



Tax Enforcement, Inequality and Redistribution in sub-Saharan Africa

Four Essays in Empirical Economics

Léo Czajka

Dissertation pour l'obtention du grade de *Doctorat en sciences économiques et de gestion*

5 juillet 2024

Promoteur:

Prof. William Pariente - Université catholique de Louvain

Jury:

Prof. William Pariente (UCLouvain), promoteur, secrétaire du jury

Prof. Gonzague Vannoorenberghe (UCLouvain), président du jury

Prof. Amma Panin (UCLouvain), évaluatrice interne

Dr. Oyebola Okunogbe (World Bank), évaluatrice externe

Prof. Mazhar Waseem (University of Manchester), évaluateur externe

Remerciements - Acknowledgements

Mes premiers remerciements vont à mon promoteur, sans qui tout cela n'aurait pas été possible. Merci William d'avoir cru à toute cette affaire. J'étais loin de savoir dans quoi je me lançais alors, peut-être n'y aurais-je pas cru moi-même. Merci pour les très nombreux conseils, toujours donnés avec la plus grande bienveillance. Pour les cafés impromptus, sous l'œil curieux du roitelet huppé. A l'ombre d'une thèse, reposent aussi les vestiges des projets abandonnés ou reportés. Merci de m'avoir accompagné dans leur élaboration, puis aidé à prendre les décisions difficiles qui s'imposaient alors. Merci enfin d'avoir toujours prêté une oreille attentive aux nouvelles plus personnelles, certaines heureuses, d'autres moins, tu as toujours eu les mots justes.

Thank you, Anne Brockmeyer, for welcoming me at the IFS. Thank you so much for giving me the opportunity to present my work and share ideas with you and the many researchers at the Institute. I only wished I would have stayed longer. I am extremely grateful for your help on the job market, as for all the precious career advises. I am really looking forward to opportunities for us to collaborate in the future.

Thank you, Amma Panin, not only for agreeing to take part in my thesis jury, but even long before that, for showing curiosity about my work, despite a presentation that was chaotic, to say the least. I have always admired the ever-renewed eye you have casted over my work, whatever its stage of progress. Merci à toi Gonzague Vannoorenberghe de participer aussi à ce jury. Merci pour ton soutien tout au long de ces années. Je crois que tu n'as pas raté une seule de mes présentations, et je peux confesser maintenant qu'il m'est arrivé d'en sortir en pensant que tu comprenais mes projets mieux que moi. Thank you Oyebola Okunogbe and Mazhar Waseem for taking the time to read and review my work. Your comments have already contributed to significantly improve it, and will further in the future, as I manage to address more of them!

Cette thèse est aussi en partie le fruit d'un héritage parisien qui m'a été transmis alors que j'étais encore assistant de recherche au *Word Inequality Lab*. Merci à vous, Denis Cogneau et Thomas Piketty, de m'avoir inspiré, fait confiance, et récemment défendu dans la jungle du *job market*. Je vous dois beaucoup.

Les idées de recherches ne mûrissent que si elles sont confrontées, et aucun de ces projets n'aurait pris forme sans ses co-auteurs et co-autrices. Merci, pour l'énergie et la patience, les critiques et les fulgurances. Bass, pour tous les projets, le séjour à Louga, la vision. Amory, ceux qui ont travaillé avec toi le savent déjà, ceux qui n'ont pas eu cette chance risquent de l'apprendre à leurs dépens, véritable Lucky Luke du clavier, tu es le chercheur le plus prolifique

depuis Aristote. Aroop, I did not know how good a flat white could taste before we took one together in London, I also thought I knew a bit about South Africa's history before I met you. Maybe a wealth tax will come. Mattea, that was *so much* work, let's make the most of it – and cheers to Leonora and her model dad! Je ne désespère pas, Florence, de voir émerger des résultats de cette deuxième expérimentation d'ici la fin de l'année. J'ai développé mon sens de la mesure sous ta supervision, Denis, j'aurais aimé piquer ta plume. A quand ce projet sur les fonctionnaires, Kenneth ? Anne-Sophie, Akim, Mahdi: si seulement la micro-simulation pouvait intéresser la recherche autant que les gouvernements. Fabrizio, Awa, looking forward to Oxford! Pierre et Justine, ne rien lâcher, on en rediscute en septembre. En parlant de septembre, je ne saurais cacher ma hâte de nouer de nouvelles amitiés de recherche au sein d'*EU-Tax Observatory* !

Et si vos co-auteurs ne sont pas disponibles, il y aura toujours quelqu'un à l'IRES. Merci en particulier à Jean-François Maystadt, Luca Pensieroso, Joseph Gomes, Bruno Van der Linden, Fabio Mariani, David de la Croix, Vincent Vandenberghe et Vincent Bodart. Aux « post-doc » devenue amies puis professeures (ou équivalent) : Esther, ou l'audace, Stefanija, beaucoup trop modeste, Elsa, aussi efficace qu'inventive, et Dorothee, I miss your sense of humour. Merci enfin à tous mes compagnons d'infortune, ou de train SNCB (c'est la même chose), celles et ceux des plus anciennes générations d'abord : Elisa, partie si vite, Boris, la classe, Henrik, le juste, François, le gardien du temple, Riccardo, de loin le plus vif, Marion, déjà ?, Francesco, si on a des questions, et Annalisa, merci encore pour la reco. Un merci tout particulier à Daniele, vous ne trouverez pas co-bureau plus drôle et fin - *the question is what is the question*. Et à Yannik enfin, pour les émotions partagées : du doute au kiff, et retour. Puis celles et ceux qui sont arrivés ensuite, au fil des années : Esmeralda, Laura, Nathan, Margaux, Chiara, Luigi, Rossana, Filippo, Diego, Natalia, Sylvia, Charles, Mathilde, Vincent, Giulia, Florian et Tiziano. Comme un goût de trop peu. Hommage à ce dernier d'ailleurs, pour l'organisation des visites de Gand en kayak et autres réjouissances. Visiting the IFS, the conversation freely continued with: Vedanth, that rainy bike ride in secret London on our way to Icelandic choir music, Ross, until you crossed the line with stuffed croissants, Harriet, it's all about the right roast, Yani, thank you so much for the flat!, Andrea, for the wisdom, Maxime, for the long strolls back home, and Daniel and Dario for all the comments and advises, I was so lucky to meet you!

La moitié de cette thèse repose sur une collaboration de bientôt huit années avec l'administration fiscale du Sénégal durant lesquelles j'ai contracté des dettes avec plus de personnes que ces remerciements n'en peuvent contenir. Merci à Bassirou Samba Niasse, pour l'idée de départ, et sans qui rien n'aurait été possible. Merci Mor, le plus brillant d'entre nous, pour le soutien sans faille, on se voit à Thiès la prochaine fois ? Merci mille fois Hady, puissent les projets être à la

hauteur de ton engagement à les défendre et les porter. Merci Mabeye, à la fois si tranquille et si efficace. Merci à Alioune et Guisse pour tout l'administratif et la chaleur de l'accueil. Et merci à Oumar et Diangar de nous aider à louvoyer dans cet océan d'information. Merci enfin à toutes celles et ceux qui ont fait le travail dit de « terrain », et qu'on retrouve le plus souvent en note de bas de page dans les publications. Anaëlle, Eva, Hasnae, c'est votre tour ! On aurait aimé que tu restes davantage avec nous Ousseynou, et la famille d'ailleurs ? Et Mouhamed, de retour du Canada ? Aïssatou, ton sens de l'analyse nous manque. Dieynaba, ça gcolapse ? Heureusement tout le monde n'est pas parti : Binta, la Coordinatrice, Mbacké, le Rigoureux, Amadoune, l'Aîné. Quel parcours depuis toutes ces années ! De la numérisations des états 1024 à la coordination des interventions, du déploiement des répertoires au nettoyage des données : les projets vous doivent tant. Et la fonction publique serait bien inspirée de vous embaucher une bonne fois pour toutes !

Tout le monde ou presque a fini par en entendre parler, si bien que chacun s'est senti le besoin de me soutenir à un moment ou à un autre. Merci Sébastien d'avoir montré le chemin. Alex, comment fait-on pour décider de finir ? Je ne t'ai pas assez remerciée, Clara, pour l'accueil à Londres. Niklas, I am so happy we met again, as if for a second time, that day in Dakar. À Laetitia, n'attendons pas aussi longtemps avant la prochaine fois. Dis Mariona, quand est-ce qu'on retourne faire du ski ensemble ? Alessandro, si j'en crois WhatsApp, il s'agirait de « catch up » depuis le 11 janvier – my bad. Mon grand Clément, à Copenhague promis ! Merci Rozenn de trouver les mots justes auxquels on n'aurait pas pensé. Merci Kévin d'avoir perdu une ou deux fois. Merci encore Marion pour le coup de boost. So, so, so happy for you Pepe. Merci Paul D. pour le soin que tu mets à chaque chose. Merci Laura, pour l'exemple déjà, les courses à Londres, au Bois de la Cambre ou que sais-je, et les rires bien sûr. A Paul B., pour l'interlude londonien, les photos de rouge-gorges et les histoires. A toi mon Nico, rien que pour tes beaux yeux. Et Loulou, pour l'air de la campagne, les feux de cheminée, l'autre vie possible.

Pao, t'as quitté Bruxelles, c'est comme s'il n'y avait pas eu d'été. Esther et Maxime vous ne valez pas mieux. Aaron, maybe not on our fridge anymore, but always in our hearts. Je ne referai les 20km qu'avec toi, Juju. Et voilà que c'est déjà notre tour maintenant de clôturer le cycle Bruxellois. La gorge se serre. Merci mille fois Clémentine pour le souffle, l'art du vivre avec nous, les messages de la dernière chance, et l'aide à domicile dans les heures les plus sombres. Merci à vous Ryan et Amélie, pour tous les brunchs, toutes les soirées chantées et même les meubles ! Merci Ju de nous avoir aidés à apprivoiser la ville et ses quartiers, pour les conversations plus deep et l'indulgence pour mes jeux de mots. Greg, mon pote, mon camp de base, j'aurai fait durer autant que j'ai pu, j'te jure ... allez, allez ... embrassez Zouzou et Jo. AG pasta forever !

Un immense merci à la famille Gontard, pour sa générosité sans limite et son sens de la fête. À mes parents, pour la tendresse, notamment quand les pépins de santé s'enchaînaient et que je n'y voyais plus très clair. On se sera bien amusés avec la chaise roulante tout de même ! Papa, tu trouveras peut-être que je ne vais pas assez loin, Maman, tu aurais probablement regardé tout ça de plus près. A celles et ceux aussi qui vous précédèrent, de Carhaix à Starzawa en passant par Tours, qui sait vraiment de qui l'on tient ? Merci à toi Ana, soeur à jamais complice et témoin privilégiée de ma vie entière.

Et bien sûr, Nolwenn. Si l'on avait su qu'on allait vivre aussi longtemps à Bruxelles, on aurait acheté. Merci d'être venue, merci d'être restée. La dizaine de voyages à Dakar, les chutes et rechutes, le semestre à Londres, et les péripéties d'une dernière ligne droite qui n'était jamais la dernière : tu es peut-être la plus patiente des deux finalement. Merci pour ton soutien quotidien, toutes ces années durant. Merci aussi pour tout ce que ma vie a de moins académique grâce à toi, de la découverte de Karafun à celle du papier cadeau transparent, du brouillard des Abruzzes aux croquis du jour. La route n'a pas toujours le tracé souple et régulier de celle qui mène aux Grands-Jardins, mais j'ai la chance de la faire avec toi. Et regarde qui nous sourit dans le rétro !?

Résumé

Dans leur forme actuelle, les États restent les principaux garants de la sécurité extérieure, écrivent le droit et assurent son respect, produisent et entretiennent des biens publics que des acteurs privés ne pourraient pas financer eux-mêmes, et assurent à leurs citoyens et citoyennes un accès à des services d'éducation et de santé de base et – dans les pays riches – de protection sociale. Pour remplir ces fonctions, les États se financent principalement en prélevant des impôts. Or les pays les plus pauvres tendent à collecter moins d'impôts, en pourcentage de leur produit intérieur brut, que les pays riches. S'il n'existe pas dans la littérature scientifique de consensus autour du taux de pression fiscale optimal, les conditions de vie observées dans ces pays laissent peu de doutes quant à la nécessité de financer davantage de biens publics. L'idée que ces pays devraient augmenter leurs recettes fiscales plutôt que d'avoir recours à l'endettement ou d'aide internationale, pour pallier les insuffisances de leurs finances publiques, fait de plus en plus consensus. La question reste de savoir comment une telle transition est possible. La première partie de cette thèse s'attache à apporter des éléments de réponses à cette question en étudiant comment un pays comme le Sénégal peut collecter davantage d'impôts sans modifier les règles fiscales, mais en renforçant stratégiquement leur application.

L'approche méthodologique employée dans les deux premiers chapitres est celle de l'évaluation d'impact contrefactuelle. Cette méthode consiste à estimer, à l'aide d'outils statistiques, l'effet causal d'un changement (ici une politique publique) en comparant les individus affectés (« traités ») par le changement en question avec un groupe d'individus témoins, qui n'ont, eux, pas été touchés mais dont il est crédible de penser qu'ils reflètent ce qu'auraient été les individus traités s'ils ne l'avaient pas été. Les deux premiers chapitres sont le fruit d'une collaboration de plusieurs années avec l'administration fiscale sénégalaise, durant lesquelles j'ai pu collecter et consolider des données administratives d'une grande richesse. Cette longue et régulière collaboration m'a aussi donné la chance d'observer le fonctionnement de l'administration « de l'intérieur » et de m'entretenir de nombreuses fois avec le personnel administratif – tout échelon confondu. A quelques exceptions près, ces données qualitatives ne sont pas exposées dans cette thèse, mais elles ont significativement contribué à l'élaboration des projets et l'interprétation de leurs résultats.

L'histoire des pays riches suggère que ces derniers n'auraient pu atteindre leur taux de pression fiscale actuel sans instaurer systématiquement la déclaration des revenus par les tiers – règle administrative obligeant, par exemple, les employeurs à déclarer la rémunération de leurs employés. Ce type de stratégie permettrait à l'administration de récolter plus d'informations

sur d'éventuels fraudeurs, mais aussi de réduire les opportunités de sous-déclaration chez ceux dont les revenus sont déclarés pas des tiers. Dans le premier chapitre – coécrit avec Bassirou Sarr et Mattea Stein – nous montrons que ce type d'instrument est en vigueur au Sénégal mais qu'il est en encore appliqué de manière imparfaite. Chaque année les entreprises sénégalaises sont tenues de déclarer le montant qu'elles ont versé à chacun de leurs prestataires de service, en indiquant l'identifiant numérique du prestataire. Une inspection systématique des données collectées sur la période 2018-2020 auprès des quatre centres fiscaux les plus importants du pays, révèle cependant que près de 90% des entreprises omettent d'indiquer l'identifiant de leurs prestataires. Le recueil de témoignages d'inspecteurs des impôts en charge de mener les audits suggère que cet état des choses résulte, en partie du moins, d'une vérification encore trop peu systématique de ces documents. En appliquant la méthode des essais aléatoires contrôlés, nous avons coordonné avec l'administration fiscale une campagne de communication auprès de ces entreprises afin de les dissuader d'omettre des informations dans leurs déclarations ultérieures. L'intervention s'avère particulièrement efficace. La part des prestataires renseignés par leur client avec un identifiant numérique valide augmente de 24.7 points de pourcentage en moyenne, soit une différence de 52% par rapport à la moyenne dans le groupe témoin. La disponibilité de cette information permet alors de retrouver les déclarations soumises par les prestataires eux-mêmes et ainsi de comparer la somme des revenus qu'ils ont perçue selon leur(s) client(s), au chiffre d'affaires qu'ils ont eux-mêmes déclaré. Nous pouvons ainsi montrer que l'intervention augmente significativement le taux de détection de prestataires qui sous-déclarent leurs revenus. Une comparaison de la distribution des revenus non déclarés par les prestataires informels avec celle de ceux sous-déclarés par des fournisseurs dûment enregistrés aux impôts révèle qu'à court terme, contrairement aux stratégies parfois mises en avant, l'administration aurait moins à gagner en enrôlant les entreprises informelles qu'en renforçant sa surveillance des contribuables déjà enregistrés. Finalement, nous estimons qu'en utilisant cette information supplémentaire pour auditer en priorité les 300 à 400 prestataires qui sous-déclarent le plus, l'administration augmenterait ses recettes fiscales directement tirées des audits de plus 100%. Les résultats de cette étude suggèrent qu'en s'appuyant sur les plus grandes entreprises, les administrations de pays comme le Sénégal peuvent encore, à relativement peu de frais, récolter de grandes quantités d'informations utiles à la détection de fraudes fiscales conséquentes.

