

Technical Quality of IMF Research

47. This chapter examines the technical quality of the different types of IMF research. It assesses the soundness of the analysis and policy conclusions in different product lines, as well as the clarity of exposition in relation to their intended audience. The evaluation criteria were adjusted to take into account the different goals and intended audiences of each category (and product line) of research.¹⁵ Surveillance-oriented research was assessed on the basis of how well it explained the relevance of the policy issue being examined, the appropriateness of the analytical framework and data used to address the question posed, and the clarity of the policy conclusions. In addition, academic-style output was judged on the basis of whether it generated new knowledge or broadened the understanding of policy frameworks.

48. Overall, the evaluation found that the analytical chapters of the *WEO* and the *GFSR* as well as external publications were of high quality, while the quality of other product lines was mixed, with great variability within products and across themes. One common weakness was that policy conclusions were not always well linked to specific analysis, giving the impression that the IMF was mechanical in its policy recommendations and that it did not take into account changing circumstances or the features of different country groups. Authorities, staff, and other stakeholders considered the quality of IMF research to be at least as good as that of other international organizations, but views differed on whether it was at par with that of some mon-

etary authorities.¹⁶ All agreed that IMF research was not comparable with research produced in academia because it was much more policy oriented.

A. Surveillance-Oriented Output

World Economic Outlook

49. A peer review was conducted of the *WEOs*' 30 analytical chapters issued between 2004 and 2008 (see Kiguel, 2011). On average, *WEO* chapters were highly rated, particularly for the clarity of presentation and the technical analysis. The topics addressed were judged relevant as they focused on issues that dominated international policy discussions. Specific examples of strong analytical work included the discussions of inflation targeting in emerging market economies (September 2005), global imbalances (September 2005), and decoupling (April 2007). The country coverage achieved a good balance between developed and emerging market economies, but there was little coverage of ECF-eligible countries.

50. Some areas of weakness in the *WEOs* were identified. It was found that policy advice was often vague or too general to be of practical use, for example, arguing in general terms for fiscal adjustment, freer capital flows, and stable and predictable monetary policies. Also, conclusions were not always clearly drawn from the analysis, giving the impression of being message-driven.

¹⁵ The chapter is based on six background papers (summarized in Annex 4) that present the findings on technical quality from peer reviews conducted by external experts on each of the main product lines of research, as well as on a citation review of WPs. In addition, it presents the findings from semi-structured interviews and surveys of authorities, staff, and other stakeholders. The peer reviews focused on major qualitative dimensions of the research: the clarity of the questions posed, the appropriateness and proficiency of the technical analysis, whether the conclusions were firmly grounded in the analysis, and the policy relevance of the conclusions (see Annex 2 for a detailed discussion of the methodology).

¹⁶ Authorities considered the technical quality of IMF research better than that produced at the World Bank and the OECD, but not as good as research from the BIS, the U.S. Federal Reserve, or the ECB. Authorities from ECF-eligible countries rated IMF research favorably compared to institutions in their own countries, but those from advanced and large emerging market economies were more ambivalent. IMF staff considered that its research compared favorably to that of the OECD and World Bank and was at par with that of the BIS and ECB, but was not as good as that of the U.S. Federal Reserve.

51. Country authorities and staff also rated the *WEO* very highly, consistent with the peer review. These views were echoed by most external researchers interviewed. Staff attributed the higher quality of the *WEO*, relative to other research products, to the greater resources devoted to its production, to the contributions of external consultants, and to the thorough review to which it was subjected.

Global Financial Stability Report

52. A peer review of the *GFSR* examined the 20 analytical chapters written between 2004 and 2008 (see Kiguel, 2011). It found that the quality of the *GFSR* improved over time and that by the end of the evaluation period it was as good as that of the *WEO* chapters, as the content and analytical framework of the report improved. Still, the *GFSR* policy recommendations were often too numerous or not specific enough to be of practical use.

53. Country authorities, staff, and academics also rated highly the technical quality of the analytical chapters of the *GFSR*, only slightly below that of the *WEO*.

Regional Economic Outlooks

54. A peer review of the technical quality of the *REOs* reviewed all 44 of these publications issued from 2003 to 2009, focusing on analytical quality and exposition (see Montiel, 2011). Overall, the technical quality of *REOs* was assessed as being lower than for other publications, although their quality had improved over time. The quality of the analysis suffered because it was often based on pooled data from countries with very different circumstances. While many *REOs* were found to be insightful and well-grounded in empirical work, many more were judged to be too prescriptive and weakened by a tendency to advocate policies with little mention of options and trade-offs. There were also many instances of unsubstantiated claims, and missing or incoherent analysis.

55. Most of the country authorities, staff, and academics interviewed also found the quality of *REOs* to be much lower than that of the *WEOs* and other research outputs.

Selected issues papers

56. A peer review of the technical quality of SIPs examined a sample of 60 papers issued during

2004–08, taking two papers from each of 30 randomly selected countries (see Selowsky and Škreb, 2011). It found that a majority of these papers were good enough for the purpose they served, but that their quality varied widely. A significant number of papers were of high quality, but many were totally unsatisfactory. SIPs for advanced countries were better than those for emerging markets, and quality was lowest for ECF-eligible countries. Good SIPs addressed well-defined and relevant questions and showed familiarity with country context. The weak papers, on the other hand, showed limited knowledge of the country's basic institutional context and seemed to have been hurriedly prepared. Some SIPs applied quantitative techniques without explaining their appropriateness or discussing data-related and other limitations. Many used aggregate cross-country data, even when country-specific analysis would have been more appropriate.

