Public Policies in Decentralised Contexts: The Multi-level Approach”, *OECD Regional Development Working Papers*, No. 2011/04, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5kg883pkxkhc-en>.
- Charbit, C. and M. Michalun (2009), “Mind the Gaps: Managing Mutual Dependence in Relations among Levels of Government”, *OECD Working Papers on Public Governance*, No. 14, OECD Publishing, Paris, <http://dx.doi.org/10.1787/221253707200>.
- Davoodi, H. and H. Zou (1998), “Fiscal Decentralization and Economic Growth: A Cross Country Study”, *Journal of Urban Economics*, 43, pp. 244-257.
- Dollery and Robotti (eds) (2008), *The theory and practice of local government reform*, Edward Elgar Publishing.
- Eberts, R. and T. Gronberg (1990), “Structure, Conduct, and Performance in the Local Public Sector”, *National Tax Journal*, 43(2), 165-173.
- Grassmuck G. and M. Shields (2010), “Does Government Fragmentation Enhance or Hinder Metropolitan Economic Growth?”, *Papers in Regional Science*, 89(3), pp. 641-657.

- Hamilton, D., D. Miller and J. Paytas (2004), "Exploring the Horizontal and Vertical Dimensions of the Governing of Metropolitan Regions", *Urban Affairs Review*, 40(2), pp. 147-182.
- Hirschman, A. (1970), *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, Harvard University Press, Cambridge MA.
- Iimi, A. (2005), "Decentralization and Economic Growth Revisited: An Empirical Note", *Journal of Urban Economics*, 57, pp. 449-461.
- Lars, P.F., H. Zimmermann and T. Döring (2004), "Federalism, Decentralization, and Economic Growth", *Philipps-Universität Marburg Working Paper No. 30/2004*, Marburg.
- Lin, J.Y. and Z. Liu (2000), "Fiscal Decentralization and Economic Growth in China", *Economic Development and Cultural Change*, 49, pp. 1-21.
- Nelson, A. and K. Foster (1999), "Metropolitan Governance Structure and Income Growth", *Journal of Urban Affairs*, 21(3), pp. 309-324.
- North, D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge University Press.
- Oates, W.E. (1999), "An Essay on Fiscal Federalism", *Journal of Economic Literature*, 37(1), pp. 1120-1149.
- OECD (2009a), "Regional typology: Updated statistics", www.oecd.org/gov/regional/statisticsindicators.
- OECD (2009b), *How Regions Grow: Trends and Analysis*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264039469-en>.
- OECD (2010a), *Regional Development Policies in OECD Countries*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264087255-en>.
- OECD (2010b), *Strategies to Improve Rural Service Delivery*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264083967-en>.
- OECD (2012), *Promoting Growth in All Regions*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264174634-en>.
- OECD (2015), *The Metropolitan Century: Understanding Urbanisation and its Consequences*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264228733-en>.
- OECD (2016), *OECD Regions at a Glance 2016*, OECD Publishing, Paris, http://dx.doi.org/10.1787/reg_glance-2016-en.
- OECD/KIPF (2012), *Institutional and Financial Relations across Levels of Government*, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264167001-en>.
- Paytas, J. (2001), *Does Governance Matter? The Dynamics of Metropolitan Governance and Competitiveness*, Carnegie Mellon Center for Economic Development, Pittsburg, PA.
- Rodriguez-Posé, A. and R. Ezcurra (2010), "Is Fiscal Decentralization Harmful for Economic Growth? Evidence from the OECD Countries", *Journal of Economic Geography*, 11, pp. 619-643.
- Stansel, D. (2005), "Local Decentralisation and Local Economic Growth: a Cross-sectional Examination of US Metropolitan Areas", *Journal of Urban Economics*, 57(1), pp. 55-72.
- Thornton, J. (2007), "Fiscal Decentralization and Economic Growth Reconsidered", *Journal of Urban Economics*, 61, pp. 64-70.
- Tiebout, C.M. (1956), "A Pure Theory of Local Expenditure", *Journal of Political Economy*, 64(5), pp. 416-424.
- Treisman, D. (2002a), "Decentralization and the Quality of Government", UCLA, mimeo.
- Treisman, D. (2002b), "Defining and Measuring Decentralisation: A Global Perspective", UCLA, mimeo.
- Treisman, D. (2006), "Fiscal Decentralization, Governance, and Economic Performance: A Reconsideration", *Economics and Politics*, 18 (2), pp. 219-235.
- Williamson, O.E. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press, New York NY.
- Williamson, O.E. (2010), "Transaction Cost Economics: the Natural Progression", *American Economic Review*, 100, pp. 673-690.
- Wolman, H., A. Levy and D. Hincapie (2011), "Government, Governance, and Regional Economic Growth", *Working Paper No. 044*, George Washington Institute of Public Policy, Washington, DC.
- Zhang, T. and H. Zou (1998), "Fiscal Decentralization, Public Spending and Economic Growth in China", *Journal of Public Economics*, 67, pp. 221-240.

Appendix

A1. Impact of municipal fragmentation by OECD countries

The empirical analysis provides evidence of a different impact of municipal fragmentation in urban and rural regions. Table 4 shows the marginal effect of fragmentation (as estimated in our model) on per-capita GDP growth, with regions grouped by countries. The marginal impact is plotted against regional population density, showing that the effect becomes negative the more densely populated is the region.

Table 4. Impact of fragmentation on per-capita GDP growth, by country

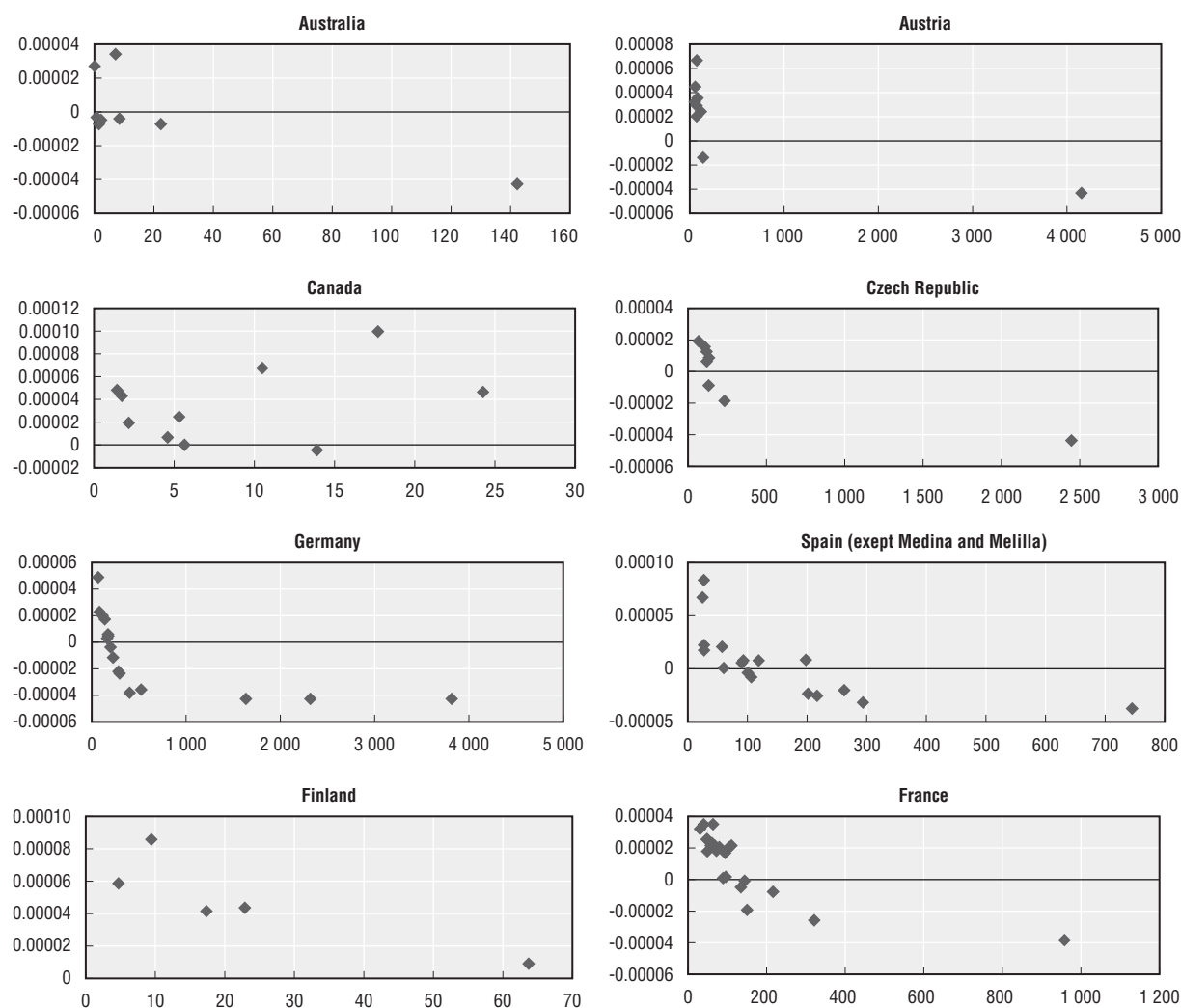
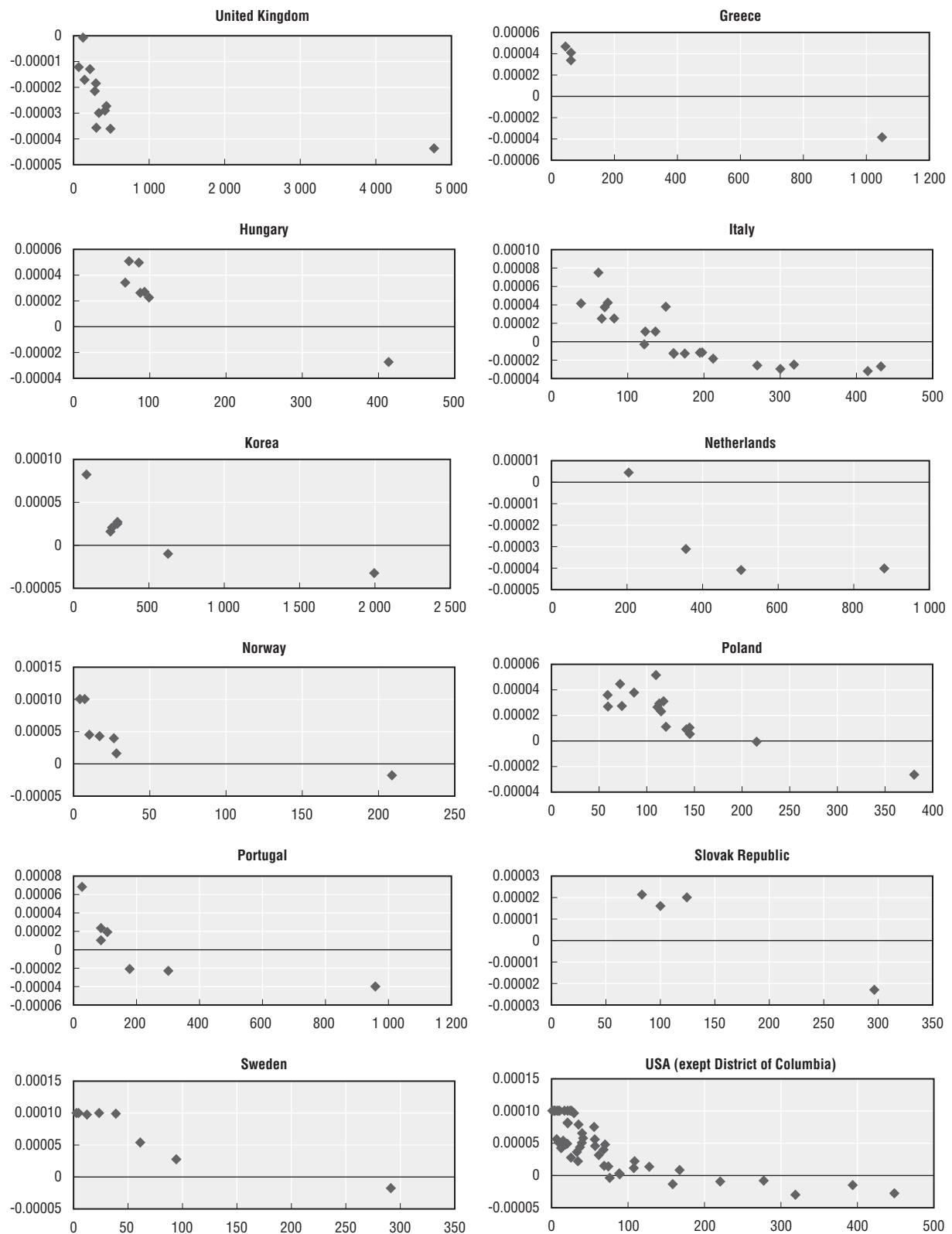


Table 4. **Impact of fragmentation on per-capita GDP growth, by country (cont.)**

Note: Data on Belgium, Luxembourg, and Ireland are not displayed as a plot, because they consist of less than four observations.

Source: Author's calculations.

These plots show that within countries the marginal impact of fragmentation is positive for some regions (less densely populated) and negative for other. In most countries it is difficult to predict the overall effect of a reform that reduces (or increases) municipal fragmentation.

A2. Administrative territorial reforms may result in asymmetries between urban and rural areas

The first 15 years of the 21st century had been characterised by an intense activity of reform of the sub-national administrative structure in many OECD countries (see Table 1). In some cases, the process has received a further stimulus by the economic crisis and the need to pursue fiscal consolidation, which forced many governments to reduce the number of subnational administrative.

Table 5 shows a wide range of reforms implemented by OECD countries. Although they are country specific and differ in many aspects, one common feature is the homogenous treatment of urban and rural areas. As documented by the empirical analysis, using a space-blind policy may lead to different results in different regions.

Table 5. **Territorial administrative reforms in selected OECD countries**

Country, year	Municipal fragmentation	Multilevel structure	Note
Latvia, 2009	Reduction of the number of municipalities from 527 to 119	Abolished the intermediated level of government (26 districts), leaving only one level of sub-national government (the municipal level)	The reform did not distinguish between urban and rural areas.
Germany, 2006-11	Trend of mergers between municipalities leading to a decline from 16 216 in 1990 to 11 327 municipalities in 2012 (a drop of 30%)		The trend of mergers affected some Federal States more than others; for instance, in 2011 the number of municipalities in Saxony-Anhalt was reduced from 840 to 219 (where 44% of the population lives in rural areas)
Italy, 2014	No mergers of municipalities, but inter-municipal co-operation is encouraged for municipalities with less than 5 000 inhabitants	Transformation of the intermediate level of government (provinces) into an assembly of mayors. Creation of ten metropolitan governance bodies (<i>città metropolitana</i>) for the administration of large cities	Some provinces are transformed in metropolitan bodies while other in assembly of majors. This may create an asymmetry in terms of functions and competences.
France, 2014		In December 2013 was passed a law creating new governance structure for the top three metropolitan areas (Paris, Lyon, and Aix-Marseille) as well as for 11 other urban areas of more than 400 000 inhabitants on a voluntary basis. According to a draft law the competences of the intermediate level, <i>départements</i> , may be modified, and merges among them and among regions considered. The reform of the Regional level of government was adopted in 2014	The reform implicitly introduces an asymmetric governance structure between big urban centres and rural regions
Greece, 2010-11	In January 2011, the number of municipalities decreased from 1 033 to 325 (a three-fold drop)	The 54 departments were replaced by 13 democratically elected regions, including two metropolitan regions (Attica and Thessaloniki)	The so-called <i>Kallikratis</i> reform, adopted as part of the 3852/2010 law and operational since 1 January 2011, is both a territorial reform and institutional reform.
Denmark, 2007	As a result of the territorial administrative reform the number of municipalities dropped from 271 to 98, with an average population of 55 000 inhabitants.	Replacement of 13 counties by 5 newly created regions (at the NUTS2 level)	The reform included a new distribution of tasks between levels of government, and a new financing and equalisation system. The merger of municipalities was achieved by imposing a limit of 20 000 inhabitants for any municipality.

Source: Various sources compiled by OECD; Dexia (2012) Sub-national public finance in the European Union, Dexia editions, Paris; Nam, Chang-woon, "Subnational Government System in the EU and Its Recent Reforms", CESifo DICE Report 11 (4), 2013.

A3. Post-estimation test of multicollinearity

The set of dependent variables should be linearly independent; otherwise estimation results risk to be highly unstable. In fact, the presence of a regressor that is the linear combination of another regressor poses a problem of multicollinearity. In order to test for multicollinearity it is possible to use the variance inflation factor (VIF). This indicator provides a measure of the degree of collinearity among dependent variables. As a rule of thumb a value of the inverted VIF below 0.1 is a sign of multicollinearity. In our case, all dependent variables display a factor higher than 0.1, showing no sign of multicollinearity.

Table 6. Post estimation test of multicollinearity

Variable	VIF	1/VIF
rurfrag	5.1	0.196243
frag	4.55	0.219779
g1996	1.85	0.541459
rur	1.83	0.547912
federal	1.54	0.650064
patents	1.49	0.670392
popdens	1.42	0.706543
tertiary	1.3	0.767956
primary	1.27	0.788658
Mean VIF	2.26	

National testing policies and educator based testing for accountability: The role of selection in student achievement

by
William C. Smith

Increasingly accountability in education has linked student test scores to teacher and school evaluations. The underlying assumption behind this educator based accountability is that the high stakes linked to student test scores will prompt behavioral change, thus improving student learning and education quality. This study conducts a cross policy analysis using pooled data from the 2009 PISA, categorizing participant countries of the 2009 PISA into three national testing policies based on what type of educator based accountability is applied in the country. Results indicate that initial differences between national testing policy categories are not significant once school types and school practices that select on the student are included. This suggests that potential gains from more stringent accountability may be an artifact of schools under pressure engaging in practices that shape their testing pool, such as admitting only relatively high achieving students or transferring out lower achieving students.

JEL classification: I21, I24, I25, I28

Keywords: Education, PISA, accountability, testing, equity

Proponents of educator based testing for accountability often argue that linking student test scores to teacher and school evaluations prompt behavioural change that will improve student learning and thus education quality. Both market and behaviourist approaches are used in educator based testing for accountability with the former using publicly disseminated school aggregate test scores to inform consumers (parents), encouraging them to choose high performing schools and forcing schools to compete for students, and the latter connecting punitive measures to school aggregate test scores, enticing educators to adapt their behaviour to avoid sanctions or attain rewards. This study explores the relationship between educator based testing for accountability using the Smith's (2016d) recently released National Testing Policy (NTP) categories. Using pooled data from the 2009 Programme for International Student Assessment (PISA) this cross policy analysis addresses two research questions. Firstly, how are student PISA scores related to the type of testing for accountability the school is embedded in? Secondly, how is the relationship between NTP and student test score related to school practices that select on students?

The study starts with a review of the literature on accountability in education, the motivation behind testing for accountability, and the associated student outcomes. This is followed by an introduction to NTP categories, other included variables and the primary analytic approach. As students are nested in schools which are nested in countries a three level hierarchical linear model (HLM) is used to predict student mathematics scores on the 2009 PISA. The results section addresses the two research questions and is followed by a conclusion connecting the results with previous literature and discussing limitations and next steps.

Literature review

Accountability in education

Accountability is the process of identifying who is responsible for what. It generally starts with a desired outcome such as clean water or safer streets but is often only noticed by the public and emphasized when that desired outcome is not reached. During these shortcomings accountability systems can be readjusted so the motivational levers in place change the behaviour of the responsible actor to ensure the desired outcome. Within the system, individuals delegated responsibility “have to account for what they are doing with this authority or responsibility” (OECD, 2011, p. 430). Their success in reaching the desired outcome is often evidenced in simplified and quantified measures. For example, safer streets may be judged solely on the number of violent crimes and clean water measured by toxins per millilitre.

The increased accountability over the past few decades in social sectors like education and health has been identified by Hopmann (2008) as the ‘age of accountability’. During this ‘age of accountability’ governments are accountable to the public for their investments (Smeed and Victory, 2010); “the public has a right to expect that its resources are being used responsibly and that the public institutions are accountable for caretaking the public trust”

(Supovitz, 2009, p. 215). Since arguments for education have become entrenched in the human capital rhetoric that first appeared in the 1970s (Becker, 1962; Schultz, 1961), education has been largely viewed as an investment with the public interested in issues of efficiency, effectiveness, and return on investment (Dorn, 2007). The increased emphasis in education and the acknowledgement of education's importance for personal and national well-being at the turn of the 21st century led Volante in 2007 to claim that internationally “accountability systems has been one of the, perhaps the most powerful, trend in education policy in the last 20 years” (p. 4).

Although still in dispute, many would argue that one of the primary desired outcomes of education is student learning. During this intensification of accountability in education, educators (administrators and teachers) are increasingly held responsible for student learning (Smith, 2014; Wiliam, 2010). This removes some of the responsibility from parents and redirects attention away from other social correlates of student learning, such as poverty and student health. Educators, namely teachers, are central in accountability policy as they play a dual role as policy implementer and student influencer. Therefore, motivational levers that change educator behaviour have the ability to impact large numbers of students at once.

