INTERNATIONAL MONETARY FUND

INDEPENDENT EVALUATION OFFICE

RESEARCH AT THE IMF: RELEVANCE AND UTILIZATION

BACKGROUND PAPERS

MAY 20, 2011

Papers

I. Evaluating the Quality of IMF Research: A Citation Study

II. Review of IMF Research on Monetary Policy Frameworks

III. Review of IMF Research on Tax Policy

IV. An Examination of the Quality of a Sample of 60 Selected Issues Paper

V. An Evaluation of the Research Chapters of the IMF's *World Economic Outlook* and *Global Financial Stability Report*

VI. Review of the IMF's *Regional Economic Outlook Reports*, 2003–09

VII. Macro-Financial Linkages in IMF Research

IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/11/01

Evaluating the Quality of IMF Research: A Citation Study

Joshua Aizenman, Hali Edison, Larissa Leony, and Yi Sun

May 20, 2011

Address. 700 19th Street, N.W., Washington D.C. 20431, U.S.A. Telephone. +1 202 623 7312 Fax. +1 202 623 9990 Email. feedback@ieo-imf.org Website. www.ieo-imf.org

 BP/11/01

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

Evaluating the Quality of IMF Research: A Citation Study

Prepared by Joshua Aizenman, Hali Edison, Larissa Leony, and Yi Sun

May 20, 2011

Abstract

This study compares the performance of IMF research to that of seven peer institutions in terms of publication in refereed journals and of citations in working papers of the comparator group during the period 1999–2009. Publication and citations are commonly used as indicators of quality of research. The comparator group is composed of Bank of Canada, the Board of Governors of the Federal Reserve, Federal Reserve Bank of New York, Federal Reserve Bank of San Francisco, Inter-American Development Bank, Organization for Economic Cooperation and Development, and the World Bank. IMF working papers were cited less often than those of the Federal Reserve, but more than those of the other comparators. Citations by the IMF and other international organizations are underestimated because the data used for the analysis excluded publications in developing countries and in languages other than English.

JEL Classification Numbers: A10, A14, A30

Keywords: International financial institutions, citation counts, publication counts, quality comparisons

Author's E-Mail Address: jaizen@ucsc.edu; hedison@imf.org

Contents Page

ABBREVIATIONS

BCA	Bank of Canada
BIS	Bank for International Settlements
CBs	central banks
CEPR	Center for Economic Policy Research
ECB	European Central Bank
FedBG	Federal Reserve Board of Governors
FedNY	Federal Reserve Bank of New York
FEDs	Federal Reserve Board of Governors, Federal Reserve Bank of New York, and Federal Reserve Bank of San Francisco
FedSF	Federal Reserve Bank of San Francisco
G-Scholar/GS	Google Scholar
IDB	Inter-American Development Bank
IFIs	international financial institutions
IMF	International Monetary Fund
NBER	National Bureau of Economic Research
OECD	Organization for Economic Cooperation and Development
RePEc	Research Papers in Economics
UCSC	University of California at Santa Cruz
WB	World Bank
WP	working paper

I. INTRODUCTION

1. Research and its timely dissemination are critical to the success of international financial institutions (IFIs) and central banks (CBs) in fulfilling their missions. Research provides the conceptual and empirical basis for better policymaking, and for better communication of policies to affected countries and the public. Consequently, many IFIs and CBs have invested significant resources in recent years to improve their capabilities for research and its dissemination. This study evaluates the quality, relevance, and utilization of IMF research, seen in the context of research by other leading policy institutions. The premise of the study is that research is critical to the IMF's successful operations because it contributes to the development and updating of the analytical tools that the Fund needs to discharge its responsibilities.[1]

2. A significant and growing literature deals with the scientific evaluation of academic research, evaluating and ranking the quality of academic papers, individual scholars, and institutions. A substantial literature on the use and usefulness of citations of academic journal articles for measuring research quality and researcher productivity follows the work of Cole and Cole (1973) and Merton (1973). Studying the quantity and pattern of citations of published research findings has become an important approach for measuring the impact and diffusion of academic research. Article citations are often used as a measure of research quality and utilization in the assessment of individual researchers, university departments, or academic journals. Data on citations of publications may also be used to trace the influence and evolution of knowledge, following the approach of Jaffe and Trajtenberg (2002), who used data on patent citations to study the diffusion of technological knowledge.

3. The usefulness of citations for measuring research impact is also the subject of a large literature in the sociology of science that emphasizes the role of academic culture and the incentives to cite. Much of the interest in citation patterns and the motives to cite stems from the use of citations to evaluate scientific careers and compensate academic researchers. Posner (2000) offers a lucid analysis of the incentives of scholars to cite the work of predecessors. He points out that citations play a key informational role in the written presentation of research, and acknowledge the priority of contributions by others, but can also be used strategically.[2] Citations can eliminate the replication of results, derivations, and arguments that are already known. They can also be used to place new findings or ideas in context, using readers' familiarity with a literature to reduce what needs to be written and read. Because scholarship is costly, information costs and networking may be other important

---

[1] See the March 2010 IEO Issues Paper, "Research at the IMF: Relevance and Utilization," available at www.ieo-imf.org.

[2] While citations serve the purpose of conveying information in academic exposition, the rewards to being cited and the role of others in the peer reviewing process affect the incentives to cite.

factors for understanding citation behavior (see Aizenman and Kletzer (2008) for further discussion and references).

4. Recognizing that there are differences among journals, efforts to assess the quality of research often apply two broad measures of quality. One of these is a publication count, weighting the publication record of an institution (or an individual) by the relative quality of the journals in which its publications have appeared. The other measure is a citation count, sometimes weighting each citation by the relative quality of the citing journals. Coupé (2003) applied such a measure to provide a comprehensive ranking of economics departments. His ranking methodology is based on the citations-weighted journal ranking by Laband and Piette (1994), and is used to assess the output of individual researchers and then, according to their affiliation, compute the departmental rankings.[3]

5. The literature that rates the research produced by central banks has applied a similar approach, focusing mostly on publications in refereed journals.[4] But, while informative, this approach frequently ignores the unique dimension of policy institutions, whose mandate goes beyond research and may include the design and implementation of domestic policies (as in the case of central banks); the establishment of rules, institutions, and procedures to regulate the international financial system at the multilateral level (as at the IMF) and Bank for International Settlements (BIS); or fostering the development of poor and emerging markets (as at the World Bank (WB) and regional development banks). The policy role of policy institutions implies that the fast dissemination of their research and position working papers (WPs) is at least as important as the longer-run impact of these papers on knowledge via their ultimate publication in refereed journals.

6. Our paper studies the research impact of the eight policy institutions, focusing on their working papers and using two distinct approaches. First, we evaluate the short-term impact of a working paper by tracing the citations of new working paper in subsequent working papers written within the eight institutions (and in all publications). We use this citation count as a measure of the revealed policy impact (and general impact) of that paper. Looking at the panel patterns of the citation count of a working paper provides a measure of the relevance and the speed of dissemination of research within the peer institutions, and the

---

[3] Coupé (2003) is one of four papers sponsored by the European Economic Association to rank European research centers; the other three are Combes and Linnemer (2003), Kalaitzidakis and others (2003), and Lubrano and others (2003).

[4] Jansen (1991) and Rapoport and Yi (1997) compare the publication records of the various components of the U.S. Federal Reserve System; Eijffinger and others (2002) and Jondeau and Pagès (2003) focus on those of central banks in Europe. Goodfriend and others (2004) provide a detailed assessment of the economic research activities of the European Central Bank (ECB), and St-Amant and others (2005) measure the policy relevance of central banks' research.

impact of the research outside the peer institutions. Second, we evaluate the longer-term impact of a working paper by tracing its ultimate publication, if any, in policy-oriented and academic journals. Aggregations of the citations and the publication records of the working papers published by the policy institutions provide measures of these institutions' overall research performance. Such measures are imperfect, due to various aggregation and heterogeneity issues discussed below. Yet, short of better measures, changes in the relative ranking of these institutions provide useful information about the relevance, utilization, and quality of their research.

7. The paper is structured as follows. After Section II, on data sources, Section III outlines the methodology applied. Section IV presents the results for the two approaches outlined above—the citation counts achieved by the working papers and the working papers' record of being published in journals—drawing conclusions about the strengths and weaknesses of Fund research relative to that of peer institutions. Section V summarizes the main findings and concludes. Statistical tables are grouped in the annex.

II. Data Sources

8. Evaluating the quality, quantity, and relevance of the research undertaken by policy institutions is made more difficult by the lack of a single efficient platform where all the questions can be answered in a timely manner at reasonable expense. The sheer volume of such a study, dealing with thousands of working papers, renders manual procedures impracticable.

9. These constraints led us to use two platforms: Research Papers in Economics (RePEc) and Google Scholar (G-Scholar or GS). While these two platforms are subject to various limitations, taken together they provide useful insights about the peer institutions' research.

- *RePEc* (http://RePEc.org) is an initiative to create a public-access database that promotes scholarly communication in economics and related disciplines. All RePEc information is freely available from the web-based RePEc services. A stated limitation of this service is that due to procedural limitations, RePEc thus far covers so far only three-fourths of the relevant electronic documents posted on the Internet. RePEc provided us with a data set that allows analysis of the citations received by all the working papers that were published by the eight policy institutions during 1999–2009.

- *Google Scholar* is an open and rich data source, aggregating and ranking the citations of scholarly papers according to Google's algorithms. The convenient features of GS, and the presumed quality associated with the Google brand, have made this a popular service. We used GS to establish which of the well-cited working papers issued by the peer institutions had eventually been published in professional journals.

10. It is important to note that the use of RePEc and Google Scholar may exclude publications in developing countries that are not captured by these two platforms, particularly if they are in foreign languages. This may generate some bias in the results that are discussed later.

III. Methodology

11. A common approach for evaluating institutions (or scholars) is benchmarking: comparing the citations and publications record of an institution (or a scholar) to a group of peer institutions (or peer scholars) that share similar characteristics. Short of having a simple metric for measuring the absolute value of research, benchmarking allows the performance of an institution to be assessed relative to that of comparable peer institutions operating in a similar environment.[5]

12. An example of applying a benchmarking approach in evaluating the performance of policy institutions is St-Amant and others (2005). These authors measure the policy relevance of research done by CBs by counting the number of citations of central bank working papers in the publications of BIS. They chose BIS for this purpose because BIS pursues interests in the policy issues that are relevant to CBs, and because its websites are conducive to citation searches. Their methodology is based on the premise that having a paper cited in a publication of a policy institution is a good indicator of its usefulness for addressing relevant policy issues. Our methodology is a natural extension of St-Amant and others (2005) and the references in that article.

13. We study the research impact of the IMF and seven other policy institutions selected for their commonality of purpose and data considerations. As well as the Fund, the group comprises the Bank of Canada (BCA), Inter-American Development Bank (IDB), Organization for Economic Cooperation and Development (OECD), U.S. Federal Reserve Board of Governors (FedBG), Federal Reserve Bank of New York (FedNY), Federal Reserve Bank of San Francisco (FedSF), and the WB.[6]

14. As reported by RePEc, the IMF and WB each issued similar numbers of working papers during 1999–2006—about 260 per year—while the three Fed institutions together

---

[5] See Camp (1989) and Boxwell (1994) for further discussion about benchmarking.

[6] The U.S. Federal Reserve System comprises the Board of Governors and 12 regional district offices. To prevent overrepresentation of the Federal Reserve in our study, we chose the Board and the two regional Federal Reserve District Banks with the highest citation counts. These turned out to be the Federal Reserve Bank of New York and the Federal Reserve Bank of San Francisco (we applied several criteria, including the total citations for all papers; the total citations for the top 50 papers, and the largest number of papers with at least one citation). We did not include the ECB in our sample because of some measurement errors caused by RePEc's automated procedures: ECB WPs each include a list of previous working papers, and the automated procedures count those lists as part of the citation statistics.

published about 320 working papers per year and the other three institutions (BCA, IDB, and OECD) together published about 240 working papers per year (see Annex Table A1).

A. Achieving a Stable Panel for Comparison

15. Ideally, we would like to have a stable panel comparison across the eight policy institutions during the period 1999–2009, whereby each working paper has its citations and ultimate publication, if any, measured over the same time horizon (i.e., period of the same length) as the other papers.[7] The wish to have a stable panel of this kind poses a dilemma: longer citation and publication horizons reduce the panel length, because the information is truncated in 2009. As described in detail below, we deal with this dilemma by observing from the data that the citation count per year peaks about two years after a working paper is issued, and that the average working paper gets more than half of its total citations within two years from its date of issue. Hence, we set the citation horizon to be two years from the working paper's date of issue.[8] Similar considerations apply to the analysis of publications. Here the data show, as detailed below, that most of the publications of working papers as journal articles take place within three years from the working paper's date of issue. Therefore, we set the journal publication horizon at three years.

16. Given these choices, the durations of the panels in our analysis reported in Section IV depend on the context of the discussion. Figures and tables reporting on citations and publication of working papers within five years of their date of issue cover those papers that were issued during 1999–2003, while those dealing with citations of working papers within two years of their date of issue cover those papers issued during 1999–2006.

Defining a lag structure: citations

17. To define a lag structure that will allow a stable panel for comparing citation rates among the peer institutions during 1999–2009, we use the RePEc database to compare the citation dynamics across the eight institutions over time. The goal is to select the same time horizon for each working paper to measuring its citations and ultimate impact.

18. As shown in Figure 1, which plots the citation rate in the years following the working papers' dates of issue, we find that the curves follow an asymmetric inverted U-shape, with a fast takeoff and slower decay. Information dissemination is substantially slower among

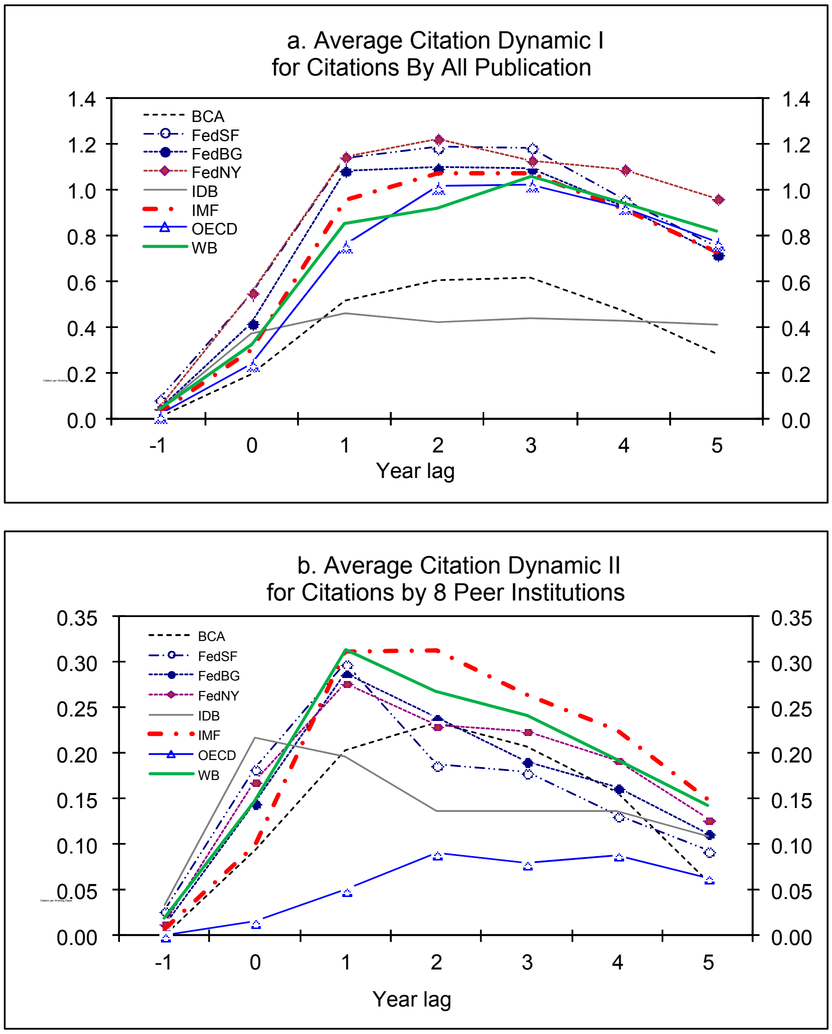
---

[7] Thus, when measuring the number of citations received, a working paper issued in 2000 should be accorded the same time horizon as a working paper issued in 2004; similarly, when measuring whether or a not a working paper has been published, a working paper issued in 1999 should be accorded the same time horizon as a working paper issued in 2002.

[8] That is, from t–1 to t+2, where t stands for the paper's year of issue.

publications at large than it is within the peer institutions: the citation curve counting all citations (i.e., including those outside the peer institutions) peaks about three years after a working paper's issue date (Figure 1a), thus a year or two later than the peak of the citation curve within the peer institutions (Figure 1b).[9]

Figure 1. Citations Per Paper, for Working Papers Issued in 1999–2003



Note: Compiled using data provided by RePEc. Panel *a* reports the citations of a WP in all the publications in the RePEc database. Panel b reports the citations of a WP in the publications of the peer institutions.

[9] For the same group of papers, issued during 1999–2003, the ratio of citation outside the eight institutions versus inside the eight institutions increased from 2.53 (=8228/3253) two years after their issue dates to 3.26 (=18850/5776) five years after their issue dates (see Annex Tables A6 and A7), indicating the broadening of the WPs' influence from policy institutions to academia, practitioners, and other institutions.

19. We focus our analysis on the citation dynamics during a paper's first two years (i.e., from t–1 to t+2, where t stands for the paper's year of issue), thus covering a year past the peak of the average citation curve for citations by the eight peer institutions. This choice provides us with a longer panel study of possible trends for papers that were written during 1999–2006. Hence, while we use a citation horizon of five years in studying the citation curves shown in Figure 1 (therefore covering working papers issued during 1999–2003), we shorten the citation horizon to two years in some of the subsequent discussions in order to get a longer panel. Working with such a longer panel helps to identify trends.

Defining a lag structure: publications

20. The second measure we use to compare the quality and relevance of the Fund's research with that in the other peer institutions is the rate of publication of working papers in top journals. To begin calculating this rate, we fed Google Scholar with a list of those working papers that were issued by the eight institutions during 1999–2006 and had received at least one citation in RePEc within two years of their issue date. Using the G-Scholar searching program, we identified the papers in our working paper list that had eventually been published in top journals. The list of top journals was extracted from Combes and Linnemer (2003).[10]

21. To define a lag structure that will allow a stable panel for comparing the publication rates across policy institutions, we try to identify the typical number of years between the date that a working paper is issued and the date that it is published in a journal (if indeed it is published). Figure 2 reports the rates of publication in top journals for the working papers that were issued between 1999 and 2003. The "top journals" are defined as those with impact weights equal or above 0.33 in Combes and Linnemer (2003).

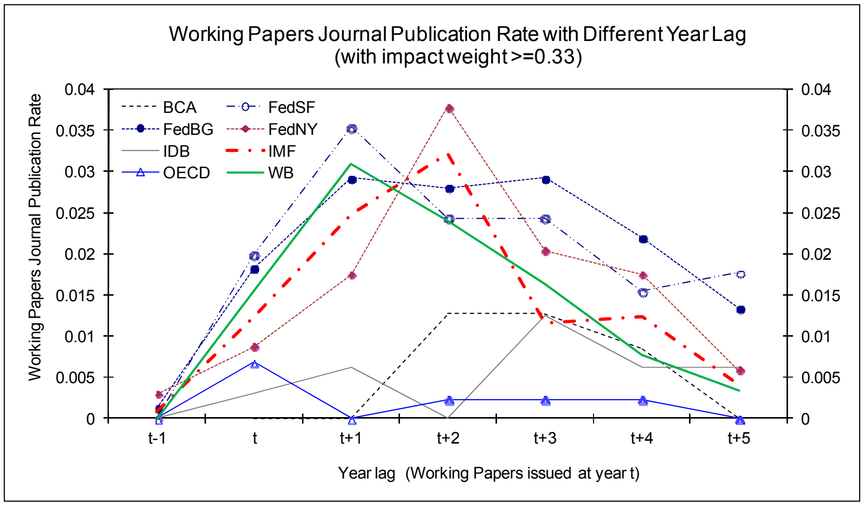
22. We find that most of the publication in journals took place within three years from the working papers' date of issue: as shown in Figure 2, publications in journals during the first three years from issuing a working paper account for respectively 81 percent (IMF), 71 percent (OECD), 75 percent (BCA), 58 percent (IDB), 75 percent (FedNY), 73 percent (FedBG), 79 percent (FedSF), and 86 percent (WB) of their overall journal publications.

---

[10] Combes and Linnemer (2003) ranked top journals with impact weights of 1, 0.67, 0.5, 0.33, 0.17, and 0.08. We focused on the top four groups of journals (weighted 1, 0.67, 0.5, 0.33), neglecting journals with impact weights below 0.17. *IMF Staff Papers* is not included in Combes and Linnemer's (2003) journal ranking, presumably because their article focuses on academic scholarship, and therefore on journals that are open to outside academicians. We added *IMF Staff Papers* to their list of journals, attaching to this journal a weight of 0.33. This weight is consistent with the journal ranking reported by Kalaitzidakis and others (2003), and equals the weight attached by Combes and Linnemer (2003) to the *World Bank Economic Review*, a journal that is ranked by Kalaitzidakis and by others as comparable overall to *IMF Staff Papers*.

Thus in the analysis reported in Section IV we focus on journal publications within three years.

Figure 2. Working Paper Publication Rates in Top Journals, by Different Year Lags



Note: Compiled using data provided by G-Scholar. The figure reports the publication rate (i.e., number of publications/number of cited WPs) by year lag for WPs issued during 1999–2003. The data include only those WPs that had achieved at least one citation within two years after their issue date. The top journals include only those journals with impact weights >=0.33, as defined by Combes and Linnemer (2003).

B. Identifying the Effects of Institutional Size

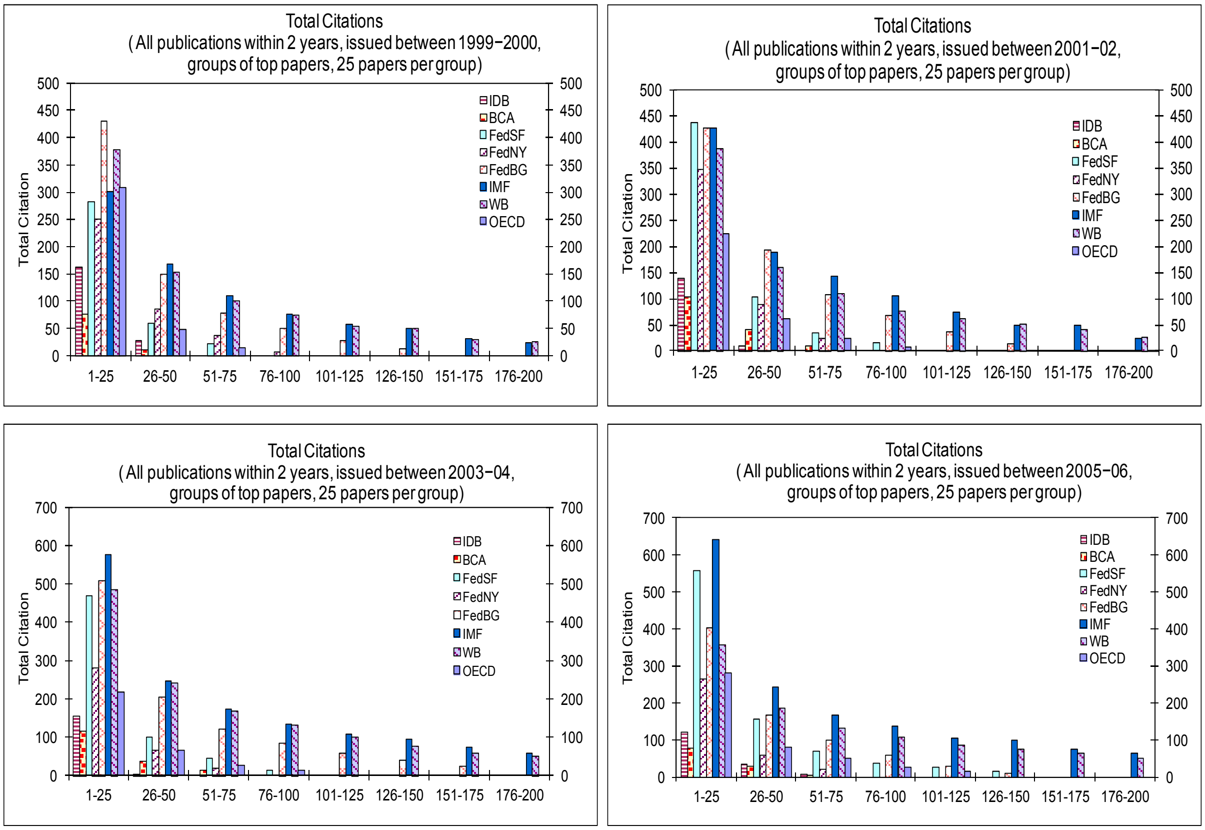
23. The wide range in size of the eight policy institutions may matter for their research culture, quality, and citations, due to scale effects.[11] To investigate systematically the effects of institutional size, we report in Figure 3 the citation patterns of the 200 most cited papers published by each of the peer institutions, grouped in bins of 25. Note the exponential decay of the citation curves.[12] These curves highlight the importance of the size of the peer institutions, as well as allowing us to compare the quality of the top papers across institutions.

---

[11] Previous studies found that size affects recognition (the "Matthew effect"): a scientific community experiences a nonlinear increase in the amount of recognition it receives as its size increases. See Merton (1968, 1973). A noisy indicator of size is the number of registered scholars at the RePEc web page (as of August 9, 2010) citing a peer institution. The numbers are as follows: IMF (226), WB (265), OECD (138), BCA (53), IDB (62), FedBG (81), FedNY (67), FedSF (28).

[12] The exponential decay of research productivity is consistent with models of knowledge. See Jaffe and Trajtenberg (1996, 2002).

Figure 3. Total Citations, By Groups of Top Papers



Note: Based on data provided by RePEc. The figure reports the top 200 cited papers of each institution, ordered in bins of 25.The citations covered are those achieved within two years of a WP's issue date, and include citations in all publications, not just those of the peer institutions.

IV. Institutional Comparisons: IMF Research Performance in Context

24. This section compares the achievements of the working papers of the IMF with those of the seven other policy institutions, first in terms of citation rates and then in terms of their publication rates in top journals. We then evaluate the Fund's research performance over time, relative to that of the peer institutions, using research quality and importance ratios. Finally we compare the journal-publication impact ratios of the IMF and the other policy institutions, as a measure of the presumed long-term impact of their research.

A. Citation Rates

25. The IMF's record of citations is relatively good. Counting citations by all sources, we find that about half the working papers issued by the IMF, WB, and the three Fed institutions had been cited within two years of their issue date. The IMF leads other IFIs and CBs in this respect, with more than 60 percent of its working papers receiving citations within two years (Annex Table A2, column 4). The citation rates are relatively low for BCA, IDB, and OECD:

on average, only about one-third of their working papers had received citations within two years.[13]

26. The IMF's number of citations per cited paper (4.37) is lower than that of the three Federal Reserve institutions (FEDs, 5.84), but higher than those of the WB (3.87) and the other three policy institutions (3.64), confirming the overall high quality of IMF's research papers (Annex Table A2, column 5).

27. Counting only the citations in working papers prepared by the 8 peer institutions, we find that more than 40 percent of the IMF working papers had been cited within two years of being issued (Annex Table A2, column 8). This is a rate well above those of all the seven other institutions, which range from 20 percent to 36 percent.[14]

28. The IMF's number of citations by peer institutions per cited paper (2.16) is lower than that of the three FEDs (2.31) but higher than that of the WB (2.13) and that of the BCA and IDB combined (2.07) (Annex Table A2, last column).[15]

B. Publication Rates

29. The publication rate of IMF working papers in professional journals is somewhat low compared with those of the other policy institutions. Looking across the eight institutions at the publications that took place within three years of working papers' issue dates, we find that on average 5–20 percent of the cited working papers had been published in a top journal: 14 percent (IMF); 22 percent (three FEDs); 15 percent (WB); and 5 percent (BCA/IDB/OECD combined) (Annex Table A3, column 4).[16]

30. Which top journals most commonly published the working papers from the peer institutions? For the IMF, articles in *IMF Staff Papers* far outnumbered those in any other top journal; the other main publishing journals were *Journal of International Economics* and

---

[13] The RePEc database has problems with the records of citation by OECD: most of the OECD WPs do not have a reference list in the RePEc database. This may cause a downward bias of the citation rate for OECD WPs. Therefore in Annex Table A2 we report both the subtotal of "the other three institutions" (BCA, IDB, and OECD), and the subtotal of BCA and IDB without the OECD.

[14] Due to the RePEc database's problem with OECD reference records, OECD WPs' citation count, at 8 percent, is relatively low compared with those of other institutions.

[15] In Section IV we look at the "self-citation" patterns—the citations of an institution's working papers in other working papers issued by the *same* institution. We find significant variation in the self-citation rates across institutions.

[16] Over a time horizon of five years from a working paper's issue date, the comparable rates are 10.1 percent (IMF), 10.0 percent (WB), 11.1 percent (FedSF), 13.9 percent (FedNY), 14.5 percent (FedBG), 1.6 percent (OECD), 3.4 percent (BCA), and 3.7 percent (IDB).

Journal of Development Economics (Annex Table A4). Across all eight institutions, the two top journals most commonly used were the *Journal of International Economics* and the *Journal of Money, Credit and Banking*.

31. We find that some of the institutions published many of their working papers in their own sponsored journals. For example, in the case of the FedNY, more than half of its published working papers appear in its own sponsored journal, *Economic Policy Review*. In the analysis reported below, we consider only the journals that have impact weights equal or above 0.33, according to Combes and Linnemer (2003), in order to eliminate instances of self-publication in relatively low-ranked journals.[17]

C. Research Performance Over Time

32. To evaluate IMF research performance over time, we first trace the evolution of the quality ratio and the importance ratio of the Fund's working papers relative to those of the other peer institutions during 1999–2006. We then apply a range of approaches to check the robustness of our findings, and lastly assess the evolution of the journal-publication impact ratio.

Quality and importance ratios

33. Quality and importance ratios provide a means to control for the presence of the significant fraction of working papers that are not cited within two years of being issued. The average quality ratio of an institution's working papers is defined as the percentage of working papers that have been cited within two years of being issued, and the average importance ratio of an institution's working papers is defined as the average number of citations that a cited paper achieves within two years of being issued.[18] Hence, the importance ratio measures the average feedback received by those working papers that are counted in the quality ratio. The quality ratio and the importance ratio provide information about the overall quality of an institution's working papers.

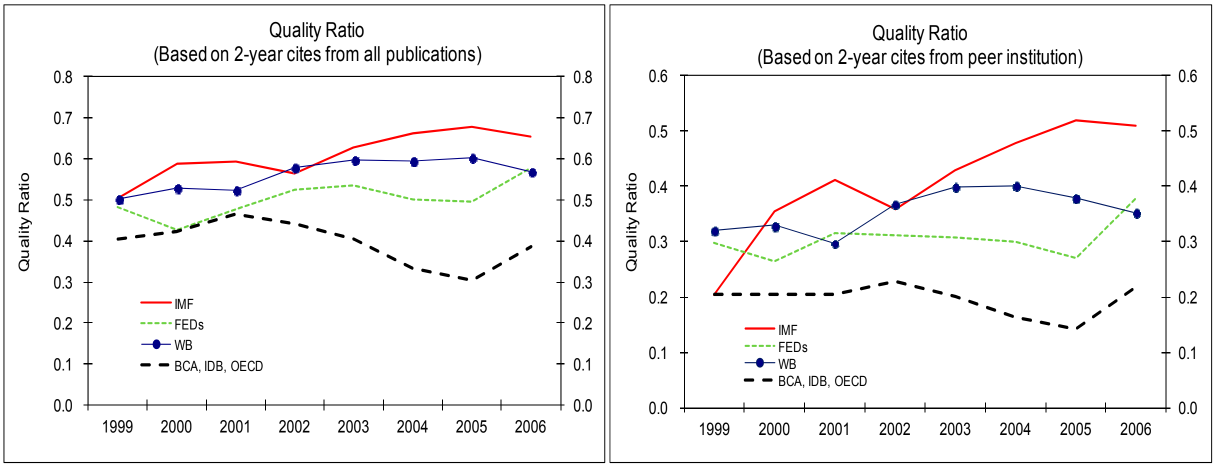
34. The quality and importance ratios show that the IMF's research performance was at about the mean for the policy institutions in 1999, and improved remarkably during the next seven years (Figures 4 and 5). An apparent deterioration in the last year was not large enough to negate the overall improvement during the period studied. Note that the overall performance of the peer institutions other than the IMF is trendless, both for citations within the peer institutions and for all citations.

---

[17] We also used other criteria, the *World Economic Outlook* citation ranks, and obtained similar results.

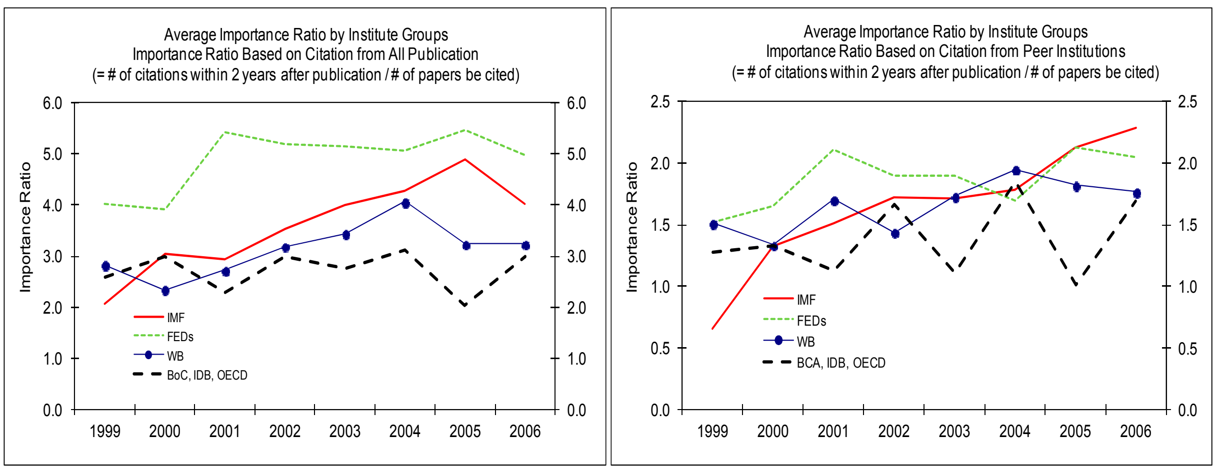
[18] The choice of a two-year window reflects the shape of the citation curves reported in Figure 1. Similar results hold for longer time windows.

Figure 4. Quality Ratios: IMF and Non-IMF, 1999–2006



Note: Compiled using data provided by RePEc. We define the average quality ratio of an institution's WPs as the percentage of WPs that have been cited within two years of being issued. In the averages for the combined three Fed institutions and the combined BCA, IDB, and OECD curves, the weights of the individual institutions are based on the numbers of WPs they issued each year.

Figure 5. Importance Ratios: IMF and Non-IMF, 1999–2006

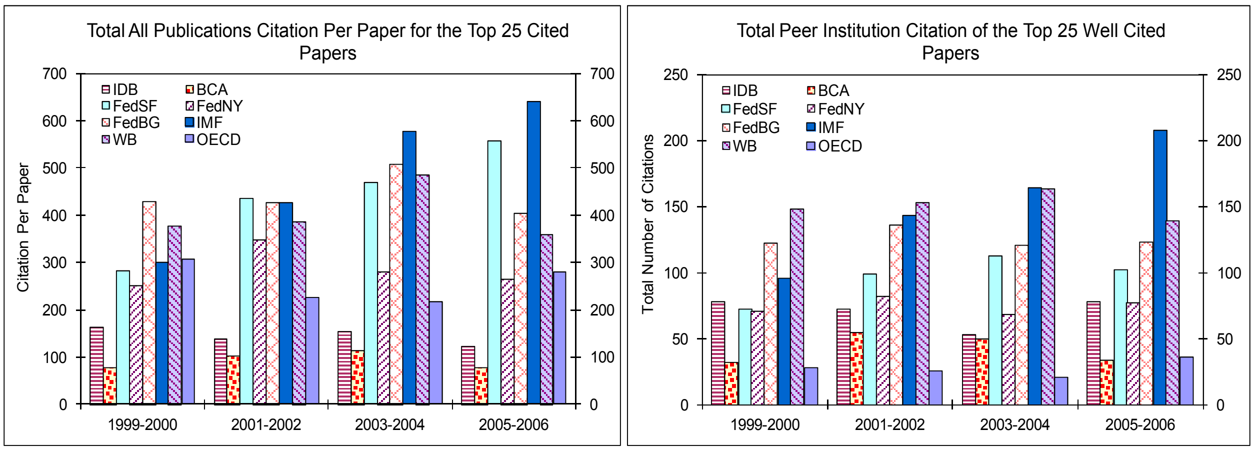


Note: Compiled using data provided by RePEc. We define the average importance ratio of an institution's WPs as the average number of citations that each cited paper achieves within two years of being issued. In the averages for the combined three Fed institutions and the combined BCA, IDB, and OECD curves, the weights of the individual institutions are based on the numbers of WPs they issued each year.

35. To control for the effect of institutional size on the citation rate, and thus on our understanding of the relative performance of the different institutions, we compare the citations of each institution's 25 most cited papers. Looking at the citations of these papers in all publications, we see that the IMF improved its performance relative to that of the WB during 1999–2006 (Figure 6, left panel). In 1999–2000, the FedBG was the top performer, followed by the WB, OECD, and IMF. Over the next six years this ranking changed considerably, with the IMF leading in 2005–06, followed by the FedSF, FedBG, and the WB. The IMF, though a much smaller organization than the WB, "out-performed" the WB by this measure across the entire range of papers during the later period of the sample, 2005–06.

36. Looking at the citations within the eight institutions (Figure 6, right panel), the results are similar, showing a significant improvement in the performance of the 25 most cited papers at the IMF relative to the peer institutions.

Figure 6. Top 25 Well-Cited Papers: Total Citations and Peer Institution Citations

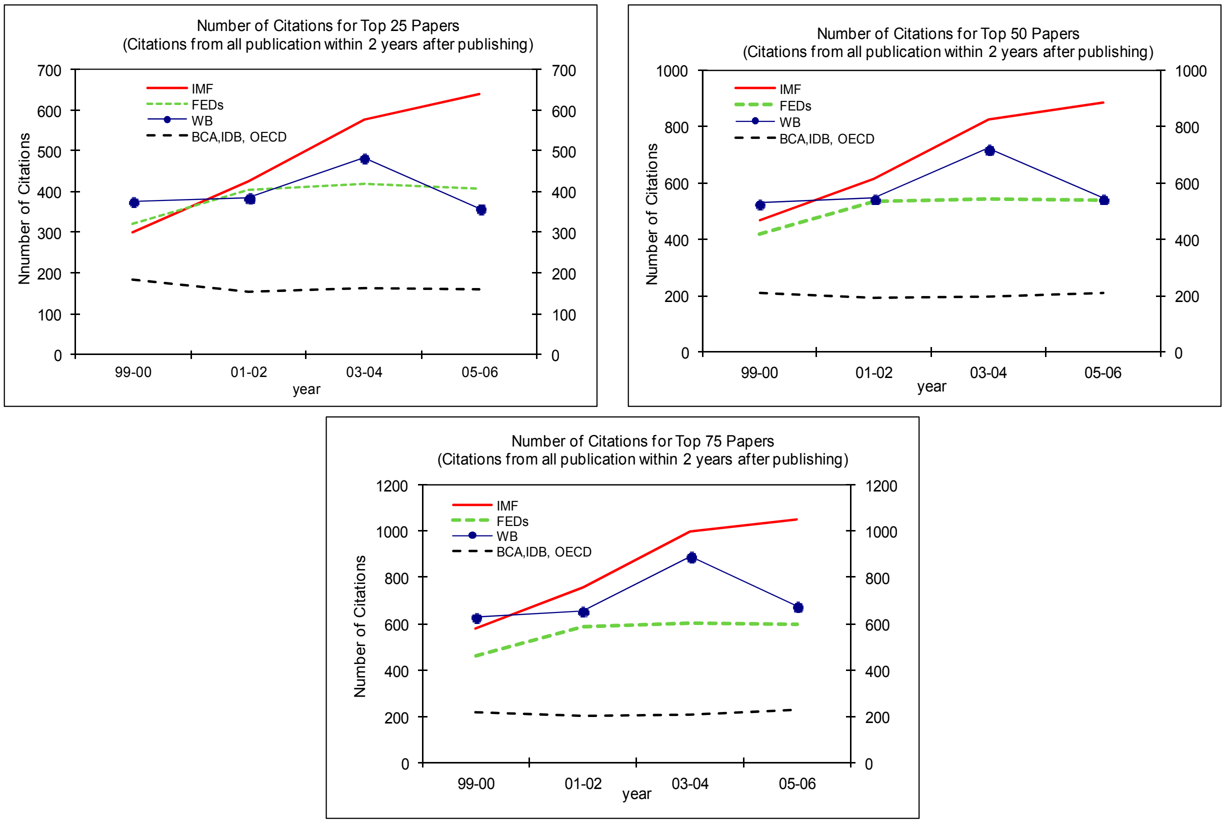


Note: Compiled using data provided by RePEc. Covers citations achieved by WPs within two years of their issue date.

37. Figure 7 provides a robustness check on these findings, verifying that the improvements over time in the citation rates of IMF working papers have depth, and apply well beyond the tranche of the top 25 papers. The figure shows the citation trends of the 25, 50, and 75 most-cited papers of the IMF in comparison to the trends for the other peer institutions, looking at citations in all publications. The IMF's performance improved remarkably between 1999 and 2006, although the rate of improvement declined in the later part of the sample.

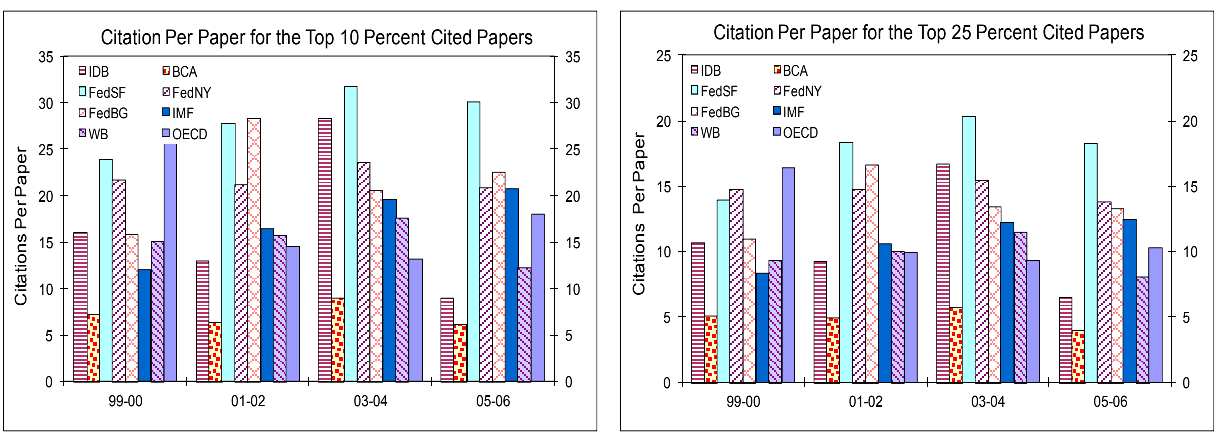
38. Trends in the average number of citations achieved per top paper show that over the study period the IMF steadily improved its research quality relative to that of its peer institutions. Among the Fund's most-cited 10 percent of papers, the citation rate increased from 12 citations per paper in 1999–2000 to 20 per paper in 2005–06 (Figure 8, left panel). As a result, in 2005–06 the Fund's citation rate was higher than those of all the other institutions except for the three FEDs. Similar results apply for the top 25 percent of papers (Figure 8, right panel).

Figure 7. Top Papers of IMF and Non-IMF: Numbers of Citations



Note: Compiled using data provided by RePEc. Covers citations achieved by WPs within two years of their issue date.

Figure 8. Citations Per Paper for the Top Well-Cited Papers

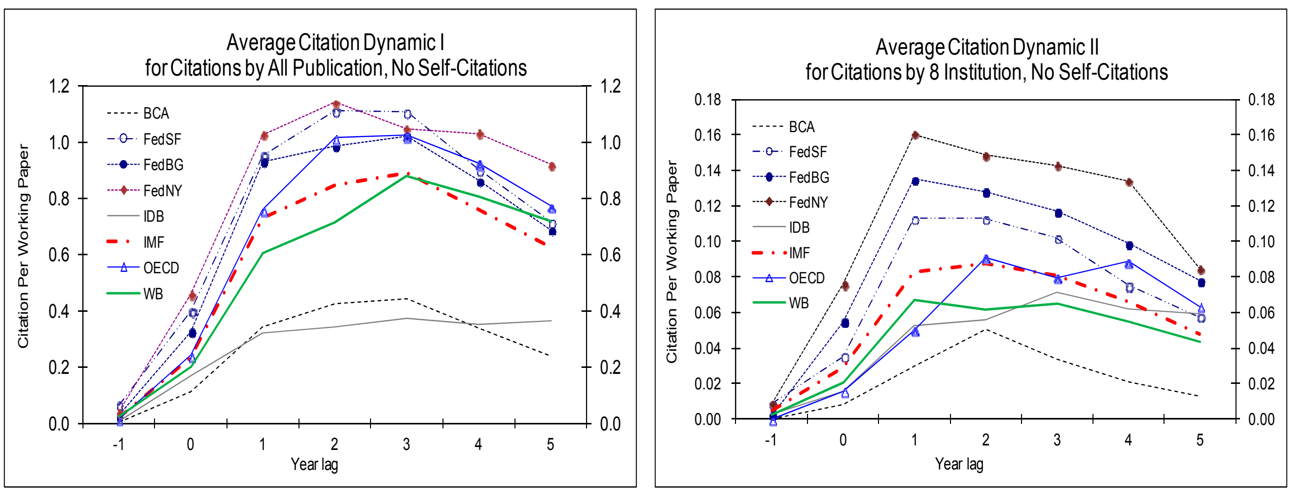


Note: Compiled using data provided by RePEc, covering citations in all sources within two years of the WP's issue date.

39. To further check the robustness of this record of improvement in IMF research quality, we look at what proportion of the citations of all working papers are "self-citations," that is, citations in other papers issued by the same institution. We find that on average during a working paper's first two years, about two-thirds of all the citations it receives are self-citations (Annex Table A5, last column).[19] There is a large range within this average, from zero (OECD)[20] to 0.43 (FedNY), 0.53 (FedBG), 0.61 (FedSF), 0.72 (IMF), 0.78 (IDB), 0.8 (WB), and 0.83 (BCA) (Annex Table A5). Similar results hold when we look at the citations within five years of issue date achieved by the papers that were issued over the same period (Annex Table A6).

40. When self-citations are excluded, the performance gap widens between the three Fed institutions and the other policy institutions including the Fund. On this basis, the three Fed institutions received the most citations, with the IMF performing below the Feds but above the other peer institutions (Figure 9). Recall that when self-citations are included, the IMF appears to be at par with the Fed institutions, as was seen in the lower panel of Figure 1 above.

Figure 9. Citations Per Paper, Excluding Self-Citations, for Working Papers Issued in 1999–2003



Note: Compiled using data provided by RePEc, covering WPs issued during 1999–2003. "Self-citations" is defined as citations in papers produced within the same institution.

[19] Using Annex Table A5 as a benchmark, Annex Table A6 reports the five-year citation rates for papers that were issued during 1999–2003, and Annex Table A7 reports the two-year citation rates for papers that were issued during 1999–2006. These tables allow one to see how changing the time horizon for citation affects the self-citation rate and how the citation and self-citation rates change over time.

[20] Due to the RePEc database's problem with OECD records, data on OECD working papers do not show self-citations.

41. Why should the IFIs show a greater tendency to self-cite, in comparison with the Federal Reserve institutions?[21] Annex Table A8 lists the top 10 working paper series or journals that cite the working papers produced by the 8 peer institutions, and shows that NBER and CEPR are the 2 main working paper series that cite the working papers of the 8 peer institutions. The relatively large numbers of citations of Fed working papers by these 2 working papers series help to explain the relatively low self-citation rates of the 3 FEDs.