Au cours des deux dernières décennies, l'accélération considérable de la production et de la diffusion des technologies de l'information a permis aux gouvernements de rapidement moderniser leurs administrations. Les pays à faibles revenus ont largement participé à cette transition et investissent désormais dans des services autorisant la déclaration et le paiement des impôts en ligne. Dans des contextes où la conformité fiscale est encore faible, les gouvernements qui adoptent ces nouveaux outils espèrent voir leurs recettes augmenter en rendant plus

efficaces leurs administrations, en diminuant les coûts de conformité pour les contribuables et en limitant les opportunités de collusion ou d'extorsion facilitées par les interactions directes entre administrateurs et contribuables. Cela étant, les résultats espérés pourraient ne pas se matérialiser. Les gains d'efficacité générés par l'administration pourraient servir d'autres objectifs qu'un renforcement de la lutte contre l'évasion fiscale, tandis que les contribuables qui n'ont plus à se déplacer pourraient ne pas modifier pour autant leur comportement fiscal. Par ailleurs, la diminution des cas d'extorsion et de collusion n'aura d'impact significatif sur les recettes qu'à condition que ces cas soient suffisamment fréquents *ex-ante*. Il est difficile par conséquent de comprendre et anticiper les conséquences de la déclaration et du paiement en ligne sans examen empirique. C'est ce à quoi s'attache le deuxième chapitre de cette thèse, coécrit avec Awa Diouf et Fabrizio Santoro. Début 2018, le Sénégal a rendu obligatoires la déclaration et le paiement en ligne pour les grandes entreprises. En exploitant l'univers des déclarations mensuelles de 2017 à 2019, nous appliquons la méthode des doubles différences, afin de comparer l'évolution des grandes entreprises avec celle de contribuables n'ayant pas adopté la réforme. Dans les deux à trois mois qui suivent la réforme, l'adoption de la télédéclaration diminue la probabilité de déclarer de 2.5 points de pourcentage, probablement du fait des coûts de transition, tant pour l'administration que pour les contribuables. A plus long terme, la réforme ne semble pas avoir d'impact significatif sur la probabilité de déclarer, ni sur la probabilité de déclarer un impôt dû non nul. La déclaration en ligne contribue en revanche à améliorer considérablement la qualité des informations collectées : la probabilité qu'une valeur soit manquante diminue de 5 à 7 points de pourcentage. Enfin, la réforme ne semble pas avoir eu d'impact sur la probabilité de payer (marge extensive), mais nous estimons un impact significatif sur le montant moyen payé (marge intensive). En conservant l'échantillon traité au complet, l'impact moyen pourrait s'élever jusqu'à USD 20,000 par mois, par entreprise – soit une augmentation des recettes fiscales agrégées de 6.5% sur l'année. Cependant, la magnitude de ces résultats dépend crucialement de quelques très grands contribuables. En leur absence, l'effet moyen sur les paiements converge vers USD 5,000. Le recueil de témoignages au sein du centre des grandes entreprises suggère que cette augmentation des paiements effectifs pourrait résulter de la diminution des chèques impayés, ou de l'augmentation des procédures d'avis à tiers détenteurs.

L'augmentation des recettes fiscales n'est souhaitable que si elle permet au gouvernement d'allouer les ressources de manière plus efficace ou plus équitable que ne le feraient les marchés privés. Cela étant, nous disposons d'encore très peu de connaissances empiriques sur les niveaux d'inégalités de richesse ou de revenu dans les pays à revenu faible ou moyen, et *a fortiori* sur l'incidence globale de l'intervention des gouvernements sur le niveau de vie des individus. La deuxième partie de cette thèse, coécrite avec Aroop Charterjee et Amory Gethin, contribue à combler ces lacunes en étudiant le cas de l'Afrique du Sud.

L'approche méthodologique de ces deux derniers chapitres ne consiste plus à identifier et quantifier un lien causal entre une politique et ses résultats attendus, mais à mettre en cohérence des données micro-économiques d'une part (données d'enquêtes ou déclarations fiscales) avec des données macro-économiques d'autre part (comptes nationaux, budgets du gouvernement) afin de mesurer l'évolution des inégalités de richesse et de revenu. Prise à part, chaque source est incomplète : les données d'enquête manquent ou sous-estiment des composants du revenu national, les données fiscales ne portent que sur une partie de la population, et les données macro-économiques ne renseignent que la taille des agrégats dont il s'agit justement de savoir comment ils sont distribués. Dès lors, la stratégie générale employée dans les deux derniers chapitres consiste à partir des agrégats pour les distribuer aux individus en fonction d'informations présentes dans les enquêtes. Le choix d'étudier l'Afrique du Sud s'est imposé pour deux raisons. C'est tout d'abord le pays d'Afrique subsaharienne pour lequel existent les données les plus riches. Dans la plupart des pays du continent, les sources restent encore trop limitées ou imparfaites pour autoriser ce type d'exercice. C'est ensuite parce qu'à l'aune de l'extrême violence de son histoire, la question de l'évolution récente des inégalités y est particulièrement pertinente. L'Afrique du Sud a en effet été marquée par les inégalités raciales durant des siècles. Dès sa création même comme dominion de l'empire britannique, la minorité blanche au pouvoir adopta le Native Land Act, lequel restreignait la population noire africaine à ne pouvoir acheter ou louer de terres que dans des dites "réserves" constituant moins de 10% du territoire du pays. Le racisme institutionnel culmina avec l'instauration du régime d'apartheid de 1948 à 1994 déclinant la ségrégation raciale à toutes les dimensions de la vie sociale (mariage, travail, logement, circulation etc.).

Le troisième chapitre porte sur la mesure des inégalités de patrimoine en Afrique du Sud depuis la fin de l'apartheid. Parce que les sources mesurant le capital individuel sont rares, il est souvent beaucoup plus difficile de mesurer les inégalités de patrimoine que celles de revenu. L'Afrique du Sud ne fait pas exception à cet égard, le patrimoine étant encore très mal renseigné dans les enquêtes. Nous observons néanmoins certains revenus du capital dans ces mêmes enquêtes (loyers, dividendes, retraites, etc.), et la banque centrale sud-africaine publie ses estimations du patrimoine total (et des dettes) détenu par les ménages. En distribuant les stocks agrégés en proportion des flux individuels, nous parvenons à réconcilier les deux sources et reconstituer une distribution du patrimoine aussi complète et cohérente que possible. Nos estimations mettent à jour des niveaux extrêmes d'inégalités. En 2017, les 10% les plus riches possèdent 86% de la richesse nationale. Plus haut dans la distribution, les 3,500 les plus riches (top 0.01%) concentrent 15% du patrimoine des ménages, soit autant que les 90% les plus pauvres (31,860,000 individus). Ces niveaux de concentrations touchent toutes les formes d'actifs : le logement, les fonds de pension et les actifs financiers. Si l'on observe bien une

évolution du patrimoine moyen selon l'âge, les niveaux inégalités de patrimoine à l'intérieur de chaque tranche d'âge restent parfaitement identiques à ceux mesurés pour la population totale. Ces résultats sont cohérents avec l'hypothèse selon laquelle l'héritage déterminerait non seulement les niveaux de départ mais aussi les trajectoires individuelles d'accumulation du patrimoine. Nous n'observons aucun signe de diminution des inégalités de richesse depuis la fin de l'apartheid.

Dans le quatrième et dernier chapitre, nous nous attachons cette fois-ci à mesurer la répartition des revenus avant et après l'intervention de l'état. Dans la logique du chapitre précédent, il s'agit une fois encore de réconcilier de très nombreuses sources de données afin de produire une image statistique aussi fiable et complète que possible de l'objet d'intérêt. La difficulté tient ici non pas tant à la mesure du revenu, pour lequel nous avons d'assez bonnes sources directes, qu'à la modélisation de l'incidence des impôts et transferts sur ces revenus. Nous procédons en quatre étapes. Nous commençons par produire une distribution des revenus avant toute forme de redistribution, dont les concepts et les agrégats sont cohérents avec les comptes nationaux. A partir de cette distribution, nous en estimons une deuxième en retranchant les contributions pour les assurances retraite et chômage, et en ajoutant les allocations correspondantes. Ces formes d'assurances ne bénéficiant qu'à une petite partie de la population, elles ne jouent qu'un rôle mineur dans la redistribution des revenus en Afrique du Sud. Puis nous produisons une distribution des revenus après impôts, en imputant chaque impôt (sur le revenu, les bénéfices, le patrimoine, la valeur ajoutée, etc.) selon des méthodes d'allocation aussi proches que possible de ses règles de calcul et d'incidence. L'étape finale consiste à récupérer une distribution après impôts et transferts, en distribuant la dépense publique, composante par composante, à chacun de ses bénéficiaires estimés. Ceci implique d'allouer non seulement les transferts monétaires directs mais aussi les transferts en nature (des services d'éducation et de santé aux dépenses d'administration générale, en passant par la police et les transports). Nos résultats montrent que si le revenu national par habitant a augmenté de près 40% depuis la fin du régime d'apartheid, les inégalités de revenus avant impôts et transferts se sont aggravées. Le top 1% perçoit 28% des revenus en 2019 contre 23% en 1993, tandis que le *bottom* 50% continue de n'en percevoir que 3% environ. Cela étant, le système, déjà redistributif en 1993, l'est devenu de plus en plus à partir de 2000, à mesure que se généralisaient les transferts monétaires directs, puis après la crise des *sub-primes* suite à une forte expansion budgétaire à la faveur de dépenses d'éducation, de santé et des gouvernement locaux. Cette tendance a permis de contrebalancer entièrement l'augmentation des inégalités de revenus avant impôts et transferts. Les inégalités raciales, quant à elles, ont significativement diminué, au bénéfice, principalement, des 10% les plus riches de la population noire africaine et de celle dite "*coloured*".

Abstract

In their current form, States remain the main guarantors of national defense; they write the law and ensure that it is respected; they produce and maintain public goods that private actors could not finance themselves, and they provide their citizens with access to basic education and health services and – in rich countries – social protection. To carry out these functions, governments finance themselves mainly by levying taxes. However, the poorest countries tend to collect less tax as a percentage of their gross domestic product than the richest countries. Although there is no consensus in the scientific literature on the optimal amount of tax to collect, the living conditions observed in these countries leave little doubt as to the need to finance more public goods. There is a growing consensus that these countries should increase their tax revenues rather than resorting to debt or international aid to make up for the shortcomings of their public finances. The question remains whether and how such a transition is possible. The first part of this thesis seeks to provide some answers to this question by studying how a country like Senegal can collect more taxes, not by changing the tax rules, but by strategically strengthening their application.

The methodological approach used in the first two chapters is that of counterfactual impact assessment. Using statistical tools, this method consists of estimating the causal effect of a change (in this case, a public policy) by comparing the individuals affected ('treated') by the change in question with a group of control individuals, who were not affected but who can credibly be assumed to reflect what the treated individuals would have been had they not been affected. The first two chapters are the fruit of a collaboration of several years with the Senegalese tax authorities, during which I was able to collect and consolidate a wealth of administrative data. This long and regular collaboration has also given me the chance to observe the operational routine of the administration "from the inside" and to talk to administrative staff on numerous occasions and at all levels. With a few exceptions, this qualitative data is not presented in this thesis, but it contributed significantly to the development of the projects and the interpretation of their results.

The history of rich countries suggests that they could not have achieved their current tax burden without systematically introducing third-party reporting - an administrative rule requiring employers to declare their employees' remuneration, for example. This type of strategy would enable the authorities to gather more information on potential tax evaders, but also to reduce the opportunities for underreporting among those whose income is declared by third parties. In the first chapter – co-written with Bassirou Sarr and Mattea Stein – we show that this type of instrument is in force in Senegal but that it is still imperfectly applied. Each year, Senegalese

companies are required to declare the amount they have paid to each of their service providers, indicating the provider's numerical identifier. However, a systematic inspection of the data collected over the period 2018-2020 from the country's four largest tax centres reveals that almost 90% of companies fail to indicate the identifier of their service providers. Testimonies from tax inspectors responsible for carrying out audits suggest that this situation is due, at least in part, to the fact that these documents are still not checked systematically enough. Using the randomised controlled trial method, we coordinated a communication campaign with the tax authorities to dissuade these companies from omitting information from their subsequent returns. The campaign has proved particularly effective. The proportion of service providers whose clients provided valid numerical identifiers increased by an average of 24.7 percentage points, a difference of 52% compared with the average in the control group. The availability of this information makes it possible to trace the declarations submitted by the service providers and thus to compare the amount of income they received from their client(s) with the turnover they themselves declared. We can thus show that the intervention significantly increases the rate of detection of providers who underreport their income. A comparison of the distribution of informal service providers' unreported income with that of income under-reported by suppliers duly registered with the tax authorities reveals that, in the short term, contrary to the strategies sometimes put forward, the administration would have less to gain by enrolling informal businesses than by stepping up its surveillance of already registered taxpayers. Finally, we estimate that by using this additional information to prioritise audits of the 300 to 400 service providers who underreport the most, the authorities would increase their tax revenue directly from audits by more than 100%. The results of this study suggest that by relying on the largest companies, administrations in countries such as Senegal can still, at relatively little cost, gather large quantities of information useful for detecting substantial tax fraud.

Over the last two decades, the considerable acceleration in the production and dissemination of information technologies has enabled governments to rapidly modernise their administrations. Low-income countries have played a large part in this transition and are now investing in services that allow tax to be declared and paid online. In contexts where tax compliance is still low, governments that adopt these new tools hope to increase their revenues by making their administrations more efficient, reducing compliance costs for taxpayers and limiting opportunities for collusion or extortion facilitated by direct interaction between administrators and taxpayers. That said, the hoped-for results may not materialise. The efficiency gains generated by the administration could serve different objectives, other than strengthening the fight against tax evasion, while taxpayers who no longer must travel may not change their tax behaviour as a result. Furthermore, the reduction in cases of extortion and collusion only has a significant impact on revenue if these cases are frequent enough. It is hence difficult to

understand and anticipate the consequences of online filing and payment without empirical examination. This is the focus of the second chapter of this thesis, co-authored with Awa Diouf and Fabrizio Santoro. At the beginning of 2018, Senegal made online declaration and payment mandatory for large companies. By exploiting the universe of monthly returns from 2017 to 2019, we apply the double-differences method, in order to compare the evolution of large companies with that of taxpayers who did not adopt the reform. In the two to three months following the reform, the adoption of electronic filing reduces the probability of filing by 2.5 percentage points, probably because of the transition costs for both the administration and taxpayers. In the longer term, the reform does not appear to have a significant impact on the probability of filing tax returns, nor on the probability of filing a non-zero tax liability. On the other hand, online filing considerably improves the quality of the information collected: the probability of a value being missing decreases by 5 to 7 percentage points. Finally, the reform does not appear to have had any impact on the probability of paying (extensive margin), but we estimate a significant impact on the average amount paid (intensive margin). Keeping the full treatment group, the average impact could be as much as USD 20,000 per month, per company - an increase in aggregate tax revenue of 6.5% over the year. However, the magnitude of these results depends crucially on a few very large taxpayers. In their absence, the average effect on payments converges to around USD 5,000. Testimonies from the large business centre suggest that this increase in actual payments could be the result of a reduction in unpaid cheques, or an increase in forced recovery through the banking system.