57. The feedback on SIPs from different sources was somewhat inconsistent, partly reflecting the large dispersion in the quality of these papers. In interviews, most authorities said that the quality of SIPs varied widely. Many pointed to insufficient country context, and noted that SIPs tended to cite only other IMF research and did not acknowledge research done by local economists. In the survey, however, a majority of authorities rated the overall quality as “somewhat good” though weaker than for most other IMF research products. Similarly, in the survey a majority of staff rated the quality of SIPs as “somewhat good.” But in interviews, staff was much less positive, with some comparing SIPs to “term papers.” Staff indicated that the quality was affected by the fact that often SIPs had to be produced very quickly and that they needed to be closely aligned with the timing and policy directions of the bilateral surveillance process of which they were an integral part.

Occasional papers and staff position notes

58. The evaluation team reviewed a small sample of the IMF's other policy-oriented research, which included occasional papers, policy discussion papers, and staff position notes. Generally, these papers were found to be well written, articulating the policy relevance of the findings and providing advice to policymakers in simple, clear language. However, they sometimes lacked the analytical and empirical detail found in WPs and other academic-style products.

B. Academic-Style Output

Working papers

59. The evaluation conducted two peer review assessments of random samples of 60 WPs each, one on monetary frameworks and the other on fiscal revenues (see Kuttner and others, 2011 and Boadway and others, 2011, respectively).¹⁷ These panels found a wide dispersion in the quality of WPs. In both assessments, about 10 percent of the working papers reviewed received the highest rating, while about one-third of the papers was considered to be of low quality and 5 percent was rated unacceptable. The best WPs typically offered original or innovative findings and a critical assessment of the results and their robustness, and drew policy implications. They were well focused, included a thorough literature review, and used appropriate statistical techniques. The weaker WPs were a larger and more diverse group with a range of shortcomings. They lacked a coherent conceptual framework and in some instances used inappropriate empirical approaches. Many were superficial, had poor documentation, and lacked robustness checks. In some, the conclusions were not well grounded in the analysis and lacked appropriate caveats. WPs produced by the Research Department, FAD, and MCM were rated highest, while many of the weakest had previously been issued as SIPs.¹⁸

60. The evaluation also conducted a study comparing the publication and citation records of IMF WPs with those of a group of central banks and other international organizations (see Aizenman and others, 2011).¹⁹ More than one-third of IMF WPs were subsequently published in professional journals within three years of their issuance—similar to the share of pub-

lications in the comparator institutions. On average, 40 percent of IMF WPs received citations within the comparator group and 60 percent received citations overall. The number of citations received by each paper varied widely. Though about 40 percent of the IMF working papers were not cited, some of those cited received a large number of citations—again a pattern similar to that of WPs issued by comparator institutions and academia. Excluding self-citations (i.e., citations in other publications from within the same institution), IMF WPs were cited more often than those of other international organizations, but not as often as those of the various U.S. Federal Reserve banks.²⁰ During the review period, there was an increase in the number of citations for the most cited IMF working papers.

61. For WPs as for SIPs, the feedback from the surveys and interviews was somewhat inconsistent, once again most likely reflecting the wide dispersion of quality. More than half of the country authorities responding to the survey rated the technical quality of WPs as “very good.” However, in interviews both authorities and external researchers expressed much more negative views; both groups reported a significant unevenness of quality. Some observed that it was hard to compare IMF WPs to papers produced by academic researchers because the IMF papers focused on policy issues that are often difficult to model. In some ECF-eligible countries and emerging markets, authorities and local researchers believed that IMF WPs were too technical. At the same time, some academics, especially from advanced economies, noted that the IMF’s empirical WPs often lacked a coherent conceptual and theoretical framework. These papers tended to use reduced-form regression analysis where the variables were loosely linked to theory, making the results difficult to interpret.

62. Staff were more critical of the technical quality of WPs. Only 20 percent of survey respondents rated WPs as “very good” while about 15 percent rated them as “very poor” or “somewhat poor.” Negative views were also expressed in interviews.

63. Ensuring high and consistent quality of WPs is more important for the IMF than for academic and other institutions because, as mentioned above, most country authorities and other readers saw the IMF’s

¹⁷ These topics were selected because they are at the core of the IMF mandate and expertise. Each review was conducted by a panel of three academics with recognized expertise on the corresponding topic.

¹⁸ During the period under review, there was a gradual increase in the number and share of WPs prepared by area departments which was linked to a perception among staff that producing WPs had become an important element for promotion. This increase led to an increase in the quality dispersion of WPs.

¹⁹ Such comparisons are a common tool for measuring the quality of research. Data on the citation of WPs were obtained from the RePEc project and from Google Scholar. All RePEc information is freely available from their website (www.repec.org). The benchmark institutions were: Bank of Canada, U.S. Federal Reserve Board, 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. The methodology used in this background study was similar to the Bank of Canada study by St-Amant and others (2005), which investigated the relevance and utilization of central bank research.

²⁰ The citation count does not include publications in developing countries, particularly in languages other than English, and thus underestimates the citations of research produced by international organizations relative to the U.S. Federal Reserve. Also, the frequency of IMF papers to cite other IMF work was at par with other international organizations, but higher than for the U.S. Federal Reserve.

WPs as final outputs and as broadly representing the views of the IMF. Many IMF staff also reported that they saw WPs as final outputs that they did not intend to revise nor submit for publication in external journals.

IMF Staff Papers and external publications

64. Papers published in *IMF Staff Papers* and external professional journals were of high technical

quality—not surprisingly since these papers had undergone a refereed review by the corresponding journals. However, except for studies published in *IMF Staff Papers*, the IMF did not get much recognition among authorities and other stakeholders for the research in these external publications, because most officials did not follow professional journals and often, when they did, they did not focus on the author’s affiliation. A majority of these papers had previously appeared as WPs or as *WEO/GFSR* chapters.