Test scores and accountability

Student test scores have been employed to evaluate student learning or achievement. Test scores, as simplified, quantified measures, have been embraced by proponents for their objectivity, ability to provide standardized metrics across diverse environments, and utility in overcoming the principal-agent problem, providing the information necessary for school choice. Since the Chinese Civil Service exam, many have considered tests or exam scores “more objective and less open to particularistic influence” (Ho, 1964, p. 15) than other available measures of quality. There is also a sense that quantified measures, such as test scores, are more accurate and trustworthy, scientific, statistically valid, and reliable (Hopmann, 2008; Paris and McEvoy, 2000; Wiseman, 2010). By providing standardized measures, test scores provide comparable indicators of quality across heterogeneous environments, making their use common in decentralized education systems (Woessman, 2007) to ensure that everyone is held to the same high expectations (Stotsky, 2000).

Test scores can also provide information to overcome the principal-agent problem (Figlio and Loeb, 2011). The principal-agent problem starts with the assumption that the principal (i.e. parents) has hired an agent (i.e. educators) to perform a service. If the interests of the principal and the agent do not align the service may not be performed efficiently: “for if an agent’s interests diverge from those of the principal, and if the principal is not fully informed about the agent’s real performance, then the agent may pursue his own interest instead of those of the principal, without the latter becoming aware of this behaviour and thus without him being able to sanction it” (Woessman, 2004, p. 5). Supporters of making test scores publicly available believe the score provides the common metric for evaluating the educator’s real performance. The expanding use of publicly available comparable information on performance is correlated with increasing school choice policies (Smith and Rowland, 2014) where test scores allow for greater “customer sovereignty” (Carl, 1994). The foundational justification for school choice comes from the earlier work of Friedman (1955, 1962) and Chubb and Moe (1990) who believed that markets are the solution to a failing system. Education would then become more efficient and effective as public schools were situated in a marketplace and forced to compete with private schools. With test scores

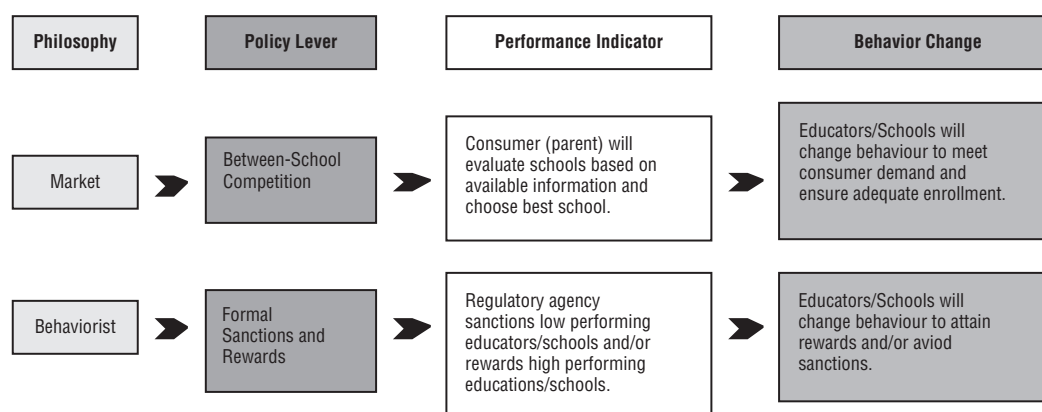
providing parents with the necessary “price information” (Woessman, 2007) to make informed choices, schools would have to compete for student enrolment.

Given the taken for granted notion that test scores are the only legitimate measure of quality in education (Booher-Jennings, 2005) it is not surprising that globally we are seeing the ubiquitous use of student test scores in accountability policies (Hanushek and Raymond, 2004). Testing has, in essence, become synonymous with accountability in education (Smith, 2016b) and testing for accountability – the “application of formal or informal, positive or negative consequences on educators dependent on their student’s performance” (Smith, 2014, p. 6) – has been adopted as a common solution to education problems around the world (Sahlberg, 2010). The engrainment of testing for accountability and other student focused high-stakes standardized tests at the national and international level has recently been described as the Global Testing Culture (Smith, 2016b). Within the Global Testing Culture schools align their practices “with the institutional environment in order to achieve legitimacy and ensure their survival” (Booher-Jennings, 2005, p. 234). Nationally, ministries increasingly participate in international testing and develop national assessment systems to legitimate their education systems to their peer countries as well as their own citizens (Kamens and McNeely, 2010; Kijima and Leer, 2016).

Motivation behind testing for accountability

Countries adopt testing for accountability approaches based largely on two philosophies (see Figure 1). The market philosophy provides comparable school level test scores to promote competition between schools. This approach is driven by a belief that the invisible hand of the market will improve the quality of schools (Apple, 1999). Parents, informed and acting in the best interest of their child, will base their school choice on the available test score and educators will change their behaviour to meet consumer demand and ensure their professional livelihood. There is some evidence that countries are moving towards more testing for accountability (Smith, 2014) with Teelken (1999) arguing that all nations were effectively moving “closer to the market mechanism” (p. 298).

Figure 1. **Testing for accountability philosophies**



Source: Adapted from Smith (2016d).

The behaviourist philosophy suggests that rewards or sanctions can shape behaviour (Woessman, 2007) and that serious consequences are needed to transform the education system (Braun and Kanjee, 2006). It builds on research indicating the concern for teachers

is their personal and professional survival (Gilles, Cramer and Hwang, 2001) and that educators have a heightened survival response when faced with testing for accountability (Nichols and Berliner, 2007). From this approach educators change their behaviour to either attain rewards or avoid punishments.

Outcomes associated with testing for accountability

Past research has indicated that testing for accountability measures have been successful in altering educator behaviour (Booher-Jennings, 2005; Smith, 2016d). At times the competition and sanctions that have fuelled behaviour change has been associated with marginally higher test scores (Levin and Belfield, 2003) and more efficient outcomes (Levin, 1992; Levin and Belfield, 2003). Some, however, have found no effect of testing for accountability systems on student test scores (Natriello and Pallas, 1999) or diminished effects in lower stakes “audit tests” relative to the higher stakes tests linked to the accountability policy (Figlio and Loeb, 2011). Past meta-analyses by Lee (2008) and Phelps (2012) have both found small positive effects of testing for accountability on student test scores, however, Lee cautions that the minimal effect size “does not lend strong support for claims for school accountability” (p. 616).

When the two approaches to testing for accountability are compared, past research indicates that the behaviourist model of applying formal sanctions or rewards is more effective in eliciting greater test score gains (Bishop et al., 2001; Hanushek and Raymond, 2005). Although Springer (2008) found a positive association in systems that publicly compared their school level results he recognized that the “practical significance of this gain is negligible” (p. 5).

As these results indicate, motivational levers through testing for accountability may produce behavioural change in educators that lead to genuine gains in student learning. Alternatively, past research cautions that the incentives in place when a testing for accountability approach is applied may motivate educators to game or manipulate the system in order to increase the likelihood of high test scores. Outside of examples of outright cheating, with educators manipulating test results or completing student exams (Gipps, 2003; Jacob and Levitt, 2003; Lashway, 2001), educators have been known to game the system through narrowing the curriculum, teaching to the test, teaching to the “bubble”, and shaping the testing pool. The first three generally occur at the classroom level and may be motivated through targeted professional development. Under the pressure of testing for accountability, instructional time is often shifted toward tested subjects and basic test preparation (McNeil and Valenzuela, 2001; Rentner et al., 2006). Moreover, pedagogical practices tend to focus on surface level rote memorization instead of deep understanding (Froese-German, 2001; Somerset, 2016; Supovitz, 2009). The test atmosphere can also be mimicked through classroom set up and practicing with similarly formatted tests (Yarema, 2010). Additionally, with poorly performing students less likely to make the desired mark on the test some educators focus on those in the “strike zone”, described by some as the “bubble kids” who have scored near the desired mark in the past and are more likely to attain it in the future (Booher-Jennings, 2005; Gillborn and Youdell, 2000; Lipman, 2004; Reback, 2008).

Shaping the testing pool changes the group of students subject to testing, generally excluding low performing students to increase the average test score (Fallabela, 2014; Hanushek and Raymond, 2004; Ooghe and Schokkaert, 2013). Unlike the other gaming behaviours, which can be done independently at the classroom level, shaping the testing pool generally requires some amount of coordination at the school level. Shaping the testing pool

includes reclassifying students into non-tested groups, selectively admitting students into the school, and temporarily or permanently removing low performing students from the school. In more intense accountability systems, low achieving students have been reclassified as working on an alternative certificate (Lewin and Meina, 2003) or in special education (Booher-Jennings, 2005; Cullen and Reback, 2006; Haney, 2000). Internationally, special education students are often not included in school aggregate test scores (Booher-Jennings, 2005; Eurydice, 2009). Selective admission policies involve admitting only students that have performed well in the past. The practice, often known as “cream skimming” (Gerwitz et al., 1995), is, as Fallabela (2014) explains, “one of the most powerful contemporary school improvement strategies” in which “selective schools guarantee good test results and institutional reputation in a more efficient and rapid manner” (p. 12). Under pressure, schools select in some students while selecting out others (West et al., 1998). This may include transferring, excluding, or expelling low achieving students. In the three years following the advent of league tables in the United Kingdom the permanent exclusion rate increased by 300% (Gillborn, 1996). Furthermore, to avoid the stigmatization of exclusion some parents have been coached to remove their low performing children from school, in a process known as “constructive exclusion” (Gerwitz et al., 1995). Temporary exclusion during the testing period has also been noted, with discipline practices disproportionately targeting low performing students (Figlio, 2006). Finally, retention is increasingly used in testing for accountability systems with the likelihood that a student is retained, and therefore not allowed to test with their peer cohort, greater in the year prior to the test (Kornhaber, 2004; Hursch, 2005).

This study

To examine the relationship between testing for accountability approaches and student test scores this study uses Smith’s categorization of 2009 Programme for International Student Assessment (PISA) participants into National Testing Policy (NTP) categories. Through a three level hierarchical linear model (HLM) two research questions are addressed. 1) How are student PISA scores related to the type of testing for accountability the school is embedded in? and 2) how is the relationship between NTP and student test score related to school practices that select on students? Measuring testing for accountability at the national level is important as “national ministries of education typically act as agents imposing this activity [testing] on schools and education systems” (Kamens and McNeely, 2010, p. 6) and previous research measuring educator based testing for accountability often relies on single school responses to whether they publicly post their test score results (Rosenkvist, 2010; Ikeda, 2011) which fail to capture the larger accountability environment (Smith, 2016c).

National testing policies

Outlined in Figure 2 are Smith’s (2014, 2016d) National Testing Policy categories. For more technical details on the categorization process see Smith (2016d).¹ This study uses his categorization of the participants of the 2009 PISA into NTP categories and compares those applying the Summative policy to the two using the testing for accountability approaches; Evaluative and Punitive (see Appendix A for full categorization). Policies were categorized through a careful examination of national and international policy documents as of the 2008-09 school year, based on national and regional tests conducted at the ISCED 1 and ISCED 2 level, and dummy coded. In countries with multiple tests the test with higher levels of accountability (i.e. toward Punitive) was included as the country example. Multiple sources were used for data triangulation when possible. Following initial categorization, national

Figure 2. **National testing policy categories**

National Testing Policy	Summative	Evaluative	Punitive
Underlying Philosophy		Market	Behaviourist
Description of National Testing Policy	The use of national or regional examinations as a tool that summarizes student learning and is shared with parents; if disseminated more broadly it is only at national or regional level.	The use of national or regional examinations as a tool that summarizes student learning and is disseminated at the school level to allow for between-school comparisons.	The use of national or regional examinations as a tool that summarizes student learning is disseminated at the school level, with school/class level results used to apply rewards or sanctions.

Source: Adapted from Smith (2016d).

experts from within the author's network were consulted to clarify ambiguous country categories. An updated country categorization was then sent out to the Organization for Economic Co-operation and Development (OECD) Group of National Experts on Evaluation and Assessment for review. Their feedback provided additional support for the categorization through the identification of individual country nuances. Countries with large regional variance in their NTP are coded into multiple categories when the variance divides the country by NTP and data available in the 2009 PISA can be divided by region (i.e. Belgium, Canada, and the United Kingdom).

Evaluative policies aggregate student test scores at the school level and make school level scores publicly available. The policy is linked to the previously described market mechanism in which informed parents use school level test score to evaluate their options and choose the best option for their child. Within this policy "responsibility is devolved to the individual consumer and the aggregate of consumer choices provides the discipline, of accountability and demand, that the producer [school or educator] cannot escape" (Ball, 1995, p. 69). The classic example of publicly stigmatizing schools through school aggregate student test scores can be seen in the league tables of England. Proponents believe that public pressure through identifying poorly performing schools with a 'scarlet letter' (Harris and Herrington, 2006) or 'naming and shaming' (de Wolf and Janssens, 2007) will encourage educators and schools to change the behaviour and ensure the desired outcome of student learning.²

The Punitive testing policy applies a more intense form of accountability and moves one step beyond the Evaluative policy. Countries that apply a punitive policy aggregate and disseminate school level test scores and then apply rewards or formal sanctions based on the results. This may include increasing or decreasing school autonomy or funding (Figlio and Loeb, 2011). This policy is linked with the behaviourist approach previously mentioned and is best illustrated in the United States during the period of No Child Left Behind (NCLB). During NCLB multiple "turnaround" models were applied to poorly performing schools including closing the school altogether or replacing the majority of the staff. Although technically countries may prompt behaviour change through a behaviourist approach that applies only rewards to high performing or improving schools, in Smith's categorization he found that any country that included rewards as part of its accountability policy also included sanctions.

Summative policies, in contrast, do not use testing for accountability. Instead countries that use Summative policies typically conduct sample based, not census based, tests and

aggregate scores at the national level. Comparisons are not made between schools but between disaggregated subgroups such as students with disabilities or students in the lowest income quintile. The aggregation at the national level shifts some of the responsibility back to the federal government who must then consider equity issues that come to light through disaggregation.

Data and methods

PISA 2009

Data from the 2009 PISA was used for this study. The 2009 PISA was administered by the OECD to 15 year olds in 61 countries and 4 sub-regions or participating economies. Like any data set the use of PISA has both benefits and limitations. As PISA is administered to age 15 students NTP categorization only accounted for testing at the primary and lower secondary level. As secondary leaving or exit exams are one of the most common forms of testing the extension of NTP categories to include the upper secondary level may adjust country classifications. Additionally, the cross-sectional nature of PISA limits the claims able to be made in this study. The correlational relationships identified in this study should not be mistaken for causal claims as temporal ordering between NTP and student achievement cannot be identified and a true counterfactual is not present.

As PISA is not tightly linked to a country's curriculum it can be considered an "audit test" or lower stakes soft test (Barry et al., 2010). This study focuses on student math scores as student learning in mathematics is less likely to be conflated with parental background than reading and is more likely to be associated with a testing for accountability policy (Figlio and Loeb, 2011; Wong, Cook, and Steiner, 2009). The sample for this study pools data from all 2009 PISA participating countries and economies to perform a cross-policy analysis. Nested units (i.e. schools) with two or fewer cases (i.e. students) sampled are omitted from the analysis to prevent the overestimation of explained variance and the exaggeration of model fit (Subedi, 2005). The remaining missing data were imputed using the imputation for chained equations (ice) command in Stata statistical software, version 12. With exclusion criteria and missing data procedures taken into account, the effective sample size for this study is 446 330. The sample is further reduced to 411 529 after countries without a national testing policy are removed.

Variables

Primary explanatory variables for this study include NTP and measures that limit attendance to select students. To identify whether a school may be shaping their testing pool by selecting in higher achieving students while selecting out low achieving students, schools with selective admission and transfer policies were identified from the school questionnaire. The former identifies schools where student achievement, including test scores, are included in the admission criteria (1= Sometimes or Always; 0 = Never). The latter identifies whether a student is likely to be transferred out due to low academic achievement (1 = Likely to Very Likely; 0 = Not Likely). As private schools have greater autonomy in their student selection process a dichotomous dummy variable for school type (1 = Private; 0 = Public) was included as an additional variable capturing student selection.

Student level control variables included gender, SES, home language, and immigrant status. Gender is a dichotomous variable coded 1 for female and 0 for male. SES is taken from the PISA created student level index of economic, social and cultural status (ESCS). ESCS is a

composite variable created through principle component analysis. Included in the composite are variables that identify home possessions, the highest parental occupation, and the highest parental education in years of education. It is a standardized variable with an OECD mean of zero and standard deviation of one. Median factor loadings for the observed variables in OECD countries ranged from 0.70 (home possessions) to 0.80 (parental education). Median scale reliability for OECD countries was 0.65 (OECD, 2012).

Home language is a dichotomous control variable taken from question number 19 of the student questionnaire: "What language do you speak at home most of the time?" The variable is coded 1 if the student does not speak the test language as their primary home language and 0 if the test language is their primary home language. Immigrant status is taken from the PISA created immigrant status variable and coded 1 if the student is a first or second generation immigrant (i.e. they were not born in the country they were tested) and 0 if they are native to the test country.

Education expenditure and market competition are both extracted from the World Bank Databank to capture additional differences due to country context. Education expenditure is a continuous variable that is used as a proxy for national investment in education. To ensure that education expenditure accurately represents national investment in education, given the global recession, a five year average (2005 to 2009) of public expenditure in education as a percent of GDP is used. Market competition is a continuous variable that is operationalized as the percentage of total enrolment attending a private school and, therefore, acts as a proxy for competition between schools. Market competition is calculated as the average percentage of 2009 private school enrolment at the elementary and lower secondary level.

Analytic approach

To account for the nested nature of the data with students nested in schools nested in countries three level Hierarchical Linear Modeling (HLM) was the primary analytic approach used in this study. HLM recognizes the hierarchical data structure and adjusts standard errors to provide a more robust estimate (Guo and Zhao, 2000; Raudenbush and Bryk 2002). The full HLM model addressing both research questions is illustrated in Equation 1 which predicts student achievement for i student in j school in k country. School level intercept (γ_{00k}) is predicted by a national level intercept (δ_{000}), predictor variable (δ_{001}), context variables ($\delta_{002} - \delta_{003}$) and error term (v_{00k}). The model controls for school level variables ($\gamma_{01k} - \gamma_{02k}$) and student level variables ($\beta_{0jk} - \beta_{3jk}$), and includes a cross level interaction (γ_{03k}) and school and student level error terms ($u_{0jk}; e_{ijk}$).

$$\text{Student Achievement}_{ijk} = \delta_{000} + \delta_{001} (\text{NTP}) + \delta_{002} (\text{Education Expenditure}) + \delta_{003} (\text{Market Competition}) + \gamma_{01k} (\text{School Type}) + \gamma_{02k} (\text{Shaping the Testing Pool}) + \gamma_{03k} (\text{School Type} \times \text{NTP}) + \beta_{0jk} (\text{Gender}) + \beta_{1jk} (\text{Student SES}) + \beta_{2jk} (\text{Immigrant Status}) + \beta_{3jk} (\text{Home Language}) + v_{00k} + u_{0jk} + e_{ijk}$$

The analysis was adjusted for student and school survey weights and completed using HLM software. Plausible values for student mathematics achievement were combined using the rules of Rubin (1987) to take into account their imputed nature (OECD, 2012) and provide adjusted standard errors.

Results

Prior to exploring student achievement by NTP category, an initial analysis was conducted to examine differences between achievement in countries with any form of

testing for accountability and the Summative category. HLM results controlling for student level variables indicated no significant difference between student's math scores in Summative countries and their peers in testing for accountability countries ($p = .160$).³ Given this result and the prior research indicating that a more behaviourist approach may have a greater influence on student achievement (Bishop et al., 2001; Hanushek and Raymond, 2005), the following analyses focus on comparisons between Summative, Evaluative, and Punitive NTP categories.

Descriptive statistics comparing included variables across NTP category can be found in Table 1. A one-way ANOVA was used to test for significant differences between NTP and the two continuous variables (Education Expenditure, Market Competition, Student Achievement and Student SES) while chi-square tests were used when comparing categorical variables. There is no significant difference between student test scores in Punitive or Evaluative systems and the reference group (Summative). Evaluative and Punitive countries have a significantly higher percentage of students enrolled in private schools with the percent of private school enrolment 50% higher in testing for accountability countries. Significant differences were found in student SES with students in Punitive countries scoring approximately half a standard deviation lower than their peers in Summative and Evaluative systems ($F(2,411526) = 7118.51$, $p < .001$). In addition to having lower SES, students in Punitive countries are more likely to speak the testing language at home ($\chi^2 = 6.6e + 03$, $df = 2$, $p < .001$) and be an immigrant ($\chi^2 = 587.78$, $df = 2$, $p < .001$) than students in Summative countries. At the school level, schools situated in Punitive countries are more likely to be private ($\chi^2 = 2.8e + 03$, $df = 2$, $p < .001$), and practice selective admission ($\chi^2 = 4.7e + 03$, $df = 2$, $p < .001$) and transfer ($\chi^2 = 1.1e + 04$, $df = 2$, $p < .001$).