Increased tax revenue is only desirable if it enables the government to allocate resources more efficiently or more equitably than would private markets. However, we still have very little empirical knowledge about the levels of wealth or income inequality in low- and middle-income countries, let alone about the overall incidence of government intervention on individual living standards. The second part of this thesis, co-authored with Aroop Charterjee and Amory Gethin, contributes to filling these gaps by studying the case of South Africa.

The methodological approach of these last two chapters is no longer to identify and quantify a causal link between a policy and its expected results, but to reconcile microeconomic data (survey data or tax returns) with macroeconomic data (national accounts, government budgets) in order to measure changes in wealth and income inequalities. Taken separately, each source is incomplete: survey data misses or underestimates components of national income, tax data covers only part of the population, and macro-economic data only provides information on the size of the aggregates whose distribution we are interested in. Consequently, the general strategy employed in the last two chapters consists of starting from the aggregates and distributing them to individuals on the basis of information available in the surveys. South Africa was chosen

for two reasons. First, it is the sub-Saharan African country for which the richest data exists. In most of the continent's countries, the sources are still too limited or imperfect to allow this type of exercise. Second, because of the extreme violence of its history, the question of recent trends in inequality is particularly relevant. South Africa has been marked by centuries of racial inequality. From its very creation as a dominion of the British Empire, the white minority in power adopted the Native Land Act, which restricted the Black African population to being able to buy or rent land only in so-called "reserves" accounting for less than 10% of the country's territory. Institutional racism culminated in the establishment of the apartheid regime from 1948 to 1994, which extended racial segregation to all aspects of social life (marriage, work, housing, movement, etc.).

The third chapter looks at the measurement of wealth inequality in South Africa since the end of apartheid. Because sources measuring individual capital are scarce, it is often much more difficult to measure wealth inequality than income inequality. South Africa is no exception in this respect, as wealth is still very poorly recorded in surveys. We do, however, observe some capital income (rents, dividends, pensions, etc.) in these same surveys, and the South African central bank publishes its estimates of total wealth (and debt) held by households. By distributing aggregate stocks as a proportion of individual flows, we are able to reconcile the two sources and reconstitute a distribution of wealth that is as complete and consistent as possible. Our estimates reveal extreme levels of inequality. In 2017, the richest 10% owned 86% of national wealth. Further up the distribution, the richest 3,500 (top 0.01%) concentrate 15% of household wealth, as much as the poorest 90% (31,860,000 individuals). These levels of concentration affect all forms of assets: housing, pension funds and financial assets. While average wealth does change with age, the levels of wealth inequality within each age group remain perfectly identical to those measured for the total population. These results are consistent with the hypothesis that inheritance determines not only starting levels but also individual trajectories of wealth accumulation. We find no evidence of a decline in wealth inequality since the end of apartheid.

In the fourth and final chapter, we focus this time on measuring the distribution of income, before and after state intervention. Following the logic of the previous chapter, the aim is once again to reconcile a large number of data sources in order to produce a statistical picture of the subject of interest that is as reliable and complete as possible. The difficulty here lies not so much in measuring income, for which we have relatively good direct sources, but in modelling the impact of taxes and transfers on this income. We proceed in four stages. We begin by producing a distribution of income before any form of redistribution, whose concepts and aggregates are consistent with the national accounts. From this distribution, we estimate

a second distribution by subtracting pension and unemployment insurance contributions and adding the corresponding benefits. As these forms of insurance only benefit a small proportion of the population, they play only a minor role in income redistribution in South Africa. We then produce an after-tax income distribution, imputing each tax (on income, profits, wealth, value added, etc.) according to allocation methods that are as close as possible to its calculation rules and incidence. The final stage consists of recovering a distribution after taxes and transfers, by distributing public spending, component by component, to each of its estimated beneficiaries. This involves allocating not only direct monetary transfers but also in-kind transfers (from education and health services to general administration expenditure, police and transport). Our results show that while per capita income has risen by almost 40% since the end of the apartheid regime, income inequality before taxes and transfers has increased. The top 1% received 28% of income in 2019, compared with 23% in 1993, while the bottom 50% continued to receive only around 3%. That said, the system, which was already redistributive in 1993, became increasingly so from 2000 onwards, as direct cash transfers became more widespread, and then after the sub-prime crisis following a strong budgetary expansion in favour of education, health and local government spendings. This trend has entirely offset the increase in income inequalities before taxes and transfers. Meanwhile, racial inequalities, have fallen significantly, mainly to the benefit of the richest 10% of the Black African and 'Coloured' population group.

Contents

1	Fraud Detection Under Limited State Capacity: Experimental Evidence From Senegal	27
1.1	Introduction	28
1.2	Institutional Context	33
1.3	Experimental Design	35
1.4	Data, Estimation and Baseline Statistics	37
1.4.1	Data	37
1.4.2	Estimation	38
1.4.3	Selection and Baseline Statistics	39
1.4.4	Assessing the take-up	40
1.5	Results on Misreporting and Fraud Detection	40
1.5.1	Misreporting of suppliers list significantly decreases	40
1.5.2	New information increases the detection of tax evading suppliers	42
1.6	Implication for the distribution of tax avoidance and the targeting of future deterrent measures	43
1.6.1	How many and how large are newly identified suppliers?	44
1.6.2	What do we learn about the distribution of tax evasion?	45
1.6.3	How much additional tax revenue can be expected from exploiting the new information to target future audits?	46
1.6.4	Cost-Benefit Analysis for the Tax Administration	49
1.7	Conclusion	50
1.8	Figures	52

1.9	Tables	65
	Appendices	77
A	Baseline data, Sample Selection and Randomisation	77
B	Baseline enforcement of supplier lists' declarative rules	78
C	Results on the Withholding Tax on Service (WTS)	79
D	Impact on suppliers	81
E	The Senegalese Tax System	84
F	Couldn't the client be strategic as well?	86
G	Detailed Message Content	88
H	Complementary Figures and Tables	93
2	Does Electronic Filing and Payment Increase Tax Compliance?	99
2.1	Introduction	100
2.2	Institutional Context	103
2.3	Assumptions and Expected Impacts	104
2.31	General behavioral assumptions	105
2.32	Expected impacts	105
2.4	Data and Empirical Strategy	108
2.41	Data	108
2.42	Selection	109
2.43	Empirical Strategy	110
2.5	Results	113
2.6	Conclusion	116
2.7	Figures	117
2.8	Tables	161
3	Wealth Inequality in South Africa, 1993–2017	165
3.1	Introduction	166
3.2	Data sources	168

3.21	Data sources on aggregate wealth	168
3.22	Data sources on the distribution of household wealth	172
3.3	Bridging the micro-macro gap: the distribution of wealth in South Africa	179
3.31	Methodological approaches: direct measurement, rescaling and capitali- sation	179
3.32	Measuring wealth inequality using survey data	182
3.33	Measuring wealth inequality using tax data	184
3.4	The distribution of wealth in South Africa: key results and comparative perspectives	186
3.5	Conclusion	192
4	Redistribution without Inclusion? Inequality in South Africa Since the End of Apartheid	195
4.1	Introduction	196
4.2	Data and Methodology	200
4.21	Conceptual Framework: Distributional National Accounts	201
4.22	From Reported Household Income to Factor National Income	202
4.23	From Factor National Income to Pretax National Income	205
4.24	From Pretax National Income to Posttax National Income: Taxes	206
4.25	From Pretax National Income to Posttax National Income: Transfers	208
4.26	Household Expenditure and Household Wealth	209
4.3	The Distribution of Factor Income	210
4.31	National Income Growth in South Africa since 1993	210
4.32	The Distribution of Factor National Income	210
4.33	Decomposing Factor Income Inequality: Labor versus Capital	212
4.4	The Distribution of Taxes and Transfers	212
4.41	From Factor to Pretax Income	213
4.42	The Distribution of Taxes	214
4.43	The Distribution of Transfers	216
4.44	The Overall Impact of the Tax-and-Transfer System	217

4.5	The Evolution of Racial and Spatial Inequality	219
4.51	Racial Inequality	219
4.52	Spatial Inequality	221
4.6	Conclusion	222
4.7	Figures	223
4.8	Tables	250
	Appendices	254
A	Harmonization of Macroeconomic Aggregates	254
B	Construction of DINA Microfile	256
C	Distribution of Taxes and Transfers	258

List of Figures

1.1	Timeline - Intervention and Main Declarations	53
1.2	Impact on the Share of Correctly Reported Suppliers - 2022 Suppliers Lists - Heterogeneity by Size and Age	54
1.3	Number of Detected Tax Evading Suppliers - by Quantiles of Unreported Amounts	55
1.4	Decomposition of 2022 Payments to Suppliers by Experiment Group, Suppliers Type and Decile of the Distribution of Payments	56
1.5	Number of Civil Servants per Taxpayers - by Tax Center and Civil Servant Category : 2016-2021	57
1.6	Distribution of the 2,975 Under-reporting Suppliers of the Treatment Group and Revenue Share	58
1.7	Baseline decomposition of the distribution of payment according WTS liability status - 2018	59
1.8	Baseline decomposition of the distribution of payment according to suppliers' sector - 2018	60
1.9	Baseline decomposition of the distribution of payment according to clients' tax centre	61
1.10	Baseline decomposition of the distribution of payment according the supplier' type of person	62
1.11	Returns expected from auditing the largest under-reporting formal suppli- ers according to the treatment group compared to estimated returns when targeting	63
1.12	Hypothetical Audit Returns and Short-Term Cost-Benefit Analysis	64

1.913	Spillover Effects on Suppliers' Fiscal Behavior - Heterogeneity by Decile of Average Revenue over 2018-2020	85
2.71	Adoption of Etax	118
2.72	Probability distribution with respect to taxpayers size	120
2.73	Sensitivity to sample selection	122
2.74	Submits at least one declaration	123
2.75	Some declaration has missing information	125
2.76	Some declaration has inconsistent information	127
2.77	Some declaration has implausible information	129
2.78	Liability declared is positive	131
2.79	Liability declared (USD)	133
2.710	Tax paid is positive	136
2.711	Amount paid (USD)	138
2.712	Turnover (USD)	145
2.713	Payment as a percentage of total liability paid	150
2.714	On measurement and missing information	152
2.715	COVID shock on the Probability to submit at least one declaration	155
2.716	COVID shock on the probability to pay a positive amount	157
3.21	The evolution of household wealth in South Africa, 1975-2018	170
3.22	Share of financial assets held through trusts, 1975-2018	178
3.43	The composition of assets by wealth group in 2017	189
3.44	Average wealth by age relative to average wealth per adult, 2010-2017	190
3.45	Wealth inequality within age groups, 2010-2017	190
3.46	South African wealth inequality in comparative perspective: Top 10 per cent wealth share	193
3.47	South African wealth inequality in comparative perspective: Top 1 per cent wealth share	193
4.71	Average national income per capita, 1993-2019	224

4.72	The distribution of factor national income, 1993-2019	225
4.73	Decomposing top factor income inequality	227
4.74	From factor to pretax income: net transfers operated between factor income groups by the pension and unemployment insurance systems	229
4.75	Government revenue in South Africa	230
4.76	Taxes paid by pretax national income group: 1993 versus 2019	231
4.77	Taxes paid as a percentage of pretax income	233
4.78	Government expenditure in South Africa	235
4.79	The rise of social transfers	236
4.710	The overall impact of taxes and transfers on inequality	238
4.711	Redistribution, inequality, and growth: cumulated income growth by per- centile, 1993-2019	240
4.712	Demographic Decomposition by Racial Group	241
4.713	Racial inequality and top incomes	243
4.714	The structure of racial inequality in 2019	246
4.715	Spatial inequality and redistribution	248

List of Tables

1.1	Reporting rules for suppliers list - examples for each type of lines	66
1.2	Rules and examples for the definition of the under-reporting (formal) suppliers	67
1.3	Suppliers' lists Declaration and Sample Selection by Tax Units	68
1.4	Sectoral Decomposition of the Selected Sample	69
1.5	Baseline patterns of declarations and payments of WTS	70
1.6	Summary Statistics and Balance tests - Full Sample (N = 3,487)	71
1.7	Direct Responses - Decomposition by Tax Centre	72
1.8	Impact on Misreporting, Information, Fraud Detection and Remittance of Withholding Tax on Informal Service Suppliers (WTS)	73
1.9	Identified Suppliers in 2022 - Summary Statistics by Intervention Group and Suppliers Compliance Status	74
1.10	Distribution of Unreported amounts - Suppliers of the Treatment Group - 2022	75
1.11	Underreporting Rates in 2022 by Decile of Declared Revenue	76
1.912	Impact on Revenue declared by Suppliers identified ex-ante	83
1.913	Summary Statistics and Balance tests - Conditional on Declaring 2022 Suppli- ers List (N = 2,633)	94
1.914	Summary Statistics and Balance tests - Conditional on Declaring 2021 Suppli- ers List (N = 2,608)	95
1.915	Conditional Impact on Misreporting, Information, Fraud Detection - comparing 2021 and 2022 suppliers lists	96
1.916	Comparing WTS declared in Yearly Suppliers List with monthly WTS Remit- tances - 2022	97

1.917	Treatment Heterogeneity on Monthly Declaration and Remittance of the Withholding Tax on Informal Service Suppliers (WTS)	98
2.81	Number of Civil Servants by treatment status and qualification - 2016-2022 .	162
2.82	Main Results	163
2.83	Main Results - excluding the transition period	164
3.21	Ownership rates and coverage of household balance sheets by asset class in NIDS	173
3.22	The coverage of selected national accounts components in South African surveys	174
3.23	The coverage of owner-occupied housing, mortgage debt and other debt in South African surveys	175
3.24	The coverage of capital income in the tax microdata	177
3.35	Estimating the distribution of personal wealth in South Africa: a mixed approach	181
3.36	Shares of household wealth held by groups in South Africa: survey-based results	182
3.37	Shares of household wealth held by groups in South Africa: results from tax microdata and survey combined	184
3.48	The distribution of personal wealth in South Africa in 2017	187
3.49	Share of total assets held by wealth group by asset class, 2017	187
4.81	The distribution of factor national income and pretax national income	251
4.82	The distribution of taxes	252
4.83	The distribution of factor income in South Africa in 2019	253

Chapter 1

Fraud Detection Under Limited State Capacity: Experimental Evidence From Senegal

with Bassirou SARR and Mattea STEIN

Abstract Tax administrations in low-income countries face widespread tax evasion and high enforcement costs. They thus need information to detect where tax evasion is most severe, and allocate scarce resources accordingly. This paper shows that leveraging large firms' trading network to collect information about their suppliers is a cost-efficient way to detect tax evasion and increase future audit returns. We collaborate with the Senegalese tax administration on a vast data collection effort to digitise lists of payments submitted by the largest firms and show that 88.6% of these firms provide incomplete information about their suppliers. This prevents any cross-checking against income declared by the suppliers themselves. We then randomise a low-cost communication campaign across all 3,487 misreporting firms, to discourage future misreporting. The intervention increases the prevalence of suppliers' identification information by 52%. In aggregate, this allows to uncover \$145.5 million in unreported revenue (i.e. 0.5 % of GDP). Most of it accrues to a few tax-registered suppliers, as opposed to informal ones. A simulation exercise shows that exploiting the newly available information to target the largest under-reporting suppliers would increase audit returns by at least 100%.