Journal-publication impact ratios

42. On the premise that a paper will have a bigger impact if it is published in a top-tier journal, we define an institution's journal-publication impact ratio as its total weighted number of journal publications divided by its total number of working papers in the G-Scholar paper list. The weights are based on the rankings in Combes and Linnemer (2003): each journal is assigned a weight corresponding to four tiers: 1 (5 top-tier journals); 0.67 (16 second-tier journals); 0.5 (39 third-tier journals); 0.33 (69 fourth-tier journals); and zero (other journals).[22]

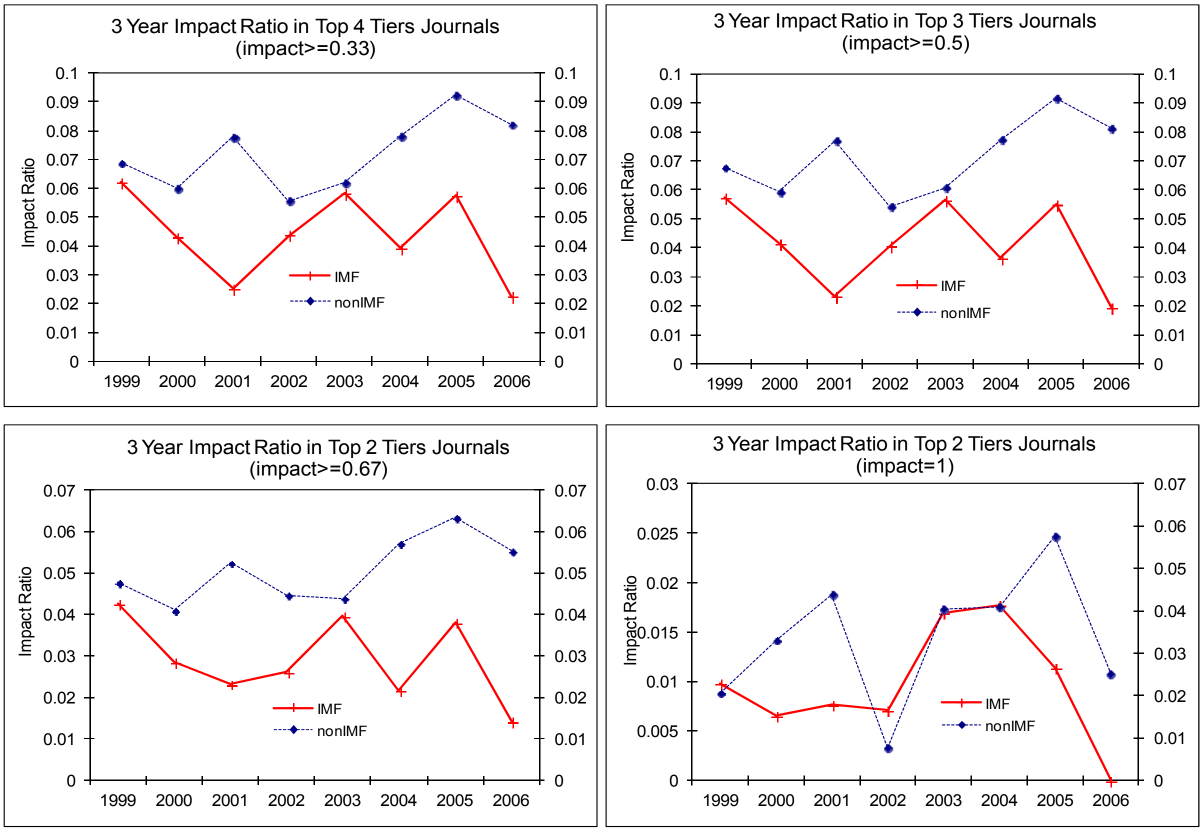
43. Compared with the average for the other seven policy institutions, the IMF has published fewer of its working papers in top journals. For the Fund, and for the other peer institutions taken as a group, Figure 10 reports the journal-publication impact ratio of working papers issued during 1999–2006 that had received at least one citation within two years of their issue date. On average, the IMF's ratio has been below the average of the other seven peer institutions. It approached the average in 2003 but dropped towards the end of the sample, in 2006.

---

[21] There are a host of issues associated with explaining the variance in self-citation ratios across institutions. Size matters: larger institutions tend to cite more of their own working papers. Cultural issues and the mission of the institution, too, play a role: some institutions that focus on narrower tasks may cite more of their own papers. In some institutions, the research done by different departments may be highly interdependent. In addition, differences in the corporate culture of institutions are likely to affect the frequency of citation of, and relating to, peer scholars in the same institution.

[22] Hence, a weight of 0.1 at year t implies that on average, a working paper published in year t, and cited within the next two years, had an average journal impact of 0.1 during the years t to t+3. This procedure attaches a journal-impact ratio of zero for papers that were not published in the window [t, t+3].

Figure 10. Journal-Publication Impact Ratios: IMF and Average Non-IMF



Note: Compiled using G-Scholar data covering WPs issued during 1999-2006. We count only journal publications within three years after the WP's issue date. Journal impacts are based on Combes and Linnemer (2003).

44. When the other seven peer institutions are disaggregated into three groups, as in Figure 11, the overall performance of the IMF appears to be comparable to that of the WB, well below the combined average of the three FEDs, and well above the combined average of the BCA, IDB, and the OECD. During 2003–05, the IMF performed noticeably better in the top journal tier than did the WB, but this improvement did not last.

Figure 11. Journal-Publication Impact Ratios: IMF and Peer Institution Groupings

Note: Compiled using G-Scholar data covering WPs issued during 1999-2006. We count only journal publications within three years after the WP's issue date. Journal impacts are based on Combes and Linnemer (2003).

V. Conclusion

45. This study has compared the citation and publication records of the IMF to those of seven peer institutions during 1999–2009, focusing on their research contained in working papers. We used the citation count of a working paper in other papers issued by the eight institutions (and in all publications) as a measure of the revealed policy impact (and general impact) of that paper. We also evaluated the longer-term impact of the policy institutions' working papers by tracing their eventual publication, if any, in policy-oriented and academic journals.

46. The relative performance of the IMF is quite sensitive to including or not including self-citations. The reason is that the self-citation rate is higher in the Fund than the average of the eight institutions, and in particular relative to the 3 FEDs' (about 80 percent in the Fund versus about 60 percent for the FEDs). For example, during the period 1999–2009, the IMF led other IFIs and CBs in the number of cited papers as a ratio of total papers (the quality ratio), with more than 60 percent of its working papers receiving citations within two years. However when self-citations are excluded, the three FEDs received the most citations, with

the IMF performing below the FEDs but above the other peer (policy institutions) institutions. This pattern is also observed for the number of citations that a cited paper achieves within two years of being issued (the average importance ratio).

47. For an institution of its size, the IMF has published relatively few of its working papers in top journals. Looking at the publications that took place within three years of working papers' issue dates, the Fund's publication rate of 14 percent compares with 22 percent at the selected three Fed institutions and 15 percent at the WB, though it is well above the combined average of the BCA, IDB, and the OECD. In addition 40 percent of Fund papers in the top journals are published in *IMF Staff Papers*. However this was also true for other institutions. For example 60 percent of the Federal Reserve Bank of New York publications in the top journals are accounted by articles in the institution's *Economic Policy Review.*

48. By several measures, the Fund's research performance improved over the evaluation period. Both in terms of the quality ratio and the importance ratio the IMF's research performance was at about the mean for the policy institutions in 1999 and improved significantly during the next seven years, particularly in terms of the later indicator. An apparent deterioration in 2006 was not large enough to negate the overall improvement during the evaluation period.

49. There are well-known limitations to the information content of citations. While citations provide signals about the research value of a paper, they also reflect other factors such as research networking, and are subject to known biases and challenges. Because this paper compares the research records of relatively large peer institutions, concerns regarding biases related to the idiosyncrasies of an individual scholar and paper are less relevant. The benchmarking approach applied has the advantage of providing a relative ranking of the peer institutions operating in a similar environment. Yet the approach is still subject to possible biases related to differential scales, heterogeneity, and the varying missions of the eight policy institutions studied. Finally, one would also expect that much of the publications in the developing world, in particular those in foreign languages, are not captured by the two data platforms being used. To the extent these publications may tend to cite more often research published by the IFI than the FEDs the use of these two platforms may have underestimated the total citations of the former relative to the latter.

ANNEX. STATISTICAL TABLES

Table A1. Numbers of Working Papers Issued By Policy Institutions in 1999–2006

Institutions	1999	2000	2001	2002	2003	2004	2005	2006	Total
IMF	202	260	219	248	282	255	260	290	2,016
WB	231	255	239	193	248	287	335	304	2,092
FedSF	86	86	89	101	90	105	138	143	838
FedBG	176	157	148	144	196	143	180	103	1,247
FedNY	81	74	65	61	62	57	74	45	519
Fed subtotal	343	317	302	306	348	305	392	291	2,604
BCA	38	37	41	59	61	63	58	65	422
IDB	71	87	56	53	55	43	103	119	587
OECD	52	81	94	88	125	134	152	175	901
Others subtotal	161	205	191	200	241	240	313	359	1,910

Note: This table is based on RePEc data, and includes both cited and noncited WPs.

Table A2. Citations of Working Papers Issued in 1999–2006[1]

		Citations by All Sources[2]				Citations by Peer Institutions[3]			
	Number of WPs in Sample	Number of WPs Cited	Total Citations	Number of Cited WPs/Total Number of WPs	Cites per Cited paper	Number of WPs Cited	Total Citations	Number of cited WPs/Total Number of WPs	Cites per cited paper
IMF	2,016	1,237	5,405	0.61	4.37	838	1,808	0.42	2.16
WB	2,092	1,182	4,576	0.57	3.87	748	1,591	0.36	2.13
FedSF	838	390	2,442	0.47	6.26	240	541	0.29	2.25
FedBG	1,247	583	3,392	0.47	5.82	351	856	0.28	2.44
FedNY	519	292	1,550	0.56	5.31	176	374	0.34	2.13
Fed subtotal	2,604	1,265	7,384	0.49	5.84	767	1,771	0.29	2.31
BCA	422	213	514	0.5	2.41	129	204	0.31	1.58
IDB	587	162	657	0.28	4.06	116	303	0.2	2.61
OECD	901	349	1,466	0.39	4.2	72	111	0.08	1.54
BCA and IDB subtotal	1,009	375	1,171	0.37	3.12	245	507	0.24	2.07
BCA, IDB, OECD total	1,910	724	2,637	0.38	3.64	317	618	0.17	1.95

Note: Based on RePEc data. The RePEc database has problems with the records of citation by OECD: most of the OECD WPs did not have a reference list in the RePEc database. This may cause a downward bias of the citation rate for OECD WPs. Therefore we report both the subtotal of “the other three institutions” (BCA, IDB, and OECD), and the subtotal of BCA and IDB without the OECD.

[1] “All sources” and “peer institutions” citations include only citations within two years of the WP’s issue date (i.e., from *t*–1 to *t*+2, where *t* stands for the WP’s issuing year).

[2] “Citations by all sources” are citations from all sources (WPs and journals) in the RePEc data.

[3] “Citations by peer institutions” includes self-citations. Self-citations are citations in other WPs issued by the same institution.

Table A3. Top-Journal Publication of Working Papers Issued in 1999–2006

	Total Number of WPs Issued[1]	Number of WPs in G-Scholar Database[2]	Top Journal Publication Within Three years[3]		
			Number of WP Publication	Conditional Publication Ratio	Overall Publication Ratio
	(1)	(2)	(3)	(4)=(3)/(2)	(5)=(3)/(1)
IMF	2,016	1,237	173	0.14	0.09
WB	2,092	1,182	179	0.15	0.09
FedSF	838	390	89	0.23	0.11
FedBG	1,247	583	142	0.24	0.11
FedNY	519	292	45	0.15	0.09
Fed subtotal	2,604	1,265	276	0.22	0.11
BCA	422	213	17	0.08	0.04
IDB	587	162	13	0.08	0.02
OECD	901	349	6	0.02	0.01
BCA IDB subtotal	1,009	375	30	0.08	0.03
BCA IDB OECD total	1,910	724	36	0.05	0.02

Note: Based largely on G-Scholar data.

[1] "Total number of WPs issued" is based on RePEc data.

[2] The G-Scholar data include only those WPs with at least one citation by any source covered by G-Scholar within two years of the WP's issue date.

[3] Top journal publications include only publications within three years of the WP's issue date. Top journals are defined as journals with impact weights >=0.33, as defined in Combes and Linnemer (2003).

Table A4. Numbers of Working Papers Published in Top Journals in 1999–2009

BCA	Number of Publications	Weight	IDB	Number of Publications	Weight
Canadian J. of Economics	8	0.5	J. of Development Economics	5	0.5
J. of Money, Credit, and Banking	3	0.67	Economics Letters	3	0.5
J. of Monetary Economics	2	0.67	J. of International Economics	2	0.67
J. of International Economics	2	0.67	J. of Money, Credit, and Banking	1	0.67
International Finance	2	0.17	J. of Policy Modeling	1	0.17
Quarterly J. of Economics	1	1	Economic J.	1	0.5
European Economic Review	1	0.67	J. of Banking and Finance	1	0.5
Open Economies Review	1	0.33	Review of Economics and Statistics	1	0.67
Review of Financial Studies	1	0.33	J. of Law, Econ, and Organization	1	0.5
Applied Economics	1	0.33	International Economic Review	1	0.67
FedSF	**Number of Publications**	**Weight**	**FedBG**	**Number of Publications**	**Weight**
Economic Review (FED SF)	22		J. of Money, Credit, and Banking	30	0.67
J. of Monetary Economics	14	0.67	J. of Monetary Economics	24	0.67
J. of International Economics	9	0.67	J. of Banking and Finance	17	0.5
J. of Money, Credit, and Banking	8	0.67	International Finance	15	0.17
J. of Econ Dynamics and Control	8	0.5	Review of Economics and Statistics	12	0.67
J. of the Japanese & Int'l Econ	8	0.33	J. of International Economics	10	0.67
American Economic Review	6	1	J. of International Money and Finance	8	0.33
Economic J.	6	0.5	J. of Economic Dynamics and Control	8	0.5
Quarterly J. of Economics	5	1	American Economic Review	7	1
Macroeconomic Dynamics	5	0.17	J. of Business and Economic Statistics	7	0.67
FedNY	**Number of Publications**	**Weight**	**OECD**	**Number of Publications**	**Weight**
Economic Policy Review (FED NY)	73	0.17	International Finance	2	0.17
J. of Money, Credit, and Banking	9	0.67	J. of International Economics	1	0.67
J. of Monetary Economics	8	0.67	Review of Economics and Statistics	1	0.67
J. of International Economics	7	0.67	Economic Policy	1	0.33
Review of Economics and Statistics	4	0.67	American Economic Review	1	1
American Economic Review	4	1	Brookings Papers on Econ Activity	1	0.33
J. of Int’l Money and Finance	2	0.33	Labour Economics	1	0.33
J. of Econ Dynamics and Control	2	0.5	World Development	1	0.33
J. of Finance	2	0.67	World Economy	1	0.33
Quarterly J. of Economics	2	1	J. of Economic Growth	1	0.33
WB	**Number of Publications**	**Weight**	**IMF**	**Number of Publications**	**Weight**
J. of Development Economics	24	0.5	IMF Staff Papers	66	0.33
World Bank Economic Review	22	0.33	J. of International Economics	21	0.67
World Development	14	0.33	J. of Development Economics	16	0.5
J. of Banking and Finance	12	0.5	J. of International Money and Finance	10	0.33
J. of International Economics	11	0.67	J. of Banking and Finance	7	0.5
World Economy	10	0.33	Review of Economics and Statistics	7	0.67
World Bank Research Observer	8	0.17	American Economic Review	6	1
Econ. Dev. & Cultural Change	8	0.17	Applied Economics	6	0.33
J. of Int'l Money and Finance	7	0.33	J. of Money, Credit, and Banking	6	0.67
J. of Money, Credit, and Banking	7	0.67	International Finance	6	0.17

Note: Based on G-Scholar search results in Combes and Linnemer (2003). Impact weights as in Combes and Linnemer (2003).The table covers those WPs issued during 1999–2003 that were published in journals with impact weights equal to or greater than 0.17. This is a more inclusive definition of “top journals” than used in the rest of our analysis, Journals with impact weights less than 0.17 are not included in our publication impact analysis.

Table A5. Two-Year Citation Rates of Working Papers Issued in 1999–2003[1]

			Number of Citations Cited by Different Institutions									Citations by the 8 Peer Institutions			
WP Issuing Institution	Number of WPs in Sample	Number of WPs cited by 8 Institutions	BCA	FedSF	FedBG	FedNY	IDB	IMF	OECD	WB	Citations by others, outside 8 peer institutions	Total number of cites by 8 peer institutions	NSC: non-self-citing[2]	NSC ratio = NSC/total citations in 8 peer institutions	SC ratio = 1–NSC.ratio[2]
BCA	236	73	104	1	6	1	–	11	–	2	187	125	21	0.168	0.832
FedSF	452	134	16	194	77	8	–	16	–	5	1,020	316	122	0.386	0.614
FedBG	821	226	22	104	299	52	6	52	1	26	1,600	562	263	0.468	0.532
FedNY	343	112	19	18	45	102	6	31	–	16	779	237	135	0.57	0.43
IDB	322	73	1	7	3	1	147	12	–	17	229	188	41	0.218	0.782
IMF	1,211	433	21	45	29	12	28	637	–	112	1,967	884	247	0.279	0.721
OECD	440	41	8	3	13	1	8	14	–	22	824	69	69	1	0
WB	1,166	399	9	15	24	2	50	77	–	695	1,622	872	177	0.203	0.797
Total	4,991	1,491	200	387	496	179	245	850	1	895	8,228	3,253	1,075	0.33	0.67

Note: Based on RePEc database (i.e., from *t*–1 to *t*+2), sorted by citing institution.
[1] The table reports the sources of citations by the peer institutions within two years of the paper's issue date.
[2] NSC stands for "non-self-citing" (i.e., citations by publications of the other seven peer institutions); SC stands for "self-citing."

Table A6. Five-year Citation Rates of Working Papers Issued in 1999–2003[1]

WP Issuing Institution	Number of WPs in Sample	Number of WPs Cited by 8 Institutions	Number of Citations Cited by Different Institutions: BCA	FedSF	FedBG	FedNY	IDB	IMF	OECD	WB	Citations by others, outside 8 peer institutions	Citations by the 8 Peer Institutions: Total number of cites by 8 peer institutions	NSC: non-self citing[2]	NSC ratio = NSC/total citations in 8 peer institutions	SC ratio = 1–NSC.ratio[2]
BCA	236	110	188	2	13	3	–	17	–	2	411	225	37	0.164	0.836
FedSF	452	158	40	271	114	16	–	47	–	11	2,144	499	228	0.457	0.543
FedBG	821	277	54	153	441	97	16	125	1	58	3,468	945	504	0.533	0.467
FedNY	343	138	32	39	78	165	18	59	–	33	1,682	424	259	0.611	0.389
IDB	322	89	2	9	5	2	208	37	–	48	517	311	103	0.331	0.669
IMF	1,211	570	64	72	46	26	61	1,175	–	214	4,480	1,658	483	0.291	0.709
OECD	440	80	11	14	18	2	14	51	–	61	1,917	171	171	1	0
WB	1,166	526	18	26	28	6	122	168	–	1,175	4,231	1,543	368	0.238	0.762
Total	4,991	1,948	409	586	743	317	439	1,679	1	1,602	18,850	5,776	2,153	0.373	0.627

Note: Based on RePEc database.

[1] The table reports the sources of citations by the peer institutions within five years of the paper's issue date (i.e., from *t*–1 to *t*+5, where *t* stands for the WP's issuing year).

[2] NSC stands for "non-self-citing" (i.e., citations by publications of the other seven peer institutions); SC stands for "self-citing."

Table A7. Two-Year Citation Rates of Working Papers Issued in 1999–2006[1]

WP Issuing Institution	Number of WPs in Sample	Number of WPs Cited by 8 Institutions	Number of Citations Cited by Different Institutions									Citations by the 8 Peer Institutions			
			BCA	FedSF	FedBG	FedNY	IDB	IMF	OECD	WB	Citations by others, outside 8 peer institutions	Total number of cites by 8 peer institutions	NSC: non-self citing [2]	NSC ratio = NSC/total citations in 8 peer institutions	SC ratio = 1–NSC.ratio[2]
BCA	422	129	170	2	9	3		18	–	2	310	204	34	0.167	0.833
FedSF	838	240	42	298	102	28	4	49	–	18	1,901	541	243	0.449	0.551
FedBG	1,247	351	34	154	451	80	7	100	1	29	2,536	856	405	0.473	0.527
FedNY	519	176	24	35	65	177	6	45	–	22	1,176	374	197	0.527	0.473
IDB	587	116	1	11	5	1	230	31	–	24	354	303	73	0.241	0.759
IMF	2,016	838	40	82	43	32	63	1,346	–	202	3,597	1,808	462	0.256	0.744
OECD	901	72	15	3	14	1	8	40	–	30	1,355	111	111	1	0
WB	2,092	748	10	21	28	4	83	151	–	1,294	2,985	1,591	297	0.187	0.813
Total	8,622	2,670	336	606	717	326	401	1,780	1	1,621	14,214	5,788	1,822	0.315	0.67

Note: Based on RePEc database.
[1] The table reports the sources of citations by the peer institutions within two years of the paper's issue date.
[2] NSC stands for "non-self-citing" (i.e., citations by publications of the other seven peer institutions); SC stands for "self-citing."

Table A8. Top 10 Sources of Citations for Working Papers Issued in 1999–2006

BCA Journal or paper series name		IDB Journal or paper series name	
Bank of Canada Working Papers	154	RES Working Papers	222
IMF Working Papers	18	NBER Working Papers	50
NBER Working Papers	12	IMF Working Papers	31
Fed Kansas,Proceedings	12	Policy Research Working Paper Series	24
Bank of Canada, Technical Reports	12	Central Bank of Chile WPs	19
CEPR Discussion Papers	11	CEPR Discussion Papers	18
Central Bank of Chile WPs	8	William Davidson Institute WPs	14
Res.Bank of New Zealand WPs	8	Universidad Torcuato WPs	10
Borradores de Economia	8	MPRA Paper	7
Fed Atlanta Working Paper	8	Boston College WPs in Economics	7
OECD Journal or paper series name		**FedSF** Journal or paper series name	
IZA Discussion Papers	73	NBER Working Papers	259
CEPR Discussion Papers	58	CEPR Discussion Papers	139
MPRA Paper	53	ECB Working Paper Series	97
IMF Working Papers	39	Fed SF WP in Applied Economics	95
NBER Working Papers	38	FED BOG Finance and Economics	57
CSLS, Int'l Productivity Monitor	35	Fed SF Working Paper Series	55
Policy Research WP Series	30	MPRA Paper	54
CESifo Working Paper Series	29	Fed SF, Proceedings	52
WIFO, working paper	24	IMF Working Papers	48
Kiel Working Papers	22	Fed SF, Economic letters	44
FedBG Journal or paper series name		**FedNY** Journal or paper series name	
FED BOG Finance and Economics	302	NBER Working Papers	158
NBER Working Papers	293	Fed NY Staff Reports	130
ECB Working Paper Series	147	CEPR Discussion Papers	88
CEPR Discussion Papers	142	ECB Working Paper Series	46
FED BOG International Finance	139	IMF Working Papers	45
IMF Working Papers	99	FED BOG Finance and Economics	37
Fed NY Staff Reports	56	FED BOG International Finance	26
Fed SF WPs in Applied Economics	53	Fed NY Economic Policy Review	24
EconWPA Macroeconomics	49	MPRA Paper	23
Fed SF, Proceedings	42	Bank of Canada Working Papers	22
WB Journal or paper series name		**IMF** Journal or paper series name	
Policy Research WPs	1294	IMF Working Papers	1334
NBER Working Papers	219	NBER Working Papers	342
CEPR Discussion Papers	182	CEPR Discussion Papers	218
IMF Working Papers	147	Policy Research Working Paper Series	202
(UNU-WIDER) Working Papers	103	MPRA Paper	132
MPRA Paper	97	EconWPA Macroeconomics	88
IZA Discussion Papers	81	ECB Working Paper Series	87
RES Working Papers	80	Central Bank of Chile Working Papers	75
Central Bank of Chile WPs	49	CESifo Working Paper Series	73
William Davidson Institute WPs	49	William Davidson Institute WPs	65

Note: Based on RePEc data. We only include citations within two years after a WP is issued (i.e., from t–1 to t+2, where t stands for WP's issuing year.

References

Aizenman, Joshua, and Ken Kletzer, 2008, "The Life Cycle of Scholars and Papers in Economics—the 'Citation Death Tax,' " NBER Working Paper No. 13891 (Cambridge, Massachusetts: National Bureau of Economic Research). Forthcoming, *Applied Economics.*

Boxwell, Robert J. Jr., 1994, *Benchmarking for Competitive Advantage* (New York: McGraw-Hill).

Camp, Robert C., 1989, *Benchmarking: The Search for Industry Best Practices That Lead to Superior Performance* (Milwaukee, Wisconsin: ASQC Quality Press).

Cole, Jonathan R., and Stephen Cole, 1973, *Social Stratification in Science* (Chicago: University of Chicago Press).

Combes, Pierre-Philippe, and Laurent Linnemer, 2003, **"**Where Are the Economists Who Publish? Publication Concentration and Rankings in Europe Based on Cumulative Publications," *Journal of the European Economic Association*, Vol. 1, No. 6, pp. 1250–1308**.**

Coupé, Tom, 2003, "Revealed Performances: World Wide Rankings of Economists and Economic Departments: 1990–2000," *Journal of the European Economic Association*, Vol. 1, No. 6, pp. 1309–45.

Eijffinger, S.C.W, J. de Haan, and K. Koedijk, 2002, "Small Is Beautiful: Measuring the Research Input and Output of European Central Banks," *European Journal of Political Economy*, Vol. 18,No. 2, pp. 365–74.

Goodfriend, M., R. König, and R. Repullo, 2004, "External Evaluation of the Economic Research Activities of the European Central Bank," paper prepared for the Executive Board of the European Central Bank (Frankfurt).

Jaffe, Adam, and Manuel Trajtenberg, 1996, "Flows of Knowledge from Universities and Federal 39 Labs: Modeling the Flow of Patent Citations over Time and Across Institutional and Geographic Boundaries," NBER Working Paper No. 5712 (Cambridge, Massachusetts: National Bureau of Economic Research).

________, 2002, "*Patents, Citations, and Innovations: A Window on the Knowledge Economy* (Cambridge, Massachusetts: MIT Press).

Jansen, D.W., 1991, "Ranking Federal Reserve System Research Departments by Publications in Professional Journals," *Journal of Macroeconomics*, Vol. 13, No. 4, pp. 733–42.

Jondeau E. and Pagès H., 2003, "Benchmarking Research in European Central Banks," Working Paper (Paris: Banque de France).

Kalaitzidakis, P., T. Mamuneas, and T. Stengos, 2003, "Rankings of Economic Journals and Economic Departments for the 1995–2000 Period," *Journal of the European Economic Association*, Vol. 1, No. 6, pp. 1346–66.

Laband, David, and Michael Piette, 1994, "The Relative Impact of Economic Journals," *Journal of Economic Literature*, Vol. 32, pp. 640–66.

Lubrano, M., L. Bauwens, A. Kirman, and C. Protopopescu, 2003, "Ranking Economics Departments in Europe: A Statistical Approach," *Journal of the European Economic Association*, Vol. 1, No. 6, pp. 1367–1401.

Merton, Robert K., 1968, "The Matthew Effect in Science," *Science*, Vol. 159, pp. 56–63. Available at: www.garfield.library.upenn.edu/merton/list.html.

________, 1973, *The Sociology of Science* (Chicago: University of Chicago Press).

Posner, Richard A. 2000, "An Economic Analysis of the Use of Citations in the Law," *American Law and Economics Review*, Vol. 2, No. 2, pp. 381–406.

Rapoport, D., and K.-M. Yi, 1997, "Research Output by Federal Reserve System Economists Employed as of January 1997," internal paper (New York: Federal Reserve Bank of New York).

St-Amant, Pierre, Greg Tkacz, Annie Guérard-Langlois, and Louis Morel, 2005, "Quantity, Quality, and Relevance: Central Bank Research, 1990–2003," Bank of Canada Working Paper 2005-37 (Ottawa: Bank of Canada, December).

IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/11/02

Review of IMF Research on Monetary Policy Frameworks

Kenneth Kuttner, Petra Geraats, and Refet Gürkaynak

May 20, 2011

Address. 700 19th Street, N.W., Washington D.C. 20431, U.S.A. Telephone. +1 202 623 7312 Fax. +1 202 623 9990 Email. feedback@ieo-imf.org Website. www.ieo-imf.org

 BP/11/02

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

Review of IMF Research on Monetary Policy Frameworks

Prepared by Kenneth Kuttner, Petra Geraats, and Refet Gürkaynak

May 20, 2011

Abstract

This study examines the technical quality of a sample of 60 IMF working papers on monetary policy frameworks. It found that the quality and value-added of IMF research on monetary policy frameworks varied considerably. Most of the working papers issued in 1999–2008 posed interesting policy-related questions and many were very skillfully executed. Some were cited extensively and made major contributions to the literature. Yet many of the papers were substandard, raising the possibility that some of the IMF's advice might rest on less than rock-solid research. Many of the flaws in the weaker papers stemmed from a lack of proficiency with empirical methods, or from failure to articulate a well-focused research question within the context of a coherent and appropriate theoretical framework, or from less than full and detailed description of the data and methods used to generate the results. The paper offers recommendations on screening, feedback, and documentation that could help address these weaknesses.

JEL Classification Numbers: E42, E52, E58, E61, and F33

Keywords: exchange rate regimes, inflation targeting, monetary policy transmission, policy rules, central banking

Author's E-Mail Address: kenneth.n.kuttner@williams.edu

Contents **Page**

I. Overview

1. The IMF's research agenda is diverse and ambitious. Even within the papers devoted to monetary policy frameworks, the research poses a wide range of interesting and policy-relevant questions. Many of the papers included deal with conventional topics such as monetary policy transmission, inflation targeting, and policy rules. Others address more idiosyncratic issues such as corruption, bimetallism, West African Economic and Monetary Union, and currency board design. Research methods range from pure theory to data description, and everything in between. Moreover, in line with the Fund's mission, IMF research deals with policy-relevant issues in virtually every country in the world. With its depth of expertise, institutional knowledge, and access to data, the IMF is uniquely situated to do much of this research—and in fact, some of it could have been done *only* at the IMF.

2. This report evaluates the quality of IMF working papers dealing with the general topic of monetary policy frameworks. In performing our evaluation, we were cognizant of the wide range of research topics and methods, the special challenges encountered in doing research on developing countries, and the special role that research plays in the formulation of the IMF's policies.

3. Our review led us to the following conclusions. First, the working paper series contains a number of first-rate papers, comparable in quality to those published in top-tier professional journals—in fact, many of them have been published in prestigious journals. Second, the working paper series contains many papers that, while not suitable for publication in academic journals, are nonetheless worthwhile and well executed. Third, the quality of the analysis varies greatly across papers. While the IMF produces some very good research, a significant proportion of the working papers in monetary policy frameworks suffer from some serious weaknesses. Those focusing on a particular country or region, most of which are based on selected issues papers (SIPs), tended to be among the weaker papers. Fourth, a common problem among the less well written papers is a lack of care in the presentation of quantitative material (charts, tables, etc.). And fifth, data sources and empirical methods are often not carefully documented. Most of these weaknesses are easy to remedy, however, and we believe ample opportunities exist for raising the average quality of the working papers and the research they contain.

4. Our recommendations fall under three headings. The first is quality control: more screening should be applied to the papers submitted for inclusion in the working paper series. We suggest a number of ways to do this: clear articulation of standards, consistent application of these standards to all types of papers, greater selectivity, a greater degree of accountability for supervisors signing off on submissions, and internal or external reviews. Our second recommendation is for more feedback and constructive criticism prior to submission to the working paper series. There is a great deal of research and writing expertise within the institution, and this should be harnessed more effectively to bring the weaker papers up to the IMF's standards. This involves institutionalizing mechanisms for

communication, collaboration, and constructive criticism involving research and country staff. Guidelines for clear exposition and presentation could also be helpful. The third heading is documentation. Because of its important role as an input to policy, IMF research should be as transparent and replicable as possible. Every paper published in the working paper series should therefore contain enough information on data sources and methods to allow other researchers to understand and correctly interpret the results. To that end, the IMF might consider establishing a repository for nonproprietary data at least to allow internal and external researchers to access the data used in the studies.

5. The production of high-quality research is an integral part of the IMF's mission, and the creation of working papers is an important component of the Fund's research output. Although one of our recommendations is to do more screening, ultimately the aim of our recommendations is to raise the quality of research, not just to keep weaker work away from the public eye. IMF working papers should be high quality and representative of the institution's research output.

6. The report is organized as follows. Section II describes the evaluation criteria, and how the panel interpreted them in the context of the IMF's research mission. Section III explains the categories into which the papers were grouped for comparison purposes. Section IV provides an overview of the evaluation ratings given by the panel and of other quantitative indicators of the quality of the 60 papers. Section V highlights a number of features shared by the strongest papers in the sample, and calls attention to some common sources of weakness. Section VI provides conclusions and recommendations. The annex summarizes the procedures used in conducting the review.

II. Evaluation Criteria

7. Each of the 60 papers had both a primary and a secondary reviewer (see Annex 1). The reviewers evaluated each paper according to 10 evaluation criteria, which were grouped into three categories, as shown in Table 1: first, the coherence and focus of the framework; second, the quality of the analysis; and third, the quality of the output. The framework category, for example, included assessments of the paper's contribution to the literature, and whether the research question was clearly focused. The analysis category included ratings of the proficiency with which the empirical and/or theoretical methods were applied, and whether the paper satisfactorily addressed robustness issues. The output category included criteria that included writing style and the degree to which the conclusions were grounded in the analysis and were relevant for policy purposes.

Table 1. Research Quality Indicator Form

Evaluation criteria	Rating[1/]				
	S	AA	A	BA	U
Framework					
1. Question is well posed and clearly focused					
2. Places work within the context of existing literature					
3. Specifies contribution to existing literature					
Analysis					
4. Uses an appropriate theoretical/conceptual framework[2/]					
5. Uses appropriate data and empirical methods proficiently					
6. Includes critical discussion and/or robustness analysis of results					
Output					
7. Writing is clear and well organized					
8. Adds value relative to existing research					
9. Conclusions are firmly grounded on the analysis					
10. Articulates policy relevance of findings					
Overall rating					

[1] The rating scale is as follows: "superior" (S);"above average" (AA); "average" (A); "below average" (BA); "unacceptable" (U).

[2] This includes whether there was excessive use of technique relative to the question being posed.

8. Papers were given marks ranging from "superior" (S) to "unsatisfactory" (U) on each of the ten criteria, with "above average" (AA), "average" (A), and "below average" (BA) falling in between. Each reviewer then aggregated those scores into a single score representing his or her judgment as to overall paper quality. The reviewers were not asked to use a uniform scheme in aggregating the individual scores, but the three categories of evaluation criteria received roughly similar weights.

9. The panel members agreed that the IMF's emphasis on practical policy-oriented research called for emphasizing different aspects of quality than those typically used in evaluating academic research. One dimension we paid close attention to in our review was the inclusion of an appropriate theoretical or conceptual framework. We were not as demanding on this criterion as we might have been if we were evaluating academic work, and we did not insist that every paper include a fully worked-out theoretical model. But in our review it became clear that the most effective papers were those that were able to frame their analysis with reference to a coherent theoretical framework. Those papers that did not include an explicit model invariably benefited from a thorough summary of the relevant theoretical considerations.

10. Technical proficiency is essential to quality research, of course. But some methodological qualities that are highly valued in academic research, such as theoretical novelty or technical innovation, received less weight in our evaluation than they would have in an academic context. We recognized that focused, well-executed research is highly valuable in a policy setting, even when the research is too incremental in nature to be of interest to an academic journal. Technique was valued to the extent that it furthered the analysis, and the use of "excessive" technique was not viewed favorably.

11. In judging the robustness of their analysis, we held IMF working papers to a higher standard than we would have applied to academic research. Our view was that since it is likely to be used as an input into policy decisions, research issued under IMF auspices need to be clear about the limits to the analysis, and what caveats apply. "Overselling" results is never advisable, but is especially inappropriate in policy research. The papers' use in the policy process also makes it essential that their conclusions and policy recommendations be based firmly on the analysis. Papers with conjectural, overgeneralized, or boilerplate conclusions were marked down.

12. We put a premium on clarity of writing and presentation, again reflecting the imperative to communicate the findings to policymakers and nonspecialists. We also expected empirical working papers to fully report their estimation methods and diagnostic tests, and to document the data sources used in the analysis. Enough information should be provided to allow a reader to correctly interpret and replicate the results.

III. Paper Classification

13. To facilitate comparison, we divided the 60 papers into four categories. The first category consists of 24 empirical papers. Among these, 17 emphasize formal statistical analysis and the other 7 are primarily descriptive. Several of the descriptive papers involve the creation of new data sets, and their main purpose is to document the construction and describe the data set contents and characteristics.

14. Thirteen working papers can be classified as theory papers. Among them, eight can be characterized as "pure" theory, in the sense that they work out the implications of an original model, or develop an extension to an existing theoretical model without empirical testing or simulation of results. Five involve the simulation of calibrated macroeconomic models, most often to evaluate the performance of monetary policy rules.

15. The third category comprises 19 studies of particular countries or regions. Typical issues addressed in this set of papers included the suitability of inflation targeting, or the

choice of the appropriate exchange rate regime. Fourteen of these country studies were adapted from SIPs, and some even retain the SIP formatting and style.[1]

16. All the country studies involve either regression analysis or data description, and therefore could have been included in the empirical category. The reasons for breaking them out as a separate category are threefold. First, the purpose of country studies typically differs from that of more academically oriented research papers, in the sense that they are commissioned to address a specific issue of particular relevance to IMF policymaking. In that sense, they are even more important for the IMF than pure research that addresses questions in the general literature. Second, much country-specific research encounters obstacles such as limited or poor data. And third, because the country studies tend to share a similar set of weaknesses, it is useful to discuss them collectively.

17. The final category consists of four hard-to-classify papers, which we will refer to as "thought pieces" for lack of a better term. The four contain little or no original empirical or theoretical research; all involve some combination of literature survey and commentary.

18. Needless to say, some of the working papers, except for the thought pieces, could have been put into more than one category. A number of the papers develop theoretical models and then proceed to estimate or test the models empirically. These papers were classified according to our judgment about their primary contribution.

IV. Overview of Quantitative Indicators

19. Table 2 shows the panel's overall ratings for the 60 papers by both the primary and secondary reviewer, broken down by evaluation criteria, with each paper receiving two ratings. The overall rating, shown in the last column, was 3.2, slightly above the 3 score corresponding to the "average" rating. The median score fell into the "average" category. Most of the papers on the low end of the distribution were deficient on more than one of the criteria we identified. But even many of the highly ranked papers on our list also fell short on some criteria.

20. The uneven quality of the papers is reflected in the distribution of the scores, and the standard deviation of 0.9 suggests a considerable amount of variability.[2] The dispersion is immediately evident in the histogram of paper scores shown in Figure 1. The good news in the picture is the high proportion of papers scoring 4 ("above average") or higher: 20 papers

---

[1] The set of papers evaluated included 15 former SIPs. One of these was classified as an empirical paper, rather than as a country study.

[2] In order to eliminate the variance introduced by discrepancies between the ratings of the two reviewers for each paper, the average of the two reviewers' scores was used in calculating the standard deviation and the histogram displayed in Figure 1.

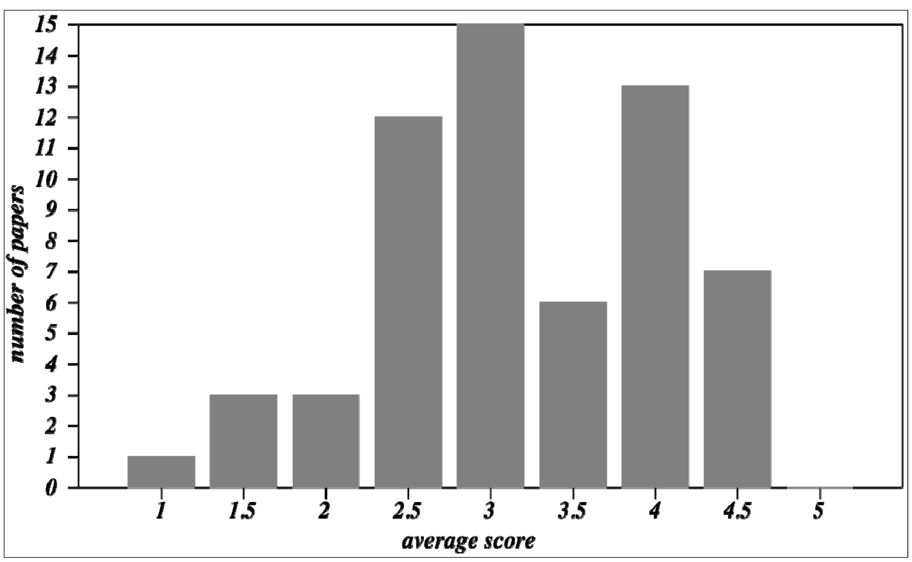
(33 percent) fell into this upper tail. The bad news is that 7 papers (12 percent) scored 2 ("below average") or below.

Table 2. Summary of Overall Ratings

Rating	Empirical	Theoretical	Country studies	Thought pieces	All
Superior	7	4	1	0	12
Above average	18	12	4	1	35
Average	15	8	17	1	41
Below average	7	2	14	4	27
Unsatisfactory	1	0	2	2	5
Total number of papers	**48**	**26**	**38**	**8**	**120**
Average[1]	**3.5**	**3.7**	**2.7**	**2.1**	**3.2**
Standard deviation[2]					**0.9**

[1] Based on a scale where S=5 and U=1.
[2] Standard deviation across papers of the average of the two reviewers' scores.

Figure 1. Distribution of Average Paper Scores



A. Ratings by Category of Paper

21. The breakdown by category reveals some additional insights. The average scores for papers in the theoretical and empirical categories were 3.5 and 3.7, respectively, approximately halfway between the "average" and "above-average" ratings. The median for both groups was "above-average." The country studies tended to score considerably lower. Their average numerical score of 2.7 puts these papers collectively somewhere between the

"average" and "below-average" ranges. Another revealing metric is the number of scores falling into the "below-average" and "unsatisfactory" bins. Among the empirical papers that number is 4 out of 24 (17 percent), and among the theoretical papers, 1 out of 13 (8 percent). Among the country studies, by contrast, 8 out of the 19 ratings (42 percent) were either "below average" or "unsatisfactory."

22. Because the papers in the "thought piece" category contained little or no original empirical or theoretical research, it was difficult to rate them using the same criteria as those used for the other 56 papers. Many of the evaluation criteria listed in Table 1 did apply, however, and in our review we relied more heavily on attributes such as having a clearly focused question and using an appropriate theoretical or conceptual framework. It should also be noted that because this category contains only four papers, our findings may not be representative of the broader population. Having said that, we judged the "thought piece" papers to be considerably weaker than those in the other categories. Our average score was only 2.1, and the median rating was "below average."

23. There was less than complete agreement among the panel members on the strongest and weakest papers among the 60, and relatively little overlap between the lists of papers that were nominated in each category. To some extent, this lack of agreement reflected the different weights that the individual reviewers assigned to specific strengths and weaknesses. But with few exceptions, the papers nominated for the "best paper" category by one reviewer tended to be rated highly by the other reviewer. The same was true for nominees in the "weak paper" category.

24. What is more interesting is the distribution of the "best" and "worst" papers across the 4 categories of paper we identified. Seven on the "best paper" list came from the theory category, 12 came from the empirical category, and 3 were country studies. None was a "thought piece." The country study and "thought piece" categories were overrepresented on the "weakest paper" list, with 7 and 3, respectively. Among the weakest papers, 6 came from the "empirical" category, and only 1 was a theoretical paper, suggesting perhaps that the barriers to entry inherent in this kind of work prevent substandard papers from being written.

B. Citation Counts and Publication Success of the Sampled Papers

25. Publication data and citation counts are often used in academia to gauge research quality. These are imperfect metrics for judging IMF research for the simple reason that they undervalue quality research that might be highly valuable from a policy perspective and yet is insufficiently novel or technically innovative to appeal to the referees and editors of academic journals.

26. The citation and publications statistics do contain some information, however, at least with regard to the visibility of the more academically oriented papers in the sample. Figure 2 is a histogram of citation counts tabulated by Research Papers in Economics (RePEc), and Figure 3 is a comparable histogram using data from Google Scholar. As might be expected,

both are highly skewed to the right. In the RePEc data, the median citation count is two and the mean is six; for Google Scholar, the comparable statistics are thirteen and thirty-two. Still, the numbers suggest that IMF research *is* being cited, and also that a respectable number of papers are getting a great deal of attention.

Figure 2. RePEc Citation Counts

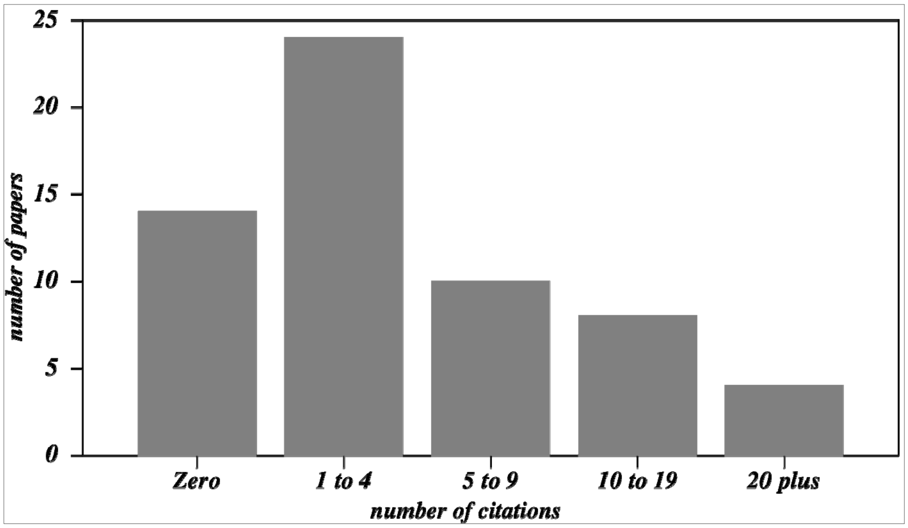
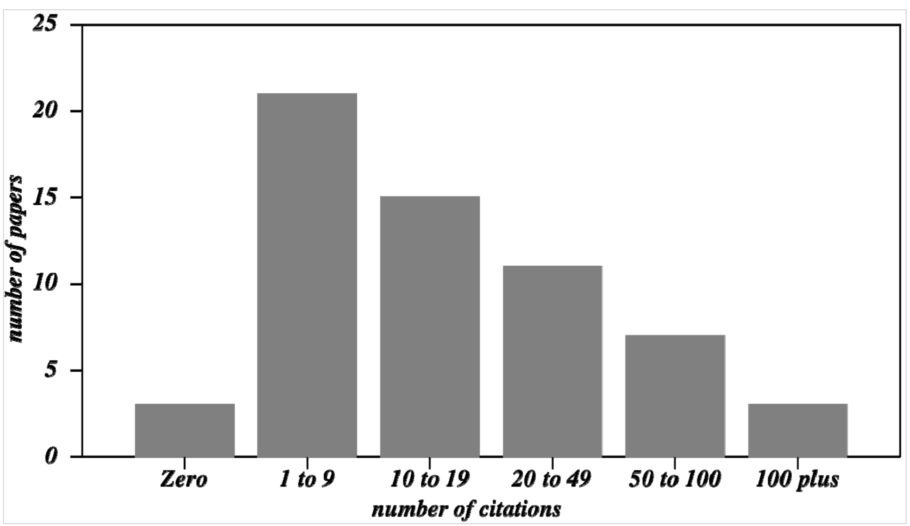


Figure 3. Google Scholar Citation Counts



27. It is also worth nothing that 24 of the 60 papers (including 3 originally issued as SIPs) have been published in journals (including *IMF Staff Papers*) or as book chapters. This may overstate publishing success to some extent, because it includes some non-refereed venues. On the other hand, it may understate the number of publishable papers, because many of the more recently written papers are surely still under review.

28. Another point worth mentioning is that although our paper rankings were intended to measure attributes other than publishability, there is a link between our rankings and the citation counts. To illustrate this, we regressed the citation *ranking* (i.e., the most highly cited paper was a "one") on the overall rankings given by the primary and secondary reviewers, plus a constant, for both the RePEc and the Google Scholar indexes. Both scores had the correct (negative) sign, and were significant at the 5 percent level or better in the RePEc regression.

29. The adjusted R-squared from the RePEc regression was only 0.23. A low R-squared is not surprising, however, given our use of a broad range of quality metrics. A set of criteria that focused more narrowly on academic "publishability" would likely have resulted in a better fit. Indeed, several of the papers scoring highly on our criteria had very modest citation counts. (A few of these were more recent papers, for which low citation counts are to be expected.) One widely cited paper received only average marks.

C. Disaggregation by Department, Affiliation, and CoAuthorship

30. Our ratings suggest some differences in paper quality across IMF issuing departments, although it would be hazardous to make strong inferences based on such small samples. The European Department (EUR) was most highly represented, contributing 12 of the 60 papers. Monetary and Capital Markets (MCM) was next with 11, followed by Research (RES) with 10.[3] Six papers came from the IMF Institute (INS). The area departments (other than EUR) individually accounted for relatively few papers, but they collectively contributed 21.

