Cote du document:	EC 2020/109/W.P.4
Ordre du jour:	5
Date:	22 mai 2020
Distribution:	Publique
Original:	Anglais

F

Investir dans les populations rurales

Analyses de sensibilité de l'évaluation de l'impact concernant FIDA10 et incidences sur FIDA11

Note à l'intention des membres du Comité de l'évaluation

Responsables:

Questions techniques:

Sara Savastano
Directrice
Division recherche et évaluation de l'impact
téléphone: +39 06 5459 2155
courriel: s.savastano@ifad.org

Alessandra Garbero
Économiste principale
téléphone: +39 06 5459 2458
courriel: a.garbero@ifad.org

Transmission des documents:

Deirdre Mc Grenra
Cheffe
Gouvernance institutionnelle et relations avec les États membres
téléphone: +39 06 5459 2374
courriel: gb@ifad.org

Comité de l'évaluation — Cent neuvième session
Rome, 19 juin 2020

Pour: **Information**

I. Contexte

1. Lors de sa cent vingt-septième session, le Conseil d'administration a examiné le Rapport d'évaluation de l'impact dans le cadre de la Consultation sur la Dixième reconstitution des ressources du FIDA (FIDA10), ainsi que les commentaires du Bureau indépendant de l'évaluation du Fonds (IOE). Le Conseil a remercié la direction d'avoir procédé à cette évaluation dont elle a accueilli les résultats avec intérêt, mais a aussi noté que la direction devait étudier les insuffisances de la méthode actuelle et s'employer à l'améliorer. Il lui a demandé, plus précisément, de soumettre la méthode employée à une revue par les pairs et de la renforcer avec l'appui d'un expert extérieur, et il lui a suggéré de la présenter pour examen, en même temps que la méthode d'échantillonnage, avant de procéder à l'évaluation de l'impact dans le cadre de FIDA11.

2. La direction a recruté à cette fin un expert[1], qu'elle a chargé de deux missions: évaluer la méthode utilisée pour l'établissement du Rapport d'évaluation de l'impact dans le cadre de FIDA10 de manière à déterminer si la sélection des projets constituant l'échantillon avait biaisé les résultats; et, sur la base de cette évaluation, confirmer la méthode de sélection pour l'évaluation de l'impact dans le cadre de FIDA11.

3. Étant donné le volume et la complexité des données de l'évaluation de l'impact dans le cadre de FIDA10, des analyses de sensibilité détaillées ont été consacrées à la démarche suivie pour réaliser cette évaluation. Les résultats des analyses et la validation de l'échantillon établi pour FIDA11 sont présentés dans l'appendice au présent document.

4. Le FIDA a procédé à la sélection de l'échantillon suivant le protocole approuvé par le Conseil d'administration pour le Cadre relatif à l'efficacité en matière de développement[2]. Comme l'a suggéré le Conseil par la suite, cette opération a été complétée par une analyse de sensibilité visant à tester la robustesse de l'échantillon. L'analyse de sensibilité a établi que les biais de sélection éventuels étaient négligeables et, par conséquent, que les résultats présentés dans le Rapport d'évaluation de l'impact dans le cadre de FIDA10 étaient valides. La valeur ajoutée par la méthode suivie pour l'établissement des rapports de l'institution l'emportait sur les biais qui pouvaient exister, et ces derniers ne devraient pas remettre en cause les efforts déployés aux fins de l'établissement des rapports du FIDA qui conféraient un caractère exceptionnel à l'institution. Cette même méthode servira à valider l'échantillon utilisé pour FIDA11. La validité de l'approche suivie par le FIDA pour évaluer les rapports institutionnels a été confirmée.

5. L'IOE et le Conseil d'administration ont eu raison de soulever la question d'un biais éventuel dans l'échantillon des projets. Il s'agit là d'une préoccupation légitime pour tous les échantillons, mais surtout lorsqu'il est impossible de suivre un processus d'échantillonnage aléatoire. La nécessité de considérer la possibilité d'un biais au stade de la sélection des projets est une leçon fondamentale tirée des analyses de sensibilité. La procédure de sélection utilisée pour FIDA11 étant conforme au protocole accepté par le Conseil, les résultats seront validés au moyen d'une analyse de sensibilité. D'autres méthodes seront considérées pour FIDA12.

6. Selon les conclusions des analyses, la méthodologie utilisée pour FIDA10 est valide. Il est de surcroît raisonnable d'utiliser le processus de sélection de l'échantillon – qui suit le protocole du Cadre relatif à l'efficacité en matière de

[1] Stefano Gagliarducci, professeur d'économie, Département de l'économie et des finances de l'Université de Rome Tor Vergata, et adjoint de recherche à l'Institut Einaudi de l'économie et des finances. Il a publié des articles dans des revues économiques de premier plan et a procédé à des travaux de recherche sur les biais de publication.

[2] Voir le Cadre du FIDA relatif à l'efficacité en matière de développement – Conseil d'administration, décembre 2016 (paragraphe 58).

développement – pour les futures activités d'évaluation de l'impact. Cette approche n'entraîne aucun risque de réputation pour le FIDA.

II. Résumé des constatations

7. Un certain nombre d'analyses systématiques, conçues pour détecter la présence éventuelle de biais dans les résultats des évaluations de l'impact institutionnel, ont été réalisées. Les biais pouvant exister concernent l'échantillon de projets sélectionnés pour la période de FIDA10 et la mesure dans laquelle ces projets sont représentatifs du portefeuille de projets clôturés durant cette même période. Le FIDA réalise des évaluations de l'impact de manière à promouvoir la responsabilité et l'apprentissage, de sorte que la méthode employée doit impérativement respecter les principes de représentativité, de rigueur et de transparence.

8. Les conclusions des analyses peuvent être récapitulées comme suit:

i) **Le biais dont peut être entachée la sélection des projets pour l'évaluation de l'impact dans le cadre de FIDA10 est négligeable.** L'échantillon sur lequel a porté l'analyse de l'impact dans le cadre de FIDA10 a donné lieu à une évaluation systématique conçue pour déterminer la présence éventuelle d'un biais de sélection ainsi que la nature, l'orientation et l'ampleur de ce biais. Toutes les variables qui auraient pu influencer la sélection ont été examinées dans cette optique, notamment les notes de la performance au niveau de l'exécution, c'est-à-dire les notes disponibles au stade de la sélection (juillet 2016), qui sont les seules variables susceptibles d'influencer le processus. Des tests ont été appliqués pour déterminer l'existence d'écarts significatifs sur le plan statistique entre l'échantillon constitué de 15% des projets (soit 19 projets) et le reste du portefeuille (88 projets) pour l'univers de 107 projets achevés durant FIDA10, et 24 notes de performance au niveau de l'exécution. Le test n'a détecté aucun biais pour la grande majorité des variables et a déterminé que les écarts entre les notes moyennes ne sont significatifs que pour seulement deux variables: i) le taux de décaissement, ii) les fonds de contrepartie. Une opération approfondie de validation a donc été réalisée pour ces deux variables de manière à détecter la présence éventuelle de biais.

Des méta-analyses des sous-groupes ont donc été effectuées dans le but de déterminer si l'ampleur de l'impact, tel que mesuré dans les évaluations au niveau des projets, était associée à la classe de notation de la performance de l'exécution (notes 1 à 6) du projet. Ces analyses permettent de vérifier s'il existe une association entre l'ampleur de l'impact et les notes attribuées. On pourrait en effet s'attendre à ce que les projets jugés satisfaisants sur la base des taux de décaissement et des fonds de contrepartie aient des impacts plus importants que les projets notés modérément satisfaisants ou non satisfaisants. De tels résultats auraient indiqué l'existence d'une relation positive entre la performance et l'impact. Les observations de la direction ne permettent toutefois pas d'établir l'existence d'une relation claire entre la notation et les estimations de l'impact, en particulier en ce qui concerne la note attribuée sur la base des taux de décaissement.

L'objectif stratégique d'accès aux marchés offre un bon exemple en ce domaine: les projets dont la performance en matière de décaissement a été jugée satisfaisante ou mieux (notes 1 à 3) ont l'impact le plus faible (57% contre 80% pour les projets notés modérément satisfaisants) tandis que les projets notés insatisfaisants enregistrent les impacts les plus forts (89%). Les résultats sont pratiquement analogues pour d'autres objectifs stratégiques. Dans le cas des capacités de production et de la variable de notation des fonds de contrepartie, les projets notés modérément satisfaisants (note 4) affichent un impact de 18% par comparaison aux projets notés non satisfaisants (notes 1 à 3). La direction a conclu, sur la base de cette

évaluation, qu'une note élevée – une performance plus solide – ne correspondait pas à un impact plus important.

ii) **Les corrections effectuées au titre du biais de sélection ont montré que ce biais était marginal, ce qui signifie que les résultats présentés dans le Rapport d'évaluation de l'impact dans le cadre de FIDA10 sont valides.** Une deuxième étape a été conçue dans le but de valider plus précisément ces résultats grâce au recours à deux méthodes supplémentaires, qui ont pour objet de déterminer s'il est nécessaire d'ajuster les résultats des méta-analyses au titre d'un éventuel biais de sélection. La première méthode – à savoir la correction du biais de sélection au moyen du modèle d'Heckman – donne lieu au calcul de la probabilité de chaque projet d'être sélectionné aux fins de l'évaluation de l'impact dans le cadre de FIDA10 et, si ledit projet est sélectionné, à l'ajustement des estimations de l'impact institutionnel au titre de ce biais. Le calcul de la probabilité prend en compte des facteurs de sélection tels que la performance relative aux décaissements, aux fonds de contrepartie et à d'autres variables importantes pour laquelle des observations peuvent être réunies[3]. La deuxième méthode dite "supprimer et remplacer" (*trim-and-fill*)[4] se fonde sur les études des méta-analyses et a initialement servi à vérifier la présence d'un biais dû à l'exclusion délibérée de certains projets ou études dans le cadre d'une méta-analyse. Cette méthode présente un inconvénient bien connu, qui tient au fait qu'elle est susceptible de corriger un biais de publication qui n'existe pas, en sous-estimant l'ampleur des effets (les résultats, qui sont peu fiables, sont présentés dans l'appendice)[5]. La première méthode a permis d'établir que le biais de sélection n'était que marginalement significatif (au seuil de 10%) dans le cas de l'objectif stratégique de l'accès au marché. Les estimations de l'impact corrigées au moyen de la méthode d'Heckman ont montré que, toutes choses étant égales par ailleurs, l'impact était dans ce cas surestimé de moins de 15%[6].

iii) **Les projets sélectionnés pour constituer l'échantillon de l'évaluation de l'impact dans le cadre de FIDA11 ne seront pas source de biais.** La même approche a servi à valider l'échantillon relatif à FIDA11. Étant donné les résultats produits par les analyses de sensibilité effectuées pour FIDA10, il a été décidé d'appliquer une méthode similaire pour valider l'échantillon retenu pour les évaluations de l'impact dans le cadre de FIDA11 et de déterminer s'il existait un biais de sélection au stade de la sélection des projets (juillet 2018) en considérant plusieurs caractéristiques de l'univers du portefeuille, notamment les notations de la performance. Le processus suivi pour le programme d'évaluations de l'impact dans le cadre de FIDA11 a donné lieu à la sélection de 24 des 112 projets réalisés en vue de la poursuite d'une évaluation d'impact rigoureuse; ces projets constituaient 21,4% du portefeuille, 20,9% du montant total des financements et 25,6% des financements du FIDA. L'examen a fait intervenir, comme les analyses de validation effectuées pour FIDA10, 24 notes de performance au niveau de l'exécution ainsi qu'un certain nombre d'autres aspects objectifs du portefeuille (nombre de bénéficiaires et financements, par exemple) de manière à déterminer l'existence d'un éventuel biais de sélection. Aucun écart significatif sur le plan statistique n'a été détecté entre l'échantillon de projets retenus (24) et les autres projets de l'univers (88 projets clôturés durant

[3] Cette démarche a été poursuivie conjointement à la réalisation de méta-régressions et de méta-analyses.
[4] La méthode dite "*trim-and-fill*" qui consiste à supprimer et à remplacer des éléments est un moyen couramment employé pour détecter et corriger les biais de publication.
[5] Terrin N, Schmid CH, Lau J, Olkin I (2003) *Adjusting for publication bias in the presence of heterogeneity. Statistics in Medicine* 22: 2113-2126.
[6] Cette surestimation est de 2% pour la mobilité économique, de 15% pour l'accès au marché, de 10% pour la production et de 6% pour la résilience. Les résultats relatifs à la nutrition ne sont pas modifiés.

FIDA11). Seule la performance des systèmes de suivi et d'évaluation s'est révélée significative sur le plan statistique. Ces résultats montrent que l'échantillon de FIDA 11 n'est pas non plus entaché d'un biais de sélection.

iv) En conclusion, la direction peut confirmer, sur la base de ces validations statistiques et de ces analyses de sensibilité, que les échantillons de projets sélectionnés pour les évaluations de l'impact institutionnel ne sont entachés d'aucun biais de sélection, que ce soit dans le cadre de FIDA10 ou dans celui de FIDA11.

Appendix: Peer review of IFAD10 Impact Assessment Methodology

Stefano Gagliarducci, University of Rome Tor Vergata and EIEF

In collaboration with Alessandra Garbero and Sara Savastano, IFAD

Contents

1. Introduction

During the discussion of the IFAD10 Impact Assessment Report at the 127th Executive Board meeting on September 11, 2019, the Board recommended to conduct Sensitivity analyses to assess the robustness of the corporate impact estimates and verify the results. This recommendation was made in light of the comments received by the Independent Office of Evaluation (IOE) and other Stakeholders, indicating the possible presence of a bias in the meta-analysis estimates and projections, raising concerns around the credibility of the findings. Such bias concerned the choice of the IFAD10 sample of projects selected for an Impact Assessment (IA) and notably, that such projects are, according to IOE, not representative of the portfolio of projects completing during IFAD10.

IOE's argument was based on a descriptive analysis of the performance ratings[7] at completion (or project completion reports ratings, in brief PCR). Their conclusion was that the projects selected for an impact assessment during IFAD10, seemed to include a large percentage of higher performing ones, therefore "potentially" yielding "biased" estimates of impact and possibly implying an overly optimistic vision of IFAD10 aggregate impact performance.

Systematic sensitivity analyses were therefore conducted to assess whether bias existed, and then investigate its magnitude. This is justified on transparency grounds, and on the fact that IFAD strongly believes in demonstrating accountability and learning, through rigorous methods.

Broadly speaking, sensitivity analysis is a process that allows the analyst to prove that the findings from a meta-analysis are not dependent on arbitrary or unclear decisions. In practice, they are aimed at repeating the meta-analysis, substituting alternative decisions or ranges of values for decisions that were arbitrary or unclear. For example, if the eligibility of some studies in the meta-analysis is dubious because they do not contain full details or are not representative, sensitivity analysis may involve undertaking the meta-analysis twice: first, including all studies and second, only including those that are definitely known to be eligible/representative. In this context and, through a weighting procedure, sensitivity analysis address the robustness of the results to the explicit inclusion of selection bias into the estimates, whereby this bias is assumed to be originated by the inclusion of a large number of projects with high performance ratings at completion, in the sample of the IFAD10 IAs.

The document presents the results of these analyses and is structured as follows. Section 3 recapitulates the background of IFAD approach to corporate reporting as stated in the Development Effectiveness Framework. Section 4 first introduce some definitions, section 5 presents a descriptive analysis, section 6 a literature review on the possible strategies to address bias, section 7 the results from the sensitivity analyses, section 8 concludes on IFAD10 and section 9 presents some implications for IFAD 11 and the corresponding validation of the IFAD11 sample of impact assessments.

2. Background

IFAD carries out project-level impact assessments (IAs) on a selection of projects (about 15 per cent) that are representative of the portfolio, to be able to measure corporate impact or aggregate development effectiveness. The latter requires a methodology that can attribute IFAD impact at the corporate level, e.g. provide an estimate of aggregate impact for the corporate indicators laid out in the IFAD Strategic Framework 2016-2025. The approach used is systematic, comprehensive, transparent, and builds upon the

[7] Since 2005, in line with the practice adopted in many other International Financial Institutions (IFIs) and United Nations organizations, IOE uses a six-point subjective rating system (where 6 is the highest score and 1 the lowest score) to evaluate projects. In addition to reporting on performance based on the six-point rating scale, in 2007 IOE introduced the broad categories of "satisfactory" (rating coded 4 to 6) and "unsatisfactory" (rating coded 1 to 3) for reporting on performance across the various evaluation criteria.

IFAD9 Impact Assessment Initiative methodology as well as the IFAD10 Development Effectiveness Framework.

IFAD's Development Effectiveness Framework (DEF)[8], approved by the Board in September 2016 lays out the selection protocol to assess projects suitability to undergo an impact assessment, specifying to the following criteria:

(i) potential to learn lessons;
(ii) feasibility of conducting a scientifically rigorous impact assessment;
(iii) buy-in from the government and IFAD;
(iv) the capacity of a project to represent IFAD's portfolio and
(v) the relevance of the impact assessment for subsequent project phases.

A key factor of impact assessment, in addition to accountability, is learning; and learning needs to inform the design of new projects in the same country or elsewhere. This provides a public good for policymakers. Therefore, a major recommendation approved by the Board - in the Development Effectiveness Framework -– stated that "impact assessments should have been selected and structured to facilitate and maximize learning while recognizing the need for corporate reporting, and that an impact assessment agenda should be a multi-stakeholder and participatory process to ensure relevance" (IFAD, 2016[9] pag.1).

Consequently, projects selected for IFAD10 IAs had to both display the potential for learning (innovative approaches or a clear evidence gap), while maintaining feasibility and have buy-in from the government.

In order to allow for adherence to the IFAD10 selection protocol, a working group was created to ensure that the selected projects were representative of the portfolio and revealed gaps for additional assessments, with a view to gaining an understanding of how projects fit into the portfolio. The expectation was that selected projects would have ultimately reflected the thematic and regional coverage of IFAD projects.

This led to a participatory process, finalized in September 2016, whereby projects selected for impact assessments were chosen in collaboration with IFAD's regional divisions to maximise this learning criteria. The divisions provided a list of projects suitable for inclusion based on the criteria specified according to the selection protocol. Subsequently, an appraisal was done to determine the impact assessments' feasibility in consultation with the regional divisions and relevant country directors.

Concerning corporate-level impact, IFAD's methodology to estimate aggregate development effectiveness involves a two-steps procedure whereby a meta-analysis of individual project-level impact assessment estimates is conducted in the first stage to compute aggregate corporate impacts, and a projection is conducted in the second stage to extrapolate impacts to the rest of the portfolio and estimate number of people benefiting across the portfolio[10].

[8] The DEF was developed based on the lessons learned from the experience in demonstrating impact as part of the IFAD9 Impact Assessment Initiative. See EB 2016/119/R.12
[9] International Fund for Agricultural Development (IFAD), 2016. Development Effectiveness Framework (EB 2016/119/R.12).
[10] As far as the projection approach is concerned, this refers to a methodology that allows the estimated impact to be extrapolated to the whole IFAD portfolio, in order to obtain an assessment of the number of people that have benefited from IFAD investments. The corporate impact is interpreted as percentage change gain in each of the Strategic Objectives (SOs) and on IFAD's overarching goal. To translate this into the number of beneficiaries who benefited from IFAD's investments, distributional assumptions are needed to extrapolate the corporate estimates to the universe of beneficiaries in the portfolio.. The IFAD10 projection universe includes 107 projects, and is defined as the total number of projects completing during the replenishment period (2016-2018). As the projection require estimates of beneficiaries reached across the whole universe, the additional challenge has been to aggregate the number of beneficiaries for the overall portfolio. The information on the number of beneficiaries in the IFAD10 portfolio can be extracted from project documentation and IFAD internal reporting systems. Projected beneficiaries impacted are calculated based on the number of actual beneficiaries belonging to the universe of 107 projects. The latter amount to around 65.3 million beneficiaries. At the basis of the extrapolation, there are two main assumptions. One concerns the distribution of impacts, where the assumption is that corporate impacts are normally distributed with means and standard errors corresponding to the ones estimated empirically while obtaining aggregate impact estimates from the 17 impact studies covering 19 projects (equivalent to 18 per cent of the universe, actually). The second

The aggregation is systematically done via a meta-analysis, a statistical procedure for combining data from multiple studies. Meta-analysis was pioneered in medical studies in the late seventies and then exponentially applied to clinical research. The meta-analysis is a study design used to systematically assess previous research studies to derive conclusions about a specific drug/treatment/research (or in our case policy) question. Outcomes from a meta-analysis may include a more precise estimate of the effect of treatment or risk factor for disease, or other outcomes (Haidich, 2010)[11]. More broadly, meta-analysis is defined as "the statistical analysis of a large collection of results for the purpose of integrating the findings" (Glass 1976[12]). In other words, it is "a quantitative summary of statistical indicators reported in similar empirical studies" (Brander et al. 2006[13]).

In the context of IFAD10 IA, the meta-analysis is a statistical procedure that aggregate the results of the 15% of projects on which an individual study is conducted. The outcome of the analysis is a proxy for an average effect (the treatment effect, or effect size) of the impact of IFAD's overall portfolio. Once aggregated, corporate impacts were computed as percentage changes over the comparison group for each Strategic Objective (SO), notably production, market access or participation, and resilience, and for the overall IFAD goal of increased economic mobility.

3. Definitions

In this section, some definitions that are going to be useful for an understanding of the remainder of the document are provided.

3.1 Representativeness

First, the concepts of representativeness, population or universe, and sample is defined: a representative sample is one that matches some characteristic of the underlying population, usually the characteristic that one is targeting with the research. In the context of IFAD10, the population refers to the population of projects that are in the universe of projects completing during IFAD10 (around 107 projects). The sample under analysis is defined as the 19 projects chosen for an impact assessment study.

As mentioned before, the IFAD10 selection protocol for the 19 IFAD10 Projects was based on a number of criteria to ensure representativeness of the portfolio. To what extent such criteria were as good as random is open to question. They could be quasi-random, in the sense that ex-ante, or at the time of selection, it was not possible to ascertain all of them.

In practice, it is almost impossible to ensure randomness due to a number of factors, such as feasibility of the impact assessment itself, knowledge asymmetries, political considerations and stakeholders buy-in, among others. It is worth recalling that the issue

assumption is about defining what benefiting means in terms of exceeding a certain threshold. The projected number of beneficiaries impacted by IFAD's investments can be obtained by setting a threshold of at least 20 per cent for impact gains. Using estimates on the aggregate impacts and knowledge of the portfolio, one can then obtain projected number of beneficiaries benefiting above a 20 per cent threshold. In summary, projected beneficiaries impacted are obtained by randomly drawing a normal distribution of impacts with means and standard errors centred to the ones empirically estimated from aggregate impact distributions, thereby assuming that benefits are randomly and normally distributed and are above a specific threshold.

[11] *"Important medical questions are typically studied more than once, often by different research teams in different locations. In many instances, the results of these multiple small studies of an issue are diverse and conflicting, which makes the clinical decision-making difficult. The need to arrive at decisions affecting clinical practice fostered the momentum toward "evidence-based medicine. Evidence-based medicine may be defined as the systematic, quantitative, preferentially experimental approach to obtaining and using medical information." Haidich AB, 2010, Meta-analysis in medical research. Hippokratia. 2010 Dec.14 (Suppl 1):29-37.*

[12] Glass, G. (1976). Primary, Secondary, and Meta-Analysis of Research. Educational Researcher, 5(10), 3-8.

[13] Brander L.M., et al. (2007). The recreational value of coral reefs: A meta-analysis, Ecological Economics, Vol. 63, Issue 1, 2007, 209-218.

of representativeness of the impact assessment sample was also raised during the previous replenishment cycle (IFAD9). Maintaining the integrity of the random selection conducted during IFAD9, was extremely difficult due to the above-mentioned factors. In that instance, projects were selected according to a number of criteria: 1) feasibility (suitable for an ex post impact assessment); 2) with the overarching aim of measuring the poverty reduction impact and, 3) statistically representative of the portfolio of activities undertaken by IFAD during IFAD9.

Therefore a representative sample of projects to be evaluated was determined by drawing a stratified random sample (a total of 41 projects, i.e. 26 first-choices and 15 reserves) from the universe of projects (with available datasets) closing between 2010 and 2015.

However, maintaining the integrity of the random sample proved difficult, as some randomly selected ones had to be replaced owing to both political and practical concerns (conflict setting, absence of PMU or key informants essential to gather retrospective information about projects, and impossibility to determine a counterfactual, among others). An internal consultation (in 2012) with IFAD Regional Directors and divisional representatives was then conducted to endorse the list of randomly selected projects to be evaluated by the external research partners. This process led to the replacement of 11 randomly selected projects with a set of purposively selected ones (purposive evaluations), given their strategic relevance and overall performance across the portfolio. Two of the purposively selected projects were dropped (namely, those in India and Senegal) after discussing the feasibility with internal staff. This last factor showed that even for "cherry picked" projects feasibility of an impact assessment was not guaranteed.

Notwithstanding these issues, IFAD sought to maintain the integrity of the representative sample and decided to maintain the randomly selected projects excluded from the final list of ex post evaluations and conduct the ex post assessments in-house with secondary datasets (14 Shallow dives).

Regarding IFAD10, and as noted above, a selection protocol was followed to ensure representativeness of the portfolio. The rationale for using a protocol is similar to what is normally conducted in the medical field, which is to randomly assign patients into treatment and control groups. As such, these protocols have features of quasi-randomness – as patients are selected into treatment – across a population of eligible patients[14].

3.2 Sampling/selection bias

Selection bias is problematic because it is possible that a statistic computed of the sample is systematically erroneous. Selection bias can lead to a systematic over- or under-estimation of the corresponding parameter in the population. Selection bias occurs in practice as it is practically impossible to ensure perfect randomness in sampling (see before). If the degree of misrepresentation is small, then the sample can be treated as a reasonable approximation to a random sample. Also, if the sample does not differ markedly in the quantity being measured, then a biased sample can still be a reasonable estimate.

Selection bias is mostly classified as a subtype of selection bias, sometimes specifically termed sample selection bias, but some classify it as a separate type of bias. A distinction,

[14] There is currently a heated debate around the topic of randomized controlled trials and whether they should be generally considered the gold standard, namely the best method to infer causality. It is worth recalling that in medical studies, researchers often choose not to randomize the intervention for one or more of the following reasons: (1) ethical considerations, (2) difficulty of randomizing subjects, (3) difficulty to randomize by locations (by region in the case of IFAD portoflio), (4) small available sample size (Harris et al. 2006).

albeit not universally accepted, of selection bias is that it undermines the external validity of a test (the ability of its results to be generalized to the entire population), while selection bias mainly addresses internal validity for differences or similarities found in the sample at hand. In this sense, errors occurring in the process of gathering the sample or cohort cause selection bias, while errors in any process thereafter cause selection bias. In this specific case, it refers to selection bias.

However, selection bias and selection bias are often used synonymously.

4 Descriptive analysis and selection bias

In this section, a first assessment of the presence of selection bias is provided by presenting descriptive statistics that characterize the universe of IFAD10 projects, compared to the sample chosen for impact assessments. These descriptives are essential to understand the extent and the severity of the bias based on observable features. In presence of a selection bias, one should expect the two groups to differ significantly over an array of observable dimensions.

Recall that the sample is made of 19 projects, and that the universe is composed of 107 IFAD10 projects slated to close during IFAD10 at the time of selection. After projects were selected for assessment, a subset of the selected projects were extended such that their closing dates are now in IFAD11[15].

The main argument against lack of representativeness cited by IOE was that the sample of IFAD10 IAs included a large majority of high performing projects as displayed by IOE's analyses of performance indicators at completion (PCR ratings).

Thus, Table 1 reproduces the one presented by IOE in their summary document for the Evaluation Committee Session (EC) held in September 2019. Average performance indicators at completion (PCR ratings) are displayed for the sample of 19 projects evaluated as part of IFAD10 IAs and the remaining 88 projects not evaluated out of the total universe of 107 projects.

PCR ratings at completion are subjective ratings with a six-point measurement scale system ranging from 6 to 1, with 1 being the lowest score across each criterion. At the time of IFAD10 selection, when the sample of 15% of projects was identified, none of these scores were available for consultation nor officially available within IFAD official system[16].

While comparing the two tables, a number of issues became apparent. The first, is the lack of definition of the universe of projects analysed e.g. the total number of observations (projects), in IOE's document. As a consequence statistics for unselected projects completing during IFAD10 (columns 3 and 4) varied across some indicators, while results for the IFAD10 sample (namely columns 1 and 2) coincided with IOE calculations. Also, PCR scores were not available for all the unselected projects, therefore the total number

[15] Notably Bangladesh (CCRIP), Kenya (SCDP), Sao Tome (PAPAC), Rwanda (PRICE) will now close in IFAD11.