1.1. Introduction

Governments in low-income countries (LICs) lack the resources to sustainably fund public goods and often collect little tax revenue compared to other countries.¹ At the same time, tax evasion is widespread in LICs. At the extensive margin, evidence from household survey data shows that less than 10% of the adult population is subject to personal income tax (Bonnet, Vanek, Chen, et al., 2019). Recent evidence also suggests that most households are often subject to very little indirect tax, as they mainly buy from informal markets (Bachas, Gadenne, and Jensen, 2020b). Moreover, using administrative micro-data, some studies document very prevalent tax evasion among the tax-registered, i.e. at the intensive margin (Best, Shah, and Waseem, 2021a).² Altogether, this evidence suggests that LICs could improve tax collection by strengthening enforcement at both the extensive and intensive margins, without changing existing policy rules that define tax bases or rates. However, enforcement is costly, especially in LICs where the lack of a systematic address system, imperfect transport infrastructure, lower levels of digitisation, and partial financial inclusion all increase the cost for the administration to reach out to taxpayers (and vice versa). Tax enforcement is also more difficult in more fragile states, where low levels of trust in government institutions can foster tax resistance (Besley, 2020). Face with widespread tax evasion and high enforcement costs, tax administrations in LICs do not have the resources to register and monitor every non-compliant actor and need to strategically allocate scarce resources. Yet, doing so requires access to reliable information about taxpayers and their tax base, which is particularly difficult to collect in an environment where statistical capacity is weak, several institutional procedures are still paper-based, and national registries may be flawed.³ In the context of Senegal, this paper provides evidence that leveraging large firms' trading network to collect information about their suppliers is a cost-efficient way to detect tax evasion and increase future audit returns.

By law, Senegalese firms must withhold a 5% tax on service suppliers (henceforth *withholding tax on services* - WTS), which are not tax registered.⁴ Similar tax instruments are in place in other sub-Saharan African countries including Mauritania, Côte d'Ivoire, Ghana and Nigeria,

¹Using comparable tax revenue statistics (excluding social contribution) from OECD's [Global Revenue Statistics Database](#), and the most recent [World Bank income group classification](#), we estimate that the average tax-to-GDP ratio was 32.2% among (46) high-income countries, 20.8% among (31) upper-middle-income countries, 17.9 % among (30) lower-middle-income countries and 12.8% among (11) low-income countries - with sub-Saharan African countries accounting for 100% and 36% of the LICs and the LMICs group respectively.

²See also: [Shimeles, Gurara, and Woldeyes \(2017\)](#) for Ethiopia, [Mascagni et al. \(2022\)](#) for Rwanda, and [Almunia et al. \(2022\)](#) for Uganda, all of which classified as LICs.

³See [World-Bank, 2021](#), [Czajka et al., 2022](#), [Clark, Metz, and Casher, 2022](#), [OCDE, 2020](#).

⁴With the exemption of foreign suppliers, transactions below USD 40 and suppliers operating in the Hospitality sector.

with the objective to reach informal actors through formal ones. In Senegal, the WTS accounts for slightly less than 1% of all fiscal revenue in 2019. In practice, firms must collect, declare and remit the tax to the administration on a monthly basis (henceforth *monthly WTS*) but without providing the details of the transactions. At the end of each calendar year, remitting firms must also declare the list of all their service suppliers (henceforth *suppliers lists*), whether tax-registered or not, indicating, for each of them, the amount paid and their identification number (henceforth *ID*). Tax inspectors may use this detailed source to (a) verify whether the 5% tax on informal service suppliers has been duly remitted, (b) check these charges' eligibility for deduction, as well as to (c) compare payments received by tax-registered suppliers to the self-declared income tax statement of the later to potentially detect tax evading suppliers. If performed at all, such verification are always carried out on a case by case basis.

We collaborated with the tax administration (TA) on a vast, multi-year, data collection effort to assemble a unique dataset of 6,706 suppliers lists over 2018-2020,⁵ covering 75% of the taxpayers of the four largest tax centres.⁶ These tax centres monitor all of the country's largest firms, they account for 90% of all tax revenue, despite representing only 10% of all taxpayers. Our in depth analysis of this new database revealed that reporting rules of suppliers lists were seldom observed. Indeed, 88.6% of the taxpayers in our sample had misreported their suppliers list at least once during the period. On average, misreporting taxpayers omitted the IDs of 83.5 % of their suppliers. Without IDs, tax inspectors cannot use suppliers lists to perform the above mentioned verification. Other available information (names or addresses) are insufficiently standardised to allow precise and automated matching with other sources such as the tax-registry or other taxpayers' suppliers lists. This implies, for instance, that distinguishing tax-registered suppliers from informal ones, or even simply counting the number of clients by supplier, is exceedingly complex.

We identified all 3,487 firms who have misreported suppliers lists for at least one year during 2018-2020 and randomised an official communication campaign with the objective of deterring future misreporting. The message, individualised and conveyed by repeated emails and confirmation phone calls, consisted in (a) recalling reporting rules, (b) notifying that irregularities had been detected in past declarations and asking for rectifications, (c) urging them to do better in the future. We expected the intervention to operate through two mechanisms: (a) by providing information to possibly ignorant taxpayers, (b) by increasing taxpayers' perception of the likelihood of sanction in case they persist in unlawful behaviors. As large firms tend to make more and larger payments, targeting this group should maximise the expected gains and

⁵We digitised the first 4,797 ones, combining OCR technology and manual data entry. Then we collaborated with the TA to design a platform allowing taxpayers to directly submit their report online.

⁶The remaining 25%, most probably never submitted any suppliers list over that period.

make the intervention as cost-efficient as possible. In addition, because the administration more frequently interacts with them, their email address, phone number and postal address are more likely to be up-to-date, and they are probably less likely to ignore our message. The first email was sent just before the deadline for the declaration of 2021 suppliers list, in January 2022. Confirmation phone calls and subsequent email reminders ran through the year of 2022. Eventually, we collected evidence that the message was received: 60% of the treated group submitted a rectified declaration for the past years and we could confirm message reception for another 33% thanks to follow-up verification calls. Only 7% of the treated remained unreachable.

In this paper, we present three important sets of findings. First, we show that the intervention considerably improves the quantity and quality of information submitted. We analyze the impact on 2022 suppliers lists, submitted soon after the end of the full communication campaign (in January 2023). The proportion of taxpayers misreporting all their suppliers decreases by 19.5 percentage points, while the proportion of taxpayers perfectly complying with declarative rules increases by 10.1 percentage points. This essentially results from a large increase in the prevalence of IDs. The share, per client, of suppliers associated with a formally valid ID increases by 52% on average, while the share of misreported suppliers decreases by 42% on average. Such positive effects on reporting are more pronounced for larger firms. Treated firms were not more likely to report suppliers satisfying the conditions to be exempt from the WTS, nor to fill missing fields with intentionally wrong IDs. Interestingly, the intervention also increased the declaration of suppliers' phone numbers, an element which was not mentioned in the communication campaign nor is mandatory by law. Altogether, these results strongly suggest that firms did try to comply with the message instructions without circumventing the rules. Importantly, the intervention did not affect the number of suppliers, nor the distribution of payments, therefore the large changes observed in the prevalence of IDs do reflect modifications of declarative behavior rather than real changes in the service supply network.

Second, we show that the intervention increases the detection of tax evading suppliers. Using suppliers' identification information, we cross-check suppliers lists with the tax registry to split Senegalese suppliers into formal (tax-registered) and informal (unregistered) categories. Next, we create a revenue database for all tax-registered actors by assembling the universe of tax statements that contain information on total gross revenue.⁷ By matching formal suppliers to this revenue database, we can check whether the sum of all payments they received is superior to the total gross revenue they self-declared. We then distinguish *compliant* formal suppliers from *under-reporting* ones, according to each intervention group. With these definitions at hand,

⁷This consists of five types of declaration: one is declared every month (the value added tax), others are declared on a yearly basis (the corporate income tax, the personal income tax, the business property tax and the synthetic tax for small business).

we show that the intervention increases the share of payments going to identified suppliers classified as formal but under-reporting, by 2.5 percentage points, i.e. 45%, and the share of payments going to identified suppliers classified as informal by 17.5 percentage points, i.e. 76.4%.

Finally, we show that exploiting the newly available information to strategically allocate scarce audit resources could raise tax revenue collected upon audit. Aggregating data at the supplier level first allows us to draw further conclusions on the distribution of non-compliant suppliers according to each intervention group. Information declared by taxpayers of the treatment group allows us to identify 2,795 under-reporting formal suppliers and 23,141 informal suppliers. Comparing these figures to those derived from suppliers lists declared by the control group, we estimate that the intervention increased the detection of tax evaders by 60% on the intensive margin (formal but under-reporting suppliers), and by almost 200% on the extensive margin (informal suppliers). In aggregate, the intervention uncovers USD 145.5 million in unreported revenue, i.e. the equivalent of 0.5 % of GDP. Importantly, the information declared by taxpayers in the treatment group helps to identify more non-compliant firms at all levels of unreported income. In a simulation exercise, we compare the tax revenue expected from auditing the largest under-reporting suppliers according to the treatment group, to tax revenue expected when applying four alternative selection methods.⁸ Under conservative assumptions on unreported revenue uncovered by audits, our simulations show that selection based on newly available information outperforms all alternative methods. In particular, we estimate that using this information to target the 200 largest under-reporting taxpayers would increase the amount collected by at least 100%. Finally we show that even under very conservative assumptions, implementing our intervention and carrying the follow-up audits would help collect at least USD 5.7 million in tax revenue, inducing a benefit-to-cost ratio superior to 9.

Our paper makes three important contributions to the literature on tax compliance and enforcement in LICs. First we show that, even in LICs, communication campaigns can be an efficient deterrent tool to enforce third-party reporting, especially among the largest firms. Consistently with experimental studies from various context ([Kleven et al., 2011](#), [Brockmeyer and Hernandez, 2022](#)), historical and cross-country evidence suggest that the adoption of third-party reporting is instrumental to increase tax ratio along the development path ([Jensen, 2022](#), [Besley and Persson, 2014](#)). But the literature also shows that in LICs, where such tools are in place, significant improvements are yet to be made to reach full compliance ([Almunia et al., 2022](#), [Mascagni et al., 2023](#)). In this regard, our paper illustrates that the digitisation of

⁸Method 1: select the largest under-reporting suppliers according to the control group. Method 2: select a random sample from all the tax-registered. Method 3: select the largest taxpayers only. Method 4: select the largest taxpayers, by centre (i.e. same proportion from each tax centre)

third-party reports helps systematic detection of misreporting, which in turn allows the scale-up of individualised communication campaigns via direct mail technology. Our intervention demonstrates the effectiveness of such campaigns despite being a relatively mild method compared to other deterrent alternatives. These results contrast with that of recent studies which find no (Santoro et al., 2020) or limited (Mascagni and Nell, 2022, Almunia et al., 2023) impact on tax payments following communication interventions in Eswatini, Rwanda and Uganda. In a context where tax evasion is widespread, we assume this could be explained by taxpayer resistance: while communication campaigns are efficient to collect better information about taxpayers other than those directly targeted, they may be insufficiently threatening to induce directly targeted taxpayers to pay more revenue.

Second, we show that information reported by the largest firms reveal a significant amount of tax evasion. In that regard, our paper contributes to a recent literature emphasizing the central role of firms for tax collection (Slemrod and Velayudhan, 2018). Previous studies show that firms, especially the largest ones, are efficient withholding agents that tax administrations can rely on to outsource part of the costs related to tax collection (Garriga and Tortarolo, 2022, Milanez, 2017). Here we show that large firms can also be used as informants to detect large, non-compliant, actors, both at the intensive (formal) and extensive (informal) margins. This result is especially relevant to LICs where most actors operate in the informal sector and administrations can only rely on a few compliant firms to find the next margin worth expanding to.

Third and last, our results imply that strengthening tax enforcement among already tax-registered actors would help collect more revenue than enrolling new ones. Previous studies on the distribution of tax evasion have focused on tax-registered individuals (Alstadsæter, Johannesen, and Zucman, 2019, Londoño-Vélez and Ávila-Mahecha, 2021), or firms (Best, Shah, and Waseem, 2021a, and micro evidence on the informal sector in LICs exclusively derives from surveys. Thanks to our unique administrative data on both tax-registered and informal suppliers, we are able to more precisely determine whether it is strategic for revenue-maximizing administrations to widen their tax base, as can be advocated (Ohnsorge and Yu, 2022).⁹ Due to the high concentration of unreported revenue detectable, before audit, among tax-registered suppliers, our results suggest that greater gains should be expected by auditing the largest under-reporting taxpayers, at least in the short-term.

⁹In particular, since 2020, one of the core pillar of Senegal's mid-term revenue strategy for 2020-2025 has been to broaden the tax base. As emphasised in a recent IMF country report communicating about the approval of USD 1.8 billion loan extension (IMF, 2023), this objective is still given very high policy priority: "On the revenue side, the implementation of the medium-term revenue strategy should be accelerated [...] notably through the streamlining of VAT exemptions and broadening the tax base."

The rest of the paper is structured as follows: section 2.2 provides a detailed description of the context of the experiment, section 1.3 explains the design, section 1.4 presents the data, the econometric model, the balance tests, and results on take-up. Section 1.5 and C present and interpret the results on misreporting, fraud detection and remittances. Section 1.6 discusses the implications for the distribution of tax evasion and audits targeting. Section 8 concludes.

1.2. Institutional Context

Senegal is a low income country from West Africa, with stable democratic institutions, a population of 16.5 million people, and a GDP per capita of about PPP 2011 \$ 3,500. In 2019, its tax to GDP ratio was 18.4%.¹⁰, i.e. 2 p.p above the average for the 31 African countries with comparable statistics (OCDE, 2022) The TA collects about 62 % of all fiscal revenue, the rest coming from the Customs Authority and the Treasury.¹¹ The tax base is small. Overall, the TA directly manages about 30,000 taxpayers, one third of which are legal person, others being self-employed individuals. Tax registered firms are the cornerstone of tax collection: in addition to the taxes they are directly subject to (e.g the Corporate Income Tax), they also remit to the TA taxes collected through withholding applied to the payments made to employees, shareholders, administrators and informal or foreign service suppliers.¹² Directly managed taxpayers are distributed across 20 tax centres according to business size, sector, and location. Taxpayers with a turnover exceeding USD 3,200,000 for three consecutive years are managed by the Large Taxpayer Centre (LTC), irrespective of their headquarters' location. Smaller taxpayers operating within the region of Dakar are distributed across two medium-sized taxpayers centres (MTC1, MTC2), one unit for firms operating in regulated sectors (RSC) and seven municipal tax centres.¹³ The rest of the taxpayers are spread across 11 regional centres. Most importantly, tax collection is heavily concentrated:¹⁴ with approximately 900 taxpayers, the LTC alone accounts for 80% of all tax revenues collected by the TA, while the smallest 16 centres monitor more than 31,000 taxpayers but collect less than 10% of all fiscal revenues. Our intervention strictly

¹⁰Including social contributions, which are equal to 1.2 % of GDP

¹¹The treasury collects taxes on civil servants' wages, as well as local taxes, the Custom Authority collects import duties and import VAT, the TA collects all other taxes (Personal Income Tax, Corporate Income Tax, Value Added Tax etc.