Table 1. **Descriptive statistics by national testing policy**

	Overall	Summative	Evaluative	Punitive
Student achievement	466.96 (100.84)	465.66 (101.32)	467.87 (102.07)	468.73 (95.49)
Number of students	411 529	199 998	135 157	76 374
Number of schools	14 383	6 919	4 845	2 619
Number of countries	60	29	22	9
Education expenditure	4.63% (1.13)	4.52% (0.99)	4.93% (1.35)	4.63% (0.90)
Market competition*	16.35% (19.60)	12.48% (13.58)	19.27% (19.94)	18.68% (25.90)
School type – Private*	18.16%	16.39%	17.08%	24.84%
Admission decision*	56.23%	55.80%	51.22%	66.62%
Transfer decision*	34.83%	38.79%	24.32%	43.79%
SES*	-0.30 (1.13)	-0.23 (1.09)	-0.15 (1.11)	-0.73 (1.23)
Female	50.71%	50.44%	50.85%	51.18%
Home language – Other*	11.25%	14.75%	10.11%	4.08%
Immigrant status*	9.30%	8.20%	10.07%	10.82%

* Indicates statistically significant difference ($p < .05$) between at least two NTP categories based on one-way ANOVA or chi-square test, as appropriate. Standard deviations in parentheses.

Taking the nested nature of the data into account reveals a very different picture of student achievement. As illustrated in Table 2, in the unconditional three level HLM, adjusted for student and school weights, there is a strong positive relationship between the more intense Punitive accountability system and student PISA scores, relative to students in

Summative systems. The non-significant relationship between Evaluative and student achievement largely mirrors the past research suggesting that it is the formal sanction or reward, not the market pressure that is associated with changes in student test scores. Model 2 adds student level control variables to the unconditional model. The coefficient for Punitive is reduced slightly and remains significant, although at the marginal $p < .10$ significance level. Additional significant differences are found in three of the four student control variables with males scoring, on average, approximately 15 points higher than females, non-immigrants outscoring immigrants by 22 points, and students with a one standard deviation higher SES outpacing their peers by 11 points. The non-significance of the home language category may due to the overlap between students that speak another language at home and immigrants – over 40% of immigrants in the full sample speak another language at home, compared to 8% of non-immigrants – forcing the variables to compete for explanatory power.

Table 2. **Three level HLM predicting student mathematics score on the 2009 PISA**

	Model 1	Model 2	Model 3	Model 4
Evaluative	37.87 (30.23)	33.25 (28.64)	27.98 (27.46)	13.02 (17.26)
Punitive	61.77** (26.81)	52.05* (27.21)	39.15 (25.03)	13.60 (15.89)
Education expenditure (% of GDP)				18.67*** (6.52)
Market competition				-1.69*** (0.56)
Admission decision			10.57*** (3.93)	10.61*** (3.93)
Transfer decision			16.61*** (4.27)	16.59*** (4.27)
Private			5.54 (12.32)	5.70 (12.24)
Evaluative x Private			38.67** (17.86)	38.63** (17.75)
Punitive x Private			38.07** (15.03)	38.32*** (14.17)
SES		11.31*** (2.24)	11.00*** (2.29)	10.99*** (2.33)
Female		-14.81*** (2.33)	-14.83*** (2.20)	-14.83*** (2.20)
Home language – Other		-2.68 (4.51)	-2.58 (4.54)	-2.56 (4.54)
Immigrant status		-21.58*** (4.63)	-21.60*** (4.65)	-21.61*** (4.65)
Intercept	410.93 (22.90)	430.84 (22.08)	417.94 (23.80)	369.75 (31.31)
Deviance	4 525 512	4 509 489	4 508 132	4 508 103

* $p < .10$ ** $p < .05$ *** $p < .01$ Robust standard errors in parentheses. Student and school weights applied. Results use all five plausible valuations with standard errors adjusted according to the rules of Rubin (1987).

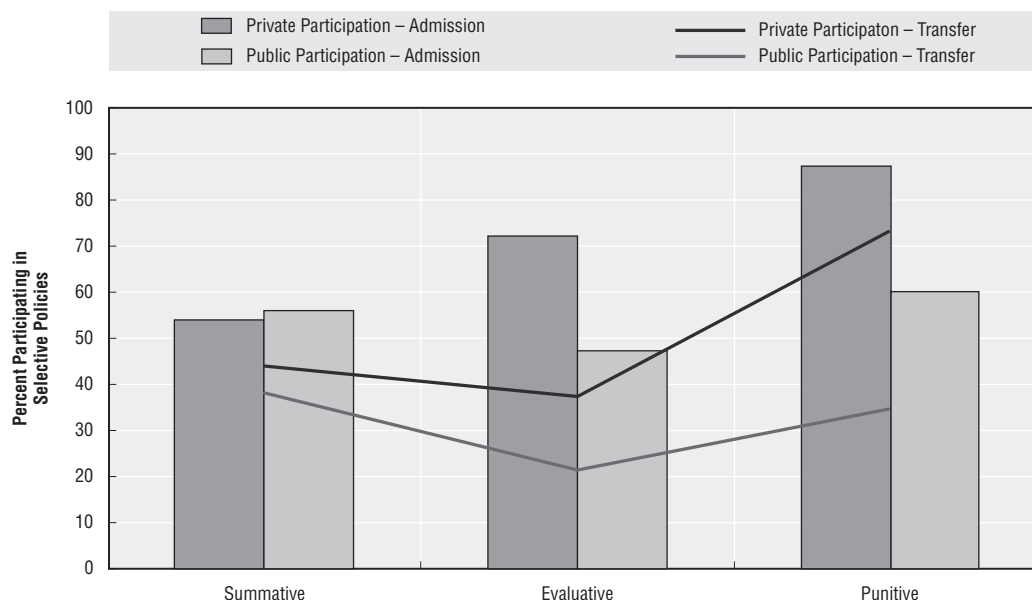
Model 3 adds school practices that select on students. Two interesting findings are clear in the results. First, the inclusion of these selection variables changes the significance level of the Punitive coefficient, suggesting that when comparing students in public schools that do not use student test scores to admit students in or transfer students out, there is no significant difference in scores between Punitive and Summative categories.

Second, an interesting interaction is present between school type and NTP. While there is not a significant difference in test scores between students in private and public schools in Summative countries, in Punitive and Evaluative countries students in private schools score nearly 0.4 standard deviations above their peers in public schools. Not surprisingly, all selection variables are positively associated with 2009 PISA scores.

The final model adds country level variables to model 3. The significance and magnitude of most previously included variables is maintained when education expenditure and market competition is included. The noticeable exception is the size of the coefficient for Evaluative and Punitive. The substantial reduction of the coefficients of these NTP categories suggests that, in addition to school level selective practices, it is important to consider country context when evaluating differences by accountability policy. Specifically, after controlling for school policies and student demographics, an increase in the percent of GDP spent on education is associated with a 19 point increase in student math scores. Private school enrolment share appears to work in the opposite direction. A 10% increase in number of students enrolled in private schools, a measure of market competition, is related to a 12 point decrease in student math scores.

In examining the school level selection policies some may argue that selective admission and transfer practices only occur in private settings, as public institutions cannot discriminate in this manner. Further analysis suggests otherwise. Figure 3 makes it clear that selection based on student achievement does not only occur in private schools. Although a greater percentage of private schools practice selective policies, a substantial number of public schools participate, including nearly 60% of public schools in Punitive systems that admit students based, in part, on their past test scores. Selective admission practices are more common than selective transfer policies in both private and public schools across NTP categories. Private schools in Punitive countries appear to be highly restrictive with 87% of schools participating in selective admission and 73% participating in selective transfer, 66% of schools participate in both.

Figure 3. **Participation in selective policies by NTP and school type**



Limitations

In addition to the limitations of cross-sectional data and the exclusion of upper secondary from NTP categorization, it is possible that some countries have moved to a different classification since 2009. For example, South Korea has moved towards a Punitive system by connecting school evaluations to student aggregate test scores (Chung and Chea, 2016). Follow up research using other rounds of PISA would better speak to whether findings are linked to NTP categories and their respective accountability approaches or are unique to a specific conglomeration of countries at a given time. Investigations over time can also use shifts in national policies to conduct a more quasi-causal analysis. Although descriptive, Smith's (2014) examination of three countries that changed their policies between 2000 and 2009 found mixed results. Hungary, which quickly moved to a Punitive policy saw a 0.5% increase in PISA math score between 2000 and 2009. The mean score for the United States increased slightly in 2003 following the implementation of the punitive NCLB policy. The 2006 and 2009 scores, however, fell below those from 2000. Finally, South Korea, who transitioned from a Summative to an Evaluative policy in 2007, saw a 2% decrease in their PISA math scores between 2006 and 2009.

Conclusion

Results of this study indicate that the initial significant association between Punitive systems and higher student test scores may be an artifact of policies that shape the student body. Specifically, once controlling for the school policies (selective admission and transfer) and types (private) associated with inflated scores, there is no significant difference in test scores between students in Punitive and Summative systems. In addition to the greater autonomy in admission practices typically found in private schools, a large number of both private and public schools participate in selecting in or transferring out lower achieving students. This is clear in Punitive countries where 60% of public schools and 87% of private schools admit students based, in part, on their achievement scores. By selectively admitting their students, schools in more intense accountability systems are able to avoid punishments by artificially increasing their school aggregate test score. This mirrors the work of Fallabela (2014) who identifies selective admission policies as one of the "most powerful contemporary school improvement strategies" (p. 12).

Across all models there is no significant difference between PISA scores in Evaluative systems, which provides the public with school aggregate scores and rely on market pressures to motivate behaviour change, and Summative systems, which do not use test scores for educator based accountability. This result suggests that accountability designed to unleash the market to fuel change through customer choice does not produce significantly different results than systems which do not aggregate and disseminate test scores at the school level. The inability of the invisible hand to create efficiency in education may be because compulsory education is incompatible with essential market assumptions, such as having no barrier to entrance and exit or having flexible pricing without artificial constraints (Beal and Beal, 2013; Diedrich, 2012).

The change in significance and magnitude of the Punitive coefficient as selective policies and country context variables are included is interesting given that students mathematics score is the chosen dependent variable. Past research suggests that if an effect of accountability is to be found, it would be on math achievement (Figlio and Loeb, 2011; Wong, Cook, and Steiner, 2009). The non-significant relationship once selective school

policies are controlled for calls into question whether testing for accountability is associated with other subjects more prone to parental instructional assistance, such as reading.

The full model indicates that country context matters and, in addition to school measures that select on the student, must be taken into consideration in explaining the effects of accountability policies. Both Evaluative and Punitive systems have a significantly higher portion of students enrolled in private schools. Furthermore, the decrease in the magnitude of both coefficients when education expenditure and market competition are included further suggests that NTP categories may capture some national trends. Relative to past research, the positive and significant relationship between education expenditure and student test score may seem first contradictory. Some studies using international large scale assessments (including PISA) find no relationship or even negative associations between school resources and student test score (Hanushek, 2003; Woessman, 2016). However, when expenditure allocated to education per pupil was included, this general spending measure was positive and significantly related to student achievement (Woessman, 2016, see Table 1); an association that is more likely to be found when aggregate data that does not capture within country differences is used (Hanushek, 2003). Additional research should be conducted exploring which national level characteristics are associated with more intense forms of testing for accountability. Finally, upon controlling for selective school policies and student demographics, results suggest that increases in spending on education are associated with better student performance while increases in private school enrolment share are related to lower student test scores.

The results point to a number of equity concerns. If, in more intense accountability systems, few schools admit low performing students, these students may end up concentrated in under-resourced, underperforming schools, under the constant threat of sanctions. This is compounded by prior research which indicates that, relative to low income schools in Summative systems, selective transfer and admission practices are more common in lower income schools in Punitive systems (Smith, 2016d). Additionally, the interesting interaction between NTP and school type (Table 2, Models 3 and 4) suggests that a large and significant gap is present between private schools and public schools in testing for accountability countries. This same difference is not found in Summative countries. More research needs to be conducted on why students in public and private schools are generally equal performers in Summative countries, but in countries that aggregate test scores at the school level and use market or punitive behavioural approaches for accountability, private school pupils greatly outpace their public school peers.

Notes

1. Note that in some versions of NTP categorization the Punitive category is identified as Formal Sanction and/or Reward.
2. Six countries or economic regions did not have an identified national testing policy at the ISCED 1 or ISCED 2 level in 2009. Results from the group without a national testing policy are not reported here but are available from the author upon request. This group includes the Czech Republic, Dubai-UAE, Greece, Liechtenstein, Panama and Switzerland.
3. An additional analysis comparing Summative, Evaluative, and Punitive to students in countries without a national testing policy (no NTP) also found non-significant differences. Full analysis based on the larger testing for accountability category, or using the no NTP category is available from the author upon request.

References

- Apple, M. (1999), "Rhetorical reforms: Markets, standards, and inequality", *Current Issues in Comparative Education*, 1(2), 6-17.
- Ball, S.J. (1995), "Parents, schools and markets: The repositioning of youth in United Kingdom education", *Young*, 3(3), 68-79.
- Barry, C., J. Horst, S. Finney, A. Brown and J. Kopp (2010), "Do examinees have similar test taking effort? A high-stakes question for low-stakes testing", *International Journal of Testing*, 10, 342-363.
- Beal, B. and H.K.O. Beal (2013), "Rethinking the market metaphor: School choice, the common good, and the National Football League", *Journal of School Choice International Research and Reform*, 7(4), 471-497.
- Becker, G.S. (1962), "Investment in human capital: A theoretical analysis", *The Journal of Political Economy*, 70(5), 9-49.
- Bishop, J., F. Mane, M. Bishop and J. Moriarity (2001), "The role of end-of-course exams and minimal competency exams in standards-based reforms", *Brookings Papers in Educational Policy*. Washington, DC, Brookings Institution.
- Booher-Jennings, J. (2005), "Below the bubble: 'Educational triage' and the Texas accountability system", *American Education Research Journal*, 42(2), 231-268.
- Braun, H. and A. Kanjee (2006), "Using assessment to improve education in developing nations", in Cohen, J.E., D.E. Bloom and M. Malin (Eds.), *Improving Education through Assessment, Innovation, and Evaluation* (pp. 1-46), Cambridge, MA: American Academy of Arts and Sciences.
- Carl, J. (1994), "Parental choice as national policy in England and the United States", *Comparative Education Review*, 38(3), 294-322.
- Chubb, J. and T. Moe (1990), *Politics, Markets, and America's Schools*, Washington, DC, Brookings Institution Press.
- Chung, P.J. and H. Chea (2016), "South Korea's accountability policy system and national achievement test", in W.C. Smith (Ed.) *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice* (pp. 249-260), Oxford, Symposium Books.
- Cullen, J. and R. Reback (2006), "Tinkering towards accolades: School gaming under a performance accountability system", in Gronberg T. and D. Jansen (Eds.) *Advances in Applied Microeconomics* (pp. 1-34), London, Emerald Group Publishing Limited.
- De Wolf, I. and F. Janssens (2007), "Effects and side effects of inspection and accountability in education: An overview of empirical studies", *Oxford Review of Education*, 33(3), 379-396.
- Diedrich, M. (2012), *False Choices: The Economic Argument against Market-Driven Education Reform*, Saint Paul, MN: Minnesota 2020, www.mn2020.org/assets/uploads/article/False_Choices_2012.pdf.
- Dorn, S. (2007), *Accountability Frankenstein: Understanding and Taming the Monster*, Charlotte, NC, Information Age Publishing.
- Eurydice (2009), *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*. Education, Audiovisual and Culture Executive Agency, European Commission.
- Falabella, A. (2014), "The Performing School: The Effects of Market and Accountability Policies", *Education Policy Analysis Archives*, 22 (70).
- Figlio, D. (2006), "Testing, crime, and punishment", *Journal of Public Economics*, 90, 837-851.
- Figlio, D. and S. Loeb (2011), "School accountability", *Handbook of the Economics of Education*, 3, pp. 383-421.
- Freidman, M. (1962), *Capitalism and Freedom*, Chicago, IL: University of Chicago Press.
- Freidman, M. (1955), "The role of government in education", in Solo, R. (ed.) *Economics and the Public Interest*, New Brunswick, NJ, Trustees of Rutgers College in New Jersey.
- Froese-Germaine, B. (2001), "Standardized testing + high-stakes decisions = educational inequity", *Interchange*, 32(2), 111-130.
- Gerwitz, S., S. Ball and R. Bowe (1995), *Markets, Choice and Equity in Education*. Buckingham, Open University.
- Gillborn, D. (1996), *Exclusions from School*, London, University of London.
- Gillborn, D. and D. Youdell (2000), *Rationing Education: Policy, Practice, Reform and Equity*. Buckingham, UK, Open University Press.

- Gilles, C., M. Cramer and S. Hwang (2001), Beginning teacher perception of concerns: A longitudinal look at teacher development, *Action in Teacher Education*, 23(3), 92-98.
- Gipps, C. (2003), "Educational accountability in England: The role of assessment", Paper presented at the Annual Meeting of the American Education Research Association, New York, NY.
- Guo, G. and H. Zhao (2000), "Multilevel modeling for binary data", *Annual Review of Sociology*, 26, 441-462.
- Haney, W. (2000), "The myth of the Texas miracle in education", *Education Analysis Policy Archives*, 8(41).
- Hanushek, E.A. (2003), "The failure of input-based schooling policies", *The Economic Journal*, 113(485), F64-F98.
- Hanushek, E. and M. Raymond (2005), "Does school accountability lead to improved student performance?", *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Hanushek, E. and M. Raymond (2004), "The effect of school accountability systems on the level and distribution of student achievement", *Journal of the European Economic Association*, 2(2-3), 406-415.
- Harris, D. and C. Herrington (2006), "Accountability, standards, and the growing achievement gap: Lessons from the past half-century", *American Journal of Education*, 112(2), 209-238.
- Ho, P. (1964), *The Ladder to Success in Imperial China: Aspects of Social Mobility*, New York, Science Editions.
- Hopmann, S.T. (2008), "No child, no school, no state left behind: Schooling in the age of accountability", *Journal of Curriculum Studies*, 40(4), 417-456.
- Hursh, D. (2005), "The growth of high-stakes testing in the USA: Accountability, markets and the decline of educational equality", *British Educational Research Journal*, 31(5), 605-622.
- Ikeda, M. (2011), "School autonomy and accountability: Are they related to student performance?", *PISA in Focus*, 2011/9.
- Jacob, B.A. and S.D. Levitt (2003), "Rotten apples: An investigation of the prevalence and predictors of teacher cheating", *Quarterly Journal of Economics*, 118(3), 843-878.
- Kamens, D.H. and C.L. McNeely (2010), "Globalization and the growth of international educational testing and national assessment", *Comparative Education Review*, 54(1), 5-25.
- Kijima, R. and J. Leer (2016), "Legitimacy, state-building, and contestation in education policy development: Chile's involvement in cross-national assessments", in W.C. Smith (Ed.) *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice* (pp. 43-62), Oxford, Symposium Books.
- Kornhaber, M.L. (2004b), "Appropriate and inappropriate forms of testing, assessment, and accountability", *Educational Policy*, 18(1), 45-70.
- Lashway, L. (2001), *The New Standards and Accountability: Will Rewards and Sanctions Motivate America's Schools to Peak Performance?* Washington, DC, ERIC Clearinghouse of Educational Management.
- Lee, J. (2008), "Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies", *Review of Educational Research*, 78(3), 608-644.
- Levin, H. (1992), "Market approaches to education: Vouchers and school choice", *Economics of Education Review*, 11(4), 279-285.
- Levin, H.M., and C.R. Belfield (2003), "The marketplace in education", *Review of Research in Education*, 27, 183-219.
- Lewin, T. and J. Medina (2003), "To cut failure schools shed students", *The New York Times*, p. A1.
- Lipman, P. (2004), *High-Stakes Education: Inequality, Globalization, and Urban School Reform*, New York, Routledge.
- McNeil, L. and A. Valenzuela (2001), "The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric", in Orfield, G. and M.L. Kornhaber (Eds.) *Raising Standards or Raising Barriers? Inequality and High-Stakes Testing in Public Education* (pp. 127-150), New York, Century Foundation Press.
- Natriello, G. and A.M. Pallas (1999), "The development and impact of high stakes testing", presented at the High Stakes K-12 Testing conference, New York, NY, 4 December 1998.
- Nichols, S. and D. Berliner (2007), *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*, Cambridge, MA, Harvard Education Press.
- OECD (2012), *PISA 2009 Technical Report*, PISA, OECD Publishing <http://dx.doi.org/10.1787/9789264167872-en>.
- OECD (2011), *Education at a Glance 2011: OECD Indicators*, OECD Publishing, <http://dx.doi.org/10.1787/eag-2011-en>.