31. With mean scores of 3.7 and 3.6, respectively, RES and MCM stand out as producers of high-quality research. Papers from EUR also tended to be rated highly, earning a mean score of 3.3. INS received a low mean score of 2.5, but more than the usual amount of caution is warranted in generalizing this number given the small number of papers on which it is based. The area departments (other than EUR) tended to produce the weakest papers, and their average was 2.5. This finding is not surprising, given that most of the low-scoring country studies originate from these departments.

32. Disaggregating the data according to authors' affiliation and coauthorship provides some evidence for a connection between collaboration and paper quality. The 16 papers that were written jointly with at least one outside coauthor received a mean score of 3.6, while those written solely by IMF staff received 2.8. Internal collaboration was also associated with higher paper quality. Of the 40 papers written exclusively by IMF staff, coauthored papers averaged 2.9, while sole-authored papers averaged 2.7. And although the small sample size

[3] The figures for MCM also cover the former Monetary and Exchange Affairs Department (MAE) and Monetary and Financial Systems Department (MFD).

makes it hard to generalize, collaboration with IMF staff may benefit outsiders as well: the 4 papers written by external authors earned a mean score of 3.3, compared with 3.6 for papers involving collaboration between IMF staff and external authors.

V. Strengths and Weaknesses

33. The numeric ratings summarized above conceal as much as they reveal. A more useful critique would involve a comprehensive cataloging of strengths and weaknesses of the papers we reviewed. Naturally, this is hard to do in detail for all 60 papers. But we were able to identify a number of attributes shared by most of the papers we rated highly. Similarly, the weaker papers tended to share many of the same flaws.

A. Common Attributes of Strong Papers

34. **Theoretical framework.** The best papers we reviewed all had a coherent conceptual or theoretical framework. Naturally, this was the case for all the papers in the theory category. The stronger empirical papers were also grounded in economic theory. In some cases, the framework was developed explicitly in an optimizing structural model that could then be applied to the data. In one study, this involved estimating a fully specified dynamic stochastic general equilibrium (DSGE) model. Another paper that was quite successful in this regard used some relatively simple theory to discuss the persistence of inflation under alternative exchange rate regimes, and then tested those implications empirically.

35. Not all empirical analysis is amenable to such a high degree of integration with theory, of course, but it was often possible to bring theory to bear in a looser way. One study successfully used a standard macro model to motivate the identifying restrictions on a Galí-style structural vector auto-regression (VAR). A perfectly standard monetary exchange rate model served as a satisfactory foundation for another paper's empirical analysis. In other cases, an informal but careful reference to existing theory proved effective in motivating and interpreting the empirical analysis. One very good empirical paper on the credit channel of monetary policy transmission framed its analysis with a concise but thorough reference to the conclusions from others' theoretical research. We found that high marks on this dimension tended to carry over into other areas as well. Papers with a well-developed framework tended to be focused, and easier to interpret. The discipline of a model, even an implicit one, usually helped to keep the conclusions firmly grounded in the analysis.

36. **Appropriate empirical methods.** The use of appropriate statistical techniques was another hallmark of the strongest empirical papers in our sample. We emphasized the appropriateness of the method, rather than the sophistication, because in many cases relatively simple econometric methods were perfectly adequate for the question at hand—especially given the data limitations often encountered in IMF work. Many papers using simple ordinary least squares or common panel data methods received high marks on this criterion. In fact, the primary reviewer of one of our most highly ranked papers wrote that it had "not much fancy econometrics," and that it "used the right data and the simplest methods

possible." Another well-regarded paper in the portfolio was awarded high marks more for its development of an original data set than for its econometric analysis.

37. Of course more sophisticated methods were sometimes appropriate, as in the application of panel cointegration techniques to exchange rate behavior, and Bayesian DSGE methods to monetary policy transmission. Similarly, several of the calibrated macro modeling exercises received high marks.

38. **Robustness checks, diagnostic statistics, and caveats.** As mentioned earlier, one criterion to which we paid close attention was whether the paper included enough information for a reader to be able to assess the robustness of the analysis. Good empirical papers provided robustness checks and diagnostic statistics, and took care to acknowledge their limitations. One particularly thorough paper presented two sets of estimates, one based on maximum likelihood and another using the generalized method of moments, along with a Monte Carlo simulation. Others presented structural VAR estimates based on alternative identifying assumptions.

B. Common Weaknesses

39. Not surprisingly, most of the papers on the low end of the distribution were deficient on more than one of the criteria we identified. But even many of the highly ranked papers on our list also fell short on some dimensions. Table 3 gives a breakdown of the ratings by evaluation criteria.

Table 3. Distribution of Ratings by Evaluation Criterion

Evaluation criterion	Distribution by rating (In percent)					Average Rating
	S	AA	A	BA	U	
1. Question is well posed and clearly focused	28	43	17	7	5	3.8
2. Places work within the context of existing literature	20	28	20	27	5	3.3
3. Specifies contribution to existing literature	17	25	28	25	5	3.2
4. Uses an appropriate theoretical/conceptual framework	15	12	30	22	22	2.8
5. Uses appropriate data and empirical methods proficiently	10	24	27	27	12	2.9
6. Includes critical discussion and/or robustness analysis of results	7	18	15	42	8	2.7
7. Writing is clear and well organized	12	27	35	25	2	3.2
8. Adds value relative to existing research	13	28	23	27	8	3.1
9. Conclusions are firmly grounded on the analysis	10	23	27	33	7	3.0
10. Articulates policy relevance of findings	12	23	32	27	7	3.2
Overall distribution[1]	**10**	**29**	**34**	**23**	**4**	
Overall average rating						**3.2**

[1] The distribution by rating criterion is for the primary reviewer only, whereas the overall rating is based on the average of the two reviewers' scores.

40. One of the most common flaws among the empirical papers was the absence of a coherent conceptual or theoretical framework, as reflected in the below-average score of 2.8 on this criterion. The papers scoring poorly on this criterion all involved reduced-form regression analysis. Their choice of variables was sometimes loosely motivated by theory, but the tenuousness of the theoretical link usually rendered their results unintelligible. A paper on equilibrium exchange rates, for example, cited theory to justify the inclusion of half a dozen or more variables in a cointegrating relationship, but the results from this "kitchen sink" regression were impossible to interpret. Another study neglected the likely endogeneity of one of its regressors—a problem that would have been immediately obvious had the analysis been framed with reference to a simple macro model. A third neglected to discuss the theory of optimal currency areas in its assessment of the desirability of moving to a fixed exchange rate regime.

41. A disturbing number of papers contained conclusions that bore at best a tenuous relationship with the core analysis. The mean score for this criterion was 3.0, but the primary reviewer rated one-third of the papers "below average" or below on this dimension. Some papers made little effort to connect their conclusions with the analysis, instead offering a set of unsubstantiated boilerplate conclusions based on conventional wisdom—for example, that fiscal imbalances should be addressed; that the exchange rate should be allowed to float; or that central bank policy should be predictable. One reviewer remarked that one paper's concluding section seemed to have been written first, with the rest of the paper added later. In other cases, the conclusions were overgeneralized, ignored inconvenient findings, or lacked appropriate caveats. In one paper, two of the three policy conclusions that were stated in the abstract could not be substantiated by the results without significant qualifications. Other papers exaggerated or "spun" their policy implications. One study, for example, extrapolated from a reduced-form analysis of trade flows to broader statements about the effects of the credibility of exchange rate pegs. At least two papers presented their findings as bearing on the implications of globalization, though they offered no analysis relating specifically to that issue.

42. Another common problem among empirical papers was a lack of technical proficiency, or the choice of an inappropriate empirical framework. The mean score for this evaluation criterion ("uses appropriate data and empirical methods proficiently") was 2.9, and every paper receiving a "below-average" or "unsatisfactory" overall rating from its primary reviewer received low marks on technical proficiency. Some papers used inappropriate methods given the data problems their authors faced—which in the case of developing economies are often severe. Standard econometric tools, such as VARs and cointegration methods, were sometimes applied mechanically with little attention to their limitations or their suitability to the available data. Issues arising from obvious breaks (due in one case to a civil war) or other anomalies in the data were sometimes inadequately addressed. Issues of trends and nonstationarity generally received insufficient attention.

Endogeneity and sample selection problems were often left unacknowledged and unaddressed, even in otherwise well-regarded papers. An example is a paper that excluded a certain de facto exchange rate regime from its analysis, but failed to discuss the sample selection issues that might have resulted from that decision.

43. Given the emphasis on policy implications, and the data problems inherent in doing this kind of research, our view is that a careful examination of robustness issues is essential. Several papers were marked down on this criterion, including some of those that scored well overall, and the mean score on this criterion was only 2.7. Papers involving model simulations were sometimes prone to this problem. One paper presented simulations based on only one set of parameters, and failed even to consider the possibility that the resulting estimated shock variances may have been exaggerated by measurement error. Another used de jure exchange rate regimes rather than the de facto classifications, leaving unanswered the question of whether similar results would have been obtained with the de facto regimes. Regression-based analyses often lacked robustness checks. Few papers rigorously assessed the results' sensitivity to outliers. For example, one paper claimed to have found evidence for a nonlinearity involving a threshold rate of inflation, ignoring the fact that only four countries in the sample exceeded the threshold. A large number of papers neglected to report the basic diagnostic tests that would have been required to correctly interpret their results.

44. Most of the papers were clearly written, and the mean score on this criterion was 3.2. But some of the weaker papers—and even a few of the good ones—were quite hard to follow. These expositional issues often surfaced in their introductions, which were sometimes long, verbose, or hard to follow. In at least one paper, the symbols used in the model were never defined, leaving the reader to guess their meaning. One paper with more than 50 equations regularly referred to equations several pages back without specifying their equation numbers.

45. A number of papers suffered from slipshod presentation of tables and graphs. Axis labels and units were omitted. Some tables were incomplete, others disorganized. Some papers, particularly those involving model simulations, presented their results as vast numerical tables when a graphical presentation would have been more effective. At least one paper referred to a nonexistent table. Sometimes the description of results in the text was inconsistent with the numbers shown in tables. Another common problem was the inadequate description and interpretation of tables and graphs. For instance, one paper devoted only one short paragraph to the empirical results for 9 out of the 11 countries in its sample.

46. Most papers articulated a well-defined research question, and indeed the mean score on this criterion was 3.8. A few did not, however. Papers lacking a clear focus also tended to be weak on other dimensions, such as the inclusion of a coherent conceptual framework. The "thought piece" papers tended to be quite weak in this regard. Several of the country studies suffered from a vaguely specified research question, and ended up doing little more than

providing a narrative account of a country's monetary and/or exchange rate policies, some reduced-form regression results, and some generic recommendations.

47. Finally, an issue common to many of the papers (again, even some of the highly ranked ones) was inadequate documentation of data and methods. One paper, for example, mentioned that it used a foreign price index in its regression, without specifying exactly what the data represented or where they came from. Another failed to specify the countries included in the sample, the data frequency, and the data source. Replicability is the litmus test for good scientific analysis, and all too many working papers fell short on this criterion.

VI. Conclusions and Recommendations

48. The IMF is clearly doing a lot of things right when it comes to research. Most of the papers pose interesting policy-related questions, and many are executed with great skill. Some have been cited extensively and have made major contributions to the literature.

49. Many of the papers in the working paper series are substandard, however, and the inclusion of these papers in the series reflects badly on the institution. Even more worrisome is the thought that the IMF's advice might rest on less than rock-solid research. But the good news is that much low hanging fruit is waiting to be picked. Some modest interventions would have strengthened most of the papers, and brought them up to respectable standards of quality. Very few were completely unsalvageable.

50. Just getting the basics right will help a great deal. Doing the simple things correctly in 50 percent of the papers would do more to bring up the average than the application of leading-edge methods in 5 percent. A refresher course on basic panel data methods, for example, would yield a bigger "bang for the buck" than a workshop on panel cointegration. In most cases, core first-year Ph.D. level theory would be more helpful in providing empirical papers with the necessary theoretical foundations than the latest advances in DSGE modeling. This is not to say that the more sophisticated methods should be discarded, of course, and it is very important for the IMF to be doing state-of-the-art research. Indeed, even among the strong papers there is scope to move a little closer to the methodological frontier. But in allocating resources optimally, it should be kept in mind that the bulk of IMF's research has and will continue to be practical, applied, policy research relying on tried-and-true econometric methods.

51. With these considerations in mind, our recommendations encompass three elements: screening, feedback, and documentation.

A. Screening

52. Quality control is a serious problem in the working papers on monetary policy frameworks, but it can be remedied with a more systematic screening process. Some of the papers included in the series probably should never have been written. Others merited

inclusion but needed a great deal more work in order to bring them up to a minimal standard of quality. Here are some suggested ways in which the screening process might be improved.

- **Standards.** The IMF should establish a clear standard for quality, perhaps broken down using the same criteria we used for our evaluation. For example, authors should know that a coherent conceptual or theoretical framework is a requirement for publication. Writing and presentation standards should also be clearly articulated. These might include the requirement that all tables and graphs be self-contained, that the relevant diagnostic tests be reported, and that the data and methods be adequately documented.

- **Accountability.** The supervisor signing off on a paper should be responsible for ensuring that the quality standards are met. The signature should not merely be pro forma.

- **Referee reports.** An internal refereeing process should be established to complement supervisory approval. Especially for the more technical papers, it may be more appropriate to have the research evaluated by an in-house expert, not just the immediate supervisor. The European Central Bank sends its working papers out for external referee reports, and the IMF might consider instituting a similar procedure.

- **SIPs.** The IMF should be more selective in turning SIPs into working papers. Not all SIPs will be suitable for inclusion in the working paper series. Those that are should be held to standards comparable to those that apply to other working papers.

- **Nonresearch papers.** The IMF should exclude from the working paper series those papers that contain little in the way of original research. Applying this criterion would weed out the weak "thought pieces" we found in our sample, whose uneven quality reflected badly on the overall quality of the papers we reviewed. Exceptions might be appropriate for survey papers, if these were judged to add value relative to the existing literature.

B. Feedback

53. The second theme in our set of recommendations involves feedback and constructive criticism. Many of the flaws we identified in the weaker papers stemmed from a lack of proficiency with empirical methods, or the failure to articulate a well-focused research question within the context of a coherent and appropriate theoretical framework. This should not happen at an institution with the intellectual depth of the IMF. The expertise required for generating consistently competent papers is already in place. What is needed is to make that expertise more accessible.

- **Collegial feedback.** Our sense in reading the 60 papers is that many would have benefited greatly from the comments of one or two colleagues with relevant expertise.

This does not happen automatically, of course, especially since the people with the relevant expertise may work in a different part of the organization. For that reason, we suggest that the IMF create forums or institutionalize processes that would facilitate peer-level feedback. This could take the form of informal internal seminars—preferably set up in such a way as to cut across organizational boundaries.

- **Editorial feedback.** The editorial process is another possible channel for providing constructive criticism. In addition to (or perhaps instead of) the normal sign-off process, the IMF should consider appointing an editor, or creating an editorial board, for the working paper series. Though one job of the editor or the board would be to weed out inappropriate or unsalvageable papers, an equally important function would be to provide feedback and to make suggestions about any needed changes or improvements. Ideally, the editor would be able to direct the author to someone else within the institution who could give the necessary technical or methodological guidance.

- **Collaboration.** Many of the papers we reviewed showed great depth of knowledge about local and institutional features but were weak in terms of technical proficiency. Others had the opposite problem: strong in methods but weak in local knowledge. This suggests that there are unexploited gains from trade: experts on time series econometrics could work with country authorities to refine the empirical work, for example, or someone who has studied monetary unions in other contexts could be enlisted to work on a project involving the CFA franc zone, strengthening the conceptual framework and facilitating a more comparative analysis.

- **Writing resources.** Judging from the working papers we reviewed, many economists might benefit from some additional resources dedicated to the improvement of writing skills. Simply distributing a style guide for IMF publications could be an effective way to take the rough edges off some of the working papers. This could include templates for formatting regression output and tabular material, and guidelines for annotating and labeling graphs. And because everyone's writing can benefit from the suggestions of an editor, the IMF might consider dedicating a staff member to perform this function. Some of the presentation flaws, such as the failure to define the symbols used, could easily have been detected and corrected by a proofreader.

C. Documentation

54. Our third recommendation concerns documentation. The credibility of any kind of research rests on its reproducibility, which in turn requires a full and detailed description of the data and methods used to generate the results. Credibility is particularly important for an institution like the IMF, whose research forms the foundation for its policy recommendations.

55. Many of the papers we reviewed fell short on this criterion. The IMF is not alone in this regard, of course, and authors of academic publications are often remiss when it comes to documenting data and methods. In an effort to remedy this, authors of IMF working papers should know that adequate documentation is an essential ingredient of good research. Checking for this documentation should be an integral part of the editorial or administrative approval process.

56. The IMF could do even more in this regard. Academic best practice is for authors of published papers to make available online, or by request, the data and programs used in the analysis. This kind of openness is particularly valuable for the IMF's country studies, where locally based economists may have a keen interest in reproducing and extending the IMF's analysis. Undoubtedly, a lot of information is already being shared informally. But as part of a broader effort to improve the documentation of its research, the IMF might consider establishing an online repository for nonproprietary data of potential interest to outside researchers.

Annex. Evaluation Procedures

The IEO forwarded to the review panel 60 randomly selected from the 187 IMF working papers produced since 1999 on the topic of monetary policy frameworks. Each of the three panel members was designated as the primary reviewer for 20 papers and the secondary reviewer for another 20. All papers were therefore read by two panel members. Thus, each paper had both a primary and a secondary reviewer. Primary reviewers prepared a brief report on each assigned paper and completed a Research Quality Indicator Form (reproduced as Table 1 in main text). Secondary reviewers provided a numerical evaluation, and in some cases, a supplementary narrative critique. The panel members were also asked to identify the 4 strongest and the 4 weakest papers from both their primary and secondary lists.

The panel convened on Monday, March 15, 2010 to discuss the paper evaluations. The first substantive topic was an examination of the strongest and weakest papers. Next was a discussion of any major discrepancies between the primary and secondary reviewers' scores. Although the panel members broadly agreed on the papers' merits and demerits, 11 of the papers were assigned scores that differed by 2 or more increments. Our discussions of these papers led in many cases to the convergence of views, and often one of the two reviewers was willing to concede that an important feature of the paper had been overlooked. In only a few cases did a significant divergence of views persist after going over the paper in question, and most of the major rating discrepancies would have been eliminated had there been a formal reconciliation process.

The third task of the panel meeting was to compare notes on each of the 60 papers. This gave the panel an opportunity to discern recurring themes and issues in the body of research as a whole. The meeting concluded with a general discussion of the strengths and weakness of the papers examined, and some preliminary planning for the drafting of the panel's report.

IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/11/03

Review of IMF Research on Tax Policy

Robin Boadway, Christopher Heady, and Henrik Kleven

May 20, 2011

 BP/11/03

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

Review of IMF Research on Tax Policy

Prepared by
Robin Boadway, Christopher Heady, and Henrik Kleven

May 20, 2011

Abstract

This study examines the technical quality of a sample of 60 IMF working papers that focus on revenue and tax policy. It found significant variability in the quality of those papers. The papers were generally well motivated and focused on policy issues that were relevant for many countries. The papers were generally well written and mostly set within the context of the relevant literature. But many fell short in the analytical execution of the research, including the formulation of the model, the innovativeness of the approach, and the empirical or theoretical analysis. This resulted in lower scores for value added than for exposition. Fund researchers' reliance in some areas of research on a limited number of established and sometimes dated approaches may reflect an overly inward-looking approach to research. The study offers recommendations for research program management.

JEL Classification Numbers: A1, E6, H2, H3, H6

Keywords: tax policy, fiscal policy

Author's E-Mail Address: boadwayr@econ.queensu.ca

Contents **Page**

I. Overview

1. To assess the technical quality of tax policy research produced by the IMF over the period 1999–2008, the review panel evaluated a sample of 60 IMF working papers drawn randomly from the 120 such papers that were issued during the period on tax and revenue issues, including fiscal policy papers focusing on tax issues.[1]

2. The panel concluded that the best IMF working papers in tax-policy-related areas are very good by international standards, as reflected by the fact that much of this research is published in strong international journals. At the same time, the top papers by such standards represent only 25 percent of the working papers. There is a large group of average papers in the middle that are mostly unpublished, and some papers that are considerably below publication standards. The top papers come mainly from a small number of researchers, an above-average proportion of whom are from two IMF departments—Fiscal Affairs (FAD) and Research (RES). In the middle and lower categories, there is limited diversity in empirical approaches, and almost no use of techniques that are now becoming commonplace in academic policy-related research. Much of the Fund's tax policy research uses rather mechanical applications of long-established approaches that have limited applicability to real-world policy problems. This suggests a tendency to be too inward-looking in topic selection. There is limited use of outside expertise, which when used often produces fruitful results.

3. We recognize the constraints that IMF researchers face. Much of the research emanates from country missions and advisory work, which means that topics are constrained and so is the time available. These limitations pose challenges for improving the quality of papers in the middle and lower quality ranges.

4. This said, the credibility of the research program would be much improved if the proportion of papers published in refereed outlets were raised significantly and if more of the papers were jointly written with external coauthors. To help Fund researchers keep current with developments in the economics literature, the panel recommends more use of opportunities for training, seminars, and interaction with outside scholars. To foster collaborative research and possibly reduce the cost of empirical research in the longer term, the panel recommends establishment of a data repository for empirical papers.

---

[1] The three international scholars who made up the review panel have diverse backgrounds and experience but share a common interest in tax policy research and analysis. They have published widely in the major academic journals, especially in public economics. Their experience includes posts as editors and associate editors of major journals, as well as experience in tax and fiscal policy research at the Organization for Economic Cooperation and Development. Together, they have skills in public economic theory and policy analysis and modern empirical approaches to tax policy, including recent micro-data-based approaches and natural experiments. Despite their different individual backgrounds, experience, and perspectives, considerable consensus emerged both in their initial evaluations and their overall evaluation of the research.

5. This report is organized as follows. Section II describes the evaluation criteria and how the panel interpreted them in the context of IMF's research mission. Section III is an overview of the evaluation ratings given to the 60 papers by the review panel, and of other quantitative indicators of the papers' quality. Section IV explains the categories into which the papers were grouped for comparison purposes. Section V discusses the concerns that surfaced from the evaluation and Section VI offers conclusions and recommendations. Annex 1 outlines the evaluation procedures used and Annex 2 provides illustrative data.

II. Evaluation Criteria

6. The quality of the research was judged partly by the extent to which it added value—whether by generating new knowledge, broadening the understanding of fiscal policy, or developing new policy frameworks. To assess the value added we rated the soundness of the motivation of the analysis; the relationship of the analysis to the existing literature; the use of state-of-the-art analytical techniques, both theoretical and empirical; the quality and originality of the analysis; the implications of the analysis for policy purposes; and an awareness of the sensitivity of the results to the assumptions and parameters used in the analysis. These are the same criteria as are used to evaluate similar literature found in peer-reviewed outlets, and especially in reputable economics journals that publish public economics and fiscal policy papers.

7. We also considered that the research should be judged against the role and objectives of the IMF. Thus the motivation of the papers and their potential relevance for policy advice and understanding should reflect the Fund's policy interests, and the research should include adequate country and institutional knowledge and an understanding of which research questions are relevant and which policy recommendations can be feasibly implemented. While academic research might be motivated by a search for pure knowledge or innovation, and be rewarded accordingly, IMF research presumably complements the institution's role in evaluating country policies and offering advice. The outcome of that advice has more riding on it than academic output, so IMF researchers must be prudent on that account. This reasonably constrains the choice of topics and the strong orientation toward policy research.

8. Each of the 60 papers had both a primary and a secondary reviewer (see Annex 1). The reviewers rated each paper on ten evaluation criteria using a rating scale of one to five, where the ratings were labeled "superior" (S), "above average" (AA), "average" (A), "below average" (BA), and "unacceptable" (U). As shown by the Research Quality Indicator Form reproduced in Table 1, the evaluation criteria fell into three categories: Framework (referring to the framing of the research question in the context of the literature); Analysis (referring to the theoretical or conceptual framework, the empirical analysis, and the critical discussion and robustness analysis); and Output (referring to the clarity of writing, the value added by

the research, the conclusions, and the policy relevance). In addition, reviewers were asked to provide an overall rating for each paper.[2]

Table 1. Research Quality Indicator Form

Evaluation criteria	Rating[1/]				
	S	AA	A	BA	U
Framework					
1. Question is well posed and clearly focused					
2. Places work within the context of existing literature					
3. Specifies contribution to existing literature					
Analysis					
4. Uses an appropriate theoretical/conceptual framework[2/]					
5. Uses appropriate data and empirical methods proficiently					
6. Includes critical discussion and/or robustness analysis of results					
Output					
7. Writing is clear and well organized					
8. Adds value relative to existing research					
9. Conclusions are firmly grounded on the analysis					
10. Articulates policy relevance of findings					
Overall rating					

[1/]The rating scale is as follows: "superior" (S); "above average" (AA); "average" (A); "below average" (BA); "unacceptable" (U).
[2/]This includes whether there was excessive use of technique relative to the question being posed.

III. Overview of Quantitative Indicators

9. It is useful to begin with a summary of the quantitative evaluations of the 60 papers, based mainly on the ratings given by the primary and secondary reviewers.

A. Overall Ratings

10. The overall rating given by all three primary reviewers to the 60 papers was 3.1, only slightly above an "average" rating. Two of the three reviewers gave average overall ratings of "below average" (2.7 and 2.9), while the third gave a rating "above average" (3.6). The ratings were slightly higher among the secondary reviewers.

[2] The individual reviewers chose how to interpret these ratings, and slight differences in their interpretations are apparent: two of the reviewers tended to judge the papers more positively than did the third reviewer, both in their primary and secondary reviews of the papers (see Annex 2). This likely reflected different interpretations of the letter grades.

11. The primary reviewers gave the 60 papers an overall score slightly above "average" (3.1). They rated 7 of the papers as "superior" overall; 10 as "above average;" 24 as "average;" 17 as "below average," and 2 as "unsatisfactory;" this was a single-peaked distribution in which 40 percent of papers were judged "average" and almost 30 percent were judged "below average." The secondary reviewers gave a similar overall rating of 3.2, but the distribution here was much flatter: 90 percent of the papers were roughly equally distributed across the "above average," "average," and "below average" categories (Table 2).

Table 2. Average Ratings by Primary and Secondary Reviewers Across Evaluation Criteria

Evaluation Criteria	**Primary Reviewer**	**Secondary Reviewer**
Question focus	3.7	3.7
Literature review	3.1	3.3
Contribution	3.1	3.2
Framework	3.1	3.1
Data	3.0	3.1
Robustness	2.9	2.9
Clarity	3.5	3.5
Value added	2.9	3.0
Conclusions	3.0	3.1
Policy relevance	3.1	3.2

12. The average judgments mask some differences across the three reviewers. The primary and secondary reviewers largely agreed on which were the top three papers: of the nine papers that were ranked in the top three by primary reviewers, six were also ranked in the top three by secondary reviewers, and two of the other three had similar overall ratings.[3] There was more disagreement about the nine lowest-ranked papers, with very little overlap between those ranked lowest by the primary and secondary reviewers. Nonetheless, the raw overall ratings for lowest-ranked papers were generally similar for both primary and secondary reviewers.

13. Three exceptions—papers to which the primary and secondary reviewers gave widely different overall ratings—highlight some of the differences among reviewers in judging quality. One was an empirical paper that uses panel data to estimate an aspect of government fiscal behavior. One reviewer rated this paper "above average," based on the relevance and motivation for the analysis, the somewhat innovative policy question being posed, and what

[3] In one anomalous case, a primary reviewer ranked a paper in the top three with a rating of "superior," while the secondary reviewer rated it "below average." This was a paper whose contribution lay mostly in devising a statistical technique for revenue-forecasting purposes. It was made difficult to judge by the fact that the only two references it made to the economics literature were two econometrics textbooks, along with two references to documents from public organizations. This failure to place the contribution of the paper in the context of the literature made it difficult to judge its quality and importance.

he considered to be the novel choice of some political explanatory variables. The other reviewer rated it "below average," based on deficiencies in econometric practices—and particularly an inadequate treatment of the potential endogeneity of some of the explanatory variables used, including the political explanatory variables to which the first reviewer attached importance.

14. The second paper was a case study of an administrative aspect of tax policy. One reviewer rated this paper "above average," based on the view that, although it was to a large extent descriptive, it outlined in detail the many policy issues that arise with respect to improving tax administration in a particular developing economy and evaluated the alternatives in light of one country's experience; in that sense, it was a useful input to the policy process in other developing countries. The second reader, who judged it "below average," took the view that despite the paper's policy relevance, the level of economic analysis and value added was limited.

15. The third exception was a paper that provided Monte Carlo simulations of an important policy issue in resource-dependent developing countries. One reviewer rated it "above average," based on the quality and novelty of the statistical analysis. The second reviewer found it "below average," on the basis that the policy objective embedded in the policy analysis was not a reasonable way to manage temporary resource revenues, since it assumed away the need to build up a stock for use by future governments and generations.

B. Ratings by Evaluation Criterion

16. Thus far we have focused on the overall ratings. It is also useful to look at the average ratings for each of the ten evaluation criteria. Using the scale of 1–5 for the ratings U–S, average ratings for each criterion can be calculated. The criteria on which the papers consistently scored highest were criterion 1 (question well posed and clearly focused) and criterion 7 (clear and well organized writing). Primary reviewers gave slightly higher than average ratings to criterion 10 (articulation of policy relevance). Crucially, the lowest ratings tended to be given for criterion 8 (value added) and criterion 6 (critical discussion and robustness analysis). Ratings on the remaining criteria were about average.

C. Ratings by Department

17. Though only one IMF department (FAD) produced half the papers in the sample (the other 30 were produced by 8 other departments), it is instructive to look at the origins of the papers that the panel rated highest and lowest.

18. Of the 9 highest-rated papers, FAD contributed a disproportionately large number: 6 were from FAD (out of the 30 that FAD produced), 2 were from RES (out of the 6 that RES produced), and 1 was from the Asia and Pacific Department (APD) (out of the 2 that APD produced). FAD and RES together produced a significant proportion of the working papers that were eventually published.

19. A further property of the nine highest rated papers is worth noting. One IMF staff member from FAD was an author or coauthor of three of them, while another FAD staff member was coauthor of two. One RES staff member was coauthor of two top papers.

20. Of the nine papers that were ranked lowest, four were from FAD, three were from Western Hemisphere Department (WHD) (out of four it produced), and one each were from the African Department (AFR) and the former Middle Eastern Department (MED). Thus, FAD produced disproportionately few of the lowest-rated papers, while WHD produced significantly more. (None of the papers from RES featured among the lowest ranked.)

D. Other Indicators of Quality

21. Of the 60 working papers reviewed, 15 were eventually published formally in journals and books (some of the more recent working papers might be published and would raise this number). The 15 represent only 25 percent of the working papers—a surprisingly low proportion compared with publication rates in most academic economics departments.

22. Of the papers published, 10 were published in journals outside the IMF,[4] 3 in the internal journal *IMF Staff Papers*, and 2 others in books of readings. Of the 10 journal articles, 5 were produced by FAD (3 of them by 1 author), and 4 by RES (2 of them by 1 author). Three were written jointly with outside authors, and the rest were written wholly by IMF staff members. These results suggest that research by FAD and RES tends to be viewed more positively, by academic standards, than research produced elsewhere in the Fund.

23. Citation counts listed by Research Papers in Economics (RePEc) and Google Scholar (GS) were provided for all 60 assigned papers. Only 5 papers had received more than 10 RePEc citations; all were FAD papers and most of them were published. Nine papers had received no citations at all, and 25 had received between 1 and 3. Among the 12 papers that were rated highest by either a primary or secondary reviewer, 1 had received 36 citations, 1 had received 6, 1 had 5, 3 had 3, and the remainder had received fewer than 3. Not surprisingly, the lowest-rated papers had much lower citation counts, with only 1 having received more than 2 citations. The pattern of GS citations is somewhat similar. Five papers had received more than 100 citations (the same 5 papers from FAD that had received more than 10 RePEc citations). A further 11 had received between 20 and 100 citations, and 31 had received fewer than 10. Among papers in the top-rated category, 1 had received 197 citations (the same 1 that had received 36 RePEc citations), while others had received 58, 33, 29, and 20, and the remainder 12 or fewer. The lowest rated papers showed much lower GS citation counts.

---

[4] Of the outside journals, one is a top general journal (*Economic Journal*), three are top field journals *(Journal of Environmental Economics and Management*, *Journal of International Economics*, and *Journal of Public Economics)*, two are in a good tax policy journal *(National Tax Journal)*, and four are other specialist journals *(Journal of International Money and Finance*, *Journal of Policy Modeling*, *Japan and the World Economy,* and *Review of World Economics)*.

24. It is also useful to look at the average ratings of papers by number and type of authors. External coauthorship seems to produce better papers: the average overall score given by primary and secondary reviewers to the 13 papers produced in collaboration with an external coauthor was 3.5, which is well above the average for all 60 papers. The papers with external coauthors were concentrated in the research category, as distinct from country studies or survey papers. Another 21 papers were single-authored by IMF staff members; they received an average overall rating of 3.0, slightly below the average for all 60 papers.

IV. Characterization of Research in Working Papers

25. The working papers on tax and fiscal policy are of different types, though some of them contain elements of more than one approach. Those of the first type are primarily research-oriented, though usually motivated by fiscal policy concerns. Most of them are empirical and some develop models that are calibrated to country data. Examples include studies of tax reforms or fiscal policy actions, evaluation of development aid, studies of tax smoothing and tax effort, and analyses of stabilization policies, including automatic stabilization. Many of the papers have potential policy applicability to groups of countries such as developing countries, resource-rich countries, transitional economies, or federal economies. They vary in their degree of innovation and value added, but in principle they are candidates for publication in refereed journals.

26. A second type of papers offers policy analysis. These are typically country-specific and use known research tools. Examples include studies of the intertemporal management of natural resource revenues, tax evasion studies, and analyses of country-specific tax reform and fiscal consolidation. They are often related to IMF mission work, and may have been released earlier as selected issues papers (SIPs). They are less likely to be publishable in good journals because they are not innovative, but to the extent that they use state-of-the-art analytical tools, they are valuable as inputs into the policy process of specific countries.

27. Survey and best-practice papers make up a third type. Examples include surveys of resource taxes and intergovernmental transfers, case studies in tax administration and tax reform, and critical discussions of different mechanisms for managing or financing public debt, such as collateralization of future revenues. Many are primarily descriptive, illustrating best practices as they have evolved in policy reforms in various countries, although good ones are also evaluative and emphasize lessons learned from country practices. Some address topics that do not lend themselves easily to standard theoretical or empirical analysis, for example because they involve complex administrative or institutional processes. Survey/best-practice papers typically do not appear in journals, but such papers are potentially very valuable to policymakers and researchers because they synthesize the relevant literature as it applies to particular policy issues in specific countries or country groupings. They can also be valuable for informing nonspecialists about broad policy areas and indicating alternative policy approaches to a given problem.

28. Each type of paper was judged to be appropriate for the IMF working paper series and was represented in the sample reviewed. Of the 60 papers in the evaluation sample, 27 were determined to be research papers, 21 to be country studies, and the remaining 12 to be survey or best-practice papers. Many of the papers were rated as good as they either advance the boundaries of knowledge or apply good analysis to inform the policy process.

29. On average, the papers that can be thought of as research papers are of significantly higher quality than survey or best-practices papers, which in turn are of higher quality than country studies. Based on the ratings assigned by the first and second reviewers, the average rating of the research papers was 3.5, that of the country studies was 2.6, and that of the survey papers was 2.9 (Table 3). Thus, the research papers were rated distinctly above average and the country studies well below average.[5]

Table 3. Average Ratings by Type of Paper

Type	**Number**	**Average Rating**
Research	27	3.5
Country study	21	2.6
Survey/best practices	12	2.9
External coauthor	14	3.5
Multiple internal coauthors	21	3.0

30. Perhaps not surprisingly, many of the papers were based on country-specific policy analysis. Fourteen of the 60 sample papers (about 23 percent) had been issued as SIPs before becoming working papers. Of the former SIPs, 10 were produced by area departments and the rest by FAD; a significant proportion were ranked "below average" by one or both their reviewers. The only one of the former SIPs to be eventually published in a journal received an "above-average" rating.

31. The papers that were formally published, especially those that appeared in the better journals, tended to be research papers, though two of the survey/best practices papers were published in good journals.

32. A number of themes repeated themselves in the research, often along with repetition of analytical approaches and techniques. One example is the management of revenues from nonrenewable resources, where several papers applied versions of the so-called permanent

[5] These relative ratings of the three types of papers were confirmed by classifying the highest and lowest rated papers. Of the 12 papers that were rated among the highest 3 by either the primary or secondary reviewer, 10 were research papers and 1 each was a country study and a survey paper. Of the 16 papers that were rated among the lowest 3 by a first or second reviewer, 9 were country studies, 4 were research studies, and 3 were survey papers. These findings reinforce the differences in quality between the research papers on the one hand and the survey and country studies on the other.

income hypothesis to determining the sustainable level of government spending out of finite-lived resource revenues, as discussed further below. Another was the empirical estimation of tax effort. A third was tax smoothing studies based on representative-agent models.

V. Evaluation of the Research

33. The quality and value added of the research varied considerably. The research was generally well motivated and focused on tax and fiscal policy issues that were relevant for many countries. The papers were also generally very well written, with conclusions carefully drawn where appropriate. Usually, the papers were set within the context of the relevant literature. Where many fell short was in the analytical execution of the research, including the formulation of the model, the innovativeness of the approach, and the empirical or theoretical analysis. This resulted in lower scores for value added than for exposition.

34. The relatively few theoretical papers tended to receive ratings above average. Most of the empirical papers were competent, although many of them shared a number of shortcomings.

35. The best of the papers—which included roughly those rated in the top three by the primary and/or secondary reviewers—comprised about one-quarter of the papers (roughly the number that were published). Almost all of them were research papers, either theoretical or empirical, although one was a best-practice paper dealing with an important issue of tax administration. Some were written in collaboration with outside researchers. All reported original or innovative findings that were almost all of value for policy purposes. Most had a well-articulated theoretical framework and used it either to study alternative policy interventions from a theoretical perspective or to test the predictions of the theory empirically, using good techniques. These papers compare with what one might find in the working paper series of a good research-oriented university, and they typically warranted publication in good, though not the top, international journals (the best of them could aspire to top journals). Compared with a good university working paper series, their orientation is much more toward policy and policy-applicable research results, and the proportion of publishable papers is considerably smaller. Of course, some of the papers that are not suitable for journal publication are nonetheless of some value to the international policy community.

36. The large group of papers in the middle-rated group, comprising at least half the papers, hovered around the “average” rating. Many of them were reasonably solid and well motivated but added rather limited value, and a good number had questionable implications for policy. Relatively few of them were published or publishable, and they would not compare favorably on that criterion with the average papers in working paper series at good research-oriented academic departments.

37. The weakest papers were a very diverse group. Though overall their policy motivation was well conceived, their execution and value added were wanting. They suffered

from different faults. Some were largely descriptive, using statistics to describe fiscal problems that particular countries faced, and offered limited value for policy purposes. Others mechanically applied readily available statistical tools to study a problem faced by a country, with little concern for analytical foundations or causal relationships. A few were theoretical in nature, but used theoretical models that were not well motivated and tended to be somewhat overspecified for the problem being addressed.

38. A few in the weakest group of papers seemed not to be current with theoretical findings in the literature or state-of-the-art empirical approaches. A small number of them were judged to be below the quality one would expect of a research-oriented working paper series, while others might have been made more suitable by some revision. The number of below-the-line papers was not excessively large, although their presence might detract from the overall reputation of the series. Several of the lowest rated papers were single-authored, perhaps revealing the value of collaborative work for quality control. The presence of low-quality papers might also indicate the lack of peer interaction and of oversight by department managers in ensuring that papers that are approved for publication have been developed sufficiently to meet minimal standards.

39. The panel members recognize that comparisons with university research papers may not be wholly appropriate. As noted at the beginning of this review, the goals of the Fund's research agenda differ in some important respects from those of academic research. And, at the same time, IMF researchers tend to have other responsibilities that restrict the time they can take doing research. More important, they may be more sheltered from the cutting edge of academic research and not completely up to date on the most recent techniques and criticisms of conventional empirical approaches. An understandable temptation is to use standard "canned" techniques, especially those that are not too time-consuming. Interaction with outside researchers can help alleviate these problems, as some of the papers illustrate.

40. The panel nonetheless identified a number of specific concerns with the research as a whole, and especially the papers that were judged below the top 20–25 percent. These concerns are discussed below.

A. Concerns About the Choice of Empirical Approaches

41. The breadth and diversity of IMF applied research approaches differs from what may be found in the academic world. In the view of the panel, the gap is greater than it should be.

42. A common problem in the empirical strategies followed by the papers was a neglect of, or failure to deal adequately with, problems of identification. This was a general problem in the many cross-country panel studies, but it also applied to some of the within-country analyses. Explanatory variables were often endogenous—for example due to simultaneity or omitted variables—and although this problem was sometimes acknowledged, it was typically not rigorously discussed and dealt with, and the instruments and identification assumptions chosen were not convincing. Where it was dealt with using instrumental variables techniques,

the instruments chosen were not always appropriate. Examples of empirical studies suffering from important identification problems include work on the determinants of tax revenue ("tax-effort studies"), the effect of foreign aid on growth, the effect of tax policy on income distribution, and the determinants of tax amnesties. In some cases, innovative arguments were used to justify the choice of particular explanatory variables—for example, political economy determinants of policies—but poor treatment of endogeneity concerns undermined the validity of the findings. Related to this, some of the empirical work seemed to be driven by the ready availability of a data set and standard canned techniques for estimation. A more useful approach would be to first think about an important question, and then search for the data and identify the variation needed to shed light on that question.

43. In general, too much emphasis was placed on cross-country studies and macro evidence, at the expense of within-country studies and micro evidence.

44. These features of the papers mean that many of the results reflect correlations rather than behavioral effects or causation. This indicates a need for some caution in inferring policy implications and conclusions from the analysis. Appropriate caution was not always used in drawing conclusions from suspect empirical analysis.

45. The mix of empirical approaches used in the sample papers does not reflect the evolution of empirical research in the economics literature. In most applied research fields, including development economics, public economics, labor economics, international economics, and industrial organization, research in the past two decades has adopted a so-called design-based approach, where the search for identification of causal effects takes center stage. Typically, this involves seeking natural experiments or randomized experiments where causal inference is made by comparing treatment and control groups and applying difference-in-differences estimation and other techniques. The innovation in such research partly involves finding actual experimental situations that allow one to analyze phenomena of interest, and this takes both time and ingenuity. The payoff comes in more reliable estimates of causal effects. But there were no examples of these approaches in our assigned papers.

46. Other approaches can also be used besides natural experiments. Structural estimation can make empirical analysis more persuasive by explicitly identifying behavioral relationships that are suppressed in reduced-form estimation. These approaches are becoming more common in industrial organization, labor economics, and macroeconomics, where structural estimation can be embedded in a calibrated model. They have some advantage in policy evaluation. Structural estimation has drawbacks, but does provide useful evidence complementary to that from other empirical approaches.

47. A focus on identifying variation and causal inference typically requires a micro-approach. It is very hard to identify causal effects of, for example, tax and transfer policy using aggregate data. Micro-data studies have been used increasingly in areas such as public economics, where more aggregated studies would mask important details. Such analyses are

more demanding because they can be time consuming and the needed data are not always readily available, sometimes because of confidentiality problems. While confidentiality problems may be a genuine constraint, increasingly researchers have found imaginative ways to obtain micro-data that might be used for policy analysis. This holds even in the field of development economics where the data limitations are most severe. The IMF may have a comparative advantage in getting access to country-based micro-data.

B. Lack of Diversity of Methodological Approaches

48. As noted above, some modeling themes and approaches tend to be repeated, in many cases quite uncritically, in about 50 percent of the sample of papers we were assigned to review. This tendency might reflect some inward-looking bias and if so, it calls for greater efforts to interact with outside researchers.

49. One example of such repetition is the adaptation of the permanent income hypothesis (PIH)—which was designed originally to explain consumer behavior in the face of lumpy and uncertain income—to the management of revenues from nonrenewable resources to finance government services. The question is what sustainable level of government expenditures, and by implication sustainable fiscal deficits, can be achieved given a finite-lived stream of resource revenues. Several of the working papers address the question of how to spread the benefits of these revenues into the future, especially the distant future when future generations are alive. They adapt the PIH rather simplistically to the problem by suggesting that the present value of revenues should be annuitized to yield a permanent flow of finance for government spending, albeit taking account of some complications like uncertainty and differences in views about the weight to be given to future generations. While there is nothing very wrong with doing such a calculation and the results are suggestive, the approach tends to be accepted rather too uncritically given the many other issues that are relevant—such as the need for public investment and poverty alleviation, as well as the commitment issues a government is likely to face. Some of these papers do address some of the other issues that come with resource wealth, such as the design and management of a resource revenue fund, policies for diverting resource revenues to the public sector, sharing of natural resource revenues between levels of government, and the implications of natural resources for economic development and governance, including well-known resource-curse effects.

50. A second such theme is intertemporal tax smoothing, where representative-agent models inspired by Barro's long-ago analysis are used to study the optimal path of fiscal balances. In representative-agent models, convex deadweight costs of taxation are the main criterion for evaluating the path of fiscal policies in response to shocks; issues of unemployment are completely suppressed; intergenerational redistributive issues are largely suppressed; and the aggregative nature of the models means that components of both the expenditure and revenue sides of the budget cannot be distinguished. From either an

academic or a policy advisory point of view, the value added of these exercises seems rather limited.[6]

51. Fiscal policy analyses based on the PIH approach or the classic tax-smoothing model are largely mechanical applications of existing theory. The use of existing theory is not a bad thing per se. Problems arise when a relatively limited set of theories dominates, so that the overall relevance of the research for informing policymaking is limited, and when the theories used are themselves outdated. For example, studies that rely on the currency-demand approach to estimating tax evasion are unduly restrictive, given that other approaches have superseded this. Again, the Fund researchers' reliance on a limited number of established and sometimes dated approaches may reflect an overly inward-looking approach to research topics.

52. A third example is estimating so-called tax effort using cross-country data. Overall, the tax-effort papers do not add much to understanding of fiscal capacity in developing countries. These studies generally involve the addition of political and institutional explanatory variables into a basic tax-effort regression equation, relating the tax/GDP ratio to variables such as per capita income, share of agriculture in the economy, and economic openness. They do this in a rather ad hoc manner, with no explicit recognition of the processes that the basic equation is intended to capture or of whether the additional variables make sense in this context. The fact that several papers produced in quick succession do almost the same thing—adding slightly different political and institutional variables and adopting different (and sometimes conflicting) identification assumptions—suggests that there could have been some benefit to coordination among the authors.

C. Concern About Ideological Perspective

53. Some, but not all, of the papers seemed to have an ideological predisposition, if only implicitly. Most commonly, this took the form of faith in market outcomes (e.g., with respect to evaluating taxes on financial transactions), especially in macro-based models in which unemployment plays a surprisingly small role, even in those studying responses to shocks. There was often a strong emphasis on efficiency as a criterion of policy, to the exclusion of equity, perhaps partly as the result of the representative-agent models that were used. We have already mentioned the emphasis on tax smoothing as the researchers' main criterion for evaluating dynamic fiscal policy. Sometimes, analyses were done from a normative point of view with limited recognition of governments' ability or interest to implement such policies; for example, commitment issues were often assumed away in dynamic models, including those dealing with the management of nonrenewable-resource revenues. Some papers

[6] The tax-smoothing model is still used in some academic research—an example being political economy models of debt that take account of political processes and constitutional constraints like balanced budget requirements. However, in these studies the emphasis is more on understanding government behavior in realistic public choice settings, and the tax smoothing model is a convenient first approach to evaluating outcomes, as opposed to being an end in itself.

showed a tendency to judge outcomes in non-U.S. economies, including those of the European Union and Russia, using institutional features of the U.S. economy; one example is the use of the U.S. pattern of labor mobility as a norm to judge labor mobility elsewhere. Given these examples, some observers might perceive the research and policy advice of the IMF as disproportionately emphasizing market-based or efficiency-based solutions in policy evaluation and prescription.

54. The risk of this perception should not be overestimated, however, and there are good reasons for not trying to eradicate ideological bias proactively. Normative approaches to policy analysis do in principle form a good basis for policy prescription—even if researchers understand well that normative prescriptions may not be politically palatable—and they provide useful benchmarks against which to judge actual policy outcomes. What is most important is that normative approaches should not neglect important market failures (such as unemployment, limited labor mobility, etc.) that can constrain the choice of optimal policies. Moreover, normative approaches should also be sensitive to the full range of possible normative objectives, considering both efficiency and equity. Given the different views that reasonable people can have about equity objectives and the importance that most governments attach to these objectives, addressing the sensitivity of results and policy recommendations to alternative views is warranted.

VI. Conclusions and Recommendations

55. The research we reviewed has both strengths and weaknesses. The range of types of research is healthy, given the role of research as a complement to the other activities of the IMF, and encompasses policy-oriented applied and theoretical research, policy analysis and evaluation, and survey papers emphasizing best practices. Almost all the papers are very well written, with care taken to relate the research to existing literature and to draw lessons where appropriate for policy. On the other hand, with several exceptions, the value added in terms of innovation and original insights was limited, and this was reflected in the small proportion of the papers that were eventually published in good journals. There was considerable replication, often seemingly uncoordinated, of topics and approaches.