¹²Using personal income tax statements, withholding data on employees, suppliers lists as well as lists of shareholders and administrators, we estimate that about 350,000 physical persons pay tax, either directly or indirectly through withholding, i.e less than 5 % of the adult population.

¹³Regulated sectors are medical and legal professions, accountants, architects, state agencies and N.G.O.. Firms within the MTC1 (MTC2) have a turnover between USD 1,600,000 and USD 3,200,000 (USD 800,000. USD 1,600,000). Sector prevails over size for fiscal assignment. Dakar's seven municipal centres are Dakar-Plateau, Ngor-Almadies, Parcelles Assainies, Dakar Liberté, Grand Dakar, Pikine-Guédiawaye and Rufisque.

¹⁴This has been shown also in other contexts like the US, or India Slemrod and Velayudhan, 2018

concentrates on taxpayers from the four most important centres: LTC, MTC1, MTC2 and RSC. By law, tax registered (henceforth *formal*) firms buying services from other economic actors must withhold a 5% presumptive tax on all transactions superior to XOF 25.000 paid to unregistered (henceforth *informal*) Senegalese service suppliers (except restaurants and hostels). The objective of this policy is to integrate into the tax net all, otherwise typically not taxed, physical persons who provide services to tax-registered firms but are not employed by the latter.¹⁵ This tax represents a little more than 1% of all taxes collected by the TA.¹⁶ This withholding tax on services (henceforth *WTS*) must be collected, declared and remitted on a monthly basis by the clients. On the monthly declaration, remitting firms indicate information on the *total* amount spent on external service provision and the *total* amount withheld to be transferred to the TA. As a complement of information, remitting firms are required to submit suppliers lists, i.e. third-party reporting statement listing *each* of their service providers – including both tax registered and unregistered service suppliers, exempt or liable to the WTS, Senegalese or foreign – with their names, ID, contact information, the amount of transaction (summed up over the year) and the WTS amount withheld.

Suppliers lists have three functions for tax enforcement. First they should allow to check whether the total WTS remitted on a monthly basis along the year corresponds to the sum of what should have been withheld from informal suppliers. Second they are used by tax inspectors to assess whether payments to service suppliers can rightfully be claimed by the clients as deductible charges. Third, and most importantly, tax inspectors also exploit it by cross checking information declared by the clients with information present in the self-declared tax statements submitted by identifiable formal suppliers. For either function to be performed, tax inspectors must be able to distinguish tax-registered actors from unregistered ones, hence the requirement for IDs.

There are two types of IDs: *Firm-IDs* and *National-IDs*. All tax-registered actors, irrespective of their legal form, have a *Firm-ID*. This ID is used by the TA as the central tax identification number. On the other side, the vast majority of unregistered actors are self-employed individuals with only a *National-ID* for numerical identification purposes (even though some may also have

¹⁵Note that, by definition, this does not prevent informal *firms* (i.e with several employees) to be liable to the WTS as long as they satisfy the above conditions

¹⁶More generally on the decomposition of tax revenue: most of what is collected by the TA derives from the value added tax (26%), the personal income tax (23%) the corporate income tax (15%) and special taxes on telecommunication and petroleum products (13%). The rest comes from capital income taxes (3.5 %), revenue stamp and registration fees (5 %), and tax on financial operations (4%). Authors calculation based administrative reports.

a *Firm-IDs*).¹⁷ As a result, tax-registered clients are expected to provide the *firm-IDs* of their tax-registered suppliers and mostly the *National-ID* (occasionally the *firm-ID*) of their informal suppliers. Overall, if suppliers lists were duly filed we would only find two types of Senegalese suppliers:

- *formal (tax-registered) suppliers*: no WTS and a *firm-ID* matching the tax registry
- *informal (non tax-registered) suppliers*: some WTS (unless exemption rules apply) and a valid *national-ID* (or a *firm-ID* matching the national business registry but not the tax registry)

Yet our inspection of the suppliers list reveals very prevalent reporting issues. Out of the 6,706 suppliers lists we collected from the largest tax centres over the 2018-2020 period, 6,170 (i.e 92%) had been misreported. Table 1.1 classifies and illustrates each type of issues, all of which results from missing information. We indicate average prevalence of misreporting issues over the 2018-2020 period (pulled and per client). Overall, well reported lines (i.e. client-supplier pairs) represent only 15.5 % of all observed lines ((a)+(d)+(f)+(g)), while 83.2% are misreported because neither the *Firm ID* nor the *National-ID* have been filled (type (c)+(e)+(h)). This provides evidence that reporting rules of suppliers lists were barely enforced before our intervention (see appendices B for a discussion on what could explain such low level of enforcement).

1.3. Experimental Design

Following hypothesis underlying classic frameworks developed by Allingham and Sandmo, 1972 to model decisions related to tax evasion, we assume that, for each supplier, tax-registered firms take a binary decision - to comply or not with reporting rules - simply by comparing compliance and non-compliance expected costs. Compliance costs typically capture everything related to bookkeeping past transactions, collecting reliable and up-to-date identification and contact information from suppliers, staying up-to-date with reporting rules, and eventually filling suppliers lists. To the extent that misreporting suppliers lists can also be a strategy for

¹⁷This comes from the fact that *Firm-IDs* are not produced by the TA it-self but by the National Agency for Statistics and Demographics (NASD) in charge of maintaining the firms and associations national registry. As such, *Firm-IDs* are also used by the Immigration Services, the Custom, Banks, the Treasury or local governments. In such instances, economic actors can derive short term benefits from the interacting with the institution (namely a visa, an access to foreign markets, a loan or a public contract) hence prompting some of them to ask the NASD a *firm-IDs*. On the other hand, they probably expect only disadvantages in the short run from interacting with the TA. In fact more 90% of the firms and associations registry entity are not tax-registered

tax-registered firms to hide transactions with informal suppliers, compliance cost also include losses from reduced opportunities to evade the WTS. On the other hand, non-compliance cost will incorporate all penalties prescribed by law for misreporting suppliers lists and, possibly, evading the WTS, weighted by the perceived probability for these sanctions to materialise. To account for the fact taxpayers can be ignorant about fiscal and reporting rules, we further assume that the taxpayers decision's is based on taxpayers believes about compliance and non-compliance costs, may the believes be true or not.

The intervention we test essentially consists of a communication campaign spanning over a year from the TA to taxpayers that we identified as having recently submitted incomplete suppliers lists. Building from the communication experiment literature which emphasises the prominent role of deterrence ([Hallsworth, 2014](#), [Slemrod, 2019](#), [Antinyan and Asatryan, 2019](#)), this campaign was designed with the primary objective to induce firms to submit more complete information in the future and explicitly conveys the following messages:

1. The TA has checked recent suppliers list, detected misreporting issues and demands rectifications of the incomplete reports within 20 days, otherwise the taxpayers may be subject to sanctions as set by the fiscal code¹⁸ The TA invites the taxpayers to submit complete suppliers list in the future.
2. Fiscal rules related to the withholding tax on services suppliers and declarative obligations in suppliers list are recalled, citing precise elements of the fiscal code.

Given our above assumptions on taxpayers reporting choice, we expect our intervention to increase taxpayers probability to comply to the message's instructions via two mechanisms. The first message should on average increase the perceived probability of sanctions for future declarations: it demonstrates the administration's ability to detect misreporting issues in the past and makes explicit warning for future declarative behavior. According the law, taxpayers may be subject to a fine equal to 200,000 XOF (USD 320) for not submitting suppliers list, or submitting incorrect information. This is arguably small but there could also be risks (real or perceived) associated with having being detected as not compliant. The second part of the email may improve taxpayers' knowledge and understanding of reporting obligations. In this regard, following [De Neve et al., 2021](#) on the role of simplification for tax compliance, we endeavoured to express in the clearest possible way, what taxpayers would have to do to comply to suppliers list's reporting rules. Overall, a taxpayers perfectly complying to the message would: (a) submit a new, corrected, suppliers list for each year pinned down in our message; (b) submit future suppliers list free of misreporting issues.

¹⁸The years for which misreporting has been detected are individualised.

Figure 1.1 displays the general timeline of the intervention, together with the tax calendar for the relevant ex-post tax declarations. The message was first sent by email during January 2022, right before 2021 suppliers list were normally due. We used automated direct mail technology to send individualised emails by batch. Follow-up calls were systematically made to each taxpayers to verify that the email had been received, read and understood. In May-June 2022 additional calls were made to taxpayers who had not yet submitted rectifications - in accordance with what the TA would typically do in such situation. In July, another email was sent to the entire treatment group simply reminding taxpayers about all declarative obligations related to the suppliers lists and inviting them to duly comply in the future (but without referring again to past non-compliance). A last email was sent in December 2022 similar to that of July but with two additions : a final reminder to those who had still not submitted any rectifications and a paragraph recalling the necessity for yearly suppliers lists to be consistent with (the sum of) monthly declarations and remittances of the WTS.

Within the time frame of the intervention, no sanctions were applied for not submitting rectification. Relative to pure communication intervention, applying systematic sanctions would have added significant burden on the TA. The administration preferred to not use sanction, arguing that checking whether future suppliers lists were better filled was a more strategic use of its resources than spending time getting access to improved versions of past declarations. Considering this, the objective of the individualised reminders we sent in May-June 2022, and then in December 2022 to the gradually decreasing subset of firms which had not yet submitted any rectifications was to limit as much, and as long, as possible the general sentiment in those recalcitrant taxpayers that threats were not credible.

We did not send any specific message to the control group. Front office staff did not know about the existence of the control group and taxpayers could not know they were part of an experiment. Each communication has been done through communication canals normally used by the TA, in strict accordance with what is allowed by the tax code. As such, the TA has full ownership over all tools used for the implementation of the experiment, would it wish to replicate it or scale it up.

1.4. Data, Estimation and Baseline Statistics

1.4.1. Data

This project would not have been possible without extensive collaboration with the tax administration (TA) to collect, understand and assemble vast amount of administrative records (Czajka

et al., 2022). We produced four databases, three of which result from the consolidation of administrative sources. Our first database gathers suppliers lists submitted by clients from the largest tax centres. We digitise all suppliers lists submitted on paper since 2018 by taxpayers registered in the four largest tax centres, and complete this with the universe of suppliers lists declared online, via a platform we contributed to design. We use this at baseline for sample selection (Section 1.4.3) and to measure our outcomes of interest when estimating the impact on misreporting and fraud detection (Section 1.5). Our second database consists of all monthly declarations submitted for the Withholding Tax on informal Service suppliers (WTS) together with detailed information recording actual remittances of the WTS to the TA (all of which are stored into TA's central database). We use this source to analyse the impact on withholding (see Section C). We build a third database by extracting, for all 20 tax centres, all five types of declarations where total gross revenue are reported: value added tax (monthly), corporate income tax, personal income tax, business property tax and the synthetic tax for small business. This is the data we exploit to measure self-declared revenue by tax-registered suppliers. Confronting this source to information from the suppliers lists mentioned above allows to measure the impact on fraud detection (Section 1.5.2) and its implication for tax evasion (Section 1.6). Finally, we produce ourselves information based on immediate response from taxpayers targeted by the intervention. That last data simply allows to document in detail the effective reception of the message among the treated (Section 1.4.4, Table 1.7).

1.4.2. Estimation

Our main model estimates the average treatment impact at the taxpayer (client) level i , the statistical unit with respect to which randomisation was executed. $Treat_i$ is the treatment dummy, δ_s are fixed effect for each strata s used for the randomisation (Section A), and $\mathbf{X}_i$ is a vector of potential additional control variables. The outcomes y_i may take various forms, e.g. the total number of suppliers reported by client firm i , the probability to perfectly comply to reporting rules, or the share of foreign suppliers, or the total amount to formal suppliers.

$$y_i = \alpha + \beta Treat_i + \gamma \mathbf{X}_i + \delta_s + \epsilon_i \quad (1.1)$$

Note that we cannot estimate models at the supplier level as long as they would require using information from follow-up suppliers lists that may have been influenced by the intervention. In particular, as our intervention increases the prevalence of IDs (Section 1.5.1), ex-post identifiable suppliers of the treatment group are very likely to be different from suppliers identifiable based on information provided by the clients in the control group. Therefore, in the absence of

alternative sources to identify suppliers ex-post, any estimation at the supplier level comparing ex-post (sub-)samples of suppliers of each intervention groups is likely to suffer from selection bias. This implies for instance that the data does not allow to test whether some types of client-supplier pairs at baseline are more or less likely to be present after our intervention. Nevertheless we can explore impact, at the suppliers level, when restricting to the sample of suppliers that could be identified ex-ante, as long as we do not use ex-post suppliers list to measure the outcomes of interest. We do so to test whether suppliers of the treatment group tend to self-declare higher income now they their clients more thoroughly report their spending to the administration (see Appendix D)

1.4.3. Selection and Baseline Statistics

Based on our analysis of the 6,706 suppliers lists submitted by taxpayers from the four largest tax centres over the period 2018-2020, we select all firms who have misreported at least once during those years, as per the definition of misreporting provided in table 1.1. Our sample consists of 3,487 taxpayers; 799 from the Large Taxpayers Centre (LTC), 439 from the Middle-sized taxpayers Centre 1 (MTC1), 1000 from the Middle-sized Taxpayers Centre 2 (MTC2) and 1249 from the Regulated Sectors Centre (RSC), i.e 91.4%, 77.4%, 88.6% and 91.4% of each centre respectively (Table 1.3). On average we observe 1.7 suppliers lists per selected taxpayers at baseline over 2018-2020. We randomly allocate half of the sample to a treatment group by stratifying according to tax centre, the number of yearly suppliers (split in 4 groups) and the share of misreported suppliers (split in 2 groups). This leads to 32 strata with the least represented cell accounting for 14 taxpayers (see appendix A for more details on sample selection and randomisation).

Table 1.6 displays summary statistics for key covariates and outcomes, together with β coefficients and standard errors of balance tests. With the exception of the number of employees, treatment and control group do not differ significantly from one another. Overall, 70 % of the sample are registered as legal persons, the majority has been interacting with the TA for 13 years or more, 32.6% had started declaring suppliers list online before the intervention and we observe at least 2 suppliers lists at baseline for the 2018-2020 period for more than half of the sample. It is a very heterogeneous sample in economic terms, which reflects vast disparities across taxpayers in Senegal (Czajka and Sarr, 2023). The average turnover is USD 5.48 million, with a standard deviation of USD 30.678 million and a median at USD 407,000. Some firms have no formally registered employees, while a few others have more than 100,000. The average number of suppliers is about 40 but standard deviation is 152 and the median is 10. However the sample is more homogeneous in terms of misreporting patterns. On average

about about 86% of lines from suppliers list are misreported, 98.9% of suppliers are Senegalese.