- Ooghe, E. and E. Schokkaert (2013), "School accountability: Can we reward schools and avoid pupil selection?", *IZA Discussion Paper*, No. 7420.
- Paris, S. and A. McEvoy (2000), "Harmful and enduring effects of high-stakes testing", *Issues in Education*, 6(1), 145-159.
- Phelps, R.P. (2012), "The effect of testing on student achievement, 1910-2010", *International Journal of Testing*, 12(1), 21-43.
- Raudenbush, S.W. and A.S. Bryk (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Thousand Oaks, CA, Sage Publications.
- Reback, R. (2008), "Teaching to the rating: School accountability and the distribution of student achievement", *Journal of Public Economics*, 92, 1394-1415.
- Rentner, D.S., C. Scott, N. Kober, N. Chudowsky, V. Chudowsky, S. Joftus et al. (2006), *From the Capital to the Classroom: Year 4 of the No Child Left Behind Act*, Washington, DC, Center on Education Policy.
- Rosenkvist, M.A. (2010), "Using student test results for accountability and improvement: A literature review", *OECD Education Working Papers*, No. 54, Paris, OECD Publishing.
- Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York, John Wiley.
- Sahlberg, P. (2010), "Rethinking accountability in a knowledge society", *Journal of Educational Change*, 11, 45-61.
- Schultz, T.W. (1961), "Investment in human capital", *The American Economic Review*, 51(1), 1-17.
- Smith, W.C. (Ed.) (2016a), *The Global Testing Culture: Shaping Educational Policy, Perceptions, and Practice*, Oxford, Symposium Books.
- Smith, W.C. (2016b), "An introduction to the global testing culture", in Smith, W. (Ed.), *The Global Testing Culture: Shaping Educational Policy, Perceptions, and Practice* (pp. 7-24), Oxford, Symposium Books.
- Smith, W.C. (2016c), "Exploring educator based testing for accountability: National Testing Policies and student achievement", in Burns, T. and F. Koester (Eds.), *Governing Education in a Complex World* (pp. 73-91), Paris, OECD Publishing.
- Smith, W.C. (2016d, Forthcoming), "National testing policies, school practices, and student outcomes: An analysis using data from PISA 2009", *Education Working Papers #xx*. Paris, OECD Publishing.
- Smith, W.C. (2014), "The global transformation toward testing for accountability", *Education Policy Analysis Archives*, 22(116), <http://dx.doi.org/10.14507/epaa.v22.1571>.
- Smith, W.C. and J. Rowland (2014), "Parent trigger laws and the promise of parental voice", *Journal of School Choice*, 8(1), 94-112.
- Smeed, J. and M. Victory (2010), "Testing for accountability", *TLN Journal*, 17(3), 28-29.
- Somerset, A. (2016), "Questioning across the spectrum: Pedagogy, selection examinations and assessment systems in low-income countries", in Smith, W.C. (Ed.), *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice* (pp. 171-192), Oxford, Symposium Books.
- Springer, M.G. (2008), "The influence of an NCLB accountability plan on the distribution of student test score gains", *Economics of Education Review*, <http://dx.doi.org/10.1016/j.econedrev.2007.06.004>.
- Stotsky, S. (2000), *What's at Stake in the K12 Standards War?*, New York, Peter Lang Publishing.
- Subedi, B.R. (2005), "A demonstration of the three-level hierarchical generalized linear model applied to educational research", *Dissertation*, Department of Educational Psychology and Learning Systems, Florida State University.
- Supovitz, J. (2009), "Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform", *Journal of Educational Change*, 10, 211-227.
- Teelken, C. (1999), "Market mechanisms in education: School choice in the Netherlands, England and Scotland in a comparative perspective", *Comparative Education*, 35(3), 283-302.
- Volante, L. (2007), "Evaluating test-based accountability perspectives: An international perspective", Paper presented at the Association for Educational Assessment – Europe, Stockholm, Sweden.
- West, A., H. Pennell and P. Noden (1998), "School admissions: Increasing equity, accountability and transparency", *British Journal of Educational Studies*, 46(2), 188-200.
- Wiliam, D. (2010), "Standardized testing and school accountability", *Educational Psychologist*, 45(2), 107-122.

- Wiseman, A. (2010), "The uses of evidence for educational policymaking: Global contexts and international trends", *Review of Research in Education*, 34, 1-24.
- Woessman, L. (2016), "The importance of school systems: Evidence from international differences in student achievement", *IZA Discussion Paper No. 10001*.
- Woessman, L. (2007), "International evidence on school competition, autonomy, and accountability: A review", *Peabody Journal of Education*, 82(2-3), 473-497.
- Woessman, L. (2004), "The effect heterogeneity of central exams: Evidence from TIMSS", TIMSS-Repeat and PISA, *CESifo Working Paper No. 1330*.
- Wong, M., T.D. Cook and P.M. Steiner (2009), "No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series", *Northwestern University Institute for Policy Research Working Paper Series WP-09-11*.
- Yarema, C.H. (2010), "Mathematics teachers' views of accountability testing revealed through lesson study", *Mathematics Teacher Education and Development*, 12(1), 3-18.

APPENDIX A

National Testing Policy: Classification of 2009 PISA Participants

Summative Testing Policy	Evaluative Testing Policy	Punitive Testing Policy
Albania	Australia	Chile
Argentina	Azerbaijan	Hong Kong-China
Austria	Belgium (Flemish)	Hungary
Belgium (French and German)	Brazil	Latvia
Bulgaria	Canada (Ontario)	Macao-China
Canada (except Ontario)	Colombia	Mexico
Chinese Taipei	Denmark	Portugal
Croatia	Estonia	Shanghai-China
Finland	Iceland	United States
France	South Korea	
Germany	Lithuania	
Indonesia	Netherlands	
Ireland	Norway	
Israel	Poland	
Italy	Qatar	
Japan	Romania	
Jordan	Singapore	
Kazakhstan	Slovenia	
Kyrgyz Republic	Sweden	
Luxembourg	Trinidad and Tobago	
Montenegro	Turkey	
New Zealand	United Kingdom	
Peru		
Russian Federation		
Serbia		
Slovak Republic		
Spain		
Thailand		
Tunisia		
Uruguay		

Source: Smith (2016d).

Efficiency and contestability in emerging market banking systems

by

Christian Daude and Julien Pascal*

This paper explores some of the potential determinants of efficiency and contestability in the banking systems of major emerging countries, using a sample of the 24 countries over the period 2004-13. Stochastic and data envelopment analyses are used to estimate national levels of efficiency, while market contestability is captured through the H-statistic. Panel data econometric methods are used to determine potential drivers of both efficiency and market contestability, which provides the basis for an evaluation of potential complementarities and trade-offs between these two dimensions.

JEL classification: E44, G14, G21, G28, L11

Keywords: Banking, competition, efficiency, emerging markets

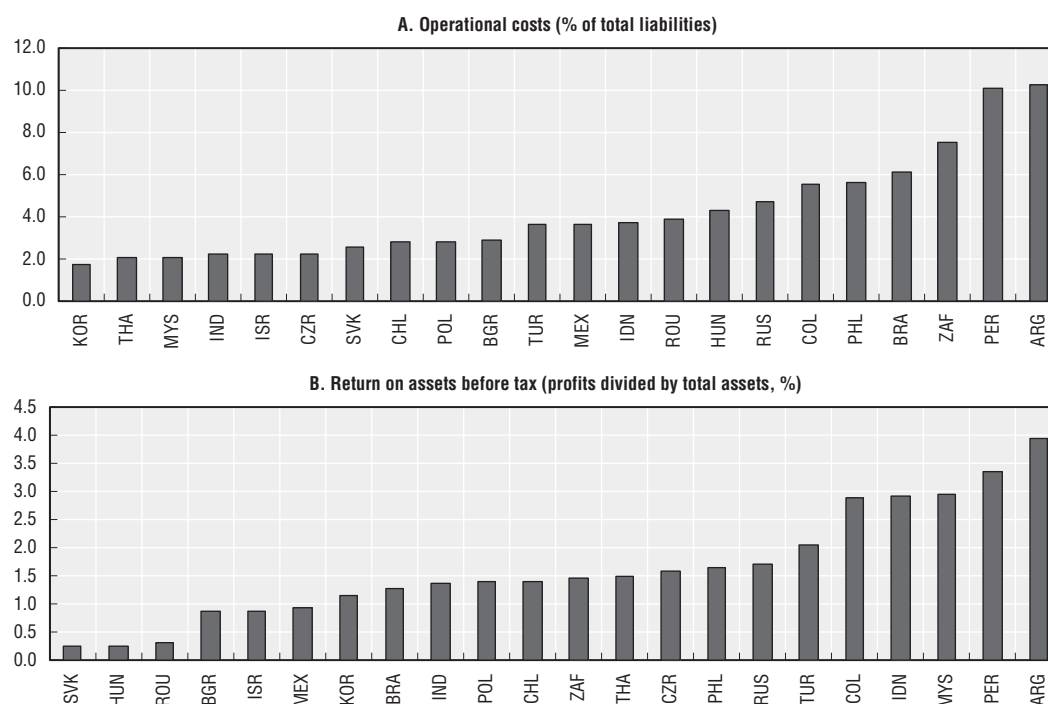
* The authors are from the OECD Economics Department (christian.daude@oecd.org) and Sciences Po (julien.pascal@sciencespo.fr), respectively. They would like to thank Jens Arnold, Andrés Fuentes, Eduardo Olaberria, Alvaro Pereira, Piritta Sorsa, participants at the brownbag seminar of the OECD Economics Department and two anonymous referees for helpful comments and suggestions. The views expressed are those of the authors and do not reflect those of the OECD or its member countries.

1. Introduction

A growing body of evidence suggests that financial markets have a powerful effect on economic development (Demirguç-Kunt and Levine, 2008). By linking investors and savers, banks play a crucial role in the allocation of resources, diversifying risks and reducing information friction in credit markets. Limited access to loans for individuals and firms hinders investment in human and physical capital, innovation and productivity growth. Besides long-run growth, the empirical evidence also shows that access to finance plays an important role in shaping economic disparities between individuals (Cihák et al., 2012). Restrictions to financial access reduce economic opportunities for those lacking of parental wealth, social status and political connections. Furthermore, a well-functioning financial system usually reduces the gap between the lower and the upper tail of the income distribution in developing and emerging market economies (Beck et al., 2007).

International comparisons shed light on potential low efficiency and contestability of the banks operating in major emerging countries relative to a selected sample of OECD countries with similar levels of development. Operational costs for banks belonging to the first category are high when compared with the second category (Figure 1, Panel A), which might reflect an efficiency gap. At the same time, return on assets before tax in non-OECD countries generally

Figure 1. **Operational costs and profitability in emerging and selected OECD countries in 2011**



Source: Authors' calculations based on Bloomberg and SNL Financial; World Bank, GFDD, IMF for Chile in Panel B.

exceeds the levels reached by OECD banking markets (Figure 1, Panel B). While high levels of profitability are generally a sign of stability and solvency of the banking system, an excessive capacity to make profits is usually also a sign of low contestability. While both indicators are useful to get first insights on different levels of efficiency and contestability among countries, they ignore country-specific factors and do not fully capture banks' inputs and outputs. This paper aims at delving further into the estimations of efficiency and market contestability, as well as identifying the channels through which they interact.

Contestability measures the degree to which barriers to entry and exit restrict new entries, such that contestable markets have limited barriers. This generally ensures competitive behaviour by market participants. More contestability usually ensures a higher level of efficiency, defined as the combination of technical, allocative and scale efficiency – terms defined below. The levels and evolution of efficiency in major emerging countries, as well as their drivers, are explored in Section 2, using stochastic and data envelopment analyses on a panel of commercial banks from 23 countries between 2004 and 2013. Section 3 evaluates the degree of contestability for the same sample. The last section explores the link between efficiency and contestability and summarises the main findings.

2. Cost efficiency estimation

This section presents estimates of technical efficiency for a sample of commercial banks for the period 2004-13 in 23 countries using a stochastic frontier model. A data envelopment analysis is also performed to check the robustness of the resulting technical and allocative efficiency. Both methods highlight the existence of a gap in technical efficiency for non-OECD countries with respect to OECD countries.

2.1. Technical Efficiency, Allocative Efficiency and Scale Efficiency

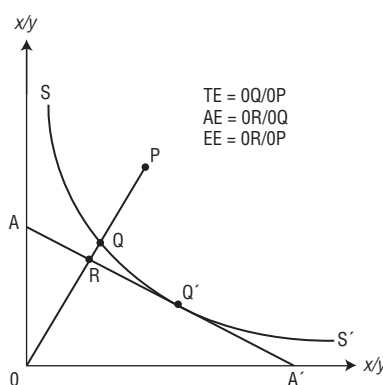
It is important to give a clear definition of efficiency. Modern efficiency measurement begins with Farrell (1957), who decomposes efficiency into two components: *technical efficiency* (TE) and *allocative efficiency* (AE). For the rest of the paper, we adopt the following definitions:

- Technical efficiency is observed when, given a set of inputs, a firm is not able to increase further the amount of output produced, meaning that this firm is located on the production possibility frontier.
- Allocative efficiency is observed when, among all possible combinations of inputs within the production possibility set, a firm chooses the particular combination that minimizes total costs. Therefore, for any amount of output that a firm wishes to produce, it chooses the optimal allocation of inputs.

To illustrate this two definitions, one could consider a firm using two inputs x_1 and x_2 to produce a single output y under constant returns to scale (Figure 2). The fully efficient firm, represented by the unit isoquant SS' , allows one to determine technical and allocative efficiencies. A firm producing one unit of y , using quantities of inputs represented by the point P , is incurring a loss in terms of technical efficiency because it could reduce its inputs to point Q while maintaining its level of production. In this example, technical efficiency is measured by the ratio:

$$TE = OQ/OP.$$

A firm with inputs represented by the point Q could still improve its efficiency by moving to point Q' , where the absolute value of the marginal rate of substitution is equal to

Figure 2. **Technical and Allocative Efficiencies**

Note: For a more developed description of efficiency, see Coelli (1996) on which the previous example is based.

Source: Coelli (1996).

the ratio of input prices represented by the slope of the line AA' . This capacity of improvement is measured by the ratio (AE), which is the allocative efficiency of firm operating at P:

$$AE = OR/OQ.$$

The (total) economic efficiency captures both sources of improvement by indicating how much cost could be reduced if the production were to occur at the technical and allocative efficient point Q' :

$$EE = \frac{OR}{OP} = \frac{OQ}{OP} \times \frac{OR}{OQ} = TE \times AE.$$

Figure 2 illustrates the case in which a firm is producing under constant returns to scale. If a firm has variable returns to scale, this gives rise to a new question: what is the optimal scale of production? The definition of scale efficiency captures this dimension:

- Scale efficiency is observed when a firm produces at the most efficient scale of output or, in other words, when it cannot increase its profits by increasing or decreasing its production level. Scale efficiency is not a matter of allocating different inputs, but defining the optimal scale of production of the final good.

In practise, the efficient frontier SS' represented in Figure 2 is unknown, and parametric as well as non-parametric methods are commonly used to estimate it. The subsequent analysis relies first on a parametric method using a stochastic frontier model to determine technical efficiency in national banking sectors. Second, a non-parametric data envelopment analysis is used to assess the robustness of results and to quantify potential gains from scale efficiency.

2.2. An international comparison based on stochastic frontier estimates

A widespread approach to measuring the efficiency of banks is to estimate a stochastic frontier model, which is motivated by the fundamental idea that no economic agent can exceed the efficient frontier and that deviations from this ideal set of input-output combinations represent the individual inefficiencies in the production process. Stochastic frontier models allow individual observations to be above the efficient frontier because of idiosyncratic disturbances. Thus, the efficiency frontier is not just fixed by the most efficient unit in the sample, but also by taking into account the stochastic feature of the data generating process. Hence, these models decompose the error term in two components: one capturing the noise associated with individual observations, and another related to

production inefficiencies. Since, Meeusen and van den Broeck (1977) and Aigner et al. (1977), this class of models has become a popular tool for efficiency analysis.

In the empirical literature on bank efficiency, two dominant approaches have been used to model cost functions depending on the theoretical view concerning the role of deposits. On the one hand, the *production approach* sees banks as using labour and capital to produce deposits and loans, considering them as an output in the production process. On the other hand, the *intermediation approach* puts emphasis on the brokerage role of financial institutions, creating links between savers and investors. According to the intermediation approach, banks are collecting deposits to transform them into financial instruments exchangeable on financial markets. Deposits are then considered as an input. While the production approach is usually considered to be well-suited for modelling branches within the same bank, the intermediation approach is considered more relevant for performing efficiency comparisons (Kosak and Zoric, 2009). Therefore, the present paper, which aims at measuring and comparing efficiency across national banking systems, adopts the intermediation approach. Deposits are an input, used to produce financial goods (earning assets, which are defined as the sum of total gross loans and securities).

In a first step, technical efficiencies of banks are estimated using a time-varying inefficiency model proposed by Battese and Coelli (1992), which is well-suited for panel data analysis. In particular, the following Cobb-Douglas cost function¹ specification is assumed:

$$\ln(TC_{it}) = \beta_0 + \sum_{j=1}^3 \beta_j \ln(p_{it}^j) + \beta_4 \ln(Y_{it}) + \theta_1 \frac{\text{total equity}_{it}}{\text{total assets}_{it}} + \theta_2 \frac{\text{total loans}_{it}}{\text{deposits}_{it}} + \text{time trend} + v_{it} + u_{it},$$

TC = operating expenses + interest expenses, a measure of total cost,

$p^1 = \frac{\text{personnel expenses}}{\text{total assets}}$, proxy for price of labour,

$p^2 = \frac{\text{operating expenses} - \text{personnel expenses}}{\text{total assets}}$, proxy for price of equipment,

$p^3 = \frac{\text{interest expenses}}{\text{total liabilities}}$, proxy for price of fundings,

Y_{it} = earning assets, capturing banks' output,

$v_{it} \sim \text{iid} N(0, \sigma_v^2)$ $u_i \sim \text{iid} N^+(0, \sigma^2)$, [1]

$u_{it} = \exp[-\eta(t - T)]u_i$, with T being the last period of the sample.

The idiosyncratic error (v_{it}) is associated with the randomness of the data generating process, while u_{it} is a non-negative half-normal random variable of capturing the distance from the cost efficient frontier, that is the degree of technical inefficiency.² Furthermore, the model imposes further structure on the evolution of the degree of technical inefficiency, as the non-negative inefficiency effects u_{it} , decrease, remain constant or increase as t increases if $\eta > 0$, $\eta = 0$ or $\eta < 0$. The measure of cost efficiency is provided by $CE_{it} = e^{-u_{it}}$, ranging from 0 for fully inefficient firms to 1 for firms on the cost efficient frontier.³ More precisely, this cost efficiency measures *technical efficiency*, as previously defined. The specification is fairly close to other studies on bank efficiency (see for instance Kosak and Zoic, 2009 and Tabak et al., 2011). However, the price of equipment used is slightly different. It is common to use fixed assets instead of total assets in the denominator of p_2 , as it represents tangible assets – property, plant and equipment. However, fixed assets in the

sample appear to be poorly measured for Latin American countries. For example, the correlation between the number of offices and fixed assets is -0.009, while the correlation between the number of automatic teller machines and fixed assets is just 0.1. Using total assets instead of fixed assets, these correlations increase to 0.48 and 0.44, respectively. Besides, using fixed assets creates a high dispersion in the proxy for equipment prices. It is hard to believe that banks are facing extremely dispersed prices for their properties, plants and equipment. Therefore, the normalisation by total assets was chosen as a second best solution.

Further to considering a well-behaved cost function, one should impose homogeneity in factor prices⁴ and subtract $\ln(p_{it}^3)$ from both sides of [1] to obtain the equation used in the estimation:

$$\ln(TC_{it} / p_{it}^3) = \beta_0 + \sum_{j=1}^2 \beta_j \ln(p_{it}^j / p_{it}^3) + \beta_4 \ln(Y_{it}) + \theta_1 \frac{\text{total equity}_{it}}{\text{total assets}_{it}} + \theta_2 \frac{\text{total loans}_{it}}{\text{deposits}_{it}} \quad [2]$$

+time trend + v_{it} + u_{it} .

A common cost frontier is estimated for commercial banks of the 23 countries included in the sample. Of course, the cost frontier analysis depends on the units included in it, as the degree of inefficiency is measured by the distance to the cost efficient frontier, which is influenced by the best performing units. Therefore, not all OECD countries were included, but a selected sample of OECD countries that have closer levels of overall development to other emerging countries.⁵ The sample of OECD members includes Chile, Czech Republic, Hungary, Israel, Poland, Slovakia, Slovenia, South Korea and Turkey. While to some extent these choices were influenced by the quality of the data available on the mentioned countries, their respective banking sectors have average returns on assets higher than the OECD average, closer to emerging countries' figures, indicating a closer proximity to the former ones. Besides, to control for heterogeneity across countries in terms of regulatory conditions and risk preferences, the ratio of equity to assets and the ratio of total loans to deposits are included (Tabak et al., 2011). Furthermore, a time trend is included to allow for technical progress (Battese and Coelli, 1992). Finally, in order to test the robustness of the estimation, a second specification for the price of labour is implemented by normalising labour expenses by the number of employees.