56. Most striking was the limited variety of theoretical and, especially, empirical approaches. Dominant theoretical approaches included the use of the representative-agent model with no room for redistribution or unemployment; tax smoothing as an objective of intertemporal fiscal policy; the permanent income hypothesis approach to managing natural resource revenues; and tax effort studies. Empirical strategies were dominated by panel estimation, often in a cross-country context; time-series estimation, including causation studies; and reduced-form estimation. Problems of endogeneity and causation were often not addressed satisfactorily or convincingly.

57. Working with macro-data will typically not be sufficient for studying the kinds of taxation issues that are of policy importance. The panel found almost no examples of the kind

of empirical techniques that are increasingly common in research elsewhere, such as natural experiments, micro-data based studies, and structural estimation. These approaches are typically more time-consuming than their precursors and require more ingenuity. Researchers must seek out data and situations where causal effects can be isolated by means of quasi-controlled experiments.

58. Given the critical role that policy advice plays at the IMF, quality control is a relevant consideration. Working papers are the face of IMF research to the outside world, and their quality reflects the credibility of the research the Fund is undertaking. The reviewers judged that very few papers were below the line of acceptance for the series, and in that sense, the Fund's vetting process seemed to work. The bigger problem concerns the average quality of most of the working papers: most of the papers were rated as probably not publishable in good journals, and those that were formally published were produced by a small number of departments and researchers.

59. The credibility of the research program would be much improved if the proportion of working papers published in refereed outlets were raised significantly, and if the number of researchers publishing papers were to increase. Mere changes in procedure like pre-refereeing working papers would have limited payoff, although the payoff might be raised if outside reviewers were involved. Encouraging outside participation in research would help, given that academic researchers have strong incentives for publication.

60. The quality of research is intimately related to the quality of researchers in the institution. Presumably, care is taken to hire the best possible new staff, typically in competition with comparable institutions and universities.

61. Even the best trained new researchers need to continue to grow in experience and knowledge, and must keep current with developments in the economics literature at large. Exposure to these ideas can be facilitated by opportunities for training, seminars, and interaction with outside scholars. There are already mechanisms within the Fund, such as the IMF Institute seminars, for keeping researchers current with new developments in economic research, including new methodologies or tools of analysis. The fact that some of the more successful research papers were written in collaboration with outside researchers suggests that these collaborations should be nurtured.

62. In the end, the quality of research is likely to be strongly influenced by the research culture of the institution. To change this culture requires a combination of strong leadership and adequate incentives and rewards for research performance. An atmosphere of collegiality and interaction is also helpful. Some of this is fostered by collaborative research. It might also be helped by such seemingly minor innovations as the establishment of a data repository for empirical papers. This would allow studies to be replicated. It would also encourage internal information sharing and could reduce the cost of empirical research in the longer term.

ANNEX 1. EVALUATION PROCEDURES

The IEO forwarded to the review panel 60 randomly selected IMF working papers produced since 1999 in the general area of tax and fiscal policy. Each of the three panel members was designated as the primary reviewer for 20 of the working papers and the secondary reviewer for another 20. Thus, each paper had both a primary and a secondary reviewer. Primary reviewers prepared a brief report on each assigned paper and completed a Research Quality Indicator Form (see Table 1 in main text). Secondary reviewers completed a Research Quality Indicator Form for each assigned paper. The panel members were also asked to indicate the best and worst three papers from both their primary and secondary lists.

The panel visited the IMF on Friday, February 26, 2010. They spent the day at a meeting with members of the IEO team evaluating IMF research, chaired by Hali Edison, Lead IEO Evaluator, and attended by Larissa Leony and Scott Standley. The panel was addressed by Ruben Lamdany, Deputy Director of IEO.

The meeting began with a discussion of the top and bottom rated papers of the panel members, including their strengths and weaknesses. Particular attention was given to cases where primary and secondary reviewers disagreed on top and bottom choices.

This was followed by a discussion of each of the 60 papers in chronological order by year of publication. For each paper, the primary reviewer gave a brief evaluation, and then there was a general discussion of the paper.

This was followed by a general discussion of the quality of the research overall, and a discussion of the emphasis of this report. The mechanics of drafting the panel's report were agreed to.

Staff at the IEO prepared a dossier of background material that included the following items:

- A summary of all research output produced by the IMF in all research areas over the period 1999–2008, of which the largest categories were working papers and selected issues papers (about 40 percent each).

- The numbers of working papers produced by each IMF department and by broad topic over the same period.

- The numbers of fiscal policy working papers by department and by broad topic.

- The numbers of fiscal policy working papers by department, focusing on broad tax policy issues from which papers for this review were selected, consisting of 120 papers of which half were from the Fiscal Affairs Department (FAD); the 120 papers included 101 from the revenue and tax category and the remainder from the general fiscal policy area that focused on tax issues.

- The allocation across departments of the 60 working papers randomly selected from the previous list of 120.

- A summary of which of the 60 working papers selected (a) had originally appeared as selected issues papers; (b) had been formally published; and (c) had been written by IMF authors alone, by external authors alone, or with external coauthor(s).

ANNEX 2. ILLUSTRATIVE DATA

Table A.2.1. Ratings for Papers by Department

	Primary Reviewer							
Department	Number	S	AA	A	BA	U	Average	
AFR	7	0	0	4	3	0	2.6	
APD	2	1	1	0	0	0	4.5	
EUR	4	1	0	2	1	0	3.3	
FAD	30	4	7	13	6	0	3.3	
INS[1]	3	0	1	1	1	0	3.0	
MCD/MED[2]	5	0	0	1	3	1	2.0	
RES	5	1	1	3	0	0	3.6	
WHD	4	0	0	0	3	1	1.8	
	Secondary Reviewer							
Department	Number	S	AA	A	BA	U	Average	Overall Average
AFR	7	0	3	3	1	0	3.3	2.9
APD	2	0	1	0	1	0	3.0	3.8
EUR	4	0	1	1	2	0	2.8	3.0
FAD	30	2	10	9	9	0	3.2	3.2
INS[1]	3	1	0	1	1	0	3.3	3.2
MCD/MED[2]	5	0	2	0	3	0	2.8	2.4
RES	5	3	1	0	1	0	4.2	3.9
WHD	4	0	0	1	3	0	2.3	2.0

[1] IMF Institute.
[2] The Middle Eastern Department became the Middle East and Central Asia Department in 2003.

Table A.2.2. Types of Papers

	Number	Percent of Total
SIPs that became WPs	14	23.3
WPs published	15	25.0
WPs with external coauthor	14	23.3
WPs with only external authors	3	5.0
WPs with only IMF affiliation	43	71.7

Table A.2.3. Overall Ratings by Primary and Secondary Reviewers

Rating	Primary Reviewer		Secondary Reviewer	
	Number	Percent of Total	Number	Percent of Total
S	7	12	6	10
AA	10	17	18	30
A	24	40	15	25
BA	17	28	21	35
U	2	3		0
Average[1]	**3.1**		**3.2**	

[1] Based on a scale where S=5 and U=1.

IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/11/04

An Examination of the Quality of a Sample of 60 Selected Issues Papers

Marcelo Selowsky and Marko Skreb

May 20, 2011

Address. 700 19th Street, N.W., Washington D.C. 20431, U.S.A. Telephone. +1202 623 7312 Fax. +1 202 623 9990 Email. feedback@ieo-imf.org Website. www.ieo-imf.org

 BP/11/04

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

An Examination of the Quality of a Sample of 60 Selected Issues Papers

Prepared by Marcelo Selowsky and Marko Skreb

May 20, 2011

Abstract

This study reports on an assessment of the technical quality of a sample of 60 Selected Issues Papers (SIPs), which were prepared as part of IMF Article IV consultations. To be effective, these papers need to address policy issues in a way that can be understood by the economic community in the country in question. About one-third of the evaluated papers were found to be better than satisfactory by both readers; they included very good and excellent papers. Good papers addressed well-defined and relevant questions and exhibited knowledge of country context—they made intuitive use of economics and the technique used matched the question. Approximately half of the papers were judged as satisfactory but exhibited specific elements of weakness. Finally, 12 percent of papers were judged to be unsatisfactory by both readers. SIPs prepared for advanced countries were typically found to be better than those for low-income countries. Common factors were identified in weak papers: they had a cryptic definition of the issue to be addressed and the relevance to the country was often not convincing; they showed a weak knowledge of basic institutional country context and often lacked the minimum data needed to address the issue; they exhibited an excessive eagerness to apply a quantitative technique without a good explanation of the economics behind the technique; and they seemed to be prepared with little time and with authors having not spent enough time in the respective country. The evaluation offers some recommendations to improve the management of SIPs.

JEL Classification Numbers: A1, F5

Keywords: Selected issues papers, IMF bilateral surveillance, quality comparison, research evaluation

Author's E-Mail Address: mselowsky@imf.org

Contents **Page**

I. INTRODUCTION[1]

1. This background paper for the IEO evaluation of IMF research examines the technical quality of the IMF's selected issues papers (SIPs), based on a sample of 60 such papers that were issued during 2004–08. SIPs serve as analytical background for key policy issues discussed during Article IV consultation missions, which are part of the Fund's surveillance mandate. The analysis contained in SIPs plays an important role in the provision of policy advice and therefore differs somewhat from research contained in working papers. This difference is an important element in shaping the judgments of this assessment. The introduction describes how the evaluation sample was drawn, the key dimensions of quality being assessed, and the evaluation criteria. Section II presents the overall results. Section III discusses the patterns of strength and weakness in the sample papers; and Section IV offers recommendations.

A. Sample Selection

2. The sample of papers that were evaluated was drawn from a universe of 1,110 SIPs issued by the IMF during 2004–08. The final sample excludes the SIPs for 21 countries that are featured in the in-depth case studies for the IEO evaluation of IMF research. The universe of papers was classified into three groups: low-income countries (LIC), 290 papers; middle-income countries (MIC), 472 papers; and high-income countries (HIC), 348 papers.

3. To draw the evaluation sample, 12 countries were selected from each of the first 2 country income groups and 6 from the third group, making sure that the selection included countries covered by different area departments in the IMF and deliberately overrepresenting the LIC and MIC groups. For each selected country, those SIPs issued in 2004–08 were identified. Using a random-number generator, the SIPs for each country were ranked and the first 2 were selected. The evaluation sample thus contained 24 SIPs from each of the first 2 country groups and 12 from the third country group, yielding a total of 60 SIPs.

B. Evaluation Criteria

4. The template reproduced in Table 1 shows the seven evaluation criteria, which fall into three categories. The first category of criteria refers to how clearly the question addressed in the paper is posed, and how well its relevance—particularly its policy relevance to the country in question—is explained. The paper needs to convince the reader of the importance of the question, particularly for the authorities and think tanks/academics in the country. Criteria in the second category examine how the question is addressed. Does the analytical framework and data being used match the question? Here the analytical framework is defined to include the technique being used as well as the economic reasoning behind it;

[1] We wish to thank Hali Edison, Ruben Lamdany, and Peter Montiel for useful inputs and suggestions.

technique alone does not suffice—the underlying economics reasoning must be clear and intuitive. A discussion of limitations and robustness is important to make the results credible and usable by the economic community in the country in question, and also to encourage further work on the topic. The third category of evaluation criteria examines how the conclusions are delivered. Are they clearly presented? Do they follow logically from the earlier analysis and are their implications for policy well articulated?

Table 1. SIP Quality Assessment

Evaluation criteria	**E**	**VG**	**S**	**MU**	**U**
The question					
Question is clearly posed and its relevance to the country well articulated					
Analysis					
Uses an appropriate theoretical/conceptual framework					
Uses appropriate data and empirical methods proficiently					
Includes critical discussion and/or robustness analysis of results					
Output					
Writing is clear and well organized					
Conclusions are firmly grounded in the analysis					
Articulates clearly the policy relevance of findings for the country in question					

5. Five ratings were used to assess the papers on each evaluation criterion: “excellent” (E), with a score of 5; “very good” (VG), with a score of 4; “satisfactory” (S), with a score of 3; “moderately unsatisfactory” (MU), with a score of 2; and “unsatisfactory” (U), with a score of 1. The average score of the paper is the simple average of the scores given to each of the quality dimensions. Each paper was read and scored independently by the two authors (readers).

6. In scoring the papers, judgments were made that take into account the particular role and major audience of SIPs. These papers accompany Article IV consultations and ought to address issues of high policy relevance. Their basic audience is in the country where the consultation is taking place and consequently SIPs need to address policy issues in a way that can be understood by the economic community in the country. Authors of SIPs need to be sensitive to the ability of economists in the country to absorb technical material—these economists themselves will probably not come forward and voice such problems.

II. Results

7. On the basis of the seven scores given to each paper each reader computed an average score for each paper ranging between 1 and 5. Using that average score, each reader independently ranked the 60 papers. As expected, the specific ranking of papers tended to differ between the readers. However, in ranking, the two readers independently identified

three common sets of papers: a top group of papers considered above "satisfactory" by both readers, including "very good" and "excellent" papers; a middle group of papers that ranged from "satisfactory" to "moderately unsatisfactory," namely having some weaknesses that reduced their potential; and a bottom set of papers considered "unsatisfactory." The groups are described below:

- ***The best 20 papers (33 percent of the sample)***. Both readers agreed independently that about one-third of the papers reviewed could be considered above "satisfactory." Papers in this group passed the test of being in the best third of papers of each reader. These papers had scores higher than 3, namely those either "satisfactory," "very good," or "excellent." Furthermore, both readers agreed independently on the top 9 papers (15 percent of the sample). These papers were judged to be "very good" or "excellent."

- ***The middle group (55 percent of the sample)***. This group included papers with scores that according to both readers ranged between 2 and 3, namely those between "moderately unsatisfactory" and "satisfactory." Readers often differed on the ranking of specific papers within the group: Sometimes one reader believed the paper to be below "satisfactory," while the other reader believed the paper to be above "moderately unsatisfactory."

- ***The bottom group of 7 papers (12 percent of the sample). These papers were considered "unsatisfactory."*** Scores in this group ranged from 1.5 to 2.0.

Some patterns

8. A further examination of the groups yields some interesting insights:

- ***There were large differences in quality of papers across income groups.*** The HIC papers were significantly more represented in the top group, with 8 papers. These 8 papers account for two-thirds of the HIC group. The MIC group was also represented with 8 papers, but these 8 papers account for only 40 percent of the MIC group. The results were significantly less favorable for the LIC group**.** Only 4 of 20 papers in the LIC group were included in the top one-third of papers.

- ***A clearly posed and relevant question was an important feature of the top one-third of the papers.*** Of the 40 combined scores assessing this dimension in the top 20 papers 36 scores were 4 or above. An important feature of the good papers was a clear articulation of the issue being addressed and its policy relevance.

- ***The unsatisfactory papers were rated "unsatisfactory" in basically all the criteria of quality***. Out of a total 98 possible grades given to this group (7 papers, 7 quality dimensions, 2 reviewers), 89 were scored 2 and below.

- ***On the other hand, the top 9 papers tended to be rated "very good" or "excellent" in most criteria of quality.*** Almost 90 percent of the 126 possible grades given to this group (9 papers, 7 quality dimensions, 2 reviewers) were above 4.

III. Qualitative Issues

9. This section discusses some problems that explain differences in the quality of papers, in particular the differences between the papers in the middle group and the papers in the top group. Weaknesses in addressing some common themes are also discussed.

The lack of precision in identifying the issue or why a question needs be addressed seemed adversely to influence the later stages of the paper

10. Sometimes many objectives were specified and it was not obvious why they were critical, for example, the impact of a relatively small policy change today on many (endogenous) variables over both the short and medium term, which is obviously an ambitious task. Sometimes it was not obvious why a problem was a welfare problem, be it a small deviation in inflation rates with respect to the target or a small correction needed in the fiscal balance. This lack of precision in defining the rationale for the problem being addressed made it difficult to assess the robustness of results and the policy conclusions.

Too much eagerness to quickly apply a particular technique with which the author was familiar; in some instances the question being addressed seemed to be an afterthought

11. Several of the weaker papers had a very cryptic definition of the problem to be addressed and its relevance, but then they moved quickly into specifying a complex econometric model (usually of a reduced form) with little intuitive explanation of its structure and why that model was critical to the problem at hand. Some papers attempted to derive a set of results for a large (ambitious) number of endogenous variables that were not intuitively linked to the question. Other papers in this category did not go through a process of explaining why that technique was critical to the question and what the economic forces were behind that reduced form.

Tendency to estimate country-wide/multi-sector econometric models when the issue could be addressed by a simpler model but with better and richer country-specific data

12. Some papers tried to estimate the impact of developments or shocks in a very large country (the U.S. or the rest of the world) on a smaller country (where the Article IV consultation was taking place) which is basically a "price taker" from the larger country (or the rest of the world). These papers tried to model fully *both* countries, using very aggregate data and many parameter assumptions. As a consequence, it is difficult for readers to assess the robustness of the results. An alternative could have been to simply simulate a specific shock emanating from the larger country/rest of the world on the small country, for example, taking the shock from the large country as given. In this case the modeling could

focus on the smaller country and concentrate the effort in further disaggregation and collection of country-specific data (that the author could have collected during his/her mission).

13. In several cases there appears to be an inclination to assess the impact of a disturbance on an excessively large number of variables. As a consequence many parameter assumptions need to be made. Why not focus on a few outcomes for which a simpler but more robust model could be estimated? One gets the impression of an eagerness of authors to go to "general equilibrium" too quickly, without first carefully posing the need for such an approach given the problem at hand.

Very quick use of cross-country econometric estimates then applied to the country in question

14. Some papers estimated cross-country relationships with very aggregate cross-country data and then applied the results (sometimes only at the end of the paper and rather cryptically) to the country in question (where the Article IV consultation was taking place). This was done without first examining whether such a relationship could have been estimated within the country by using time series data. Authors should make extra effort to improve their country database during their visits based on the specific hypotheses they want to test. There is nothing wrong with cross-country analysis but a chance should be given to time series country data analysis, at least for an initial exploration of country hypotheses.

Sometimes papers showed few efforts to collect data/hypothesis within the country; sometimes it seemed as if authors had seldom visited the country

15. The quick use of aggregate cross-country data for econometric analysis rather than sharpening specific hypotheses and disaggregating the analysis within a country was a major syndrome found in some papers. Many papers used very aggregate data/indicators collected by other institutions. One gets the impression authors did not try to sharpen the quality of country data or get better data during their country visits. This may call for lengthier country visits or for stronger collaboration with local researchers. For example, several papers discussed the need to reduce obstacles for the private sector and improve the business environment. Published data from business indicators from the World Bank and UN were used. One would have expected that authors would have sharpened such indicators during their visits, unbundling them more, or would have deepened their understanding of what ultimate factors were behind those obstacles. They should have focused on some specific and critically important constraints and examined why they had not been addressed in the past. For example, papers elaborate very little on the political economy of reform: if reforming such obstacles was so beneficial why had these reforms not taken place in the past? There seems to be insufficient institutional knowledge and one is sometimes left with the impression that authors seldom visited the country.

Weaknesses in addressing common themes

16. Some papers tried to estimate *output gaps* on the basis of past time series and using different methodologies. Presumably the purpose of the exercise was to provide the authorities with a forward-looking tool that would allow modulating fiscal/monetary policies to the possible emergence of such a gap in the outer years. However, these estimates were sometimes explained too cryptically. Estimates should have been more qualified and the robustness of results more carefully assessed. Papers should provide examples on how to use such estimates in a forward-looking context and discuss the implications of possible structural breaks. The policy implications of these results should be better discussed.

17. Several papers estimated *real exchange rate (RER) equations* based on time series data, relating RER levels to economic fundamentals. Presumably, this would allow judging possible RER misalignment situations in the future. Sometimes these papers did refer to future sustainability/risks issues but without explicitly linking these topics to the equation being estimated. Future sustainability of specific RER levels presumably depends on the future sustainability of the fundamentals—fiscal deficits, foreign aid, terms of trade, etc. The exercises usually do not put these estimates in "future motion," discussing sustainability in terms of future developments in the fundamentals. For example, the estimated coefficients could be used to assess how RER will change according to future aid flows—this may be an important area in aid dependent countries. In deciding what past data on RER to use (in estimating these equations), it is important to be aware of periods where the exchange rate (ER) regime has been characterized by multiple exchange rates or segmented ER markets. This calls for careful decisions on what constitute the "binding" ERs in building the appropriate time series.

18. A topic of several SIPs was the analysis of *sources of growth*. The papers emphasized the importance of increases in education and skills. However, these papers showed a weak familiarity with the well-known literature on how to build a labor skills index based on changes in the educational distribution of the labor force in order to measure the contribution of education (Griliches and Jorgenson, 1967).[2] In much of the literature such correction has reduced the "residual" and explained an important share of growth in developed and developing countries. Authors could have easily built such indices from wages by educational levels and data on changes in the educational distribution of the labor force.

19. *Clearer separation between once and for all effects in price levels from sustained inflation trends*. Several papers tried to analyze the impact of shocks (world food or energy prices, increased political risk) on inflation, but they did not spell out the mechanism by which these impacts might translate into sustained periods of inflation. Again, this was the

[2] Zvi Griliches and Dale W. Jorgensen (1967), "The Explanation of Productivity Change," *Review of Economic Studies,* Vol. 34, No.3, pp. 249–83.

result of using reduced form equations without a clear discussion of the underlying causal mechanism. Several papers were not sensitive enough to the difference between trends in inflation and once and for all changes in the price level.

Common attributes of strong papers

20. Many of the good papers relied on simpler analytical tools or simpler models incorporating good economic analysis. All these papers had very specific, well-defined, and convincing welfare questions. The authors seemed to be familiar with the country context and country data, and in many cases they also made special efforts to collect relevant data. Some of these papers focused on fiscal issues such as institutional aspects and reforms of the fiscal framework, fiscal rules, debt sustainability issues, competitiveness in the context of expansions in external flows, specific issues in the banking systems, and specific issues of tax reform.

IV. CONCLUSIONS AND RECOMMENDATIONS

21. During the period covered by the evaluation, the Fund produced about 220 SIPs a year, accounting for about 40 percent of total research output.[3] Among the Fund's research outputs, SIPs probably have the most significant potential to enhance the policy advice of the Fund as part of its Article IV surveillance mandate. These papers also carry important reputational responsibilities for the institution. Furthermore, in the lower income countries, SIPs are particularly important given the scarcity of policy-oriented research. Given the resources and talent available to the institution, weak papers could easily be improved to become as good as the top one-third of papers identified by this evaluation.

The following are specific recommendations to help improve SIPs:

(i) *The Fund should better clarify the exact objective and function of SIPs* and how they differ from other analytical outputs of the institution such as working papers, technical assistance activities, etc. If the objective is to assist country authorities and the economic community in addressing major policy areas, the institution should make this clear and draw its implications (below). SIPs should not be vehicles whose main purpose is to test techniques or specific models. Such objectives should be served by other research outputs.

(ii) *The Fund should pay special attention to how SIP topics are selected*. A lack of consultation on topics to be addressed by future SIPs was a major finding that emerged from the country visits that were undertaken for the research evaluation. Such consultations should improve the relevance of the issues addressed and increase

[3] IEO Background Document II: "IMF Research: Taking Stock" will be available at: www.ieo-imf.org.

the sensitivity of IMF staff to the way SIPs are written and disseminated in order to be more effective. Collaboration with local institutes in low-income countries could be important to learn about the country context and institutions as well as to improve the quality of data.

(iii) *The Fund should focus on areas of examination where it has expertise and a mandate*. Too wide a scope of topics has the risk of lowering the quality of SIPs and opens questions of overlap with other institutions.

(iv) *The Fund should pay more attention to the structure of a SIP and how to make it more "reader friendly."* Some suggestions are given below:

- First, clear executive summaries are needed. Authors should keep in mind that they are addressing policymakers (among others) with limited time.

- Second, econometric or modeling techniques should not occupy the center of attention in SIPs, but should be included in appendices or reserved for the working papers series. SIPs could then extract from such working papers the most intuitive economic analysis and conclusions.

- Third, SIPs should be careful in giving policy advice as many emerging/low-income countries have limited institutional capacity for reforms. SIPs should try to be focused and specific in their recommendations and avoid being vague in their advice (like strengthening tax revenues, improving financial stability, etc.). Stronger collaboration with authorities and institutions will allow a better understanding of specific conditions in a country.

- Fourth, many papers were based on weak data and authors should make explicit the implications for the robustness of results. Again, this may be particularly important for low-income countries. If the topic to be selected is critical but the required data is poor, Fund staff should work with researchers in the country to improve such data.

(v) *A major conclusion of the evaluation is that in order to produce a good SIP more time should be spent at all levels of the process*. This would include more time with authorities in identifying issues, more collaboration in the field to increase country knowledge of context and institutions, and stronger efforts to obtain a minimum quality of data. In the context of a given resource envelope this could be achieved by reducing the number of SIPs per year and focusing more effort on low-income countries. For example, producing only one SIP per country every two years may allow for a much better crafted paper. This could be accompanied by a stronger quality control process within the institution.

IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/11/05

An Evaluation of the Research Chapters of the IMF's *World Economic Outlook* and *Global Financial Stability Report*

Miguel A. Kiguel

May 20, 2011

Address. 700 19th Street, N.W., Washington D.C. 20431, U.S.A. Telephone. +1202 623 7312 Fax. +1 202 623 9990 Email. feedback@ieo-imf.org Website. www.ieo-imf.org

 BP/11/05

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

An Evaluation of the Research Chapters of the IMF's *World Economic Outlook* and *Global Financial Stability Report*

Prepared by Miguel A. Kiguel

May 20, 2011

Abstract

This study examines the technical quality of the analytical chapters of the two IMF flagship reports, the *WEO* and the *GFSR*. It found that most of these chapters provided good quality analysis of relevant macroeconomic and financial topics. It found that *WEO* chapters were stronger when they dealt with areas within the core mandate of the Fund, and that the *GFSR* chapters had greatly improved since 2007. The weaker chapters in both reports tended to lack clear analytical frameworks. The policy advice in both reports tended to be predictable and did not always follow from the analysis. Neither the *WEO* nor the *GFSR* chapters achieved an integrated view of macroeconomic and financial developments and neither foresaw the severity of the 2007–08 crisis.

JEL Classification Numbers: E0, F3, F5, G0

Keywords: IMF research, IMF multilateral surveillance, *World Economic Outlook*, *Global Financial Stability Report*, quality comparison, research evaluation

Author's E-Mail Address: mkiguel@econviews.com

Contents **Page**

ABBREVIATIONS

BRIC	Brazil, Russia, India, and China
GFSR	*Global Financial Stability Report*
IMF	International Monetary Fund
NGOs	nongovernmental organizations
OECD	Organization for Economic Cooperation and Development
WEO	*World Economic Outlook*

I. Introduction and Main Findings[1]

1. This evaluation covers the research or analytical chapters of the IMF's *World Economic Outlook* (*WEO*) and *Global Financial Stability Report* (*GFSR*) issued during the evaluation period, 2004–08. It covers 30 chapters of the *WEO* and 20 chapters of the *GFSR* and is based on a variety of criteria, including the quality of the analysis and the balance between theory and evidence, the relevance of the topics for policymakers, their timeliness, the soundness of the policy recommendations, and whether the presentation is clear, with well-focused questions posed at the outset, with understandable language and with a good use of tables and graphs.

2. The overview chapters of the *WEO* and the *GFSR* provide an evaluation of current developments and of the main risks in the world economy and in the international financial markets. The analytical chapters of the two reports address specific topics that are relevant for current policy discussions in more depth, treating them in a way that is accessible to a broad audience.

3. The two reports are flagship IMF publications that are broadly distributed and intended to have an important influence on policymakers and the media, and in some academic and think tank circles, in order to foster the debate on current policy issues. Most of the media coverage focuses on the initial chapters, while the analytical chapters receive less attention, at least at the time the publications are released.

4. The *WEO* has a broader scope than the *GFSR*, mainly because it addresses a large number of macroeconomic issues that represent the core work at the IMF, while the *GFSR* deals mainly with financial issues, which have only recently received much attention from policymakers. In addition, the *WEO* has a longer trajectory and over the years it has built a clear reputation as one of the leading reports on current economic developments and prospects. The *GFSR* is a more recent publication whose importance is rising along with the recognition that financial and macroeconomic issues are closely interconnected. The analytical chapters of the *WEO* in the evaluation period seem to be more developed products, while those of the *GFSR* are more uneven. Only in more recent years has the *GFSR* developed a structure that seems suitable for the product.

5. This evaluation is organized as follows. The remainder of this section provides a summary overview and evaluation ratings, with scores for different aspects of the two reports. Section II looks in more depth at the *WEO*, analyzing aspects such as structure, choice of topics, and policy conclusions, and Section III presents a similar analysis for the *GFSR*. Section IV presents conclusions and some reflections for the future.

---

[1] I am grateful to the participants at the evaluation workshop that took place in Washington, in April 2010, to Hali Edison for her patience in providing me tips about how to approach the report, to Marcelo Selowsky for his useful comments, and to Federico Bragagnolo for research assistance.

A. The *World Economic Outlook*

6. On the whole, the research chapters of the *WEO* achieve the objective of providing a good analytical framework combined with evidence to discuss current economic policy issues. Nevertheless, there are exceptions. The chapters are generally stronger when they address traditional macroeconomic issues (such as global macro imbalances,[2] exchange rate issues,[3] or inflation[4]) than when they address issues (e.g., labor markets,[5] workers' remittances,[6] institutions,[7] or climate change[8]) that fall outside the Fund's core areas of research. This is not surprising, as the stronger chapters correspond to those areas in which the IMF has a "comparative advantage" and where it spends more research effort.

7. There are exceptions, though, as some chapters, especially those that do not address core IMF issues, have weak analytical frameworks, while others get lost in technicalities—especially some that include large econometric models to simulate macroeconomic outcomes.

8. The conclusions and discussions of policy implications tend to be quite predictable and somewhat vague, especially when they argue in broad terms in favor of a prudent fiscal stance, a credible and sound monetary policy, a more open economy, or more market-oriented policies. It would certainly be useful to present more specific policy recommendations, targeted to the issues being discussed, perhaps including some examples of specific countries or type of countries.

9. The evaluation uses ten criteria that fall into four groupings:

- **Structure and presentation**, which looks at whether the chapter is well organized, whether the questions are clearly posed, and whether there is a good balance and coherence between empirical questions, analytical issues, and policy discussions;

---

[2] See for instance, Chapter 3 and Box 1.4, April 2005; Chapter 2 and Appendix 1.2, September 2005; Chapter 2 and Box 1.4, April 2006; Box 2.1, September 2006; Chapter 3, April 2007; Box 1.3, October 2007; Box 1.3, April 2008.

[3] See for instance, Chapter 2, September 2004; Chapter 3, April 2007.

[4] See for instance, Box 1.1, September 2004; Chapter 4, September 2005; Chapter 3, April 2006; Chapter 3, October 2008.

[5] See for instance, Chapter 5, April 2007.

[6] See for instance, Chapter 2, April 2005.

[7] See for instance, Chapter 3, September 2005.

[8] See for instance, Chapter 4, April 2008.

- **Choice of topics**, assessing especially whether the topics are related to the Fund's work and whether they address policy issues that were relevant at the time the report was issued;

- **Quality of analysis**, looking especially at whether the chapter has a strong (though not necessarily complicated) analytical framework, whether the data presented are relevant for the issues being discussed, and whether the conclusions follow from the analytical framework and the evidence presented; and

- **Policy conclusions**, mainly assessing their relevance and their relation to the analysis presented in the chapter.

10. For each criteria, judgments were made that take into account relevant factors, example given, did the chapter address issues of high policy relevance. We use a rating system that ranges from 1 to 5 (where 5 is the best "grade"). Table 1 summarizes the results of the evaluation of the *WEO* research chapters overall.

Table 1. *World Economic Outlook* Summary Evaluation

Report Chapter	**Overall Average**		
	Core (24)	**Noncore (6)**	**All (30)**
Structure and presentation			
1. The questions are well posed and clearly focused	4.2	3.2	4.0
2. There is a sound and clear discussion of the analytical framework	3.7	2.8	3.5
3. There is a good balance between analytical discussions, empirical evidence, and policy implications	3.7	3.2	3.6
4. Writing is clear and well organized	4.2	3.5	4.1
5. Is the empirical evidence presented in an intuitive and convincing way?	3.5	2.8	3.4
Choice of topics			
6. The issue is timely and relevant	3.8	2.5	3.5
7. The issues are important for policy discussions at the time	3.5	2.0	3.2
Quality and analysis			
8. Does it use appropriate analytical tools and relevant data?	3.8	2.8	3.6
Policy advice			
9. Would it be useful for policymakers?	3.4	2.7	3.3
10. Are the policy conclusions convincing?	3.4	2.7	3.2
Overall rating report	3.7	2.8	3.5

Note: The number of chapters analyzed is given in parentheses. The rating system ranges from 1 to 5 where 5 is the best rating.

11. The evaluation shows that overall the *WEO* is a high-quality product—a finding that is not surprising given the reputation of this report and the influence that it has when it is

released. There is a question, though, regarding whether some of the more technical portions of the analysis, as well the use in some cases of rather complex econometric techniques, are really within the grasp of policymakers, especially the top policymakers who constitute an important part of the intended audience for the report.

12. We divided the sample of 30 *WEO* analytical chapters into two groups: 24 chapters that address traditional macroeconomic issues ("core issues") such as inflation, exchange rate policy, or global imbalances; and 6 chapters on issues that fall outside the Fund's core research (e.g., labor markets, institutions, or climate change) (Table 2).[9] With the world economy performing as well as it did during most of the evaluation period, it might have been difficult for the Fund staff to identify timely research topics, and there might have been an understandable tendency to write about issues outside the core of the Fund's work. Nevertheless, we found an important difference in the quality of the chapters depending on whether they address "core" IMF issues or others, in which the Fund does not have the same level of expertise. This is reflected in the differences in the average score, 2.8 for the noncore chapters versus 3.7 for the core chapters.

13. The problems generated by the U.S. current account deficit and global imbalances were at the center of the policy discussion on macroeconomic issues during the evaluation period and hence it is not surprising that they received significant attention in the *WEOs* during those years.[10] Other issues that received attention were related to inflation and inflation targeting, exchange rate policy, and housing prices. The major themes that dominated the macroeconomic debates during these years were the end of inflation, the increasing importance of the BRICs (Brazil, Russia, India, and China) in general and of China in particular, global imbalances, asset price booms (especially in housing) and the rise in commodity prices.

14. The subprime crisis began in July/August 2007. Though the last three *WEOs* included in the sample were produced too early to consider the impact of the global crisis, the *WEOs* issued since 2007 do consider the potential impacts of a crisis and instruments of political economy with which to address them. For instance, the April 2007 report includes a chapter on the potential impact on the global economy of a recession in the United States. Chapter 3 of the April 2005 report addresses the implications of the housing price decline for monetary policy, and the latest report in the sample (October 2008) includes a chapter addressing the relation between financial stress and economic downturns and another one discussing the possible utilization of fiscal policy as a countercyclical tool. The most recent reports in the sample also contain boxes on issues related to the crisis, such as the macroeconomic

---

[9] There is some arbitrariness in this distinction. For instance, one could argue whether the functioning of labor markets fall within the main mandate of the IMF. Our view is that it broadly belongs to the category of structural issues that are not necessarily a central part of IMF-supported programs.

[10] See for example, Chapter 2, April 2004; Chapter 3, April 2005; Chapter 2, September 2005.

implications of the financial market turmoil and the risks that housing markets posed for global growth.

Table 2. *World Economic Outlook* Chapters Evaluated

Year		**Chapters**
2004	April	(ii) The Global Implications of the U.S. Fiscal Deficit and of China's Growth (iii) Fostering Structural Reforms in Industrial Countries (iv) Are Credit Booms in Emerging Markets a Concern?
2004	September	(ii) Three Current Policy Issues: The Global House Price Boom; Learning to Float: The Experience of Emerging Market Countries Since the Early 1990s; Has Fiscal Behavior Changed Under the European Economic and Monetary Union? (iii) How Will Demographic Change Affect the Global Economy?
2005	April	(ii) Two Current Issues Facing Developing Countries: Workers' Remittances and Economic Development; Output Volatility in Emerging Markets and Developing Countries (iii) Globalization and External Imbalances (iv) Will the Oil Market Continue To Be Tight?
2005	September	(ii) Global Imbalances: A Saving and Investment Perspective (iii) Building Institutions (iv) Does Inflation Targeting Work in Emerging Markets?
2006	April	(ii) Oil Prices and Global Imbalances (iii) How Has Globalization Affected Inflation? (iv) Awash with Cash: Why Are Corporate Savings So High?
2006	September	(iii) Asia Rising: Patterns of Economic Development and Growth (iv) How Do Financial Systems Affect Economic Cycles? (v) The Boom in Nonfuel Commodity Prices: Can It Last?
2007	April	(iii) Exchange Rates and the Adjustment of External Imbalances (iv) Decoupling the Train? Spillovers and Cycles in the Global Economy (v) The Globalization of Labor
2007	October	(iii) Managing Large Capital Inflows (iv) Globalization and Inequality (v) The Changing Dynamics of the Global Business Cycle
2008	April	(iii) The Changing Housing Cycle and the Implications for Monetary Policy (iv) Climate Change and the Global Economy (v) Globalization, Commodity Prices, and Developing Countries
2008	October	(iii) Is Inflation Back? Commodity Prices and Inflation (iv) Financial Stress and Economic Downturns (v) Fiscal Policy as a Countercyclical Tool (vi) Divergence of Current Account Balances Across Emerging Economies

Note: Highlighted chapters are considered "noncore" issues.

15. An issue that has been missing but would have been useful to address is the link between credit markets and the business cycle. While there are some discussions of the booms in asset prices, including the housing bubble, the chapters fail to describe a link from the fall in asset prices to the impact on credit and on the business cycle. While this weakness is easy to detect today, given the crisis that eventually took place, it would have been a natural extension of many of the issues that were raised in different chapters.

B. The *Global Financial Stability Report*

16. The importance of the *GFSR* as a flagship IMF publication has been growing, though the *GFSR* is less well known and has a shorter track record than the *WEO*. The first chapter of the *GFSR* quickly began to be recognized among policymakers and practitioners for its useful and comprehensive reviews of the main developments in world financial markets and overall assessments of financial risks. This overview chapter typically presents some technical issues, such as topics related to credit default swaps or related to the markets of different derivatives. It is written relatively simply, and is accessible to a broad audience.

17. The *GFSR* typically devotes two chapters to research topics. Compared with the opening chapter, these analytical chapters have not achieved the same level of maturity, though since 2007 they have been improving. By and large, those produced until 2007 do not show a good balance between strong analytical work and well-structured evidence. They provide a good description of the stylized facts and raise relevant policy issues, but they certainly do not have the same analytical depth as their corresponding overview chapters.

18. Relative to the *WEO*, the research chapters of the *GFSR* on the whole are more descriptive and much less technical. They primarily provide stylized facts and trends about the evolution of the world financial markets, including the role of different players or products. They typically lack a well-defined analytical framework and a clear focus, and in general they do not pose a set of questions that they attempt to answer using theory and/or evidence. Their conclusions are unnecessarily lengthy and do not give the reader a clear sense of priorities; their long lists of "to do's" tend to overwhelm the reader.

19. However, starting in 2007, and coinciding with the first signs of softness in the world financial markets, improvements can be seen in the content, the analytical frameworks, and the presentation of the *GFSR* research chapters. Questions are posed more clearly at the outset, the analysis presented is better integrated with that presented in the first chapter of the reports, and a better balance is achieved between the topics discussed in the first chapter and the evidence and theoretical analysis in the research chapters.

20. Our review is based on 20 analytical chapters of the *GFSR* that were written between 2004 and 2008 (Table 3).

Table 3. *Global Financial Stability Report* Chapters Evaluated

Year		Chapters
2004	April	(iii) Risk Transfer and the Insurance Industry (iv) Institutional Investors in Emerging Markets
2004	September	(iii) Risk Management and the Pension Fund Industry (iv) Emerging Markets as Net Capital Exporters
2005	April	(iii) Household Balance Sheets (iv) Corporate Finance in Emerging Markets
2005	September	(iii) Aspects of Global Asset Allocation (iv) Development of Corporate Bond Markets in Emerging Market Countries
2006	April	(ii) The Influence of Credit Derivative and Structured Credit Markets on Financial Stability (iii) Structural Changes in Emerging Sovereign Debt and Implications for Financial Stability
2006	September	(ii) Household Credit Growth in Emerging Market Countries
2007	April	(ii) Changes in the International Investor Base and Implications for Financial Stability (iii) The Globalization of Financial Institutions and Its Implication for Financial Stability
2007	September	(ii) Do Market Risk Management Techniques Amplify Systemic Risks? (iii) The Quality of Domestic Financial Markets and Capital Inflows
2008	April	(ii) Structured Finance: Issues of Valuation and Disclosure (iii) Market and Funding Illiquidity: When Private Risk Becomes Public
2008	October	(ii) Stress in Bank Funding Markets and Implications for Monetary Policy (iii) Fair Value Accounting and Procyclicality (iv) Spillovers to Emerging Equity Markets

21. The evaluation uses the same methodology and set of criteria as for the *WEO* and the same rating scale, ranging from 1 to 5 (where 5 is the best "grade"). Table 4 summarizes the evaluation of the *GFSR* analytical chapters.

22. For the *GFSR* we divided the sample into two subperiods, the first ranging from 2004 to 2006 and the second from 2007 to 2008. Of the 20 chapters in the sample, 11 correspond to the earlier period. We calculated the overall average scores of the chapters for each subperiod. Dividing the sample in this way reveals important changes in both the substance and the presentation of the chapters.

23. The overall results indicate that in general the *GFSR* chapters perform as well as the *WEO*. However, the difference is startling when we consider the subperiod 2004–06, when the *GFSR* research chapters were particularly weak, their scores are significantly below the *WEO*. But the quality of *GFSR* chapters improved sharply in the later period and became in some cases better than that of the *WEO* chapters.

Table 4. *Global Financial Stability Report* Summary Evaluation

Report Chapter	Overall Average		
	2004–06 (11)	2007–08 (9)	**2004–08 (20)**
Structure and presentation			
1. The questions are well posed and clearly focused	2.9	3.9	**3.4**
2. There is a sound and clear discussion of the analytical framework	2.6	4.1	**3.3**
3. There is a good balance between analytical discussions, empirical evidence, and policy implications	3.1	4.0	**3.5**
4. Writing is clear and well organized	2.9	3.9	**3.4**
5. Is the empirical evidence presented in an intuitive and convincing way?	3.0	4.0	**3.5**
Choice of topics			
6. The issue is timely and relevant	2.9	4.3	**3.6**
7. The issues are important for policy discussions at the time	2.5	4.0	**3.2**
Quality and analysis			
8. Does it use appropriate analytical tools and relevant data?	3.1	4.2	**3.6**
Policy advice			
9. Would it be useful for policymakers?	2.7	4.0	**3.3**
10. Are the policy conclusions convincing?	2.9	3.8	**3.3**
Overall rating report	**2.9**	**4.0**	**3.4**

Note: The number of chapters being analyzed is given in parentheses.

II. The *World Economic Outlook*

A. The World Economy and Financial Markets During the Evaluation Period

24. During most of the evaluation period the world economy enjoyed high and synchronized growth, low rates of inflation, and financial stability, in a period that has sometimes been called "the Great Moderation." In fact, 2003–07 was probably the best economic period of the last four decades. Low rates of inflation generated unique financial conditions. Credit spreads were low and there were no apparent problems in financial institutions.

25. These favorable economic conditions, which are discussed at some length in the first chapters of the various issues of the *WEO*, probably affected the Fund's selection of research topics for this report. As there were no "hot issues" that needed to be discussed, chapters focused on some idiosyncratic choices such as workers' remittances, institution building, or demographic changes, which do not appear to be core topics for an IMF flagship publication.

26. The main themes that dominated the macroeconomic debates during these years were the end of inflation, the increasing importance of the BRICs in general and of China in particular, global imbalances, asset price booms (especially in housing), and the rise in commodity prices. The subprime crisis started to surface in July/August 2007, initially without large effects on the global economy. The financial crisis became apparent only in the first quarter of 2008, with the collapse of Bear Sterns, coinciding with the publication of the April 2008 *WEO*, and hence there was not much time to include any discussion of the issues till the October 2008 reports.

27. The topics that dominated the policy discussions during those years were covered in the research chapters of the *WEO*. As discussed below, the treatment was sometimes uneven; macroeconomic and financial issues were not well integrated; and though the policy recommendations were generally in the right direction they mostly failed to give the reader a clear message about priorities or about whether the issues were really urgent.

B. Evaluation Findings

Structure and presentation

28. The *WEO* chapters are generally well structured. They show discipline in clearly and concisely stating their main questions at the outset, and this definitely helps the reader to understand their purpose. Typically, a well-defined part of the chapter presents the main stylized facts and the analytical framework and is followed by a section that tends to be more technical, many times including references to econometric models or to quantitative evidence. The chapter concludes by summarizing the main findings and presenting the policy implications. There has been some improvement in the organization and clarity of the chapters over time.

29. Some individual *WEO* chapters cover two or three different topics, which many times are unrelated to each other and as a result are difficult to tie together to find a common theme.[11] A case in point is Chapter 2 of the April 2004 *WEO*. This analyzes three topics—the housing boom, exchange rate flexibility in emerging markets, and fiscal policy in the euro area—that have no obvious common theme; by and large they relate to problems observed in different regions and their economic and policy links are far apart. From a reader's perspective it would have been better to treat each in a stand-alone chapter, even if a short one.

30. The *WEOs* are prepared for the Spring and Fall meetings of the Board of Governors of the IMF and, as stated in some documents that we had access to, their main audience is policymakers attending those meetings. The nature of the material and the technical

[11] See for instance, Chapter 2, April 2004; Chapter 2, September 2004; and Chapter 2, April 2005.

presentations of some issues in the various chapters suggest that they cater to a broader audience than just the high public officials—namely mid-career staff at ministries and central banks, and probably equally importantly to researchers and to think tanks, nongovernmental organizations (NGOs), universities (including students), and people generally interested in macroeconomic issues. The material presented in the chapters is generally useful for policy discussions and for use in courses in colleges and universities (mainly for undergraduate economics and in some schools that deal with public policy issues).

31. The research chapters balance a relatively nontechnical introduction and conclusions with a more technical main body that uses somewhat more technical language. Given the multiple objectives of these chapters and the variety of audiences that tend to make use of them, the current approach seems reasonable.

32. The chapters generally present an analytical framework that helps the reader to understand the issue under discussion. Presentation of the framework is a useful and necessary component and it works best when it is done in a relatively simple and readily understood way, linking the analysis to the existing literature. The chapters generally succeed in this respect, though in a few cases the analytical frameworks are not made clear. An example of a good analytical framework is "Exchange Rates and the Adjustment of External Imbalances" (Chapter 3, April 2007), which poses a clear set of questions and describes the approach to addressing them, along with links to relevant work presented in previous *WEOs*.

33. There are several examples of chapters with a weaker framework. For example, in "Fostering Structural Reforms in Industrial Countries" (Chapter 3, April 2004) the question is presented at the outset but analytical and methodological problems weaken the results and conclusions. The indices of structural reforms used in the chapter are based on very few variables, and could be misleading; for instance, in the case of trade reform the chapter looks only at the average tariff (and does not consider nontariff barriers), while in the case of financial reform it looks only at variables like controls on credit or on interest rates, which have become less pertinent over the years as indicators of the quality of the financial markets. The interesting and somewhat worrisome result is that these two areas are precisely those in which the authors find that there has been more progress. Something similar occurs in "Building Institutions" (Chapter 3, September 2005), which offers abundant interesting evidence about changes in institutions and the historical background, but no clear guide to policymakers regarding what specific policies can be adopted to improve institutions. The long "shopping list" that is offered does not help the policymaker know where to start.

34. Another example is the section on workers' remittances in Chapter 2, April 2005, in which the analysis does not pose clear questions at the outset and provides policy conclusions that are too vague and seem to overestimate the virtues of remittances. The lack of a clear analytical framework weakens the analysis and the conclusions. The discussion presents a large number of alleged benefits of remittances (e.g., that they help to increase human and physical capital or that they reduce the risks of financial crisis), which in our view tend to

overstate the importance of remittances, and the discussion lacks a clear focus. In addition, there is almost no discussion of their volatility caused by problems in the host country—an issue that became apparent following the 2008 recession. The international financial positive effects of remittances appear to be overstated. For example, "in extreme cases, remittances might reduce the probability of financial crisis" (p. 73) seems farfetched, as financial crises are typically related to problems with the allocation of stocks of assets, while remittances represent a flow. Besides, it is not clear whether the authors have in mind problems related to the stability of the banking system or to potential problems that could arise in the balance of payments (i.e., a run on the currency).

35. The more technical sections of the chapters sometimes take for granted that the reader has some knowledge of econometrics. One example is Chapter 4, April 2007, where p. 137 mentions the "cross-country and cross-region set of vector autoregressive analysis (VAR)." We believe that for many readers of the *WEO* there should be at least a footnote explaining what a VAR is, and what is the purpose of using this VAR analysis relative to other techniques.