1.4.4. Assessing the take-up

In this section, we discuss results from the descriptive analysis of direct responses from the treatment group to assess to what extent treated firms have actually received and understood the message conveyed by the communication campaign. Some firms sent an acknowledgement of receipt right away, but most did not. We used email tracking technology to check whether emails sent in bulk had been received and eventually opened. Still emails could be left unread, misunderstood or land in the mail box of somebody who is not anymore working at the targeted firm. To avoid such situations, after each email campaign, we also systematically called targeted taxpayers and ran quick interviews to verify whether the email had been read and understood. Still, some firms did not answer, or never was the right person available. Next to this, we also collected meta-data on rectified suppliers list submitted after the 2022 intervention for years 2018, 2019 and/or 2020 - evidence which strongly suggest that not only the message had been received but also that firms were willing to comply. Combining all these elements, table 1.7 decomposes immediate response rates overall and by centre. Overall, 60.8 % of all treated taxpayers did submit a rectified declaration. About 32% did not, but we could still confirm the message had been well received. Interestingly, when last asked why they had still not submitted any rectifications: 16.4% claimed they needed more time, 7.1 % explained they could not retrieve the requested information and 3.5 % asked for assistance on the online platform. Eventually, we have no proof of reception for 125 taxpayers, i.e 7.1% of the treatment group.

1.5. Results on Misreporting and Fraud Detection

1.5.1. Misreporting of suppliers list significantly decreases

In this subsection, we estimate and interpret the intervention's impact on reporting behavior for 2022 suppliers lists, which were due in January 2023, i.e after the full communication campaign.¹⁹

The intervention was very successful in deterring misreporting in follow-up suppliers list (Table

¹⁹The measurement of most outcomes related to the presence of information in suppliers list requires that taxpayers have declared the list. 22% of our entire sample did not submit any suppliers lists for 2022 but the difference in attrition rates across groups is not statistically significant (see table 1.8). In this section, we therefore restrict our sample to taxpayers having declared a 2022 suppliers lists. Table 1.913 reproduces descriptive statistics balance tests conditional on having declared a 2022 suppliers.

1.8, Panel (A)). The share of misreported suppliers decreased by 23.3 percentage points, i.e. 42% compared to the control group's average. The proportion of taxpayers fully complying with reporting rules increased by 10.1 percentage points, while that of taxpayers misreporting all their suppliers decreased by 19.5 percentage points. This impact is essentially driven by the large increase in the prevalence of ID. Indeed, the share of suppliers with a formally valid ID increased by 24.7 percentage points on average (+52%). Importantly, the intervention had no significant impact on the prevalence of intentionally misreported ID ²⁰, so the intervention not only increased the *quantity* but also the average *quality* of the available information. Interestingly, we also estimate a strong impact on the presence of suppliers phone number (+9.2 p.p, +22%), despite this not having being mentioned during the campaign, nor legally mandatory. Last, suppliers of treated firms are not more likely to be eligible of WTS exemptions. Altogether, these results strongly suggests that treated firms did attempt to comply with the law and collaborate with the administration without circumventing the rules. Interestingly, these positive effect on reporting are more pronounced for larger firms (Figure 1.2). This is robust to any measure of firm size (turnover, number of employees, number of suppliers and number of years since creation).

Yet, the intervention was not sufficient to induce perfect reporting compliance. Indeed, results from Table 1.8, Panel (A) imply that about 9% of the treated taxpayers improved only part of their declaration. As a result, 71% of the treated firms still misreport at least one supplier ex-post. This is consistent with the possibility for compliance cost to depend on suppliers willingness to share information with their client. This interpretation is also supported by direct feedback received from treated firms complaining about the difficulty to collect the information from reluctant suppliers (see table 1.7).

Treated firms collected and reported more information about their suppliers, but a subsequent question is whether this occurred simultaneously with a change in the suppliers network. The first email was set in early 2022, but 2022 suppliers list are due by January 2023, hence leaving most treated firms with almost a year to possibly adapt their trading relations to facilitate reporting compliance. Our intervention could lead to real changes in the network for essentially two reasons. First, firms trading with informal suppliers without complying to the WTS rules may choose to trade less with informal suppliers and more with formal ones, which are exempt from the WTS. This could not only affect the composition of suppliers but also their number, e.g a firm could replace several individually contracted informal security guards by a single formal firm specialised in security services. Second, firms could modify their trading partner to better comply with the information requirement, looking for suppliers that agree to share a

²⁰An ID is considered as intentionally misreported if its structure is obviously wrong (ex : 999999999) or if it is provided for many different suppliers by the same clients for the same year

correct numerical ID. Obviously, several other factors will affect this decision such as search and matching costs, or differences in prices, quality and risk across suppliers type (formal/informal). In some case, a substitution may be hard to operate, at least in the short run, and the clients may as well decide to end contract without replacement.

Unfortunately we cannot disentangle real changes in suppliers network from changes in reporting behavior whenever the measurement of those changes requires the information affected by improvement in declarative behaviors (Section 1.4.2). Nevertheless, we can explore how the intervention affected some other dimensions of the suppliers network. Estimates from table 1.8 show that the communication campaign had no significant impact on the overall structure of the service supply. The average number of suppliers remained unchanged, so did the total, the average, and the standard deviation of the payments. The median may have slightly decreased but this coefficient is only marginally significant. We still cannot discard that some real changes in the network may have occurred during 2022, but given our results on the distribution of suppliers network, this would have had to taken place mostly by replacing a supplier by another, for a transaction of similar in size. Contrasting this with the widespread changes in information availability documented above, it seems reasonable to assume that, even in 2022, the impact we observe essentially reflect improvements in information sharing rather than real changes in suppliers composition.

Our interpretation is further reinforced by results obtained on 2021 suppliers list. Result on these lists - which had to be declared by the 31st of January 2022 - can only be due to improvement in reporting behavior, because the treated firms received the first email after December 2021, thus after all reported transactions had already been completed, leaving not time to introduce real changes in the service supply structure. However our intervention increased the probability to submit 2021 suppliers lists, thus potentially introducing selection bias due to attrition in the control group. To account for this, we follow Lee et al., 2009 and estimate treatment-effect bounds for non random sample selection. As shown in the comparison table 1.915, results are very consistent across wave, both in term of sign and significance level. They differ only in magnitude, with average impact slightly larger in the second wave.

1.5.2. New information increases the detection of tax evading suppliers

In the short run, access to new information on suppliers may be valuable to tax revenue-maximising administration only if it allows to better identify non-compliant suppliers. Yet this needs not to be the case. Indeed, the additional information could only help tracking transactions going to firms that already duly declare all their revenue to the TA. In this section,

we show that the newly available information helps uncover tax evading suppliers, both at the extensive and intensive margin.

First, we distinguish tax-registered suppliers from informal ones, among all those for which we have a formally valid ID. We identify formal suppliers either directly - when their Firm-ID matches the tax registry - or indirectly - in case their national-ID matches the national business registry where it can then be linked to a Firm-ID that in turn matches the tax registry. We consider informal all suppliers with a formally valid national-ID²¹, or with a Firm-ID that matches the national business registry but not the tax registry. Using these definitions, the intervention increases the share of suppliers identified as formal (informal) by 6.9 (17.5) percentage points, which translates into a 30% (76%) augmentation (Table 1.8, Panel (C)).

To the extent that informal suppliers are not tax-registered, we consider them as tax evaders, irrespective of whether their client withheld the 5% tax or not (see Appendix C for a discussion on the detection of under-withholding clients). Formal suppliers, on the other hand, may be perfectly compliant. To verify this, we confront the sum of payments they received - according to their clients' suppliers list - to the total revenue they have indicated for the same period in their self-declared annual income tax. More specifically, we define a binary outcome classifying each transaction t^T (resp. t^C) paid by a client from the treatment (resp. control) group as going to an *under-reporting* formal suppliers S if and only if the sum of all transactions t^T (resp. t^C) going to S is superior or equal to the total revenue self-declared by S in its annual income tax statement (see Table 1.2 for detailed examples). Under this definition, the same supplier may therefore be considered as under-reporting according to one group but not the other. This definition prevents capturing externalities outside of the treatment group due to network effects. Most importantly, it better reflects what to expect if we were to scale up our intervention to all firms. With this definition at hand, we estimate that the intervention increases the share of payments going to under-reporting formal suppliers by 2.5 percentage points on average, i.e. 45% (Table 1.8, Panel (C)).

1.6. Implication for the distribution of tax avoidance and the targeting of future deterrent measures

We now want to explore the implication regarding tax avoiding suppliers. Results from section 1.5 and C aggregate client-supplier pair data at the *client* level to make statistical inference on

²¹To be considered formally valid, national-IDs must respect a strict structure. As the suppliers' age can be inferred from the national-IDs, we also invalidate national-IDs implying age below 15.

average treatment impact. But, as suppliers may have several clients, within and/or across experimental groups, this perspective is not sufficient to discuss implications at the *supplier* level. In this section, we thus turn to the descriptive analysis of the ex-post distributions of suppliers level data.

1.6.1. How many and how large are newly identified suppliers?

Detecting tax evasion is only valuable to tax collection if newly detected non-compliant suppliers (formal or informal) are not too small and thus worth targeting with follow-up deterrence interventions. To assess fraud detection's fiscal potential, we therefore need to estimate its implication for the ex-post distribution of tax evading suppliers. By definition, suppliers identified as informal are not tax-registered. From the perspective of the TA, they represent tax evasion at the extensive margin. We hence classify all of them as tax evaders. On the other hand, formal firms are tax-registered and classified as non-compliant if and only if the sum of payments they received according to information in their clients' suppliers list is larger than what they have claimed in their self-declared income tax statement. As in section 1.5.2, we define tax evasion *relative* to each experimental group: a formal supplier can thus be under-reporting according to one group and not the other (see Table 1.2 for detailed examples).

Table 1.9 compares the population of suppliers identified as tax evaders based on information coming from the control group or the treatment group. Information retrieved from the treatment group 2022 suppliers list allows to identify 29,317 suppliers, i.e 2.4 times more than for the control group. The sum of all payments declared by clients of the treatment group and going to identified suppliers (either formal or informal) equals to USD 1.56 billion, i.e 5.63 % of Senegal's GDP for that year (Table 1.9, column (4) and (5)). Comparing this to the control's group equivalent figure, we estimate that our intervention allowed to identify more than USD 1 billion worth of transaction, i.e 3.8 % of GDP.

Further splitting suppliers, we observe that, information coming from the treatment group flags 2,795 under-reporting formal suppliers and 23,141 informal ones (against 1,742 and 7,750 respectively for the control group). Measured this way, i.e. in terms of tax evaders' population size, fraud detection increased by 60% at the intensive margin and by almost 200% at the extensive margin. Under-reporting formal suppliers of the treatment group received USD 118,823 on average in 2022 from their (treated) clients, compared to USD 61,951 for the control group. This implies that formal tax evaders detected thanks to the treatment group are not only more numerous but also larger than those detected thanks to information provided by the control group. However, at the extensive margin, newly identified informal suppliers decreased

average amount received by informal suppliers by 40%, suggesting that the informational gains induced by the intervention shed more light on the bottom of the distribution comparatively. Bringing all these figures together, we establish that information from the treatment group helps revealing USD 145.5 million in additional unreported revenue, i.e. 0.5 % of GDP

To better assess the distribution of these informational gain, we compute the number of tax evading suppliers identified by the treatment group, the control group or both, by decile of *unreported amounts* (Figure 1.3). For informal suppliers, all payments received that we detect go unreported by definition. For formal but under-reporting suppliers we define unreported amounts relative to treatment (control) as the sum of payments received from clients of the treatment (control) group, minus total revenue declared to the TA. Interestingly, information from treated clients allows to identify more tax evading suppliers across all level of unreported income, both at the intensive and extensive margin, even though the difference in detection rate tends to decrease with unreported income.

1.6.2. What do we learn about the distribution of tax evasion?

On average, more than 50% of the suppliers of the control group have no valid ID (see Table 1.8 and Figure 1.4.A), therefore preventing any match with the tax registry, which in turn implies that we cannot determine whether they are formal or informal, tax-avoiding or compliant suppliers. With so much information missing, any attempt to describe the distribution of tax evasion is likely to suffer from sampling biases. But because treated firms did submit much more complete reports about their suppliers (see Figure 1.4.A and 1.4.B for a graphical illustration), restricting our analysis to suppliers of the treatment group allows to more confidently explore this question. We document three stylised facts about tax evasion which the prevalence of unidentified suppliers was previously hiding (Tables 1.9 and 1.10).

First, the average number of clients of tax evading suppliers is rarely superior to 2, which strongly suggests that detected under-reporting is only a fraction of real total income, despite the prevalence and magnitude of tax evasion already observed here. Second, conditional on being identified as formal, under-reporting is highly prevalent: 45% of the treatment group's formal suppliers have been detected as under-reporting. They represent a bit less than 10% of all tax-registered actors identified as active in 2022 (see Table 1.11). Third, the observed distribution of unreported revenue by tax evading suppliers is extremely skewed. Most are very small: 90 % have a yearly observable net unreported income lower than USD 11,000. But unreported amount among the largest 1% tax evaders reaches USD 639,570 on average. Overall, net under-reported amounts among formal suppliers are significantly larger than

amounts received by informal ones. Below the 96th percentile (USD 31,798) tax evading suppliers are mostly informal, while above, tax evading suppliers are mostly formal. Under-reporting formal suppliers within the top 1% of the distribution of tax evaders account for 52% of all unreported income.

Considering this and the fact that auditing already tax-registered suppliers is likely to be much less costly than to go after fully non-compliant actors, our results suggest that, in the short run at least, an administration maximizing fiscal revenue should target under-reporting formal firms rather than informal ones, at least above a certain threshold. It is worth stressing however that, in the longer run, this strategy will reinforce inequities between those who contribute and those who do not. To avoid this, we believe tax administrations facing large informal sectors should always invest resources also to enrol new economic actors even if they do not have the resources to go after *all* informal actors. Our results speak to this issue as information in supplier lists allow to identify large informal suppliers worth registering.

1.6.3. How much additional tax revenue can be expected from exploiting the new information to target future audits?

We run a simulation exercise to estimate the expected *additional* tax revenue collected would the administration exploit the new information retrieved from suppliers lists submitted by the treatment group. To do so, we compare tax revenue expected when using this information to select firms to be audited, with tax revenue expected based on alternative selection method that the TA could use. Following the conclusion of Section 1.6.2, we concentrate exclusively on the formal sector, and define our benchmark selection method (S_0) as the audit of the largest under-reporting formal suppliers according to information reported by the treatment group. Ideally, we would like to compare tax revenue collected with our benchmark selection method to those collected when using the method that the TA currently applies to select taxpayers to be audited. Current audit selection procedure is very decentralised. Executed at the tax centre level mostly, it mixes discretionary decisions made by individual tax inspectors based on elements of suspicion, with more rule-based decisions validated by a college of higher ranked administrators (Bachas et al., 2021, Czajka et al., 2022).²² Given this, we decide to confront S_0 to four, very different, alternatives:

S_0 ; audit the largest under-reporting formal suppliers according to information reported

²²Exhaustive and reliable data and information on audit selection and execution is not systematically collected and consolidated into standardised format. Our description here derives from qualitative and observational evidence.

by the *treatment* group (benchmark selection method)

S_1 : audits of the largest under-reporting formal suppliers according to information reported by the *control* group

S_2 : audits of randomly selected taxpayers

S_3 : audits of the largest taxpayers (based on declared income)

S_4 : audits of the largest taxpayers, by centre (based on declared income)²³

Let n be the number of firms to be audited and r the effective audit rate of return, i.e tax revenue collected upon audit as a percentage of unreported amounts. For each selection procedure S_k (with $k \in [0,4]$) listed above, we define $T_k^{nr} = \sum_{i=1}^n (r \cdot u_i^k)$, where $(u_i^k)_{0 \leq i \leq n}$ are unreported incomes of the n selected taxpayers. T_k^{nr} represents the total expected tax revenue collected by auditing the n taxpayers selected through the procedure S_k , assuming an effective audit rate of return of r . For some n , r and selection S_k (with $k \in [1,4]$), we are interested in estimating $\Delta_k^{nr} = (T_0^{nr} - T_k^{nr})/T_k^{nr}$. For simplicity, we can assume that r does not vary with selection method, hence it can be removed from the equation which thus simplifies into $\Delta_k^n = (U_0^n - U_k^n)/U_k^n$ where, for any $k \in [0,4]$, $U_k^n = \sum_{i=1}^n u_i^k$, i.e the sum of all revenue unreported by the n selected taxpayers.