[16] Specifically, regarding features that might have driven the IFAD10 IAs selection process, and alter the representativeness of the sample of the IFAD10 projects portfolio, the following ones were available at the time of the selection, notably in 2016: the project type or sector, the region of implementation, the size of outreach, the disbursement performance, and the implementation performance indicators. As projects were ongoing, project completion report ratings (the one verified by IOE) were not available to inform the selection.

of observations by indicator varies between 64 to 88 projects (column 3). Last, and similar to IOE's table, final PCR ratings are only available for 13 of the 19 projects that underwent an IA. The unavailability of the PCR ratings for the whole IA sample, is due to the fact that PCRs cannot be finalized if projects completion dates are extended. This was the case of 6 projects out of 19, whereby their completion dates were extended into IFAD11.

Therefore, Table 1 shows the difference across PCR ratings as presented by IOE in their comments. T-tests were run for the statistical significance of the difference in means (balance tests).

Before commenting the table, it is important to highlight that, while this is certainly an informative exercise, the latter should be taken with caution. As stated in the DEF, ex-post impact assessment should ideally occur prior to the closure of the project, so project completion reports can benefit from the impact assessment findings. If so, PCRs ratings incorporate IA findings when available – hence potentially influencing the direction of the final rating. Therefore, from a statistical standpoint, PCR ratings should not be used to assess the presence of selection bias, as they are positively affected by the mere virtue of a project being under evaluation.

Table 1: Balance tests: PCR ratings

	Average PCR ratings (IFAD10 IA sample)		***Average PCR ratings (completing IFAD10 projects 2016-2018)***		***Sample - Unselected***	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	N. projects	**Mean**	**N. projects**	**Mean**	**Diff. in Means**	**P-score**
Relevance	13	5.2	75	4.6	0.6	0.005
Effectiveness	13	4.8	76	4.2	0.6	0.005
Efficiency	13	4.4	76	3.8	0.6	0.026
Sustainability	13	4.4	76	3.8	0.5	0.017
Project performance	13	4.7	75	4.1	0.6	**0.002**
Rural poverty impact	13	4.8	75	4.1	0.7	0.001
Gender equality and women's empowerment	13	4.6	88	0.6	0.2	0.38
Innovation	13	4.8	76	4.4	0.4	0.111
Scaling up	13	4.8	75	4.4	0.5	0.066
Environment and natural resource management	13	4.5	73	4.1	0.4	0.056
Adaptation to Climate Change	11	4.5	64	4.1	0.5	0.022
IFAD performance	13	4.8	76	4.3	0.5	0.01
Government performance	13	4.7	76	4.1	0.6	0.014
Overall project achievement	13	4.8	75	4.2	0.6	0.004

Source: Calculations based on IFAD10 IA sample and data extracted from IOE ratings database.

Table 1 shows average subjective rating scores across 14 mandatory criteria[17], used by IOE to evaluate projects at completion. However, means of selected and unselected projects are based on the universe of projects as defined by Management in the IFAD10 Report (107 completed projects).

Given the above concern with the use of PCR ratings, in what follows, the significance of differences in pre-determined characteristics is tested. These essentially are baseline characteristics and include objective features and 24 implementation ratings as measured at the beginning of the project (i.e., the first indicator of performance that is available in the system). Implementation ratings are monitored during the lifespan of the project. These are the ones that, effectively, should have informed projects' selection at the beginning of the IA process. Note that while the first three indicators in the table are objective, notably project duration, number of beneficiaries and total approved funding, performance indicators are self-assessed and are expressed on a rating scale (1-6) ranging from unsatisfactory, to highly satisfactory[18].

Zooming in, note how the projects selected for impact assessments were similar on average in terms of financing and number of actual beneficiaries to the universe of projects. The average approved financing across the sample of IAs was $51.7 million, and the average in the universe was of $50.9 million. In terms of beneficiaries, the average number of beneficiaries was 610,556 in the universe and 490,339 in the IA sample, but this difference is not statistically significant. In almost all performance ratings categories, the IAs performed slightly better than the universe of projects, on average. However it is

[17] Based on IOE Manual (2015) pp. 38 -40. These definitions build on the OECD/DAC Glossary of Key Terms in Evaluation and Results-Based Management; the Methodological Framework for Project Evaluation agreed with the Evaluation Committee in September 2003; the first edition of the Evaluation Manual discussed with the Evaluation Committee in December 2008; and further discussions with the Evaluation Committee in November 2010 on IOE's evaluation criteria and key questions. Rural poverty impact is defined as the changes that have occurred or are expected to occur in the lives of the rural poor (whether positive or negative, direct or indirect, intended or unintended) as a result of development interventions. Project performance is an average of the ratings for relevance, effectiveness, efficiency and sustainability of benefits. Relevance measures the extent to which the objectives of a development intervention are consistent with beneficiaries' requirements, country needs, institutional priorities and partner and donor policies. It also entails an assessment of project design and coherence in achieving its objectives. An assessment should also be made of whether objectives and design address inequality, for example, by assessing the relevance of targeting strategies adopted. Effectiveness is the extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance. Efficiency is a measure of how economically resources/inputs (funds, expertise, time, etc.) are converted into results. Sustainability of benefits (or simply sustainability) is the likely continuation of net benefits from a development intervention beyond the phase of external funding support. It also includes an assessment of the likelihood that actual and anticipated results will be resilient to risks beyond the project's life. Gender equality and women's empowerment measures the extent to which IFAD interventions have contributed to better gender equality and women's empowerment, for example, in terms of women's access to and ownership of assets, resources and services; participation in decision making; work load balance and impact on women's incomes Nutrition and livelihoods. Innovation and scaling up (OR scaling up) measures the extent to which IFAD development interventions: (i) have introduced innovative approaches to rural poverty reduction; and (ii) have been (or are likely to be) scaled up by government authorities, donor organizations, the private sector and others agencies. Environment and natural resource management represents the extent to which IFAD development interventions contribute to resilient livelihoods and ecosystems. The focus is on the use and management of the natural environment, including natural resources defined as raw materials used for socio-economic and cultural purposes, and ecosystems and biodiversity – with the goods and services they provide. Adaptation to climate change is the contribution of the project to reducing the negative impacts of climate change through dedicated adaptation or risk reduction measures. Performance of Partners (IFAD and Government): This criterion assesses the contribution of partners to project design, execution, monitoring and reporting, supervision and implementation support, and evaluation. The performance of each partner will be assessed on an individual basis with a view to the partner's expected role and responsibility in the project life cycle. Finally, overall project achievement provides an overarching assessment of the intervention, drawing upon the analysis and ratings for rural poverty impact, relevance, effectiveness, efficiency, sustainability of benefits, gender equality and women's empowerment, innovation and scaling up, as well as environment and natural resources management, and adaptation to climate change.

[18] Ratings of project performance should be consistent with the findings of progress reports and of the supervision mission report. By rating each indicator, different criteria are applied as explained below, however in general the ratings are:
(6) Highly satisfactory. Targets/requirements met or exceeded. Considered as best practice.
(5) Satisfactory. Targets/requirements met with only minor delays or set-backs.
(4) Moderately satisfactory. Most targets/ requirements met but delays or set-backs experienced.
(3) Moderately unsatisfactory. Some targets/ requirements met but issues/constraints have negatively affected implementation.
(2) Unsatisfactory. Few targets/requirements met. Issues/constraints remain unresolved. Delays have seriously undermined implementation.
(1) Highly unsatisfactory. Almost no targets/ requirements met. Consideration should be given to cancellation/suspension.

important to note that these differences are not statistically significant for the majority of indicators presented – except for the following ratings:

- Assessment of the Overall Implementation Performance* : significant at 10% level.
- Acceptable Disbursement Rate**: significant at 5% level.
- Counterparts Funds** significant at 5% level.
- Coherence between AWPB and Implementation*: significant at 10% level.

Table 2: Balance tests: implementation performance ratings (baseline characteristics)

	IFAD10 IA Sample		***IFAD10 projects***	***Unselected***	***Sample - Unselected***	
	N. projects (1)	**Mean (2)**	**N. projects (3)**	**Mean (4)**	**Diff. in Means (5)**	**P-score (6)**
Project Duration	19	8.16	88	8.27	-0.12	0.863
Beneficiaries	19	490 339	88	636 512	-146 173	0.732
Approved Funding	19	51 712 292	88	50 700 000	986 098	0.957
Assessment of the Overall Implementation Performance	19	4.11	88	3.85	0.25	0.059
Likelihood of Achieving the Development Objective	19	4	88	3.98	0.02	0.857
Effectiveness	12	3.83	69	3.88	-0.05	0.74
Targeting and Outreach	19	4.26	88	4.11	0.15	0.224
Gender equality & women's participation	19	4.05	88	3.99	0.06	0.666
Agricultural Productivity	15	4.13	71	3.94	0.19	0.147
Adaptation to Climate Change	2	4	11	3.91	0.09	0.863
Institutions and Policy Engagement	18	4.11	78	4.01	0.1	0.549
Human and Social Capital and Empowerment	15	4.13	77	3.92	0.21	0.177
Quality of Beneficiary Participation	19	3.95	88	4.06	-0.11	0.401
Responsiveness of Service Providers	19	3.89	88	3.97	-0.07	0.592
Environment and Natural Resource Management	2	4	13	3.77	0.23	0.607
Exit Strategy	11	4.09	58	3.97	0.13	0.387

Potential for Scaling-up	14	4.21	72	4.07	0.15	0.366
Quality of Project Management	19	3.95	88	3.85	0.1	0.631
Knowledge Management	16	4.19	76	4.03	0.16	0.243
Coherence between AWPB and Implementation	17	4.12	81	3.79	0.33	0.064
Performance of M&E System	19	3.74	87	3.83	-0.09	0.574
Acceptable Disbursement Rate	19	4.21	88	3.43	0.78	0.028
Quality of Financial Management	17	4.12	79	3.9	0.22	0.218
Quality and Timeliness of Audit	19	4.11	87	4.01	0.09	0.582
Counterparts Funds	19	4.42	88	4.01	0.41	0.031
Compliance with Loan Covenants	19	4.21	88	4.02	0.19	0.189
Procurement	19	4.16	88	4	0.16	0.292

Table 3 and Table 4 show the distribution of the sample of IAs projects by IFAD's region and project sector or type. In the universe, 30 projects were in the Asia and Pacific Region (APR) followed by 26 in Western and Central Africa (WCA), 20 in Eastern and Southern Africa (ESA), 18 in Latin America and Caribbean (LAC), and 13 in the North East and Northern Africa region (NEN). As far as the IAs Projects' Sample is concerned, the majority of IAs (six) were conducted in ESA, while five were conducted in APR, four in WCA, three in LAC, and one in NEN. Table 22 in the Annex presents the mean performance by region and shows similar results i.e. that none of the mean ratings are statistically different across the IA sample and the unselected projects, although there is more variation, largely due to the lower number of overall projects with each region.

Turning to the project sector or type, a variable that is quite broad in the current classification system, the majority of projects in the universe are classified as agricultural development (37), rural development (34), and credit (14). However, no credit projects were selected for assessment in IFAD10 and over 40% of all IAs were of rural development projects. Nevertheless, because the project sector categorization is extremely broad, contains considerable overlap among categories, and is insufficiently informative about the true nature of the project, it lacks utility for the conduction of rigorous sensitivity analysis or bias estimation.

Table 3: Distribution of Projects in the Universe and in the IAs sample by Region

Universe by Region			***IAs by Region***		
BU	**Projects**	**%**	**BU**	**Projects**	**%**
APR	30	28.04	APR	5	26.32
ESA	20	18.69	ESA	6	31.58
LAC	18	16.82	LAC	3	15.79
NEN	13	12.15	NEN	1	5.26
WCA	26	24.30	WCA	4	21.05
Total	107	100.00	Total	19	100.00

Table 4: Distribution of Projects in the Universe and in the IAs sample by Sector or Project type

Universe by Project			***IAs by Project***		
Sector	**Projects**	**%**	**Sector**	**Projects**	**%**
AGRIC	37	34.58	AGRIC	6	31.58
CREDI	14	13.08	CREDI	0	0.00
FISH	2	1.87	FISH	0	0.00
IRRIG	7	6.54	IRRIG	1	5.26
LIVST	4	3.74	LIVST	2	10.53
MRKTG	6	5.61	MRKTG	1	5.26
RSRCH	3	2.80	RSRCH	1	5.26
RURAL	34	31.78	RURAL	8	42.11
Total	107	100.00	Total	19	100.00

Finally, implementation performance ratings were combined (Table 5) to assess and test for differences across proportions/percentage of projects rated satisfactory both in the IA sample(19) and in the universe of projects completing during IFAD10 (107). Note that there was not much variation in project scores in either the universe or the IA sample with

the highest density of projects around scores of four and five out of six. Specifically, when the indicators are transformed to indicate whether a project scored a satisfactory (4-6) or unsatisfactory (1-3) rating, it is apparent that in both the IFAD10 IA sample and the universe the majority of projects received satisfactory ratings. In the IA sample, between 84 and 100 percent of projects received satisfactory ratings and in the universe between 75 and 98 percent of projects did.

Although there are differences in the relative frequency of unsuccessful projects, it is clear that the majority of portfolio projects receive satisfactory scores. As such, it is reasonable that a high proportion of the sample would be high performing projects.

In summary, a conclusion that can be drawn from this analysis, is that bias is absent for the majority of baseline indicators presented – except for the following two ratings:

- Acceptable Disbursement Rate**: significant at 5% level.
- Counterparts Funds** significant at 5% level.

Further analyses are therefore conducted on these variables in question in the following sections.

Table 5: Balance tests: Proportions of Projects Rated Satisfactory

	Average performance ratings	*Average performance ratings*	*Difference in IAs and Universe Means*	*Proportion of IAs Rated*	*Proportion of Universe Rated*	*Sample-Universe*	*Significance*
	IFAD10 IA Sample	*(Completing IFAD10 projects 2016-2018 : universe(107))*		*Statisfactory*	*Satisfactory*		
	Mean (1)	**Mean (2)**	**Diff. in Means (3)**	**Proportion (4)**	**Proportion (5)**	**Diff. in Proportion (5)**	**P-score (6)**
Assessment of the Overall Implementation Performance	4.11	3.9	0.21	100%	92%	8%	0.436
Likelihood of Achieving the Development Objective	4	3.98	0.02	100%	84%	16%	0.561
Effectiveness	3.83	3.88	-0.04	100%	83%	17%	0.957
Targeting and Outreach	4.26	4.14	0.12	100%	98%	2%	0.336
Gender equality & women's participation	4.05	4	0.05	89%	89%	1%	0.93
Agricultural Productivity	4.13	3.98	0.16	100%	84%	16%	0.569
Adaptation to Climate Change	4	3.92	0.08	94%	90%	5%	0.685
Institutions and Policy Engagement	4.11	4.03	0.08	100%	82%	18%	0.537
Human and Social Capital and Empowerment	4.13	3.96	0.18	95%	87%	8%	0.483
Quality of Beneficiary Participation	3.95	4.04	-0.09	100%	95%	5%	0.833

Responsiveness of Service Providers	3.89	3.95	-0.06	95%	87%	8%	0.221
Environment and Natural Resource Management	4	3.8	0.2	100%	87%	13%	0.582
Exit Strategy	4.09	3.99	0.11	84%	75%	9%	0.356
Potential for Scaling-up	4.21	4.09	0.12	100%	90%	10%	0.355
Quality of Project Management	3.95	3.87	0.08	84%	82%	2%	0.851
Knowledge Management	4.19	4.05	0.13	100%	85%	15%	0.293
Coherence between AWPB and Implementation	4.12	3.85	0.27	84%	69%	15%	0.317
Performance of M&E System	3.74	3.81	-0.07	78%	72%	7%	0.993
Acceptable Disbursement Rate	4.21	3.57	0.64	95%	65%	29%	0.025
Quality of Financial Management	4.12	3.94	0.18	89%	79%	11%	0.764
Quality and Timeliness of Audit	4.11	4.03	0.08	89%	93%	-3%	0.486
Counterparts Funds	4.42	4.08	0.34	79%	74%	5%	0.194
Compliance with Loan Covenants	4.21	4.06	0.15	100%	87%	13%	0.492
Procurement	4.16	4.03	0.13	89%	74%	16%	0.681

5 Strategies for addressing the selection bias

Given that there are statistically significant differences in only two observable features of the projects subject to an impact assessment, two main strategies are considered to assess the need for adjusting for the possible selection bias.

5.1 Modelling selection bias for impact assessment using observables features

Since some information on the universe is available (including implementation performance), the Heckman approach can be adapted (1979)[19] to compute the likelihood of a project being selected for an assessment compared to the rest of the universe, conditional on the available observed characteristics. The meta-analysis is then run once again after reweighting each project by its probability of being selected into an IA.

The success of this approach, of course, rests on the number and on the importance of the observed factors in driving the selection into an IA. If one can plausibly argue that selection of IA projects depends mostly on the observable (rather than unobservable) characteristics that are observed, such as project type, region, financing and implementation performance ratings, the meta-analysis results can be adjusted for selection bias based on observables.

5.1 Publication bias

The second approach considered draws on the meta-analysis literature and treat our sampling issue as a classical publication bias problem. The latter refers to the distortion of meta-analysis outcomes due to the higher likelihood of publication of statistically significant studies rather than non-significant studies. This is similar to the problem at hand – where impact assessment estimates are only available for the projects evaluated, hence less performing projects are not observed – hypothetically – in the sample. Therefore, presenting this kind of sensitivity analysis would be useful here too.

In order to test for the presence or absence of publication bias, first, a funnel plot can be used. In essence, studies are plotted on a scatter plot with effect size on the x-axis and precision or total sample size on the y-axis. If the points form an upside-down funnel shape, with a broad base that narrows towards the top of the plot, this indicates the absence of a publication bias. On the other hand, if the plot shows an asymmetric shape, with no points on one side of the graph, then publication bias can be suspected. Second, to test publication bias statistically, Begg and Mazumdar's rank correlation test or Egger's test can be used. If publication bias is detected, the trim-and-fill method can be used to correct the bias (Shi et al, 2019).

A known limitation of Trim-and-Fill is that it can correct for publication bias that does not exist, underestimating effect sizes (Terrin et al, 2003). Results of this method are therefore optional and presented in Annex IV.

[19] Heckman, J. (1979). Sample Selection Bias as a Specification Error. Econometrica, 47(1), 153-161.

6 Results from the Sensitivity Analyses

6.1 Subgroup meta-analyses

Subgroup meta-analyses are presented to assess whether the magnitude of the impact is associated with the rating class to which a project belongs, and verify whether there is a gradient between impact magnitude and rating scale. The rationale for this is due to the fact that in in Section 4 it was shown how two baseline ratings were in fact unbalanced between the IFAD10 IA sample of projects and the rest of the universe.

Specifically, two features (Acceptable Disbursement Rate and Counterparts Funds) were significantly different at 5% level between projects evaluated and the unselected ones in the universe. Therefore, the extent of bias is investigated, notably whether there is a relationship between positive ratings and impacts. If bias existed, one would expect to see patterns or gradients such as the following, namely the higher the rating the higher the impact in the IFAD10 estimates of the aggregate effect sizes, as this is the main argument by IOE, notably that impact is overestimated due to higher performing projects.

Results across the Strategic Objectives (SOs), notably Market access, Resilience, Production as well as the cross-cutting theme Nutrition and the overarching IFAD goal of Economic mobility, are presented in the Annex (). Forest plots, the visual representations commonly used for meta-analytic results are employed for the purpose. The reader needs to focus on the diamonds, which represent the size of the effect – grouped according to ratings: unsatisfactory (below 3), moderately satisfactory (4), and satisfactory plus (ratings higher than 5). The "overall" diamond represents our pooled effect size, the one presented to the Board in September.

It is remarkable to see how there is no clear relationship between the ratings class and the impact estimates, particularly in the case of disbursement performance ratings. This is reassuring and corroborates the absence of bias in the impact estimates due to selection of higher performing projects.

For example, note how in the case of the market access SO domain, projects rated unsatisfactory in terms of disbursement performance ratings, display the highest impact (1.89 equivalent to 89%), compared with satisfactory plus (57%) and moderately satisfactory ones (80%). This is also largely true for the other SOs.

Also in the case of the performance rating "counterpart funds", note how projects rated as unsatisfactory have a higher impact (29 %) compared with the moderately satisfactory ones (18%), in the domain of Production. Tables are presented in Annex II.

Therefore, a major conclusion that can be drawn from this analysis is that the direction of the ratings e.g. stronger performance, does not correspond to a larger impact.

6.2 Sample selection bias correction á la Heckman

In this section, one of the sensitivity analyses described in Section 6, palatable for adjusting for the bias, is presented. The second method – trim and fill – adjusts for bias non-parametrically and given its low reliability, is presented in the Annex.

In the approach presented here, a three-stage estimation procedure is applied, whereby the results are corrected for the presence of observable and unobservable selection using an approach a la Heckman (1979), combined with meta-regression and meta-analysis.

In the first stage, a probit regression model is estimated over the universe of projects on the variables determining the selection, notably the ones that are assumed as observable drivers for the selection into IA (project sector, region, and the significant performance ratings). These estimates are used to obtain an Inverse Mills Ratio (IMR), and introduce the latter in a meta-regression in order to estimate corrected standard errors. The variance associated with this corrected standard error is then computed and summed to the original variance, to obtain final "corrected standard errors" (second stage). These final corrected standard errors are then meta-analyzed along with the original effect sizes to derive a pooled effect size adjusted for bias (third stage).

It is important to note that IMR is only marginally significant (at 10% level) in the meta-regression pertaining to the market access SO domain. For the rest of SOs it is not significant - this implies that the hypothesis of selection bias is rejected in the case of these SOs, and that it weakly holds in the context of market access.

Results are summarized in Table 6, where "observed" refers to the original impact estimates presented in the official IFAD10 Report and "adjusted" refers to the ones corrected for sample selection bias. The full set of tables is presented in the Annex III.

Note how the results maintain the integrity of the baseline random-effect meta-analysis model - the one presented in the original IFAD10 Report - and show minimal discrepancies. Specifically, results based on this scenario also remain largely positive, and indicate that, ceteris paribus, impact is overestimated by 2% in case of economic mobility, 15% in case of market access, 10% in case of production, and 6% in the case of resilience. Nutrition results remain unchanged.

However, given the lack of significance of the IMR in the second stage regression, Management concludes that there is no selection bias in the corporate impact estimates and that bias adjusted estimates are not needed and represent an over-correction.

Table 6: Results from the Sample selection bias correction

Production				
	N. projects	ES	Lower CI	Upper CI
Observed	17	1.44	1.26	1.64
Adjusted	17	1.33	1.24	1.43
Market Access				
	N. projects	ES	Lower CI	Upper CI
Observed	16	1.76	1.45	2.14
Adjusted	16	1.51	1.32	1.72
Resilience				
	N. projects	ES	Lower CI	Upper CI
Observed	17	1.13	1.02	1.25
Adjusted	17	1.06	1.02	1.1
Nutrition				
	N. projects	ES	Lower CI	Upper CI
Observed	16	1.01	0.99	1.03

Adjusted	16	1.01	0.99	1.02
Economic Mobility				
	N. projects	ES	Lower CI	Upper CI
Observed	17	1.74	1.51	1.97
Adjusted	17	1.72	1.42	2.02

7 Conclusions on IFAD10

In this document, a number of sensitivity analyses are presented, to assess the presence, direction and magnitude of the possible selection bias inherent in the sampling of projects chosen to be evaluated under IFAD10.

Results highlight that the bias is absent and, if anything, negligible. Through a detailed descriptive analyses it is shown that almost all the pre-determined e.g. baseline features of the IFAD10 IA sample and of the ones of the unselected projects are largely balanced – e.g. they are not statistically different - with the only exception of a couple of implementation ratings. Upon further investigation, it was found that the direction of the ratings does not imply a larger estimated impact, allowing one to conclude that projects rated highly unsatisfactory on certain attributes exhibit higher effect sizes compared with satisfactory projects. This finding strongly hints that corrective actions are put in place by implementers across the project lifetime to influence ratings towards more positive ones, particularly at completion. This factor corroborates Management's choice of not employing ratings at completion (PCR ratings) for an assessment of selection bias as the latter are endogenous (e.g., influenced by the evaluation process) and may be inflated by many reasons, the first being that ratings may reflect corrective actions by implementers, and second, that ratings do incorporate the findings of the impact assessments when available.

This analysis is complemented by an assessment of the need to correct for sample selection bias. To this end, two approaches are considered, notably the sample selection bias correction a la Heckman and the trim-and-fill approach. These are meant to more formally assess the presence and the magnitude, respectively, of any possible sample selection bias.

Given that information about the observable factors that might influence selection are available in the system, a sample selection bias correction a la Heckman is the preferred approach and is applied in the meta-analytic context.

Results based on this scenario show that selection bias does not hold and it is weakly present only in the estimations of corporate impact for market access. After computing bias adjusted estimates – the latter remain largely positive, and indicate that, ceteris paribus, impact was overestimated by 2% in case of economic mobility, 15% in case of market access, 10% in case of production, and 6% in the case of resilience. Nutrition results remain unchanged.

The trim-and-fill method is instead a popular tool to detect and adjust for publication bias, in other words the bias originated by ad-hoc inclusion of projects/studies in the meta-analysis. However this approach is strongly criticized by the literature (Terrin 2003, Simonsonh et al 2014) whereby meta-analysts are not recommended to perform the trim-and-fill method when using meta-analysis software programs (Shi et al, 201920), as outliers and the pre-specified direction of missing studies could have influential impact on the trim-and-fill results. In addition a known limitation of Trim-and-Fill is that it can correct for publication bias that does not exist, underestimating effect sizes (Terrin et al 2003).

Although results adjusted using this approach remain largely positive they are presented in the Annex for the above mentioned considerations.

[20] Ref : need to put all references in footnote.

8 Implications for IFAD11

Turning to IFAD11 – what are the implications moving forward? In the following sections the process for selecting the IFAD11 IA is summarized, and descriptive analyses are presented, to assess for the presence of selection at the time of the projects' choice in the IFAD11 context, using performance ratings and other features of the portfolio universe.

Management has formalized the process of identifying candidate IFAD-supported projects to undergo ex post impact assessments. All regional divisions have been requested to identify and select potential countries and projects to conduct impact assessments from a list of all projects scheduled to close during IFAD11 (between 2019 and 2021) as of July 2018. Projects have been identified and selected through a participatory approach which involved Management and specifically the Research and Impact Assessment Division (RIA) and each of the five regional divisions, similar to the one implemented during IFAD10.

A first screening was done in July 2018 based on disbursement rate, timing of the project, and type of project. After this first screening, further identification was conducted based on learning potential, feasibility of conducting impact assessment given the eligibility and targeting criteria and project implementation, quality of M&E data, number of beneficiaries, type of interventions, and buy-in from country and project teams. RIA staff met with representatives with each regional division to select IFAD11 ex-post impact assessments.

During this meeting, each regional division received a list of projects that RIA staff had pre-screened[21]. RIA staff requested each regional division to identify six projects as candidates for impact assessments during IFAD11 (two projects per replenishment year).

Subsequently, a validation exercise was conducted through follow-up meetings in collaboration with each regional division and projects received clearance from both Country and Regional Directors. Additionally, RIA held internal discussions to ensure that projects selected were representative of the IFAD11 portfolio in terms of both regional distribution and sector.

8.1 IFAD11 Sample Validation

As part of the IFAD11 impact assessment agenda, 24 out of 121 projects have been selected for rigorous impact assessment equalling 19.8% of total projects, 20.7% of total financing, and 23.3% of total IFAD financing. Of the 121 projects belonging to the IFAD11 universe, nine[22] were projects already part of evaluations initiated during IFAD9 and IFAD10 whose closing dates now fall during IFAD11. This gives a final universe of 112 projects eligible for evaluation in IFAD11. Considering the latter, the projects selected to be evaluated during IFAD11 account for 21.4% of the portfolio, representing 20.9% of total financing and 25.6% of IFAD financing.

[21] The number of pre-screened projects scheduled to close between 2019 and 2021 that the RIA team had initially offered to each regional division were as follows: 26 projects for APR, 21 projects for ESA, 23 projects for LAC, 17 projects for NEN, and 16 projects for WCA.

[22] The universe of "121 projects" include all projects CLOSING during IFAD11. However, upon further scrutiny it appeared that in the UNIVERSE of 121 – there are 9 projects whose evaluations were carried out during IFAD9 and IFAD10. The IFAD9 & IFAD10 projects are those whose closing dates were extended into IFAD11. The projects evaluated during IFAD10 are: Sao Tome and Principe PAPAC 1100001687; Senegal PAFA (extended as PAFA-E) 1100001693; Rwanda PRICE 1100001550; Kenya SDCP 1100001305; Nepal HVAP 1100001471; Bangladesh CCRIP 1100001647. The projects evaluated in IFAD9 then extended are : Uganda VODP2 1100001468; Ghana GASIP 1100001678; Bangladesh PACE 1100001648.

Table 7 presents the IFAD11 projects selected for impact assessment and their distribution by region, country, and project sector or type.