36. Two useful aspects of the current format should be maintained. First, the boxes, which are used quite extensively, are a helpful instrument to illustrate specific points or to address some technical issues that are targeted to specific audiences. The advice here would be to try to prevent them becoming a distraction to the main argument, and to try to keep them short. Second, the chapters are presented in a helpful format, accompanying the column of text with a column of supporting graphs and tables.

37. Regarding the presentation of evidence, most of the time it is better to show graphs presenting the partial correlation between two variables than to provide tables with multiple regressions that could be difficult to interpret.

38. In our view, the Fund staff should make a big effort to present in a simpler, nontechnical way the important findings from its research. Event analysis and case studies are very useful means to illustrate the points made. Chapter 4 of the April 2007 *WEO* (on decoupling) offers a good balance between "event studies" and econometrics as well as the presentation of data. Econometrics is a useful tool, though it is probably a "black box" for many readers, who will simply need to have faith in the results presented. A number of *WEO* chapters rely on econometrics and get lost in technicalities about the econometric or simulation techniques and do not pay enough attention to the interpretation or implications of the results. One of these is Chapter 3 of the April 2005 *WEO*, which uses the Global Economic Model (GEM) "to investigate the implications of globalization for the adjustment of external imbalances" (p. 133). This chapter simulates several scenarios (instead of one base case and alternatives to the base cases) and does not present them clearly enough to convey a good sense of their likelihood of occurring, or of the major risks that the world economy faces. The discussion in the chapter on climate change and the global economy

(Chapter 4, April 2008), using the Global Integrated Monetary and Fiscal Model, also becomes unnecessarily technical.

39. In our view the *WEO* should aim to reach a broad audience, most of whom are expected to be familiar with economic terms and concepts but not necessarily knowledgeable in econometrics or on more technical quantitative issues. As a result it would be useful to always provide an intuitive interpretation of the econometric results and their implications. Econometrics and other quantitative methods should be only used when they can help to quantify the effects of certain policies on outcomes (say, the effect of a 10 percent real depreciation on the trade balance on the adjustment of the current account), and the results should be presented in a simple and readily understandable way.

40. The audience expects the IMF to quantify some of the effects discussed in the analytical chapters. Nevertheless, these chapters should make clear that their quantitative estimates are mainly scenarios, rather than forecasts, and should explain the underlying assumptions and the limitations of the empirical results.

Choice of topics

41. The analytical chapters of the *WEOs* typically address macroeconomic issues and policies that are relevant at the international level and that are part of the policy debate in the months prior to the publication of the reports. It is difficult for them to be fully up to date with macroeconomic policy issues, because often the choice of research topics is based on research commissioned months earlier or in other cases ongoing research conducted at the Fund.

42. The issues covered by the *WEO* in the evaluation period are by and large those that dominated the policy debates and the concerns of policymakers during those years. They include global imbalances, including issues related to current account balances and fiscal deficits; inflation; exchange rate and monetary policies; capital flows; and commodity prices and their impact on the world economy. In addition, there is some discussion of the links between the financial sector and the real economy as well as of some specific issues that could affect macroeconomic performance, such as housing prices or credit booms.

43. A number of chapters address the importance of globalization. In itself, globalization is important for the IMF, but some of the chapters discuss aspects of globalization that are not central to the Fund's work; an example is the discussion on globalization and inequality (Chapter 4, October 2007). The analysis of how globalization has affected inflation (Chapter 3, April 2006) is clearly a mainstream issue for the IMF, while the one on globalization and labor markets (Chapter 5, April 2007) falls somewhere in between.

44. Other topics fall less clearly within the expected range; they include "Building Institutions" (Chapter 3, September 2005), "Fostering Structural Reforms in Industrial

Countries" (Chapter 3, April 2004), and "Climate Change and the Global Economy" (Chapter 4, April 2008), which are not part of the core work of the Fund.

45. A number of issues covered in the *WEO* have an overlap with the *GFSR*. Examples are the analysis of housing booms, though in general the *WEO* chapters on this topic address the macroeconomic effects and monetary policy, which means that they fit naturally within the *WEO* mandate. Another example, with a less convincing fit, is the chapter on "Awash with Cash: Why Are Corporate Savings So High?" (Chapter 4, April 2006). Here the analysis is mainly on corporate financial decisions, and although it has some implications for overall saving and investment, it would have been more appropriately included in a report on corporate finance.

46. One of the central themes of the *WEOs* is the issue of global imbalances. The treatment of these imbalances focuses mainly on the U.S. economy, dealing both with the current account deficit and possible mechanisms for adjustment and with the budget deficits and their possible reduction. Most of the discussion could not have anticipated that in the end the effects of the 2008 crisis would eventually lead to a reduction in these imbalances.

47. Other recurrent themes are the fall of inflation rates and the experience with exchange rate flexibility in emerging countries. In these two topics the Fund clearly has significant expertise, and by and large they are well analyzed in the sample chapters. The chapter that asks whether inflation targeting works in emerging markets (Chapter 4, September 2005) is well structured and contains good analysis. Another chapter (Chapter 3, April 2006) analyzes whether globalization of the real and financial sectors has affected inflation. This topic has received a good deal of attention, especially as many developed economies have started to rely on outsourcing. The analysis here is well done, and the policy conclusions argue in favor of the importance of monetary policy. Perhaps the authors should have been more emphatic about the fact that globalization affects price levels but not necessarily the rate of inflation. Chapter 3 of the October 2008 *WEO* discusses the effects of the rise in commodity prices on inflation. While this was a "hot topic" at the time, the report's publication coincided with the crisis and the topic became quickly outdated.

48. Issues related to exchange rate policies are not discussed as extensively for developed economies as for emerging countries in the sample chapters, though the April 2007 report contains a chapter on exchange rates and adjustment of external imbalances. That analysis is quite useful as it discusses several episodes in which it relates devaluations to improvements in the current account.

49. Another chapter that presents timely issues and contributed to the policy discussions of the time is that on decoupling (Chapter 4, April 2007)—an issue that will probably be further analyzed in the future. The chapter provides a good framework for future discussions.

What was missing?

50. Though the *WEOs* offer a fair amount of discussion about the risks posed by increases in housing prices and about the possibility of some speculation in the market, as well as in the commodity markets, they did not foresee the size of the problem and the potential interconnections it had with credit markets and the stability of the financial sector.

51. With hindsight, it seems clear that more research was needed on the links between developments in the financial markets and credit on the one hand, and economic activity on the other. The most recent chapters that try to put together financial and macroeconomic issues (such as "The Changing Dynamics of the Global Business Cycle" (Chapter 5, October 2007)) acknowledge that the increases in housing and commodity prices pose some macroeconomic risks but do not explore deeply enough the connections between the financial markets and macroeconomic variables.

52. The chapters on credit booms (Chapter 4, April 2004 and Chapter 2, September 2004) emphasize the risks that a reversal of the booms could pose for consumption and aggregate demand, and the possible impact on growth and the business cycle. However, they fail to address the possible connections between potential problems in consumer credit, the banking sector, and overall financial stability. Neither do they explore the financial channel in depth. Chapter 4, April 2004 focuses on credit booms in emerging markets, and a similar analysis is not offered for the developed countries, nor did the *WEO* apparently consider that the growth in nonbank credit was a concern.

53. Similarly, the chapter that discusses the possible existence of a bubble in the housing market (Chapter 2, September 2004) presents compelling evidence of the existence of overvalued housing prices in many countries, and of the risks of a large fall in real estate prices, but its discussion of the implications of the price correction concentrates on the macroeconomic effects, such as a possible reduction in consumption and aggregate demand, or GDP growth. There is no discussion of the implications for the quality of bank assets or the possible risks to the portfolios of securitized mortgages.[12]

[12] Discussing the effects of financial stress on the economy, the same chapter argues, using historical evidence, that episodes in which there is bank distress are more likely to be associated with severe and protracted recessions than are episodes in which the problems center in the securities or foreign exchange markets. This statement appears to be only half right, however, as the recent crisis initially centered on the investment banks and moved only later to the commercial banks—in effect, it was a credit crisis that affected liquidity levels in the commercial paper and repo markets, which resemble much more the credit or securities markets than the traditional banks.

Timeliness

54. Most of the topics covered in the *WEOs* are those that were being discussed at the time in think tanks, in academia, and among policymakers: global imbalances, inflation, exchange rate management, the sustainability of growth, and housing prices among others.

55. A number of the topics featured do not appear to be central in the policy discussions; they include workers' remittances, corporate savings, and the analysis of demographic change. As argued above, the inclusion of such topics probably reflects the unusually low inflation and high growth that prevailed in most of the evaluation period.

56. Since research projects take time and are typically started well in advance of publication, often it is difficult to adjust quickly the content of a report such as the *WEO* to accommodate the most recent developments. This was probably the case in the April 2007 issue, and to some extent in both 2008 issues (especially the chapters on climate change, the changing dynamics of the business cycle, and commodity prices and inflation). These chapters were published in the midst of the financial crisis when the dominant concerns were about the severity of the recession and the policy responses needed to avoid a new Great Depression.

57. The lesson from this period is that the research section of the *WEOs* should probably address timely and relevant macroeconomic policy issues, but cannot by its very nature be expected to address the most recent macroeconomic developments with the same degree of timeliness as Chapter 1.

Final reflections on the choice of topics

58. On the whole, the *WEO* research chapters focus on relevant topics and themes, reflecting the concerns of the international policy discussions of the time. There is a good balance between the coverage of developed and emerging markets, with perhaps a little more attention to the problems of developed countries. Among the latter group, it seems that the United States receives more attention than Europe or Japan while other Asian member countries of the Organization for Economic Cooperation and Development (OECD) in general are left behind. Asia receives more attention than Latin America or Eastern Europe, and Africa receives very little.

59. A disproportionate emphasis on issues related to the United States is clearly visible in the discussion of fiscal deficits, global imbalances, and decoupling. It is not obvious whether this is because the United States has a bigger role regarding these issues or is simply because the IMF is located in Washington.

60. A few issues appear to have been chosen "out of the blue." The discussions of climate change or income do not fit within the core mandate of the IMF, which is to discuss

macroeconomic issues. Some of them could be better addressed by other international development institutions.

61. The impression given is that the chosen themes are driven by the research that is being conducted by the IMF Research Department and are not necessarily oriented to the issues that are discussed in Chapter 1 of the *WEO*. As a result, some of the chapters are more in line with the themes presented in Chapter 1, while others appear to come out of the blue.

62. Finally, the use of chapters that group together two or three unrelated issues does not appear to be a good option. This practice seems to have been discontinued since 2006.

Quality of analysis

63. The quality of the analytical work is solid in the *WEO* research chapters reviewed and in most cases it fulfills the objectives of the *WEO*. The score that we estimate for the sample as a whole is 3.5 out of five, which is a good figure. The score is stronger for the chapters that concentrate on the core issues of the Fund's work (3.7) and weaker for those that deal with noncore issues (2.8).

64. In general the research chapters of the *WEO* have good analytical frameworks, as they convey to the reader the issue that they plan to discuss, they link it to the theoretical and empirical literature, and they motivate the discussion giving examples that are related to current economic events. This presentation is then used to stimulate the discussion of policy issues or important macroeconomic topics, and to motivate the discussions with evidence and results of statistical and econometric analysis.

65. Examples of good chapters include the discussion of inflation targeting in emerging market economies (Chapter 4, September 2005), which is backed by a good analytical framework and useful evidence and is relevant for policymakers; the discussion of decoupling (Chapter 4, April 2007), whose main contribution is the presentation of evidence and the spillovers from developed to emerging markets; and the chapter on savings and investment and the global imbalance (Chapter 2, September 2005), which enhances understanding of the issue at hand.

66. There are also some weak chapters, having either a flimsy analytical framework, or unsatisfactory use of the data or the evidence. One example is the chapter on structural reforms in industrial countries (Chapter 3, April 2004), whose conclusions rely on a very fragile and questionable data set about the evaluation of actual policies. For instance, in assessing the need of financial reform the chapter looks at whether there are controls on interest rates and credit allocations on international financial transactions, instead of assessing the quality of banks and of other financial intermediaries and the quality of the financial regulation and supervision. This view of financial reform seems outdated. Similarly, in the case of trade reform the chapter only looks at the effective tariff rate and not at nontariff barriers.

67. Another disappointing chapter is "How Do Financial Systems Affect Economic Cycles?" (Chapter 4, September 2006) which, instead of discussing the pro-cyclical behavior of credit or the transmission mechanisms of liquidity and monetary policy through the financial sector, addresses a relatively specific microeconomic issue: what difference does it make whether financing is channeled through the banks or through the capital markets? Finally, the chapter on remittances (Chapter 2, April 2005) suggests that remittances are much more important than they seem to be in practice, and probably exaggerates the positive and negative effects.

68. Chapters that address the core IMF topics tend, on average, to be analytically stronger but dealing with a core IMF issue does not always ensure good quality. Global or external imbalances, for example, are treated better in some chapters than in others. "The Global Implications of the U.S. Fiscal Deficit and of China's Growth" (Chapter 2, April 2004) is well organized, its questions are clearly posed and addressed, its discussion of the issues makes use of a solid analytical framework, and its answers and their policy implications are supported by relevant and easy-to-understand evidence. Besides, many of the outcomes predicted in this chapter have since materialized. By contrast, "Globalization and External Imbalances" (Chapter 3, April 2005) promises to provide an interesting discussion of globalization: but most of its analysis actually focuses on whether more integrated world capital markets are a plus or minus to correct external imbalances. While this might be an interesting topic in itself, the chapter does not convince one that it is important for understanding external imbalances. Besides, the chapter offers no discussion of important factors such as sudden stops of capital flows (which have been widely discussed in the case of emerging market countries), or the larger role that sovereign wealth funds and central banks play in the allocation of capital flows—two issues that would have helped to gauge the overall risks.

Policy advice

69. Each of the sample chapters ends by presenting the policy implications of the analysis. This is a key part of the discussion, although often very brief. The Fund's view on policies is taken seriously and is an important basis for policy discussions among policymakers, analysts, and academics.

70. Ideally, the advice presented should help policymakers by clarifying the options for dealing with the issues under discussion. The recommendations should be as specific as possible in order to help governments design appropriately targeted measures. When the policy recommendations are vague or presented in general terms, such as "governments should follow prudent macroeconomic policies" or "market liberalization is useful to foster competitiveness and growth" they are not very useful. The policy advice needs to be specific and tailored to policymakers' needs: it should help them identify priorities and detect the most critical issues that need to be addressed, and should propose specific measures that could be implemented to provide an adequate response to the issue at hand.

71. The *WEO* in general fails to achieve this important objective because although its advice is generally in the right direction it is not specific enough to make an important difference in the design of policies. True, many of the issues that are discussed involve problems that are general in nature (for instance, inflation targeting in emerging markets or decoupling in the global economy) and hence one could not expect the *WEO* to provide very specific recommendations.

72. One example of the lack of specificity of advice follows the discussion of Dutch disease in "Managing Large Capital Inflows" (Chapter 3, October 2007). The chapter's policy conclusions are: "In addition to the macroeconomic policy instruments discussed in Chapter 2 [which refer to the primary fiscal balance or primary expenditures], the authorities have other tools at their disposal, which have not been analyzed systematically—notably, financial supervision and regulation, but also a wider range of policies such as labor and product market reforms. The role of such policies in responding to capital inflows would be an important topic for future research." These recommendations in our view are very vague, as labor reform is a very general issue and product market reform is even vaguer. In addition, it is not clear how financial supervision could affect the Dutch disease problem. Besides, in this particular case, the authors dismiss too quickly the possibility of using sterilization or some form of capital controls. Although these policies are unlikely to be effective in the medium or long term, they could be used temporarily as a way to smooth the transition. Besides, some recent work at the IMF suggests this option.

73. Another example of policy advice that is reasonable but vague comes from the April 2004 report (p. 96), which argues that emerging market economies need to increase the flexibility of their economies through structural reforms. The chapter does not specify which are the important policy reforms or which are the priorities, or which concrete policy measures the Fund would recommend to these economies.

74. Most of the policy advice offered by the sample chapters, at least until 2007, also appears to be very predictable, typically arguing in favor of fiscal adjustment, proposing policies that reduce government regulation and intervention, and more market-oriented measures, freer capital flows, and stable and predictable monetary policies that favor low rates of inflation. While it is difficult to argue with the tone of the policy recommendations, this type of general policy advice is sometimes counterproductive, especially in emerging markets, as it helps to create a view that the Fund is mechanical and that its policy recommendations do not take into account the specific features of individual countries.

III. THE *GLOBAL FINANCIAL STABILITY REPORT*

75. The objective of the *GFSR*, as stated in the preface to the April 2005 edition, is to "assess global financial market developments with the view to identifying potential systemic weaknesses. By calling attention to potential fault lines in the global financial system, the

report seeks to play a role in preventing crises, thereby contributing to global financial stability and to sustained economic growth of the IMF's member countries."

76. Most of the analysis along these lines is carried out in the first chapter, which is the core of the report and makes a clear effort to map the financial risks across financial markets and regions.

77. Each report also typically contains two or three research chapters, which treat specific issues in more depth. Some of these chapters are mainly descriptive, such as those that present the characteristics and roles of key long-term financial players such as pension funds, insurance companies, and investment, mutual, or hedge funds. Others deal with more specific issues that typically address the functioning of financial markets or assess the risks, such as the chapters that address household credit growth in emerging market economies (Chapter 2, September 2006) or the one that analyzes whether market risk management techniques amplify systemic risk (Chapter 2, September 2007).

A. Evolution of Financial Markets During the Evaluation Period

78. The evaluation period can be clearly divided into two subperiods, the first starting in 2004 and ending in the first half of 2007, and the second starting in July 2007, at the beginning of the financial crisis, and ending in 2008. The analysis undertaken in the *GFSR* chapters to a large extent reflects the different realities in these two periods.

79. In the first period, the markets enjoyed unusual financial stability, as credit spreads tightened to very low levels; there was ample funding liquidity in the markets and growing levels of financial intermediation. There were some concerns about the rise in asset and commodity prices, and continued warnings about a possible bubble in housing prices and in credit spreads. Many analysts also raised concerns about the potential financial risks generated by the growth of derivatives markets. But during those years the markets overlooked these risks and policymakers did not perceive them as a major threat. At least on the surface, the markets were working well.

80. The second period corresponds to the preface and the outbreak of the financial crisis, which peaked with the collapse of Lehman Brothers in September 2008. The first problems were observed in the U.S. subprime mortgage market in July 2007, when some funds that had invested in subprime mortgages started to face liquidity and solvency problems and the investment bank Bear Sterns (the underwriter of those funds) had to come to their rescue. These problems were followed by a number of events that clearly signaled risks in the financial markets, including the failure of Country Wide, the largest mortgage broker, in the second half of 2007, and the collapse of Bear Sterns, which was subsequently rescued by the U.S. Federal Reserve and bought by J.P. Morgan in March 2008.

81. The crisis began to escalate with the failure of Freddie Mac and Fannie Mae, though it peaked in September and October of 2008, when the financial markets were in panic and

there was a risk of a new great economic depression. The crisis was the most serious threat to international financial stability since the Great Depression. The survival of many financial institutions, including the largest investment banks, some commercial banks, and the largest insurance company (AIG) was at stake. Investment banks faced a run on their liabilities and were forced to sell many assets at fire-sale prices. In the end, the investment banks were either absorbed by other institutions or transformed into commercial banks, while the commercial banks had to be recapitalized and governments were forced to come to the rescue of many financial institutions. The crisis led to a sharp drop in asset prices, and bank credit and the commercial paper market virtually disappeared.

B. Evaluation Findings

Structure and presentation

82. The audience for the *GFSR*, as for the *WEO*, consists of policymakers, analysts at think tanks, and academics. Compared with those of the *WEO*, the analytical chapters of the *GFSR* are generally longer, more descriptive, less analytical and "model driven," and make less use of statistics. Though they do not set out to test specific hypotheses, they provide a thorough analysis of the sector or issue that they discuss, often describing the key features of a particular industry or sector and covering many of the institutional and regulatory elements that are relevant for the discussion, including cross-country comparisons. They conclude with policy recommendations.

83. The structure of the *GFSR's* research chapters has improved over time, and definite positive changes were made in 2007. The later chapters in the sample make an effort to pose at the outset the main issues and questions to be addressed, thus allowing the analysis and the messages of the reports to be sharpened.

84. The research chapters are generally written simply; they do not make much use of quantitative analysis, and hence they could be read by a broad audience that does not necessarily have a background in these issues. The chapters contain useful tables, many of which are mainly descriptive of the evolution of the financial sector or of the relative size of different financial intermediaries. At the same time they tend to be tedious to read, giving many details that are of interest mainly to people who work in the industry.

85. On the other hand, the typical overview chapter of *GFSR* makes an important contribution that is not easy to find in other places. It describes the evolution of the world financial markets, covering recent financial developments regarding capital flows, interest rates, exchange rates, and asset prices. It is useful because it describes a number of financial issues in a thorough way that are not covered in other publications. Last, but not least, it also provides a discussion of the main financial risks. At least until 2007, there was a big difference between the *GFSR's* Chapter 1 and the research chapters. In general, Chapter 1 is more interesting to read and does a much better job at combining data and analytical issues than the research chapters.

Choice of topics

86. The themes that are covered typically include a description of the main actors in the financial markets (insurance companies, mutual funds, pension funds, etc.), and of some general developments in the financial markets, such as the development of government debt or of credit derivatives. These themes are useful, but their treatment during the evaluation period did not help significantly to identify the global financial risks.

87. The sequence of themes from one report to the next during the first years of our review appears to have followed a clear pattern—which was to cover sequentially the different financial actors—and from that point of view it made sense. This decision had a drawback, namely that a number of issues, especially those related to the systemic effects of financial risks as well as other timely issues such as the housing booms and the emergence of bubbles in asset prices were left out.

88. A large part of the *GFSRs*, especially between 2004 and 2006, is devoted to discussing in great detail the implications of the process of risk transfer that resulted from the securitization of assets such as mortgages, credit card receivables, consumer loans, and other financial assets. With the benefit of hindsight it is clear that the reports overestimated the advantages of these new instruments and did not take into account the funding liquidity and credit risks associated with them.

89. Coinciding with the beginning of the financial crisis, the reports shifted towards the analysis of risks from a systemic point of view that runs across sectors. The October 2008 report contains discussions about the valuation and disclosure of structured financial products, funding liquidity, private and public sector risks, and stress scenarios in the banking system.

What is missing?

90. In our view, a number of themes are missing from the analysis. It is surprising that the research chapters of the *GFSR* did not pay enough attention to key financial issues that could potentially have large effects on the real side of the economies and on the evolution of the financial sector itself.

91. For emerging countries, it would have been useful to try to tie together the financial and macroeconomic issues. One example is the discussion about the diversification of the investor base for emerging markets (April 2004), in which the larger reliance on foreign funding is perceived as unambiguously favorable. This view ignores the risks associated with sudden stops, or those related to the vulnerabilities that external finance can introduce into these countries. This gap is especially surprising because the *WEOs* look at these risks and vulnerabilities and consider them as potential threats to financial stability.

92. For more developed markets, the *GFSR* chapters did not offer an in-depth analysis of the derivatives markets or of the risks that those could entail, or of the growth of securitization, or of the interconnections among different types of credit institutions. It is surprising that during this period no chapter analyzed the risks associated with the fall in risk premiums or the rise in asset prices (mainly for housing, or for the stock markets), especially given the possible existence of bubbles in some of these markets and the effects that a burst bubble could have on overall financial stability. Paradoxically, many of these issues were discussed in the *WEOs*, as discussed in Section II above, though the *WEOs* looked at them from a macroeconomic perspective and did not link them to the risks for financial stability.

93. The *GFSR* research chapters evaluated made a clear effort to maintain a good balance in the coverage of emerging and developed economies, and also maintained a reasonable balance in the discussion of the problems of different regions.

94. Another issue that would have been useful and timely to cover is the critical role that the credit rating agencies played in the crisis. These agencies have attracted many criticisms in the past. True, this is a difficult issue to tackle, but it is in areas such as this where the *GFSR* could make a difference. Two other missing issues that the recent crisis has brought to the fore are (1) the role of the mono-line insurance companies—companies that had an important effect on credit ratings but whose risk might not be as diversified as is generally thought (raising questions about the usefulness of this industry), and (2) the interconnections among financial intermediaries.

Timeliness

95. The evolution of the financial markets during the two subperiods and themes covered in the reports were very different. Until 2007 the themes of the *GFSR* chapters were very diverse, and it is difficult to argue that there were more pressing issues that the reports should have dealt with. Even so, it is noticeable that during the calm years, the *GFSRs* did not address potential problems such as the development of asset bubbles and their implications for financial stability, or the rapid growth of credit and the implications it could have for the financial system or, even more important, for the economy as a whole.

96. With the outbreak of the crisis it became clearer which themes should be addressed in the analytical chapters. In the April 2007 report, the themes still do not capture the importance of the crisis; the two research chapters cover issues such as the changes in the international investor base and the advantages of the globalization of financial institutions for financial stability (an issue that has become controversial in light of the crisis). The September 2007 report still did not catch up with events, as one of the chapters analyzes capital inflows to emerging market countries. The catch-up took place only in 2008: the April 2008 report deals with issues of valuation of structured finance and analyzes funding liquidity and its implications, while the October issue addresses stress in the banking system

and fair value accounting and pro-cyclicality and the spillover effects on emerging markets—three issues that were timely and important.

Quality of analysis

97. The quality of the research in the *GFSRs* is very uneven. There is an important gap between the first chapters of the *GFSRs*—which achieve a good balance between the concepts that are being presented and the evidence—and the research chapters—where the analytical issues are often vaguely presented and generally too little evidence is provided to substantiate the arguments.

98. In general, the *GFSR* research chapters are largely descriptive. The lack of a clear analytical framework that states at the outset the issues that the chapter intends to answer, the methodology that will be used, the main findings, and the policy implications that follow from these steps is a notable weakness of these chapters.

99. In our view, it would be useful to sharpen the questions that the *GFSR* research chapters are trying to analyze (as has been done to some extent since 2007), and to present more evidence to support the results. For instance, if the objective is to explain the fall in credit risk (as measured by the spreads on high yield bonds or on emerging market yields), it would be useful to show the evolution of the spreads at different times, and the determinants of the spreads (use correlations, tables, or graphs), and to give a sense of whether they are sustainable. Or, if the aim is to discuss accounting methods in the insurance industry, it would be useful to quantify the effects of different methods and to have a sense of whether there is, for example, a trade-off between accuracy (mark to market) and solvency (fair or accounting value), or some other quantification that could help the reader gauge the importance of the issue at hand.

Policy advice

100. The *GFSR* chapters, like those of the *WEO*, always end with a list of policy recommendations. Their policy recommendations tend to be more specific than those of the *WEO*, partly because the *GFSRs* look at narrower issues, such as the insurance companies or the pension fund industry, and hence they can give more concrete advice regarding the rules or regulations that are most appropriate to help financial stability.

101. Despite the greater specificity of their policy advice, however, the *GFSRs* tend to provide a "laundry list" of things that countries need to do, without giving a clear sense of priorities. Neither do they draw clear distinctions between emerging and developed financial markets.

102. As in the case of the *WEO*, the advice is almost always predictable, and the educated reader can anticipate the main recommendations. The typical policy advice is to improve supervision processes, adopt tighter regulations, or change the institutional framework.

Unfortunately, in cases in which the policy advice is more complicated, such as regarding the valuation of structured financial products (Chapter 2, April 2008), the conclusions are not clear-cut. They favor more transparency (which is difficult to argue with); urge that "rating agencies should provide investors with more analytical information;" and favor "[f]urther refinement and careful implementation of Basel II." None of these pieces of advice is likely to change a country's policy framework. Finally, about the valuation of structured financial products, the chapter says: "Greater attention to applying fair value results needs to be addressed. As experience is gained from the crisis, some weaknesses in the implementation of fair value as a valuation mechanism could usefully be addressed." This statement does not give the reader a sense of whether countries should move towards applying fair value accounting or not. This type of problem arises repeatedly in the chapters, showing that the Fund is not ready to take a strong stand on these issues.

103. A final problem, which goes beyond the research chapters of the *GFSR*, is that the policy advice does not successfully integrate the interconnections between financial markets and macroeconomic events and policies, nor does it take into account the interconnections among different financial intermediaries. This issue is addressed to some extent in Chapter 3 of the April 2008 report, but without precise recommendations. Regarding the problems with funding liquidity, this states: "Careful consideration will need to be given to these incentives in order to improve policies to reduce systemic risks in the years to come." How would a policymaker interpret this statement?

IV. Overall Evaluation of the Reports and Reflections for the Future

104. The analytical chapters of the *WEO* and of the *GFSR*, the two IMF flagship reports, are on the whole good quality chapters that address relevant macroeconomic and financial topics. The *WEO* chapters have been more mature pieces of research, to a large extent because that report has been developed over many years and because they mostly address issues that are part of the core work that the IMF has been conducting since its creation. The *GFSR* is a more recent report, whose analytical chapters have not yet reached the same level of maturity. Nevertheless, the quality of these chapters has been improving, in part because financial issues are now recognized as highly interconnected to the macroeconomic work at the IMF and in part because of a natural process of maturation.

105. The *WEO* and the *GFSR* each start with a chapter that provides an overview of recent economic and financial developments, generally presenting base-case economic and financial scenarios and identifying the main risks that could affect the outcomes. Ideally, the research chapters should provide the analytical underpinning for some of that discussion, or alternatively discuss an issue that is particularly relevant for macroeconomic or financial policies from a long-term perspective. In our view there is still a need to better integrate the two sections of the reports, including by more coordination in the selection of the topics that are covered.

106. The evaluation of the analytical chapters of both reports prompts a number of recommendations:

107. Regarding **structure and presentation,** both reports have now achieved a good level of quality. The recommendation is to insist on starting with clear questions that are linked to important policy issues and that are motivated with examples that policymakers can easily grasp. Given that the reports are intended for policymakers and analysts (including those in academia and think tanks) and the short amount of time that the researchers have to draft the chapters, the idea is that they should serve as a bridge between academics and policymakers. The chapters need to bring together the most recent analytical work and evidence and "translate" them into a language accessible to policymakers. The major value added should lie in the application of the analysis to policy issues. For this, the reports need to present clearly the policy implications of the empirical findings. This last recommendation is more relevant in the case of the *WEO*, which is the more technical publication.

108. The **choice of topics** is a critical decision, especially because the issues being addressed need to be timely and relevant for a large number of countries. In the case of the *WEO*, the most successful chapters are those that address the topics in which the Fund has clear expertise such as inflation, exchange rates, monetary and fiscal policies, or external imbalances. The treatment of structural or long-term issues is an option, but in those cases the authors need to make sure that these are related to the Fund's mandate. In the case of the *GFSR*, the selection of topics appears to be in line with the mandate of the report, though many times the analytical chapters have missed central issues that were important to understand financial risks (and that were sometimes mentioned in the overview chapter).

109. The **quality of the analysis** is on average good, but there is high variance. In the case of the *WEO*, it is better in those areas where the Fund has greater expertise and less successful in those where it has less. In the *GFSR*, it has improved over the years, though a large number of the chapters evaluated are more descriptive than analytical. The expectation is that the *GFSR* will continue to move along more analytical lines, though without becoming over-technical, choosing topics that are relevant for financial stability and macroeconomic performance.

110. An important issue is the **use of evidence** to illustrate specific points. In our view the use of boxes is very useful for the reader, and we would recommend a more intense use of partial correlations and graphs. We would also recommend relying less on econometrics, using these techniques mainly to quantify certain effects and to help the reader to understand the quantitative impact of the adoption of certain policies (e.g., fiscal multipliers) or of changes in key variables (e.g., the price of oil). The results should be presented clearly as scenarios, not forecasts.

111. The **policy advice** presented in the reports does not always follow from the analysis and is often too vague to help policymakers. In general, the advice is what one expects: in the

WEO, prudent monetary policy, avoidance of overvalued currencies, tightening in fiscal policy, better institutions, more flexible labor markets, and so forth; and in the *GFSR*, adoption of better regulatory frameworks and improvements in supervision, more transparency in financial institutions' accounting, provision of more information, and so forth. To be useful, the policy conclusions need to be more tightly related to the analytical framework and the evidence presented, and they should be more specific.

112. One possibility would be to try to distinguish different economic situations or country taxonomies and recommend different policies depending on where each country stands. For instance, the soundness of adopting an expansionary fiscal policy depends among other things on whether the economy has excess capacity or not, as well as on the size of the stock of debt and on the existing tax pressure. The policy conclusions could then differ depending on the situation of a specific country.

113. **Moving forward, one of the main challenges is to find ways to integrate and coordinate the work on the two reports**, especially regarding a better understanding of the various ways in which financial issues can affect macroeconomic outcomes and of the mechanisms through which macroeconomic developments can have an impact on financial stability. In today's world, as the recent international financial crisis has made clear, there are strong two-way connections between the financial markets and macroeconomic outcomes. A sharp deterioration in the financial markets affects the availability of credit to the private sector and thus affects economic activity. Conversely, a severe drop in economic activity can affect asset prices and the level of nonperforming loans in the banking sector and generate problems in that sector. In some extreme cases, such as in a financial crisis, governments might be forced to bail banks out through issuing debt and by providing guarantees on some of their liabilities, undertaking emergency policy measures to avoid a collapse of credit and of the payment system.

114. It seems that during the evaluation period there was essentially no coordination between the two reports regarding the topics covered and the interactions between financial and macroeconomic issues. In this regard the two reports have been working in parallel in their analysis of current economic developments, their analytical chapters, and their policy recommendations.

115. Achieving better coordination and integration will not be easy, not only because it requires a good understanding of the interaction between macroeconomic and financial issues (which is not discussed in traditional macroeconomic courses), but also because the work at the Fund and the responsibility for the preparation of the reports is in different departments.[13] While there is a review process for both reports, the bottom line is that there are not enough

[13] The *WEO* is prepared under the direction of the Research Department, while the *GFSR* is directed by the Monetary and Capital Markets Department.

interactions between the two products and each of the departments perceives the report as its own output.

116. In 2010, in an effort to integrate the analysis and policy recommendations the two reports shared a joint introduction. This is a move in the right direction but not enough to close the gap. Management could consider implementing measures to achieve more interaction between the two departments—which is particularly important at the early stages of the preparation of the reports. This process could involve some joint brainstorming meetings at the time when the topics are selected for the analytical chapters of the two reports; some joint working seminars to present the initial drafts of the chapters; and members of the two departments working together throughout the review process of both reports. In addition, the Fund could consider implementing a mobility policy between the two departments for the staff who conduct research and participation in drafting the chapters.

Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/11/06

Review of the IMF's *Regional Economic Outlook* Reports, 2003–09

Peter J. Montiel

May 20, 2011

Address. 700 19th Street, N.W., Washington D.C. 20431, U.S.A. Telephone. +1 202 623 7312 Fax. +1 202 623 9990 Email. feedback@ieo-imf.org Website. www.ieo-imf.org

 BP/11/06

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

Review of the IMF's *Regional Economic Outlook* Reports, 2003–09

Prepared by Peter J. Montiel

May 20, 2011

Abstract

This study evaluates the technical quality of the *Regional Economic Outlooks*. It covers all 44 outlooks issued between 2003 and 2009. It found the policy analysis contained in the publications to be uneven. Many of the analytical chapters were satisfactory, but some were weak. A common weakness in all *REOs* is that they fail to articulate the common features of the countries being analyzed (beyond being covered by the same administrative unit at the Fund) that justifies a "regional" approach as having value added over the global and country analysis. They were undermined by a tendency to advocate very specific, highly prescriptive policies rather than identifying policy options and the trade-offs often involved in conflicting economic objectives. Some prescribe specific policies without substantive explanations. Many of the instances of vagueness, unsubstantiated claims, missing analysis, incoherent analysis, and incorrect analysis found by the evaluation seem to arise from an uncritical acceptance of conventional wisdom. Exposing these documents to outside review before they are issued might help to address this problem.

JEL Classification Numbers: E0, F3, F5

Keywords: IMF Research, multilateral surveillance, *Regional Economic Outlook,* quality comparison, research evaluation

Author's E-Mail Address: Peter.J.Montiel@williams.edu

Contents **Page**

Tables

ABBREVIATIONS

AFR	African Department
AHF	Asian-focused hedge funds
APD	Asia and Pacific Department
ASEAN-4	Indonesia, Malaysia, Philippines, and Thailand
ASEAN-5	Indonesia, Malaysia, Philippines, Thailand, and Vietnam
CCA	Caucasus and Central Asia
CPI	consumer price index
EAC	East African Community
EU	European Union
EUR	European Department
GCC	Gulf Cooperation Council
GDP	gross domestic product
GIMF	Global Integrated Monetary and Fiscal Model
GVAR	global vector autoregressive model
HP	Hodrick-Prescott
ICR	interest coverage ratio
IEO	Independent Evaluation Office (IMF)
IT	inflation targeting
LAC	Latin America and Caribbean
LAC-7	Argentina, Brazil, Chile, Colombia, Mexico, Peru, and Venezuela
LIC	low-income country
MCD	Middle East and Central Asia Department
MENAP	Middle East, North Africa, Afghanistan, and Pakistan
NIE	newly industrialized economy
OECD	Organization of Economic Cooperation and Development
PPP	purchasing power parity
REO	*Regional Economic Outlook*
SSA	Sub-Saharan Africa
TFP	total factor productivity
WEO	*World Economic Outlook*
WHD	Western Hemisphere Department

I. INTRODUCTION

1. This background paper evaluates the 44 *Regional Economic Outlooks (REOs)* issued by the Fund's five area departments from 2003 to 2009 (Table 1 identifies the authoring departments and dates of the relevant *REOs*). Since 2003, the IMF's area departments have been publishing *REOs* to complement the *World Economic Outlook (WEO),* the Fund's flagship publication.[1] *REOs* provide regional perspectives on the general macroeconomic outlook for the region covered by each department as well as on specific economic issues that emerge as particularly important at the regional level. Preparation of the *REOs* has grown to absorb significant resources, including the creation of units within each area department dedicated mainly to the production of the *REO*. Accordingly, an outside evaluation of these reports seems timely and appropriate.

Table 1. *Regional Economic Outlook* Reports Covered in this Review

	Asia-Pacific (9)	**Europe (5)**	**Middle East and Central Asia (9)**	**Sub-Saharan Africa (13)**	**Western Hemisphere (8)**
2003					
Spring				X	
2004					
Spring				X	
Fall				X	
2005					
Spring				X	
Fall	X		X	X	
2006					
Spring	X		X	X	X
Fall	X		X	X	X
2007					
Spring	X		X	X	X
Fall	X	X	X	X	X
2008					
Spring	X	X	X	X	X
Fall	X	X	X	X	X
2009					
Spring	X	X	X	X	X
Fall	X	X	X	X	X

2. *REOs* have several functions, including to serve as a platform to engage authorities and other stakeholders in discussions on the IMF's global and regional outlook and to

[1] The *WEO* is published twice yearly by the Research Department of the Fund.

disseminate research studies focused on the region. The focus of this paper is on the technical quality of the regional outlook and studies. It also discusses presentational aspects that are important to allow country authorities and other stakeholders to assess whether the analysis and the recommendations are relevant for their own circumstances. The aim of this evaluation is to draw lessons to improve *REOs*, in particular as a tool for policy dialogue with country authorities in the different regions. Accordingly, while I note areas of strength in the documents reviewed, the focus is on areas in which I believe that there is room for improvement.

A. Evaluation Criteria

3. The reports' technical quality is assessed based on two main dimensions: general analytical quality and broad expositional principles. The general analytical quality was judged by the following criteria:

- The analysis should be clear and unambiguous in content.
- It should avoid making unsubstantiated claims.
- It should be thorough (avoiding loose ends) as well as internally consistent.
- It should have been thoroughly vetted to avoid obvious analytical errors.

4. The broad expositional principles are the following:

- The reader of these documents should be given enough information to form an independent opinion on the issues considered by the staff.
- The adoption of a regional perspective should be accompanied in each case by a description of how the region does or does not function as an economic unit.
- The economic analysis contained in these documents should be well grounded in the professional literature, with appropriate references to that literature.

5. Following these criteria, I present below an evaluation of the *REOs* produced by each of the Fund's five regional departments. I recorded a summary evaluation of each document reviewed in a ledger such as that presented below. Ratings ranged from 1 to 5, with U = "unsatisfactory" being 1, and S = "superior" being 5. Intermediate ratings were BA = "below average," A = "average," and AA = "above average" (Table 2). I used an absolute scale, e.g., the benchmark for "average" was based on my expectation of what would constitute adequacy under each criterion, rather than on a comparison across reports.

Table 2. Research Quality Indicator Form

	Ratings				
Criteria	S	AA	A	BA	U
I. Adequacy of information to form an independent opinion					
II. Provision of a regional framework					
III. General analytical quality					
IV. Grounding of analytics in the relevant literature					

B. Quantitative Indicators

6. Table 3 shows the summary of the ratings for all 44 regional economic reports evaluated, broken down by department and evaluation criteria category. The average rating across all departments was 2.0, which corresponds to the rating of "below average." This low rating reflects deficiencies in the reports on one or more criteria.

Table 3. Summary of *REO* Scores by Department and Category[1]

	Category				
Department	**I**	**II**	**III**	**IV**	**Average**
Africa	1.5	1.2	2.2	1.8	**1.7**
Asia and Pacific	2.0	3.1	2.3	1.4	**2.2**
Europe	2.0	1.4	2.0	1.6	**1.8**
Middle East and Central Asia	1.8	1.7	1.8	1.0	**1.6**
Western Hemisphere	2.0	4.0	2.9	3.1	**3.0**
Average	**1.8**	**2.3**	**2.2**	**1.8**	**2.0**

[1] Based on a scale of S = 5 and U = 1.

7. A closer look at the indicators reveals a great deal of unevenness in quality. The reports from the Western Hemisphere Department were rated the highest but only received an "average or acceptable" rating of 3.0. The reports from the African, European, and Middle and Central Asian Departments were rated below 2.0, pointing to the general weakness of the reports. However, the African and Middle Eastern reports in particular improved over time, so the low ratings received were pulled down by the earlier reports. One area of weakness revealed in many of the reports, especially from the European Department and early African and Middle Eastern reports, was that they did not provide a description of how the region functioned as an economic unit. Another common weakness, excluding Western Hemisphere, was that the reports frequently lack grounding in relevant professional literature. The later African and Middle Eastern reports provided more of a regional framework than did the early reports.

C. Outline of the Evaluation

8. The structure of this review is based on the criteria described above. In the next section, I evaluate the extent to which the *REO* documents as a group comply with my broad expositional principles. Sections III-VII evaluate the analytical quality of the *REO* documents produced by each of the Fund's area departments. Each of these sections describes the structure of the department's reports and comments on their analytical perspective as well as briefly on their general tone, before highlighting specific analytical strengths and weaknesses. Section VIII concludes the review by providing a critical discussion of the role of policy advice in *REOs*, and by identifying a set of research topics that could fruitfully be pursued at the Fund for the purpose of improving the analytical quality of future *REOs*.

II. BROAD EXPOSITIONAL PRINCIPLES

9. Do the *REOs* comply with the three broad expositional principles set out above? This section addresses this question from the perspective of the *REOs* as a group, leaving the evaluation of individual area department *REOs* for the five sections that follow.

A. Informing the Reader

10. A simple reading of the *REOs* suggests that the main intended audience consists of policymakers in the region, or the financial press, i.e., people who already familiar with the region's economy, because the reports provide very little background on the countries that comprise the region and also because they are highly prescriptive in tone. An additional role for these reports could be to engage academics and other stakeholders to think about and conduct independent research on regional economic issues that fall within the Fund's policy purview. This could better be achieved by providing more information—either within each report or in some background overview section/document—on relevant features of the individual economies and the region that could help to inform the analysis in the report. A reader who comes to these reports without this basic background information would find it difficult to form an independent view on the policy analysis provided in the *REOs* and thus to be a useful contributor to the policy conversation.[2]

[2] Even for individuals engaged in policy within each region the exposition in these documents could be much improved. While the *REOs* do not contain mathematics, some of the supporting empirical analysis is somewhat technical and (I suspect) not readily accessible to high-level policymakers and financial reporters. Probably for that reason, the reports often do not go into great detail on empirical methodologies, or provide the types of caveats that one would expect in more academic exercises. At times this makes for an awkward blend, especially when the supporting empirical analysis is described in the text, rather than in boxes or appendices. For the purposes of the policy world, I would suggest adopting a clear and consistently accessible tone in the text, while placing supporting evidence and empirical analysis in boxes and appendices (as indeed several reports do), where the methodologies employed can be described and evaluated in greater depth.

B. Implicit Regional Model

11. In addition to information on the characteristics of specific countries in each region, such as those listed above, it would be extremely helpful to offer readers a clear notion of how the staff perceives each region functioning (or not) as an economic unit—in other words, to clearly articulate the staff's implicit regional macroeconomic model up front.

12. If a region is just a collection of individual countries that are located close to each other but do not interact as an economic unit, then there is little justification for separate regional economic outlook exercises, other than spreading among the Fund's staff the task of bringing the *WEO's* global analysis down to a more local level. But in many cases there are important regional real and/or financial links among countries in a geographic area that indeed make the region a meaningful unit of economic analysis (e.g., trade and capital flows between North and South America in the Western Hemisphere, vertical integration of production between China and East and Southeast Asia, flows of bank lending between advanced and emerging European countries). At present these links, and the functioning of regional economies, emerge indirectly and almost incidentally in the *REOs*, leaving the reader to do most of the work in inferring how the region functions as an economic unit.

13. A more explicit description of the staff's perspective on this issue would help the staff sharpen its own regional analysis as well as make the *REOs* more transparent and analytically satisfying to their audience.

C. Links to the Professional Literature

14. The reports that I read often cite the professional literature on specific issues that they consider. Providing links to the literature is important for several reasons: to convey to the reader that the staff's analysis is professionally well grounded, to provide outside analytical and empirical support for some of the staff's specific views, and to provide further elaboration of the staff's analysis where that seems to be called for.

15. Unfortunately, however, the links between these reports and the professional literature tend to be deficient in three ways:

- In several cases, analysis is supported by simple made-to-order empirical exercises conducted by the staff even where more sophisticated work has already been done on the topic either inside or outside the Fund.

- References to the professional literature are not common enough. *REOs* should reflect state-of-the-art thinking by the profession on the issues being discussed by the staff. To make the reports more authoritative, references to current professional opinion and/or research done outside the Fund should be included whenever possible. I found many instances where such references were not provided (some of these instances are described in what follows).

- Some of the references that are made to the literature do not do justice to the diversity of professional opinion on a particular issue. Wherever possible, reference should be made to survey pieces that summarize the state of the art, rather than to individual pieces of research espousing a view that may have been challenged by others.

III. Asia-Pacific

16. I read nine *Regional Economic Outlook* reports prepared by the Asia-Pacific Department (APD), covering the period October 2005 to October 2009.

A. Structure

17. These reports have a uniform structure: each focuses on a general theme (e.g., "Building a Sustained Recovery" in the October 2009 report) that guides the discussion in an overview chapter and motivates the topics chosen for inclusion in two or more "analytical" chapters. Most of the analytical chapters are well motivated and relevant to the general theme; for example, Chapter 2 of the May 2009 report, on the lessons from past recessions, is pertinent, well executed, and timely.

18. Some chapters are not particularly illuminating, not necessarily topical, or not obviously relevant. Examples follow:

- Chapter 3 of the April 2007 report discuss housing prices in Asia and resource-rich countries in the Asia-Pacific region. It is textbookish in tone and not particularly informative or insightful.

- In a report issued in the midst of an international crisis (November 2008) it seems odd to have devoted the analytical chapters to developments such as the implications of commodity prices for inflation and of population aging. Undoubtedly these are important recurring issues for the region, but one suspects that the topics were chosen well before Lehman Brothers collapsed in September of that year.

- Similarly, the discussion of the Japanese "lost decade" experience in the May and October 2009 *REOs*, while useful and illuminating from the perspective of the countries at ground zero in the Great Recession, is not nearly as useful for the purpose of analyzing events within the Asia-Pacific region, where financial sector problems have not been at the heart of recent events.

19. The earlier APD reports share the problem, mentioned above, of the absence of a description of the region as an economic unit—that is, the implicit regional "model" that underpins the analysis in the overview chapter. The reports refer to many analytical country

groups,[3] which seem to be organized by common features that these economies share, rather than by an implicit regional model. In some ways this may be a sensible approach, given that the structure of economic relationships among countries in the Asia-Pacific region is quite different from that in, say, the Western Hemisphere, where the emerging-market and low-income countries (LICs) depend heavily on trade and financial flows with the two large industrial countries in the region (the United States and Canada). By contrast, Asian economies, while highly open, tend to concentrate their exports outside the region, and also interact extensively outside the region through financial flows.