For any k, i , we have: $u_i^k = c_i^k + a_i^k$ where c_i^k , is unreported income revealed by cross-checking suppliers list against self-declared income tax *before audit*, and a_i^k is unreported income revealed by the audit (via all method applicable during an audit). By definition, $c_i^0 > 0$ ($c_i^1 > 0$) for any taxpayer i who has been identified as under-reporting payments received from the treatment (control) group. But, when $k \in [2,4]$, $c_i^k = 0$, and thus $u_i^k = a_i^k$. In other words, in selection methods 2, 3, 4, the only unreported income detected is what is uncovered by the audit. To estimate U_k^n for all k , we thus need to estimate a_i^k for each taxpayer i . Unfortunately, information about a_i^k is not collected by the administration. To approximate a_i^k , we thus exploit amounts unreported by the 2,792 tax evading formal suppliers of the treatment group.²⁴ For each of them, we compute an under-reporting rate by taking the ratio between unreported income detected thanks to information declared by the treatment group, and total *real* revenue (the sum of both declared and unreported revenue). Building from [Best, Shah, and Waseem, 2021a](#), who show that in Pakistan firm size is by far the best predictor of audit-based evasion rates, we then rank the full population of taxpayers according to their declared revenue (including those 2,792 known tax evaders) and compute average under-reporting rates by decile. Column (5) of Table [1.11](#) displays average under-reporting rates *conditional on under-reporting*. As audits do not

²³This is the same as S_3 but selecting the same proportion of taxpayers from each tax centre

²⁴As further discussed below, we discard the largest 1% firms from our analysis, including the 3 under-reporting firms detected in that group.

systematically detect tax evasion,²⁵ extrapolating observed conditional under-reporting rates to all taxpayers to estimate unreported income revealed by the audits, would greatly overestimate evasion.²⁶ Instead, we use *unconditional* average under-reporting rates by decile of declared income. Before computing these rates, we discard the largest 1% firms of the sample, may they be under-reporting or not. Indeed, we detect only 3 under-reporting taxpayers among the largest 1% taxpayers, while these account for almost 75% of all (declared) revenue (Figure 1.6). Considering this, it seems reasonable to assume that the population of detected under-reporting suppliers provides too little information about this extremely different sub-group to allow any extrapolation, as a result, we prefer to discard them from the analysis. Finally, we obtain the coefficients displayed in column (6) of Table 1.11 and apply these to taxpayers' declared revenue to estimate a_i^k , and therefore u_i^k for each taxpayer i and scenario $k \in [0,4]$. This way we can compute Δ_k^n in all scenarios. We bootstrap the estimation procedure to account for potential measurement errors. Overall, the validity of our estimation will rely on the following two assumptions:

1. The audit rate of return is constant across all scenarios.
2. Unconditional under-reporting rates based on cross-checking are at least as high as unconditional under-reporting rates based on audits.

Hypothesis 1 is at worst harmless, at best conservative. Indeed, it seems plausible that, conditional on some unreported revenue detected, the selection method would not impact the share of unreported revenue recovered by the administration. Also, in case third-party evidence is perceived by under-reporting taxpayers as a sign that the tax administration is more competent than they thought it was (as evidenced by [Brockmeyer and Hernandez, 2022](#)), then they might be more likely to collaborate, possibly giving away some additional information that would increase the recovery rate (e.g. the bank account number). Relaxing hypothesis 1 to allow higher r in our benchmark scenario would lead to even more significant results.

Assessing the plausibility of Hypothesis 2 requires making sure that we do not underestimate a_i^k . As explained above, the TA in Senegal does not collect information on such parameters. Yet, there exists one study that measures unbiased tax evasion rate discovered upon audits in a LIC. Comparing the unconditional under-reporting rates we use in our estimation, to those found by [Best, Shah, and Waseem, 2021a](#), our estimates clearly stand out as upper bounds. Indeed, above

²⁵This can happen not only in cases where taxpayers do not under-report, but also if the audit procedure is not able to uncover some forms of evasion.

²⁶Indeed, we estimate that assuming audits would systematically reveal evasion rates as large as conditional under-reporting rates uncovered through cross-checking, would make aggregate unreported income almost ten times greater than Senegal's GDP

the median declared revenue, the average unconditional under-reporting rates they find are equal, or inferior, to 1.75, while ours are systematically above 2.5. Importantly also, under our assumption, the implied total estimated unreported income for the entire sample of taxpayers is USD277 millions, i.e 51% more than total unreported income detected through cross-checking information declared by the treatment group (see Table 1.9, column (7)).

Figure 1.11 displays Δ_k^n for any $k \in [1,4]$ and $n \in [100,1000]$. Selection method S_0 outperforms method S_1 by at least 80%, and up to 115% if we consider the top 100 tax evaders. It outperforms method S_2 by at least 1000%, which derives from the fact that S_2 selects too many small taxpayers. Our benchmark method also does better than method S_3 which is, to our knowledge, the closest to what the TA actually does. Finally S_0 even outperforms selection method S_4 for the top 350-400 under-reporting formal firms - beyond, under-reporting levels based on suppliers list decrease faster than estimated unreported income for the next largest firms. Overall, these results strongly suggest that newly available information would increase short-term audit returns by improving the targeting of firms to be audited.

Eventually, one could argue that if audits do not yield any return, then a relative increase of 100%, will still produce null results. This however is very unlikely. Every year, the tax administration carries out audits based on similar cross-checking exercise (although not automatised) and recovers some positive total, part of which is going to tax inspectors themselves as a form of bonus (Bachas et al., 2021). These returns might be small. Best, Shah, and Waseem, 2021a find for instance that, in Pakistan, only 2% of the uncovered evasion gets recovered upon audit. Nevertheless, it is very unlikely that, despite proof of under-reporting, results of audits based on our information would *always* be null. What is more subject to debate in the literature is the quantification of the long term impact of audits. Indeed, Kasper and Rablen, 2023 highlight that, because of the combination of learning and "crater" effects, audits can in theory have both negative or positive impact on compliance. Best, Shah, and Waseem, 2021a find that random audit have no effect in the long-run, while Kotsogiannis et al., 2024 find *some* types of audits can increase compliance the following three years.

1.6.4. Cost-Benefit Analysis for the Tax Administration

In this last section, we estimate the cost-benefit ratio of implementing our intervention and auditing the largest under-reporting formal firms detected thanks to information declared by the treatment group in their 2022 suppliers list. The total intervention costs about USD 102,720,²⁷.

²⁷This includes USD 30,720 for two years of data collection (although suppliers list are now declared online, making this amount unnecessary in the future) USD 33,600 for the one-year communication campaign, and USD 38,400 for all coordination and data analysis related costs.

Tax inspectors get paid an average salary of XOF 668,085 (USD 1,068) per month and, based on administrative reports, we estimate it takes them about 1.2 months to complete an audit.²⁸ For each $r \in [3, 30]$ and $n \in [100, 1000]$, we compute T_0^{nr} (see Figure 1.12.A) and T_0^{nr} / C_0^n (see Figure 1.12.B), where C_0^n stands for the full cost of running our intervention plus auditing the n largest under-reporting formal firms according to the treatment group.

Unfortunately, the TA does not systematically collect data on r . By law, unreported revenue should be considered as net profits and thus subject to the corporate income tax, whose statutory rate is 30%. In addition, the detection of unreported income may lead to the upward revision of the tax base for other fiscal instruments,²⁹ and thus further increase total liability. Last, financial sanctions for having committing tax fraud may also be added. Yet, during the audit process, taxpayers are often given the opportunity to revise upward deductible charges, therefore lowering net profit. Eventually, the effective r will depend on several factors that we cannot observe. To account for this uncertainty, we vary our estimates for any $r \in [3, 30]$. The 30% upper bound corresponds to a scenario where no taxpayer manages to claim new deductible charges, but also where no fine would be applied, nor any adjustment of other tax liabilities. We take 3% as a lower bound as it corresponds to the average effective tax rate (considering all taxes) with respect to total gross revenue for the full taxpayer population.

Final estimates are very sensitive to the input parameters. The nominal amount of total tax collected upon audit varies from USD 5.8 million to USD 52.2 million (see Figure 1.12.A), and the benefit-to-cost ratio goes from 9.8 to 175. (see Figure 1.12.B). Importantly also, due to the extreme skewness of the distribution of unreported amounts (see Table 1.10), benefit-to-cost ratio sharply decreases with the number of audited firms. Nevertheless, our simulation strongly suggests that, even under very conservative assumptions, implementing the intervention followed by the audits is highly cost-effective.

1.7. Conclusion

In the context of Senegal, this paper provides evidence that leveraging large firms' trading network to collect information about their suppliers can be a cost-efficient way for tax administrations in LICs to detect tax evasion and increase future audit returns. We first digitised and consolidated 6,706 lists of suppliers filled by the largest firms of the country. We documented that misreporting is highly prevalent, which prevents any systematic use of such suppliers

²⁸On average a tax inspector does about 10 audits per year. In practice audit run over a long period of time (between 3 months to a year) and inspector always work simultaneously on several cases at the time. All these figures are exclusively derived from reports produced by the Large Taxpayers Centre.

²⁹In particular: VAT, local business tax (which is based on value added), excise duties and other special taxes.

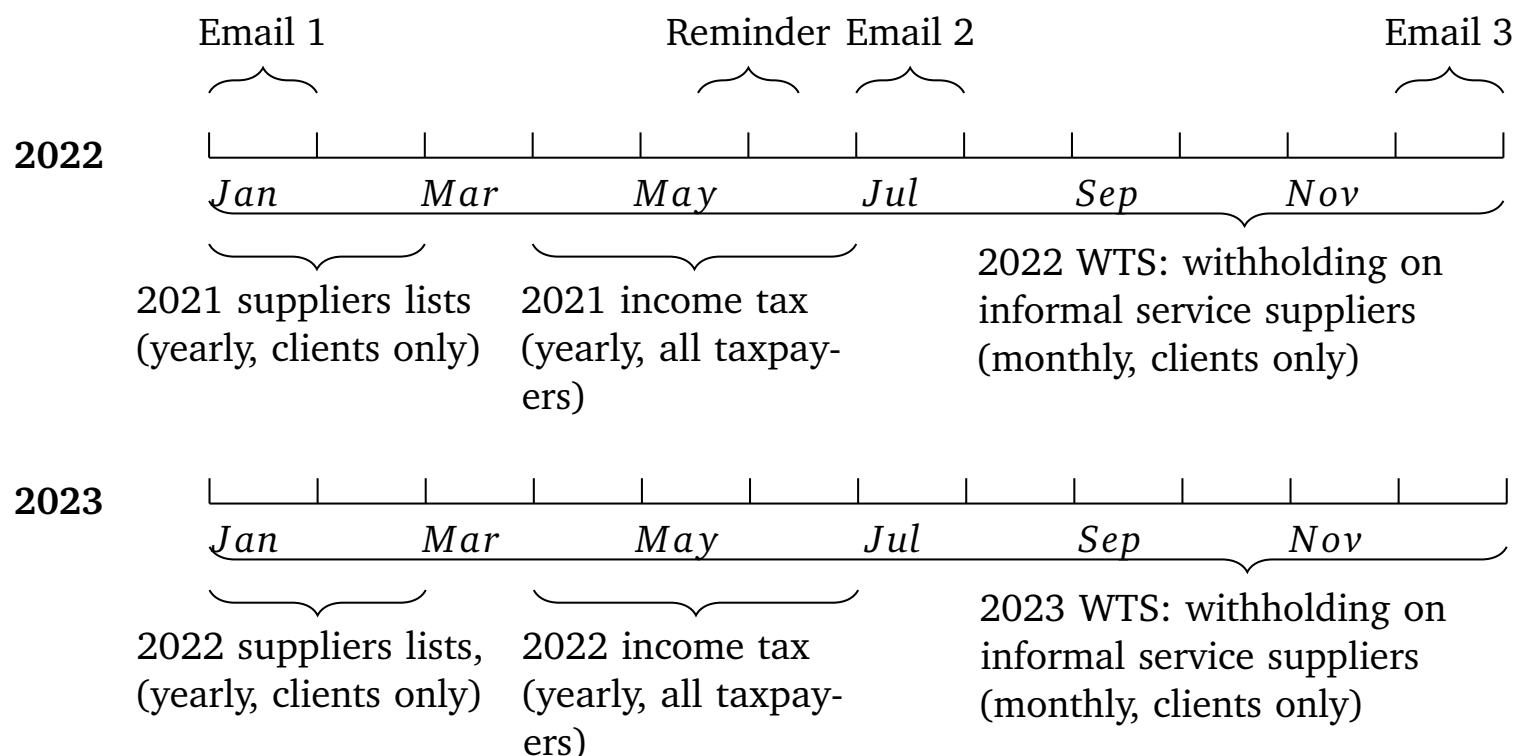
lists for audit related tasks. Building on the communication experiment literature, we then randomised a low-cost individualised communication campaign across all 3,487 firms that we identified as misreporting, to discourage future misreporting. Our intervention dramatically increases the quantity of information available in follow-up suppliers lists, without reinforcing strategically manipulative declarative behaviors. Even though the message did not specifically alert about under-withholding from informal suppliers, it also had a sufficiently credible deterrent impact to increase remittances among the taxpayers who were the most likely to have evaded the withholding tax ex-ante. The newly available information allows to better track USD 1 billions worth of transactions, i.e. 3.8% of GDP. By cross-checking these transactions with the universe of self-declared income tax statements, we show that intervention contributes to uncover 145.5 million in unreported revenue, i.e. 0.5% of GDP. A close inspection of the distribution of unreported revenue reveals extreme concentration of tax evasion, mostly accruing the largest under-reporting formal firms. A simulation exercise shows that by strategically allocating audit resources based on the newly available information, tax revenue collected upon audit could be increased by at least 100%.

This study illustrates that well targeted communication campaigns are highly cost efficient tools to induce better reporting even among the largest taxpayers, despite being relatively mild compare to alternative costlier, deterrent methods. Our results also show that large firms can be used as informants to detect large, non-compliant, actors, both at the intensive (formal) and extensive (informal) margins. This result has wide ranging policy implications in LICs where most actors operate in the informal sector and where administrations can only rely on a few compliant firms to find the next margin worth expanding to. In the case of Senegal, our estimation suggests that, at least in the short term, greater gains should be expected from strengthening tax enforcement among the tax-registered rather than widening the tax base.

Follow-up research should aim at more precisely quantifying (the distribution of) audit returns in the short and long term, as well as deepening our understanding of their determinants in the specific context of LICs. Following-up on our results, additional research should also attempt to assemble more complete information on the size and distribution of tax evasion at the extensive margin and intensive margin to better understand the trade-off for tax administrations between revenue-maximizing strategies and fiscal equity across the population.

1.8. Figures

Figure 1.1 – Timeline - Intervention and Main Declarations



Intervention: Our communication campaign consists of three waves of emails and systematic verification calls in January, July and December 2022.