Table 7: IFAD11 Impact Assessments by Region, Country, Project Sector and Name

	Region	***country***	***sector***	***Project name***		***REGION***	***country***	***sector***	***Project name***
1	APR	India	CREDI	PT-Tamil Nadu	**13**	LAC	Argentina	MRKTG	PRODERI
2	APR	Pakistan	RURAL	SPPAP - PK	**14**	LAC	Bolivia	RURAL	ACCESOS
3	APR	Papua New Guinea	AGRIC	PPAP	**15**	LAC	Peru	RSRCH	PSSA
4	APR	Philippines	FISH	FishCORAL	**16**	NEN	Djibouti	RURAL	PRAREV-PECHE
5	APR	Solomon Islands	RURAL	RDP II	**17**	NEN	Kyrgyzstan	LIVST	LMDP
6	APR	Viet Nam	RURAL	AMD	**18**	NEN	Moldova, Republic of	RURAL	IRECR
7	ESA	Kenya	AGRIC	UTaNRMP	**19**	NEN	Morocco	AGRIC	PDFAZMH
8	ESA	Lesotho	RURAL	SADP	**20**	NEN	Tunisia	AGRIC	PRODESUD II
9	ESA	Malawi	RSRCH	SAPP	**21**	WCA	Ghana	CREDI	REP
10	ESA	Mozambique	AGRIC	PROSUL	**22**	WCA	Mali	CREDI	Rural Microfinance Programme
11	ESA	Tanzania, Un. Rep. of	MRKTG	MIVARF	**23**	WCA	Mauritania	RURAL	PASK II
12	ESA	Zambia	RSRCH	S3P	**24**	WCA	Nigeria	AGRIC	VCDP

Table ***8*** compares the regional distribution of these projects selected for impact assessment to the regional distribution of projects in the universe. Specifically, six projects were selected for impact assessment in both APR and ESA, five in NEN, four in WCA, and three in LAC. In the universe of 112 projects, there are 23 projects in LAC and NEN respectively, 32 in APR, 20 in ESA, and 17 in WCA.

Table 8: Distribution of projects in the universe and in the IA sample by Region

UNIVERSE BY REGION			IAS BY REGION		
REGION	**Projects**	**%**	**REGION**	**Projects**	**%**
APR	29	25.89	**APR**	6	25.00
ESA	20	17.86	**ESA**	6	25.00
LAC	23	20.54	**LAC**	3	12.50
NEN	23	20.54	**NEN**	5	20.83
WCA	17	15.18	**WCA**	4	16.67
TOTAL	112	100	**Total**	24	100.00

Table ***9***, first column, presents the current regional distribution of impact assessments in the IFAD11 sample and compares this distribution with three others, one weighted by the actual proportion of projects in each region and two more weighted by the actual proportional allocation of projects by financing and IFAD financing in the portfolio. As

shown in the table, the distribution of projects dictated by these different weighting schemes differ slightly from the distribution of those actually selected.

Table 9: Distribution of IFAD11 IAs. Current and proportional to project numbers and financing (total and IFAD only).

	Actual IFAD11 IA Distribution	*Distribution by # of Projects*	*Distribution by Financing*	*Distribution by IFAD Financing*
APR	6	6	8	7
ESA	6	4	6	5
LAC	3	5	3	3
NEN	5	5	3	4
WCA	4	4	4	5
	24	24	24	24

IFAD projects are also classified into eight project sectors (or types) in the universe (Table 10). IFAD11 projects are concentrated in rural development (46), agricultural development (29), and credit provision (16). The number of projects in the remaining five sectors range from two to six with the lowest concentration of projects in fisheries. The projects selected for IAs are comparatively less concentrated; eight rural development projects, six agricultural development projects, and three credit projects were selected. Despite the large number of credit projects in the portfolio, an equal number of projects (3) was selected in the research category along with two market access projects and one each in fisheries and livestock, respectively. No irrigation projects were selected for impact assessment during IFAD11.

Table 10: Distribution of projects in the universe and in the IA sample by Project Sector

Universe by Project Sector			*IAs sample by Project Sector*		
Sector	**Projects**	**%**	**Sector**	**Projects**	**%**
AGRIC	29	25.89	**AGRIC**	6	25
CREDI	16	14.29	**CREDI**	3	12.5
FISH	2	1.79	**FISH**	1	4.17
IRRIG	4	3.57	**IRRIG**	0	0
LIVST	6	5.36	**LIVST**	1	4.17
MRKTG	5	4.46	**MRKTG**	2	8.33
RSRCH	4	3.57	**RSRCH**	3	12.5
RURAL	46	41.07	**RURAL**	8	33.33
Total	112	100	**Total**	24	100

Looking at Table 11, note how the distribution of the IFAD11 IAs sample by sector differs from what would be dictated by the proportion of projects by sector in the universe, as well as the sectoral distribution proportional to the amount of total financing and IFAD financing. Certainly, when considering the projects by their sectoral classifications, irrigation projects seem to be underrepresented in the IFAD11 IA selection.

Table 11: Distribution of IFAD11 IAs by project sector. Current and proportional to project numbers and financing (total and IFAD only) by sector.

	IFAD11 IAs Sample Distribution	***Distribution by # of Projects***	***Distribution by Total Financing***	***Distribution by IFAD Financing***
AGRIC	6	6	6	7
CREDI	3	3	4	4
FISH	1	1	0	0
IRRIG	0	1	1	1
LIVST	1	1	1	1
MRKTG	2	1	1	1
RSRCH	3	1	2	1
RURAL	8	10	9	9
Total	**24**	**24**	**24**	**24**

Table 12 presents descriptive statistics of the sample and the unselected projects summarizing projects by their number of beneficiaries, financing, components, and implementation score. It also presents the results of t-tests for difference in means between the sample and unselected projects to assess for the possible presence of selection bias across the various samples.

Table 12: Balance test: selected indicators for IFAD11 (baseline).

	IAs Sample	***Unselected Projects***	***Sample - Unselected***	
	Mean	**Mean**	**Diff. in Means**	**p-score**
Beneficiaries	476,109	1,033,194	-557,085	0.595
Financing	81,820,640	74,454,692	7,365,947	0.695
IFAD Financing	37,848,004	29,311,323	8,536,680	0.176
# of Financiers	2.67	2.59	0.08	0.479
# of Subcomponent	5.92	6.10	-0.18	0.753

On average, the sample of IFAD11 IAs has 476,109 beneficiaries, $81.8 million in financing, $37.8 million in IFAD financing, 2.6 types of financiers, and 6 subcomponents. Compared to the average across the universe of IFAD11 projects, there are 942,862 beneficiaries, $78.7 million in financing, $31.9 million in IFAD financing, 2.6 types of financiers, and 6 subcomponents. Note how there are no statistically significant differences across the variables presented in Table 12, across the universe and the unselected projects.

Table 13: Distribution of IFAD11 Universe by Project Type and Region

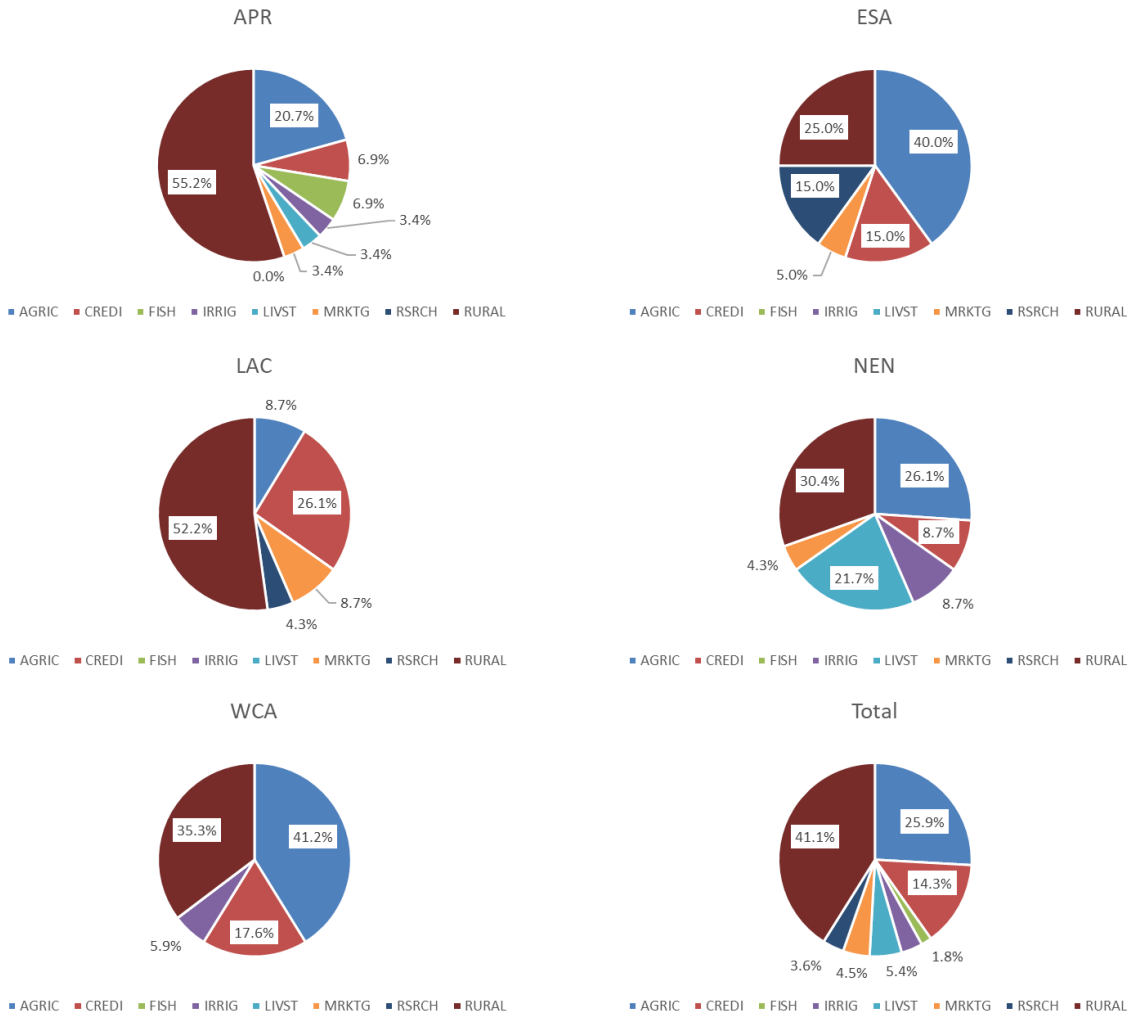


Table 14: Distribution of IFAD11 IA Sample by Project Type and Region

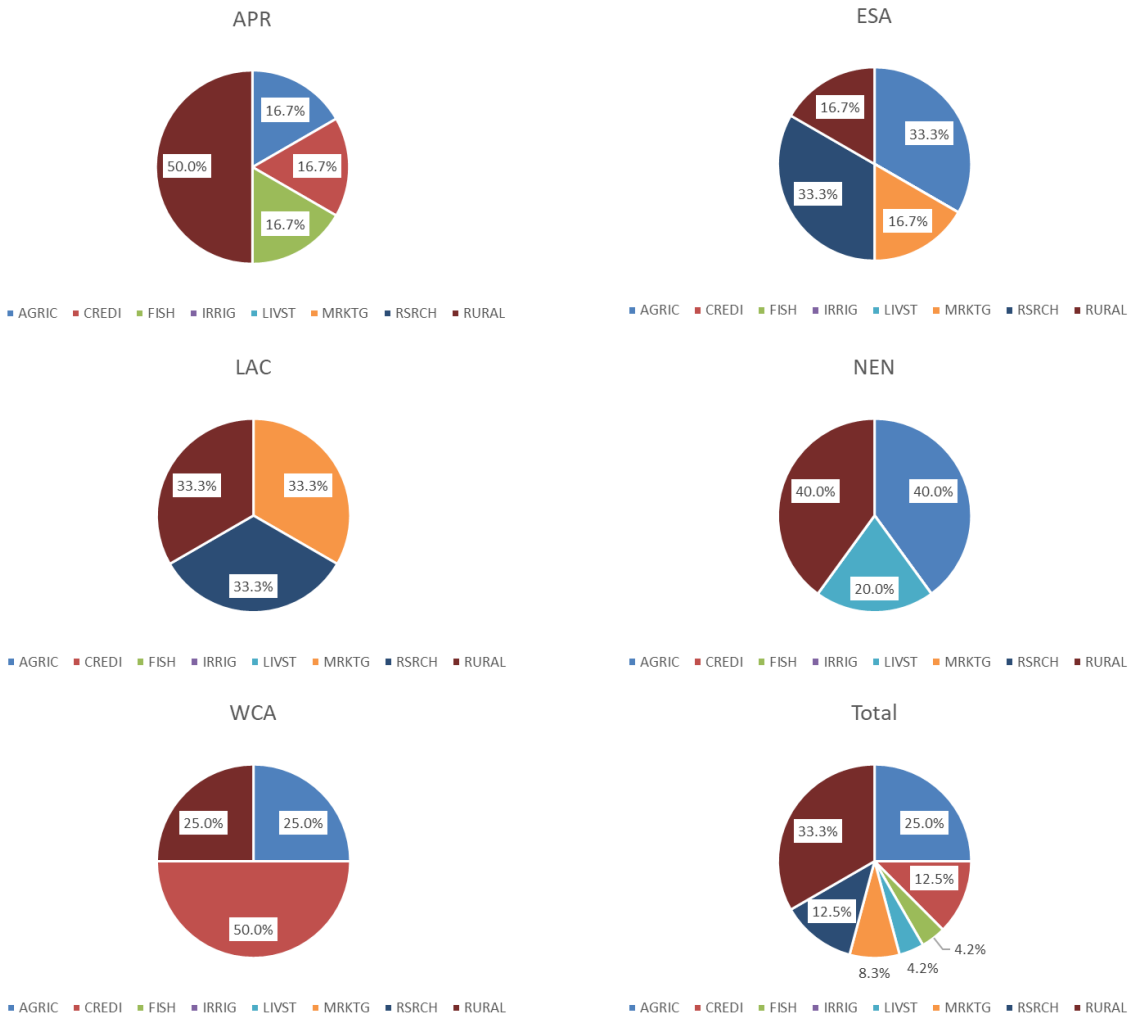


Table 15: Distribution of IFAD11 Universe by Project Type and Region Financing

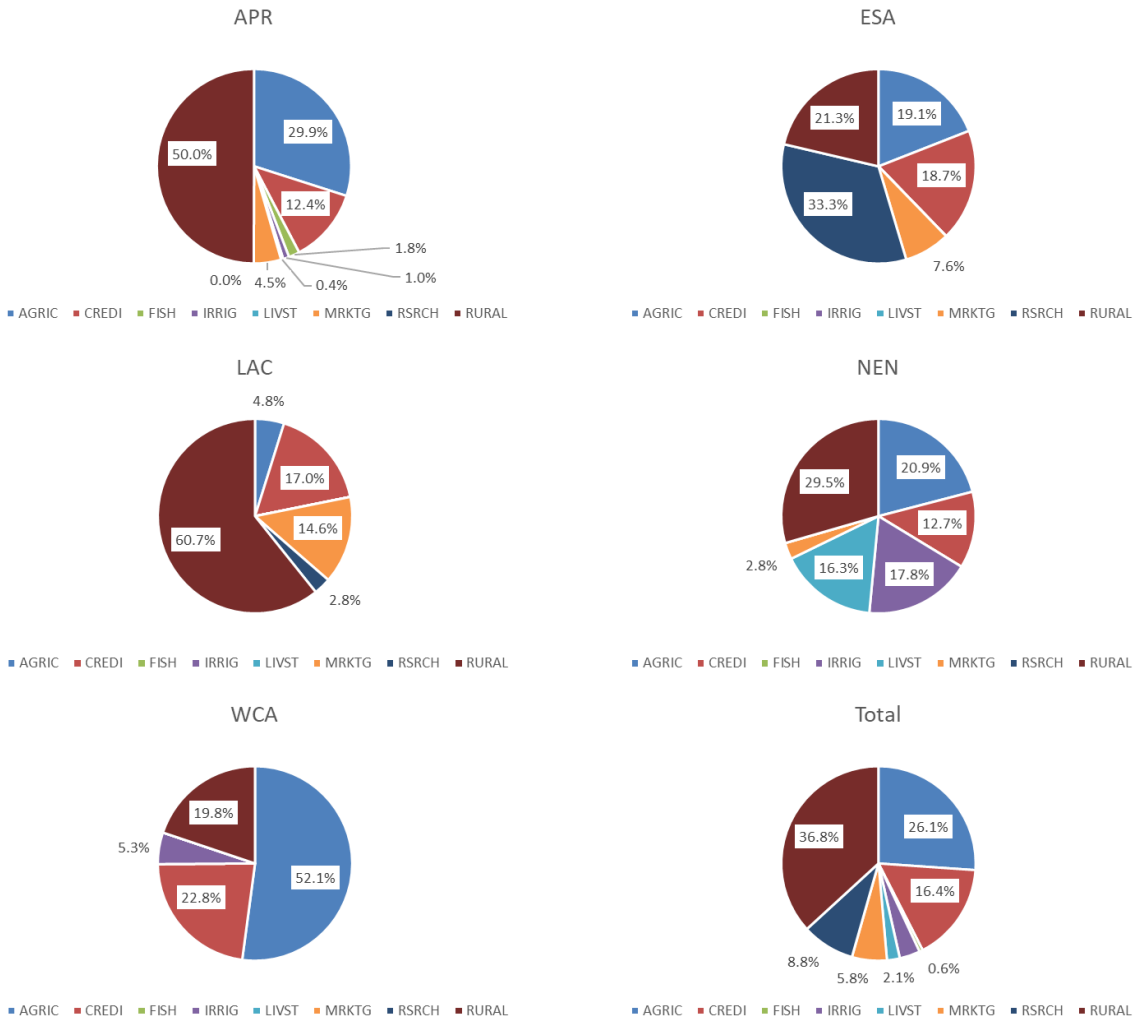


Table 16: Distribution of IFAD11 IA Sample by Project Type and Region Financing

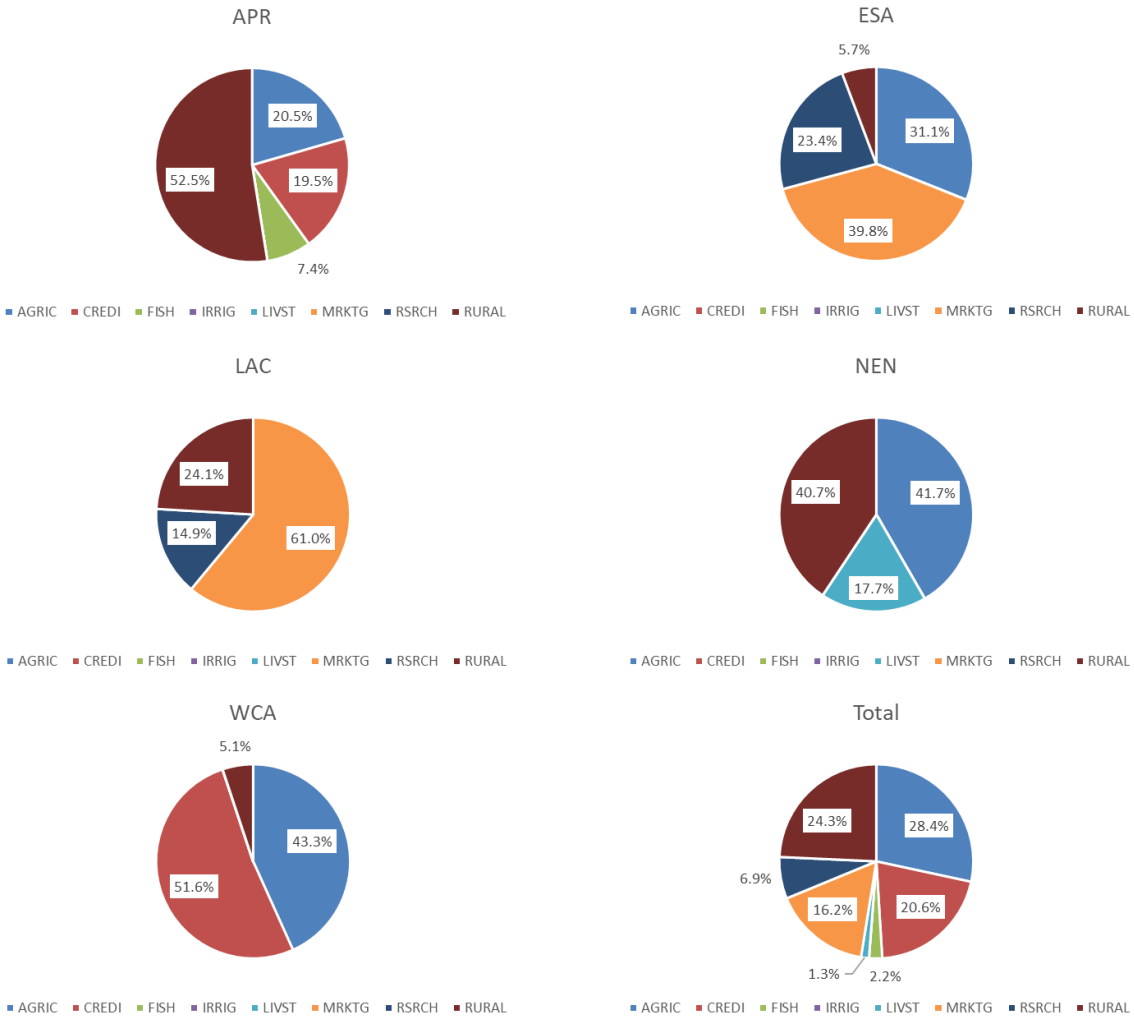
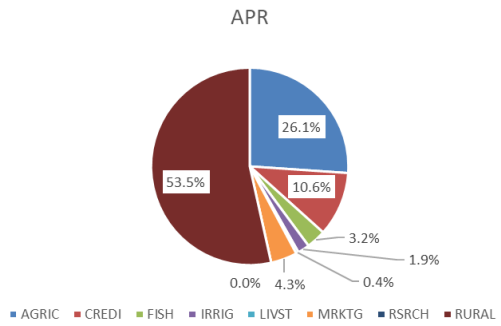
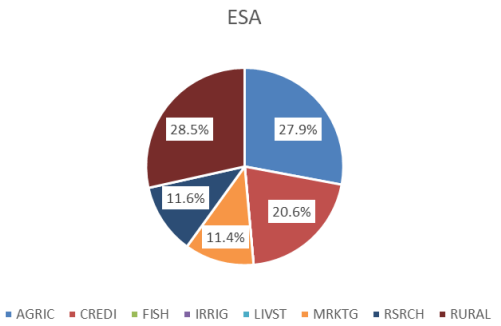
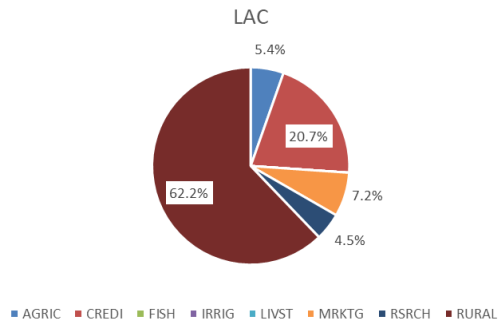
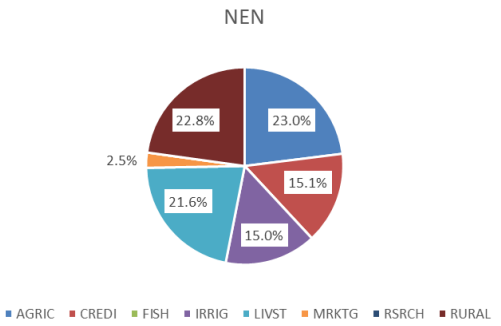


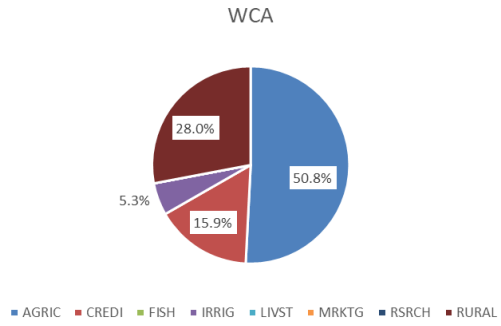
Table 17: Distribution of IFAD11 Universe by Project Type and Region IFAD Financing











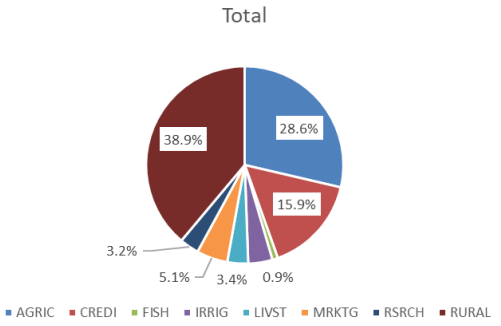
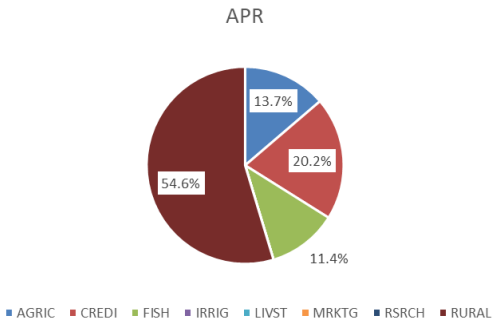
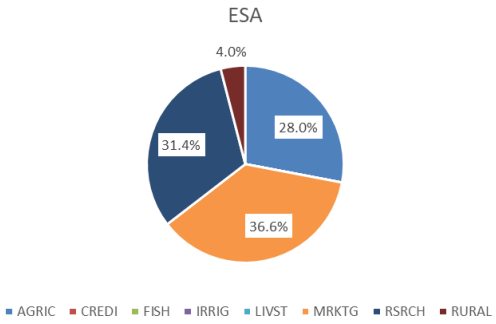
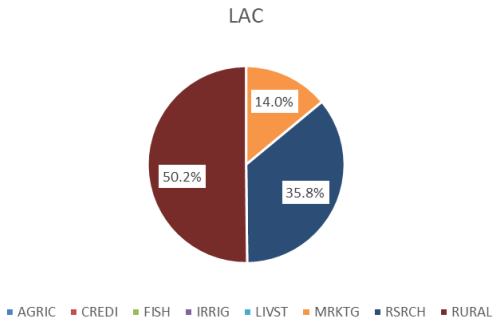
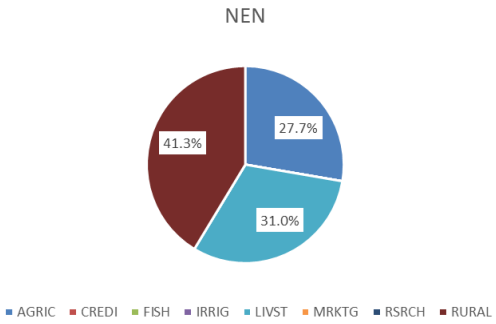


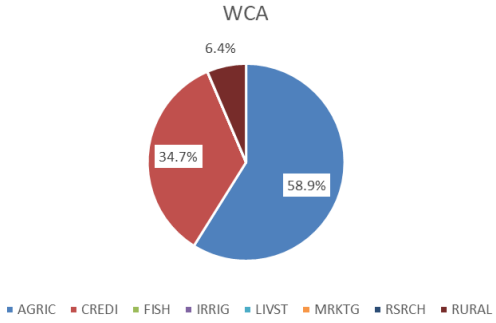
Table 18: Distribution of IFAD11 IA Sample by Project Type and Region IFAD Financing

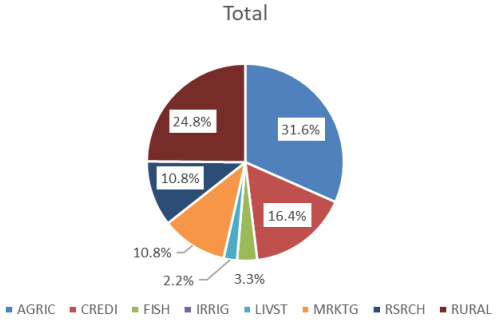












In addition to being divided into project sector and regional categories, projects can further be disaggregated into component, subcomponent and subcomponent type. The only variable that is standardized in the system (GRIPS) is the subcomponent type, which can be broken down into approximately 60 categories. Note that the component feature is not standardized across IFAD databases and therefore cannot be used in this analysis.

Each project contains multiple subcomponents funded by different sources. Nearly all projects contain a management and coordination component, but there remains considerable heterogeneity. Apart from the management subcomponent, the most common subcomponent in the universe of projects is rural financial services (44.7%), followed by local capacity building (38.39%), monitoring and evaluation (28.57%), and institutional support (27.68%). Table 19 compares the relative distribution of subcomponent types in the selected IAs and the IFAD11 portfolio.