20. The later Asia-Pacific *REOs*—starting with the September 2006 report, which describes well the increased role of China as an engine for regional growth—approach closer to a coherent analytical view of how the region fits together as an economic entity. Adopting this more coherent view enables these reports to make some intelligent, not-so-obvious, points that illustrate the value of underpinning the analysis with some sort of regional model. For example, based on its view of economic interactions within the region, APD takes a reasonable (and ex post correct) view of the limited prospects for "decoupling" that existed within the region when the international crisis broke out. It is also able to interpret the severe depth of the contraction in Asia, as well as the role played by exports and inventory cycles in the West in driving the especially strong recovery that subsequently emerged in Asia.

21. The progression from seemingly ad hoc country groupings to a unified view of the regional economy in the APD reports, and the insights that emerge from the latter, support the view that each *REO* should be buttressed by a vision of how the regional economy works.

22. A shortcoming of the structure of the APD reports is that they pay relatively little attention to the role of LICs in the region. Though these countries are mentioned in every report, including in some boxes that address issues specific to them, their overall role in the reports is very minor. The reports give more detailed treatment of emerging economies in the region than of either advanced or low-income countries (see, for example, the rather sketchy discussion of Australia and New Zealand in Box 2 of the May 2006 report).

B. Analytical Perspective

23. In some of the early APD reports the discussion in the first ("developments") chapters lacks much of an analytical framework. Some statements simply reproduce accounting identities: an example is the statement in the September 2005 report (p. 9) that reserves went up more than the current account surplus, because of capital inflows. Others are ambiguous: for example, "capital inflows have slowed considerably, because of a strengthening of the

[3] Industrial Asia, emerging Asia (subdivided into newly industrialized economies (NIEs) and the ASEAN-4/5, depending on whether Vietnam is included), and low-income countries (LICs).

U.S. dollar" (ibid.)—have they slowed because the dollar was expected to strengthen, or did the dollar strengthen because capital inflows to Asia slowed?

24. The analysis becomes more sophisticated in the later reports. By and large, I found the analytical approaches adopted in the later reports up to date and very satisfying—basically offering good solid applied macroeconomics. For example, they acknowledge the role that aggregate demand contractions have played in creating output gaps, they take appropriate account of the role of intertemporal considerations and expectations, they attempt to quantify relationships where appropriate with cross-country empirical work, and they show an awareness of the relevant professional literature. All in all, this is just the kind of work that one would hope to see from the Fund.

25. Moreover, in a number of instances these reports give a fresh and well-documented perspective on the situation in Asia. Such useful insights are provided from the earliest reports onward. Some examples follow:

- The September 2005 report points to the role of an "investment drought" rather than a "savings glut" in driving current account surpluses in emerging Asia (p. 22). Overall, the discussion of saving and investment in Asia in that report (p. 25) is well motivated, reasonable, and insightful.

- The surprising rise in income inequality in Asia is well documented in Chapter 6 of the September 2006 report, with interesting and policy-relevant patterns identified.

- Chapter 2 of the October 2007 report, on what Asia learned from the 1997–98 crisis, predicts the region's resiliency in the face of future crises, which has indeed turned out to be the case.

- The estimation of current account determinants in the November 2008 report is interesting and innovative, and gives plausible results. I consider it a very useful contribution.

- Another useful contribution is the construction of indicators such as the Financial Conditions Index, introduced in the November 2009 report (p. 17).

- The use of the Global Integrated Monetary and Fiscal Model (GIMF) (May 2009 report, p. 25) to assess the growth impact of sustained fiscal stimulus (instead of premature retrenchment) is a laudable piece of sound quantitative policy analysis. Indeed, the May 2009 report contains several good uses of quantitative indicators, including of the interest coverage ratios in Chapter 3 (p. 44) on corporate vulnerability.

26. My one reservation regarding the analytical perspective in the APD reports is the discussion of financial issues in some chapters of the earlier *REOs*. These read too much like

the financial press or financial newsletters, in that they are somewhat divorced from the underlying macroeconomic analysis and contain unnecessary jargon that tends to obscure the analysis for a general macroeconomic audience. An example is the reference to a "not unexpected" equity market correction in the September 2006 report (p. 16). The claim that stock market corrections can be widely anticipated strikes a macroeconomist as strange, and gives this writing too much the flavor of financial reporting. Box 2.1 on Asian-focused hedge funds in the same report uses too much jargon; terms such as "long-short and market neutral securities" and "family offices" should be defined for this audience. Similarly, the April 2007 report (p. 16) refers to "markets where trades had become crowded" and to "'real money' investors." The section on recent financial developments in the October 2007 report also has too much of the flavor of financial reporting for its own sake. Is it really very useful to know how Asian-focused hedge funds have fared? It may be, but the links with macro developments in the region are not established in the report.

C. Boxes

27. Most of the topics treated in boxes in the APD reports seem extremely well chosen. In the September 2005 report, clearly relevant boxes (on electronics exports, the impact of oil prices, and the effects of the removal of textile quotas) are not integrated into the text, but appear at the end of the document. The incorporation of boxes into the body of the text, beginning with the September 2006 report, provides more continuity for the reader. Particularly useful boxes in subsequent reports include:

- Empirical analysis of the trade-creating or diverting effects of Asian regional trade agreements, based on gravity models (May 2006 report, Box 9).
- An informative discussion of sovereign wealth funds in Asia (October 2007 report, Box 1.3).
- Discussions of the differential impact of the crisis on LICs, of estimating the impact of countercyclical fiscal policies on growth, and of whether policy rates have been passed through to lending rates (October 2009 report, Boxes 1.1–1.3).
- A discussion of saving in China that is timely in view of the need to stimulate domestic sources of demand (October 2009 report, Box 3.2).

D. Tone

28. I did not find the Asia-Pacific *REOs* to be excessively prescriptive in tone, unlike some of those from other regions.

29. But what comes through clearly from the Asia-Pacific *REOs* is the extent to which the Fund's policy recommendations for countries in the region are based on a *global* perspective, rather than the *single-country* or *regional* perspective that the authorities in the region are

themselves more likely to adopt. Much of the advice offered to regional economies in these documents is clearly intended to support the Fund's particular perspective on the "global imbalances" issue—a desire for Asia (especially China) to reduce its current account surpluses by increasing domestic consumption, and for these countries to adopt more flexible exchange rate arrangements (see, for example, the October 2007 report, p. 21), and allow their currencies to appreciate relative to the U.S. dollar. Advice to improve social safety nets, pursue financial liberalization, and improve corporate governance is overtly motivated by the goal of increasing household consumption and reducing corporate saving in Asia.

30. Framing its policy recommendations from a global perspective may be unavoidable, given that the Fund is a global institution. But to the extent that APD believes that these policy objectives are appropriate even from the narrow self-interest of the Asian countries themselves, the *REO* reports provide a unique opportunity to make this case. From my perspective, they did not sufficiently capitalize on this opportunity.

E. Links to the Professional Literature

31. *REOs* should be well grounded in the professional literature. At times, some of the APD reports appear to try to reinvent the wheel. For example, Chapter 5 of the May 2006 report contains a very comprehensive analysis of why China's household saving rate is so high, but it contains very few references to the professional literature on this issue. Has there been so little work on this topic outside of the Fund? If not, the absence of links with the professional literature is a serious drawback on this important topic. Box 1.4 of the September 2006 report, on remittances, is very appropriate but seems to be drawn mainly from one paper; it does not begin to do justice to the literature on remittances, and does not sufficiently identify what we still need to learn about the phenomenon. In the October 2007 report (p. 30), the claim that China's real exchange rate is undervalued is an extremely controversial one, and should have been backed with evidence and extensive references to the professional literature.

F. Analytical Strengths

32. These documents occasionally offer insightful and novel perspectives that I have not seen elsewhere, on important issues of substantial policy relevance. For example, Chapter IV of the October 2007 report, on trade patterns in Asia, is very informative, especially as regards its analyses of "decoupling" in Asia and of the impact of the crisis on economic activity in the region; it also sheds useful insight on the impacts of exchange rate changes on Chinese export prices (p. 52). The indexes of export similarity and export overlap in the same report are innovative and informative constructs, at least for someone (like myself) who is not a trade economist.[4]

[4] My one quibble here would be with the identification of "high-tech" with "high valued added" (p. 46).

G. Analytical Weaknesses

33. The APD reports have many positive features but they also have some analytical deficiencies. These are discussed below under the headings of vagueness, unsubstantiated claims, missing analysis, incoherent analysis, and (in my view) incorrect analysis.

Vagueness

34. A recurring problem in these reports is that, while reference is often made to a "baseline scenario," there is no explanation of how this baseline was compiled (see, for example, the April 2008 report). This makes it difficult for the reader to judge how much confidence to place in the baseline scenario.

35. Vagueness also characterizes some rather important judgments that have policy implications. For example, the claim in the October 2007 report that Asia remains "too dependent" on exports as an engine of growth is hard to interpret. The origin would appear to be an agenda based on "global imbalances," but the report provides no metric by which to judge "excessive" dependence. Moreover, a judgment like this seems to run counter to the Fund's past advocacy of outward-oriented growth, but this apparent contradiction is neither acknowledged nor explained.

36. A third problem is that the methodology applied in some of the underlying empirical work is left murky. Chapter 3 of the April 2008 report, on de-linking in Asia, for example, is well executed and was quite topical at the time. However, the recession scenario (run with the general equilibrium model), which combines a slowdown in U.S. activity with a decline in confidence, is artificial and not well explained. It is unclear how the scenario was constructed, or how the general-equilibrium effects of a decline in confidence were specified and quantified. It would have been preferable to describe the alternative scenario as it was actually operationalized and then claim that this is intended to capture a decline in confidence. Similarly, as I will argue in Section VII below, the use of "fiscal space" indicators is an extremely important background research issue for all of the *REOs*, and the May 2009 APD report uses such an indicator (p. 26), but leaves its construction unexplained.

Unsubstantiated claims

37. Unsubstantiated claims should be avoided, since they undermine the credibility of the policy advice offered in these reports. They are more frequent than one would like. Some examples:

- In the May 2006 report, the slowing of capital inflows into the region over the period preceding the report is attributed to narrowing interest rate spreads and waning expectations of renminbi revaluation (p. 12). The report explains the latter, in turn, by the nominal renminbi revaluation that had taken place on July 21 2005, as well as tighter monetary policy in the United States. This is a very plausible explanation of an endogenous outcome in terms of exogenous variables, but it is not backed up by evidence. Since changes in capital flows can potentially be driven by a wide range of factors, singling out any specific cause requires some justification.

- Statements about the stance of fiscal policy should always be backed up with data using a widely accepted indicator of the fiscal stance. This is not always done in these reports. For example, in the September 2006 report (p. 28) the statement that fiscal policy is broadly neutral is not backed up by data on indicators such as fiscal impulses.

- It is commendable that the April 2008 report (p. 25) acknowledges that most countries in the region have fiscal space for countercyclical policies. (Past Fund orthodoxy on this issue stressed fiscal sustainability at the expense of short-run stabilization objectives.) But the report does not indicate the basis for this conclusion.

Missing analysis

38. Probably the single most important shortcoming of the APD reports is a lack of complete or in-depth analysis of the issues discussed. Several examples will illustrate what I mean:

- The Asia-Pacific *REOs* frequently state that countries in the region should increase domestic demand in order to address world imbalances, but invariably stress that these increases need to be achieved without activist stabilization policies—that is, that they should happen through "structural reforms." It is easy to agree that such reforms (e.g., improved safety nets, financial reform, reforms in corporate governance) are likely to help, but over what time horizon? This issue is not addressed. Is the problem of global imbalances not so pressing, so that these reforms can be given time to have their effects on household spending in Asia? Or, if they are pressing, what other measures can be taken to rebalance demand? This discussion is missing.

- In accounting for the evolution of debt ratios in the region (e.g., in the May 2006 report, p. 23), decompositions of debt accumulation—into the contributions of primary deficits, real interest payments, inflation, valuation changes arising from exchange rate movements, etc.—would have been very useful in helping to understand the sources of changes in debt/GDP ratios.

- The analysis of the investment decline in ASEAN-4 countries in Box 5.1 of the September 2006 report interprets the decline as the result of credit constraints on the nontraded goods sector. But it is unconvincing, since it does not address whether the reduction of credit flows to the nontraded goods sector is the result of supply or demand contraction. The post-crisis real exchange rate depreciation that these countries experienced suggests a role for the latter, but again, this needs more investigation.

- Chapter 3 of the October 2007 report gives a clear explanation of sterilized intervention and a good survey, based on the literature, of where we stand on the effectiveness of sterilized intervention. But the tests conducted in this chapter for the effectiveness of intervention are very simple, and the reader is not told whether more sophisticated techniques are available and, if so, why they were not used here. Most importantly, while the chapter appropriately concludes with reservations about the simple methodology it employs, it gives no indication of how seriously one should take the results in light of those reservations. This is important, because if intervention is ineffective, then Asian countries would be misguided in conducting intervention policy with the objective of achieving a real exchange rate target. It is disappointing that this important implication is not brought out in the document.

- In the April 2008 report, which offers sensible projections for other variables, there is no recognition of the effects that weaker remittance flows will have on the Philippines (p. 22)—an important remittance recipient.

- Box 1.3 in the November 2008 report (p. 33), on headline and core inflation, is very useful in the context of the report, but the analysis is incomplete: the fact that households care about headline inflation does *not* call into question the usefulness of core inflation as the primary or implicit operational target. The volatility of headline relative to core inflation would require introducing more volatility into real output if monetary policy were to stabilize headline inflation, and volatility in real output is something else that households care about. This trade-off is not explicitly mentioned.

- The November 2008 report repeats the standard policy prescription for China. This is certainly understandable from a "global imbalances" perspective, but the case should instead be made that it is good for China itself, if the staff indeed believes this to be true. The fact that demographics will otherwise continue to put upward pressure on China's current account surplus, as suggested in this chapter, does not make that case, because there is no reason to believe that the resulting current account surplus would not be an optimal one.

- By contrast, the advocacy of a boost in Chinese domestic consumption in the May 2009 report is easy to justify in China's own interest, in light of the crisis-induced collapse in external demand. But here the report makes the mistake of advocating real exchange rate appreciation independently of what would be implied by stronger domestic demand—in fact such appreciation would in and of itself *discourage* consumption of domestic goods.

- The May 2009 report (p. 68) may be right that the Bank of Japan's approach to quantitative easing (buying government bonds) should use channels of influence similar to those of the U.S. Federal Reserve Board (intervening in markets that appear stressed), but it is not obviously so, and there is much confusion about exactly how these policies are supposed to work. The point deserves more consideration than it receives.

Incoherent analysis

39. In some of the reports the analysis at times lacks internal consistency, though I did not detect a particular pattern or particular area of analysis in which this problem was especially evident. Some examples follow:

- The statement (September 2005 report, p. 40) that local financial markets have not been able to keep foreign savings within the region—implying some shortcomings in these markets—is based on the view that without these shortcomings, capital inflows into the region would have resulted in larger current account deficits (or smaller surpluses). The idea seems to be that Asia is perceived as an attractive place to invest, but because Asia's financial markets are poorly developed, capital inflows into Asian countries tend to flow back out again, so they are not absorbed via current account deficits. This reasoning does not make sense to me. The region's financial markets may indeed have shortcomings, but the reason why capital inflows have not had larger impacts on current account deficits is exchange rate policy decisions (specifically, intervention in foreign exchange markets), rather than financial market performance. In the absence of reserve accumulation, those inflows would have resulted in faster real exchange rate appreciation and larger trade deficits or smaller surpluses.

- The September 2007 report provides another example of strange reasoning. It claims that "overheating and credit-fueled asset price rises remain a concern, and the scope for easing monetary policy if growth falters is much more limited" (p. 25). But if growth falters, that implies the *absence* of overheating, so overheating cannot be used to rule out expansionary policy in the event of faltering growth. This type of reasoning appears again in the April 2008 report (p. 24), regarding the scope for monetary accommodation in the event of an imported slowdown. The claim is that several countries would have little room for monetary accommodation because of rising concerns with inflation. But if slower activity is indeed imported from the world slowdown, core inflation should fall, and that should alleviate the concerns alluded to before. The statement in the same report that China would have limited room to ease in the event of an imported slowdown, because of its concern with booming investment, is similar: if the imported slowdown materializes, booming investment should certainly become less of a concern.

- The October 2007 report (p. 20) claims that a depreciation of the exchange rate would result in a loosening of monetary conditions, which may preclude countercyclical policies in response to a slowdown. I do not understand either step in that causal chain.

- The November 2008 report (p. 37) states that the emergence of high energy prices when wages are showing little indication of excess demand pressure may mean that energy prices are a better indicator of the output gap than wages are. This claim seems strange at best. After all, different sectors may reach full capacity at different times, and, as they do, one would expect to observe relative price changes. Increases in some subset of prices may thus provide a poorer indication of the economy-wide extent of excess capacity than an aggregate wage index.

- The suggestion in the May 2009 report (p. 23) that Asian central banks may have to turn to "quantitative easing" (expansion of their balance sheets) in light of policy rates being close to zero, is puzzling, since there appears to be no breakdown in domestic Asian credit markets that would suggest a need for Asian central banks to step in to provide the financial intermediation that private institutions are unable to perform. How else is quantitative easing supposed to work?

- The October 2009 report (p. 4) makes some odd statements about the consequences for Asia of lower consumption growth in the West. Specifically:

    - The need for the region to live with smaller current account surpluses is by no means an obvious consequence, since effects on the saving-investment balance are not clear.

    - The need to live with more flexible exchange rates is also not obvious, but if it is true, probably suggests more *depreciated* exchange rates—which is probably not the message the report intends to send.

Incorrect analysis

40. Some topics appear to be incorrectly analyzed in more than one instance and in more than one area department's *REO*. Examples of these in the APD reports concern the effects of exchange rate depreciation on inflation and a confusion between inflation and relative price changes:

- The May 2006 report (p. 20) expresses the fear that the depreciation of regional currencies could lead to an uptick in inflation. This is a recurring theme in many *REOs*. But this fear is not justified in the text and seems entirely unwarranted in the context at hand, especially since core inflation had not previously responded to changes in other relative prices (such as those for food and energy) in APD countries. Why should changes in import prices arising from exchange rate changes have different effects?

- Although they do this less than the *REOs* of some other regions, the APD reports occasionally confuse inflation with relative price changes (as in the November 2008 report, p. 13). The welfare effects of sharp increases in food prices stem mainly from changes in relative prices, not from an increase in the aggregate price level.

41. Other instances of incorrect analysis are more *sui generis*. Here are two examples:

- In the September 2006 report, Chapter 5 on private consumption usefully employs a cross-country regression model to gain perspective on consumption behavior in Asia, but the implementation is deficient in several respects. For example, the finding that positive terms of trade shocks tend to reduce consumption cannot be explained by consumption smoothing in the face of transitory shocks, as suggested (p. 57), since transitory shocks would tend to leave consumption unchanged. Also, the inclusion of country dummies makes the deviation of actual from fitted values uninformative about whether the fundamentals explain consumption behavior in the region; it would have been much more informative to report the country dummies for Asian countries. The report concludes that the transitory (one-year) decrease in private consumption that took place in some countries after the Asian crisis means that countries should do more to reduce the uncertainty faced by households (p. 61). This conclusion may be right, but the evidence given provides weak if any support for it. Indeed, none of the policy conclusions on p. 61 is supported by the results, and it is hard to understand why, if real exchange rate appreciation is expected to boost private consumption, that variable was not included in the regressions.

- The November 2008 report (p. 44) claims that capital should flow from slow-growing countries to fast-growing ones. This is incorrect—capital should flow from countries with a low marginal product of capital to those with a high one. This is an issue that concerns *levels*, not growth rates.

IV. European Department

42. I read five *Regional Economic Outlook* reports prepared by the European Department (EUR), covering the period from October 2007 to October 2009.

A. Structure

43. As for other regions, what is needed is some sense of how the regional economy works. Like the APD reports, the *REO* documents prepared by EUR do not open with a description of how the region functions as an economic unit, and do not refer to background material on that issue. This background information—especially about intraregional real and financial links—would have been particularly useful in the European case, given the role of the euro in the region and the aspiration of many European countries to euro accession, with associated implications for macro institutions and policy regimes.

44. Also like the APD reports, the European reports feature a "recent developments and outlook" chapter, and one or more analytical chapters focusing on an issue of topical

importance for the region. Some of these chapters are well argued and relevant and some are not. The November 2007 report contains an analytical chapter focusing on financial systems in Europe; in view of the U.S. subprime crisis that was ongoing at the time, and the rapid credit growth in emerging Europe, that focus seems very appropriate. The October 2008 report, by contrast, has analytical chapters focusing on coping with high commodity prices, on diverging credit cycles in Europe, and on cross-border labor flows. In light of what was happening in the world economy at the time, this choice of topics seems a little artificial and probably illustrates how events can sometimes get ahead of the planning for the content of the analytical chapters.

45. The October 2009 report, which contains three analytical chapters (on the effect of crisis on potential output, on the effects of changes in potential output on macro policies, and on emerging economies), illustrates a somewhat different problem: the topics of these chapters may be relevant for policy but there is simply not enough to say about them to warrant full chapter treatment, as discussed further below.

B. Analytical Perspective

46. Overall, the EUR reports are much more doctrinaire than are those of other regions about the priority that countries in the region should give to reducing inflation at the expense of other objectives, and about the role of tight macroeconomic policies as the cure for many ailments, and they give relatively little solid analysis to support these views. This perspective tended to dominate in these reports until the true dimensions of the current international crisis became evident, even after the initial slowdown in both the United States and Europe. It is possible that the evident overheating problem that existed in several European countries accounts for the dominance of this perspective, or the perspective may simply represent a departmental "fixed effect"—that is, a longstanding departmental orientation.

C. Boxes

47. Some of the boxes included in EUR reports are very illuminating and others leave much to be desired, either for reasons of relevance or quality. The better boxes include:

- Box 2 of the November 2007 report, on intra-European integration (p. 14), which estimates the growth benefits derived from intra-European trade.

- Boxes 9 and 10 of the November 2007 report, which provide specific examples of financial development in Hungary, and of its shortcomings in Ukraine. The contrast is revealing. Particularly useful is the diagnosis in Box 10 of what needs to be done to foster additional financial deepening in Ukraine.

- Box 8 of the April 2008 report (p. 44), which contains estimates of potential growth rates in Europe. While these estimates may be crude, something like them belongs where this box was placed.

48. Boxes that, in my opinion, fall short include:

- Box 6 of the November 2007 report, which argues that Latvia's small size and limited workforce have impeded the establishment of large export-oriented manufacturing projects, but at the same time that Latvia needs to shift to higher value-added products. What are these products, if not manufacturing for export? The box does not say.

- Boxes 6 and 7 of the April 2008 report, which address the real effects of financial shocks in emerging Europe. The exercise described in Box 6 is a little too mechanical, and the model simulated in Box 7 too stylized, to yield credible quantitative estimates.

D. Tone

49. These reports put almost an uncompromising emphasis on tight macroeconomic policies and they occasionally lapse into somewhat empty formalism.

50. Illustrating the first of these problems, the October 2008 report (p. 10) states that "in the current circumstances, discretionary loosening [of fiscal policy] is unlikely to be effective" in Europe. This conclusion is not explicitly justified. The emphasis on tighter policies in response to the crisis in emerging Europe (p. 16) is also surprising, given that no evidence is presented to suggest dangers to fiscal solvency in those countries. Similarly, tight monetary policy and accompanying real appreciation are also advocated, to contain second-round effects of food and energy price shocks and to reduce external imbalances. This seems completely unrealistic in the face of the projected contraction in real activity. Along the same lines, the focus in Chapter 2 of the same report (p. 19) on the risks to inflation from higher commodity prices seems highly inappropriate for October 2008. This is especially so since all of the empirical findings reported (p. 22) suggest little reason to fear pressures on core inflation in the advanced countries.

51. The October 2008 report (p. 78) provides the results of a simulation model exploring the effects of labor mobility, but all the simulated effects seem quite intuitively obvious; the model results seem to add formalism but not new understanding.

E. Links to the Professional Literature

52. While I particularly liked the treatment of financial sector issues in the EUR reports, the links to the professional literature about this topic were often weak. The November 2007 report (p. 28), for example, presents evidence on the benefits of financial development but the arguments in this section do not distinguish between domestic financial development and international financial integration. There is strong evidence on the benefits of the former, but very ambiguous evidence on those of the latter. This problem is still present in the October 2008 report, in which Box 11 (p. 61) gives a very misleading picture of the evidence on financial integration and growth. The box reads as if the verdict on integration and growth is

unambiguously positive, while the bulk of the evidence—including some prominent work done at the Fund—tends to find very ambiguous effects.

53. Along the same lines, there is a large professional literature based on international experience concerning the factors that are conducive to financial development, but this literature is barely referenced where these issues are discussed in the November 2007 report (p. 62). The analysis there relies heavily on correlates of financial development in emerging Europe, based on background work done for this chapter. Why reinvent the wheel on this heavily-researched issue? And if the answer is that a specific focus on emerging Europe is perceived as desirable, why not ground the interpretation on the international experience (to see where emerging European countries fit in) rather than restricting the sample to emerging Europe?

54. A separate instance of the same problem concerns the crisis in emerging Europe. The soft landing for emerging Europe that was projected in the April 2008 report (p. 12) contrasts sharply with what actually happened. Hindsight is, of course, 20/20, but the risks inherent in the evolution of these economies, particularly those in the Baltics, would perhaps have been more evident if the report had taken a broader international perspective and applied the lessons from crises elsewhere. A general problem is that these reports are too Euro-centric.

F. Analytical Strengths

55. The EUR *REOs* provide particularly good analysis of financial sector issues. The perspective taken in the chapters on finance in the November 2007 report, for example—that financial innovation is extremely valuable, but can get ahead of regulatory mechanisms—seems exactly right; the idea should be not to stifle financial innovation, but to ensure that it does not outpace the effectiveness of regulation. And while the advice on macro policies often tends to be more doctrinaire in the European reports than in others, this is not always true; for example, the advice given in the April 2008 report (p. 8) regarding monetary, fiscal, and financial sector policies is sensible and sufficiently nuanced.

56. The many examples of useful and high quality analysis include:

- The use of the National Institute Global Economic model to estimate the real output effects of financial market shocks (April 2008 report, p. 31).
- A valid and well-placed criticism of reduced gasoline taxes (October 2008 report, Box 6, on p. 30)—recall the discussion of this issue in the U.S. presidential contest at the time.
- A useful summary of what went wrong with securitization (October 2008 report, Box 9, on p. 40).

- The use of the global vector autoregression exercise to measure financial interdependence in Europe (October 2008 report, p. 59).

- A plausible and important argument that fiscal multipliers may have increased in size in the current crisis environment (May 2009 report, p. 31).

- The use of the GIMF model to simulate the effects of the crisis on fiscal multipliers, as well as to illustrate the potential benefits from coordinated fiscal expansions in the euro area (May 2009 report, p. 32).

- A very useful discussion of the role of the Stability and Growth Pact in allowing discretionary fiscal responses while ensuring medium-term sustainability (May 2009 report, p. 42). Overall, Chapter 2 of the May 2009 report, on fiscal issues in response to the crisis, is excellent. Chapter 3, on the crisis experience of emerging Europe, is also quite good, in particular on the determinants of bond spreads in these countries (Table 12).

- A revealing comparison of crisis dynamics in Germany, the United Kingdom, and Spain (October 2009 report, Box 1).

G. Analytical Weaknesses

Vagueness

57. Like the APD reports, the *REOs* prepared by EUR are sometimes vague on important issues. The October 2008 and October 2009 reports provide two notable examples.

58. The October 2008 report claims (p. vii) that to safeguard long-term fiscal sustainability in the region requires deficits to be kept within the limits of fiscal rules. But what does this mean? Does it mean that contingency clauses in these rules should not be activated under present circumstances, or that the fiscal response should respect those clauses? It would have been useful to clearly describe the constraints that the Stability and Growth Pact places on the policy response to the crisis, and to discuss whether those constraints allow for an adequate fiscal response. As mentioned above, a good discussion of these issues is indeed provided in the May 2009 report.

59. The October 2009 report leaves rather vague the impact of the crisis on potential output in Europe. The chapter devoted to this issue (Chapter 2) does not offer a convincing single methodology to assess the effects of the crisis on potential output, or even a convincing range of plausible estimates of these effects. It describes several approaches (growth regressions, HP filters, sectoral decompositions) but makes no attempt to reconcile them with each other, leaving any conclusions about the size of the impact rather vague. It is not clear how policymakers could have applied this material to inform themselves about likely quantitative impacts of the crisis on capacity growth.

Unsubstantiated claims

60. Much too often, these reports make claims that are stated as fact but are not supported by evidence. A partial list of examples follows:

- A claim that labor market reforms in advanced countries have been responsible for increased employment rates there (November 2007 report, p. 16). This is quite plausible, of course, but where is the evidence? Are there no other possible explanations for increased employment rates?

- A statement that lower productivity growth in the financial sector accounted for half of the economy-wide productivity growth gap between Europe and the United States during 1996–2003 (November 2007 report, p. 24); no reference is provided to the studies documenting this fact.

- How was it judged that “reforms in some parts of the region have not progressed enough to sustain current growth rates” (April 2008 report, p. 38)?

- It is not at all clear that the appreciation of the euro in 2008 was caused by “insufficient progress in resolving global current account imbalances” (October 2008 report, p. 1). A more likely explanation, in my view, was a short-run safe-haven effect. How did the authors select one cause rather than the other?

- Similarly, the statement is made that commodity price shocks played an important role in slowing economic activity in Europe (October 2008 report, p. 1). This may be so, but what is the basis for this claim? Is there supporting empirical work documenting the effects of terms of trade shocks on real economic activity in Europe?

- The October 2008 report also prescribes policy objectives that are not supported analytically. For example, on p. 80 it suggests that new member states should seek to improve labor force participation rates. Why should they do so? What are the distortions that are keeping participation rates suboptimally low? These are not explained. On p. 81 it suggests that countries should adopt policies to encourage return migration. Why? Why does the market not already produce the optimal amount of return migration? Again, the source of the problem is not identified.

- A reference to “depreciations beyond levels required to bring real effective exchange rates in line with fundamentals” (May 2009 report, p. 1) presupposes more knowledge than we really have about what the fundamental levels of real exchange rates may be.

Missing analysis

61. These reports contain many instances where one would have liked to see more analysis. The November 2007 report contains several examples:

- Why have some countries in Europe been experiencing much more rapid housing price increases than others (p. 8)? Is this related to developments in the mortgage market?

- Exactly which tax-induced distortions have contributed to the high rate of credit growth (p. 12)? This is discussed much later, but there is no suggestion here that such a discussion is forthcoming.

- The statement that high tax burdens in emerging economies imply that spending reforms are the only option for fiscal consolidation (p. 12) may be right, but it assumes that the marginal social cost of additional taxation exceeds the marginal social benefit from current expenditure programs. This case has not been made, and cannot be made solely on the basis of the size of the tax burden.

- Why have policies had limited success at slowing the pace of credit growth (p. 35)? What has been tried, and why didn't it work? There is some discussion of this later in the report (p. 44), but not nearly enough.

- Reference is made to a "blistering" rate of financial deepening (credit growth) (p. 35). But how fast is too fast? There is a curious disregard here of the experience of emerging economies outside Europe that have suffered crises because of inappropriate domestic financial regulation (e.g., Chile in the early 1980s, East Asia in 1997). Why not draw the obvious comparisons? Why not use the empirical literature on the determinants of banking crises to evaluate the prospects for problems in emerging Europe? Similarly, how does the behavior of the ratio of private credit to GDP in Table 4 (p. 38) compare to that in countries that have previously undergone rapid financial deepening? Why not see what banking crisis prediction equations say about the vulnerability of these countries to such crises?

- The report does not explain why greater flexibility in labor and capital markets implies a higher "speed limit" (p. 40). I presume that the reason must be that relative price (real exchange rate) adjustments are smoother and less costly.

- The macro effect of overheating (p. 41) is the least of the problems associated with excessively rapid financial market expansion (over-borrowing). I would judge that an even more serious problem may be the misallocation of resources by a poorly regulated financial sector. This possibility is not pursued.

- Why are there such large cross-country differences in the share of foreign-currency-denominated loans among emerging European economies (Figure 29, on p. 41)? This needs to be explained, since there are probably inferences to be drawn here about the quality of financial regulation in these countries.

62. The problem of missing analysis is not confined to the November 2007 report. The April 2008 report (p. 38) offers a growth decomposition for emerging Europe showing that growth has been driven by productivity improvements rather than by accumulation. It would have been more useful, however, to have provided a much more careful interpretation of this outcome than represented by Table 8.[5] The October 2008 report (p. 57) measures the effects of changes in the Financial Conditions Index on real output growth, but leaves the results difficult to interpret. The share of output variability explained by the three financial variables seems too high to be a plausible measure of the effects of exogenous financial shocks on real activity. Instead, it seems more likely that the exercise reflects a combination of transmission through financial variables of more primitive real shocks as well as of exogenous real shocks. The text does not offer this interpretation.

Incoherent analysis

63. There are multiple examples of analysis that I found incoherent, though, as in the APD reports, with no discernible pattern suggesting specific analytical gaps or biases:

- If the exchange rate has truly been playing a larger role in the transmission of monetary policy in emerging Europe, why has that helped these countries to weather external financial market turbulence relatively well?[6] These statements (November 2007 report, p. 7) seem contradictory.

- There is some ambiguity of language concerning monetary policy in the April 2008 report (p. 5), which claims that monetary conditions have been tightened but credit growth remains high. One interpretation would be that extensive foreign exchange market intervention has been undertaken in a context of high capital mobility, rendering tight monetary policy ineffective. Yet the subsequent narrative continually refers to monetary *loosening*, and even uses a Taylor rule (which describes the behavior of a *policy* rate) as a point of reference to judge the monetary stance.

- Why should the fact that global demand factors played an important role in the run-up of commodity prices imply that some emerging economies may need to adopt a tighter-than-usual monetary stance (October 2008 report, p. viii)? Are not these prices

[5] Does the strong total factor productivity (TFP) growth in emerging Europe suggest that the financial sector has actually been relatively efficient in allocating resources there? It would have been useful to tie the TFP findings to the extensive analysis of the financial sector elsewhere in the European *REOs*. It is also useful to note that productive capacity has grown faster in the nontraded than in the traded goods sector. But why has this been so, in the absence of real appreciation to drive resources to that sector?

[6] It is not clear to me why the particular channels through which monetary transmission works should have anything to do with how a country weathers external shocks. I suspect that the reason has to do with the advantages of exchange rate flexibility, rather than with the channels of monetary transmission. But if this is true, why have peggers been less affected than floaters?

exogenous to these economies? Why should the reasons for their increase affect how these economies should respond? This makes no sense if the report is looking at this issue from the perspective of the individual country, rather than that of the global economy.

- The October 2008 report (p. 31) states that energy price increases would tend to slow the growth rate of potential output in Europe, over some time horizon. This is undoubtedly true, but given the range of uncertainty that surrounds estimates of potential output, the estimated effect of the energy price increase on the *level* of potential output is so small that it seems hard to draw particular operational consequences for the conduct of monetary policy. The estimate seems like a flimsy basis for drawing such consequences, and makes the inferences about monetary policy on this page seem rather artificial.

- The methodology described in the October 2008 report (p. 39) for gauging the contribution of credit to growth in European economies—that is, by looking at household and corporate financial deficits—seems too crude to be useful. By definition, is this deficit not equal simply to the sum of the government's and the rest of the world's financial surpluses? What do we learn from splitting up financial deficits in this way?

- The October 2008 report takes some trouble to debunk the view that labor emigration produces overheating (see, for example, p. 76). Is this a paper tiger? If this view has serious proponents it would have been useful to document it and explain its analytical justification. Is it simply based on a contraction of potential output as the result of a contraction in the size of the labor force? How large can this effect be? And are there no offsetting demand-side effects?

- The description of the bank lending channel in the May 2009 report (Box 1, on p. 9) is confusing. It claims the supply of credit is changed "not only by changing interest rates but also by changing the supply of base money." Does the report consider these to be independent?[7]

Incorrect analysis

64. Unlike the instances of analysis that I found incoherent, many of the instances of incorrect analysis seem to have a systematic component, apparently being influenced by the direction in which the reports seem to be trying to influence policy.

---

[7] This said, I actually found the box very useful, documenting as it does the dislocations of the monetary transmission mechanism in Europe caused by the financial crisis.

65. As already indicated, the EUR reports show a particular orientation toward tight macroeconomic policies. Their analysis is sometimes led astray by the advocacy of such policies. For example, the April 2008 report (p. 16) advises emerging European economies with flexible exchange rates to adopt tight monetary policy to prevent wage growth that would erode competitiveness. But of course, the real exchange rate appreciation that would result from monetary tightening would itself produce just such a loss of competitiveness. This contradiction is not acknowledged. Page 54 of the same report uses a cyclically-balanced budget as a point of reference for fiscal policy. But this choice seems completely arbitrary. As acknowledged by the Sustainability and Growth Pact, sustainable deficits do not have to be zero over the cycle, and there may be good (optimality) reasons for a government to run deficits or surpluses over the cycle.

66. A particular concern with low inflation and wage restraint also produces some analytical glitches. There is some odd analysis in the October 2008 report (p. 30), for example, of the effects of an increase in energy prices on factor remuneration. An increase in energy prices is perceived as reducing the marginal product of labor, thus requiring a reduction in the real wage (correct), but not in the real return to capital (odd). To keep the profit rate constant, firms are advised to increase productivity (it is not clear why they would not have already done that, if they could), but somehow this increase in productivity does not feed back to increase real wages. Wage restraint is the clear message here. On p. 79 of the same report, wage inflation is said to be a consequence of international labor mobility—presumably in source rather than host countries. But it is not wage inflation (an increase in the nominal wage) that should be associated with emigration, but rather an increase in the real wage. This may or may not be associated with nominal wage inflation. Again, a concern with wage inflation distorts the analysis. On the same page, it is claimed that emigration limits the scope for keeping wage growth in line with productivity growth. Why should this be? If the domestic labor force shrinks, that would indeed be expected to increase the real wage, but it should increase labor productivity commensurately.

67. In general, the EUR reports see inflationary dangers too often. As one example, based on a calibrated model the October 2008 report (p. 32) concludes that rising energy prices driven by improved productivity in emerging economies should trigger higher inflation in Europe. But this relationship is not at all clear ex ante, since European import prices should fall in this case. I realize that the calibration could indeed produce the claimed result, but one would then want to know how sensitive the outcome is to the specific calibration. The text does not hint at possible alternative outcomes, so the resulting conclusion about monetary policy is not very convincing.

68. Not all the apparent errors in analysis stem from what I perceive as an excessive concern with the policy bottom line. For example, some of the statements in Box 9 of the April 2008 report (p. 48) are rather puzzling: since emerging Europe's current account deficits are significantly larger than those predicted by the empirical model used in this box (at least, according to the figure included in the box), in what sense can the report interpret these deficits as "justified by fundamentals?" Also, if the deficits are larger than predicted by the model, how can they be explained on the basis of demographics and net foreign asset positions?

V. Middle East and Central Asia

69. I read nine *Regional Economic Outlook* reports prepared by the Middle East and Central Asia Department (MCD), covering the period from October 2005 to October 2009.

A. Structure

70. Before 2009, the MCD reports had a different structure from those of the other regions. They tended to consist of a single chapter—containing sections on recent developments, the economic outlook and risks, and policy issues—supplemented by a lengthy statistical appendix. The analytical country classifications were based on the role of oil, with countries classified as oil-exporting or oil-importing.

71. New sub-regions were introduced in thE May 2009 report, presumably to facilitate analysis of the effects of the crisis, and subsequent MCD reports have subdivided countries into Middle Eastern oil exporters, Middle Eastern oil importers, and the Caucasus and Central Asia, and devoted one chapter to each group.

B. Analytical Perspective

72. As is true for other regions, the MCD reports leave unclear where the baseline scenario and outlook come from. The narrative in the early reports has an up-and-down character without much apparent coherence, and much of what is described is country-specific. While the narrative description in the "recent developments and outlook" section improves in later reports, the same failure to describe the implicit regional model that characterized the APD and EUR reports also weakens the analytical clarity of the MCD reports. Everything seems to be driven by oil prices and commodity prices, with some mention of remittances but little mention of other types of trade (say, with Russia and China). Developments in regional stock markets are described, but without an explanation of the role they play in these economies (see, e.g., the October 2007 report, p. 18). Frequent reference is made to capital inflows, but with little indication of the degrees to which different economies in the region are linked with international financial markets. Finally, while some countries are classified as emerging markets, this seems to make little difference in the way they are analyzed in the reports.

73. The implicit structure of the regional model becomes more apparent in later reports. The October 2008 report (p. 14), for example, notes that foreign direct investment and remittances from the Gulf Cooperation Council (GCC) countries have helped sustain demand in the oil-importing countries of the region. This kind of observation provides a useful insight into regional economic interactions, which are not as evident in these reports as a reader

might like.[8] Similarly, the May 2009 report (Box 3) gives a very useful summary of the economic relationships between the GCC and the Middle Eastern oil importers. And, for the first time in these reports, p. 26 of the May 2009 report emphasizes the important impact that developments in Russia have on the Caucasus and Central Asia (CCA). The October 2009 report contains a useful box (Box 3.2) on the role of remittances from Russia in crisis transmission. Why were these regional interactions not covered in prior reports?

74. The MCD reports frequently advocate fiscal expansion rather than real appreciation as a tool to reduce current account surpluses in oil exporting countries (e.g., the October 2007 report, p. 28). But as in the APD reports, this advocacy takes too much of a "global imbalances" perspective. Whether these countries (many of which are very poor) should try to reduce surpluses or not should depend on the investment opportunities they have at home, not on what they can contribute to reducing global imbalances.

C. Boxes

75. By and large, the MCD reports contain appropriate boxes. The May 2006 report, for example, contains boxes on the disposition of oil revenue in oil-exporting countries, the pass-through of higher oil prices in the oil importers, and developments in regional equity markets, which boomed at the time. More so than those of other regions, these reports also contain informative boxes on recent developments in specific countries. For example, the October 2007 report contains boxes on Kazakhstan (Box 4), and Armenia (Box 5). The inclusion of boxes with country capsules illustrating particular development strategies or policy frameworks is an excellent idea.

76. All is not perfect, however. For example, in the May 2008 report, which singles out Kazakhstan as the only country directly affected through its financial sector by the international financial crisis, it would have been appropriate to have included a box on events in that country, but none is provided.

D. Tone

77. Like the *REOs* produced by other area departments, the MCD reports are highly prescriptive. Box 1.1 of the October 2007 report, on the Maghreb countries, provides a good illustration.

[8] This report is notable for citing research—a practice less common in the MCD reports than in the *REOs* of APD and EUR.

E. Links to the Professional Literature

78. The MCD reports stand out from the others for their relatively scarce references to the professional literature. There are many places where such references would have been appropriate. For example:

- In the September 2005 report, the presentation in Box 3.1 gives a misleading impression of rigor. Estimating real exchange rate undervaluation for the CCA countries based on regressions of the price level on per capita incomes is both desirable—because it imposes some discipline on judgments about undervaluation—and undesirable, because it is a very crude and approximate procedure. The literature offers much more sophisticated ways to estimate misalignment.

- In the September 2006 report, Box 3.1 on remittances is an important addition, since remittances are very important for many countries in the region. But the box makes a number of assertions about the macroeconomic effects of remittances that are not backed up by evidence or references to the rather large professional literature on this subject.

- In the May 2008 report, Box 7 on the factors driving high oil prices (fundamentals or speculation) is very appropriate, since this issue was hotly debated at the time. But the box makes no reference to outside work on the topic. By comparing the price of oil to that of gold, it concludes that speculation contributed to the run-up in oil prices in 2007. It does not consider other types of evidence—for example, evidence of oil hoarding—that might bear on this question.

F. Analytical Strengths

79. By and large, I found the analysis in these reports very reasonable, and more flexible in drawing implications for policy than reports prepared by the EUR, for example. While appropriate analytical tools are not always used (see below), I found many positive features.

For example:

- The executive summary of the May 2008 report mentions the possible need to live with inflation, at least for a while, in the oil-exporting countries with dollar pegs; the need to make an exchange rate regime choice for the proposed GCC monetary union; and the urgent need to overhaul educational systems in the region.

- The October 2009 report (p. 19) usefully observes that the limited engagement of Middle Eastern oil importers in advanced manufacturing has helped insulate them from the crisis.

- To assess the appropriateness of the fiscal response to the oil boom in oil-exporting countries, it is correct to use the non-oil fiscal stance, rather than the overall fiscal outcome (September 2005 report, p. 20), but using the non-oil fiscal impulse would have been better. The October 2009 report (p. 24) uses fiscal impulses, rather than actual fiscal deficits, to assess changes in the fiscal policy stance (specifically, fiscal responses to the crisis). In the same report, I particularly liked the decomposition of the change in the primary deficit into the fiscal impulse and the effect of automatic stabilizers in Figure 3.6 (p. 35).

- Table 1 of the May 2009 report (p. 8) gives a very useful summary of the policy response of Middle Eastern oil exporters to the crisis, and this is repeated for the other groups.

G. Analytical Weaknesses

Vagueness

80. As in the reports of APD and EUR reviewed above, vagueness in some of the policy prescriptions in the MCD reports sometimes reduces these prescriptions to the level of platitudes. For example, in the September 2005 report (p. 2), the advocacy of spending in oil-exporting countries to reduce global imbalances is appropriately qualified to refer only to "quality" spending but without, unfortunately, suggesting how to recognize "quality" spending or how much "quality" spending there is to be done. Similarly, in the same report, advocacy of exchange rate appreciation to contain inflation is appropriately qualified with the need to promote productivity to preserve competitiveness, but again without indicating how, how quickly, or how effectively this can be done. The October 2008 report makes continued calls for structural reforms that could boost competitiveness in the non-oil sectors. But what exactly are these reforms? Overall, the policy advice at the end of the October 2008 report is vague and somewhat textbookish. More concreteness would have been useful here.

81. In other cases, the language leaves unclear what is exogenous and what is endogenous. For example, p. 10 of the October 2008 report contains the statement: "wages have been increased in more than half of MCD countries… ." This seems to suggest an exogenous policy decision. But what does it refer to exactly? To public sector wages? To minimum wages? Other wages would presumably be endogenous. Only later does the report make clear that the reference is indeed to wages that are set by policy.

Unsubstantiated claims

82. Some of the unsubstantiated claims in the MCD reports seem to reflect a particular policy bias or perspective. Among the instances that seem systematic to me are those that clearly reflect the Fund's view on certain issues. The September 2005 report (p. 36), for example, claims that the virtues of the Fund's preference for nominal appreciation to combat inflation in the face of capital inflows are well documented, but it neither explains nor

documents them. In the same vein, the October 2009 report (p. 27) states that "creating an environment more conducive to private business" should be the key to reducing unemployment, but it leaves unclear why this should be so. Where is the research that backs up this statement?

83. In other cases, the deficiencies in this regard appear to be more innocent:

- The September 2005 report (p. 38) claims that remittances are priced in local currency, but gives no supporting references.

- Box 1 of the October 2008 report contains very informative maps on the contributions of food and fuel prices to headline inflation. It claims that higher food and fuel prices feed through into core inflation, depending on a country's inflationary history. While this may be true, no evidence is presented and no references are given.

- The October 2009 report (p. 36) claims that there is extensive financial dollarization in the CCA countries, but provides no data to support this.

Missing analysis

84. Unfortunately some of the missing analysis in these reports concerns issues that are central to the region: poverty, income distribution, unemployment, and domestic debt markets. For example, though the earlier MCD reports repeatedly point to the persistence of poverty in the face of rapid growth in the region, not until the May 2007 report (p. 27) is it acknowledged that a failure to reduce poverty even in the face of high growth means that income distribution needs to be improved in these countries. This issue deserves much more prominence than it receives in these reports. There is considerable and appropriate concern in the MCD reports with unemployment in the region (e.g., the end of Chapter 1 in the September 2005 report), but unemployment is consistently interpreted as a growth problem. In fact, the analysis needs to recognize that a substantial part of the unemployment in the region may be structural, requiring microeconomic rather than macroeconomic policies to address it. The September 2006 report (p. 19) refers to the fact that rapid growth has not made a large dent in unemployment, not acknowledging the possibility that the natural rate of unemployment is high in many MCD countries. There is also a recurrent theme concerning the need for the Middle Eastern oil exporters to develop domestic debt markets, as well as some discussion of why this may be useful (see, for example, the concluding suggestion on p. 16 of the October 2009 report), but no identification of what constraints might limit the development of these markets.

85. Aside from such region-specific issues, analysis is also missing on more traditional fiscal, exchange rate, and financial sector policies that are the Fund's bread and butter. Regarding fiscal matters, a recurring theme in all of the MCD reports (see, for example the September 2005 report, p. 10) is that of fiscal space. Though the May 2009 report makes frequent (and appropriate) reference to the role of fiscal space in determining the fiscal

response to the downturn, it gives no indication of how the availability of fiscal space can be determined, except at the very end of the October 2009 report (and see below for a criticism of that analysis). Similarly, the advocacy of fiscal measures to put debt/GDP ratios on a declining path in several countries (p. 29 of the October 2007 report) does not seem to be based on any notion of what a desirable debt/GDP ratio should be in those countries. And while it is appropriate to point out that debt reduction achieved through privatization *may* not improve the public sector's net worth (September 2005 report, p. 10), it would have been useful to explain the conditions under which it *would* do so and to have taken a position on whether it *has* in fact done so in the countries under review.