The sample was built using three sources. Bloomberg and SNL Financial were jointly used to extract information on the balance sheet and income statements for each bank of the sample. As the number of employees was often missing, banks' financial statements were also directly consulted to increase the effective sample size.⁶ The full sample is composed of 3 407 observations. Even after examining banks' financial statements, only 2 625 observations reporting the number of employees are available. Summary statistics for the variables in equation [1] are presented in Table 1. The number of banks by country per year is reported in Table A.1 in the Appendix.

Equation [2] is estimated by maximum likelihood, where the likelihood function is expressed in terms of the variance parameters, $\sigma_s^2 = \sigma_v^2 + \sigma^2$ and $\gamma = \sigma^2 / \sigma_s^2$. In practice, the estimation was implemented using the R package *frontier* (Coelli and Henningsen, 2013). Estimates for both specifications are presented in Table 2.

Overall, cost efficiency shows signs of limited improvements during the period 2004-13 (see Figure 3). Under specification 1, the η coefficient is positive and significantly different from zero, indicating an increasing trend in efficiency, even after controlling for technological

Table 1. **Summary statistics**

	Price of funds	Price of labour proxy 1	Price of labour proxy 2	Price of equipment	Total cost	Equity/assets	Loans/deposits
Number of observations	3 407	3 407	2 625	3 407	3 407	3 407	3 407
Min	0.0005	0.0003	0.0007	0.00002	0.7897	0.0179	0.0002
Max	0.3968	0.2812	2.1938	0.2831	65129.0	0.4934	9.7430
Range	0.3962	0.2809	2.1931	0.2831	65128.2	0.4754	9.7428
Median	0.0391	0.0123	0.0260	0.0147	460.4	0.0931	0.8870
Mean	0.0429	0.0165	0.0360	0.0203	1577.0	0.1046	1.0173
Std.dev	0.0266	0.0151	0.0603	0.0209	4353.0	0.0525	0.7051

Note: Proxy 1 for price of labour = labour expenses/total assets, proxy 2 the price of labour = labour expenses/number of employees.

Source: Authors' calculation based on data from Bloomberg and SNL Financial.

Table 2. **Stochastic frontier analysis**

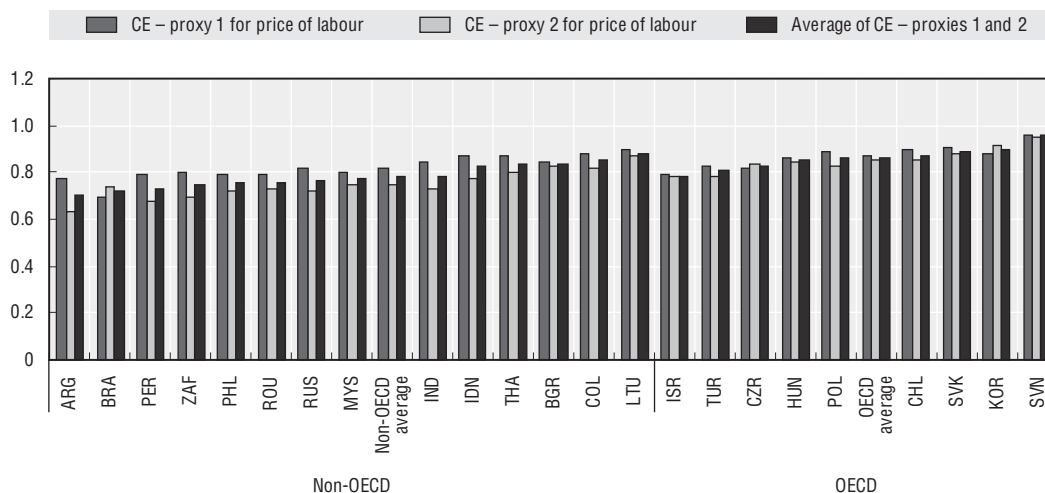
	Proxy 1 for price of labour				Proxy 2 for price of labour			
	Estimate	Std. Error	z value	Pr(> z) (%)	Estimate	Std. Error	z value	Pr(> z) (%)
(Intercept)	1.54	0.03	44.46	0.0	1.66	0.04	37.53	0.0
Log(earning assets)	0.96	0.00	249.05	0.0	0.92	0.01	169.51	0.0
Log(price labour)	0.22	0.01	26.77	0.0	0.12	0.01	17.24	0.0
Log(price equipment)	0.29	0.01	41.99	0.0	0.34	0.01	50.84	0.0
Equity/assets	-0.70	0.09	-8.12	0.0	-0.51	0.09	-5.52	0.0
Loans/deposits	-0.04	0.01	-6.98	0.0	-0.05	0.01	-6.55	0.0
Time trend	0.01	0.00	3.65	0.0	0.00	0.00	-1.88	6.0
σ_s^2	0.08	0.01	14.54	0.0	0.13	0.01	10.81	0.0
γ	0.79	0.02	49.69	0.0	0.88	0.01	72.85	0.0
η	0.04	0.01	5.39	0.0	-0.01	0.01	-0.86	38.8
Log-likelihood value	1 485.646				1 261.594			
Number of observations	3 407				2 625			

Note: Proxy 1 for price of labour = labour expenses/total assets, proxy 2 the price of labour = labour expenses/number of employees.

change (see Table 1). However, under specification 2, it is not significantly different from zero, indicating an invariant rather time-invariant cost efficiency. Given that the η coefficient is not robust to specification changes, one has to be cautious when interpreting its value.

Differences in estimates between the two specifications do not stem from a composition effect of the sample. As the number of employees is more often missing than information about total assets, the sample for proxy 1 is larger. However, when restricting the sample for both specifications to the same observations (2 625 observations) the results reported in Table 1 and Figure 3 are not altered significantly. Rankings across countries are preserved and cost efficiency scores at country's level are not significantly affected (with an average change of less than 0.01 in absolute values).

A series of robustness tests – not reported due to space limitations – show that the results presented here are robust. For example, the previous results are not affected by removing the technological trend; neither by allowing the inefficiency terms u_{it} to be a truncated normal random variable instead of a half-normal one. Robustness test also included estimating the model without controlling for differences in equity to total assets ratio and loans to deposits ratio, or taking the logarithm of these control variables. Finally, a translog cost function was used instead of a Cobb-Douglas. All these alternative

Figure 3. **Stochastic frontier analysis, commercial banking sectors' average cost efficiency, 2004-13**

Note: Cost efficiency (CE) is an input-oriented measurement of technical efficiency. It measures the relative distance from the efficient frontier. It ranges between 1 and 0 for a fully efficient and a fully inefficient firm respectively. Proxy 1 for price of labour = labour expenses/total assets, proxy 2 the price of labour = labour expenses/number of employees. National estimates are computed as simple average of domestic banks.

Source: Authors' calculation based on data from Bloomberg and SNL Financial.

specifications did not affect the overall results exposed previously, with the exception of the translog cost function which slightly increased the average cost efficiency for the whole sample, although it did not produced substantial differences in the ranking across countries. Thus, the relative rank between countries in terms of cost efficiency is probably more robust than the absolute value of the cost efficient coefficient, which depends to some extent on the functional form assumed. In this sense, a data envelopment analysis allows for fewer restrictions on the functional form.

2.3. Further insights on cost efficiency from a data envelope analysis

The previous analysis of efficiency relied on the assumption of a parametric form for the cost function. This restriction is released by evaluating the efficient frontier using a data envelope analysis (DEA), which is a non-parametric linear programming approach to frontier estimation. This second analysis gives further insights about the robustness of the stochastic frontier analysis (SFA) findings and allows quantifying potential gains from scale efficiency. In particular, the model is solved for the following three sets of linear systems:

$$\min_{\{\theta_i, \lambda\}} \theta_i \quad (\text{system 1})$$

$$\text{s.t. } -\mathbf{q}_i + \mathbf{Q}\lambda \geq 0,$$

$$\theta_i \mathbf{x}_i - \mathbf{X}\lambda \geq 0,$$

$$\lambda \geq 0,$$

$$\min_{\{\theta_i, \lambda\}} \theta_i \quad (\text{system 2})$$

$$-\mathbf{q}_i + \mathbf{Q}\lambda \geq 0,$$

$$\theta_i \mathbf{x}_i - \mathbf{X}\lambda \geq 0,$$

$$\mathbf{1}'\lambda = 1,$$

$$\lambda \geq 0,$$

$$\begin{aligned}
& \min_{\{\theta_i, \lambda\}} \theta_i \quad (\text{system 3}) \\
& -\mathbf{q}_i + \mathbf{Q}\lambda \geq \mathbf{0}, \\
& \theta_i \mathbf{x}_i - \mathbf{X}\lambda \geq \mathbf{0}, \\
& \mathbf{1}\mathbf{1}'\lambda \leq \mathbf{1}, \\
& \lambda \geq \mathbf{0}.
\end{aligned}$$

This set-up is very much in line with the existing literature on DEA.⁷ There are N inputs and M outputs for each of I firms.⁸ Firm i 's input and output vectors are represented by $\mathbf{x}_i$ and $\mathbf{q}_i$ respectively, while the I firms are represented by the $N \times I$ input and the $M \times I$ output matrices $\mathbf{X}$ and $\mathbf{Q}$. θ_i is a scalar of efficiency score for the i -th firm which satisfies $\theta_i \leq 1$ ($\theta_i = 1$, for a fully efficient firm), λ is a $I \times 1$ vector of constants and $\mathbf{1}\mathbf{1}'$ is an $I \times 1$ vector of ones.

System 1 seeks to radially contract the input vector $\mathbf{x}_i$, while remaining within the feasible input set. The inner-boundary of this set is a piece-wise linear isoquant determined by the observed data points. The radial contraction of the input vector $\mathbf{x}_i$ is a projected point on the surface of the production technology T which defines a production set that is closed, convex, and exhibits constant returns to scale (CRS).

$$T = \{(\mathbf{x}, \mathbf{q}) : \mathbf{q} \leq \mathbf{Q}\lambda, \mathbf{x} \geq \mathbf{X}\lambda\}.$$

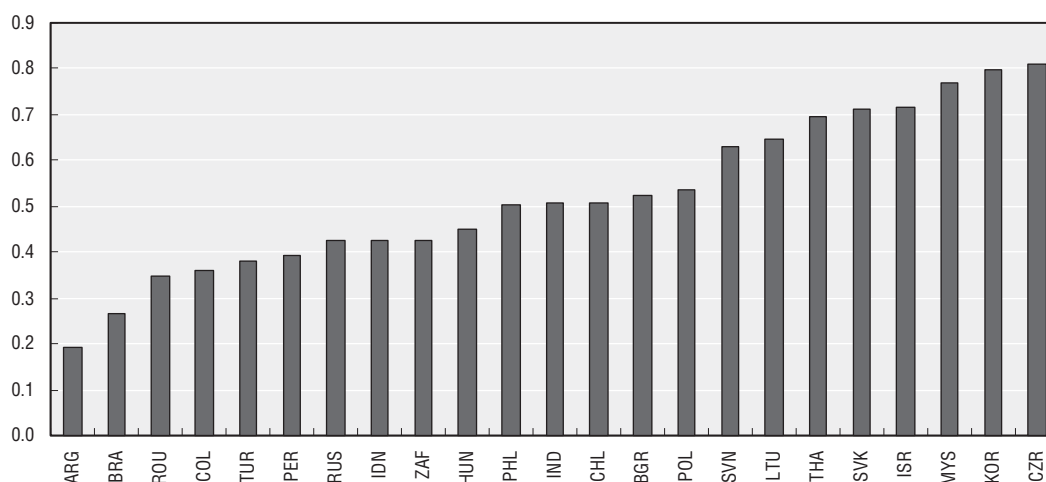
System 2 allows for variable returns to scale (VRS) by adding the convexity constraint $\mathbf{1}\mathbf{1}' = 1$, which creates a convex hull of intersecting planes that envelope the data points. This constraint ensures that inefficient firms are only compared against firms of similar size, such that the projected point of a firm on the DEA frontier is a convex combination of observed firms. Efficiency of scale is determined by conducting both a CRS (system 1) and a VRS DEA (system 2). If there is a difference in the CRS and VRS technical efficiency scores for a particular firm, this indicates scale inefficiency. To determine the nature of scale efficiency, one has to substitute the restriction $\mathbf{1}\mathbf{1}' = 1$ with $\mathbf{1}\mathbf{1}' \leq 1$ (system 3) to obtain a frontier exhibiting non-increasing returns to scale (NIRS). If VRS and NIRS technical efficiency scores are unequal (equal) for the i -th firm, then it operates under increasing (decreasing) returns to scale.

Inputs and output are defined in line with the intermediation approach adopted previously, to insure consistency with the stochastic frontier analysis in the following way:

$$\begin{aligned}
\mathbf{x}^1 &= \text{personnel expenses}, \\
\mathbf{x}^2 &= \text{operating expenses} - \text{personnel expenses}, \\
\mathbf{x}^3 &= \text{interest expenses}, \\
\mathbf{q}^1 &= \text{earning assets}.
\end{aligned}$$

The three systems of equations previously defined are solved⁹ for our dataset – the same sample that was used for the stochastic frontier analysis – on a yearly-basis (Table A.1). However, one should not consider changes in efficiency scores as measuring a trend, but rather as a series of cross-sectional comparisons.

Technical efficiency for most of the countries include in the sample appear to be smaller when measured with the DEA method (Figure 4). The “real” technical efficiency is unobservable and probably lies between the two estimates. Results presented in Figure 4 are robust to several tests. In particular, estimating the DEA using a pooled sample including

Figure 4. **Technical efficiency with variable return to scale, average 2004-13**

Note: Variable return to scale efficiency (VRS TE) is a measure of the relative distance from the efficient frontier. To control for possible returns to scale, banks are only compared against banks of similar size. It ranges between 1 and 0 for a fully efficient and a fully inefficient firm respectively.

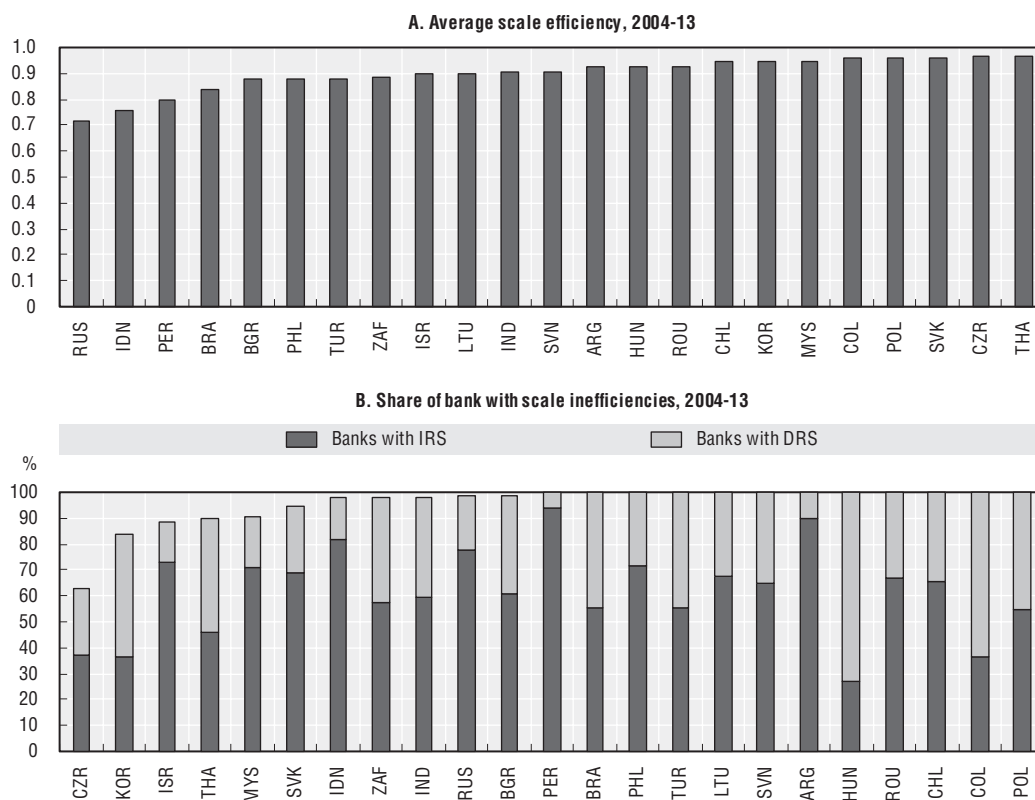
Source: Authors' calculations based on data from Bloomberg and SNL Financial.

all observations instead of the year-by-year estimation, as well as using a bootstrapping algorithm to correct for the bias induced by using observed data to compute the efficient frontier (Simar and Wilson, 1998), does not alter the results significantly. In both cases, the resulting technical efficiency is slightly smaller for the whole sample (less than 0.1 point shift downward on average), but the ranking across countries is preserved. In the first case, the small decrease in technical efficiency comes from the slight expansion of the technological frontier over time, while in the second case it comes from assuming there is noise in the frontier estimation.

A further concern may be related to the *intermediation approach* adopted in this paper. Some countries may have banking systems oriented toward generating deposits services for customers rather than producing loans and other financial products. To not penalise such countries, deposits are included as a second type of output in the production model. This produces only minor changes for countries' level of efficiency (see Figure A.1). Thus, the efficiency gap between non-OECD and OECD countries is not to be explained by considerations regarding inputs and outputs definitions.

There is a high level of scale efficiency in most countries included in the sample, indicating that the main source of differences across countries lies in technical efficiency changes rather in scaling operations differently: the loss incurred is small as the aggregate score for most emerging countries is close to 0.90 (Figure 5, Panel A). Countries for which significant gains could be achieved in rescaling their activities are Russia, Indonesia, and Peru; with a clear majority of banks with increasing return to scale (see Figure 5, Panel B).

The DEA and SFA give different pictures of the magnitude of the efficiency gaps. There are at least three reasons why this could happen. First, there is a fundamental difference in the way both methods address the estimation of the efficient frontier: the SFA model takes into account the randomness of the production process while the DEA is "deterministic".¹⁰ In SFA, *nobody* is on the frontier; while for the DEA, *somebody has to be* on the frontier. However, this would bias the DEA estimates upwards, such that the

Figure 5. **Data envelopment analysis, scale efficiency**

Note: Scale efficiency measures to what extent banks chose optimally their scale of operations. A scale efficient bank has a scale efficiency equal to 1. A scale efficiency between 1 and 0 indicates that profits could be increased by changing the scale of production.

Source: Authors' calculations based on data from Bloomberg and SNL Financials.

bias-corrected DEA estimates would indicate a lower efficiency, as illustrated above. Second, in the SFA method, one has to assume a functional form for the cost function as well as for the composite error term, while no arbitrary choice is made running a DEA. Thirdly, the SFA fully embraces the panel structure of our dataset, while the DEA used here neglects this aspect.

2.4. Determinants of cost efficiency: regulations and individual bank effects

So far the inefficiency term was unexplained; but it can be endogenised by using the approach developed by Battese and Coelli (1995). This is a one-stage procedure that estimates simultaneously the stochastic frontier, as well as the determinants of the inefficiency terms. Previous studies on banking efficiency have used a two-stage approach, in which the efficiency estimates of a first step are regressed on a vector of explanatory variables in a second stage (Nikiel and Opiela, (2002 and Bonin et al., 2005). However, this approach suffers from potentially substantial biases (Wang and Schmidt, 2002).

The econometric model relies on the Battese and Coelli (1992) model, adding the assumption that the individual inefficiency terms u_{it} – the non-negative truncation (at zero) of the normal distribution – are linearly dependent on a vector $\mathbf{z}_{it}$ of explanatory variables:

$$\ln(TC_{it}) = \beta_0 + \sum_{j=1}^3 \beta_j \ln(p_{it}^j) + \beta_4 \ln(Y_{it}) + \beta_5 \text{time trend} + v_{it} + u_{it}, \quad [3]$$

$$u_{it} = \mathbf{z}_{it}\delta + w_{it} \quad [4]$$

where the random variable w_{it} follows the truncated normal distribution with zero mean and variance σ^2 . The point of truncation is chosen such that $w_{it} \geq -\mathbf{z}_{it}\delta$, hence the inefficiency component u_{it} is a non-negative truncation of $N(\mathbf{z}_{it}\delta, \sigma^2)$. The vectors of unknown coefficients β and δ are simultaneously estimated by maximum likelihood. Cost efficiency estimates are expressed as $CE_{it} = e^{-u_{it}} = e^{-\mathbf{z}_{it}\delta - w_{it}}$.

The vector of explanatory variables used in the analysis of the present paper contains two categories: one that captures national banking structures and regulations and another one that takes into account banks individual characteristics.

The following variables are included within the bank-specific characteristics: i) the ratio of equity to assets; ii) the ratio of total loans to deposits; and iii) the net interest margin defined as the difference between interest income and interest expenses (divided by total assets). Net interest margins are closely related to the lending-deposit spread. One hypothesis was that more efficient banks operate at smaller operating costs while still securing sufficient profits; hence they can reduce their lending-deposit spread.