When projects are categorized by their largest subcomponent (in terms of current financing), the most common subcomponents among IFAD11 IAs are development funds and rural financial services each with three projects selected for IAs, respectively (table available upon request). This matches the portfolio overall where development funds and rural financial services are the largest subcomponents with thirteen and fourteen projects, respectively. However, beyond that similarity, the distribution of IAs unevenly represents the distribution in the portfolio. Finally, looking at Table 20– the distribution of projects by largest financier and region is provided in the portfolio and in the IA sample.

Two issues in this analysis hinder both our ability to draw a representative sample by sector and components distribution and assess the representativeness of our current selection. The first is that the sector variable does not reflect the true nature of the project. For example, projects in which the true intervention is related to livestock or animal husbandry may be classified as marketing if there is substantial intervention in business formation or a rural development project may have a substantive irrigation component but not be classified as such because it is combined with other interventions. Moreover, the sectors are too broad such that there is substantial heterogeneity within sectors, namely the rural and agricultural development. Secondly, the subcomponent type is too disaggregated (60 unique entries over 112 records) to be used as a possible stratification feature hence it would need to be recoded prior to be used. However, there currently exists no method for standardizing or harmonizing subcomponents within projects and each project can have multiple subcomponents all of varying sizes and relative importance.

Table 19: Distribution of subcomponent types in the IAs sample and in the IFAD11 universe

	IAs		***Universe***	
Subcomponent Type	**Freq.**	**%**	**Freq.**	**%**
Rural financial services	**10**	41.67	**50**	44.64
Local capacity building	**8**	33.33	**43**	38.39
Monitoring and evaluation	**7**	29.17	**32**	28.57
Institutional support	**5**	20.83	**31**	27.68
Technology transfer	**7**	29.17	**27**	24.11
Rural infrastructure	**5**	20.83	**25**	22.32
Business development	**2**	8.33	**21**	18.75
Development funds	**5**	20.83	**20**	17.86
Irrigation infrastructure	**1**	4.17	**19**	16.96
Community development	**5**	20.83	**18**	16.07
Crop production	**3**	12.50	**17**	15.18
Marketing: inputs/outputs	**6**	25.00	**17**	15.18

Roads/tracks	**1**	4.17	**17**	15.18
Micro-enterprises	**3**	12.50	**16**	14.29
Climate change adaptation	**4**	16.67	**14**	12.50
Rangelands/pastures	**2**	8.33	**14**	12.50
Credit	**1**	4.17	**13**	11.61
Resource mgmt/protection	**1**	4.17	**13**	11.61
Rural enterprises	**2**	8.33	**13**	11.61
Animal husbandry	**2**	8.33	**11**	9.82
Technology development	**2**	8.33	**11**	9.82
Policy Support/Development	**2**	8.33	**9**	8.04
Animal health	**1**	4.17	**8**	7.14
Crop extension services	**2**	8.33	**8**	7.14
Market information/study	**3**	12.50	**8**	7.14
Training	**3**	12.50	**8**	7.14
Communication	**1**	4.17	**7**	6.25
Market infrastructure	**3**	12.50	**7**	6.25
Soil and Water conservation	**1**	4.17	**7**	6.25
Livestock - other	**0**	0.00	**6**	5.36
Livestock post-harvest	**2**	8.33	**6**	5.36
Seed, fertilizer, pesticide	**1**	4.17	**6**	5.36
Health and nutrition	**0**	0.00	**5**	4.46
Drinking water/sanitation	**0**	0.00	**4**	3.57
Forestry	**1**	4.17	**4**	3.57
Literacy	**0**	0.00	**4**	3.57
Crop technology development	**1**	4.17	**3**	2.68
Disaster mitigation	**2**	8.33	**3**	2.68
Financing/preparation charges	**1**	4.17	**3**	2.68
Fisheries infrastructure	**2**	8.33	**3**	2.68
Irrigation management	**0**	0.00	**3**	2.68
Land improvement	**1**	4.17	**3**	2.68
On-farm storage	**1**	4.17	**3**	2.68
Animal restocking	**1**	4.17	**2**	1.79
Aquaculture	**0**	0.00	**2**	1.79
Education (primary/second)	**1**	4.17	**2**	1.79
Fisheries/marine conservation	**2**	8.33	**2**	1.79
Input supply	**1**	4.17	**2**	1.79
Knowledge management	**0**	0.00	**2**	1.79
Land reform/titles	**0**	0.00	**2**	1.79
Processing	**0**	0.00	**2**	1.79
Standards and regulations	**1**	4.17	**2**	1.79
Energy production	**0**	0.00	**1**	0.89
Fishing (capture)	**0**	0.00	**1**	0.89
Housing	**1**	4.17	**1**	0.89
Insurance/risk transfer	**1**	4.17	**1**	0.89
Legal assistance	**0**	0.00	**1**	0.89
Mechanization services	**0**	0.00	**1**	0.89
Venture capital	**1**	4.17	1	0.89

Finally, looking at Table 20 – the distribution of projects by largest financier and region is provided in the portfolio and in the IA sample.

Table 20 Distribution of Largest Financier Type by Region.

IFAD11 Universe

	APR	ESA	LAC	NEN	WCA	Total
Largest Financier						
Domestic	10	4	9	3	1	**27**
IFAD	13	13	12	18	14	**70**
International	6	3	2	2	2	**15**
Total	**29**	**20**	**23**	**23**	**17**	**112**

IFAD11 IAs Sample

	APR	ESA	LAC	NEN	WCA	Total
Largest Financier						
Domestic	1	0	1	2	1	**5**
IFAD	3	5	2	3	3	**16**
International	2	1	0	0	0	**3**
Total	**6**	**6**	**3**	**5**	**4**	**24**

Table ***21*** presents additional baseline ratings statistics based on the first available implementation performance rating available in the system - given by the corresponding supervision report. Here, only performance of M&E system seems to be statistically significant. However, given that 24 implementation ratings have been tested - these results stress the absence of selection bias in the case of the IFAD11 sample of IAs.

Table 21: Balance tests: implementation performance ratings for IFAD11 (baseline characteristics)

	IFAD11 IA Sample	*Unselected Projects (closing during IFAD11)*	*IFAD 11 Universe*	*Sample - Unselected*		*Sample - Universe*	
	Mean	Mean	Mean	Diff. in Means	p-score	Diff. in Means	p-score
Assessment of the Overall Implementation Performance	3.96	3.94	3.94	0.02	0.80	0.02	0.80
Likelihood of Achieving the Development Objective	4.04	4.02	4.04	0.02	0.82	0.01	0.94
Effectiveness	3.91	3.96	3.95	-0.05	0.61	-0.04	0.69
Targeting and Outreach	4.08	4.09	4.09	0.00	0.98	-0.01	0.95
Gender equality & women's participation	4.00	4.04	4.04	-0.04	0.71	-0.04	0.62
Agricultural Productivity	4.00	3.97	3.98	0.03	0.66	0.02	0.78
Adaptation to Climate Change	4.08	4.00	4.02	0.08	0.40	0.07	0.48
Institutions and Policy Engagement	4.00	3.97	3.99	0.03	0.73	0.01	0.92
Human and Social Capital and Empowerment	3.95	4.03	4.02	-0.07	0.39	-0.07	0.26
Quality of Beneficiary Participation	4.04	4.05	4.05	-0.01	0.94	-0.01	0.89
Responsiveness of Service Providers	4.00	3.96	3.97	0.04	0.68	0.03	0.70
Environment and Natural Resource Management	4.09	3.98	4.00	0.11	0.21	0.09	0.37
Exit Strategy	3.93	4.00	3.99	-0.07	0.40	-0.05	0.48
Potential for Scaling-up	3.95	4.07	4.04	-0.12	0.15	-0.09	0.13
Quality of Project Management	3.96	3.96	3.97	-0.01	0.97	-0.02	0.89
Knowledge Management	4.00	3.99	3.99	0.01	0.87	0.01	0.89
Coherence between AWPB and Implementation	3.92	3.77	3.82	0.15	0.24	0.09	0.35
Performance of M&E System	4.04	3.73	3.80	0.31	0.00	0.24	0.01
Acceptable Disbursement Rate	2.71	3.10	3.02	-0.39	0.31	-0.31	0.44
Quality of Financial Management	3.96	4.00	3.98	-0.04	0.71	-0.02	0.79
Quality and Timeliness of Audit	4.00	3.93	3.95	0.07	0.38	0.05	0.44
Counterparts Funds	4.08	3.99	4.01	0.10	0.50	0.07	0.64
Compliance with Loan Covenants	4.00	3.98	3.99	0.02	0.83	0.01	0.94
Procurement	3.83	3.94	3.92	-0.11	0.33	-0.09	0.53

8.2 Conclusions for IFAD11

In light of these descriptive analyses and broad considerations, Management can conclude that there is no selection bias in the IFAD11 sample of impact assessments.

However the additional following recommendation can be made, notably adjusting the regional distribution to allow for two more impact assessments in LAC.

Appendix – Annex I

Table 22: Balance test: Implementation Performance Ratings (Baseline Characteristics) by region

	APR						ESA					
	Sample		Unselected		Sample - Unselected		Sample		Unselected		Sample - Unselected	
	n	Mean	n	Mean	Difference	p-score	n	Mean	n	Mean	Difference	p-score
Duration	5	6.80	25	8.56	-1.76	0.31	6	10.00	14	8.29	1.71	0.07
Beneficiaries	5	1,226,531	25	1,108,359	118,173	0.92	6	377,717	14	464,178	-86,461	0.85
Approved Funding	5	74,052,700	25	62,145,347	11,907,353	0.64	6	68,786,299	14	83,084,966	-83,084,966	0.85
Assessment of the Overall Implementation Performance	5	4.20	25	3.88	0.32	0.07	6	4.17	14	3.71	0.45	0.17
Likelihood of Achieving the Development Objective	5	4.00	25	3.92	0.08	0.72	6	4.00	14	3.86	0.14	0.61
Effectiveness	4	3.50	19	3.84	-0.34	0.14	3	4.00	13	3.92	0.08	0.65
Targeting and Outreach	5	4.20	25	3.96	0.24	0.19	6	4.17	14	4.00	0.17	0.40
Gender equality & women's participation	5	4.20	25	4.00	0.20	0.33	6	4.17	14	4.07	0.10	0.54
Agricultural Productivity	5	4.40	19	4.00	0.40	0.00	5	4.00	14	3.71	0.29	0.32
Adaptation to Climate Change	1	4.00	3	4.00	0.00	-	-	-	-	-	-	-
Institutions and Policy Engagement	5	4.40	21	4.00	0.40	0.10	6	4.33	14	4.00	0.33	0.36
Human and Social Capital and Empowerment	4	4.25	21	3.95	0.30	0.36	6	4.17	14	3.79	0.38	0.16
Quality of Beneficiary Participation	5	4.00	25	4.04	-0.04	0.80	6	4.00	14	4.00	0.00	-
Responsiveness of Service Providers	5	4.40	25	3.84	0.56	0.01	6	3.67	14	4.07	-0.41	0.11
Environment and Natural Resource Management	1	4.00	3	4.00	0.00	-	-	-	-	-	-	-
Exit Strategy	4	4.25	19	4.00	0.25	0.03	3	4.00	3	3.67	0.33	0.37
Potential for Scaling-up	4	4.25	19	4.00	0.25	0.03	5	4.40	14	4.14	0.26	0.51
Quality of Project Management	5	4.00	25	3.84	0.16	0.61	6	4.17	14	3.64	0.52	0.24
Knowledge Management	4	4.00	22	3.86	0.14	0.45	6	4.17	13	3.92	0.24	0.14
Coherence between AWPB and Implementation	5	4.00	22	3.82	0.18	0.32	6	4.17	14	3.86	0.31	0.37
Performance of M&E System	5	3.80	24	3.88	-0.08	0.80	6	3.50	14	3.71	-0.21	0.47
Acceptable Disbursement Rate	5	4.20	25	3.32	0.88	0.14	6	4.17	14	3.86	0.31	0.62
Quality of Financial Management	5	4.40	22	3.91	0.49	0.04	5	3.80	14	3.71	0.09	0.81

Quality and Timeliness of Audit	5	4.40	24	3.92	0.48	0.07	6	4.00	14	3.86	0.14	0.53
Counterparts Funds	5	4.40	25	4.00	0.40	0.28	6	4.17	14	4.29	-0.12	0.80
Compliance with Loan Covenants	5	4.40	25	3.92	0.48	0.01	6	4.33	14	3.79	0.55	0.10
Procurement	5	4.40	25	3.96	0.44	0.03	6	3.83	14	3.79	0.05	0.88

	LAC						**NEN**					
	Sample		**Unselected**		**Sample - Unselected**		**Sample**		**Unselected**		**Sample - Unselected**	
	n	**Mean**	**n**	**Mean**	**Difference**	**p-score**	**n**	**Mean**	**n**	**Mean**	**Difference**	**p-score**
Duration	3	6.667	15	8.00	-1.33	0.35	1	7.00	12	7.92	-0.92	0.66
Beneficiaries	3	40,518	15	63,093	-22,576	0.62	1	145,600	12	92,023	53,577	0.58
Approved Funding	3	31,437,826	15	25,628,214	5,809,612	0.57	1	15,780,852	12	38,844,717	-23,063,865	0.31
Assessment of the Overall Implementation Performance	3	4	15	3.53	0.47	0.36	1	4.00	12	4.00	0.00	1.00
Likelihood of Achieving the Development Objective	3	4.00	15	3.73	0.27	0.46	1	4.00	12	4.17	-0.17	0.79
Effectiveness	1	4.00	10	3.60	0.40	0.66	1	4.00	8	3.75	0.25	0.75
Targeting and Outreach	3	4.67	15	4.13	0.53	0.20	1	4.00	12	4.42	-0.42	0.45
Gender equality & women's participation	3	3.67	15	3.67	0.00	1.00	1	4.00	12	4.08	-0.08	0.88
Agricultural Productivity	1	4.00	10	3.90	0.10	0.90	1	4.00	10	3.90	0.10	0.87
Adaptation to *Climate* Change	-	-	-	-	-	-	-	-	-	-	-	-
Institutions and Policy Engagement	3	3.33	11	3.82	-0.49	0.44	1	4.00	12	3.92	0.08	0.88
Human and Social Capital and Empowerment	1	4.00	11	3.91	0.09	0.92	1	4.00	12	3.92	0.08	0.88
Quality of Beneficiary Participation	3	3.67	15	3.87	-0.20	0.74	1	4.00	12	4.08	-0.08	0.88
Responsiveness of Service Providers	3	3.67	15	3.73	-0.07	0.83	1	4.00	12	4.08	-0.08	0.91
Environment and Natural Resource Management	-	-	-	-	-	-	-	-	-	-	-	-
Exit Strategy	1	4.00	9	3.67	0.33	0.67	1	4.00	11	4.09	-0.09	0.90
Potential for Scaling-up	1	4.00	10	3.80	0.20	0.81	1	4.00	11	4.27	-0.27	0.75
Quality of Project Management	3	4.00	15	3.67	0.33	0.60	1	3.00	12	4.08	-1.08	0.15
Knowledge Management	2	5.00	10	4.00	1.00	0.18	1	4.00	12	4.25	-0.25	0.71
Coherence between AWPB and Implementation	2	4.50	12	3.50	1.00	0.12	1	4.00	12	3.83	0.17	0.79

Performance of M&E System	3	3.67	15	3.53	0.13	0.78	1	4.00	12	4.08	-0.08	0.88
Acceptable Disbursement Rate	3	4.33	15	3.47	0.87	0.35	1	1.00	12	3.83	-2.83	0.14
Quality of Financial Management	3	4.00	13	3.85	0.15	0.83	1	4.00	11	4.18	-0.18	0.68
Quality and Timeliness of Audit	3	4.00	15	4.00	0.00	1.00	1	4.00	12	4.25	-0.25	0.76
Counterparts Funds	3	4.67	15	3.67	1.00	0.06	1	4.00	12	4.33	-0.33	0.63
Compliance with Loan Covenants	3	4.00	15	4.00	0.00	1.00	1	3.00	12	4.17	-1.17	0.02
Procurement	3	4.67	15	4.07	0.60	0.28	1	4.00	12	4.08	-0.08	0.92

	WCA					
	Sample		**Unselected**		**Sample - Unselected**	
	n	**Mean**	**n**	**Mean**	**Difference**	**p-score**
Duration	4	8.50	22	8.32	0.18	0.90
Beneficiaries	4	162583	22	897,950	-735,368	0.55
Approved Funding	4	22364480	22	40,751,001	-18,386,521	0.34
Assessment of the Overall Implementation Performance	4	4.00	22	4.05	-0.05	0.68
Likelihood of Achieving the Development Objective	4	4.00	22	4.18	-0.18	0.37
Effectiveness	3	4.00	19	4.11	-0.11	0.58
Targeting and Outreach	4	4.25	22	4.18	0.07	0.80
Gender equality & women's participation	4	4.00	22	4.09	-0.09	0.77
Agricultural Productivity	3	4.00	18	4.11	-0.11	0.57
Adaptation to Climate Change	1	4.00	3	4.33	-0.33	0.67
Institutions and Policy Engagement	3	4.00	20	4.20	-0.20	0.52
Human and Social Capital and Empowerment	3	4.00	19	4.00	0.00	1.00
Quality of Beneficiary Participation	4	4.00	22	4.23	-0.23	0.41
Responsiveness of Service Providers	4	3.75	22	4.14	-0.39	0.21
Environment and Natural Resource Management	1	4.00	5	4.00	0.00	-
Exit Strategy	2	4.00	16	4.06	-0.06	0.74
Potential for Scaling-up	3	4.00	18	4.11	-0.11	0.57
Quality of Project Management	4	3.75	22	4.00	-0.25	0.56
Knowledge Management	3	4.00	19	4.16	-0.16	0.48

Coherence between AWPB and Implementation	3	4.00	21	3.86	0.14	0.79						
Performance of M&E System	4	4.00	22	3.91	0.09	0.81						
Acceptable Disbursement Rate	4	5.00	22	3.05	1.96	0.02						
Quality of Financial Management	3	4.33	19	3.90	0.44	0.23						
Quality and Timeliness of Audit	4	4.00	22	4.09	-0.09	0.86						
Counterparts Funds	4	4.75	22	3.91	0.84	0.01						
Compliance with Loan Covenants	4	4.25	22	4.23	0.02	0.93						
Procurement	4	4.00	22	4.09	-0.09	0.68						

Appendix - Annex II

Table 23

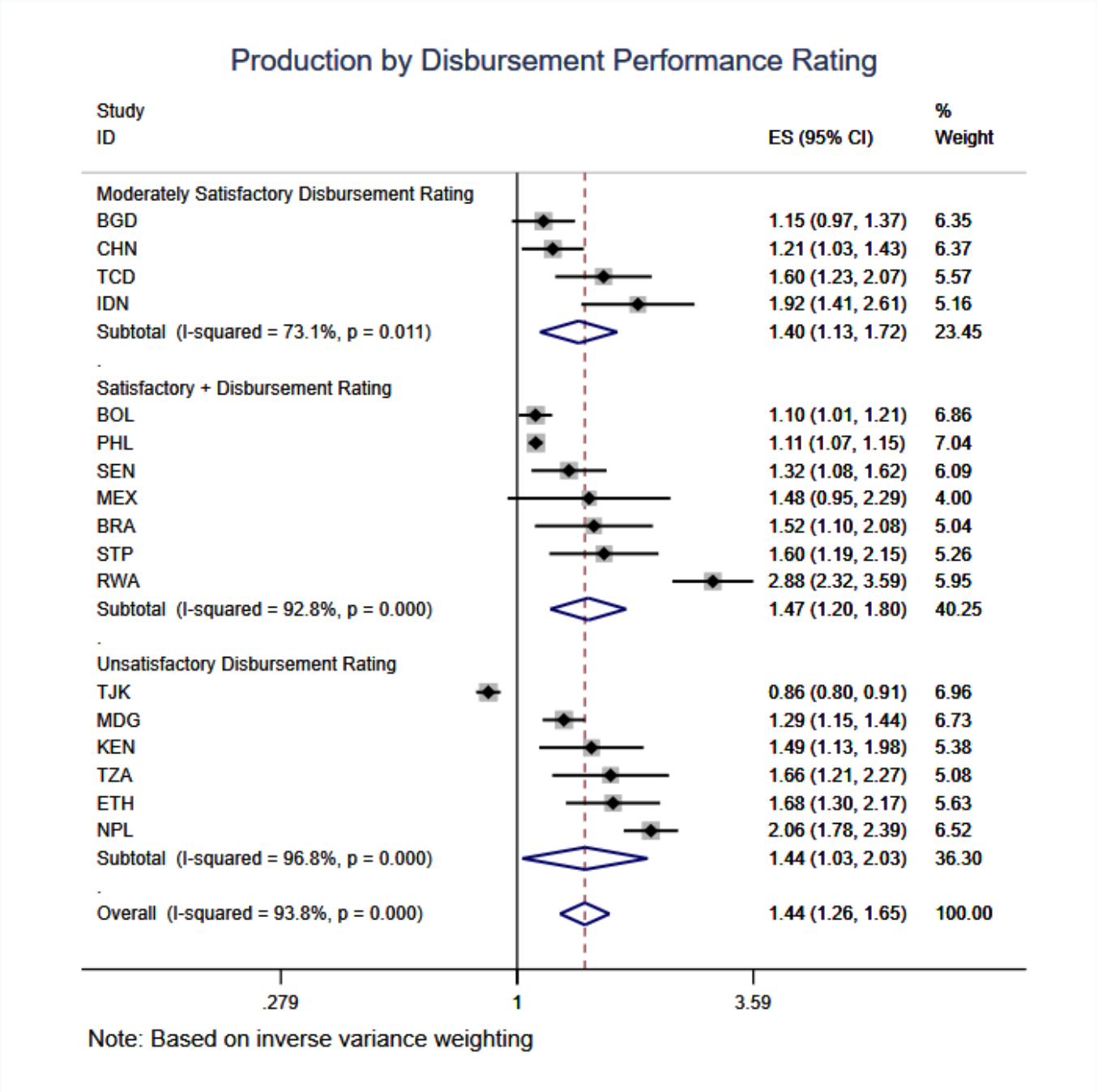


Table 24

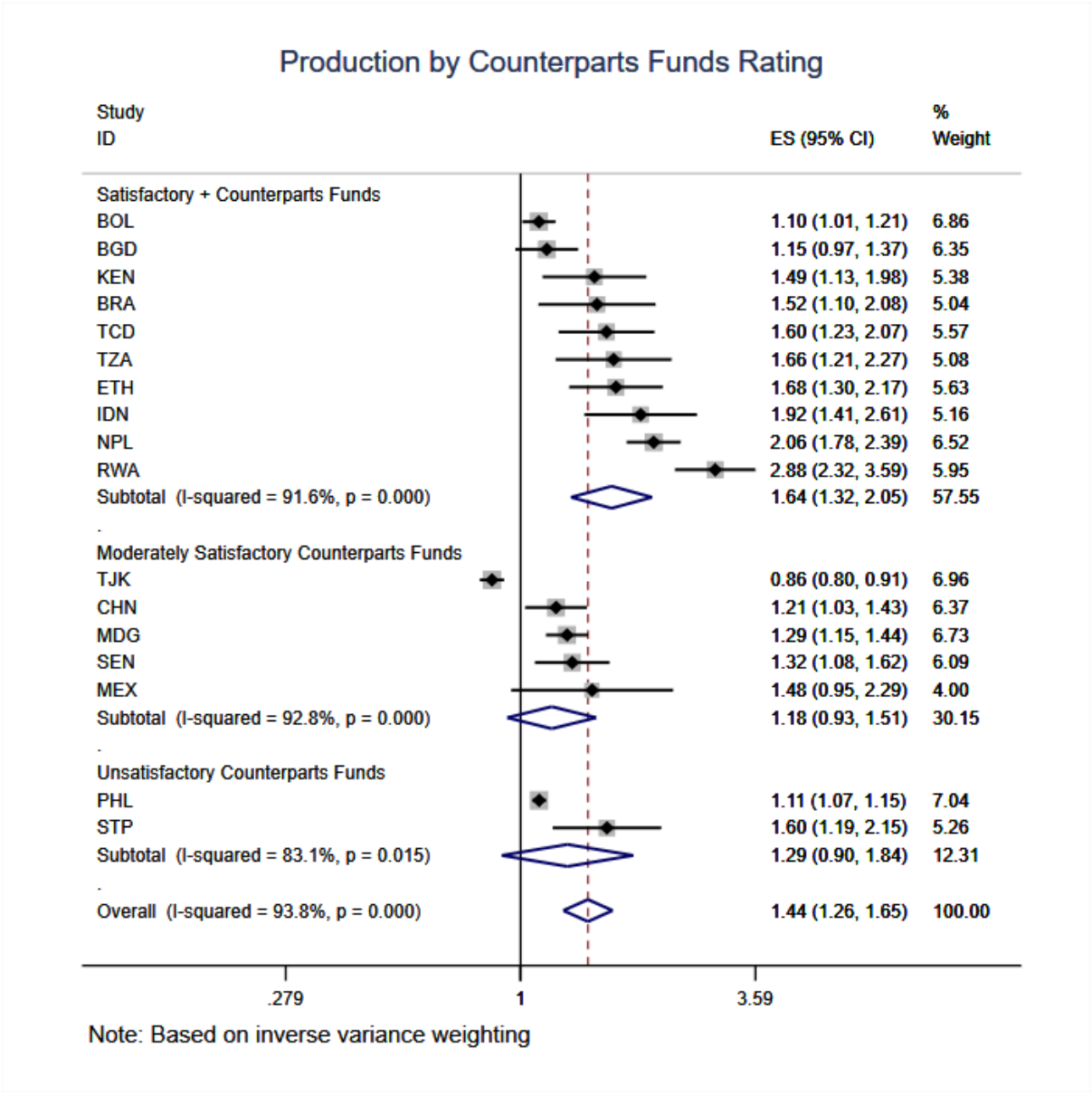

Production by Counterparts Funds Rating
Study ID
ES (95% CI)
% Weight
Satisfactory + Counterparts Funds
BOL 1.10 (1.01, 1.21) 6.86
BGD 1.15 (0.97, 1.37) 6.35
KEN 1.49 (1.13, 1.98) 5.38
BRA 1.52 (1.10, 2.08) 5.04
TCD 1.60 (1.23, 2.07) 5.57
TZA 1.66 (1.21, 2.27) 5.08
ETH 1.68 (1.30, 2.17) 5.63
IDN 1.92 (1.41, 2.61) 5.16
NPL 2.06 (1.78, 2.39) 6.52
RWA 2.88 (2.32, 3.59) 5.95
Subtotal (I-squared = 91.6%, p = 0.000) 1.64 (1.32, 2.05) 57.55
Moderately Satisfactory Counterparts Funds
TJK 0.86 (0.80, 0.91) 6.96
CHN 1.21 (1.03, 1.43) 6.37
MDG 1.29 (1.15, 1.44) 6.73
SEN 1.32 (1.08, 1.62) 6.09
MEX 1.48 (0.95, 2.29) 4.00
Subtotal (I-squared = 92.8%, p = 0.000) 1.18 (0.93, 1.51) 30.15
Unsatisfactory Counterparts Funds
PHL 1.11 (1.07, 1.15) 7.04
STP 1.60 (1.19, 2.15) 5.26
Subtotal (I-squared = 83.1%, p = 0.015) 1.29 (0.90, 1.84) 12.31
Overall (I-squared = 93.8%, p = 0.000) 1.44 (1.26, 1.65) 100.00
.279
1
3.59
Note: Based on inverse variance weighting