86. Turning to exchange rate policy, the advocacy in the October 2007 report (p. 29) of real appreciation in the CCA countries, to control inflation and avoid its harmful effects on growth, does not factor in the potential deleterious effects of real appreciation on growth. This issue is more complicated than the report supposes.

87. Analysis is also missing with regard to some financial sector issues. For example, the use of a financial stress indicator in Box 2.1 of the October 2009 report is informative, but the box does not explain what statistical procedure permits it to conclude that financial sector stress had a small effect on growth in the Middle East, North Africa, Afghanistan and Pakistan region. This issue is too important in the current context to leave the analysis incomplete.

Incoherent analysis

88. Like the cases of missing analysis, instances of incoherent analysis involve some aspects that are region-specific and others that involve areas of particular Fund competence. A recurrent region-specific issue concerns the role of so-called "flexible" labor markets in contributing to price stability in the context of a large expansionary shock in aggregate demand in oil-exporting countries, such as one arising from an oil boom. This is brought up, for example, in the September 2005 report (p. 23), and there are other references to "flexible labor markets" keeping inflation under control in the May 2006 report (p. 14) and the September 2006 report (p. 11). But why should we expect a "flexible" labor market to imply less inflation in response to an expansionary aggregate demand shock? The issue is, of course, what precisely is meant by a "flexible" labor market. I can understand the effect of trade openness in limiting inflation (under a fixed exchange rate), but the "flexible" labor market argument has to be based on labor supply elasticity rather than on wage flexibility or intersectoral labor mobility. This distinction is finally made on p. 15 of the September 2006 report, with a reference to "open product and factor market." Should this be interpreted to mean an elastic supply of labor because of migrant workers? Not until Box 2 of the May 2007 report is a clear explanation finally presented of why the real exchange rate in oil-exporting countries has not appreciated as one might have expected, and the answer is indeed not labor market "flexibility" as such, but rather labor market *openness* (i.e., the role of migrant labor).

89. There are also problems of incoherence regarding exchange rate and monetary policies. It is hard to see what real exchange rate appreciation means in the oil-importing countries—for example, whether it simply reflects an increase in inflation as the result of pass-through of higher imported oil prices (October 2007 report, p. 17). If oil prices are rising throughout the world, the direct impacts on national price levels will depend simply on the extent to which these increases are passed through to the domestic economy. There also seems to be general confusion in these reports about monetary policy in countries with managed exchange rates.

90. Regarding monetary policy, there is a contradiction in the October 2008 report (p. 16), where it is stated that oil-exporting countries with fixed exchange rates have had to follow the expansionary monetary policies of the United States at a time when their cyclical conditions call for the opposite, but also that they have undertaken open market operations to mop up liquidity. If these countries truly have no monetary autonomy, and need to follow U.S. monetary policy, they cannot effectively "mop up" liquidity. And on p. 30 of the same report it is claimed that raising interest rates in such countries would attract capital inflows, which fuel inflationary pressures. But if interest rates are successfully increased, that should reduce aggregate demand whether more capital inflows are attracted or not. Finally, an analysis of the factors contributing to the recent slowdown in credit growth in the region (October 2009 report, p. 8) is not very informative; it is basically an accounting exercise on banks' balance sheets without providing any way to tell what is driving what.

Incorrect analysis

91. Examples of weak or incorrect analysis in the MCD reports include the following:

- The region's reports typically express reserve adequacy in terms of months of import cover. This indicator is anachronistic, except perhaps for countries with no access to international capital markets. More useful in countries that are more closely linked with international capital markets are the ratios of reserves to short-term external debt or to some domestic monetary aggregate.

- The suggestion that CCA countries should move to more flexible exchange rate regimes to improve competitiveness (October 2009 report, p. 29) only works when capital is tending to flow out. When it is tending to flow in, the effect on competitiveness is just the opposite.

- The appendix on fiscal space in the October 2009 report is weak, defining fiscal space as existing only when public debt and interest rates on public debt are low, without specifying what "low" means or how these two components can be traded off for each other in assessing fiscal space. The rationale offered in the appendix for taking debt interest rates into account in this calculation is incorrect.

VI. Sub-Saharan Africa

92. The African Department (AFR) produced the earliest regional economic reports, beginning in September 2003. I read 13 *Regional Economic Outlook* reports prepared by AFR on Sub-Saharan Africa (SSA), covering the period from October 2003 to October 2009.

A. Structure

93. The structure of the SSA reports has changed dramatically over time, revealing the experimental nature of the earlier reports. The initial reports consisted of a foreword, an overview, a review of recent economic developments, a section on economic policies, and a section on prospects for the region. Later reports are more similar in structure to their contemporaries from other regions, featuring an overview chapter and usually two or three analytical chapters. Like the MCD reports, the SSA reports have consistently contained lengthy statistical appendices. Such appendices seem more useful for this region than most others, since macro data for countries in SSA are less likely to be familiar to most readers from outside the region.

94. SSA reports' analytical chapters are not always well motivated. The May 2006 report, for example, contains chapters on aid, financial sector development, and decentralization. The latter two topics did not receive much attention in earlier reports from the region, so it is unclear why they merit entire chapters in May 2006. Since these issues did not suddenly become important in 2006, they either did not receive sufficient attention in earlier reports or they receive excessive attention in the May 2006 report. I believe that these issues have been of long-standing importance for the region, and should have been more prominent in earlier reports. The point to be made here, though, is that these analytical chapters are not well connected to the ongoing narrative for the region.

95. The audience for this region's reports seems to be a little different from that of the reports from other regions, in that it includes in a very self-conscious way the international community at large (presumably policymakers, aid organizations, and civil society in industrial countries). These reports quite often call for policy action by industrial countries on behalf of countries in SSA (e.g., trade liberalization, the removal of agricultural subsidies, and additional donor assistance in fulfilling the 2005 Gleneagles accords).

B. Analytical Perspective

96. Like the *REOs* reviewed so far, the SSA reports do not explicitly provide a view of how their region functions as an economic unit—that is, of an implicit regional economic model. But Box 2.3 in the September 2006 report, on South Africa's economic links with the rest of SSA, is very appropriate and should have been in earlier reports, to give the reader some indication of the extent of intraregional economic relations, at least in Southern Africa.

97. Like the reports on other regions, the early SSA reports contain little systematic institutional or structural background on the countries in the region (e.g., on their exchange rate regimes, the nature of their links with international capital markets, the composition of their exports, their main trading partners, and so on) or on the most salient aspects of their current economic situation. For example, it would have been beneficial to inform the reader much earlier than on p. 20 of the April 2008 report that real per capita income in SSA now is about the same as in the mid-1970s, and that the poverty rate is 41 percent! The writing in the early reports is monotone and tedious, with too much emphasis on and discussion of international agreements and organizations, and the use of many acronyms.

98. The initial SSA reports also contain no analytical country groupings, but instead a hodgepodge of country-specific results. A new classification of countries was introduced in the September 2006 report: countries are classified as resource-rich (oil exporters and importers) and non-resource-rich (coastal and landlocked). This became the standard classification for most subsequent reports, and seems useful.[9]

99. A surprising aspect of the SSA reports, given the number of LICs in this region, is that little attention is paid to the development strategies being pursued—or that should be pursued—by these countries. The April 2007 report takes up the issue of development strategy for the first time, arguing that removing structural impediments to value-added industries linked to agriculture and commodities is the key to stimulating growth (p. 2).

C. Boxes

100. The first SSA report (September 2003) contains very few boxes, and its graphs are very crude. Such boxes as are included in the initial report are so vague as not to be very useful; that on p. 24, for example, is mainly anecdotal. Matters improve in successive reports, but gradually. The May 2004 report contains several boxes, but these and the figures are very poorly integrated with the text and many are deficient. Box 2, for example (p. 7), on causes and remedies for civil wars, is suitable for inclusion given the SSA context as well as the extensive amount of outside research on these issues for SSA, and it appropriately discusses some of the relevant literature, but the "findings" it discusses are primarily platitudes. Box 3 (on famine and drought) is also appropriate, but would have been more useful had it provided some evidence on incidence, severity, and impacts.

101. As is true for other regions, there are boxes that suggest themselves but are not included in the early reports. For example, the May 2004 report (p. 35) cites Benin, Burkina Faso, Cape Verde, Mozambique, Tanzania, and Uganda as countries that have implemented the "sound policies and structural and institutional reforms" that the report repeatedly calls

---

[9] The October 2009 report (p. 7), however, describes four groupings as "traditional" for the Sub-Saharan Africa *REOs*: oil exporters, MICs, LICs, and fragile countries.

for. Why not include extended boxes on these countries explaining what exactly they did, and how? This would offer an opportunity to take the policy recommendations beyond platitudes. In the September 2006 report, it would have been useful to add a box on how Madagascar managed to sustain its textile exports to the EU after textile quotas were phased out, in contrast to other SSA countries (p. 6). A box discussing Zimbabwe and its impact on the region is conspicuously absent from these reports until October 2007, and even that box contains no discussion of spillovers to other countries in Southern Africa. I presume the topic must be politically sensitive. Another opportunity is missed in Chapter 3 of the April 2008 report (on private capital flows to SSA): given that Uganda seems to have been the only SSA country that liberalized its capital account at one go, it would have been useful to have had a box distilling lessons from its experience.

D. Tone

102. There is little to distinguish the tone of the SSA reports from that of the reports prepared by other departments. Like the others, the SSA reports are highly prescriptive. Overall, the policy orientation of the SSA reports is more similar to those of APD and MCD than to those of EUR, in that they are less committed to a particular policy stance and are relatively eclectic.

E. Links to the Professional Literature

103. Like the MCD reports, the early SSA reports make very few references to the professional literature. Instead they place heavy reliance on ad hoc and crude made-to-suit evidence as support for analytical or empirical propositions. Such evidence takes the form of country classifications and simple correlations (such as comparing countries with and without Fund programs, or those with below- and above-median values on some indicator) to draw inferences, such as about the role of public investment in growth.

104. Reliance on existing research has increased over time in SSA reports. While the September 2003 report, for example, contains almost no such references (an exception is the reference to research on exchange rate pass-through in South Africa on p. 10), later reports make extensive use of the professional literature. For instance, though the chapter on aid in the May 2006 report is rather textbookish and mechanical, it does discuss some of the more prominent professional literature on the topic, and the September 2006 report (Box 2.5) contains a good box on the role of remittances in SSA, with literature references.

F. Analytical Strengths

105. I found some of the analytical chapters in the later reports to be excellent. Chapter 4 of the May 2006 report, on financial development in SSA, is an example: its analysis is sound and its policy recommendations are concrete (e.g., leasing to overcome problems with collateralization, harmonization of the financial institutional environment within monetary unions to allow for economies of scale) and reasonable. Chapter 5 of the same report, on

fiscal decentralization, gives a critical but balanced evaluation of experience to date and makes frequent reference to existing research on these issues in the region.

106. Overall, the April 2008 report is a very good one. Chapter 2 of that report (on monetary frameworks in SSA) is insightful and very useful: p. 34, for example, contains a creative and convincing argument for money targeting during the transition away from fiscal dominance, using money in a "tripwire" fashion to indicate that the stabilization is on track, and p. 35 presents a good argument for implementing some aspects of inflation targeting in countries that are already implementing a flexible monetary framework; the section containing that analysis is well thought out. Similarly, Boxes 2.4-2.6 of the same report, describing the monetary frameworks in various SSA countries, are excellent. The summary appendix table on capital account liberalization in the case study countries (p. 69) is very informative and useful. The chapter on power generation is appropriate, well-written, and thoughtful.

107. The October 2008 report, too, contains good material. On p. 23, for example, the report observes that the recent positive (pre-crisis) SSA growth experience—which was widespread across very different countries—suggests that there is little evidence of a poverty trap. This point is worth making as a counter to development pessimism, because it suggests hope for all countries. Chapter 2, on growth successes, is clearly written and well reasoned; though not analytically ambitious, it is a useful description of how much things have changed in SSA.

108. Other examples of solid analysis include:

- The use of the Global Trade Analysis Project in the September 2006 report to simulate the impact on GDP of higher oil prices.

- The appropriate acknowledgement in the September 2006 report that first-best policies in ameliorating the impact of higher oil prices on the poor are not feasible, and the proposing (p. 26) of a very reasonable set of second-best policies.

- The 2009 reports offer solid analysis of responses to the international crisis. The "four principles" listed in the April report (p. 19) to guide policy responses to the crisis are sensible and balanced, and the priority given in Box 1.5 (p. 21) to spending rather than tax cuts as countercyclical policy, and to spending on labor-intensive infrastructure projects with a low import content, is well placed. The demonstration of fiscal procyclicality in SSA (October 2009 report, p. 36), using the methodology of Kaminsky, Reinhart, and Vegh (2004), usefully illustrates the contrast with the past in SSA's current countercyclical stance, showing the virtues of fiscal prudence during good times.

G. Analytical Weaknesses

Vagueness

109. A consistent problem with the SSA regional reports is vagueness with respect to policy advice. The problem is not universal in these reports, which make many concrete suggestions, but it remains true from September 2003 to October 2009 that many good things are advocated in general terms without being translated into concrete policy advice. The September 2003 report (p. 24), for example, advocates labor market reform in South Africa and structural reforms in Nigeria, without giving any specifics. There are many other examples:

- The advice in the April 2007 report (p. 11) to deal with Dutch disease by "close coordination of fiscal and monetary policies, raising productivity and strengthening the supply side of their economies by promoting private sector activity and liberalizing their trade regimes" is (except for trade liberalization) meaningless. Similarly, Box 2.5 of the same report, on promoting private sector development, is preachy and vague.

- Chapter 5 of the April 2007 report, on domestic debt markets, is timely but contains policy advice that is generally vague. Country-specific diagnosis of the reasons why domestic debt markets have failed to develop sufficiently is notably absent. Boxes providing case studies of countries where such markets are more extensive (Kenya and Mauritius) would have been very useful.

- In the October 2007 report (Box 1.2, on p. 12), the use of the Jeanne-Ranciere model to assess reserve adequacy is commendable, but the discussion of the results is too vague and general. More specific results would have been much more informative. If these could not be published for sensitivity reasons, it would have been better not to mention them at all.

- Many of the SSA reports (including the October 2007 report, p. 17) refer frequently to the need for coordinating fiscal and monetary policies, but without much indication of what this means. Similarly, the frequent exhortations (e.g., p. 34 of the same report) to increase the productivity of traded goods sectors to mitigate the effects of real appreciation are all well and good, but how can this be done?

- In the April 2008 report (p. 10), what are the "distortions in monetary and fiscal policy that discourage bank lending?"

- Another example of vague policy advice appears in the April 2008 report (p. 20): "in some countries, exchange rate changes may help rebalance growth." Which countries? What types of exchange rate changes? Does "rebalance" mean stimulate?

- Even somewhat idiosyncratic aspects of the regional policy experience (such as the prevalence of monetary targeting) are not always explored in enough depth. The discussion of monetary targeting on p. 26 of the April 2008 report, for example, really gives no guidance as to when targeting money may be appropriate.

- Occasionally, too little information is given to enable the reader to evaluate the analysis. For example, in the April 2009 report (Box 1.4, on p. 18) the usefulness of simulations such as those reported on the effects of the global slowdown on SSA is limited without more information on how those simulations were constructed. Another example is Box 2.1 in the October 2009 report (p. 42), which is not very informative about how the debt sustainability analysis for SSA was performed.

Unsubstantiated claims

110. There are many unsubstantiated claims in the SSA reports, as in the reports of other regions. At times this seems to be explained by the authors' acceptance of conventional wisdom inside the Fund. For example, the SSA region achieved a dramatic and widespread acceleration in growth for several years before the recent crisis. The Sub-Saharan Africa *REOs* frequently attribute growth accelerations to good policy frameworks, as is done in the May 2004 report (p. 4). But how do we know? No evidence is given in that report, and this problem of attribution turns out to be a recurring issue. For example, in Box 1.5 of the October 2008 report, on what basis is it judged that growth in agricultural production from 1980–89 to 2000–05 in SSA was due to reduced conflict, greater economic stability, and reduced taxation? While plausible, this interpretation is not substantiated. Finally, where is the evidence for the statement in the April 2009 report (p. 23) that capital controls have been ineffective in SSA LICs? The May 2006 report (p. 49) states that "it already appears that past problems [with development banks] are likely to reemerge." But this statement is not expanded upon. What are the indicators? How do we know?

111. At other times, evidence is provided but is much weaker than it could be. The May 2006 report (p. 21), for example, contains the odd statement that because sustained-growth countries in SSA have avoided overvalued exchange rates through macro stability, countries that receive aid must manage their economies prudently. I think the author is trying to say that real appreciation undermines growth, so aid-receiving countries need to try to avoid Dutch disease. This is reasonable advice, but it does not have to be based on the particular correlation mentioned and on a very crude notion of overvaluation. There is plenty of professional literature on this issue that could have been cited here, and would make the point much more strongly.

112. Lack of substantiation also emerges even when the professional literature is cited, since those citations are not always offered in the most convincing way—that is, in a way that convinces the reader that what is being reported is something close to a professional consensus. On p. 4 of the October 2005 report, for instance, only one source is cited on

behalf of the proposition that increased spending on infrastructure is good for growth. But is there any contrary evidence? Why not cite the findings of a survey of the literature, rather than of an individual paper? There must be several such surveys. By contrast, the reference in the October 2009 report (p. 46) to surveys such as Nijkamp and Poot (2004) on productive public spending is very appropriate.

Missing analysis

113. Some of the policy advice provided in these reports is not backed by economic reasoning. For example, the May 2006 report (p. 14) prescribes tighter fiscal and monetary policies for fixed-exchange-rate oil-importing countries that experience an oil price increase. Why is this necessarily the right policy stance? What is the implicit objective that would be achieved with such policies? Is it just price-level stabilization? There is no unambiguous general prescription for such policies in response to an oil price shock, and no justification is given here.

114. In several instances, statements in one regional report either contradict those in earlier reports or seem to run counter to policy prescriptions in earlier reports, but the inconsistency is not acknowledged or explained. For example, the October 2005 report (p. 16) discusses a World Bank study finding that reductions in distortions in agricultural trade in OECD countries would not yield large benefits for Africa. This runs counter to claims made in earlier regional reports about the importance for Africa of advanced-country liberalization of agricultural trade, but the issue is not discussed in more depth. In another example, the April 2009 report (Box 2.6 and the material surrounding it, p. 39) seems to suggest that industrialization is a necessary concomitant of sustained fast growth. Whether this is right or wrong, it conflicts with the views expressed in an earlier report, which referred to the counterexamples of Argentina and New Zealand as countries that have grown successfully without industrializing. What is the department's view on this issue?

115. Other examples of missing analysis include:

- In the April 2007 report (p. 32), the simulation model for Nigeria (essentially a financial programming model) seems far too simple to tackle the question posed (the macro effects of a large scaling up of public spending). A more sophisticated framework is required for addressing such a question.

- Chapter 2 of the October 2007 report, on case studies in creating fiscal space for productive spending, is very appropriate. However, it suffers from lack of detail, such as on how these countries managed to increase revenue and improve the efficiency of spending. For example, how much did opening the large taxpayer office in Rwanda (p. 32) contribute to the large increase in tax revenue there? The analysis of tax reforms in Tanzania (p. 34) is much better.

- The October 2008 report (p. 11) refers to the absence of traded-nontraded price indexes and the consequent need to use CPI-based real exchange rate measures. But real exchange rate measures using foreign producer price indexes and the domestic CPI are also an option, and because they typically cover traded goods, have the advantage of not being contaminated by changes in relative prices of foreign traded-nontraded goods.

- Box 1.1 of the same report is useful and informative, but it begs the question of why—if world food prices have had a relatively small impact on domestic food prices—headline inflation in the East African Community (EAC) increased from 8 percent in 2007 to 20 percent in mid-2008.

- The April 2009 report contains several pieces of analysis that could usefully have been extended. In Box 1.1, for example, if the financial channel is indeed growing, and if this is increasing the impact of world growth on SSA's growth, these trends could have been picked up in the regressions by allowing the coefficient of partner-country growth to change over time. Also, it would have been useful to have provided measures of the fiscal impulse in response to the global growth slowdown.

Incoherent analysis

116. Several examples of analysis that I found incoherent may reflect the *REO* authors' uncritical acceptance of conventional wisdom inside the institution, while other examples appear to be less easily explainable.

117. Cases where conventional wisdom may have stood in the way of clear analysis include the following:

- In the April 2007 report (p. 18), the claim that domestic absorption can be raised by liberalizing trade (a policy frequently advocated by the Fund) is unsupported. While trade liberalization may affect the composition of absorption, why should it affect total absorption? Does this proposition simply rely on an income effect? If so, how quantitatively important is this effect?

- Box 1.4 of the October 2008 report (p. 19) advises that because the first-round inflation effects of increases in food prices may be difficult to distinguish from the second-round inflation effects, central banks may want to resist even first-round effects. This does not logically follow. At the very least, it places a large weight on inflation relative to output stability.

- I do not understand the statement in the October 2008 report (p. 22) that countries experiencing expansionary aggregate demand pressures should resist a combination of expansionary fiscal policy and loose monetary policy because "it would place excessive pressure on nominal exchange rates to facilitate external adjustment."

> Countries in this situation should avoid this policy package because they are facing excess demand, not because of what happens to the exchange rate.

118. An example of incoherent analysis that does not fit a systematic pattern appears in the October 2009 report (p. 35), which does not adequately address the tension between its advocacy of countercyclical fiscal policy and the reported finding that fiscal multipliers tend to be very small in LICs. The tension might be reconciled by emphasizing, more than is done in the report, spending that is specifically targeted to have larger impacts on both aggregate demand and supply (e.g., infrastructure investment rather than tax cuts).

Incorrect analysis

119. The September 2006 report (p. 24) erroneously refers to remittances as "private capital." Worker remittances are not capital flows, and these reports should not make that mistake.

120. The April 2007 report (footnote 32, on p. 33) makes some questionable statements. First, it claims that sterilization would reduce domestic demand. This is incorrect: it would merely prevent an increase in domestic demand. Second, it claims that selling foreign currency (causing the real exchange rate to appreciate) would "facilitate the resource transfer necessary to stimulate growth." Under what conditions does real appreciation stimulate growth?

121. In the October 2008 report (p. 35), the measurement of overvaluation is too simple and should not be taken seriously, since PPP is not an appropriate benchmark for the equilibrium rate.

122. The October 2009 report (Box 1.1, on p. 14) looks at the fiscal response to the crisis. Because output was affected by the crisis, the authors question whether normalizing fiscal deficits by output is a good idea, and instead look at spending normalized by revenues. But revenues are likely to have been affected even more than output. Why not normalize, say, by de-trended output?

VII. Western Hemisphere

123. I read eight *Regional Economic Outlook* reports prepared by the Western Hemisphere Department (WHD), covering the period from May 2006 to October 2009.

A. Structure

124. Perhaps because of the special characteristics of the regional economy, I found the structure of the Western Hemisphere *REOs* to be analytically more appealing than those of other regions. The WHD reports typically begin with an overview of the world economic situation, to describe the international environment in which regional developments play out.

Since this environment is exogenous to all of the countries in the region except the United States, this makes sense. Next, the reports consider developments in the United States. Because the United States is by far the largest economy in the hemisphere, and dominates economic events in most of the rest of the countries in the region, this is an appropriate second step. The first chapter is completed with an analysis of developments in Canada which, while itself affected by events in the United States, is also largely in the role of influencing rather than being influenced by other individual countries in the region. The second chapter analyzes the situation in Latin America and the Caribbean (LAC), taking account of spillovers from both the United States and Canada, as well as from countries outside the region.

125. In turn, LAC is subdivided into three parts: commodity-exporting countries with inflation targeting (IT) regimes, commodity-exporting countries without IT regimes, and commodity-importing countries. This structure is also appropriate because the largest economies in the region (Argentina, Brazil, Chile, Colombia, Mexico, Peru, and Venezuela, the "LAC-7") all tend to be commodity exporters, and fluctuations in commodity prices thus represent an important link between events in those countries and in the rest of the world, including the United States and Canada. These large commodity exporters are all fairly closely integrated with international financial markets (at least in normal times). Yet the large commodity exporters differ with respect to their overall macroeconomic policy orientation, in ways that are highly relevant to the analysis of domestic macroeconomic developments in those countries. Specifically, the largest subgroup of these countries (which includes Brazil, Chile, Colombia, Mexico, and Peru from the LAC-7, but also Uruguay), has made much more significant progress in reforming its domestic macroeconomic institutions (with floating exchange rates, inflation targeting, and substantial fiscal reforms, including the enactment in several of these countries of fiscal responsibility laws), than the smaller subgroup (including Argentina and Venezuela from among the LAC-7, but also Bolivia, Ecuador, and Paraguay). Accordingly, the former group has achieved more significant fiscal adjustment, lower inflation, and lower sovereign risk premiums than the latter.

126. The more recent Western Hemisphere *REOs* include analytical chapters that are meant to probe more deeply into important policy issues in the region. These chapters usually emerge as natural extensions of the two overview chapters since they address issues that arise naturally from those chapters (though, like some of those in other regions, a few of them appear somewhat forced). As an example of the former, the May 2009 report has a particularly sensible structure, with a chapter on the United States and Canada, a chapter on LAC with very relevant annexes, a chapter on banking systems in LAC (which was very appropriate given crisis-related developments at the time) and a chapter on role of foreign banks in transmission of financial shocks, which was again very appropriate in light of the origin of the crisis in the U.S. financial system.

B. Analytical Perspective

127. The analytical approaches adopted in the Western Hemisphere *REOs* appear broadly appropriate and up to date. In my view, the use of a coherent implicit model of how the region works as an economic unit helped the staff get some aspects of the economic outlook right in these reports. For example, the April 2006 *REO* emphasizes the risks posed to the region by a potential deterioration in the U.S. housing market. This turned out to be very prescient.

C. Boxes

128. As I have mentioned before, in my view boxes should be used to illuminate and/or illustrate issues that arise in the text. The boxes in the WHD documents generally do a good job on this score. Those in the November 2006 report, for instance, seem particularly well chosen.

D. Tone

129. Like the reports from other regions, the WHD reports are quite prescriptive. Rather than just simply analyze recent developments and the regional outlook, they take strong positions on policy priorities as well as on the types of policies that should be undertaken to satisfy those priorities.

E. Links to the Professional Literature

130. The links to the professional literature in the Western Hemisphere *REOs* are more similar to those in the *REOs* produced by APD and EUR than to those produced by MCD and AFR, in the sense that fairly generous references are made to professional research. Overall, I found the reports produced by WHD to be adequately grounded in the professional literature, though with some gaps. For example, the claims made in the April 2008 report (p. 10) about the impact of remittance flows on poverty and inequality are not supported by evidence from existing research.

F. Analytical Strengths

131. These reports occasionally make very perceptive observations. Two examples follow.

132. The April 2008 report (p. vi) cautions against fiscal stimulus in LAC based on "analytical work," because such stimulus would tend to undermine policy credibility. It argues that though discretionary stimulus can be appropriate when people believe that fiscal policy responds symmetrically to the business cycle, the expansionary fiscal policy that some countries adopted during the recent pre-crisis upswing in the LAC region would tend to call that symmetry into question in those countries—suggesting that fiscal expansion is more likely to be interpreted as just another instance of pro-deficit bias, which would create

uncertainty about fiscal solvency. I found this argument to be well reasoned and worth taking seriously.

133. The May 2009 report makes a courageous (and in my view sensible) pessimistic forecast for the U.S. economy. Most impressive is that the pessimism is based not just on past cyclical patterns or past experience with banking crises, as one often finds, but rather on a clear-headed forward-looking assessment of the political constraints that were likely to prove operative in the formulation of policy in the United States. In particular, the argument takes into account the likelihood of a protracted political stalemate on financial rescues and financial reforms in the specifically American context. I found such realism refreshing.

134. In addition to these perceptive observations, another attractive feature of these reports is their reliance on systematic evidence to interpret events. For example, the April 2007 report very usefully employs an empirical model from Osterholm and Zettelmeyer (2007) to explore how growth in Latin America is linked to external variables. The same model is used in subsequent *REOs*. Its application in the October 2009 report (Box 3.1) to document the region's surprising resiliency in the face of the latest crisis is very convincing.

G. Analytical Weaknesses

135. Like the reports prepared by other area departments, the WHD reports are not immune to vagueness and unsubstantiated claims, or to missing, incoherent, and incorrect analysis. To avoid repeating some of the commentary provided for the *REOs* prepared by other departments, this section simply describes how Western Hemisphere *REOs* do or do not differ from those prepared by other departments, and provides examples of each of the analytical deficiencies listed above.

Vagueness

- Like those of other regions, the Western Hemisphere *REOs* do not describe how their "baseline scenario" was generated.

- Like those of other regions, WHD reports occasionally fall into the financial reporting trap of explaining important events as the result of rather vague causes. For example, there must be more to say about the significant projected slowdown in U.S. growth in 2008 than that it is due to a "correction" in the housing market and the "effects of financial turmoil on confidence" (November 2007 report, p. 1). Does the "financial turmoil" not have any effects on the real economy on its own, but only through confidence?

- What is meant by the statement that "monetary policy is close to neutral" (November 2006 report, p. 14)?

Unsubstantiated claims

- The April 2006 report (p. 7) attributes an acceleration in credit growth in Brazil and Mexico to improved credit market infrastructure, but provides no supporting evidence. How do the authors know that this is so?

- The November 2006 report (p. 27) refers to a deficient allocation of public investment in LAC, but provides no supporting evidence.

- Why is the scope for raising revenues through higher tax rates limited in Latin America, as claimed in the November 2006 report (p. 29)? Are Latin American countries at the tops of their Laffer curves? How do we know?

- On what basis does the staff conclude, as claimed in the April 2008 report (p. 23), that most countries in Latin America need to further reduce their public debt ratios over the medium term? Similarly, in the October 2009 report (p. 30) the advocacy of reducing public debt is not backed by analysis of what public debt objectives should be, or why. Overall, though they call frequently for reducing the levels of debt, these reports are not specific about the optimal levels of debt. They should either try to argue for a specific approach to determining optimal debt levels or explain why this cannot be done and how a judgment has nevertheless been made.

Missing analysis

- Why have debt/GDP ratios been falling in Latin America (April 2006 report, p. 9)? More details are needed on this important development. Specifically, a debt accumulation decomposition exercise would have been instructive.

- The much poorer governance indicators in Latin America than in other regions, and the deterioration in these indicators during recent years (April 2006 report, p. 14), deserve much more attention (and explanation) than they receive. Can poor governance be reversed in relatively short order through policy? If so, what needs to be done, and why?

- Why were capital inflows small in Latin America during the period preceding the November 2006 report (p. 9), even though borrowing costs were very low? This needs more explanation.

- The May 2009 report contains no discussion of the role of accumulated foreign exchange reserves in easing financing constraints for fiscal stimulus in Latin America.

- The same document (pp. 34–35) also does not provide a clear and quantitative analysis of which countries have room for how much fiscal stimulus, or of what form that stimulus could take, restricting its discussion to generalities.

- The October 2009 *REO* adopts as a standard for debt sustainability the stabilization of debt/GDP ratios at end-2008 levels (p. 61). This choice seems to be completely arbitrary (especially in light of recent improvements in debt/GDP ratios), and is not justified in the text.

Incoherent analysis

- In the November 2007 report (p. 23), the box on effectiveness of central bank intervention seems to come out of nowhere, and does not fit into the surrounding analysis.

- Why should a stronger primary balance provide more room for countercyclical policies without raising concerns about debt sustainability, but a lower debt ratio *not* do so, as claimed in the October 2009 report (p. 39)?

Incorrect analysis

- A recurrent problem in the WHD reports involves an apparent confusion between inflation and a change in the relative price of food. This appears, for example, in the April 2008 report (p. 24) and in the October 2008 report (pp. 14 and 28).

- I suspect that the positive effect of lower inflation on investment reported in the October 2008 report (p. 60) probably reflects the influence of a more stable overall macro environment, rather than of inflation per se, since it is hard to see why an increase in a stable inflation rate would discourage private investment, except perhaps through distortions in the taxation of capital.

- The October 2009 report (p. 28) claims that premature tightening of macro policies is not as big a problem in the LAC region because recovery there does not depend exclusively on domestic fiscal and monetary policies. This is a *non sequitur*. The fact that external influences have a large effect on these economies does not mean that inappropriate domestic policies can do no harm.

VIII. CONCLUSIONS

136. Forty-four *REOs* were reviewed for technical quality in this evaluation. These reports were intended as policy documents and not academic pieces. At times they made use of novel analytical perspectives. Some reports supported the analysis with references to the relevant professional literature, and the policy analysis was usually reasonable. There was a clear

improvement of the technical quality over time. I will list four directions in which a reconsideration of the content of the *REOs* could go.

137. First, these documents are highly prescriptive, and often advocate very specific policies, rather than simply identifying policy options and the trade-offs that these often involve between conflicting economic objectives. There is a risk that a seemingly arbitrary policy choice strongly advocated in a *REO* may undermine the credibility of the otherwise solid analysis in the report, and make the report too easily dismissed—to the extent that the specific policy advice is rejected. Consider, for example, the policy response to a sharp increase in oil prices in an oil-importing country. Such an adverse supply shock creates a choice between stabilizing the price level, at the cost of a contraction in economic activity, or stabilizing activity, at the cost of an increase in the price level. In more than one instance in these reports, the staff recommends tight fiscal and monetary policies in response to such a shock, implicitly opting for price level stability. This may indeed be the correct choice in many cases, but it is not unreasonable for a country to make the opposite choice. *It may be more constructive for the staff to indicate the conditions under which it would be optimal for a country to do one thing or the other* (e.g., describing the roles of domestic wage indexation, the anti-inflationary credibility of the central bank, the fiscal impact of the oil price increase, and so on), *and to summarize lessons from past experience in responding to such shocks, rather than to simply make a uniform policy prescription.*

138. Second, there are certain areas in which the documents reveal specific policy predilections (many of which I strongly agree with, by the way) that are not necessarily shared by the profession at large. Examples include a preference for pro-poor policies, an emphasis on keeping inflation low, a strong concern with medium-term fiscal sustainability (which occasionally results in too strident a stance against fiscal stimulus to counter recession and a strong bias in favor of the early removal of fiscal stimulus), and a preference for exchange rate flexibility. To be sure, the preference for pro-poor policies is widely shared outside the Fund, and based on its past experience the Fund has very good reasons to be wary of inflation as well as to be sensitive to fiscal unsustainability. But many people have argued that the poor are best helped by fostering growth, that sometimes an increase in inflation may be the lesser of two evils, and that the fragility of a country's fiscal solvency can sometimes be overstated by the Fund (the Asian crisis comes to mind). The point is not that these policy predilections are unjustified, but rather that they should be questioned. *To that end, the REOs might usefully incorporate the views of a diverse panel of experts from outside the institution.*

139. Third, in cases where I have found the analysis lacking (e.g., instances of vagueness, unsubstantiated claims, missing analysis, incoherent analysis, or incorrect analysis), in my view these problems have often arisen from an uncritical acceptance of conventional wisdom inside the institution. *Again, exposing these documents to some outside review before they are issued might help to address this problem.*

140. Finally, the issues dealt with in these documents make clear a number of areas in which reviews of the professional state of the art or new research would be extremely useful in helping to inform policy choices (either of the staff or of the authorities in member countries) in the Fund's particular areas of expertise. The analysis in the *REOs* that I reviewed was weakened by absence of concrete criteria on which the staff could form objective judgments regarding:

- The desirable sustained rate of inflation for specific countries.
- The degree of real exchange rate misalignment.
- The presence or absence of output gaps.
- The effects of changes in economic activity on the inflation rate (i.e., the slope of the Phillips curve).
- The presence or absence of fiscal space.
- The efficiency of government spending.
- The degree of financial integration.
- The presence or absence of monetary autonomy under officially-managed exchange rates.
- The stance of monetary policy (i.e., the operational meaning of a "neutral" monetary stance).

141. *The Fund could usefully review the professional state of the art on each of these issues and support additional research on them where the state of knowledge is not sufficiently advanced to provide concrete guidance to the staff* (and where additional research may actually hold promise of delivering useful insights). Doing so could significantly improve the quality of the analysis in future *Regional Economic Outlook* reports.

References

Kaminsky, Graciela L., Carmen M. Reinhart, and Carlos A. Végh, 2004, "When It Rains, It Pours: Procyclical Capital Flows and Macroeconomic Policies," NBER Working Paper No. 10780 (Cambridge, Massachusetts: National Bureau of Economic Research). Available at: www.nber.org/papers/w10780 .

Nijkamp, Peter, and Jacques Poot, 2004**, "**Meta-Analysis of the Effect of Fiscal Policies on Long-Run Growth," *European Journal of Political Economy* vol 20, pp. 91–124.

Österholm, Pär, and Jeromin Zettelmeyer, 2007, "The Effect of External Conditions on Growth in Latin America," IMF Working Paper 07/176 (Washington, DC: International Monetary Fund).

IEO
Independent Evaluation Office
of the International Monetary Fund

BACKGROUND PAPER

BP/11/07

Macro-Financial Linkages in IMF Research

Gerard Caprio, Jr.

May 20, 2011

 BP/11/07

IEO Background Paper
Independent Evaluation Office
of the International Monetary Fund

Macro-Financial Linkages in IMF Research

Prepared by Gerard Caprio, Jr.

May 20, 2011

Abstract

This study reviews the research on macro-financial linkages at the IMF since 2005. It found that from 2005 to mid-2007, the amount of IMF research on macro-financial issues was limited and the rather small amount of such research that was potentially relevant was not particularly well integrated with the Fund's operational work. From mid-2007 through 2008, IMF research and operational work began to concentrate more on macro-financial-related issues. Still, even by the end of 2008, the research effort in this area was insufficient relative to its importance to the IMF's operational needs. Moreover, the study found that IMF operational work made little use of Fund research on macro-financial linkages, and the main messages from research in this crucial area remained difficult to discern.

JEL Classification Numbers: G00, G01, E0

Keywords: Macro-financial linkages, financial markets, financial crises, macroeconomics

Author's E-Mail Address: Gerard.Caprio@williams.edu

Contents **Page**

ABBREVIATIONS

AFR	African Department
APD	Asian and Pacific Department
BCP	Basel Core Principles for Bank Supervision
DSGE	dynamic stochastic general equilibrium
EUR	European Department
FAD	Fiscal Affairs Department
FDMD	First Deputy Managing Director
FSAP	Financial Sector Assessment Program
FSIs	financial soundness indicators
G-20	A grouping composed of major industrial countries and systemically important developing and emerging market countries
GFSR	*Global Financial Stability Report*
HRD	Human Resources Department
ICM	International Capital Markets Department
INS	IMF Institute
LIC	low-income country
MAE	Monetary and Exchange Affairs Department
MCD	Middle East and Central Asia Department
MCM	Monetary and Capital Market Department
MED	Middle Eastern Department
MFD	Monetary and Financial Systems Department
REO	*Regional Economic Outlook*
RES	Research Department
SIP	selected issues paper
WEO	*World Economic Outlook*
WHD	Western Hemisphere Department
WP	working paper

I. Overview and Main Conclusions

1. This paper assesses research conducted at the IMF in the period 2005–08 in the area of macro-financial linkages. The questions examined include: whether the IMF conducted research on how to integrate financial sector considerations into the standard macro analysis; did it cover linkages and feedback in both directions (i.e., from the financial sector to the macroeconomy, and vice versa); was this research adequate and potentially useful for the operational work of the IMF; and was it used in practice.[1]

2. For the purpose of this review, research includes working papers (WPs), journal articles, and books or chapters, as well as research that is included in the Fund's *World Economic Outlook* (*WEO*) and *Global Financial Stability Report* (*GFSR*). The assessment of the usage of research in operational work is limited to the Article IV documents and the selected issues papers (SIPs) produced on a sample of 15 countries.[2] It is important therefore to clarify that FSAPs, which might utilize macro-financial research, were not included in this review.

3. In evaluating the work on macro-financial issues, the focus is on the extent to which IMF research helped highlight vulnerabilities that would help staff understand the spillovers from the macroeconomy to the financial sector (and vice versa), and/or included explicit tools to assist in modeling different risks. This review also examines the extent to which research explored and highlighted the risks associated with securitized finance, the shift over the years in banks' reliance on markets for liquidity, and the dangers of rapid credit growth, or vulnerabilities that might build up after a period of low interest rates. Other examples of the type of studies sought are those that would help Fund staff understand and analyze credit crunches, or models that could help estimate the impact of financial sector variables on credit and on real activity. Given time constraints, the review omitted more international macro-financial issues, such as the transmission of shocks between countries, though this is certainly an area of interest to the IMF and one that has been featured, for example, in *GFSRs*.

4. The review shows that in the period 2005 to at least mid-2007, the amount of IMF research on macro-financial issues was limited.[3] Work had been done from the early 2000s on financial sector indicators and also on stress testing of financial institutions' balance sheets, but much of the latter was essentially micro-prudential, hence neither body of work

---

[1] Terms of Reference, June 2010.

[2] Although in the general context of IMF output SIPs are treated as research, for the purposes of this study they are treated as operational documents.

[3] Crowe and others (2010) have published a collection of research by IMF economists produced in the last ten years on macro-financial linkages, some of which are included in this review.

was able to detect systemic vulnerabilities threatening the overall economy. The relatively small amount of macro-financial research that was potentially relevant seems not to have been well integrated with the Fund's operational work. Operational documents reviewed rarely referenced IMF research and, although there were some exceptions, the selected issues papers produced by the Fund over the period tended to concentrate on a wide variety of topics other than medium-term macro-financial linkages that may be sources of vulnerability. Both explicit citations as well as clear influence by or use of concepts, tools, models and/or analytical approaches were employed as evidence of the utilization of research in operations, though only citations enable easy quantification.

5. By the second half of 2007 and even more so by 2008, IMF research and operational work on macro-financial-related issues increased significantly. For example, research (in 2008) on the impact of financial conditions in the United States, proxied by data from the U.S. Federal Reserve Board's survey of bank lending practices, explores the impact of tightening loan standards on GDP. Although it was utilized in operations that were outside the evaluation period, it is worth noting, given the possible lags between the creation and usage of research, that the findings were incorporated in the 2009 Article IV report for the United States. Other good examples of research are available, such as Segovian Basurto and Padilla (2006), Tamirisa and Igan (2008), and Classens, Kose, and Terrones (2008), but none of these were utilized during the evaluation period in country documents reviewed in this background study.

6. In operations, the emphasis on macro-financial work in the documents examined was slight during the overall period under review and in particular during 2005 to mid-2007. Part of the explanation no doubt was the economics profession's lack of attention to financial sector variables in standard macro models until well after the crisis began. However, the lack of attention to macro-financial issues during the first part of the review period seems even more pronounced than in the economics profession at large. As suggested in the internal 2007 "Report of the Taskforce on Integrating Finance and Financial Sector Analysis into Article IV Surveillance," there might be some resistance to the integration of financial variables into macroeconomic analysis in the Fund, due to staff backgrounds and training (consistent with the bias in the broader economics discipline), although it should be noted that since the Taskforce Report was issued the IMF Institute has increased its training in this area.

7. The next section discusses the types of macro-finance research reviewed and the specific outputs covered. Section III contains an examination of this research with some examples if not a paper-by-paper evaluation, while Section IV looks at the extent to which macro-finance research is utilized. The concluding section offers some observations as to how both the supply and demand side of the research production and utilization process might be improved. Fortunately the Fund appears to be in the process of significantly increasing its attention to this area; time will tell if it succeeds in becoming a leader in the area.

II. Definitions and Research Reviewed

8. This review included an examination of the IMF's macro-finance research. In particular, it examines whether research may have generated tools to help link better developments in the financial sector and its effect on the macro economy and vice versa, and the channels through which this might happen. The issues of concern include not only how risks from securitization might affect the economy—in addition to how they might spread throughout the financial system—but also the impact of rapid credit growth on household, corporate, and intermediaries' balance sheets and subsequent knock-on effects. Research on bubbles and credit crunches and, more generally, how these affect economic activity was also sought.

A. Understanding Macro-Financial Linkages

9. "Macro-financial linkages," as that term is used in this review, refers to the interaction between the financial sector and the domestic economy. Although it can also be used to examine linkages from international financial flows to an economy, or from an economy's financial system to the rest of the world, this review focuses more on domestic linkages, because it was the sense of the 2007 "Report of the Taskforce on Integrating Finance and Financial Sector Analysis into Article IV Surveillance" that the Fund's analysis of the interplay between the domestic finance system and the domestic economy was less developed than was deemed to be appropriate in light of the IMF's mission.

10. In post-crisis discussions, it is widely acknowledged that the economics discipline did not do an adequate job of integrating financial and macro variables in the toolkits of macroeconomists. To be sure, even standard texts used in undergraduate money and banking courses[4] emphasize various monetary policy transmission channels, such as balance sheet and bank capital channels. These channels posit that bank loans are a function of the net worth of the bank itself and the borrower; thus changes in asset prices affect both bank and borrower net worth, and therefore the ability and/or desire of banks to make loans. A rise in asset prices would boost the value of bank capital (e.g., by raising equity prices, including that of banks, directly and/or indirectly by reducing nonperforming loans) as well as the value of collateral, and lead to an increase in loans, while a fall in asset prices would do the reverse.

11. However, at the graduate level and in the papers in many top journals in the field, macro-financial linkages have received less attention. It is important to acknowledge that finance and macroeconomics have long proceeded in isolation from one another, in part due to the drive to use mathematical models and the related need to simplify these as much as possible. As Paul Krugman put it, "As I see it, the economics profession went astray because

[4] See those by Mishkin or Cecchetti, which together dominate this field.

economists, as a group, mistook beauty, clad in impressive-looking mathematics, for truth."[5] For years, economists were trained in models without money or a financial sector; and until the late 1990s, no text on development economics even included a chapter on financial systems. Macroeconomics courses in many graduate schools until very recently relied on dynamic stochastic general equilibrium models that contained little in the way of a financial sector, so it was possible for macroeconomists to have virtually no understanding of this sector and macro-financial linkages. Recent economics graduates would be familiar with the Bernanke, Gertler, and Gilchrist (1996) approach—in which financial market conditions can greatly amplify initial shocks (their "financial accelerator")—which for example would predict a more difficult recovery from the bursting of an asset bubble than would otherwise be the case. But those models were out of favor with the efficient-markets thinking that dominated the profession.

12. As noted, macro-financial linkages can include those coming through the international financial system via a variety of channels (interest rates, risk premia, currency or bank runs, etc.) to impact the domestic economy. Usually the domestic financial system is affected as a result. For example, if banks have borrowed in foreign exchange and lent in domestic currency, then a devaluation of the local currency could lead to their insolvency (or at least to a capital constraint on lending); if they passed this risk on to their borrowers (i.e., by lending to them in foreign currency, and hence converting foreign exchange risk into credit risk), a devaluation would hurt borrowers who lacked foreign exchange receivables and would also hurt the banks, to the extent that nonperforming loans rose. These channels have received little attention from macroeconomists, who implicitly or explicitly assumed that risks were being optimally shared. Meanwhile, many finance PhDs have been trained in the use of mathematical finance models because, in the words of a former American bank robber, "that's where the money was." These models, however, favored analyzing risks specific to individual instruments and/or institutions, and not risks to the financial system at large. The belief that risk was being divided and parceled out to those who could best bear it was especially dominant in the profession. Unfortunately, attention was not being paid to the changes in compensation practices, which are now known to have rewarded the volume of activity undertaken (loans originated, the number of securities created, or rated, etc.) without attention to the risks being generated.

B. IMF Research Reviewed

13. Table A1 in the Annex lists the research work reviewed for this study, with the working papers sorted by the year released. The research WPs and the analytical chapters of the *WEO*/*GFSR* were selected as examples of macro-finance research (or outputs that included macro-finance research) in consultation with IEO, beginning with a longer list

---

[5] Paul Krugman, "How Did Economists Get It So Wrong?" *New York Times*, September 2, 2009.

provided by IEO as those most relevant to the topic of macro-financial linkages. After examining closely those most related to macro-finance from 2005 to 2008, the list was extended to include a few additional WPs from 2004 for review, in light of the likely lag between research and its inclusion in more operational work. As it turned out, most of these early WPs were not on macro-financial linkages as defined here. An original plan was to review research (if not on this list) that was cited most in operational work, but as noted below, in practice no macro-finance research papers were being regularly cited in operations.

14. It should be clear that this review of research on macro-financial linkages is selective, so there may be relevant papers not included here. However, in searching through the titles of all 1,130 working papers released between 2005 and 2008, I only found, in addition to those listed in Table A1, 15 that looked relevant to domestic macro-financial issues. Still, just the papers listed in Table A1 were part of this review.

15. Only about half of the papers listed in Table A1 seem highly relevant to the crisis (see para 16). As a result, taking the papers from Table A1 and adding the other 15 that appeared from their titles to be possibly relevant to macro-financial issues, means that perhaps only 2–3 percent of research that the Fund released in working paper form during the review period was in the area of macro-finance. This figure appears small in relation to the potential importance of the topic for IMF operations. It might also explain the comments received in IEO interviews: when asked about macro-finance research in the Fund, common responses were that "this was a new area for the Fund," or "wasn't aware of any until recently."