Some aspects of financial regulation are included among the country-specific variables. To measure the impact of the regulatory framework on bank performances, an index capturing the official power of the supervisory of the banking sector is introduced. This index measures the extent to which the supervisor is able to take specific actions to prevent and correct problems in the banking system and institutions. The data come from three surveys by the World Bank in 2003, 2007 and 2011,¹¹ where senior banking supervisors were asked several questions relative to the power and independence of the national supervisory entity. Values were assigned to each answer to create an index ranging from 0 to 14, with higher values indicating more power for the official supervisory entity to prevent and correct issues (see Box A.1 for the questions used). In addition, an index capturing activity restrictions on banking activities based on the same survey is considered. This index measures the extent to which banks may also engage in securities, insurance and real estate activities. It ranges from 3 to 14; with higher value indicating more restrictions on banking activities (see Box A.2). While a stronger power of supervisory in theory would lead to more efficiency, the impact of activity restrictions is less clear. On the one hand, these restrictions might induce more competition and raise efficiency. On the other hand, they might prevent banks to fully benefit from economies of scope and therefore reduce cost efficiency.

Structural characteristics are likely to have an impact on efficiency scores. First, concentration might have a positive or negative impact on the efficiency of banks. On the one hand, a more concentrated market could create more monopolistic power and therefore reduce incentives for efficiency gains. On the other hand, other mechanisms could induce a short-term positive correlation between concentration and efficiency. For example through means of innovation, firms could temporarily get more market shares, which may lead to higher market concentration, at least in the short-run. Further, a more concentrated ownership may induce better monitoring and positive spill-overs towards profitability and productivity (Claessens and Djankov, 1999). To evaluate this issue, the share of assets detained by the five largest banks is included as an explanatory variable. Second, the percentages of public as well as the share of foreign banks in the national market are considered. A bank is considered as public if more than 50% of its share is owned by the government. An equivalent definition is used to define foreign ownership. The rationale for including these variables is that spill-over effects could arise in a given market. For instance,

foreign banks could facilitate the diffusion of new technologies and management knowledge, enhancing the efficiency of competitors, while the presence of public banks might reduce market competition and efficiency.

To increase the number of years covered in the sample, country-specific variables¹² for the period 2004-11 were obtained by linear interpolation. When interpolating the data between two observations from the World Bank Regulation and Supervision survey (available in 2003, 2007 and 2011), the corresponding assumption is that the underlying variables are continuous (they do not “jump” from one year to another). While this assumption makes intuitive sense for market structure variables that are slow moving objects – e.g. market concentration – one may object that variables capturing regulations objects – e.g. supervisory power – are more prone to be discontinuous. While it is true that a reform happens at a given point in time, one may argue that its implementation is going to take time and that in practice that transition from one regulatory state to another is a smooth one. In practice, since there exist an infinity of lines going through one point, only countries with at least two observations from the World Bank Survey on Financial Regulation were retained. Furthermore, in order to facilitate convergence in the maximum likelihood estimates, the homogeneity assumption in the factor prices for the cost function is relaxed. To take into account both, technical change and time-varying inefficiency effects, a time trend is included in equations [3] and [4]. In equation [3], the time trend captures Hicksian neutral technological change, while in equation [4] it means that inefficiency effects may be a linear function of time (as in Battese and Coelli, 1995). Country-fixed effects are included to control for unobservables that might be correlated with the explanatory variables.

Table 2 reports two specifications of the model described by [3] and [4] depending on the type of proxy for the price of labour used. At the bank level, the results show that the ratio of equity to assets is positively correlated with inefficiency – yet it fails to be statistically significant in the first specification. Thus, on average when a bank increases its leverage (by decreasing its equity to assets ratio), it increases its efficiency. However, it could also be a result of reverse causality: more efficient banks might be able to expand its activities. In line with this result, an increase in the loan-to-deposit ratio is associated with an increase in efficiency. More interestingly, Table 2 also suggests that a decrease in the net margin is associated with an increase in efficiency. This provides evidence consistent with the hypothesis that the high intermediation margins observable in some emerging countries are partly explained by a technical efficiency gap with respect to more efficient OECD counterparts. To put differently, to some extent, high intermediation costs can be viewed as the cost of inefficiency. The two last results are robust as the coefficients associated with loan-to-deposit ratio and net interest margin are statically significant independently of the specification chosen. As pointed out before, it is worth emphasizing that Table 2 documents correlations. To identify causality, an instrumental variable approach would be needed.

To verify to what extent these results are dependent on the interpolation described above, robustness checks were implemented. First, the model defined by [3] and [4] was estimated only using observed data (only in 2003, 2007 and 2011, when available). Second, the model was estimated on a version of the sample where country-specific variables are held fixed between two observations. This situation is equivalent to assuming that reforms are abrupt changes (going from one state to another). Globally, the results presented in Table 2 are not strongly altered. In the first case, the estimates’ signs are unchanged, but their statistical significance is reduced – which is not surprising given that the sample size is

Table 3. **Joint estimation of the cost frontier and inefficiency determinants**

	Proxy 1 for price of labour				Proxy 2 for price of labour			
	Estimate	Std. Error	z value	Pr (> z) (%)	Estimate	Std. Error	z value	Pr (> z) (%)
Stochastic frontier								
Intercept	1.12	0.04	30.10	0.0	0.84	0.06	13.27	0.0
Log(earning assets)	0.99	0.00	490.74	0.0	1.00	0.00	400.76	0.0
Log(price labour)	0.19	0.01	19.26	0.0	0.02	0.01	2.53	1.0
Log(price equipment)	0.25	0.01	29.95	0.0	0.35	0.01	43.21	0.0
Log(price funds)	0.55	0.01	78.71	0.0	0.57	0.01	63.12	0.0
Time trend	-0.01	0.00	-3.90	0.0	0.00	0.00	-1.40	16.0
Inefficiency model								
Intercept	-0.22	0.67	-0.33	74.4	-1.33	0.67	-2.00	5.0
Equity/assets	0.32	0.34	0.95	34.3	1.63	0.32	5.19	0.0
Loans/deposits	-0.66	0.09	-7.21	0.0	-1.30	0.15	-8.93	0.0
Net margin	1.11	0.47	2.36	1.9	2.18	0.31	7.13	0.0
Supervisory power	-0.17	0.04	-4.71	0.0	-0.04	0.03	-1.53	12.0
Activity restrictions	-0.07	0.03	-2.21	2.7	-0.08	0.03	-3.09	0.0
Market concentration	-0.01	0.01	-1.34	18.0	0.02	0.01	2.44	1.0
Public banks share	0.06	0.01	4.18	0.0	0.05	0.01	4.15	0.0
Foreign banks share	-0.04	0.01	-4.87	0.0	0.01	0.01	1.58	12.0
Time trend	-0.01	0.01	-0.75	45.3	-0.02	0.01	-2.23	3.0
Variance parameters								
σ_s^2	0.46	0.07	6.99	0.0	0.12	0.01	15.43	0.0
γ	0.99	0.00	671.69	0.0	0.91	0.01	120.29	0.0
Log-likelihood value	1 002.321				830.038			
Number of observations	1 694				1 694			

Note: Proxy 1 for price of labour = labour expenses/total assets, proxy 2 the price of labour = labour expenses/number of employees. The inefficiency model includes country specific effects. The inefficiency terms u_{it} has the following form $u_{it} = \mathbf{z}_{it}\delta + w_{it}$. Hence a positive sign for a component of δ indicates it tends to increase inefficiency, hence to decrease cost efficiency.

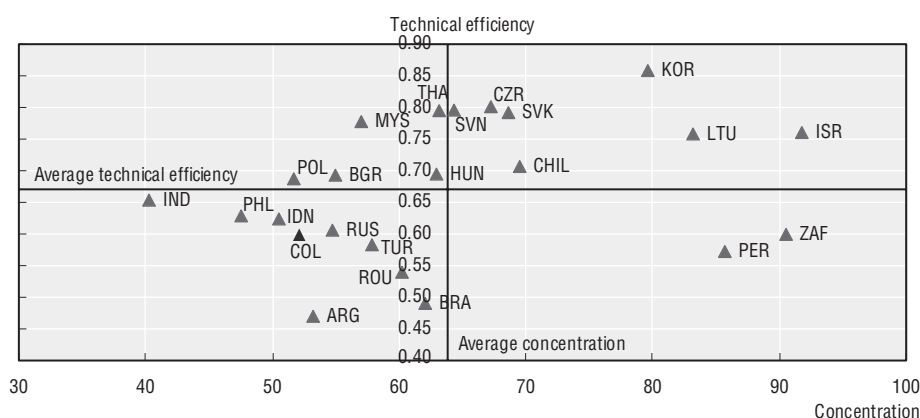
reduced to 638 observations. Only the variable measuring restrictions on banking activities remains significant at a 10% confidence level. In the second case, the statistical significance of market concentration and foreign bank shares are altered, but the estimates' signs remain the same. Overall, results presented in Table 2 are not highly sensitive to the interpolation method used.

There are also three interesting results regarding regulations and market structures. First, the power of the supervising authority is associated with higher banks efficiency. This might reflect that strong supervision helps level the playing field for all competitors and is consistent with findings for OECD economies (Ahrend et al., 2009). Second, restrictions on banking activities are found to be positively correlated with an increase in efficiency. Third, the share of public banks is associated with negative spill-overs among the competitors and higher levels of inefficiency. Thus, a decrease in the share of public banks might increase competitive pressures, resulting in efficiency gains.¹³

Interestingly, the percentage of foreign banks has no clear effect on technical efficiency, as the sign of the associated coefficient depends on the specification chosen. This reflects also the results in the literature regarding the impact of foreign entry on efficiency, which has not reached a consensus on the matter (Wezel, 2010). While in the first specification a higher share of foreign banks is associated with higher levels of efficiency, the alternative specification renders an insignificant coefficient. Thus, the evidence of spill-over effects is somewhat limited.

Finally, market concentration does not have a significant impact on banks efficiency. Concentration of the banking market entails two opposing effects: more efficient banks are more likely to gain market shares over their less efficient counterparts, resulting in higher concentration. However, the increase in concentration may reduce competitive behaviour and lessen incentives on efficiency gains. In the long run, it is hard to determine which of the two effects of concentration dominates at the national market level. Figure 6 shows a slightly positive relation between efficiency and concentration for the period 2004-11. However, the relation is not very strong.¹⁴ Some good performers have concentrated markets (Korea), while the opposite is also true (Peru, South Africa). In this sense, evaluating the differences in contestability across countries, rather than concentration (Berger et al., 2004), as it is probably a leading variable for technical efficiency, might be a more suitable test (see below).

Figure 6. **Concentration and Efficiency, 2004-11**



Note: Technical efficiency is the average of the cost efficiency estimates from the stochastic frontier and data envelopment analyses for the period 2004-11. Concentration is measured as the share of assets detained by the five largest banks of the national banking sector. The represented value is the average of the three point estimates from the World Bank Survey on Financial Regulations (2003, 2007, 2011).

Source: Authors' calculations based on data from Bloomberg, SNL Financial and World Bank Survey on Financial Regulations.

To summarise this section, both stochastic and data envelopment analyses underline an efficiency gap between non-OECD and OECD countries included in the sample. Some factors which may explain this fact have been explored. While causality was not established, interesting correlations have been documented. For instance, the public banks share is negatively correlated with efficiency, a result which is robust to specification changes. The strength of the supervisory authority, as well as the degree to which commercial can engage in securities, insurance and real estate activities, are found to have significant effects on efficiency, although these results are somewhat less robust. However, unmeasured factors might play a primordial role in determining differences across countries. Contestability is certainly one of them. To learn about the competitive structure in which banks operate, the H-statistic is a useful tool.

3. Market contestability and competition

Cost efficiency generally is influenced by the level of contestability. A more contestable market put pressure on banks, which normally react by increasing their efficiency. To determine the degree of competition in national banking systems, an econometric model

of contestability is estimated for each country in our sample. The impact of market structures and regulations are then explored and found to have a significant impact on contestability.

3.1. International comparison of banking sector competitiveness structures

This paper follows the methodology of Panzar and Rosse (1987) to compute an contestability index for a given market. Under the assumptions that firms' cost functions are homogeneous in input prices and that they operate in a symmetric long-run equilibrium, these authors show that the sum of factor price elasticities of the revenue function is equal to 1 in case of competitive equilibrium, while it must be smaller or equal to zero in case of a monopoly (or perfect cartel).¹⁵ Under monopolistic competition, the H-statistic – defined as the sum of factor price elasticities – is an index measuring the degree of contestability ranging between 0 and 1. To put it differently, the magnitude of H can be interpreted as an inverse measure of the degree of monopolistic power. This methodology is applied to each banking system of the countries included in the sample assuming the following functional form:

$$\ln(p_{it}^y) = \alpha_i + \sum_{j=1}^3 \beta_j \ln(p_{it}^j) + \sum_j \theta_j \text{BSF}_{it} + \text{macroeconomic factors}_t + v_{it}, \quad [5]$$

α_i represents bank fixed effects,

$p^1 = \frac{\text{personnel expenses}}{\text{total assets}}$ is a proxy for the price of labour,

$p^2 = \frac{\text{operating expenses} - \text{personnel expenses}}{\text{total assets}}$ is a proxy for the price of equipment,

$p^3 = \frac{\text{interest expenses}}{\text{total liabilities}}$ is a proxy for the price of funding,

$p^y = \frac{\text{interest income}}{\text{total assets}}$ is a proxy for the output price.

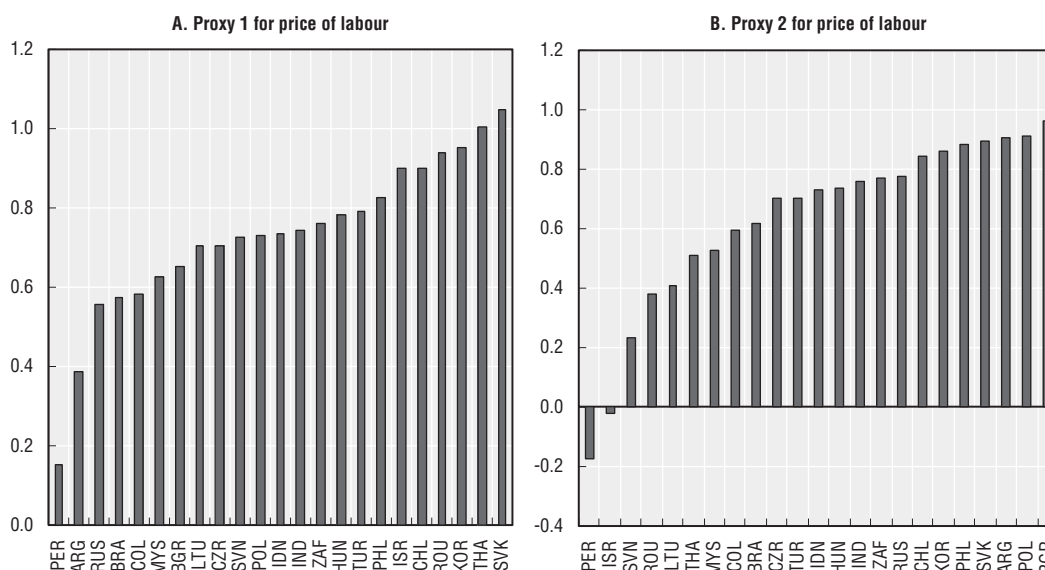
BSF_{it} captures bank specific factors.

The bank fixed effects capture unobserved long-lasting differences in managerial capacities that could affect banks' performances. Following Levy Yeyati and Micco (2007) bank specific factors (fundamentals) that could impact interest incomes without going through the price mechanism are also included. To control for differences in risk taking and capital utilisation, equity to assets, loan-to-deposit and earning assets to total assets ratios are considered. By dividing interest income by total assets, the effects of size on interest revenues are controlled. To take into account variations in country-specific macroeconomic factors that could have influenced banks' performances, the ratio of non-performing loans (capturing variations in risk premiums stemming from differences in loan quality) and GDP growth (proxy for the overall state of the economy) are included in the regressions. Finally, one should also consider the exchange rate of the local currency to US dollars as a control variable. For Latin American countries, these variations have been substantial, with potential impacts on banks' profits coming from operations in dollars. The estimation results by country are reported in Tables A.4 and A.6.

According to the estimates, nearly all banks operate in a monopolistic competition setting. A series of t-tests are performed to test the hypotheses of perfect cartel and perfect competition. Under the first specification in which the price of labour is obtained as the ratio of labour expenses to total assets, the hypothesis of perfect cartel is rejected at a 1% significance level for all countries at the exception of Argentina and Peru (see Table A.5). The

hypothesis of perfect competition cannot be rejected at a 10% significance level for six countries, in particular for Thailand, Slovakia, Romania and Israel. When using a second proxy for labour price, the perfect cartel hypothesis cannot be rejected at a 10% significance level for Israel and Peru; while the perfect competition hypothesis cannot be rejected for six countries at the same level of statistical significance (see Table A.6). Depending on which proxy of labour is chosen, the ranking of countries is significantly affected for some countries, for instance Argentina and Israel. However, given that using the number of employees reduces the number of observations, one should give more weight to the first specification described.

Figure 7. **H-Statistic panel 2004-13**



Note: Proxy 1 for price of labour = labour expenses/total assets, proxy 2 for price of labour = labour expenses/number of employees. The H-statistic captures the elasticity of bank interest revenues to input prices. The H-statistic is equal to 1 in case of perfect competition, while it ranges between 0 and 1 in monopolistic competition. Below 0, the H-statistic highlights a monopoly (perfect cartel) situation. An H-statistic above 1 is either caused by statistical noise, or by the non-respect of the underlying assumptions of the H-statistic framework.

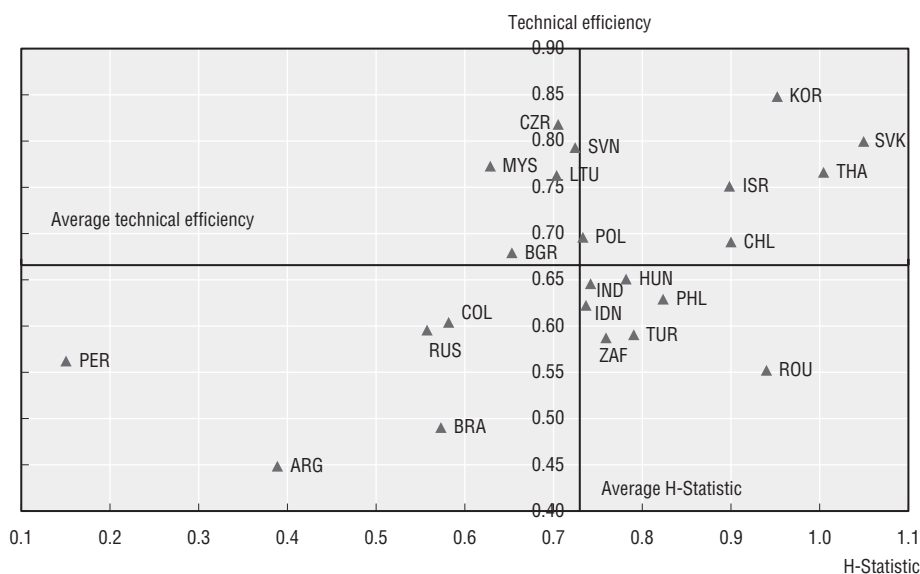
Source: Authors' calculations based on data from Bloomberg, SNL Financial.

A series of robustness tests presented in the Appendix include estimating input price elasticities without control variables using labour expenses over number of assets or using labour expenses over the number of employees as a proxy for the price of labour. They all tend to confirm that all countries are close to a monopolistic competition setting.

3.2. Links between contestability and efficiency

More contestable markets are usually thought to promote efficiency. When barriers to entry and exit are lower, the competitive behaviour of firms is fostered. The available empirical evidence seem to validate this causal relationship between contestability and efficiency gains (Berger and Hannan, 1998).

Figure 8 illustrates the positive correlation between market contestability and efficiency for the sample of countries used in this paper. It divides the contestability-efficiency plan into four quadrants to represent countries, which are above or below average with respect to the two dimensions over the period 2004-13. It shows that countries above average in terms

Figure 8. **Contestability and Efficiency, panel 2004-13**

Note: Contestability is captured by the H-statistic, measured using specification [1]. Technical efficiency is the average of the estimates from the stochastic frontier and data envelopment analyses for the period 2004-13.

Source: Authors' calculations based on data from Bloomberg, SNL Financial

of contestability for 2004-13 are also more likely to be above average in terms of efficiency. In the top-right quadrant, countries such as South Korea, the Slovak Republic, and Israel have more contestable and efficient banking markets than the sample average. A few countries including Argentina, Brazil and Russia are located in the bottom-left quadrant, presenting a below average performance in terms of contestability and efficiency.