Table 25

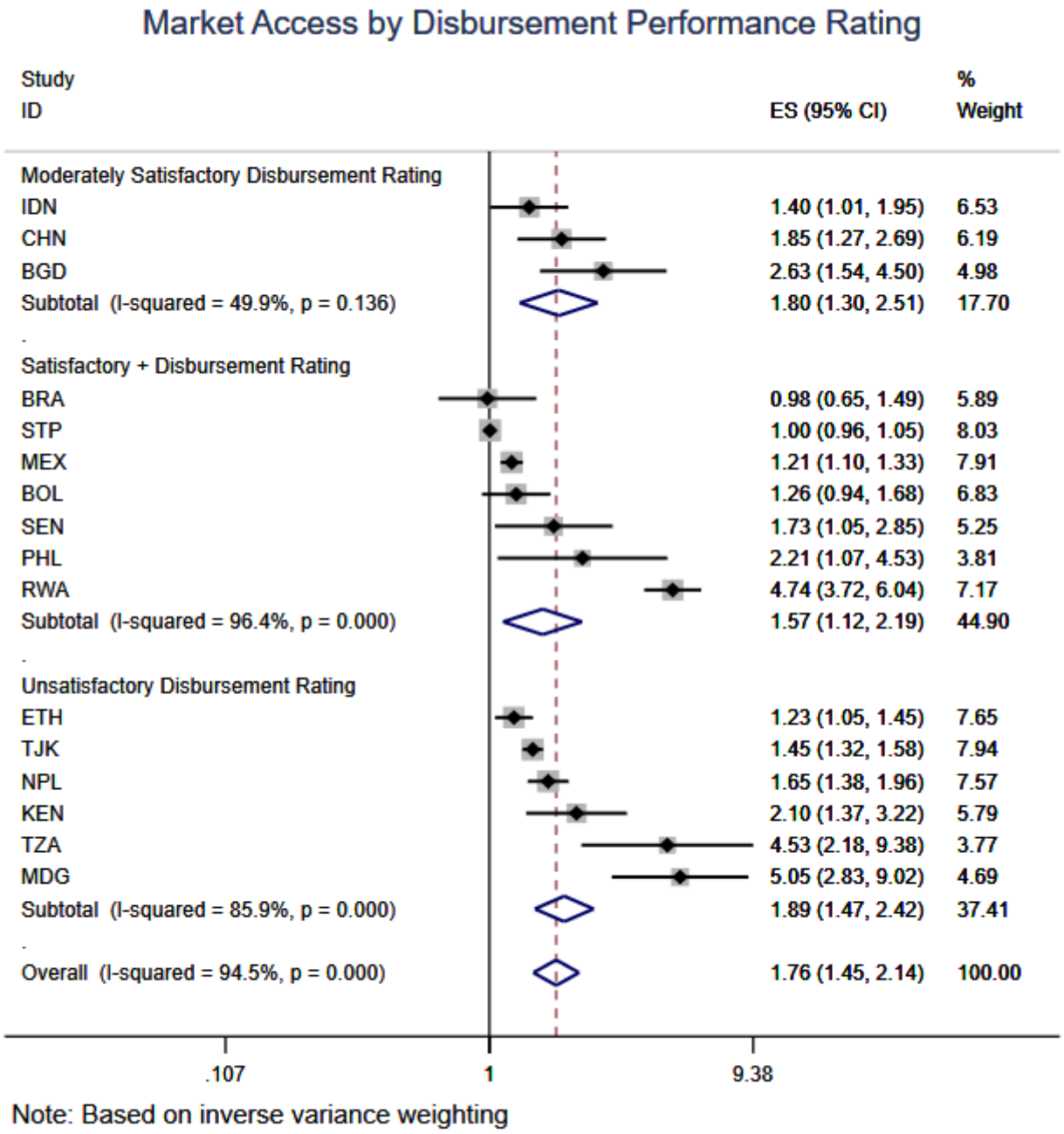

Market Access by Disbursement Performance Rating
Study ID
ES (95% CI)
% Weight
Moderately Satisfactory Disbursement Rating
IDN 1.40 (1.01, 1.95) 6.53
CHN 1.85 (1.27, 2.69) 6.19
BGD 2.63 (1.54, 4.50) 4.98
Subtotal (I-squared = 49.9%, p = 0.136) 1.80 (1.30, 2.51) 17.70
Satisfactory + Disbursement Rating
BRA 0.98 (0.65, 1.49) 5.89
STP 1.00 (0.96, 1.05) 8.03
MEX 1.21 (1.10, 1.33) 7.91
BOL 1.26 (0.94, 1.68) 6.83
SEN 1.73 (1.05, 2.85) 5.25
PHL 2.21 (1.07, 4.53) 3.81
RWA 4.74 (3.72, 6.04) 7.17
Subtotal (I-squared = 96.4%, p = 0.000) 1.57 (1.12, 2.19) 44.90
Unsatisfactory Disbursement Rating
ETH 1.23 (1.05, 1.45) 7.65
TJK 1.45 (1.32, 1.58) 7.94
NPL 1.65 (1.38, 1.96) 7.57
KEN 2.10 (1.37, 3.22) 5.79
TZA 4.53 (2.18, 9.38) 3.77
MDG 5.05 (2.83, 9.02) 4.69
Subtotal (I-squared = 85.9%, p = 0.000) 1.89 (1.47, 2.42) 37.41
Overall (I-squared = 94.5%, p = 0.000) 1.76 (1.45, 2.14) 100.00
.107
1
9.38
Note: Based on inverse variance weighting

Table 26

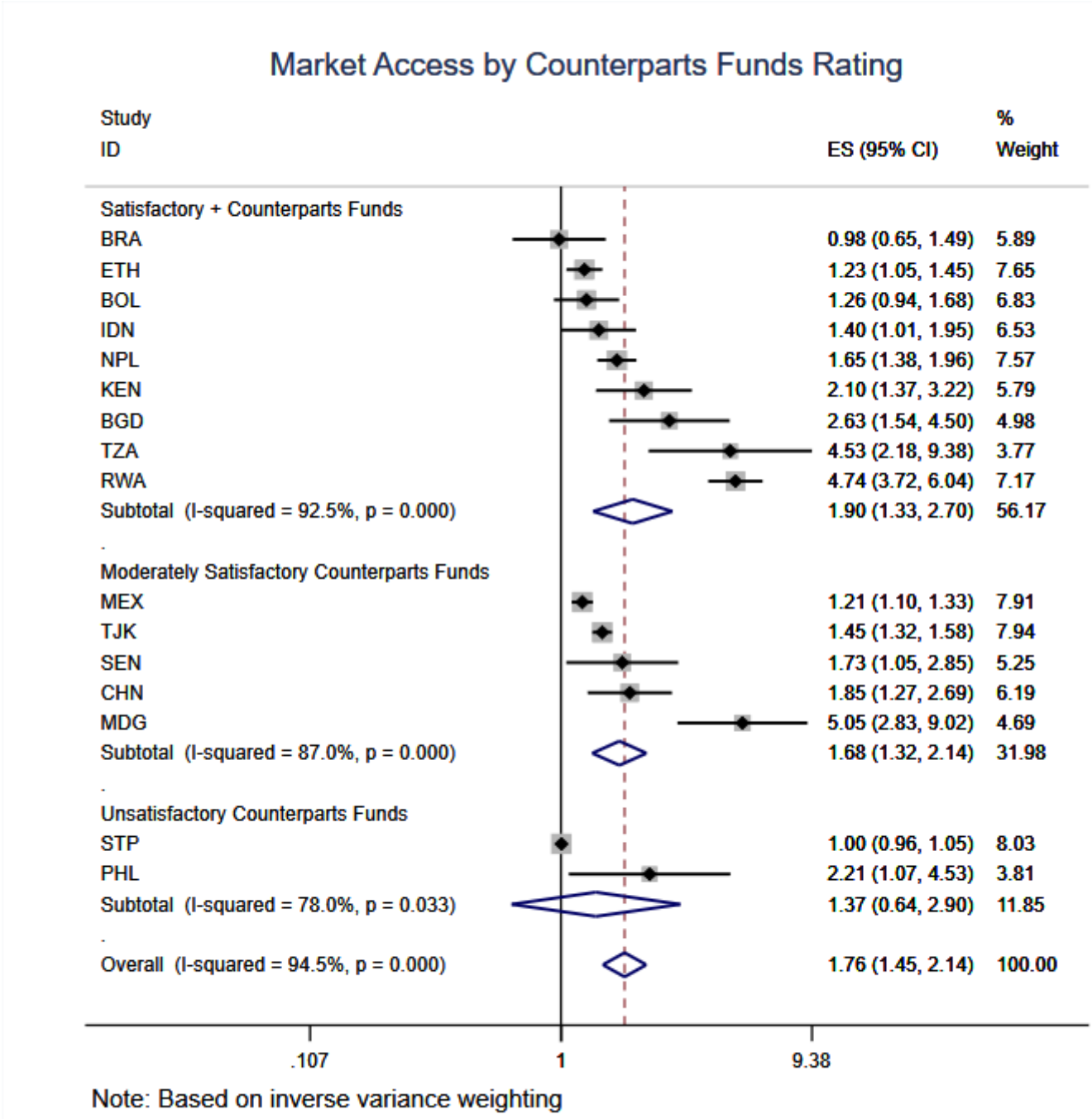

Market Access by Counterparts Funds Rating
Study ID
ES (95% CI)
% Weight
Satisfactory + Counterparts Funds
BRA 0.98 (0.65, 1.49) 5.89
ETH 1.23 (1.05, 1.45) 7.65
BOL 1.26 (0.94, 1.68) 6.83
IDN 1.40 (1.01, 1.95) 6.53
NPL 1.65 (1.38, 1.96) 7.57
KEN 2.10 (1.37, 3.22) 5.79
BGD 2.63 (1.54, 4.50) 4.98
TZA 4.53 (2.18, 9.38) 3.77
RWA 4.74 (3.72, 6.04) 7.17
Subtotal (I-squared = 92.5%, p = 0.000) 1.90 (1.33, 2.70) 56.17
Moderately Satisfactory Counterparts Funds
MEX 1.21 (1.10, 1.33) 7.91
TJK 1.45 (1.32, 1.58) 7.94
SEN 1.73 (1.05, 2.85) 5.25
CHN 1.85 (1.27, 2.69) 6.19
MDG 5.05 (2.83, 9.02) 4.69
Subtotal (I-squared = 87.0%, p = 0.000) 1.68 (1.32, 2.14) 31.98
Unsatisfactory Counterparts Funds
STP 1.00 (0.96, 1.05) 8.03
PHL 2.21 (1.07, 4.53) 3.81
Subtotal (I-squared = 78.0%, p = 0.033) 1.37 (0.64, 2.90) 11.85
Overall (I-squared = 94.5%, p = 0.000) 1.76 (1.45, 2.14) 100.00
.107
1
9.38
Note: Based on inverse variance weighting

Table 27

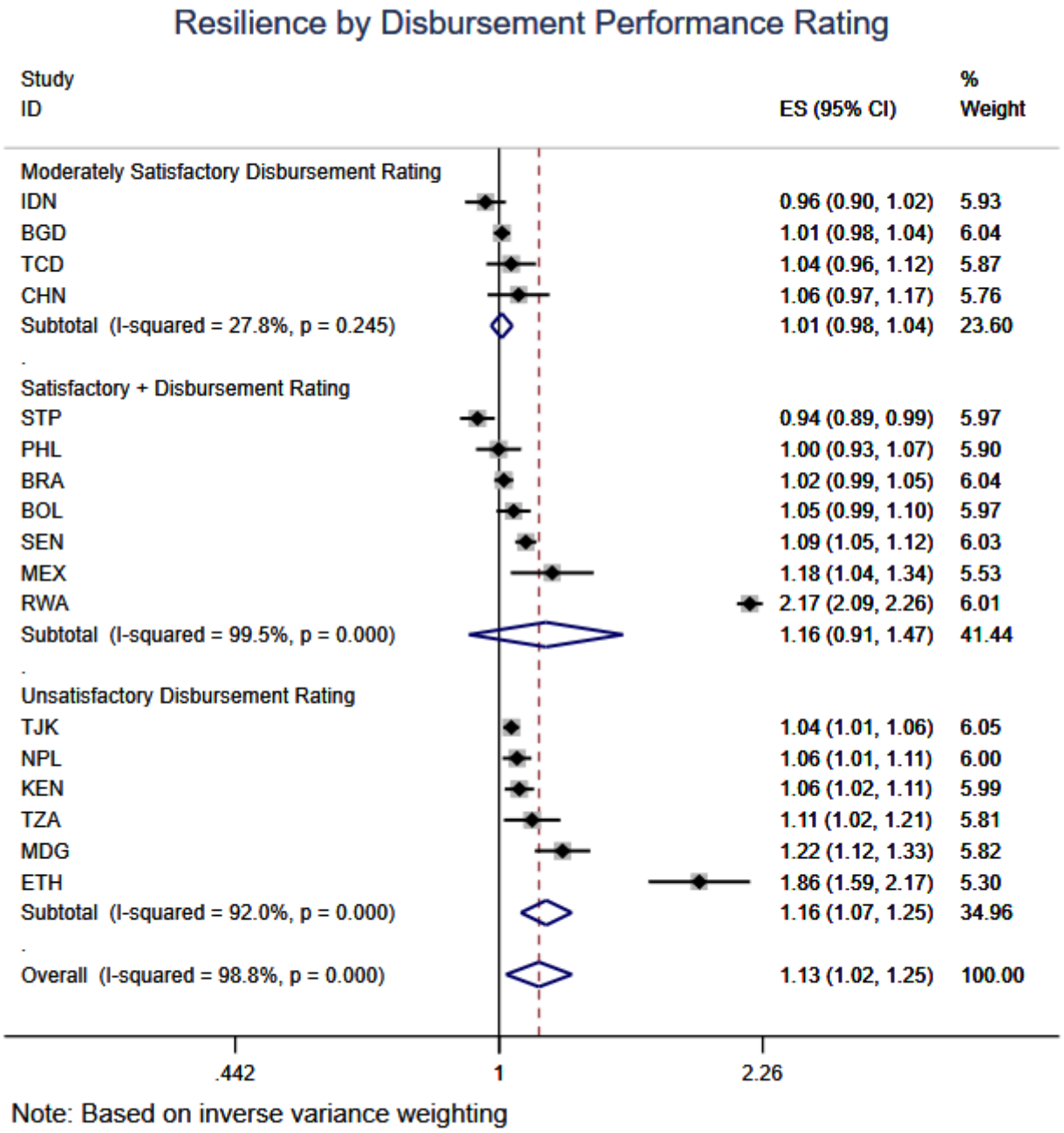

Resilience by Disbursement Performance Rating
Study ID
ES (95% CI)
% Weight
Moderately Satisfactory Disbursement Rating
IDN 0.96 (0.90, 1.02) 5.93
BGD 1.01 (0.98, 1.04) 6.04
TCD 1.04 (0.96, 1.12) 5.87
CHN 1.06 (0.97, 1.17) 5.76
Subtotal (I-squared = 27.8%, p = 0.245) 1.01 (0.98, 1.04) 23.60
Satisfactory + Disbursement Rating
STP 0.94 (0.89, 0.99) 5.97
PHL 1.00 (0.93, 1.07) 5.90
BRA 1.02 (0.99, 1.05) 6.04
BOL 1.05 (0.99, 1.10) 5.97
SEN 1.09 (1.05, 1.12) 6.03
MEX 1.18 (1.04, 1.34) 5.53
RWA 2.17 (2.09, 2.26) 6.01
Subtotal (I-squared = 99.5%, p = 0.000) 1.16 (0.91, 1.47) 41.44
Unsatisfactory Disbursement Rating
TJK 1.04 (1.01, 1.06) 6.05
NPL 1.06 (1.01, 1.11) 6.00
KEN 1.06 (1.02, 1.11) 5.99
TZA 1.11 (1.02, 1.21) 5.81
MDG 1.22 (1.12, 1.33) 5.82
ETH 1.86 (1.59, 2.17) 5.30
Subtotal (I-squared = 92.0%, p = 0.000) 1.16 (1.07, 1.25) 34.96
Overall (I-squared = 98.8%, p = 0.000) 1.13 (1.02, 1.25) 100.00
.442
1
2.26
Note: Based on inverse variance weighting

Table 28

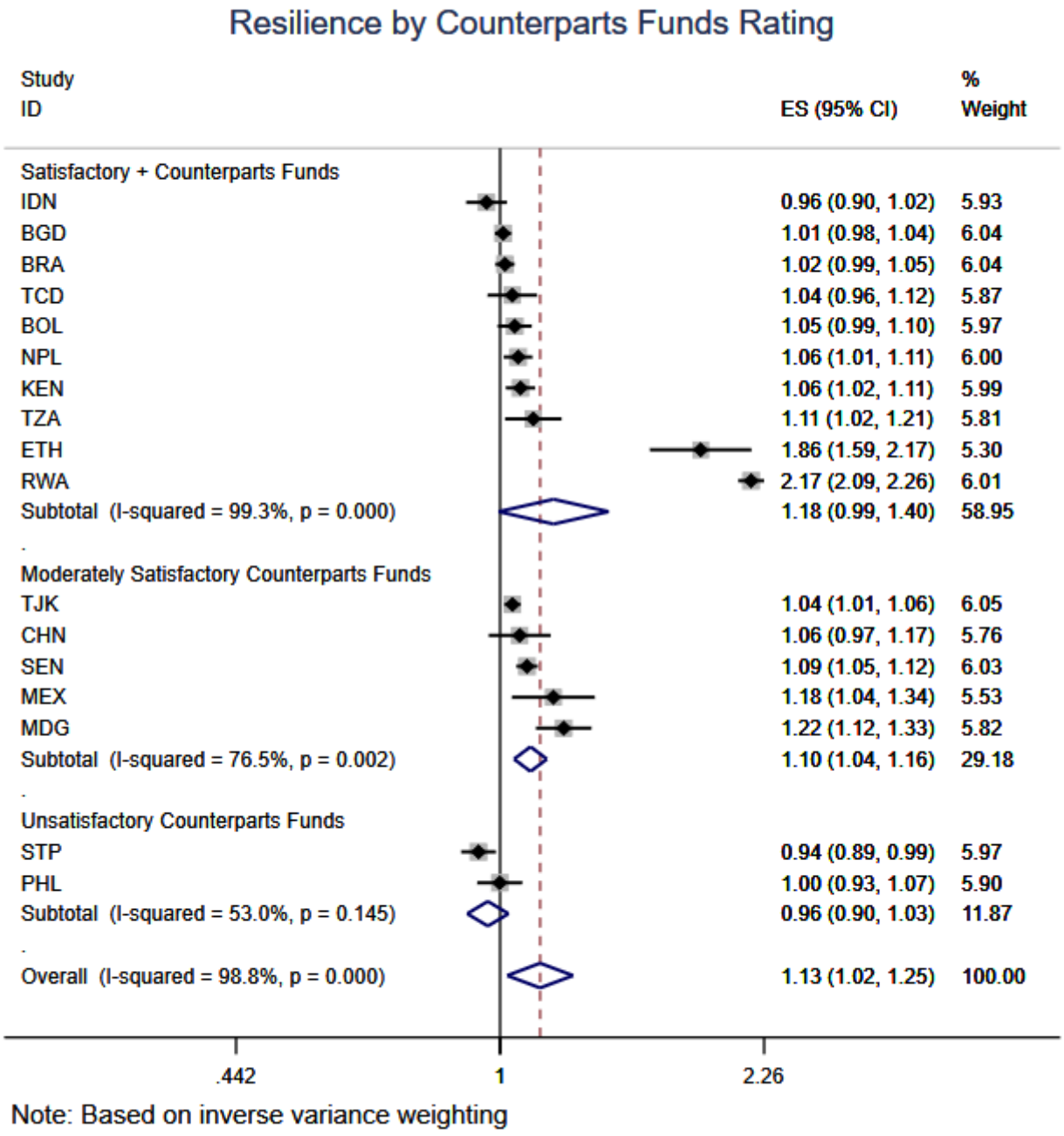

Resilience by Counterparts Funds Rating
Study ID
ES (95% CI)
% Weight
Satisfactory + Counterparts Funds
IDN 0.96 (0.90, 1.02) 5.93
BGD 1.01 (0.98, 1.04) 6.04
BRA 1.02 (0.99, 1.05) 6.04
TCD 1.04 (0.96, 1.12) 5.87
BOL 1.05 (0.99, 1.10) 5.97
NPL 1.06 (1.01, 1.11) 6.00
KEN 1.06 (1.02, 1.11) 5.99
TZA 1.11 (1.02, 1.21) 5.81
ETH 1.86 (1.59, 2.17) 5.30
RWA 2.17 (2.09, 2.26) 6.01
Subtotal (I-squared = 99.3%, p = 0.000) 1.18 (0.99, 1.40) 58.95
Moderately Satisfactory Counterparts Funds
TJK 1.04 (1.01, 1.06) 6.05
CHN 1.06 (0.97, 1.17) 5.76
SEN 1.09 (1.05, 1.12) 6.03
MEX 1.18 (1.04, 1.34) 5.53
MDG 1.22 (1.12, 1.33) 5.82
Subtotal (I-squared = 76.5%, p = 0.002) 1.10 (1.04, 1.16) 29.18
Unsatisfactory Counterparts Funds
STP 0.94 (0.89, 0.99) 5.97
PHL 1.00 (0.93, 1.07) 5.90
Subtotal (I-squared = 53.0%, p = 0.145) 0.96 (0.90, 1.03) 11.87
Overall (I-squared = 98.8%, p = 0.000) 1.13 (1.02, 1.25) 100.00
.442
1
2.26
Note: Based on inverse variance weighting

Table 29

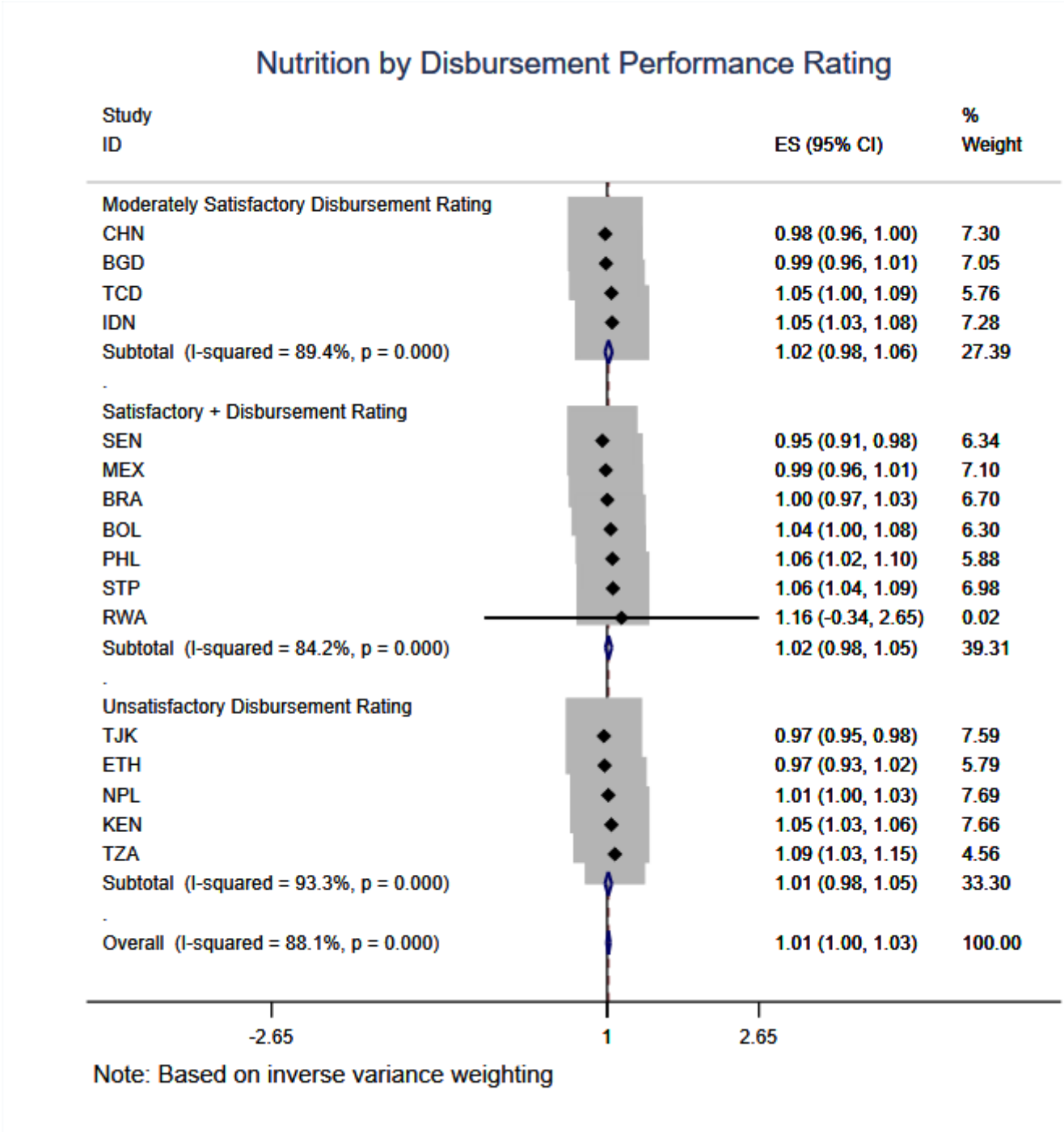

Nutrition by Disbursement Performance Rating
Study ID
ES (95% CI)
% Weight
Moderately Satisfactory Disbursement Rating
CHN 0.98 (0.96, 1.00) 7.30
BGD 0.99 (0.96, 1.01) 7.05
TCD 1.05 (1.00, 1.09) 5.76
IDN 1.05 (1.03, 1.08) 7.28
Subtotal (I-squared = 89.4%, p = 0.000) 1.02 (0.98, 1.06) 27.39
Satisfactory + Disbursement Rating
SEN 0.95 (0.91, 0.98) 6.34
MEX 0.99 (0.96, 1.01) 7.10
BRA 1.00 (0.97, 1.03) 6.70
BOL 1.04 (1.00, 1.08) 6.30
PHL 1.06 (1.02, 1.10) 5.88
STP 1.06 (1.04, 1.09) 6.98
RWA 1.16 (-0.34, 2.65) 0.02
Subtotal (I-squared = 84.2%, p = 0.000) 1.02 (0.98, 1.05) 39.31
Unsatisfactory Disbursement Rating
TJK 0.97 (0.95, 0.98) 7.59
ETH 0.97 (0.93, 1.02) 5.79
NPL 1.01 (1.00, 1.03) 7.69
KEN 1.05 (1.03, 1.06) 7.66
TZA 1.09 (1.03, 1.15) 4.56
Subtotal (I-squared = 93.3%, p = 0.000) 1.01 (0.98, 1.05) 33.30
Overall (I-squared = 88.1%, p = 0.000) 1.01 (1.00, 1.03) 100.00
-2.65
1
2.65
Note: Based on inverse variance weighting

Table 30

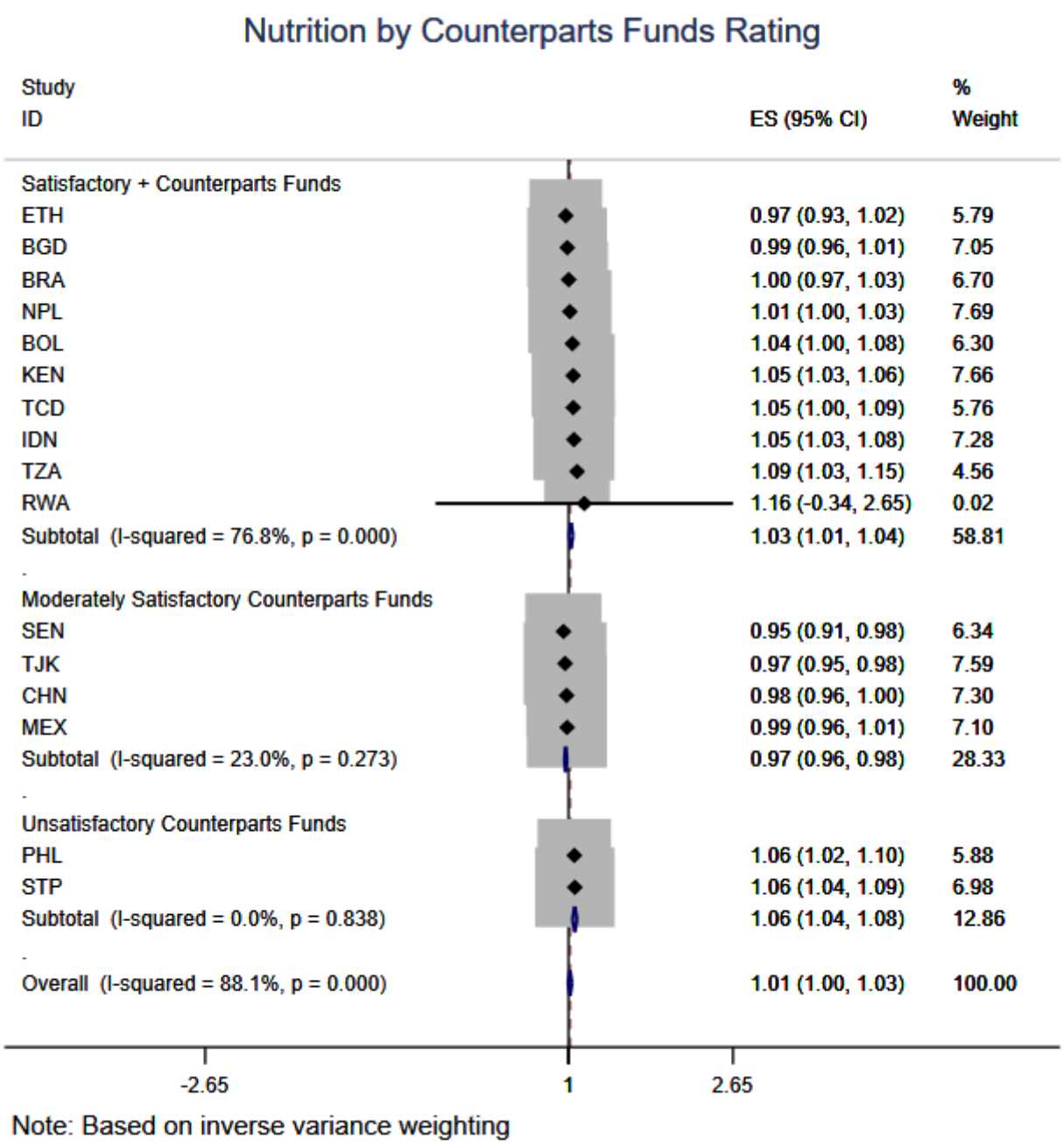

Nutrition by Counterparts Funds Rating
Study ID
ES (95% CI)
% Weight
Satisfactory + Counterparts Funds
ETH 0.97 (0.93, 1.02) 5.79
BGD 0.99 (0.96, 1.01) 7.05
BRA 1.00 (0.97, 1.03) 6.70
NPL 1.01 (1.00, 1.03) 7.69
BOL 1.04 (1.00, 1.08) 6.30
KEN 1.05 (1.03, 1.06) 7.66
TCD 1.05 (1.00, 1.09) 5.76
IDN 1.05 (1.03, 1.08) 7.28
TZA 1.09 (1.03, 1.15) 4.56
RWA 1.16 (-0.34, 2.65) 0.02
Subtotal (I-squared = 76.8%, p = 0.000) 1.03 (1.01, 1.04) 58.81
Moderately Satisfactory Counterparts Funds
SEN 0.95 (0.91, 0.98) 6.34
TJK 0.97 (0.95, 0.98) 7.59
CHN 0.98 (0.96, 1.00) 7.30
MEX 0.99 (0.96, 1.01) 7.10
Subtotal (I-squared = 23.0%, p = 0.273) 0.97 (0.96, 0.98) 28.33
Unsatisfactory Counterparts Funds
PHL 1.06 (1.02, 1.10) 5.88
STP 1.06 (1.04, 1.09) 6.98
Subtotal (I-squared = 0.0%, p = 0.838) 1.06 (1.04, 1.08) 12.86
Overall (I-squared = 88.1%, p = 0.000) 1.01 (1.00, 1.03) 100.00
-2.65
1
2.65
Note: Based on inverse variance weighting