16. Note that for research papers released after the middle of 2008, there would have been relatively little time for such work to be incorporated into operations during the period under review. Staff would have needed time to identify relevant research and acquire any necessary training to implement this work in operations. Still, a review of these later papers was useful in forming an opinion of the relevant work that was in the pipeline, and accordingly they were included in the review.

III. Overview of Fund Macro-Financial Research

17. In this section I comment on the work on macro-financial linkages that seems to be of greatest operational usefulness for member countries. As shown in Table A1, much of the research issued in 2004–06 was not on domestic macro-financial linkages but rather focused on long-run growth (Abiad and others, 2004; Edison and others, 2004), exchange rates and/or capital account opening (Bulir, 2004; Tamirisa, 2004; Mercereau, 2004; Tytel, 2004; Arvai, 2005; LeFort-Varela, 2005; Lane, 2005; and Kose and others, 2006), and fiscal policy (Jaeger, 2004). Some of the other early papers listed are more conceptual in nature—for example, on how to define and safeguard financial stability. One (Jones, 2004) provides an overview of the stress testing process, which would be relevant for readers unfamiliar with the process but would not enable them to engage in this exercise itself; moreover, the paper

asserts but does not show the value added from this process. It might be, for example, that stress tests are useful in getting bankers and/or supervisors who were not doing so to think about risk, but that their predictive value is less certain.

18. One of the early papers on the list (Dell'Ariccia, 2005), on "The Real Effect of Banking Crises," is a very good example of useful research on macro-financial linkages, as it finds evidence on the importance of the lending channel when banks are in distress. Although this research would not have been of direct applicability in operations in the run-up to the crisis, it could have sparked interest in factors that cause financial distress, leading to more focus on the warning signs thereof, etc. An understanding of the importance of the lending channel could also have helped staff appreciate the downside risks from the crisis and the likelihood of a slower recovery the greater the impairment of the balance sheets of the financial, household, and corporate sectors. The paper did not, however, provide a specific tool for operations staff, nor was this its focus. Another relatively early paper that appears relevant is that on "The Quality of Financial Policies and Financial System Stress" (Das and others, 2005).

19. Das and others (2005) and a later IMF paper on the same theme raise a broader issue about how to handle conflicting results from research. The first paper provides an index of the quality of financial policies, drawing on the assessments of compliance with the Basel Core Principles for Bank Supervision (BCP) and the IOSCO Objectives and Principles of Securities Regulation. In looking at the impact of the quality of financial policies, the dependent variable it uses is an index of financial stress that is not widely used by researchers, and the authors also are unable to take account of other institutional features. A later paper by Detragiache and Demirgüç-Kunt (2010) that uses a significantly larger sample of countries, includes institutional variables, and uses z-scores (a commonly accepted measure of risk), confirms instead other research that shows that the Basel Core Principles Assessment scores do not predict bank risk. Moreover, the most recent crisis confirmed the latter findings: distress was concentrated among the better performers on BCP assessments. According to a recent report by the Governor of the Central Bank of Ireland (Central Bank of Ireland, 2010), bank supervision performed poorly in the years leading up to the crisis, yet Ireland had strong results on its BCP assessment (and the crisis was not signaled in the 2006 FSAP Update). So when there are conflicting results on issues that are important for the Fund, IMF Management needs to weigh in with a judgment or demand more work to resolve disputes up to the point at which it feels comfortable making decisions on the basis of existing results. In this case, the Fund may or may not want to try to push for changes in the Core Principles exercise or even reallocate staff if the exercise cannot be made productive.

20. The point of this example is that when there is significant research lessons, from research carried on either within or outside the Fund, operational work that blatantly ignores research findings tells researchers that their work is not valued, making it more difficult to attract and motivate researchers in areas deemed to be important. Also, when Management ignores the operational lessons from research, based on my experience in several

organizations, the response of researchers is to work on their own priority areas, rather than those that should be the priority of the organization. Finally, neglect of research findings by Management tells operational staff that they can safely ignore research.

21. Another example about the handling of research findings concerns the pricing of deposit insurance. Research that was originally issued in 2002, prior to the crisis and prior to the evaluation period, had shown using several different methodologies that in many cases, deposit insurance is underpriced (Laeven, 2008, originally released as World Bank working paper in 2002). IMF operations in the 1990s and 2000s recommended deposit insurance but the underpricing of deposit insurance may have contributed to the worsening of the crisis, reducing market monitoring and leaving governments with a larger bill for covering losses in the banking sector. Regardless of where research originates, researchers should have the responsibility for flagging important results in areas that are critical to the IMF's mission to Management. In this case it would have been natural to investigate whether changes in the financial sector were leading to less appropriate pricing of deposit insurance, for example, whether greater risk taking and higher leverage by banks should have required a higher deposit insurance premium—an example of potential macro-financial research that could have had important policy implications for the IMF but that as far as could be determined in this review did not occur.

22. By 2006, a few more papers with macro-financial implications had been issued. Dell' Ariccia and Marquez (2006), on "Lending Booms and Lending Standards," show how a seeming reduction in information asymmetries leads to a lending boom and a lowering of bank lending standards. This idea was not new, of course, having been discussed in Kindleberger's *Manias, Panics, and Crashes* (1978).[6] But the paper's theoretical approach should have provided ammunition to counter the reigning "efficient markets" thinking, in which bank lending behavior would be assumed to be optimal. Basu (2006) proposes a framework for dynamically projecting forward the banking system's balance sheet, with some consideration of the effect of aggregate macroeconomic variables, though this paper did not include any new tools.

23. Work by Segoviano-Basurto and Padilla (2006) on quantifying portfolio credit risk and macroeconomic shocks is an important paper that comes with an application to the Danish Financial Sector Assessment Program (FSAP). Their methodology permits the probability of default to be modeled as a function of macro and financial variables, and allows for an early recognition of risk as macro conditions change. Again, it was not cited in country work but likely was incorporated into other FSAPs. It would have been interesting to see whether country teams agreed with the range of macro assumptions that were made in this approach, or whether they used it to run scenario analysis.

---

[6] And later in Reinhart and Rogoff's 2009 analysis of historical booms and busts, *This Time Is Different*.

24. The early *GFSR* and *WEO* chapters reviewed—both from Fall 2006—also did not contain research tools for operational staff. The *GFSR* chapter on household credit growth in emerging markets took the hardly controversial position that this growth was a positive development from a very low base but that care, presumably by borrowers, lenders, and authorities was needed as leverage ratios increase. The *WEO* chapter on how financial systems affect economic cycles attempted to measure and review the growth of "arm's-length" finance, looking at differences in how households use their homes to borrow, the impact of changes in asset prices on household spending, changing corporate financing patterns, etc. One of the innovations in this chapter is an index of arm's-length financing and its effect on behavior. Certainly the chapter contains a number of messages that could encourage country staff to consider how the economy on which they are working would rate, and how arm's-length financing plays out in practice. But the overall message from the chapter is a bit muted:

> "For example, under a more arm's-length system, households are able to access a larger amount of financing and seem better able to smooth consumption in the face of temporary changes in their income. This may have contributed to the reduction in consumption volatility over the business cycle. In more arm's-length systems, however, households appear to be more vulnerable to swings in asset prices, implying larger effects on demand from major asset price booms and busts. This effect, however, may be countered to some degree by the fact that the amplitude of swings in asset prices may be lower in more arm's length systems."

25. The last sentence, unfortunately, though understandable in view of the boom prior to the crisis, turned out to be incorrect—the swings were much greater directly as a result of securitization, as more information was lost in the originate and distribution process and leverage increased. Also, greater attention might have been devoted to the drivers of the more to "arm's length" financing, namely the ability to avoid capital requirements and also to generate fee income. Attention to the loss of information and the distortion of incentives in favor of growing the volume of activities, without due attention to risk, would have led to a different conclusion in the report and possibly greater attention to risks by operational staff. Concrete recommendations (e.g., implementing an overall leverage requirement that consolidates off-balance-sheet exposures, as has been done in Canada) were not made, and the chapter merely offered the general view that regulation would need to adapt.

26. It is interesting that not more papers were produced in 2007 on macro-financial linkages, particularly as the warning signs of the crisis began emerging in 2006, when real estate markets in several countries peaked. Babihuga (2007) carries on earlier work on financial soundness indicators (FSIs), looking at how those vary with macro variables. This paper says it is the "… first attempt to explore the macroeconomic determinants of FSIs using a large panel dataset." Given that these indicators have been in use since 2002, it is puzzling that it took so long to investigate their links with macro variables. Babihuga's

confirmation that FSIs fluctuate with the business cycle is not surprising, but the paper does not address the issues of which FSIs are most sensitive to the cycle, controlling for the structure of the financial sector, or of which if any might be helpful in predicting macro variables.

27. By 2008, the amount of IMF research in the macro-finance area had clearly increased. Igan and Tamirisa (2008) investigate the extent to which weak banks had led the lending boom in emerging Europe.[7] Interestingly, they do not find that rapid loan expansion weakened banks, or that weak banks were leading the way in their sample. However a critical feature is that their data set is for the period 1994 to 2004. The problem with rapid credit growth and, even more, Ponzi finance, is that the weaknesses do not show up until the boom is over (hence, Warren Buffett's famous saying, "You don't discover who is swimming naked until the tide goes out"). This is not a critique of the authors, but rather illustrates the need to extend most macro-financial research over at least a couple of business cycles before having confidence in the conclusions—and therefore the importance of sustaining macro-finance research even in a period of "Great Moderation." Nor do the authors split their sample to isolate the countries with quite high credit growth (the Baltics), finding that weak banks did indeed lead credit growth there. Thus, it might well be that in countries with more restrained credit growth, some feature of the institutional environment, such as better regulation or better market monitoring, acts as a check on weak banks' growth. This research is certainly worth pushing further.

28. In 2008, several papers by Carabenciov and others were released, using a small quarterly multi-country model with some financial sector linkages. Although a key goal of this research effort is to analyze cross-border financial and macro linkages, the work does begin to flesh out a financial sector. Unfortunately, it does so mainly through the U.S. Federal Reserve Board's Senior Loan Officer Survey of bank lending practices. Lending practices are a potentially important variable but one that still leaves much of the financial sector out of the model. Moreover, as was seen during the crisis, loan officers' views of the economy and, hence, their own behavior, are subject to very rapid revision, and one that cannot be predicted if their opinions are treated as exogenous. Still, this research effort is potentially useful, and it will be interesting to see if its conclusions are picked up by operations staff. As the papers were all issued in 2008, it would have been premature to expect an impact on operational work during the review period.

29. By the end of 2008, an important paper was issued by Claessens, Kose, and Terrones on "What Happens during Recessions, Credit Crunches, and Busts." This paper focuses on the interactions between macro and financial variables, concluding that recessions associated with credit crunches and housing market collapses tend to be significantly deeper and longer

[7] Unfortunately, they did not look at Ireland, where there would have been a very clear example of a weak bank leading the boom and even potentially encouraging other large banks to step up their mortgage lending.

than otherwise. Equity price collapses tend to be followed by shorter recessions. This line of research is fully consistent with the Bernanke view that "banks are special," dating back to his work on the Depression. An implication of this research is that stronger policy responses are called for when the banking system is in distress. Late 2008 also saw the release (actually, an updating and extension) of a database by Laeven and Valencia. These data are important because they permit a detailed investigation of crises.[8]

30. By 2008, the *WEO* and *GFSR* had increased their attention to macro-financial linkages. The April *GFSR* that year on "Market and Funding Illiquidity: When Private Risk Becomes Public," is mostly descriptive of the shift from banks holding their own liquidity to relying on the markets to supply it. This shift dates back to the decline of reserve requirements in the 1970s and 1980s, though it was significantly furthered by more recent innovations. The same chapter also discusses liquidity spirals—the phenomenon that as more intermediaries call on the market for liquidity, the supply of liquidity shrinks, because each institution fears it could be the next one to need it themselves. The authors warn that these spirals could become more prevalent in a world in which many institutions are relying on the market, and they recommend more attention to liquidity management practices in the FSAP. This paper clearly would have helped IMF staff better understand spillovers from the financial sector to the real economy. To be sure, it would have had a greater potential impact had it been issued earlier.

31. The April 2008 *WEO*, in Chapter 3, looks at "The Changing Housing Cycle and the Implications for Monetary Policy." This chapter notes the role of homes as collateral for borrowing and how this increases the financial accelerator, so that in countries where this phenomenon is occurring, housing prices will have a greater impact on the economy. In effect, as the authors note, the transmission mechanism for monetary policy thereby changes: instead of taking effect mostly through the impact on interest rates and thereby on residential investment, monetary policy operates more through housing prices. This change, for example, explains a key difference in economic forecasts today, as forecasters who put more weight on the housing-price channel expect a slower recovery where housing price decreases have wiped out collateral values and thus should be expected to restrain consumer spending. Echoing the paper of then-Federal Reserve Governor Mishkin at the August 2007 Jackson Hole conference, this increased sensitivity of the economy to housing justifies a more aggressive response of monetary policy to a downturn in prices.

32. By Fall 2008, the crisis was in full swing, following the collapse of Bear Stearns the previous March, the takeover of Fannie Mae and Freddie Mac that summer, and then the September failure of Lehman Bros., the takeover of AIG, and the creation of "bailout" programs in a number of countries. Both the *GFSR* and the *WEO* of October 2008 focused on

---

[8] In fact, the same two authors (Laeven and Valencia, 2010) have released an improved version of that database and used it to investigate bank resolution and compare interventions with earlier crises.

the impact of financial stress. The former (in Chapter 2, "Stress in Bank Funding Markets and Implications for Monetary Policy") looked at the microstructure of interbank markets, which essentially had shut down that fall, conducted an empirical investigation to see what accounted for the jump in interbank spreads, examined the impact on monetary policy transmission, and made policy recommendations. This chapter clearly speaks to helping staff understand credit crunches as well as the need for many of the innovations in monetary policy on the part of the U.S. Federal Reserve Board and the European Central Bank. A *WEO* chapter ("Financial Stress and Economic Downturns") focuses squarely on macro-financial linkages as covered in the present study, including how downturns that are accompanied by banking distress tend to be more serious than those where stress is centered in nonbank intermediaries (no doubt reflecting the 2008 paper by Claessens and others noted earlier). Picking up on the theme of arm's length versus relationship-based finance, the *WEO* chapter argues that downturns will be more severe in the former systems, and that restoring capital in banking systems is critical to crisis resolutions.

33. There are a number of issues that with hindsight are now recognized as being important subjects on which research within and outside the Fund was lacking, but two topics in particular should have been of clear importance in the run-up to the crisis and were not investigated. First was the issue alluded to earlier of compensation in the financial system and how this was distorting incentives. Second was the fact that more than half of U.S. mortgage securities were being purchased by those outside the United States. Although many or even most purchasers might have been putting their faith in ratings, the fact that these highly-rated securities were paying rates of return above those of other comparably-rated securities should have been a warning sign that they were riskier than their comparators. In addition to reporting on this red flag, there seems to have been no investigation of whether authorities were aware of, concerned about and/or taking action on these developments, whether they understood the extent to which the purchases were driven by distortions in compensation practices in the purchasing institutions, etc.

IV. Utilization in the Fund's Operational Work

34. To assess whether Fund research was reflected in the IMF operational work we examined the content of selected issues papers (which are part of the Article IV consultations) in a group of selected countries/economies: Australia, Croatia, Chile, Estonia, Hong Kong SAR, Hungary, Ireland, Korea, Latvia, Poland, Romania, Spain, Sweden, Switzerland, and the United States. Some of the countries had significant financial imbalances building up over the evaluation period and so were excellent candidates for work on macro-financial issues. It is important to underline as noted earlier that this review omits the FSAP, presumably because the macro-financial linkages were to be the focus of Article IV work, rather than the FSAP itself. Still, this omission might well leave out an important area in which research is used in operations. Moreover, it is useful to clarify that the utilization of research in operations was judged directly by citations and indirectly by the concepts and/or tools employed, though citations themselves are the most objective measure.

35. IMF research was rarely cited in IMF operational work, and those citations tended to be much more of papers produced within the same area department. In Article IV work, research was rarely cited and most such citations were of *GFSR* or *WEO* chapters. This conclusion is based on a review of at least two Article IV reports per country—typically the 2005 or 2006 report and the 2008 or 2009 report (the latter years were added in the hope that some of the macro-financial work would be featured). There is no evidence of any IMF macro-financial research being repeatedly cited—or of being cited in more than one Article IV document. Selected issues papers much more commonly cited research. Of the studies that the SIPs cited, the overwhelming majority are studies done outside the IMF, followed by the work of authors in the same area department as the operational work being examined, followed by research in the Fund's central departments.

36. The topics that were covered by SIPs suggest that, in Fund operations rather as in working papers, the period 2005 to mid-2007 was characterized by little attention to macro-financial linkages. Selected issues papers released during the evaluation period focused on a very wide variety of issues but rarely on macro-financial linkages. Table A2 lists the SIPs issued during the review period for 15 countries mentioned above (para 34). The table shows that 139 SIPs were issued over the 2005–08 period. A perusal of this list shows that of these 139, 20 papers (shown in bold in the table) treated macro-financial linkages, and 13 of the 23 were produced after the middle of 2007. However, to the extent that finance was covered, it tended to be more from the perspectives of finance and growth, pensions (due no doubt to their fiscal impact), and finance and productivity growth, rather than of analyzing macro-financial linkages with implications for vulnerability that might have helped understand and prepare for the recent crisis. And again, papers that treated the financial system in isolation—for example, the 2005 paper on the characteristics of the Swiss financial system, or that from 2006 on the derivatives market in Hong Kong SAR—without treating the links between the financial sector and the broader economy were not counted as evidence of macro-finance linkages.

37. To be sure, there are exceptions to this generalization. For example, the 2006 selected issues paper for Latvia contains a chapter on aspects of rapid credit growth, in which the macro implications are examined, and the 2006 paper for Estonia looks at balance sheet issues, giving some attention to the foreign exchange and maturity mismatches on the private sector's balance sheet. Both studies capture some macro-financial linkages.

38. Still, the foremost explanation for the dearth of references to the macro-finance literature—whether generated from IMF research or done externally—in addition to the relative dearth until recently of this research as noted above, is the topical coverage: operational work as represented in Article IV reports and SIPs had predominantly focused on a variety of other issues. Important as these issues might have been, they did not help, in understanding the macro-financial links that may have helped understand mounting countries' vulnerabilities and eventually the recent crisis.

39. For example, the 2005 selected issues paper on Ireland treats the fiscal situation, unemployment, and savings. And the 2007 paper on the same country—issued in August of that year, as the crisis was breaking in international markets, though it would be a year later until the nature of the Irish crisis became undeniable—treats trade and financial spillovers to Ireland, population aging, and the efficiency gains of private credit growth in Ireland. Yet as the Report of the Central Bank of Ireland (2010) as well as another report by Regling and Watson (2010) make clear, the Irish crisis was fully "home grown," and spillover effects affected the timing and had only some impact on the amplitude of the crisis. Unfortunately, an examination of the clearly excessive (up to 40 percent) pace of growth over the period 1999–2007 of Irish banks' balance sheets, which deserved highlighting, did not receive any attention.

40. The 2006 selected issues paper for the United States focuses on "The Attractiveness of U.S. Financial Markets: The Example of Mortgage Securitization," and concludes with "This paper suggests that U.S. financial markets have been skilful in developing tools that have helped households exploit favorable global financing conditions to boost homeownership and acquire housing wealth." As noted above, at that time there was little IMF research on macro-financial linkages that would have helped avoid this conclusion, or the implication of the Irish study that arm's-length financing (a message from the *WEO*) was desirable, with no clear mention of the risks.

41. From the last two columns of Table A1, which note whether the research item was used in operational work, what stands out is the large number of papers that were not used in the sample (see Section V below) but that could have been of use. A related issue, but one that was outside the scope of this report and in fact more suitable for review by the IMF, is the issue of what accounts for this underutilization—whether for example it relates to staff backgrounds and training, as mentioned in para 6 above.

42. What stands out from the review of the aforementioned operational work? First, limited treatment of macro-financial issues; second, a marked tendency to cite work from within the originating country department; and third, even within the 2008–09 papers examined, little citation of research from functional departments. The lack of macro-financial research or tools that are repeatedly cited is striking.

43. The amount of IMF operational work on macro-financial issues seems to be on the rise. For example, all of the six SIPs in 2008 for the United States were on macro-financial linkages or on financial issues that had direct macro implications even where these were not stressed. However, it is not clear whether once the current crisis is past the focus of operational work will remain in this area or return to a mix of selected issues. In view of the Fund's roles in averting and alleviating crises, and the centrality of macro-financial linkages in these events, some ongoing focus in this area is clearly needed.

V. Concluding Observations and Implications for Research

44. By the end of the 2005–08 evaluation period, the Fund was producing more research on macro-financial linkages, after at least a relative drought of several years. But even this later effort by the end of 2008 seems to have been inadequate relative to the importance of this area to the Fund. On the other hand, utilization in operational work of whatever research was available was low. Thus, attention to both the supply and the demand side of this issue seems essential to rectify this situation.

45. On the supply side, given the IMF's role in averting and alleviating crises, and the importance of macro-financial linkages in the crises of recent decades, a much stronger push for research in this area appears warranted. Indeed a strong case can be made that the Fund should be a leader in this field, comparable to the Bank's role in conducting and supporting research on poverty and its alleviation. Increased priority for macro-finance research means allocating increased research funding and/or staff positions to this area.

46. Fortunately, increased attention to macro-financial research seems to be in progress. Claessens, Kose, and Terrones (2008) lay out selected areas for future research that focuses on the channels through which macro and financial variables interact. The Research Department, which was lightly represented in the list of macro-financial research in Table A1, now seems to have taken on macro-financial linkages as a key research area. The department's current work program includes the following important areas[9]:

- Real estate cycles. Why do real estate cycles arise? What makes them special? What tools can be used in response to imbalances in real estate markets? What are country experiences? What policy lessons follow?

- Regulation and taxation. A compilation of the background analysis done for the G-20 as well as new work on regulation and taxation, including calibration for systemic risks.

- Designing and calibrating macro-prudential measures. What measures, which indicators, and what calibrated parameters could be used for macro-prudential purposes (e.g., if time-varying capital adequacy requirements are used, how are they adjusted to what measurable aspect of business and financial cycles)?

- Impact of regulation (and taxation) on financial intermediation and real economy. Costs and benefits of the (new) regulations, like Basel III. Using analytical/theoretical models, DSGE-type approaches, data analysis, event studies, etc.

---

[9] From correspondence with the Research Department.

- Structure of regulation and supervision. What works best for systemic stability? What are the potential conflicts of interests between say monetary policy and financial stability, or between micro-prudential supervision and financial stability? What are implications for regulatory governance (forms of accountability, legal, financial and operational independence, information sharing, integrity)?

- Financial structure and financial stability. What types of financial structures lend themselves to more financial stability while maintaining efficiency? What sources of earnings are more stable/of higher quality? What does the restructuring of the financial services industries mean for the earning potentials of different types of financial institutions?

47. This agenda makes great sense given where the literature is today. In order to achieve an impact, it will be important that this research effort is sustained and supported by senior management, and not just a short-term response to the crisis.

48. On the demand side, as the research effort on macro-financial linkages is ramped up, it also will be important to ensure that its main messages are clarified and disseminated to staff and outsiders. In addition to the finding that for part of the review period the supply of IMF macro-finance research was slim, in view of the low usage of the research that was undertaken, more needs to be done to disseminate key research messages, in particular those that research units and Management want to ensure are incorporated into operations. Yet conversations with a few staff suggest that the Fund's main messages on this crucial topic are not easily identified, and as noted earlier (para 15), a common response in staff interviews was that they were not aware of much if any macro-finance research. Moreover, conversations with a few outside leaders in the macro-finance field in academia revealed that they could not identify key macro-finance research messages that they associated with the IMF.

49. Although the *GFSRs* and *WEOs* can be used for the purpose of disseminating research messages, as they have been in the recent past, it is not always clear that this is their goal. A potentially useful complementary or alternative approach is to issue research reports that assemble the messages and/or tools from different pieces of research. Indeed, announcing a research strategy that focuses on several areas, such as the above list from RES, and aiming at high-profile research reports in the future, in the experience of this author, not only contributes to better dissemination but also helps motivate researchers. A working paper or even a journal article should be an early step in dissemination, not the last.

50. In addition to having a well defined and publicly advertised program of research in macro-finance as well as a clear and enduring commitment by senior Management that this area is a long-term focus for the Fund, Management may well want to look at other ways to improve the usage of research findings, which depends both on its availability and relevance.

Operations staff are pressed for time and asked to look at a variety of areas (hence the need for guidance on key messages), so they may need some help in sorting through priorities and messages. It would be useful to consider ways to ensure that macro-financial research is both relevant to and integrated with Fund operations.

51. Beyond the recommendations of the 2007 Taskforce Report, some regular integration of research staff into Fund missions is worth considering as a way to increase communication about the most relevant research available (whether that is generated inside the IMF or not) and about operational priorities about research. By working more than at present with country teams, research-oriented staff would have the chance to disseminate their research, learn more about the key research needs of country authorities and IMF country teams, and acquire data and even collaborators.[10] Another by-product of greater integration of country teams and researchers would be a reduction in any lag between the creation of research and its utilization in operations. In the World Bank, this integration of researchers on operational teams helps minimize the lag between the origination of research and its usage, and also helps improve research relevance. To be sure, this process itself demands oversight, as the dangers of not thoroughly testing research results are known in many fields. Still, increased communication between more operational and more research-oriented staff is likely to improve operations work as well as the quality and relevance of the research that is undertaken.

52. A more detailed look at the training and recruiting program in the macro-finance area is beyond the scope of this study but clearly of potential relevance. As noted, the IMF Institute has increased course offerings here, a promising development.

53. Regardless of how much the economics discipline moves in the same direction, the potential importance of macro-financial linkages to IMF operational work suggests that the Fund should be a leader in this area. As indicated earlier, management attention to the messages of research—and the insistence through the review process that research findings be consistently upheld in the Fund's work—will be important to ensuring that first-rate relevant research is produced and utilized. With attention to both the supply and demand side of macro-finance research, a future review of work in this area should find a more ample and relevant supply and a greater utilization of the findings in Fund operational work.

[10] Note that this recommendation is intended to apply to research-oriented staff in any of the central departments. It is my understanding that this recommendation, at least applied to staff in the Research Department, was part of the recommendations of the Mishkin report (Mishkin and others, 1999) and that although some movement has occurred, it has been limited.

ANNEX. RESEARCH AND OPERATIONAL DOCUMENTS REVIEWED

Table A1. List of Research Reviewed

Year	Issuing Department	Title	Author	Used in Operational Work (Y/N/C)[1]	Potential Use In Operational Work (Y/N/C)
2010	INS	Basel Core Principles and Bank Risk: Does Compliance Matter?	Detragiache and Demirgüç-Kunt	N	Y
2009	EUR	From Subprime Loans to Subprime Growth: Evidence from the Euro Area	Cihak and Brooks	N	Y
2008	RES	What Happens During Recessions, Crunches, and Busts?	Claessens, Kose, and Terrones	N	Y
2008	WHD	Credit Matters: Empirical Evidence on U.S. Macro-Financial Linkages	Bayoumi	N	Y
2008	APD	Real and Financial Sector Linkages in China and India	Aziz	N	Y
2008	RES	A Small Quarterly Multi-Country Projection Model with Financial-Real Linkages and Oil Prices	Carabenciov	N	Y
2008	WHD	Real Effects of the Subprime Mortgage Crisis: Is it a Demand or a Finance Shock?	Tong	N	Y
2008	WHD	Real Implications of Financial Linkages Between Canada and the United States	Klyuev	N	N
2008	WHD	A U.S. Financial Conditions Index: Putting Credit Where Credit Is Due	Swiston	Y	Y
2008	RES	A Small Quarterly Projection Model of the US Economy	Carabenciov	N	Y
2008	RES	A Small Quarterly Multi-Country Projection Model	Carabenciov	N	Y
2008	EUR	Are Weak Banks Leading Credit Booms?	Igan and Tamirisa	Y	Y
2007	MCM	Macroeconomic and Financial Soundness Indicators: An Empirical Investigation	Babihuga	N	Y/C
2007	EUR	Do Economists' and Financial Markets' Perspectives on the New Members of the EU Differ?	Luengnaruemitchai	N	N
2007	OED	Global Imbalances and Financial Stability	Xafa	N	N
2006	MFS	How Do Central Banks Write on Financial Stability?	Cihak	N	N

Table A1. List of Research Reviewed (continued)

Year	Issuing Department	Title	Author	Used in Operational Work (Y/N/C)[1/]	Potential Use In Operational Work (Y/N/C)
2006	MFD	A New Risk Indicator and Stress Testing Tool: A Multifactor Nth-to-Default CDS Basket	Avesani	N	Y
2006	MCM	Portfolio Credit Risk and Macroeconomic Shocks: Applications to Stress Testing Under Data-Restricted Environments	Segoviano–Basurto and Padilla	N	Y
2006	INS	International Financial Integration, Sovereignty, and Constraints on Macroeconomic Policies	Kletzer	N	N
2006	RES	Lending Booms and Lending Standards	Dell'Ariccia and Marquez	N	Y
2006	RES	Financial Globalization: A Reappraisal	Kose and others	N	N
2005	MFD	Quality of Financial Policies and Financial System Stress	Das	N	Y[2/]
2005	APD	The Real Effect of Banking Crises	Dell'Ariccia	N	C
2005	INS	Stock Market Liquidity and the Macroeconomy: Evidence from Japan	Choi		Y
2005	RES	Financial Globalization and Exchange Rates	Lane	N	N
2005	OED	Capital Account Liberalization and the Real Exchange Rate in Chile	LeFort-Varela	N	N
2005	EUR	Capital Account Liberalization, Capital Flow Patterns, and Policy Responses in the EU's New Member States	Arvai	N	N
2004	RES	Does Financial Globalization Induce Better Macroeconomic Policies?	Tytell	N	N
2004	MFD	Quantitative Assessment of the Financial Sector: An Integrated Approach	Worrell	N	C
2004	ICM	Toward a Framework for Safeguarding Financial Stability	Houben	N	C
2004	MFD	Stress Testing Financial Systems: What to Do When the Governor Calls	Jones	N	C
2004	APD	The Role of Stock Markets in Current Account Dynamics: A Time Series Approach	Mercereau	N	N
2004	EUR	Boom-Boom Phases in Asset Prices and Fiscal Policy Behavior	Jaeger	N	N

Table A1. List of Research Reviewed (concluded)

Year	Issuing Department	Title	Author	Used in Operational Work (Y/N/C)[1/]	Potential Use In Operational Work (Y/N/C)
2004	PDR	Do Macroeconomic Effects of Capital Controls Vary by Their Type? Evidence from Malaysia	Tamirisa	N	N
2004	PDR	Liberalized Markets Have More Stable Exchange Rates: Short-Run Evidence from Four Transition Countries	Bulir	N	N
2004	RES	International Financial Integration and Economic Growth	Edison, Levine, Ricci, and Slok	N	C
2006	MCM	*GFSR* (Fall) Chapter 2: Household Credit Growth in Emerging Market Countries	www.imf.org/External/Pubs/FT/*GFSR*/2006/02/index.htm	N	N
2006	MCM	*GFSR* (Fall) Chapter 4: How Do Financial Systems Affect Economic Cycles?	www.imf.org/external/pubs/ft/*WEO*/2006/02/	N	C
2008	MCM	*GFSR* (Spring) Chapter 3: Market and Funding Illiquidity: When Private Risk Becomes Public	www.imf.org/External/Pubs/FT/*GFSR*/2008/01/index.htm	Y	C
2008	MCM	*GFSR* (Spring) Chapter 3: The Changing Housing Cycle and the Implications for Monetary Policy	www.imf.org/external/pubs/ft/*WEO*/2008/01/	Y	Y
2008	MCM	*GFSR* (Fall) Chapter 2: Stress in Bank Funding Markets and Implications for Monetary Policy	www.imf.org/external/pubs/ft/*GFSR*/2008/02/index.htm	Y	Y
2008	MCM	*GFSR* (Fall) Chapter 4: Financial Stress and Economic Downturns	www.imf.org/external/pubs/ft/*WEO*/2008/02/	Y	Y

[1] The last two columns of the table note whether the research item was used in the operational work examined (yes/no), and whether the research could potentially have been used for detailed macro-financial work (Y/N) or to give a “big picture” or conceptual (C) overview of macro-financial issues.

[2] Note caveat in text.

Table A2. IMF: Selected Issues Papers Issued, 2005–08
(By economy, with most recent first)

Title	Year	Author	Economy
“What Goes Up Must Come Down? House Prices in the United States”	**2008**	**Klyuev**	**US**
“A U.S. Financial Conditions Index: Putting Credit Where Credit Is Due”	**2008**	**Swiston**	**US**
“Credit Matters: Empirical Evidence on U.S. Macro-Financial Linkages”	**2008**	**Bayoumi and Melander**	**US**
“The Real Effects of the Subprime Mortgage Crisis”	**2008**	**Tong and Wei**	**US**
“House Prices and Regional Cycles in the United States”	**2008**	**Estevão and Barrera**	**US**
“Analyzing the Sources of (In)stability in the U.S. Banking Sector”	**2008**	**Capuano and Segoviano**	**US**
“Summary of Foreign Entanglements: Measuring Size and Source of Spillovers Across Industrial Countries”	2007	Bayoumi and Swiston	US
“Summary of the Ties That Bind: Measuring International Bond Spillovers Using Inflation-Indexed Bond Yields”	2007	Bayoumi and Swiston	US
“Summary of Globalization, Gluts, Innovation, or Irrationality: What Explains the Easy Financing of the U.S. Current Account Deficit?”	2007	Bayoumi, Tulin, and Balakrishnan	US
“New Landscape, New Challenges: Structural Change and Regulation in the U.S. Financial Sector”	2007	Bhatia	US
“Money for Nothing and Checks for Free: Recent Developments in the U.S. Subprime Mortgage Markets”	**2007**	Kiff and Mills	**US**
“Summary of U.S. Revenue Surprises: Are Happy Days Here to Stay?”	2007	Swiston, Mühleisen, and Mathai	US
“Summary of Applying the GFSM 2001 Framework to U.S. Fiscal Data”	2007	Rial and Gorter	US
“The Attractiveness of U.S. Financial Markets: The Example of Mortgage Securitization”	2006	Mühleisen	US
“Recent Trends in Labor Supply and Demand”	2006	Tsounta	US
“Challenges Facing the U.S. Electricity Sector”	2006	Tsounta	US
“Recent Oil Price Developments and the Performance of the U.S. Economy”	2006	Ouliaris	US
“U.S. Banking: Financial Innovation and Systemic Risk”	**2006**	**Bhatia**	**US**
“Structural Change and Competition Among Auto Manufacturers and Airlines”	2006	Swiston	US
“Boom-Bust Cycles in Housing: The Changing Role of the Financial Structure”	**2005**	**Schnure**	**US**
“Explaining the Labor's Share in National Income”	2005	Guscina	US
“Why Has the U.S. Trade Balance Widened So Fast?”	2005	Justiniano	US
“A Global View of the U.S. Investment Position”	2005	Krajnyák	US
“Consequences of Fiscal Consolidation for the U.S. Current Account”	2005	Swiston	US
“Should the Fed Adopt an Explicit Inflation Objective”	2005	Kumhof, Laxton, and Muir	US
“Diagnosing the High Cost of U.S. Medical Care”	2005	Kisinbay, Roger, and Stone	US

Table A2. IMF Selected Issues Papers Issued, 2005–08 (continued)
(By economy, with most recent first)

Title	Year	Author	Economy
“Effects of Social Security and Tax Reform in the United States”	2005	Ivaschenko	US
“Spillovers to Ireland”	2007	Kanda	Ireland
“Policy Challenges of Population Aging in Ireland”	2007	Botman and Iakova	Ireland
“Efficiency Gains of Private Sector credit Growth in Ireland”	2007	Tang	Ireland
“Favorable Fiscal Outturns: Is It Just of the Luck of the Irish?”	2005	Honjo	Ireland
“The Evolution of Unemployment in Ireland: The Role of Labor Market Policies”	2005	Iakova	Ireland
“Who Saves in Ireland? The Micro Evidence”	2005	Badia	Ireland
“Coordinating Fiscal Policy in Switzerland: Issues, International Experience and Prospects”	2006	Debrun	Switzerland
“An Indicative Public Sector Balance Sheet for Switzerland”	2006	Carare	Switzerland
“A Comparison of the Swiss, Dutch, and U.K. Pension Systems, with Emphasis on the Occupational Pension Pillars”	2006	Khamis	Switzerland
“Intertemporal Policy Consistency in Switzerland: Is the Current Social Insurance System Sustainable?”	2005	Gagales	Switzerland
“The Need for Health Care Reform”	2005	Braumann	Switzerland
“The Political Economy of Adjustment and Reform in Switzerland”	2005	Braumann	Switzerland
“Characteristics of the Swiss Financial System in International Comparison”	2005	Polan	Switzerland
“Fiscal Policy and the External Balance in Spain”	2006	Catalan, Lama	Spain
“Spain's Productivity: A Cross-Country Perspective”	2006	Escolano	Spain
“Fiscal Discipline At Lower Levels of Government: The Case of Spain”	2005	Spilimbergo	Spain
“Pension Reform in Spain: Macroeconomic Impact”	2005	Catalán, Hoffmaister and Guajardo	Spain
“The Swedish Fiscal Framework: Towards Gradual Erosion?”	2005	Balassone	Sweden
“The Tax-Benefit System and Labor Supply in Sweden”	2005	Tsounta and Bonato	Sweden
“Commodity Price Booms and Fiscal Policy Options in Australia”	2008	Hunt	Australia
“Australian Banks: Weathering the Global System”	**2008**	**Rozhkov**	**Australia**
“Australia's Large and Sustained Current Account Deficits: Should Consenting Adults be Trusted?”	2006	Mercereau	Australia
“Fiscal Policy and the Terms of Trade Boom”	2006	Rozhkov	Australia
“Australia's Adaptation to a Floating Exchange Rate”	2005	Beaumont, Cui, and Mercereau	Australia
“The Proposed Future Fund: An International Comparison”	2005	Cui	Australia
“Financial Integration in Asia: Estimating the Risk-Sharing Gains for Australia and Other Countries	2005	Mercereau	Australia

Table A2. IMF Selected Issues Papers Issued, 2005–08 (continued)
(By economy, with most recent first)

Title	Year	Author	Economy
"The Residential Property Market in Hong Kong SAR"	2006	Goldsworthy	Hong Kong
"The Implications of An Aging Population for Hong Kong SAR"	2006	Leigh	Hong Kong
"Sustainability of Volatile Fiscal Revenue Items"	2006	Gruenwald	Hong Kong
"Rapid Growth and Derivatives Markets in Hong Kong SAR"	2006	Leung and Gruenwald	Hong Kong
"Rising Prices, Slowing Growth, and the Implications for Monetary Policy"	2008	Lueth	Korea
"Stress Testing Household Debt in Korea"	2008	Karasulu	Korea
"Korea's Banking Sector: Liquidity Risk Management in the Face of Structural Trends and Deregulation"	**2008**	**Khatri**	**Korea**
"Lessons and Policy Recommendations from the Financial Crisis"	2008	Dodd	Korea
"What Determines Investment in Korea?"	2008	Syed	Korea
"Achieving Long-Term Fiscal Sustainability in Korea"	2007	Feyzioglu, Skaarup and Syed	Korea
"Korea's Competitiveness in the Global Marketplace"	2007	Chensavasdijai	Korea
"Internationalization of the Won"	2007	Semblat	Korea
"The Mortgage Market in Korea: Financial Risks and Development Needs"	**2007**	**Frydl**	**Korea**
"Korea's Transition to a Knowledge-Based Economy: Prospects and Challenges Ahead"	2006	Chensavasdijai	Korea
"A Financial Big Bang in Seoul? The Development of the Financial Sector in Korea"	2006	Semblat	Korea
"Restrictions on *Chaebol* Ownership of Korean Banks: Are They Warranted?"	2006	Frydl	Korea
"A Strategy for Restructuring the SME Sector in Korea"	2006	Kang and Kim	Korea
"A Family Divided—Revisited: Income Inequality and Social Polarization in Korea"	2006	Miniane	Korea
"Long-Term Fiscal Challenges"	2006	Feyzioglu	Korea
"Supply Shocks, Inflation, and Expectations"	2008	Batini	Chile
"What is the Optimal Level of International Reserves for Chile?"	2008	Tereanu	Chile
"An Evaluation of the Welfare Implications of Chile's Macroeconomic Framework"	2008	Batini, Levin and Pearlman	Chile
"How Do Changes in Global Liquidity Affect Chile?"	2008	Matsumoto	Chile
"Sovereign Wealth Funds: An International Perspective"	2008	Papaioannou, Walsh and Singh	Chile
"Deepening Chile's Capital Markets Through Global Integration"	2008	Arvai	Chile
"Chile's Structural Fiscal Surplus Rule: A Model-Based Evaluation"	2007	Kumhof, Laxton	Chile
"Risks in the Chilean Banking System: A Contingent Claims Approach"	**2007**	**Gray and Walsh**	**Chile**
"Credit Cyclicality: A Cross-Country Analysis"	**2007**	**Soderling**	**Chile**

Table A2. IMF Selected Issues Papers Issued, 2005–08 (continued)
(By economy, with most recent first)

Title	Year	Author	Economy
“Explaining Chile's Trade Performance”	2007	Monfort	Chile
“Toward Improving the Data and Procedures Used in Current Analysis and Forecasting at the Central Bank of Chile”	2006	Reifschneider	Chile
“Deepening Liquidity in Chilean Fixed-Income Markets”	2006	Holland	Chile
“Public Sector Debt and Market Development”	2006	Holland	Chile
“The Experience of Poverty Reduction in Chile”	2006	Walsh	Chile
“Macroeconomic Effects of EU Transfers in New Member States”	2008	Allard	Poland
“What Can Poland Learn From Other European Union Countries in Terms of Labor Market Reforms?”	2008	Annett	Poland
“Inflation in Poland: How Much Can Globalization Explain?”	2006	Allard	Poland
“Credit, Growth, and Financial Stability”	**2006**	**Tamirisa and Cihak**	**Poland**
“Assessing the Flexibility of the Polish Economy”	2006	Dalgic and Klingen	Poland
“What is Driving Investments in Poland?”	2005	Murgasova	Poland
“The Labor Market in Poland”	2005	Choueiri	Poland
“The Polish Pension Reforms After Six Years”	2005	Szekely	Poland
“The Fiscal Impact of the 1999 Pension Reform”	2005	Szekely	Poland
“Would Poland Benefit from a Fiscal Responsibility Law?”	2005	Szekely	Poland
“Assessing Croatia's External Stability”	2008	Ilyina	Croatia
“Efficiency of Government Social Spending in Croatia”	2008	Jafarov and Gunnarsson	Croatia
“Economic Growth in Croatia: Potential and Constraints”	2007	Moore and Vamvadikis	Croatia
“Economic Effects of Reducing the Size of the Government in Croatia: A Note Based on the IMF's Global Fiscal Model”	2007	Gueorguiev	Croatia
“External Debt and Balance-Sheet Vulnerabilities in Croatia”	**2007**	**Hilaire and Ilyina**	**Croatia**
“Bank Stability and Credit Risk in Croatian Banks”	**2007**	**Mitra**	**Croatia**
“Medium-Term Growth and Productivity in Estonia: A Micro Perspective”	2007	Badia	Estonia
“Population Aging and Fiscal Sustainability in Estonia”	2007	Skaarup	Estonia
“Competitiveness and Sustainability in Estonia”	2006	Lutz	Estonia
“Assessment of Balance Sheet Exposures in Estonia"	2006	Choueiri	Estonia
“International Investment Positions of New EU Member States: Stylized Facts and Influences”	2005	Lutz	Estonia
“Driving Forces of Inflation in the New EU8 Countries”	2005	Stavrev	Estonia

Table A2. IMF Selected Issues Papers Issued, 2005–08 (concluded)
(By economy, with most recent first)

Title	Year	Author	Economy
"Credibility Effects of Numerical Fiscal Rules: An Empirical Investigation"	2008	Debrun and Joshi	Hungary
"Monetary Policy Responses to Real and Portfolio Shocks in Hungary"	2008	Mitra	Hungary
"Hungary: Fiscal Risks from Public Transport Enterprises"	2007	Corbacho	Hungary
"Could Hungary's Growth Deceleration Persist? Inferring Productivity Trends from Consumption Volatility"	2007	Mitra	Hungary
"What Triggers the Reforms of Budget Institutions? The Case of Hungary in the European Context"	2007	Fabrizio	Hungary
"Common and Country-Specific Components in Currency Movements: What's Holding the Forint Up?"	2006	Abiad	Hungary
"Hungary: Developing a Medium-Term Fiscal Strategy"	2006	Corbacho	Hungary
"Employment Dynamics in the New European Member States: The Case of Hungary"	2006	Fabrizio	Hungary
"Growth and Current Account Performance: Results from a Cross-Country Model"	2005	Leigh	Hungary
"Fiscal Commitment: The Role of Budget Institutions"	2005	Fabrizio	Hungary
"Inflation Persistence and Monetary Policy"	2005	Leigh	Hungary
"Integration, External Imbalances and adjustment: The Latvian Experience"	2006	Gray and Miyajima	Latvia
"Labor Outflows, Capital Inflows and Income Convergence in Latvia"	2006	Brunner	Latvia
"Latvia: Aspects of the Rapid Credit Growth"	**2006**	**Prokopenko**	**Latvia**
"Macroeconomic Implications of Domestic Credit in Latvia"	**2005**	**Stavrev**	**Latvia**
"A Balance Sheet Approach to Macroprudential Vulnerabilities in Latvia"	**2005**	**Luna**	**Latvia**
"Romania's External Stability Risks"	2008	Atoyan	Romania
"Wage-Price Setting in Romania and Other EU New Member States"	2008	Goretti	Romania
"Retooling Romania's Budget Culture"	2008	Atoyan, Christou, and Moulin	Romania
"Modeling Monetary Policy in Romania"	2007	Tiffin	Romania
"Wage Dynamics in the Romanian Economy"	2007	Christou	Romania
"The Impact of Aging on the Public Sector in Romania"	2007	Klemm	Romania
"Competitiveness"	2006	Adler and Tiffin	Romania
"Credit Growth: Developments and Prospects"	2006	Tiffin	Romania
"Real Convergence Prospects"	2006	Justice, Paliu	Romania
"The Fiscal Impact of EU Accession"	2006	Justice	Romania

References

Bernanke, Ben, Mark Gertler, and Simon Gilchrist, 1996, "The Financial Accelerator and the Flight to Quality," *Review of Economics and Statistics,* Vol. 78, No. 1, pp. 1–15.

Central Bank of Ireland, 2010, "The Irish Banking Crisis: Regulatory and Financial Stability Policy 2003-2008," Report to the Minister for Finance by the Governor of the Central Bank, May 31. Available at: www.centralbank.ie/data/NewsFiles/The%20Irish%20Banking%20Crisis%20Regulatory%20and%20Financial%20Stability%20Policy%202003-2008.pdf.

Claessens, Stijn, M. Ayhan Kose, and Marco E. Terrones, 2008, "What Happens During Recessions, Crunches and Busts?" IMF Working Paper No. 08/274 (Washington: International Monetary Fund).

Crowe, Christopher, Simon Johnson, Jonathan Ostry, and Jeromin Zettelmeyer, 2010, *Macrofinancial Linkages: Trends, Crises, and Policies* (Washington: International Monetary Fund).

Kindleberger, Charles, 1978, *Manias, Panics, and Crashes: A History of Financial Crises* (New York: John Wiley and Sons, 3rd ed., 1996).

Laeven, Luc, 2008, "Pricing of Deposit Insurance," Chapter 3 in *Deposit Insurance,* ed. by Asli Demirgüç-Kunt, Edward Kane, and Luc Laeven (Cambridge, Massachusetts: MIT Press).

________, and Fabian Valencia, 2010, "Resolution of Banking Crises: The Good, the Bad, and the Ugly," IMF Working Paper No. 10/146 (Washington: International Monetary Fund).

Mishkin and others, 1999, "Report of the External Evaluation Committee on the IMF's Economic Research Activities," (Washington: International Monetary Fund).

Regling, Klaus, and Max Watson, 2010, "A Preliminary Report on the Sources of Ireland's Banking Crisis" (Dublin: Government Publications Sales Office). Ireland. Available at: www.bankinginquiry.gov.ie/Preliminary%20Report%20into%20Ireland%27s%20Banking%20Crisis%2031%20May%202010.pdf.

Reinhart, Carmen, and Kenneth Rogoff, 2009, *This Time is Different: Eight Centuries of Financial Folly* (Princeton, New Jersey: Princeton University Press).

Segoviano Basurto, and Pablo Padilla, 2006, "Portfolio Credit Risk and Macroeconomic Shocks: Applications to Stress Testing Under Data-Restricted Environments," IMF Working Paper No. 06/283 (Washington: International Monetary Fund).

Tamirisa, Natalia T., and Deniz O. Igan, 2008, "Are Weak Banks Leading Credit Booms? Evidence from Emerging Europe," IMF Working Paper No. 08/219 (Washington: International Monetary Fund).