3.3. Determinants of market contestability

This section explores the determinants of market contestability. A two-step procedure is implemented. In a first-step, parameters of reduced form price equations are estimated using equation [6], separately for each country along the period 2004-11. Then, time-varying H-indices by country and by year (H_{it}) can be calculated using equation [7]. In a second step, the predicted H-indices are regressed on a set of explanatory variables using countries' fixed effects (equation [8]). To allow for robustness tests, H-indices were estimated with and without control variables in equation [6]. Similarly, the second step regression is estimated with and without control variables.

$$\ln(p_{it}^y) = \alpha_i + e^{\tau_i \times t} \left(\sum_{j=1}^3 \beta_j \ln(p_{it}^j) \right) + \sum_j \theta_j BSF_{it} + v_i, \quad [6]$$

$$p^1 = \frac{\text{personnel expenses}}{\text{total assets}},$$

$$p^2 = \frac{\text{administrative expenses and depreciations} - \text{personnel expenses}}{\text{total assets}},$$

$$p^3 = \frac{\text{interest expenses}}{\text{total liabilities}},$$

$$p^y = \frac{\text{interest income}}{\text{total assets}},$$

$$H_{it} = (\beta_{i1} + \beta_{i2} + \beta_{i3})e^{\tau_i \times t}. \quad [7]$$

$$\hat{H}_{it} = \alpha_i + \sum_{j=1}^n \beta_j \times x_{jit} + u_{it}. \quad [8]$$

Equation [6], which follows the approach described by Bikker and Spierdijk (2008), helps to avoid erratic changes in the H-statistic that appear when trying to evaluate its value on a yearly basis. The underlying assumption for this specification is that market structure changes are relatively gradual.¹⁶ The estimation is carried out by non-linear least squares.¹⁷

In equation [8], the explanatory variables are the country-specific variables that were previously introduced to explain variations in inefficiency: i) the power of the official supervisory authority; ii) the degree to which banks face restrictions when engaging in securities, insurance and real estate operations; iii) the concentration of the national banking market, and iv) the market share of public as well as foreign-held banks.

While the strength of the supervisory authority is found to be positively correlated with efficiency, it has no clear effect on contestability (Table 3). Using a different approach and sample, Ahrend et al., (2009) find that more power for the banking supervisory is associated with greater competition in OECD economies. A potential source of difference is probably that the degree of enforcement of regulations is weaker in emerging market economies. Therefore, legal regulations *per se* might have a weaker effect in this context. However, the results show that activity restrictions are positively related to market contestability. This result is robust to variations in specifications and statistically significant. One reason may be that when commercial banks are limited in their range of activities, they face fewer opportunities to create barriers to competition by bundling services.

Market concentration has an adverse effect on contestability. This result is robust to variations in specifications and statistically significant. When assets are detained by fewer banks, collusive behaviours are usually thought to be easier to maintain. If incumbent banks are working together to prevent new banks from entering the market, this would reduce contestability. Further, increases in concentration may deter possible competitors by increasing the cost of entry (Kessides, 1990), which could contribute to decrease contestability.

The market share of public banks also has a deterrent effect on contestability. Public banks are generally perceived as less sensitive to market pressure, and conversely to be less prone to compete for market shares than their private banks counterparts. Further, entering a market where opponents are supported by the government may be perceived as less profitable for outsiders, reducing incentives for new entries. The result is statistically significant and robust to variations in specifications. However, the result could also be driven by an omitted variable. For example, public banks might take into consideration other factors than profit maximisation, such as correcting market failures, which may results in a lower sensitivity to market prices. This could contribute to the negative variation in contestability as measured by the H-statistic.

Foreign participation is positively correlated and statistically significant with an increase in contestability in four out of the eight regressions performed. In the remaining estimations, the impact of foreign participation is small and non-significant. This somewhat weak result is consistent with the literature that finds different effects of foreign banks on contestability. For example, while Levy Yeyati and Micco (2007) find a negative relation, other studies point out a positive effect (Yildirim and Philippatos, 2007).

Table 4. **Determinants of contestability**

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Explanatory variables for H-Statistics								
Supervisory power	-0.196 (0.172)	-0.002 (0.164)	0.003 (0.142)	0.093 (0.135)	-0.057 (0.189)	0.018 (0.196)	0.034 (0.155)	0.060 (0.160)
Activity restrictions	0.576*** (0.189)	0.173 (0.192)	0.309* (0.157)	0.041 (0.164)	0.667*** (0.208)	0.511** (0.229)	0.430** (0.171)	0.321 (0.194)
Market concentration	-0.124** (0.054)	-0.099* (0.050)	-0.062* (0.033)	-0.055* (0.031)	-0.120** (0.059)	-0.110* (0.060)	-0.063* (0.035)	-0.058 (0.037)
Public banks share	-0.094** (0.043)	-0.086** (0.040)	-0.046 (0.038)	-0.046 (0.035)	-0.163*** (0.047)	-0.160*** (0.048)	-0.095** (0.041)	-0.094** (0.042)
Foreign banks share	0.096* (0.053)	0.129** (0.050)	0.086** (0.042)	0.095** (0.039)	-0.03 (0.058)	-0.018 (0.059)	-0.010 (0.045)	-0.008 (0.046)
Control variables in (6)	yes	yes	yes	yes	no	no	no	no
Interpolation technique for World Bank Survey data	Linear	Linear	Polynomial	Polynomial	Linear	Linear	Polynomial	Polynomial
Time fixed effects in (7)	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.176	0.323	0.110	0.257	0.160	0.186	0.109	0.124
Number of countries	20							
Number of observations	160							

Note: Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10. Estimates and standard errors have been multiplied by 100 to improve readability.

4. Conclusions

The analysis in this paper has focused on the influence of regulative and structural characteristics on commercial banks efficiency and contestability. A positive correlation between contestability and efficiency has been found for the banking systems included in the sample. While the present work only presents correlations banks' and not causation,¹⁸ the theory of contestable markets invites us to interpret the results in one direction: more contestable markets tend to favour more efficient outcomes. Based on this analysis, some characteristics that are likely to improve efficiency directly or indirectly (through an increase in contestability) can be identified. Table 4 offers a synthesis of the influence that the different variables have as well as their level of statistical significance.

Table 5. **Impact of structural and regulative characteristic on efficiency and contestability**

	Efficiency	Contestability
Supervisory Power	+	+
Activity restrictions	+	+
Bank concentration	°	-
% government-held banks	-	-
% foreign-held banks	°	°

Note: + (-) sign indicates that a positive (negative) effects on efficiency/contestability has been identified in more than 50% of the reported estimations (Table 2 and 3). ° indicates that no clear effect has been identified. Statistical significance: *** p < 0.01, ** p < 0.05, * p < 0.1 if the corresponding significance levels were identified in 50% of the reported estimations or more.

There are some signs pointing in the same direction in three cases – indicating that forces at play are complementary:

- The power of the supervisory agency of the banking sector tends to increase both efficiency and contestability, even if the statistical significance in the last case is not ensured. This

result is coherent with other studies (Ahrend et al., 2009) that also found that an increase in the strength of supervisor was associated with greater contestability. By enforcing a better monitoring of banking activities, the supervisory agency is fostering efficiency indirectly (by enforcing competitive-friendly legislations) and directly (by avoiding illegal practices that may deter efficiency gains in the long-run). However, some authors point out that when corruption is a concern, a powerful supervisor may have the opposing effect on efficiency (Barth et al., 2012). Hence, an increase in the supervisory authority power and anti-corruption measures are probably inseparable to achieve effective results.

- Activity restrictions have a positive effect on efficiency and contestability. It could be expected that such restrictions could be harmful for efficiency since potential economies of scope and complementarities might help reduce costs. However, these economies of scope and complementarities are apparently outweighed by the loss incurred in non-specialisation, and the decrease in contestability stemming from the creation of conglomerates more likely to enjoy market power.

The market share of public banks has a deterrent effect on both contestability and efficiency. Public banks are sometimes less efficient than their private counterparts and less sensitive to market pressure. This could reduce aggregate efficiency and market contestability. The concentration of the banking sector is one example where a potential trade-off can be identified. Efficient banks are more likely to gain market shares over than less efficient counterparts, resulting in higher aggregate efficiency in the short run. However, it would also lead to more concentration, which tends to reduce contestability. This would in turn have a dampening effect on efficiency gains in the medium run. The resulting force is ambiguous, which may explain why no direct effect of concentration could be observed. Besides, it is likely that the final result depends on the level of concentration achieved. The deterrent effects of concentration on contestability are more likely to be important at high levels of concentration¹⁹. Hence, at low levels, the first effect – efficient banks taking over their less efficient counterparts – might dominate the loss incurred from lower contestability, while the opposite is probably true at high levels of concentration. Interestingly, the present paper was not able to identify a definite impact for the penetration of foreign banks on national banking systems. While a large strand of the literature argues that foreign banks are crucial to increase efficiency and contestability (Claessens and Laeven, 2004) some authors argue that foreign bank only have a limited spill-over effect and may even be detrimental (Levy Yeyati and Micco, 2007). More research is needed to determine the impact of foreign banks on national banking systems.

Notes

1. Battese and Coelli (1992) consider an “output-oriented” measure of efficiency. To be more specific, they do not use a cost function, but a production function. Applied to the current topic, this would correspond to the *production approach* of the banking sector. Using a cost function corresponds to the *intermediation approach* that is used in this paper. If the production function displays constant returns to scale, then both approaches yield the same results (see Parmeter and Kumbhakar, 2014). Otherwise, results will be different – but proportional if the production function is homogenous.
2. The model described in [1] contains two error terms. Identification – using the maximum likelihood principle – is made possible by assuming parametric forms for both error terms. For more details about the expression of the log-likelihood, refer to the appendix in Battese and Coelli (1992).
3. The first error term, v_{it} , captures statistical noise. This is why it does not enter the expression of cost efficiency. For more insights on the link between $CE_{it} = e^{-u_{it}}$ and Figure 2, refer to Battese and Coelli (1992).

4. Homogeneity in factor prices imposes that $\beta_1 + \beta_2 + \beta_3 = 1$. This holds if the production function has constant returns to scale.
5. There is a trade-off between including more countries, as one would obtain more variance in the sample and thus sharper results in what follows, but in doing so, one has to assume that the same technology to convert funds into loans is available to each of the banks in the sample. Here a conservative approach is implemented: it is assumed that the last point only holds for countries which have a “similar” level of economic development – defined in a broad sense of the term.
6. Approximately 200 observations on the number of employees were added manually. To be consistent with Bloomberg and SNL Financial standards, part-time workers were counted as $\frac{1}{2}$ worker.
7. For a more complete treatment of DEA, see Coelli et al (2005) and Bogetoft and Otto (2010).
8. See Coelli et al. (2005), from which notations and explanations were used, for more details.
9. The estimation is realised in Stata using the command *dea* (Yong-bae and Chonnjoo, 2010).
10. Bootstrapping methods, as the one used above for robustness check purposes, allow building confidence intervals for DEA estimates (Tortosa-Ausina et al., 2004).
11. These three surveys cover the year 2001, 2005 and 2008-10.
12. Country-specific variables (only available at the country level) are the following: the power of the supervisory agency of the banking sector, restrictions on activities, market concentration, public banks and foreign banks shares.
13. The statistical significance of these results has to be taken with a grain of salt, because these variables are at the country-level, such that the degrees of freedom to estimate these parameters are much smaller than for the bank-level parameters.
14. The correlation coefficient is 0.30 if one tries to evaluate the linear relation between technical efficiency and concentration of commercial banking markets.
15. A further assumption is that differences in prices are not reflecting differences in product quality.
16. This assumption may not hold in the case of cartels. The transition from a collusive to a competition phase is typically brutal. However, as indicated in Figure 7, most banks are in monopolistic competition setting.
17. In practice, the function *nls* in R is used.
18. To establish causality, one would need an instrumental variable, a variable correlated with contestability but not with efficiency.
19. When a quadratic term for concentration was included in equation [7], the corresponding estimate was negative in the 8 regressions represented in Table 4, but not statistically significant ($p < 0.1$) in six of them.

References

- Ahrend R., J. Arnold and F Murtin (2009), “Prudential Regulation and Competition in Financial Markets”, *OECD Economics Department Working Papers*, No. 735, OECD Publishing.
- Aigner, D., C.A. Lovell and P. Schmidt (1977), “Formulation and Estimation of Stochastic Frontier Production Function Models”, *Journal of Econometrics*, No. 6, pp. 21-37.
- Barth, J.R., G. Caprio, R. Levine (2012), “The Evolution and Impact of Bank Regulations”, The World Bank, *Policy Research Working Paper*, 6288.
- Battese, G.E. and T.J. Coelli (1992), “Frontier Production Functions, Technical Efficiency and Panel Data : With Application to Paddy Farmers in India”, *The Journal of Productivity Analysis*, No. 3, pp. 153-169.
- Battese, G.E. and T.J. Coelli (1995), “A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data”, *Empirical Economics*, Vol. 20, Issue 2, pp. 325-332.
- Beck, T., A. Demirgüç-Kunt and R. Levine (2007) “Financial, Inequality and the Poor”, *Journal of Economic Growth* 12(1), pp. 27-49.
- Berger, A.N. and T.H. Hannan (1998) “The Efficiency Cost of Market Power in the Banking Industry: A Test of the ‘Quiet Life’ and Related Hypothesis”, *The Review of Economics and Statistics*, MIT Press, Vol. 80(3), pp. 454-465.

- Berger, A.N., A. Demirguç-Kunt, R. Levine and J.G. Haubrich (2004) "Bank concentration and competition: An evolution in the making", *Journal of Money, Credit, and Banking*, Vol. 36 (3), pp. 434-450.
- Bikker, J. and L. Spierdijk (2008), "How Banking Competition Changed over Time", *DNB Working Paper*, No. 167.
- Bogetoft, P. and O. Lars (2010), "Benchmarking with Dea, Sfa and R", *International Series in Operations Research & Management Science*, Vol. 157.
- Bonin, John P., I. Hasan and P. Wachtel (2005) "Bank performance, efficiency and ownership in transition countries", *Journal of Banking & Finance*, Elsevier, Vol. 29(1), pp. 31-53.
- Cihák, M., A. Demirguç-Kunt, E. Feyen and R. Levine (2012), "Benchmarking Financial Systems around the World", *Policy Research Working Paper*, No. 6175, World Bank, Washington, DC.
- Claessens, S. and S. Djankov (1999), "Ownership Concentration and Corporate Performance in The Czech Republic", *University of Michigan Business School Working Paper*, No. 227.
- Claessens, S. and L. Laeven (2004), "What Drives Bank Competition? Some International Evidence", *Journal of Money, Credit and Banking*, Vol. 36 (3), pp. 563-583.
- Coelli, T.J. (1996), "A Guide to DEAP Version 2.1: A Data Envelopment (Computer) Program", *University of New England, Department of Econometrics*, Australia.
- Coelli, T.J., D.S.R. Rao, C.J. O'Donnell and G.E. Battese (2005), "An Introduction to Efficiency and Production Analysis", Second Edition, Springer.
- Coelli, T.J. and A. Henningsen (2013), "frontier: Stochastic Frontier Analysis, R package version 1.1-0, <http://CRAN.R-Project.org/package=frontier>.
- Farell, M.J. (1957), "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society, Series A (General)*, Vol. 120, No. 3, pp. 253-290.
- Kessides, I.N. (1990) "Market Concentration, Contestability, and Sunk Costs", *The Review of Economics and Statistics*, Vol. 72, No. 4, pp. 614-622.
- Kosak, M. and J. Zoric (2009), "Bank efficiency differences in the new EU member states", *Baltic Journal of Economics*, 9(2), pp. 67-90.
- Levy Yeyati, E. and A. Micco (2007), "Concentration and foreign penetration in Latin American banking sectors: Impact on competition and risk", *Journal of Banking and Finance*, No. 31, pp. 1633-1647.
- Meeusen, W. and J. van Den Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", *International Economic Review*, Vol. 18, No. 2, pp. 435-444.
- Nikiel, E.M. and T.P. Opiela (2002), "Customer Type and Bank Efficiency in Poland: Implications for Emerging Market Banking", *Contemporary Economic Policy*, Volume 20, Issue 3, pp. 255-271.
- Panzar J.C. and J.N. Rosse (1987), "Testing for "Monopoly" Equilibrium", *The Journal of Industrial Economics*, Vol. 35, No. 4, *The Empirical Renaissance in Industrial Economics*, pp. 443-456.
- Parmeter, C.F. and S.C. Kumbhakar (2014), "Efficiency analysis: a primer on recent advances", *Foundations and Trends (R) in Econometrics*, Vol. 7, pp. 191-385.
- Simar, L. and P.W. Wilson (1998) "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models", *Management Science*, No. 44(1), pp. 49-64.
- Tabak, B.M., D.M. Fazio and D.O. Cajueiro (2011), "Profit, Cost and Scale Efficiency for Latin American Banks: concentration-performance relationship", *Working Paper Series 244*, Brazilian Central Bank.
- Wezel, T. (2010), "Bank efficiency amid foreign entry: Evidence from Central America", *IMF Working Paper*, No. 10/95, Washington, DC.
- Yong-bae, J. and L. Choonjoo (2010), "Data envelopment analysis", *The Stata Journal*, Vol. 10, Number 2, pp. 267-280.

Appendix

Stochastic Frontier and Data Envelopment Analyses

Stochastic Frontier Analysis

Table A.1. **SFA – proxy 1 for labour price – number of banks by year**

Country	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
ARG	5	7	7	7	7	7	15	16	16	14	101
BGR	3	6	5	6	7	7	14	14	14	9	85
BRA	15	19	19	20	20	18	37	39	38	30	255
CHL	6	6	6	6	6	6	14	14	14	14	92
COL	7	4	4	6	8	9	13	13	13	11	88
CZR	1	4	5	6	6	6	12	13	13	11	77
HUN	1	3	4	7	8	8	8	8	8	6	61
IDN	24	26	27	32	31	32	52	55	55	54	388
IND	30	33	35	37	39	41	43	47	47	45	397
ISR	5	7	8	8	8	8	9	9	9	9	80
KOR	3	3	3	19	19	21	18	15	23	23	147
LTU	-	3	3	5	5	5	6	6	5	5	43
MYS	10	10	11	23	23	22	40	41	42	37	259
PER	5	6	10	7	12	12	21	21	17	10	121
PHL	14	13	14	12	15	16	23	24	23	16	170
POL	6	10	12	15	15	15	19	19	17	15	143
ROU	3	7	7	9	9	8	13	14	11	5	86
RUS	12	16	17	28	34	38	42	40	39	28	294
SVK	4	4	4	4	6	6	8	8	7	4	55
SVN	2	2	5	5	5	5	11	11	10	6	62
THA	7	8	9	11	15	16	21	22	19	18	146
TUR	3	4	11	17	17	17	28	29	28	29	183
ZAF	3	3	4	5	5	5	13	13	12	11	74
Total	169	204	230	295	320	328	480	491	480	410	3 407

Table A.2. **SFA – proxy 2 for labour price – number of banks by year**

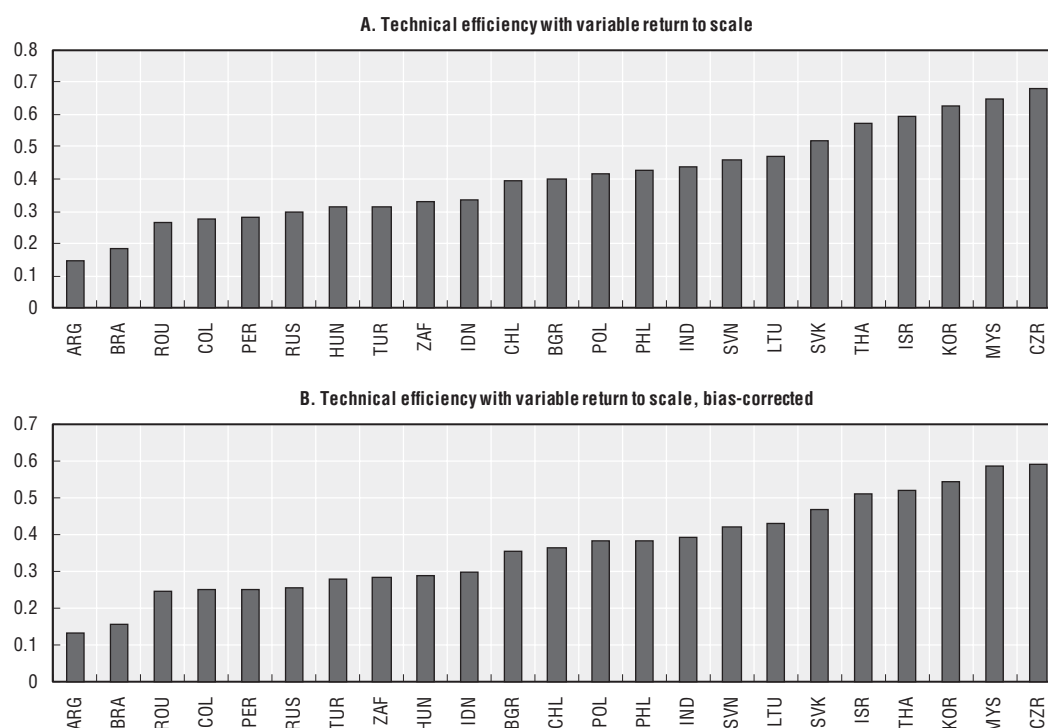
Country	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
ARG	3	4	4	4	4	5	8	8	10	9	59
BGR	2	4	4	4	5	6	14	14	14	8	75
BRA	7	10	14	15	12	13	21	21	24	18	155
CHL	3	3	4	2	3	5	10	10	11	11	62
COL	3	2	2	5	8	9	13	13	13	11	79
CZR	1	4	5	6	6	6	11	12	12	10	73
HUN	1	2	3	7	8	8	8	8	8	6	59
IDN	24	26	26	32	31	32	50	52	52	51	376
IND	25	29	24	31	31	34	32	38	39	33	316
ISR	3	6	7	8	8	8	9	8	9	9	75