Table 31

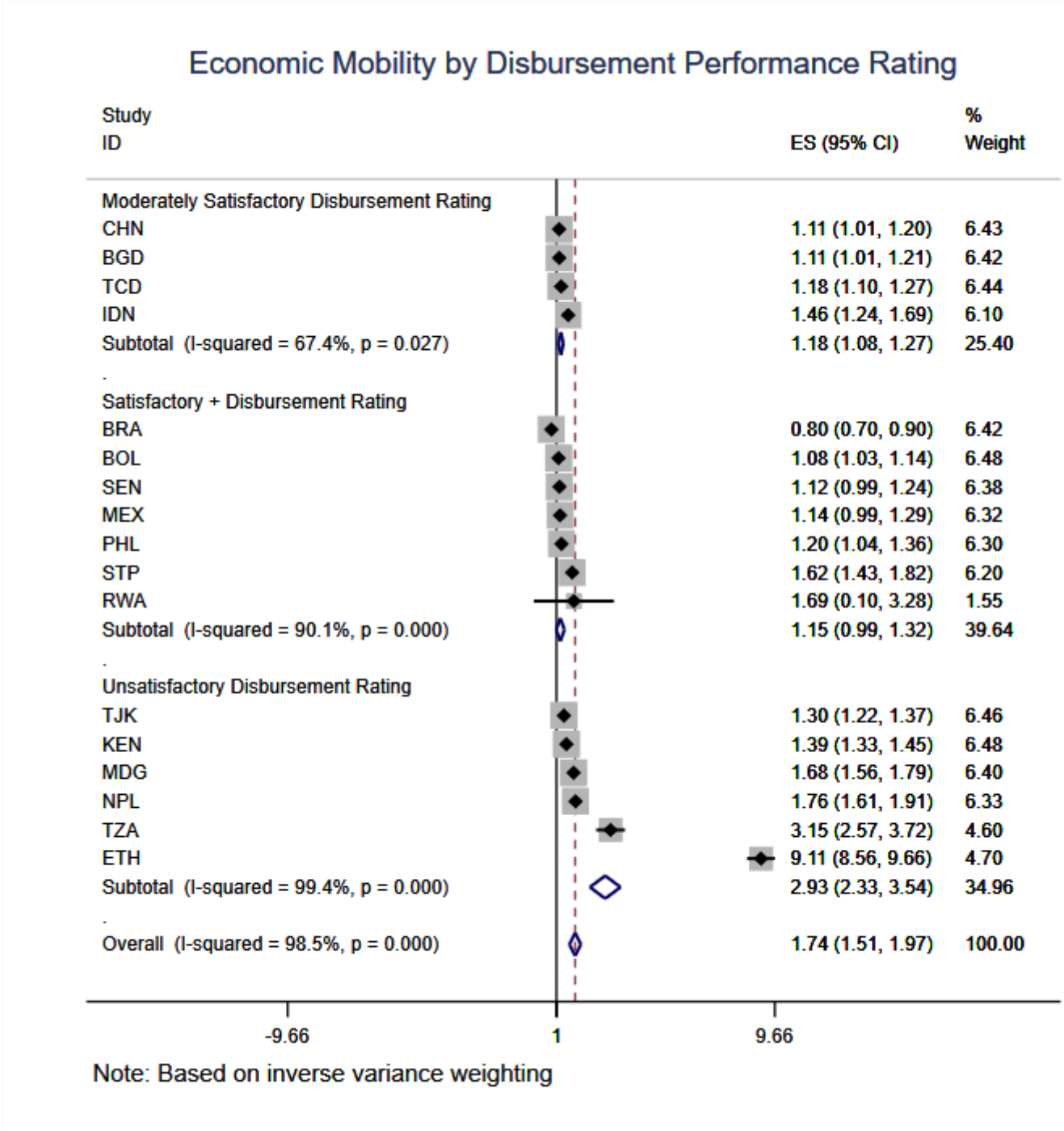
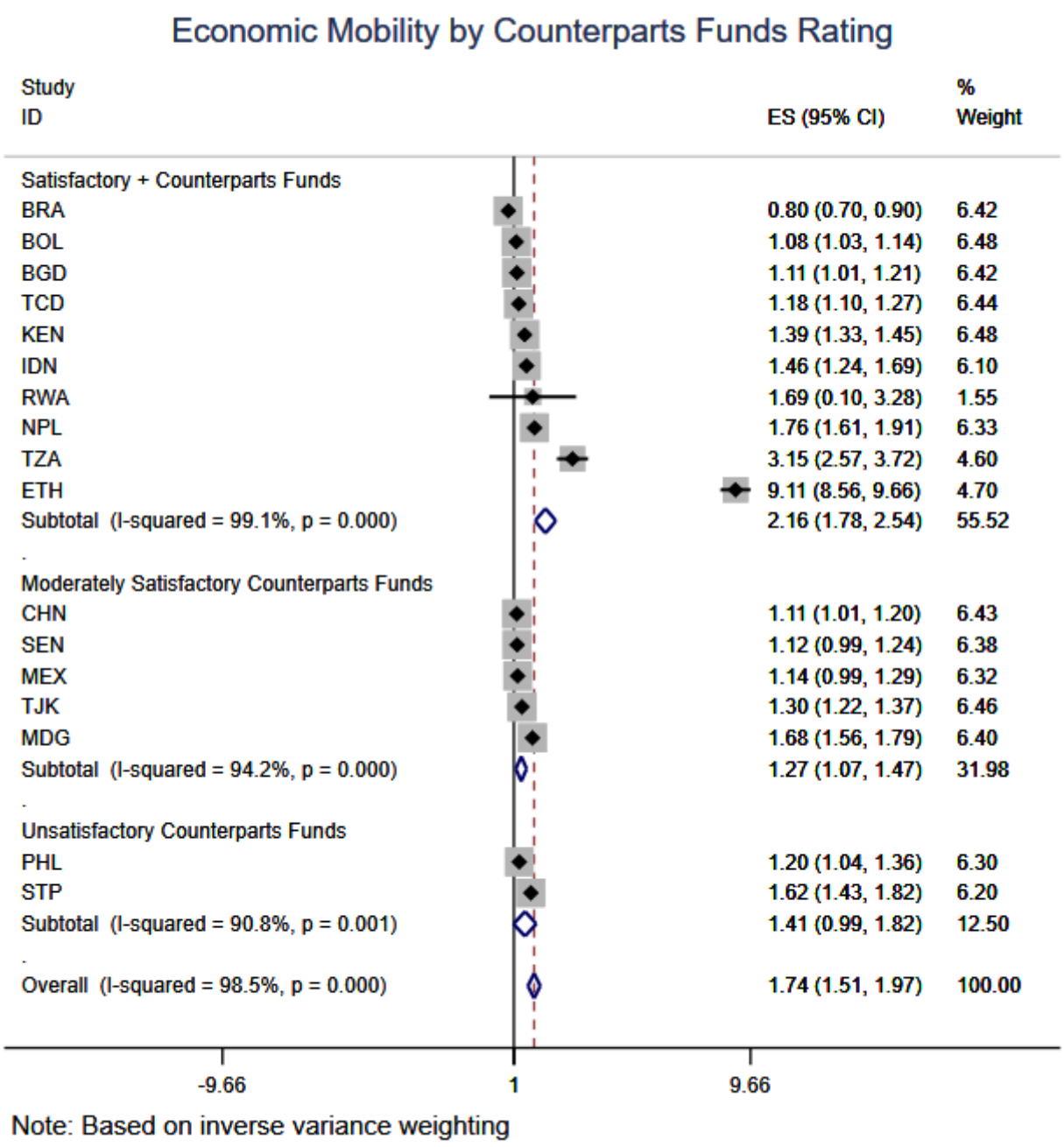

Economic Mobility by Disbursement Performance Rating
Study ID
ES (95% CI)
% Weight
Moderately Satisfactory Disbursement Rating
CHN 1.11 (1.01, 1.20) 6.43
BGD 1.11 (1.01, 1.21) 6.42
TCD 1.18 (1.10, 1.27) 6.44
IDN 1.46 (1.24, 1.69) 6.10
Subtotal (I-squared = 67.4%, p = 0.027) 1.18 (1.08, 1.27) 25.40
Satisfactory + Disbursement Rating
BRA 0.80 (0.70, 0.90) 6.42
BOL 1.08 (1.03, 1.14) 6.48
SEN 1.12 (0.99, 1.24) 6.38
MEX 1.14 (0.99, 1.29) 6.32
PHL 1.20 (1.04, 1.36) 6.30
STP 1.62 (1.43, 1.82) 6.20
RWA 1.69 (0.10, 3.28) 1.55
Subtotal (I-squared = 90.1%, p = 0.000) 1.15 (0.99, 1.32) 39.64
Unsatisfactory Disbursement Rating
TJK 1.30 (1.22, 1.37) 6.46
KEN 1.39 (1.33, 1.45) 6.48
MDG 1.68 (1.56, 1.79) 6.40
NPL 1.76 (1.61, 1.91) 6.33
TZA 3.15 (2.57, 3.72) 4.60
ETH 9.11 (8.56, 9.66) 4.70
Subtotal (I-squared = 99.4%, p = 0.000) 2.93 (2.33, 3.54) 34.96
Overall (I-squared = 98.5%, p = 0.000) 1.74 (1.51, 1.97) 100.00
-9.66
1
9.66
Note: Based on inverse variance weighting

Table 32


Economic Mobility by Counterparts Funds Rating
Study ID
ES (95% CI)
% Weight
Satisfactory + Counterparts Funds
BRA 0.80 (0.70, 0.90) 6.42
BOL 1.08 (1.03, 1.14) 6.48
BGD 1.11 (1.01, 1.21) 6.42
TCD 1.18 (1.10, 1.27) 6.44
KEN 1.39 (1.33, 1.45) 6.48
IDN 1.46 (1.24, 1.69) 6.10
RWA 1.69 (0.10, 3.28) 1.55
NPL 1.76 (1.61, 1.91) 6.33
TZA 3.15 (2.57, 3.72) 4.60
ETH 9.11 (8.56, 9.66) 4.70
Subtotal (I-squared = 99.1%, p = 0.000) 2.16 (1.78, 2.54) 55.52
Moderately Satisfactory Counterparts Funds
CHN 1.11 (1.01, 1.20) 6.43
SEN 1.12 (0.99, 1.24) 6.38
MEX 1.14 (0.99, 1.29) 6.32
TJK 1.30 (1.22, 1.37) 6.46
MDG 1.68 (1.56, 1.79) 6.40
Subtotal (I-squared = 94.2%, p = 0.000) 1.27 (1.07, 1.47) 31.98
Unsatisfactory Counterparts Funds
PHL 1.20 (1.04, 1.36) 6.30
STP 1.62 (1.43, 1.82) 6.20
Subtotal (I-squared = 90.8%, p = 0.001) 1.41 (0.99, 1.82) 12.50
Overall (I-squared = 98.5%, p = 0.000) 1.74 (1.51, 1.97) 100.00
-9.66
1
9.66
Note: Based on inverse variance weighting

Appendix - Annex III

Sample correction a la Heckman

Table 33

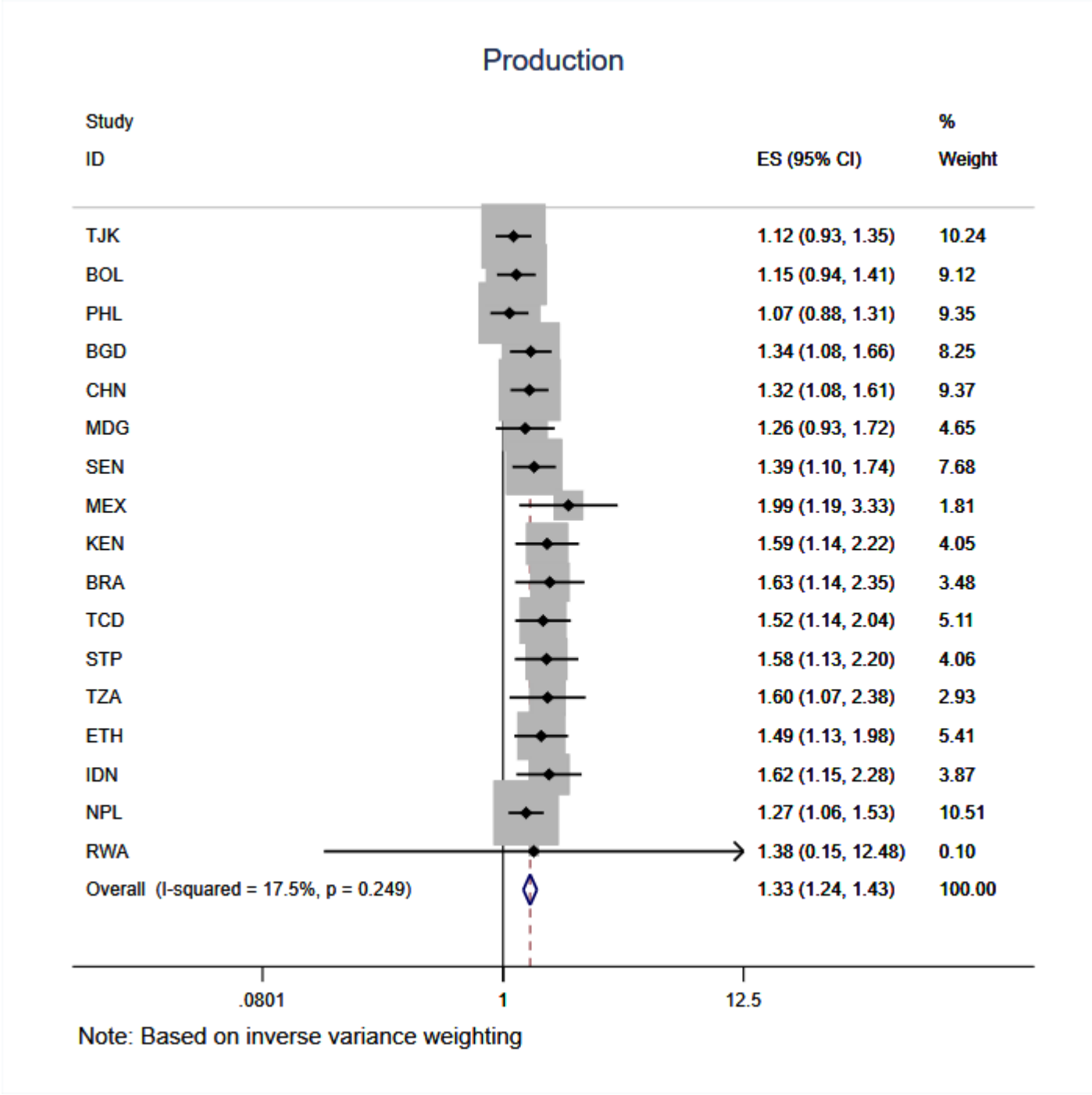


Table 34

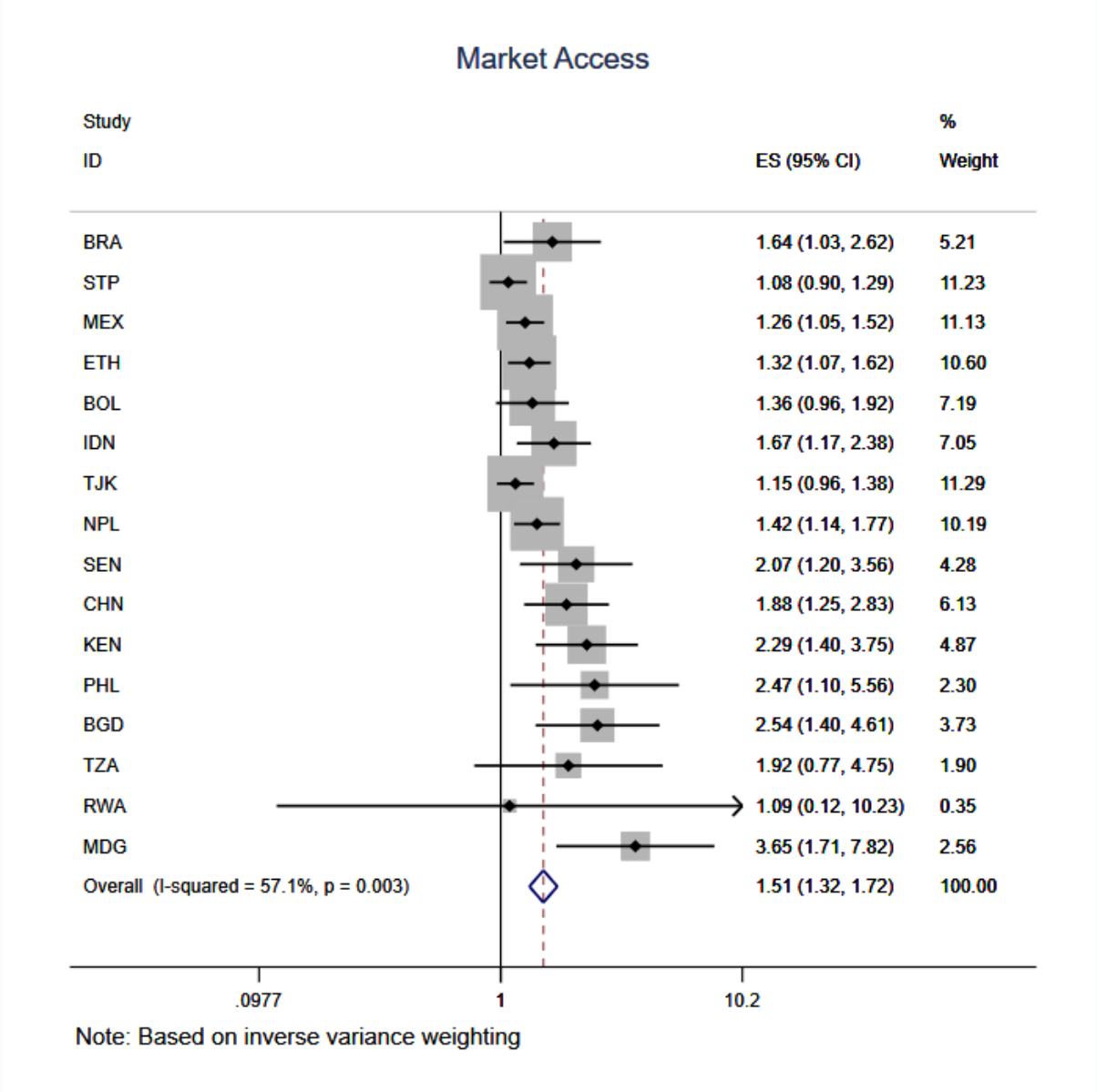

Market Access
Study
ID
ES (95% CI)
%
Weight
BRA 1.64 (1.03, 2.62) 5.21
STP 1.08 (0.90, 1.29) 11.23
MEX 1.26 (1.05, 1.52) 11.13
ETH 1.32 (1.07, 1.62) 10.60
BOL 1.36 (0.96, 1.92) 7.19
IDN 1.67 (1.17, 2.38) 7.05
TJK 1.15 (0.96, 1.38) 11.29
NPL 1.42 (1.14, 1.77) 10.19
SEN 2.07 (1.20, 3.56) 4.28
CHN 1.88 (1.25, 2.83) 6.13
KEN 2.29 (1.40, 3.75) 4.87
PHL 2.47 (1.10, 5.56) 2.30
BGD 2.54 (1.40, 4.61) 3.73
TZA 1.92 (0.77, 4.75) 1.90
RWA 1.09 (0.12, 10.23) 0.35
MDG 3.65 (1.71, 7.82) 2.56
Overall (I-squared = 57.1%, p = 0.003) 1.51 (1.32, 1.72) 100.00
.0977
1
10.2
Note: Based on inverse variance weighting

Table 35

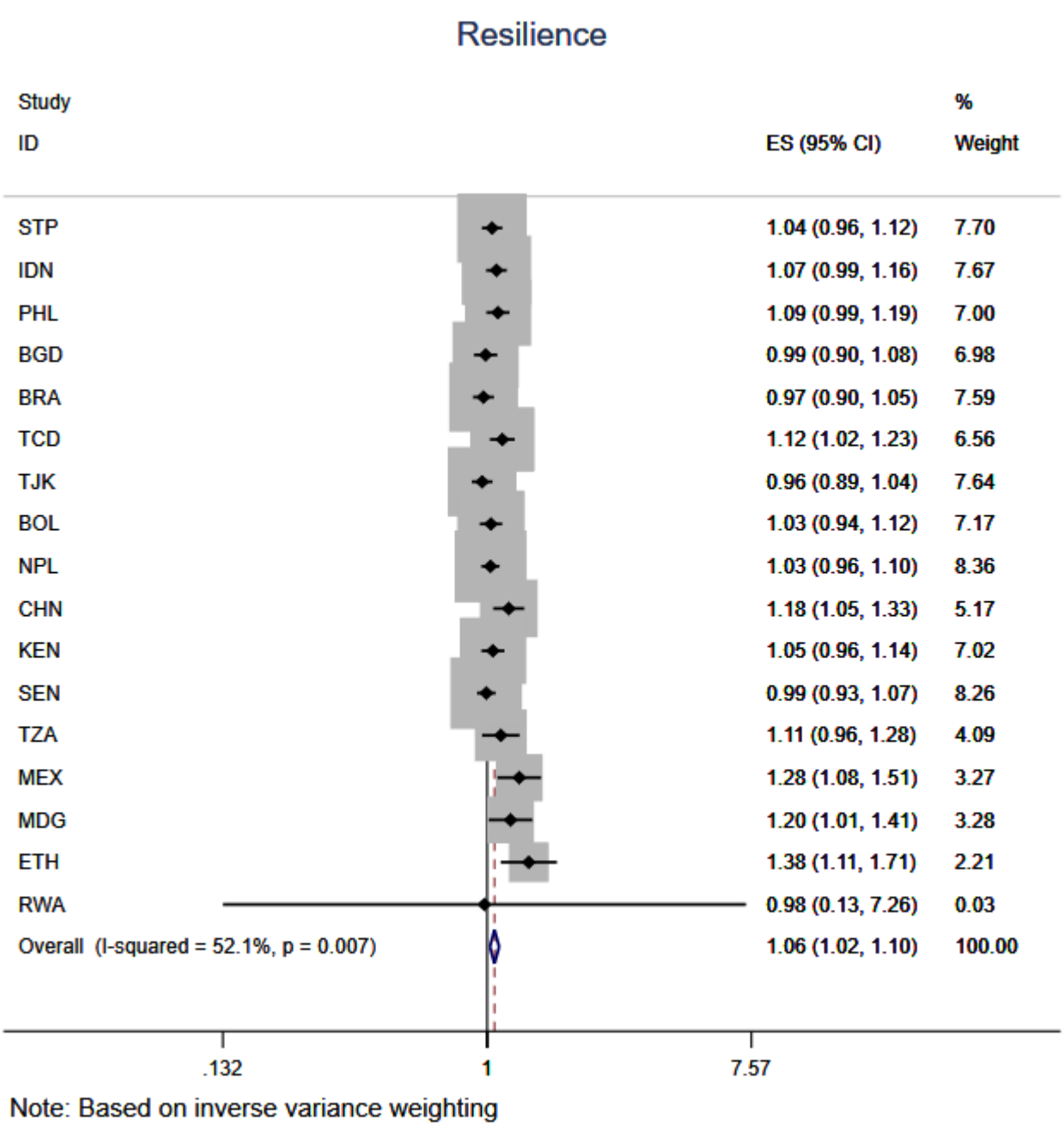

Resilience
Study ID
ES (95% CI)
% Weight
STP 1.04 (0.96, 1.12) 7.70
IDN 1.07 (0.99, 1.16) 7.67
PHL 1.09 (0.99, 1.19) 7.00
BGD 0.99 (0.90, 1.08) 6.98
BRA 0.97 (0.90, 1.05) 7.59
TCD 1.12 (1.02, 1.23) 6.56
TJK 0.96 (0.89, 1.04) 7.64
BOL 1.03 (0.94, 1.12) 7.17
NPL 1.03 (0.96, 1.10) 8.36
CHN 1.18 (1.05, 1.33) 5.17
KEN 1.05 (0.96, 1.14) 7.02
SEN 0.99 (0.93, 1.07) 8.26
TZA 1.11 (0.96, 1.28) 4.09
MEX 1.28 (1.08, 1.51) 3.27
MDG 1.20 (1.01, 1.41) 3.28
ETH 1.38 (1.11, 1.71) 2.21
RWA 0.98 (0.13, 7.26) 0.03
Overall (I-squared = 52.1%, p = 0.007) 1.06 (1.02, 1.10) 100.00
.132
1
7.57
Note: Based on inverse variance weighting