Table A.2. **SFA – proxy 2 for labour price – number of banks by year (cont.)**

Country	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
KOR	3	3	3	14	16	17	15	13	21	20	125
LTU	-	3	3	5	5	5	6	5	4	4	40
MYS	9	10	7	12	12	11	14	12	10	9	106
PER	5	5	7	7	6	6	15	17	15	7	90
PHL	13	10	12	8	10	13	13	15	15	9	118
POL	1	9	10	10	11	13	18	18	16	15	121
ROU	3	7	7	9	8	7	13	14	11	5	84
RUS	7	11	10	18	23	28	27	25	26	13	188
SVK	4	4	4	4	6	6	8	8	7	4	55
SVN	2	2	5	5	5	5	11	11	10	6	62
THA	7	7	9	9	9	8	11	12	11	11	94
TUR	2	3	7	12	15	14	24	25	25	22	149
ZAF	3	3	4	4	5	5	10	10	10	10	64
Total	131	167	181	231	247	264	361	369	373	301	2 625

Data Envelopment AnalysisTable A.3. **DEA – yearly estimates of VRS TE**

Country	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Average 2004-13
ARG	0.29	0.31	0.29	0.26	0.25	0.22	0.18	0.19	0.13	0.09	0.19
BGR	0.45	0.51	0.66	0.66	0.69	0.53	0.48	0.49	0.51	0.48	0.52
BRA	0.18	0.24	0.23	0.32	0.19	0.37	0.32	0.26	0.23	0.27	0.26
CHL	0.60	0.55	0.53	0.50	0.39	0.69	0.59	0.50	0.43	0.43	0.51
COL	0.29	0.35	0.35	0.34	0.33	0.32	0.39	0.44	0.34	0.34	0.36
CZR	0.48	0.78	0.93	0.84	0.81	0.83	0.78	0.81	0.81	0.81	0.81
HUN	0.77	0.55	0.70	0.45	0.47	0.44	0.43	0.41	0.37	0.40	0.45
IDN	0.42	0.42	0.41	0.43	0.48	0.46	0.38	0.41	0.43	0.42	0.43
IND	0.50	0.52	0.57	0.53	0.51	0.52	0.50	0.56	0.44	0.47	0.51
ISR	0.82	0.48	0.86	0.87	0.82	0.65	0.92	0.45	0.62	0.70	0.72
KOR	0.80	0.81	0.82	0.81	0.82	0.85	0.80	0.83	0.72	0.79	0.80
LTU	-	0.73	0.73	0.67	0.61	0.53	0.58	0.63	0.67	0.78	0.65
MYS	0.80	0.74	0.79	0.70	0.83	0.78	0.79	0.82	0.71	0.74	0.77
PER	0.37	0.40	0.33	0.35	0.53	0.57	0.35	0.38	0.38	0.37	0.39
PHL	0.43	0.49	0.47	0.49	0.60	0.52	0.49	0.54	0.52	0.48	0.50
POL	0.47	0.44	0.51	0.56	0.49	0.53	0.56	0.56	0.53	0.58	0.53
ROU	0.21	0.35	0.36	0.36	0.35	0.32	0.37	0.36	0.33	0.36	0.35
RUS	0.46	0.42	0.48	0.46	0.47	0.51	0.40	0.35	0.36	0.35	0.42
SVK	0.60	0.76	0.64	0.61	0.64	0.74	0.80	0.75	0.72	0.76	0.71
SVN	0.44	0.59	0.69	0.69	0.67	0.70	0.67	0.63	0.56	0.58	0.63
THA	0.91	0.89	0.71	0.68	0.79	0.73	0.71	0.63	0.57	0.53	0.70
TUR	0.24	0.33	0.37	0.40	0.34	0.41	0.40	0.39	0.37	0.38	0.38
ZAF	0.52	0.39	0.53	0.47	0.57	0.40	0.39	0.44	0.39	0.35	0.43

Figure A.1. **Technical efficiency for a dual outputs cost function**

Note: Computed on the full sample of banks over 2004-13. Country's levels are obtained as the simple average of national banks scores. Deposits and total earning assets are considered as outputs. Bias-correction based on 2 000 bootstrap replicats (Panel B).

Source: Authors' calculations based on data from Bloomberg, SNL Financial.

World Bank Survey on Financial Regulation

Box A.1. Official Supervisory Power – Survey 2011

5.10 Does the banking supervisor have the right to meet with the external auditors and discuss their report without the approval of the bank?

- a. No
- b. Yes, it happens on a regular basis.
- c. Yes, it happens on an exceptional basis.

5.9 Are auditors required to communicate directly to the supervisory agency any presumed involvement of bank directors or senior managers in illicit activities, fraud, or insider abuse?

5.12 In cases where the supervisor identifies that the bank has received an inadequate audit, does the supervisor have the powers to take actions against...

b. The external auditor

a. The bank

12.3.2 Can the supervisory authority force a bank to change its internal organisational structure?

10.5 Do banks disclose to the supervisors...?:

b. Off-balance sheet items

Box A.1. Official Supervisory Power – Survey 2011 (cont.)

- a. Full audited financial statements
 - c. Governance and risk management framework
 - d. Regulatory capital and capital adequacy ratio
 - e. Transactions with related parties
 - f. Any other material information (i.e. information which omission or misstatement could change or influence the assessment or decision of a user relying on that information for making decisions)
 - g. Scope of consolidation (including reasons for not including certain entities, where appropriate)”
- 11.1 Please indicate whether the following enforcement powers are available to the supervisory agency
- f. Require banks to constitute provisions to cover actual or potential losses
 - a. Cease and desist-type orders for imprudent bank practices
 - b. Forbearance (i.e. to waive regulatory and supervisory requirements)
 - c. Require a bank to meet supervisory requirements (e.g. capital, liquidity etc.) that are stricter than the legal or regulatory minimum
 - d. Require bank to enhance governance, internal controls and risk management systems
 - e. Require bank to apply specific provisioning and/or write-off policies
 - g. Restrict or place conditions on the types of business conducted by bank
 - h. Withdraw the bank’s license
 - i. Require banks to reduce/restructure their operations (e.g. via asset sales and branch closures) and adjust their risk profile
 - l. Suspend or remove bank directors
 - m. Suspend or remove managers
 - n. Require commitment/action from controlling shareholder(s) to support the bank with new equity (e.g. capital restoration plan)
- 11.1 j. Require banks to reduce or suspend dividends to shareholders
- 11.1 k. Require banks to reduce or suspend bonuses and other remuneration to bank directors and managers
- 11.5 Which authority has the powers to perform the following problem bank resolution activities?
- Enter the initials of the corresponding authority from the following list of options: BS = Bank Supervisor, C = Court, DIA = Deposit Insurance Agency, BR/AMC = Bank Restructuring or Asset Management Agency, OTH = Other – please specify).
- a. Declare insolvency
 - Other – please specify
 - d. Undertake bank resolution mechanisms
 - Other – please specify
 - e. Appoint and oversee a bank liquidator/receiver
 - Other – please specify
- 11.5 b. Supersede shareholders’ rights

Box A.1. Official Supervisory Power – Survey 2011 (cont.)

Other – please specify

11.5 b. Supersede shareholders' rights

Other – please specify

11.5 c. Remove and replace bank senior management and directors

Source: Bank Regulation and Supervision Survey

Box A.2. Activity restrictions on banking activities – Survey 2011

4.1 What are the conditions under which banks can engage in securities activities?

- a. A full range of these activities can be conducted directly in banks,
- b. A full range of these activities are offered but all or some of these activities must be conducted in subsidiaries, or in another part of a common holding company or parent,
- c. Less than the full range of activities can be conducted in banks, or subsidiaries, or in another part of a common holding company or parent,
- d. None of these activities can be done in either banks or subsidiaries, or in another part of a common holding company or parent.”

4.2 What are the conditions under which banks can engage in insurance activities?

- a. A full range of these activities can be conducted directly in banks,
- b. A full range of these activities are offered but all or some of these activities must be conducted in subsidiaries, or in another part of a common holding company or parent
- c. Less than the full range of activities can be conducted in banks, or subsidiaries, or in another part of a common holding company or parent,
- d. None of these activities can be done in either banks or subsidiaries, or in another part of a common holding company or parent”

4.3 What are the conditions under which banks can engage in real estate activities?

- a. A full range of these activities can be conducted directly in banks,
- b. A full range of these activities are offered but all or some of these activities must be conducted in subsidiaries, or in another part of a common holding company or parent
- c. Less than the full range of activities can be conducted in banks, or subsidiaries, or in another part of a common holding company or parent
- d. None of these activities can be done in either banks or subsidiaries, or in another part of a common holding company or parent

Source: Bank Regulation and Supervision Survey.

H-statistic**Table A.4. H-statistic, proxy 1 for labour price and control variables, main estimates**

Country	Price of funds	Price of equipment	Price of labour	H	Observations	R-squared
ARG	0.431***	0.0261	-0.0684	0.39	101	0.89
BGR	0.298***	0.0568	0.299	0.65	85	0.61
BRA	0.431***	-0.00941	0.152	0.57	255	0.47
CHL	0.516***	0.0827	0.302**	0.90	92	0.87
COL	0.315**	0.0105	0.256	0.58	88	0.76
CZR	0.340***	0.165*	0.200**	0.71	77	0.88
HUN	0.466***	-0.0195	0.336***	0.78	61	0.86
IDN	0.546***	0.0557	0.135***	0.74	388	0.53
IND	0.679***	0.0154	0.0480*	0.74	397	0.89
ISR	0.341***	0.121	0.437	0.90	80	0.74
KOR	0.766***	0.0466*	0.139***	0.95	151	0.98
LTU	0.437***	0.0796	0.187**	0.70	43	0.95
MYS	0.480***	0.0676***	0.0804**	0.63	259	0.79
PER	0.335***	-0.0299	-0.154	0.15	121	0.51
PHL	0.283***	0.233**	0.308**	0.82	170	0.49
POL	0.650***	0.0159	0.0665	0.73	143	0.71
ROU	0.476***	0.0176	0.446	0.94	86	0.76
RUS	0.316***	0.0533	0.189***	0.56	294	0.76
SVK	33.40%	-0.0361	0.752**	1.05	55	0.65
SVN	0.356***	-0.0871*	0.456***	0.72	62	0.86
THA	0.431***	0.0737	0.500***	1.00	146	0.83
TUR	0.572***	0.0870**	0.131**	0.79	179	0.89
ZAF	0.696***	0.0655*	-0.00248	0.76	74	0.86

*** p < 0.01, ** p < 0.05, * p < 0.1.

Note: Price of labour price is measured as labour expenses divided by the total assets.

Table A.5. H-statistic, proxy 1 for labour price and control variables, hypothesis testing

Country	p-value perfect cartel (%)	p-value perfect competition (%)
ARG	8.5	3.9
BGR	0.0	1.1
BRA	0.0	0.0
CHL	0.0	12.0
COL	0.1	1.1
CZR	0.0	1.2
HUN	0.0	9.7
IDN	0.0	0.3
IND	0.0	0.0
ISR	0.3	67.7
KOR	0.0	11.4
LTU	0.0	3.0
MYS	0.0	0.0
PER	22.9	0.0
PHL	0.0	9.6
POL	0.0	0.3
ROU	0.0	74.8
RUS	0.0	0.0
SVK	0.0	69.6
SVN	0.0	0.0

Table A.5. **H-statistic, proxy 1 for labour price and control variables, hypothesis testing (cont.)**

Country	p-value perfect cartel	p-value perfect competition
THA	0.0	95.7
TUR	0.0	0.1
ZAF	0.0	2.9

Note: Perfect competition: (1) $H_0: H = 1$, $H_1: H \neq 1$; perfect cartel (2) $H \leq 0$, $H_1: H > 0$.

Table A.6. **H-statistic, proxy 2 for price of labour and control variables, main estimates**

Country	Price of Funds	Price of Equipment	Price of Labour	H	Observations	R-squared
ARG	0.444***	0.281**	0.183	0.91	59	0.96
BGR	0.492***	0.0813	0.391	0.96	75	0.60
BRA	0.509***	-0.00775	0.116	0.62	155	0.49
CHL	0.508***	0.238***	0.0979***	0.84	62	0.93
COL	0.472***	0.0188	0.107	0.60	79	0.78
CZR	0.347***	0.252***	0.103	0.70	73	0.87
HUN	0.555***	0.018	0.161	0.73	59	0.81
IDN	0.551***	0.0977*	0.0814**	0.73	376	0.53
IND	0.656***	0.0419**	0.0605***	0.76	316	0.89
ISR	0.352***	0.222**	-0.594***	-0.02	75	0.78
KOR	0.715***	0.101**	0.0472	0.86	125	0.94
LTU	0.428***	0.227***	-0.247***	0.41	40	0.96
MYS	0.469***	0.0582	0.000355	0.53	106	0.84
PER	0.219***	-0.185**	-0.207*	-0.17	85	0.73
PHL	0.249**	0.594***	0.0425	0.89	118	0.50
POL	0.631***	0.0674	0.213	0.91	121	0.67
ROU	0.561***	0.0173	-0.200***	0.38	84	0.73
RUS	0.682***	0.0392	0.0579	0.78	188	0.79
SVK	0.445*	0.119	0.329**	0.89	55	0.59
SVN	0.424***	0.00827	-0.2	0.23	62	0.82
THA	0.361***	0.143*	0.00402	0.51	94	0.73
TUR	0.599***	0.105**	0.000972	0.70	145	0.89
ZAF	0.696***	0.0702**	0.00586	0.77	64	0.86

Note: Price of labour price is measured as labour expenses divided by the number of employees. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

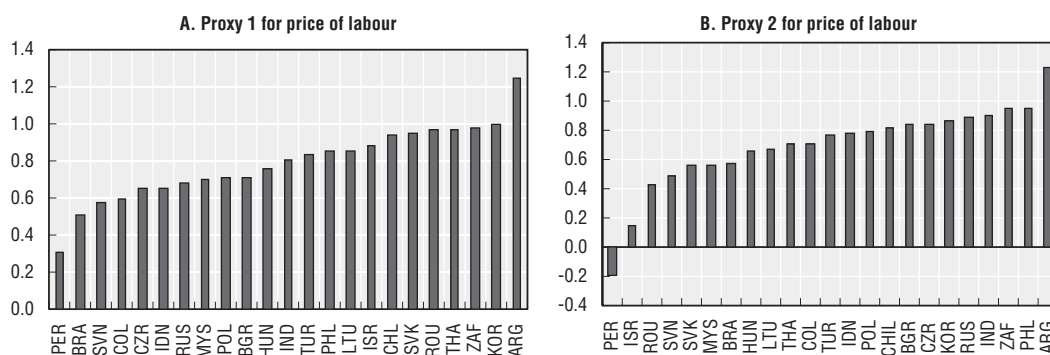
Table A.7. **H-statistic, proxy 2 for price of labour and control variables, hypothesis testing**

Country	p-value perfect cartel (%)	p-value perfect competition (%)
ARG	0.0	60.8
BGR	1.2	92.5
BRA	0.0	1.0
CHL	0.0	8.2
COL	0.2	2.9
CZR	0.0	1.3
HUN	0.3	21.8
IDN	0.0	1.1
IND	0.0	0.0
ISR	54.4	0.0
KOR	0.0	7.4
LTU	0.3	0.1

Table A.7. **H-statistic, proxy 2 for price of labour and control variables, hypothesis testing (cont.)**

Country	p-value perfect cartel (%)	p-value perfect competition (%)
MYS	0.0	0.1
PER	92.6	0.0
PHL	0.1	66.7
POL	0.1	71.5
ROU	0.1	0.0
RUS	0.0	0.9
SVK	0.0	51.2
SVN	6.9	0.0
THA	0.1	0.3
TUR	0.0	0.0
ZAF	0.0	2.0

Note: Perfect competition: (1) H0: $H = 1$, H1: $H \neq 1$; perfect cartel (2) $H \leq 0$; H1: $H > 0$.

Figure A.2. **H-statistic, estimation without control variables**

Note: Proxy 1 for price of labour = labour expenses/total assets, proxy 2 for price of labour = labour expenses/number of employees.

Source: raw data Bloomberg and SNL Financial.

Table A.8. **H-statistic, proxy 1 for labour price without control variables**

Country	Price of Funds	Price of Equipment	Price of Labour	H	Observations	R ²
ARG	0.737***	0.0205	0.490***	1.25	101	0.83
BGR	0.379**	-0.0145	0.348	0.71	85	0.38
BRA	0.488***	-0.0578	0.0824	0.51	255	0.41
CHL	0.525***	0.068	0.343***	0.94	92	0.86
COL	0.545***	0.0425	0.00405	0.59	88	0.59
CZR	0.436***	-0.0992	0.317	0.65	77	0.69
HUN	0.567***	0.0418*	0.153**	0.76	61	0.79
IDN	0.372***	0.00234	0.282***	0.66	388	0.32
IND	0.830***	-0.0304	0.00387	0.80	397	0.83
ISR	0.346***	0.0711	0.466	0.88	80	0.72
KOR	0.730***	0.0712**	0.195***	1.00	151	0.97
LTU	0.516***	-0.0639	0.403**	0.86	43	0.90
MYS	0.543***	0.109***	0.0464	0.70	259	0.75
PER	0.376***	0.00151	-0.0732	0.30	121	0.41
PHL	0.222***	0.253**	0.374***	0.85	170	0.43
POL	0.720***	-0.0541	0.0461	0.71	143	0.66
ROU	0.575***	-0.0493	0.440*	0.97	86	0.70
RUS	0.426***	0.0748*	0.177**	0.68	294	0.64

Table A.8. **H-statistic, proxy 1 for labour price without control variables (cont.)**

Country	Price of Funds	Price of Equipment	Price of Labour	H	Observations	R ²
SVK	0.402**	-0.197	0.748***	0.95	55	0.56
SVN	0.563***	-0.158**	0.172*	0.58	62	0.76
THA	0.414***	0.108**	0.450***	0.97	146	0.81
TUR	0.601***	0.0356	0.202**	0.84	179	0.82
ZAF	0.576***	0.330*	0.0703*	0.98	74	0.64

Note: Price of labour price is measured as labour expenses divided by the total assets. *** p < 0.01, ** p < 0.05, * p < 0.1.

Table A.9. **H-Statistic, proxy 2 for labour price without control variables**

Country	Price of Funds	Price of Equipment	Price of Labour	H	Observations	R-squared
ARG	0.526***	0.392***	0.306***	1.22	59	0.94
BGR	0.335	0.210**	0.294	0.84	75	0.23
BRA	0.496***	-0.0176	0.0977	0.58	155	0.46
CHL	0.471***	0.271***	0.0775**	0.82	62	0.91
COL	0.595***	0.0322	0.0853	0.71	79	0.77
CZR	0.377***	0.108	0.355***	0.84	73	0.77
HUN	0.546***	0.0522**	0.064	0.66	59	0.76
IDN	0.470***	0.116*	0.193***	0.78	376	0.33
IND	0.804***	0.00303	0.0973***	0.90	316	0.85
ISR	0.338***	0.214**	-0.405***	0.15	75	0.75
KOR	0.670***	0.120**	0.0787	0.87	125	0.93
LTU	0.537***	0.101	0.0309	0.67	40	0.85
MYS	0.517***	0.0782**	-0.0323	0.56	106	0.79
PER	0.312***	-0.204	-0.296**	-0.19	85	0.46
PHL	0.211*	0.709***	0.0303	0.95	118	0.42
POL	0.688***	0.0069	0.101	0.80	121	0.61
ROU	0.726***	-0.00549	-0.288***	0.43	84	0.64
RUS	0.765***	0.0625**	0.056	0.88	188	0.69
SVK	0.373**	0.054	0.130*	0.56	55	0.47
SVN	0.575***	-0.0954	0.0121	0.49	62	0.75
THA	0.351***	0.172**	0.186*	0.71	94	0.66
TUR	0.682***	0.0615*	0.02	0.76	145	0.81
ZAF	0.541***	0.358*	0.0484	0.95	64	0.62

Note: Price of labour price is measured as labour expenses divided by the number of employees. *** p < 0.01, ** p < 0.05, * p < 0.1.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD Journal

Economic Studies

OECD Journal: Economic Studies publishes articles in the area of economic policy analysis, applied economics and statistical analysis, generally with an international or cross-country dimension. It draws significantly on economic papers produced by the OECD Economics Department, other parts of the OECD Secretariat and the Organisation's intergovernmental committees.

We also welcome comments on articles previously published in the Journal

Consult this publication on line at http://dx.doi.org/10.1787/eco_studies-v2016-1-en.

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases. Visit www.oecd-ilibrary.org for more information.

Volume 2016/1

OECD publishing
www.oecd.org/publishing



ISSN 1995-2848
VOLUME 2016
(1 ISSUE)
13 2016 01 1 P