Table 36

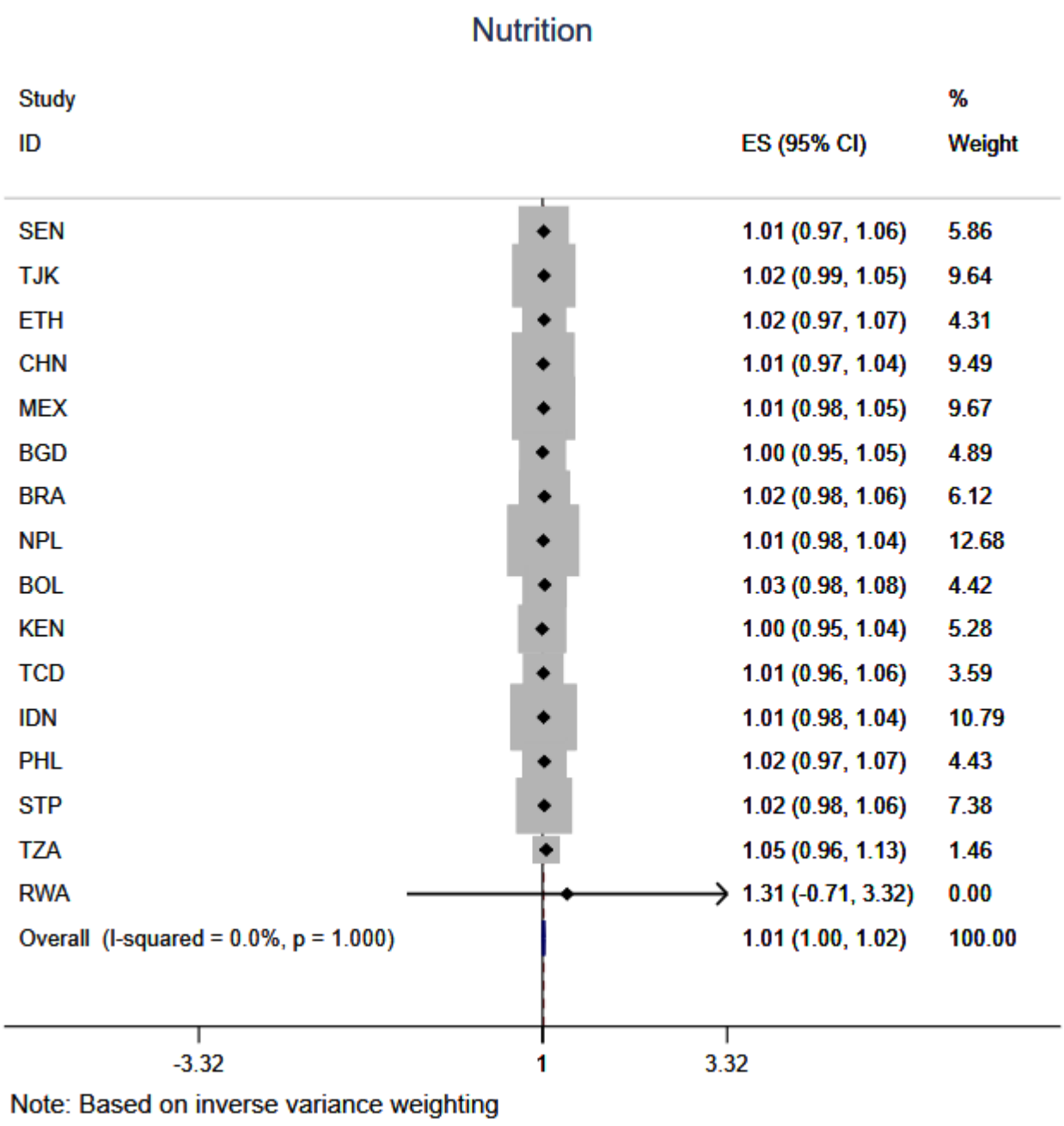
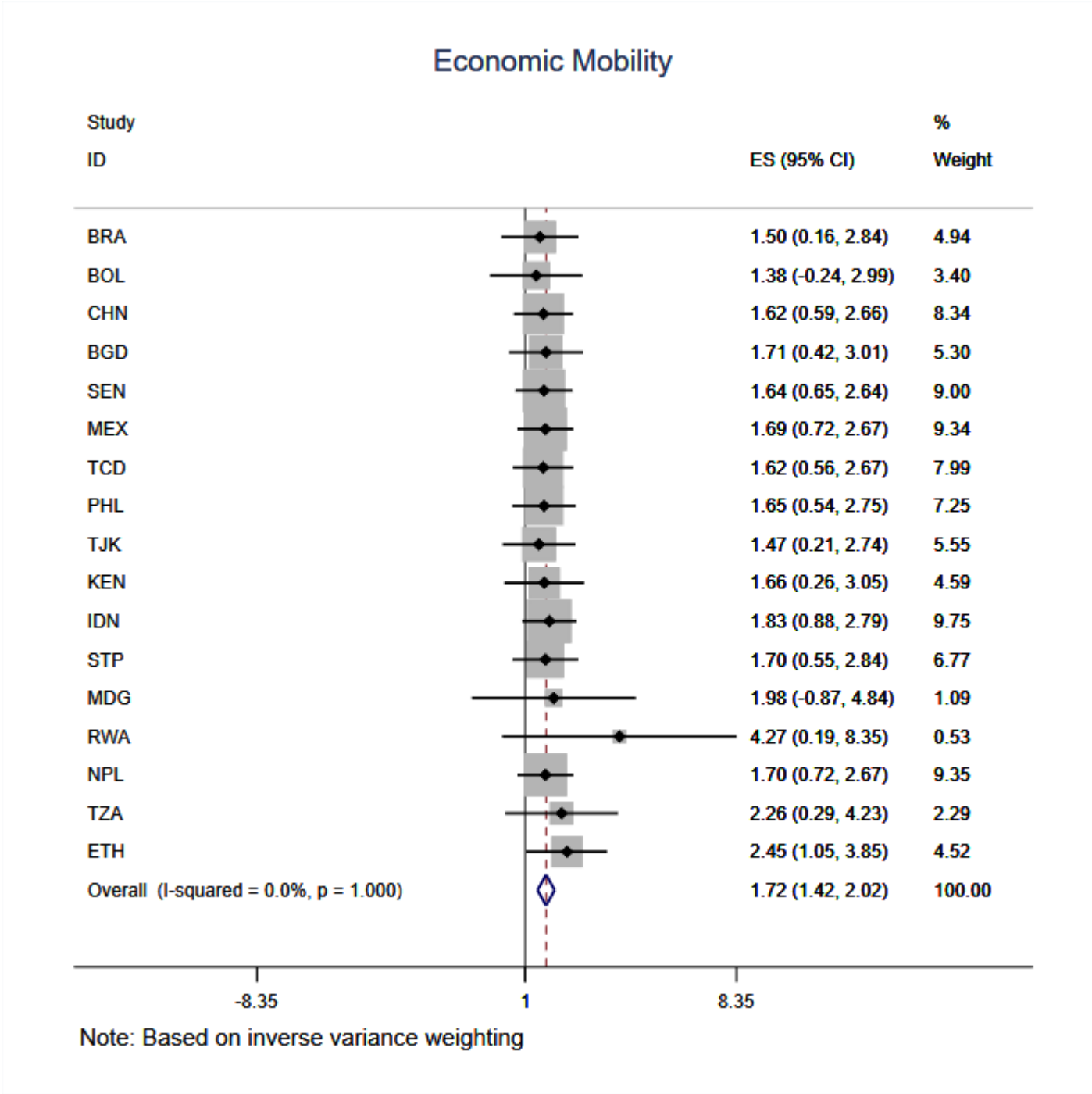

Nutrition
Study ID
ES (95% CI)
% Weight
SEN 1.01 (0.97, 1.06) 5.86
TJK 1.02 (0.99, 1.05) 9.64
ETH 1.02 (0.97, 1.07) 4.31
CHN 1.01 (0.97, 1.04) 9.49
MEX 1.01 (0.98, 1.05) 9.67
BGD 1.00 (0.95, 1.05) 4.89
BRA 1.02 (0.98, 1.06) 6.12
NPL 1.01 (0.98, 1.04) 12.68
BOL 1.03 (0.98, 1.08) 4.42
KEN 1.00 (0.95, 1.04) 5.28
TCD 1.01 (0.96, 1.06) 3.59
IDN 1.01 (0.98, 1.04) 10.79
PHL 1.02 (0.97, 1.07) 4.43
STP 1.02 (0.98, 1.06) 7.38
TZA 1.05 (0.96, 1.13) 1.46
RWA 1.31 (-0.71, 3.32) 0.00
Overall (I-squared = 0.0%, p = 1.000) 1.01 (1.00, 1.02) 100.00
-3.32
1
3.32
Note: Based on inverse variance weighting

Table 37


Economic Mobility
Study
ID
ES (95% CI)
%
Weight
BRA 1.50 (0.16, 2.84) 4.94
BOL 1.38 (-0.24, 2.99) 3.40
CHN 1.62 (0.59, 2.66) 8.34
BGD 1.71 (0.42, 3.01) 5.30
SEN 1.64 (0.65, 2.64) 9.00
MEX 1.69 (0.72, 2.67) 9.34
TCD 1.62 (0.56, 2.67) 7.99
PHL 1.65 (0.54, 2.75) 7.25
TJK 1.47 (0.21, 2.74) 5.55
KEN 1.66 (0.26, 3.05) 4.59
IDN 1.83 (0.88, 2.79) 9.75
STP 1.70 (0.55, 2.84) 6.77
MDG 1.98 (-0.87, 4.84) 1.09
RWA 4.27 (0.19, 8.35) 0.53
NPL 1.70 (0.72, 2.67) 9.35
TZA 2.26 (0.29, 4.23) 2.29
ETH 2.45 (1.05, 3.85) 4.52
Overall (I-squared = 0.0%, p = 1.000) 1.72 (1.42, 2.02) 100.00
-8.35
1
8.35
Note: Based on inverse variance weighting

Appendix - Annex IV

Sensitivity Analyses results with the trim-and-fill method

The trim and fill – adjusts for bias non-parametrically. Specifically, in order to investigate for the presence of small study effects and publication bias, visual representations such as funnel or contour enhanced funnel plots are employed.

A funnel plot shows effect sizes against measures of study precision e.g. standard error. The funnel plot is employed to explore visually publication bias or more precisely small study effect. The asymmetry is evidence and maybe the result of publication bias or may be because of other reasons (heterogeneity between studies).

The contour enhanced funnel plot, can help determine whether the asymmetry of the funnel plot is due to selection bias (e.g. publication bias). The contour lines correspond to certain levels of statistical significance. Publication bias is suspect when smaller studies are absent from the non-significant regions.

Tests for funnel-plot asymmetry are useful for detecting publication bias but are not able to estimate the impact of this bias on the final meta-analysis results. The nonparametric trim-and-fill method of Duval and Tweedie (2000a, 2000b) provides a way to assess the impact of missing studies because of publication bias on the meta-analysis. It evaluates the amount of potential bias present in meta-analysis and its impact on the final conclusion. This method is typically used as a sensitivity analysis to the presence of publication bias.

Results from the Trim-and-fill method are presented in Table 38 which summarizes the original results for the meta-analysis (the observed effect size – ES) along with the imputed one from the trim and fill results (observed plus imputed ES). The full set of tables are **Error! Reference source not found.** to Table 48.

The table shows that adjusted results remain largely positive and sometimes unaltered for some domains. Bias might affect only three coefficients: Economic mobility (1.74 vs. 1.38 equivalent to 74% and 38% respectively), Market access (1.76 vs. 1.38 equivalent to 76% and 38% respectively); and Resilience indicators (1.13 vs. 1.03 equivalent to 13% and 3%, respectively). However a known limitation of Trim-and-Fill is that it can correct for publication bias that does not exist, underestimating effect sizes (Terrin et al 2003). Recommendations from recently published literature, (Simonsohn et al 2014) argued against the use of such method[23].

[23] Simonsohn, Uri and Nelson, Leif D. and Simmons, Joseph P., P-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results (April 27, 2014). Available at SSRN: https://ssrn.com/abstract=2377290 or http://dx.doi.org/10.2139/ssrn.2377290

Table 38: Results from the Trim and Fill method

Production				
	N. projects	ES	Lower CI	Upper CI
Observed	17	1.44	1.26	1.65
Observed+Imputed	17	1.44	1.26	1.65
Market Access				
	N. projects	ES	Lower CI	Upper CI
Observed	16	1.76	1.45	2.14
Observed+Imputed	21	1.38	1.13	1.67
Resilience				
	N. projects	ES	Lower CI	Upper CI
Observed	17	1.13	1.02	1.25
Observed+Imputed	20	1.04	0.91	1.18
Nutrition				
	N. projects	ES	Lower CI	Upper CI
Observed	16	1.01	1	1.03
Observed+Imputed	17	1.01	1	1.03
Economic Mobility				
	N. projects	ES	Lower CI	Upper CI
Observed	17	1.74	1.51	1.97
Observed+Imputed	18	1.38	1.1	1.67

Table 39

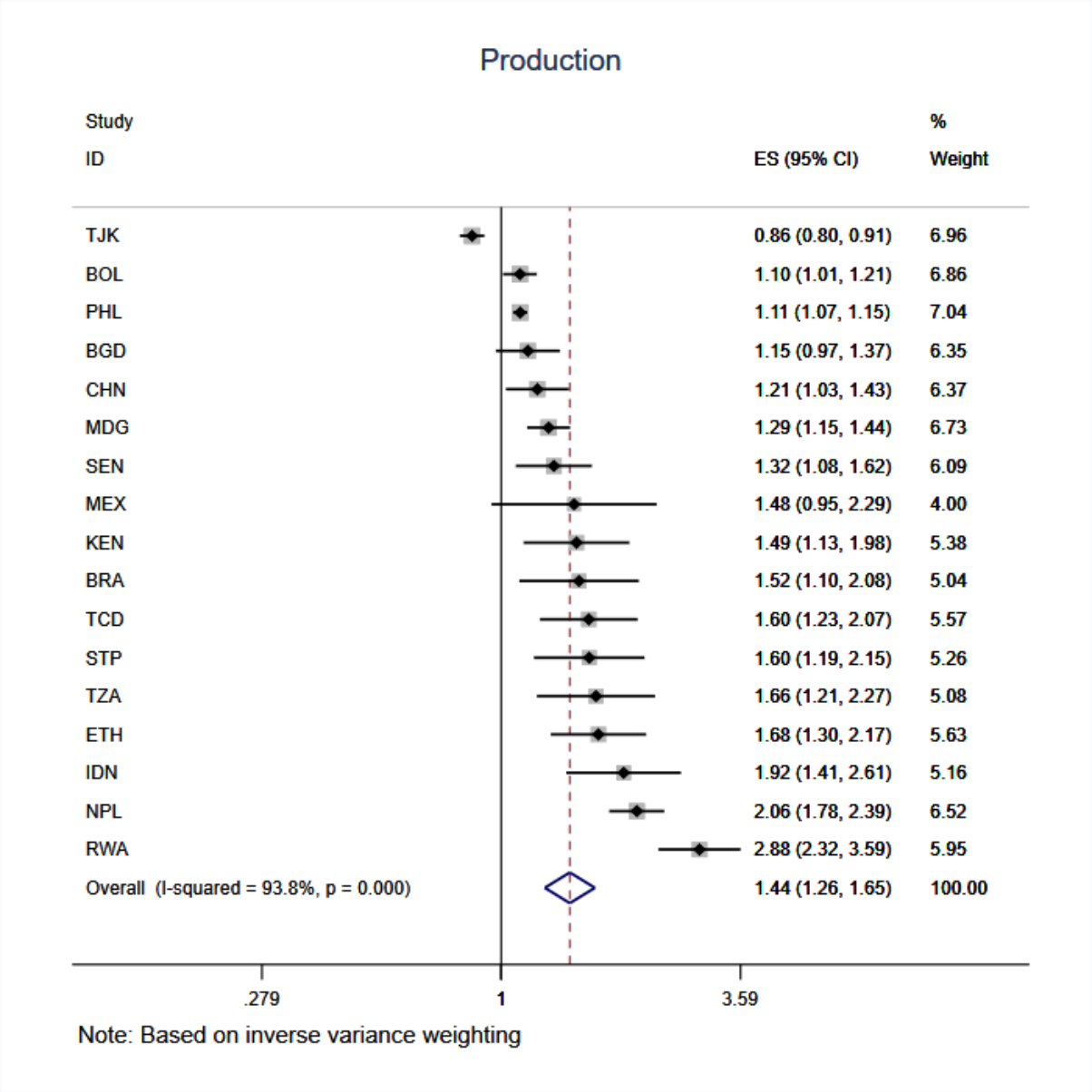

Production
Study
ID
ES (95% CI)
%
Weight
TJK 0.86 (0.80, 0.91) 6.96
BOL 1.10 (1.01, 1.21) 6.86
PHL 1.11 (1.07, 1.15) 7.04
BGD 1.15 (0.97, 1.37) 6.35
CHN 1.21 (1.03, 1.43) 6.37
MDG 1.29 (1.15, 1.44) 6.73
SEN 1.32 (1.08, 1.62) 6.09
MEX 1.48 (0.95, 2.29) 4.00
KEN 1.49 (1.13, 1.98) 5.38
BRA 1.52 (1.10, 2.08) 5.04
TCD 1.60 (1.23, 2.07) 5.57
STP 1.60 (1.19, 2.15) 5.26
TZA 1.66 (1.21, 2.27) 5.08
ETH 1.68 (1.30, 2.17) 5.63
IDN 1.92 (1.41, 2.61) 5.16
NPL 2.06 (1.78, 2.39) 6.52
RWA 2.88 (2.32, 3.59) 5.95
Overall (I-squared = 93.8%, p = 0.000) 1.44 (1.26, 1.65) 100.00
.279
1
3.59
Note: Based on inverse variance weighting

Table 40

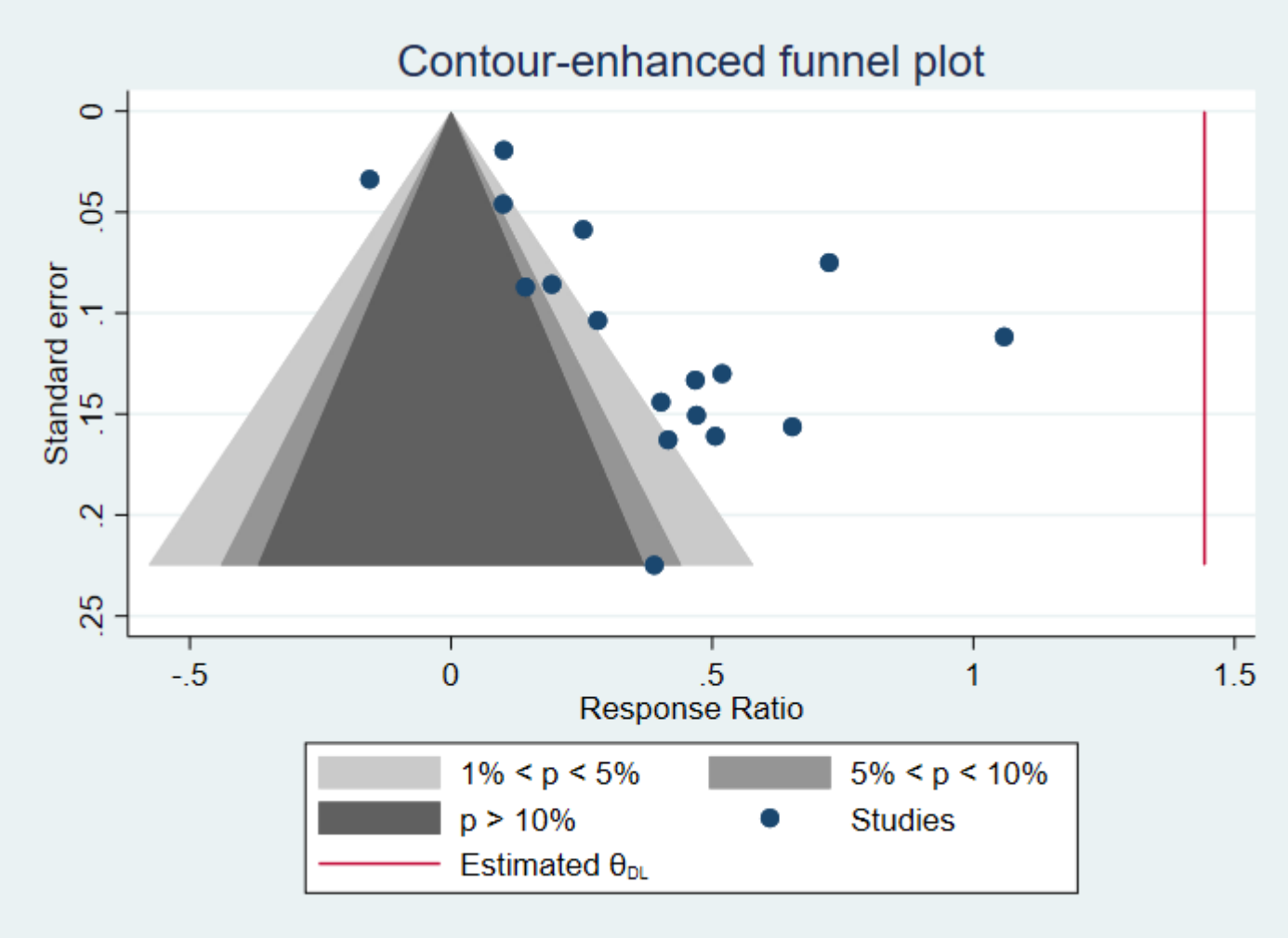

Contour-enhanced funnel plot
Standard error
0
.05
.1
.15
.2
.25
-.5
0
.5
1
1.5
Response Ratio
1% < p < 5%
5% < p < 10%
p > 10%
Studies
Estimated θDL

Table 41

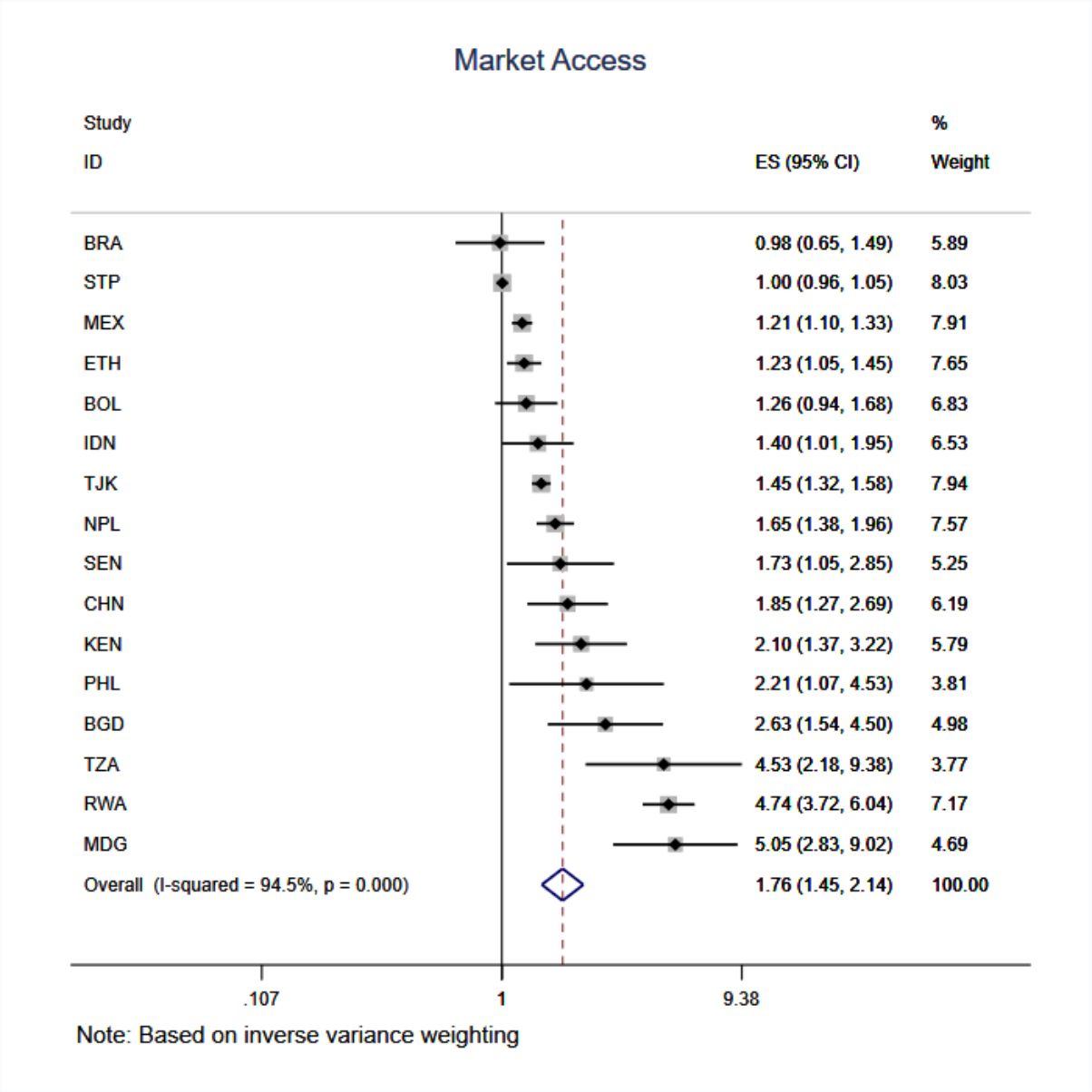

Market Access
Study
ID
ES (95% CI)
%
Weight
BRA 0.98 (0.65, 1.49) 5.89
STP 1.00 (0.96, 1.05) 8.03
MEX 1.21 (1.10, 1.33) 7.91
ETH 1.23 (1.05, 1.45) 7.65
BOL 1.26 (0.94, 1.68) 6.83
IDN 1.40 (1.01, 1.95) 6.53
TJK 1.45 (1.32, 1.58) 7.94
NPL 1.65 (1.38, 1.96) 7.57
SEN 1.73 (1.05, 2.85) 5.25
CHN 1.85 (1.27, 2.69) 6.19
KEN 2.10 (1.37, 3.22) 5.79
PHL 2.21 (1.07, 4.53) 3.81
BGD 2.63 (1.54, 4.50) 4.98
TZA 4.53 (2.18, 9.38) 3.77
RWA 4.74 (3.72, 6.04) 7.17
MDG 5.05 (2.83, 9.02) 4.69
Overall (I-squared = 94.5%, p = 0.000) 1.76 (1.45, 2.14) 100.00
.107
1
9.38
Note: Based on inverse variance weighting

Table 42

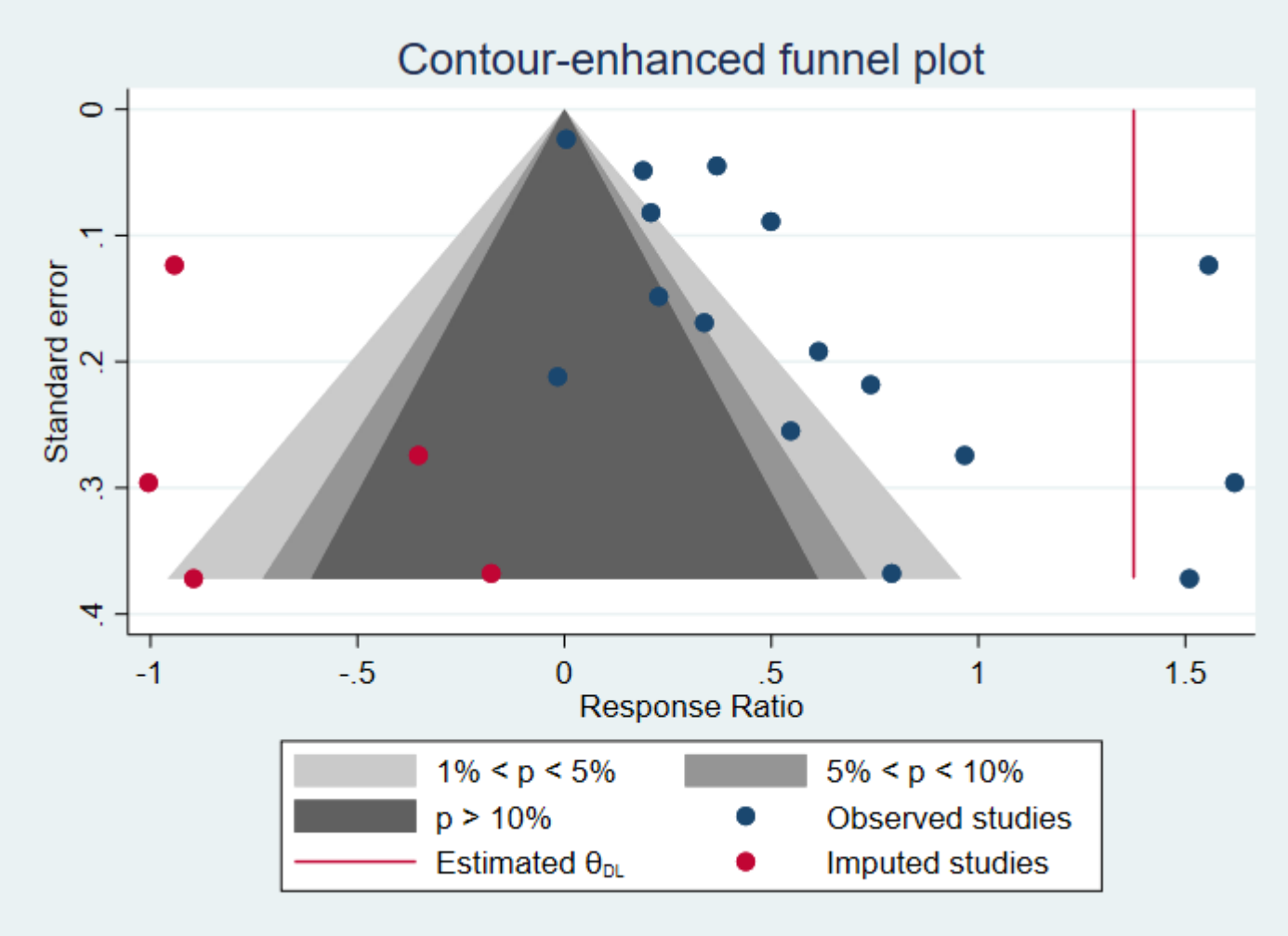

Contour-enhanced funnel plot
Standard error
0
.1
.2
.3
.4
-1
-.5
0
.5
1
1.5
Response Ratio
1% < p < 5%
5% < p < 10%
p > 10%
Observed studies
Estimated θDL
Imputed studies

Table 43

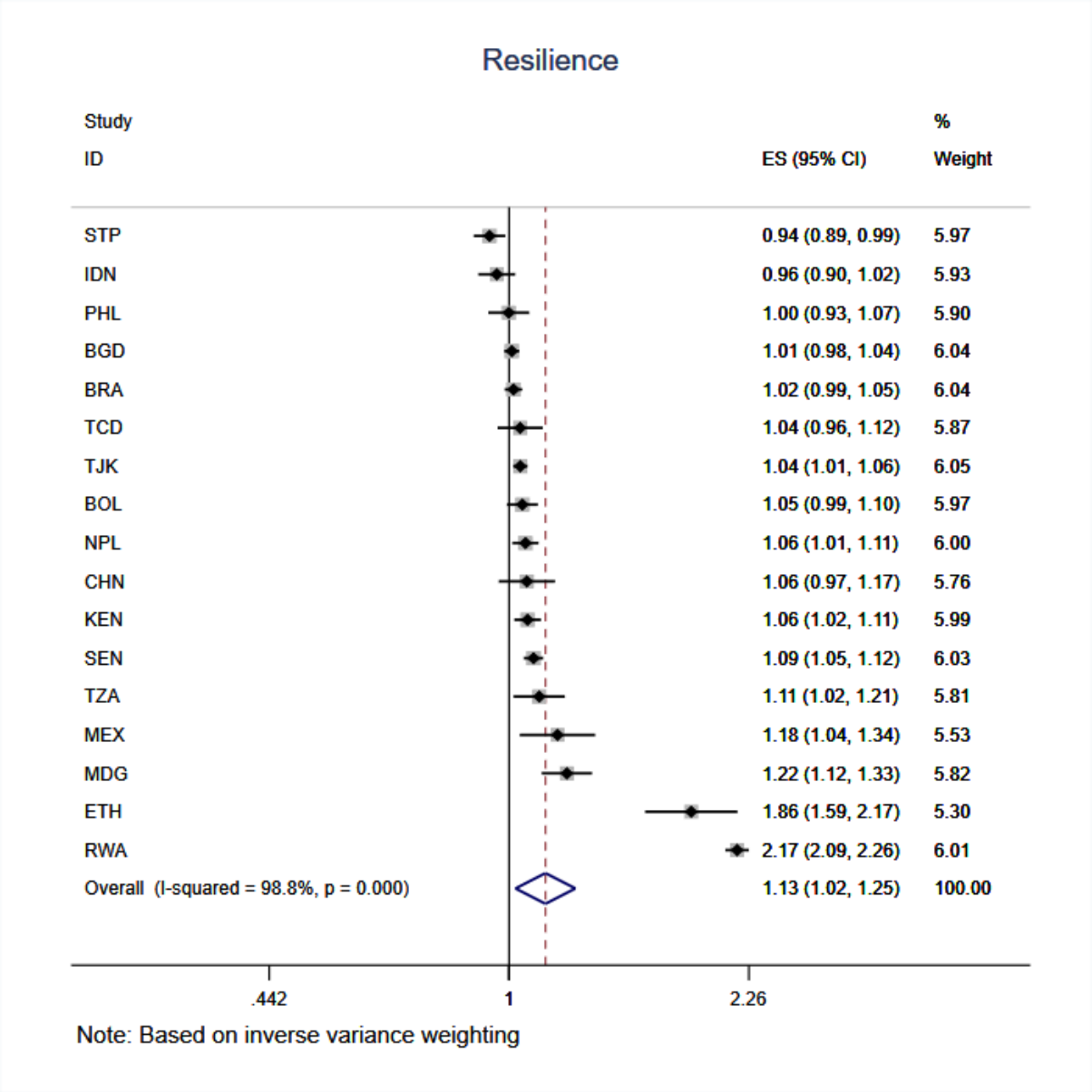

Resilience
Study
ID
ES (95% CI)
%
Weight
STP 0.94 (0.89, 0.99) 5.97
IDN 0.96 (0.90, 1.02) 5.93
PHL 1.00 (0.93, 1.07) 5.90
BGD 1.01 (0.98, 1.04) 6.04
BRA 1.02 (0.99, 1.05) 6.04
TCD 1.04 (0.96, 1.12) 5.87
TJK 1.04 (1.01, 1.06) 6.05
BOL 1.05 (0.99, 1.10) 5.97
NPL 1.06 (1.01, 1.11) 6.00
CHN 1.06 (0.97, 1.17) 5.76
KEN 1.06 (1.02, 1.11) 5.99
SEN 1.09 (1.05, 1.12) 6.03
TZA 1.11 (1.02, 1.21) 5.81
MEX 1.18 (1.04, 1.34) 5.53
MDG 1.22 (1.12, 1.33) 5.82
ETH 1.86 (1.59, 2.17) 5.30
RWA 2.17 (2.09, 2.26) 6.01
Overall (I-squared = 98.8%, p = 0.000) 1.13 (1.02, 1.25) 100.00
.442
1
2.26
Note: Based on inverse variance weighting

Table 44

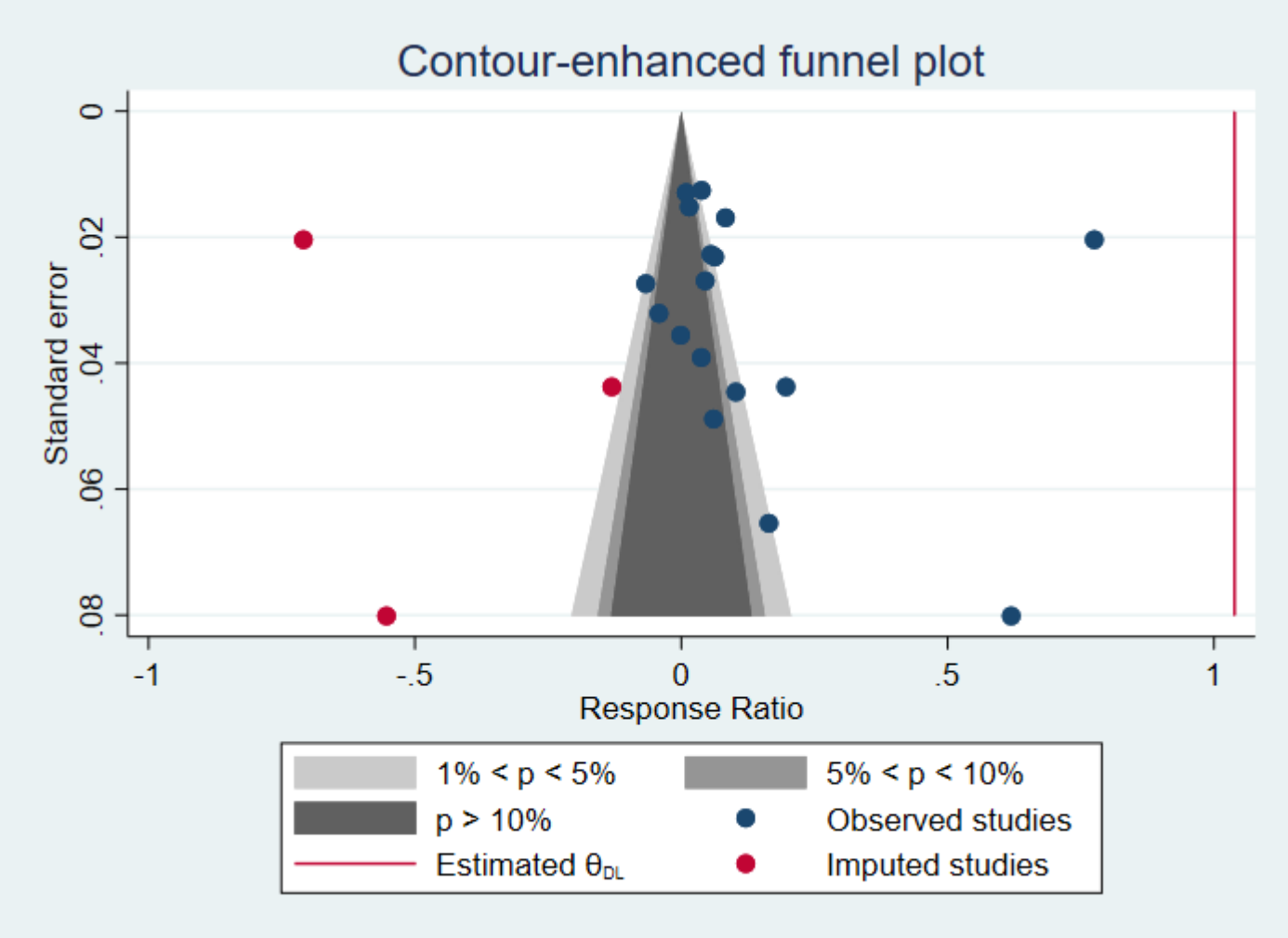

Contour-enhanced funnel plot
Standard error
0
.02
.04
.06
.08
-1
-.5
0
.5
1
Response Ratio
1% < p < 5%
5% < p < 10%
p > 10%
Observed studies
Estimated θDL
Imputed studies

Table 45

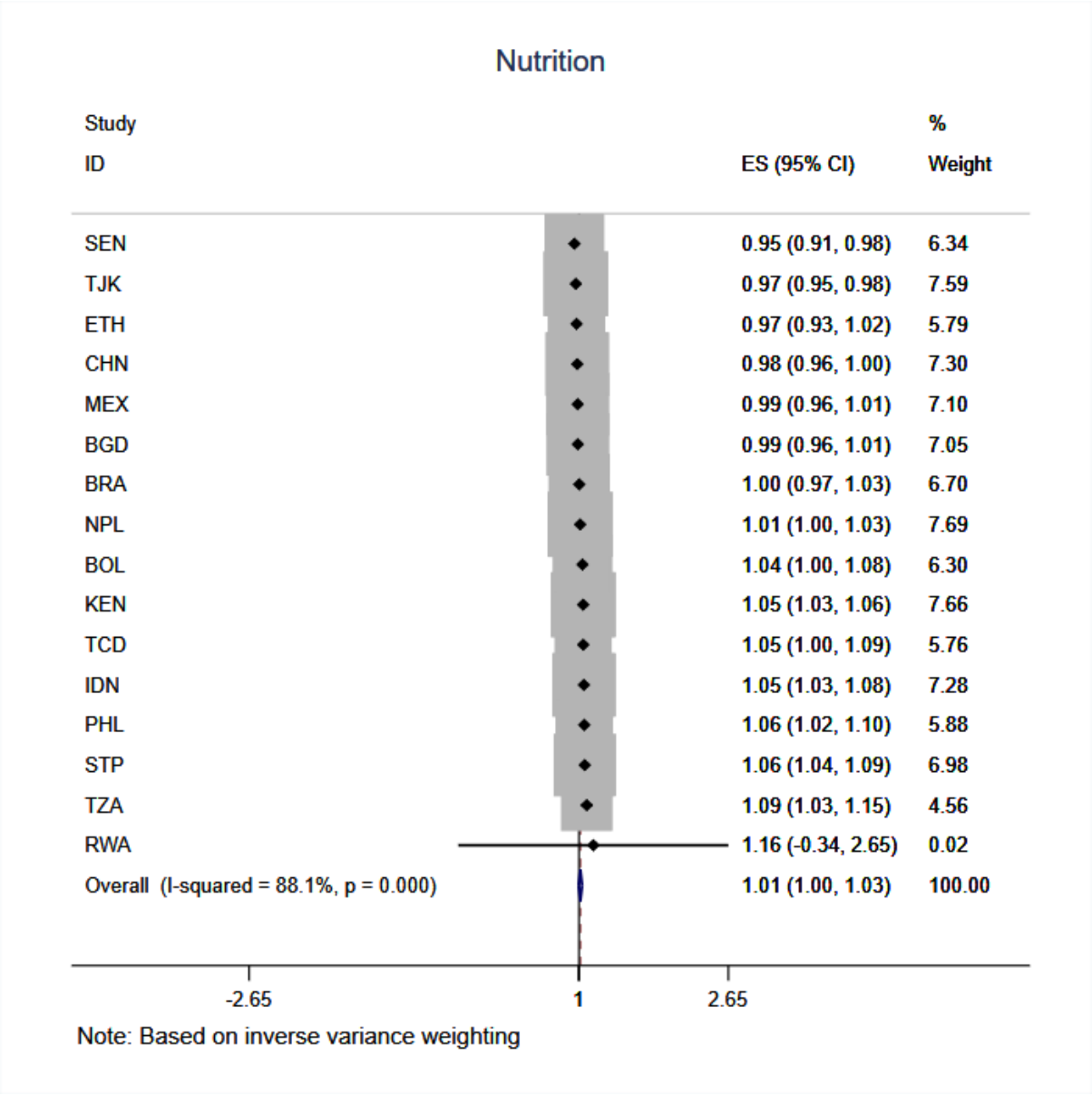

Nutrition
Study ID
ES (95% CI)
% Weight
SEN 0.95 (0.91, 0.98) 6.34
TJK 0.97 (0.95, 0.98) 7.59
ETH 0.97 (0.93, 1.02) 5.79
CHN 0.98 (0.96, 1.00) 7.30
MEX 0.99 (0.96, 1.01) 7.10
BGD 0.99 (0.96, 1.01) 7.05
BRA 1.00 (0.97, 1.03) 6.70
NPL 1.01 (1.00, 1.03) 7.69
BOL 1.04 (1.00, 1.08) 6.30
KEN 1.05 (1.03, 1.06) 7.66
TCD 1.05 (1.00, 1.09) 5.76
IDN 1.05 (1.03, 1.08) 7.28
PHL 1.06 (1.02, 1.10) 5.88
STP 1.06 (1.04, 1.09) 6.98
TZA 1.09 (1.03, 1.15) 4.56
RWA 1.16 (-0.34, 2.65) 0.02
Overall (I-squared = 88.1%, p = 0.000) 1.01 (1.00, 1.03) 100.00
-2.65
1
2.65
Note: Based on inverse variance weighting

Table 46

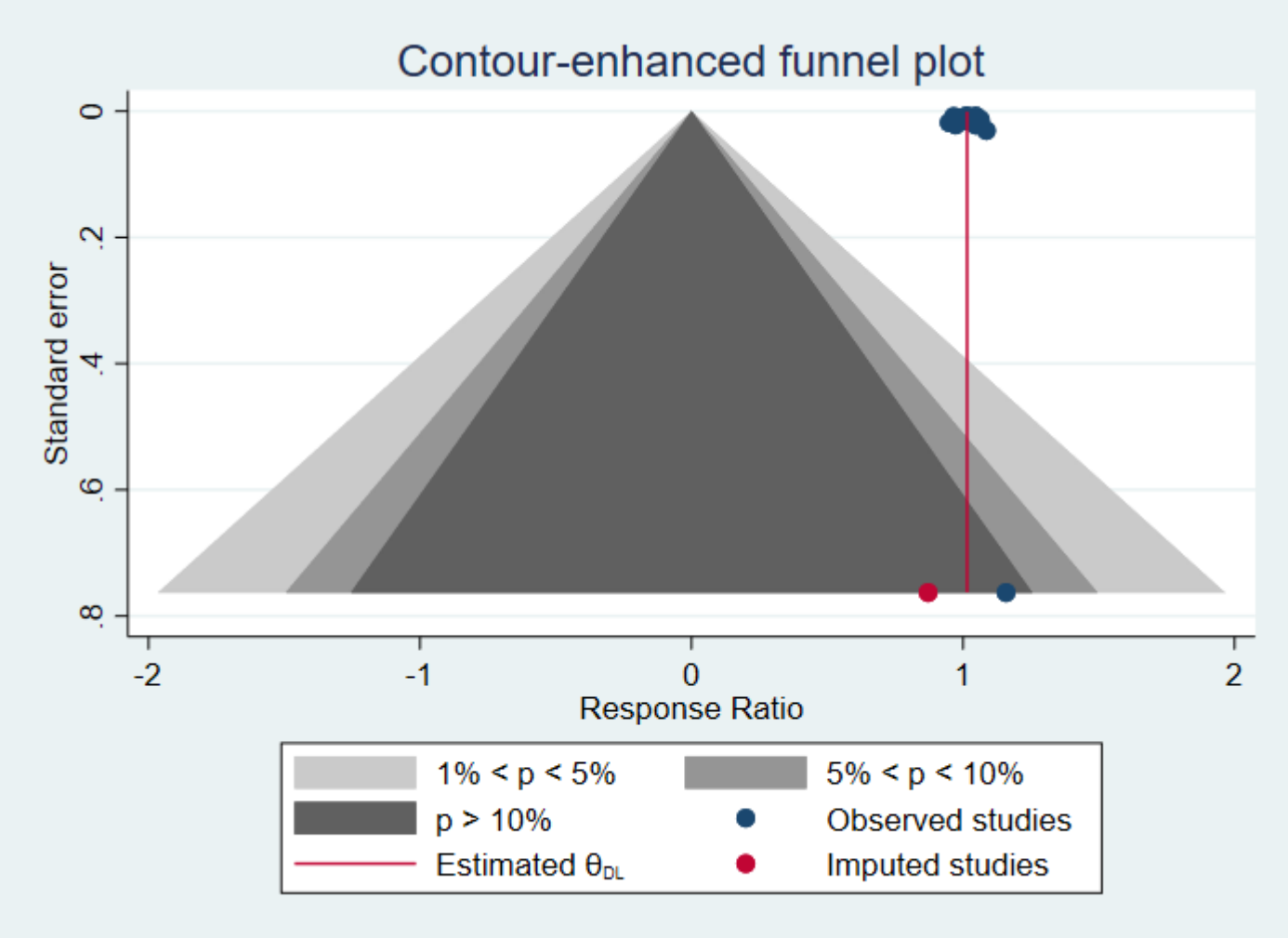

Contour-enhanced funnel plot
Standard error
0
.2
.4
.6
.8
-2
-1
0
1
2
Response Ratio
1% < p < 5%
5% < p < 10%
p > 10%
Observed studies
Estimated θDL
Imputed studies

Table 47

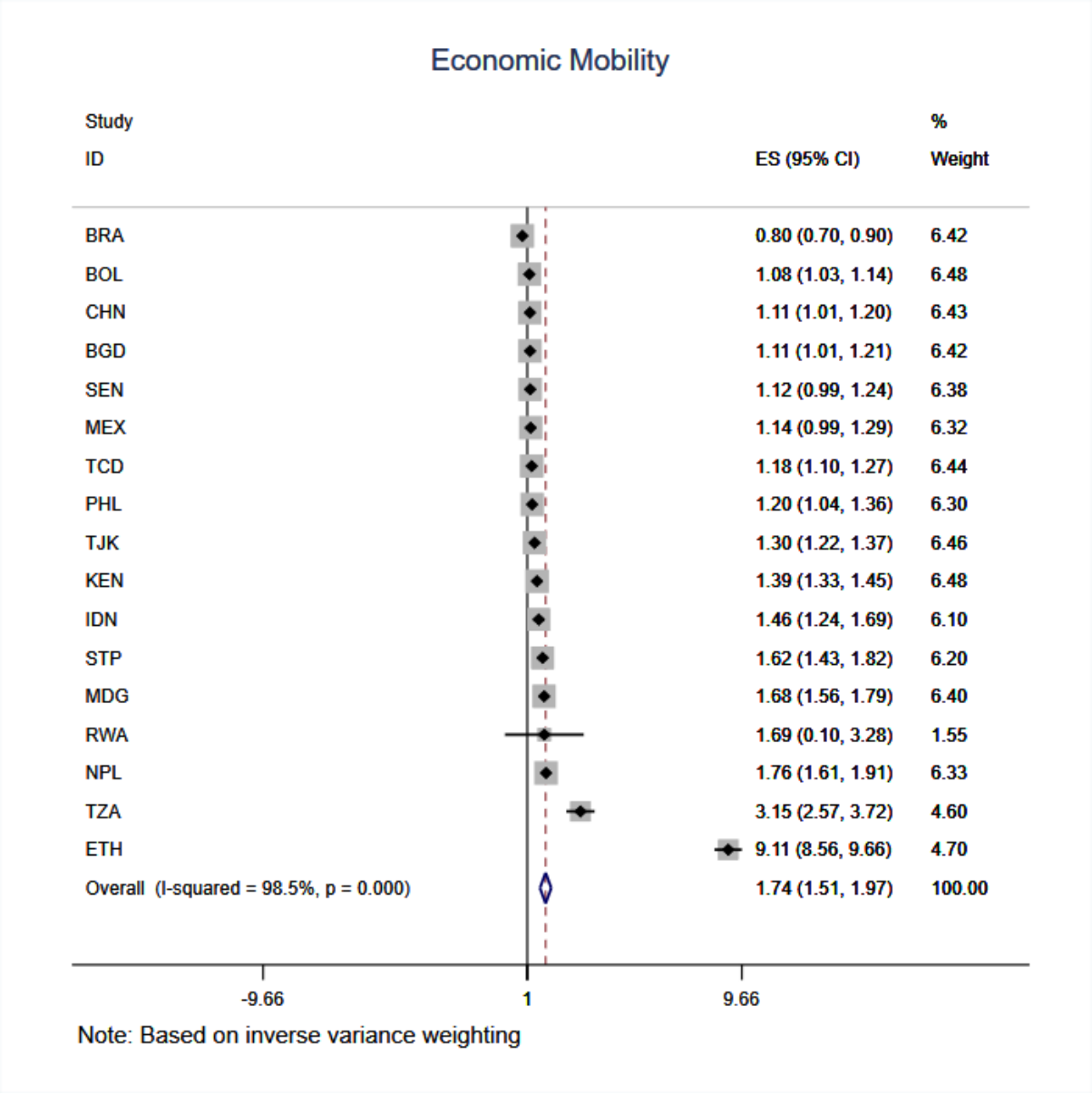
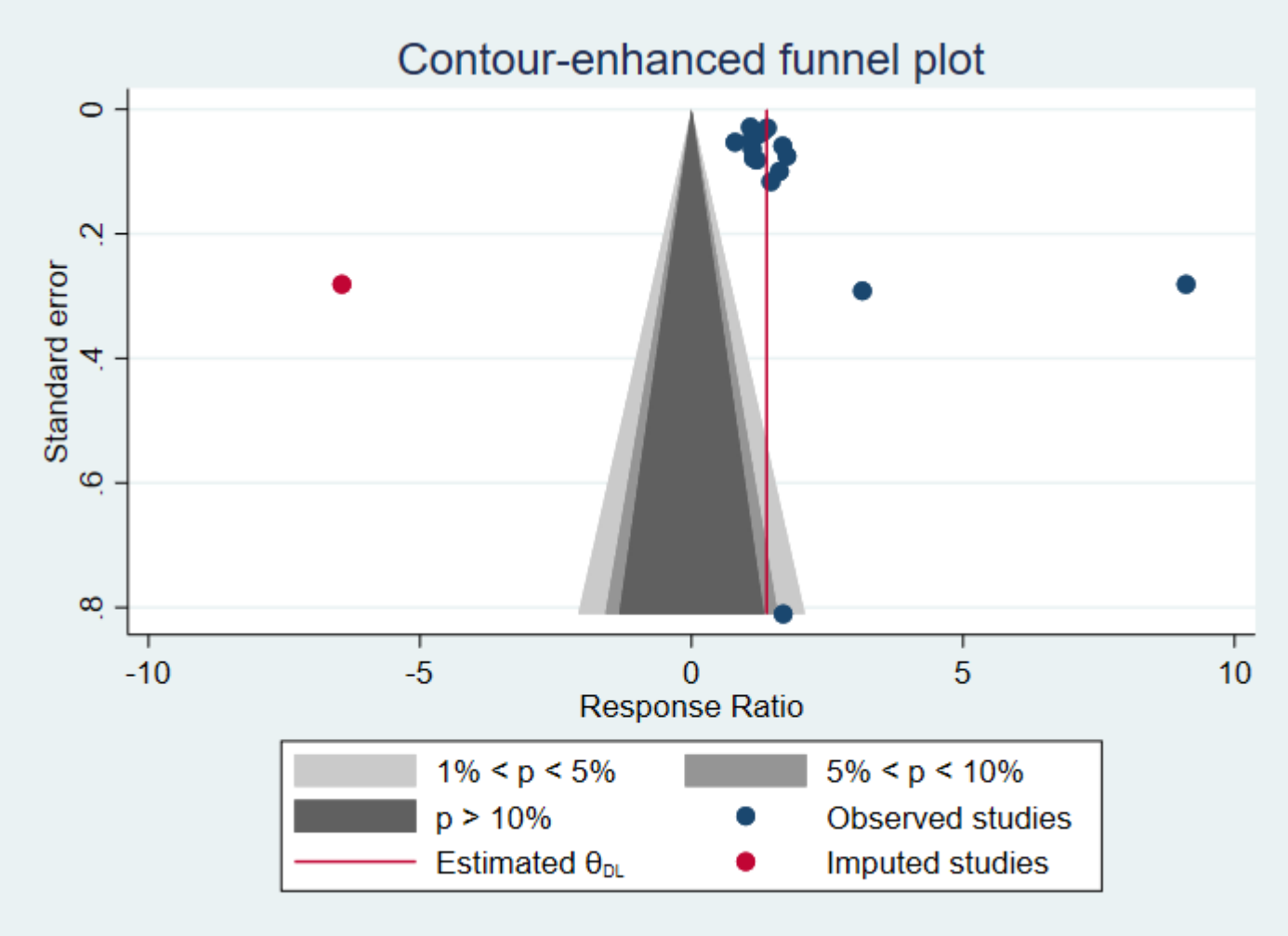

Economic Mobility
Study
ID
ES (95% CI)
%
Weight
BRA 0.80 (0.70, 0.90) 6.42
BOL 1.08 (1.03, 1.14) 6.48
CHN 1.11 (1.01, 1.20) 6.43
BGD 1.11 (1.01, 1.21) 6.42
SEN 1.12 (0.99, 1.24) 6.38
MEX 1.14 (0.99, 1.29) 6.32
TCD 1.18 (1.10, 1.27) 6.44
PHL 1.20 (1.04, 1.36) 6.30
TJK 1.30 (1.22, 1.37) 6.46
KEN 1.39 (1.33, 1.45) 6.48
IDN 1.46 (1.24, 1.69) 6.10
STP 1.62 (1.43, 1.82) 6.20
MDG 1.68 (1.56, 1.79) 6.40
RWA 1.69 (0.10, 3.28) 1.55
NPL 1.76 (1.61, 1.91) 6.33
TZA 3.15 (2.57, 3.72) 4.60
ETH 9.11 (8.56, 9.66) 4.70
Overall (I-squared = 98.5%, p = 0.000) 1.74 (1.51, 1.97) 100.00
-9.66
1
9.66
Note: Based on inverse variance weighting

Table 48



References

Andrews, Isaiah and Emily Oster. 2018. "Weighting for External Validity." National Bureau of Economic Research Working Paper 23826. http://www.nber.org/papers/w23826.

Banerjee, Amitav and Chaudhury, Suprakash. 2010. "Statistics without tears: Populations and samples." Industrial Psychiatry Journal. 19(1): 60–65.

Bown, M.J. and A.J. Sutton. 2010. "Quality Control in Systematic Reviews and Meta-analyses." European Journal of Cardiovascular and Endovascular Surgery, 40: 669-677.

Breslow NE, Day NE. 1980. Statistical Methods in Cancer Research: Vol. I - The Analysis of Case-Control Studies. Lyon: International Agency for Research on Cancer.

Copas, John and Jian Qing Shi. 2000. "Meta-analysis, funnel plots, and sensitivity analysis." Biostatistics, 1(3): 247-262.

Clarke, Kevin A, Randall W. Stone. 2019. "The Unobserved IMF." The Political Economy of International Organizations. https://www.peio.me/wp-ent/uploads/2019/01/PEIO12_paper_109.pdf

Cortes, Corinna; Mohri, Mehryar; Riley, Michael; Rostamizadeh, Afshin (2008). Sample Selection Bias Correction Theory (PDF). Algorithmic Learning Theory. Lecture Notes in Computer Science. 5254. pp. 38–53.

Cortes, Corinna; Mohri, Mehryar (2014). "Domain adaptation and sample bias correction theory and algorithm for regression" (PDF). Theoretical Computer Science. 519: 103–126.

Delgado-Rodríguez M, Llorca J. "Bias" Journal of Epidemiology & Community Health 2004;58:635-641.

Greenland, S. (2005) Multiple-bias modelling for analysis of observational data.J. R. Statist. Soc.A,168, 267–291

Heckman, James J., Sergio Urzua and Edward Vytlacil. 2006. "Understanding Instrumental Variables In Models With Essential Heterogeneity," Review of Economics and Statistics, 88(3, Aug): 389-432

Henmi M, Copas JB. 2010. Confidence intervals for random effects meta-analysis and robustness to publication bias. Statistics in Medicine 29: 2969-2983. doi: 10.1002/sim.4029

Lin L (2018) Bias caused by sampling error in meta-analysis with small sample sizes. PLoS ONE 13(9): e0204056. https://doi.org/10.1371/journal.pone.0204056

Mathur M.B & VanderWeele Tyler J. (2019): Sensitivity Analysis for Unmeasured Confounding in Meta-Analyses, Journal of the American Statistical Association, DOI: 10.1080/01621459.2018.1529598

Mavridis D, Salanti G. How to assess publication bias: funnel plot, trim-and-fill method and selection models. Evidence-Based Mental Health 2014;17:30.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." Journal of Business & Economic Statistics, 37(2):187-204.

Shi, Linyu and Lifeng Lin. 2019. "The trim-and-fill method for publication bias: practical guidelines and recommendations based on a large database of meta-analyses." Medicine, 98(23): 1-11.

Stuart, Elizabeth A., Stephen R. Cole, Catherine P. Bradshaw, and Philip J. Leaf. 2011. "The use of propensity scores to assess the generalizability of results from randomized trials." Journal of the Royal Statistical Society: Series A (Statistics in Society) 174(2): 369-386.