Suppliers lists: For each year of activity y , a detailed list of all service suppliers (tax-registered or not) must be declared by their clients before the 31st of January in year $y + 1$. The statistical unit of this data is the client-supplier pair. See Table 1.1 for a detailed review of suppliers lists reporting rules.

Income tax: For each year of activity y , all tax-registered individual or firms must declare their personal or corporate income tax statement (PIT or CIT) by mid-April of $y + 1$. This is the main source we use to measure self-declared gross revenue of tax-registered suppliers.

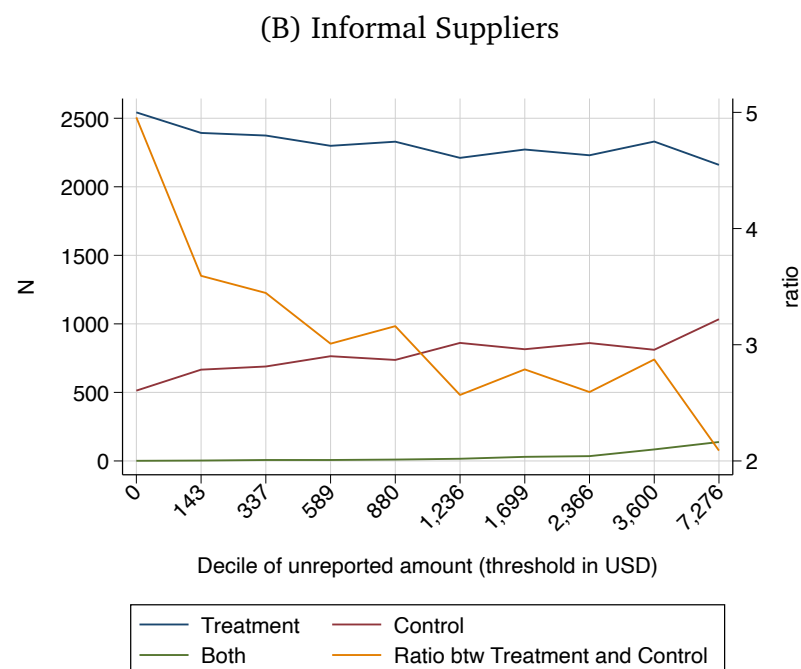
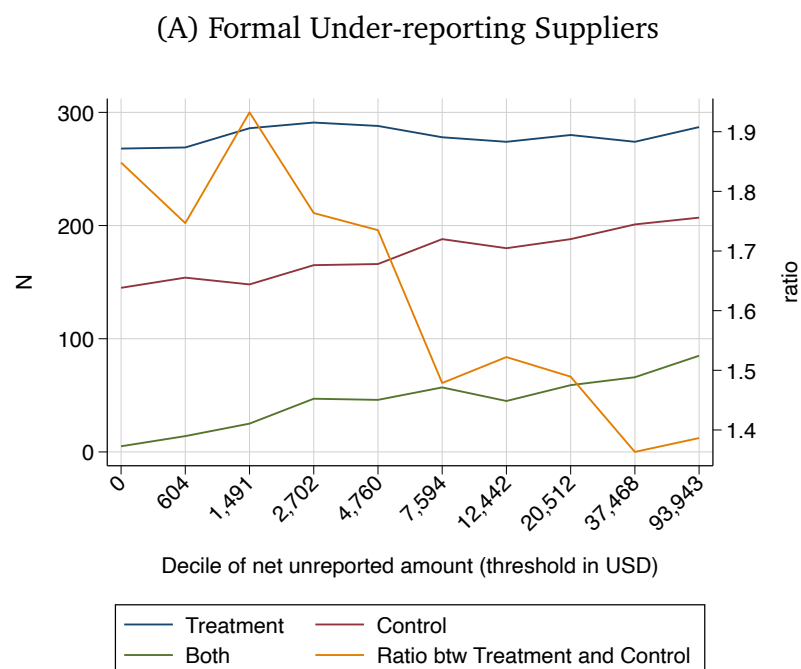
WTS: In this monthly declaration, clients of informal service suppliers indicate the total spent on such services and remit the corresponding 5% withholding tax to the administration. The statistical unit of this data is the client.

Figure 1.2 – Impact on the Share of Correctly Reported Suppliers - 2022
Suppliers Lists - Heterogeneity by Size and Age



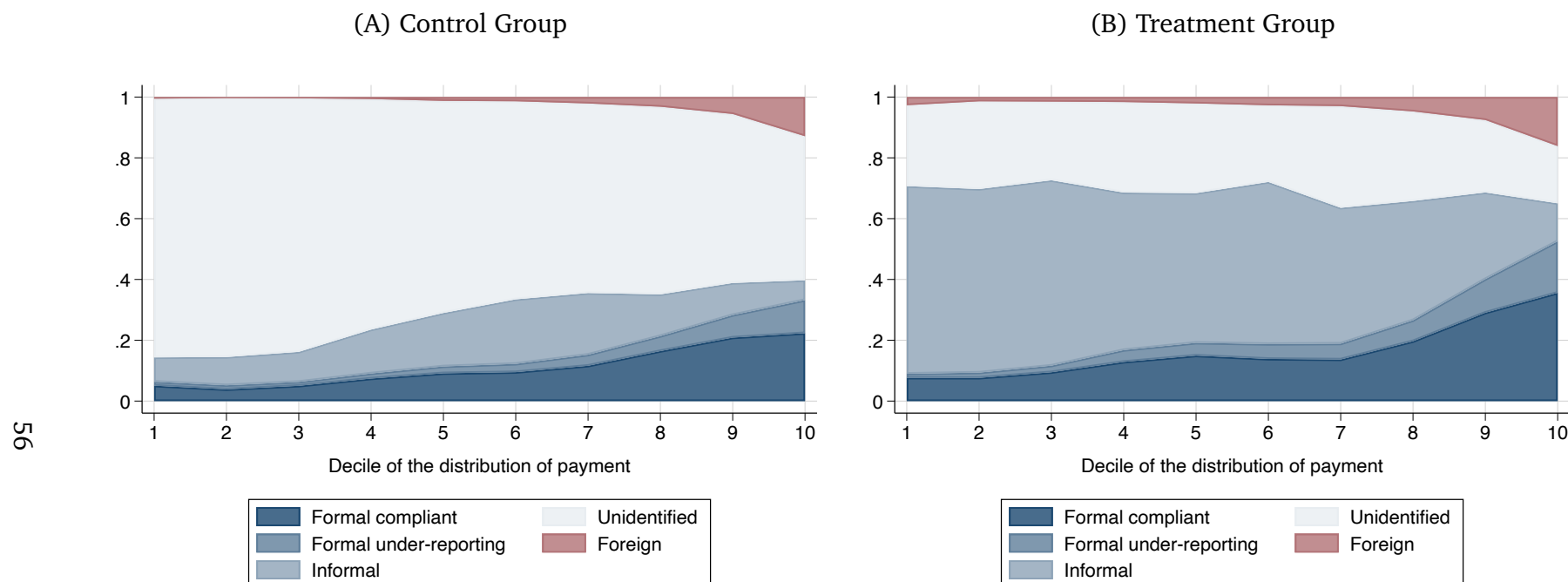
Notes: These figures display the average shares of correctly reported suppliers declared by firms in the treatment and control group, across deciles of four different covariates. For each decile, averages for the control group are the sample mean, averages for the treatment group are obtained by adding our average treatment estimates (using OLS and the model specified in 1.4.2, and variables displayed in Table 1.6 as controls) to the control group's average. Dashed lines indicate 95% intervals. Figure 1.2.(A) exploits the average Revenue declared ex-ante during the 2018-2020 period. Figure 1.2.(B) (resp. 1.2.(C)) use average number of employees (resp. service suppliers), over the same period. Labels of the x-axis for the three figure indicate the decile average of each covariate. Finally 1.2.(D) splits the sample into age groups of similar size. Age measures the number of years since the creation of the firms / start of the activity.

Figure 1.3 – Number of Detected Tax Evading Suppliers - by Quantiles of Unreported Amounts



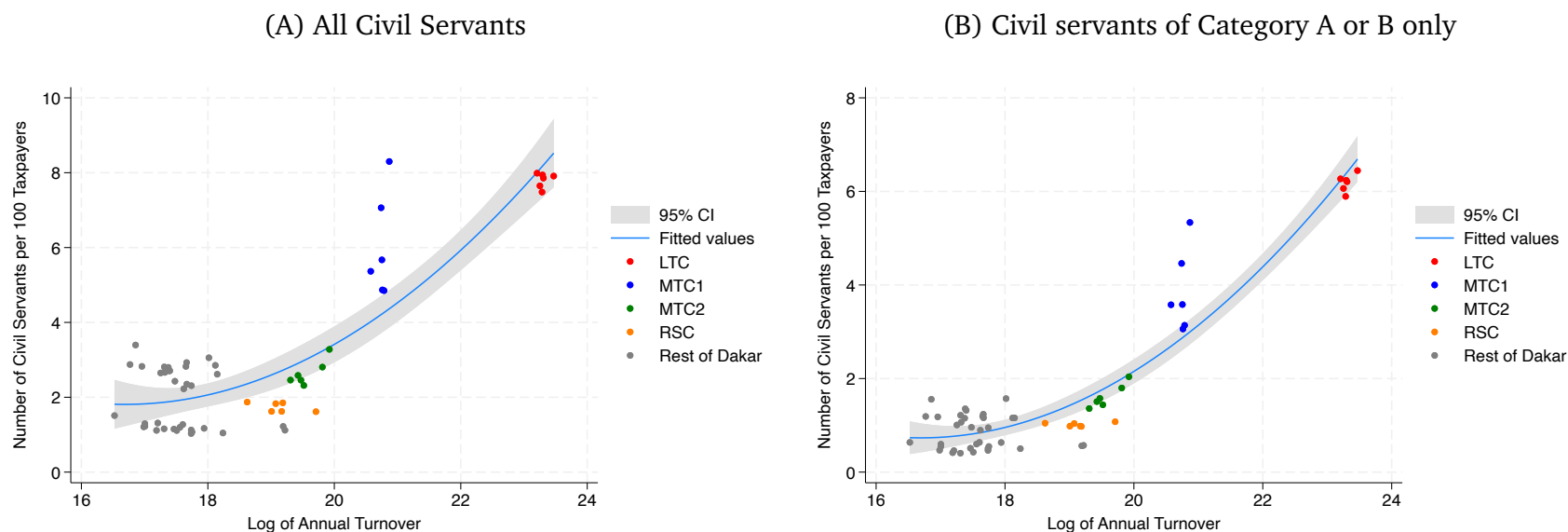
Notes: Figure 1.3.(A) and figure 1.3.(B) display the number of tax evading suppliers by decile of unreported amounts on the intensive (formal) and extensive (informal) margin respectively. By definition suppliers are classified as formal if they are tax-registered, informal otherwise. The blue (red) line represent the number of tax evading suppliers detected thanks to information coming from the treatment (control) group. The yellow line is the ratio between the two. The green line is for suppliers flagged as tax evaders by both groups.

Figure 1.4 – Decomposition of 2022 Payments to Suppliers by Experiment Group, Suppliers Type and Decile of the Distribution of Payments



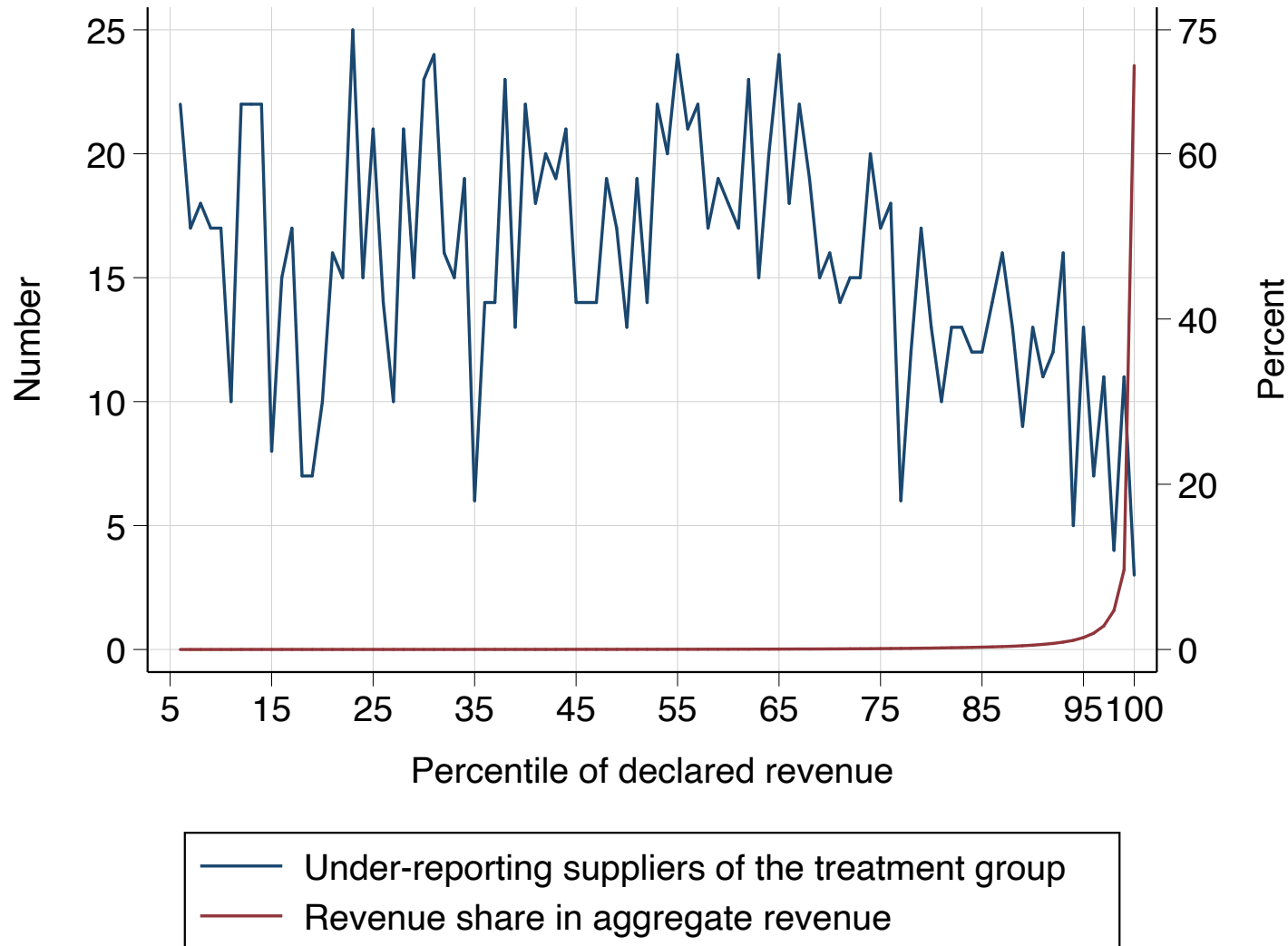
Notes: The statistical unit is the client-supplier pair as declared, by clients (control or treatment), in their 2022 suppliers list, after the full communication campaign. For both Figures, deciles are based on the within-group distribution of yearly payments for each client-supplier pair. The classification of each payment is based on the client (control or treatment) and the suppliers' type. Suppliers are classified across 5 types, according to information available in 2022 suppliers lists and cross-checks with the tax registry and suppliers' own tax declarations for 2022. Suppliers of the control (treatment) group are classified as "Formal compliant" if they are tax-registered and have self-declared total revenue at least as large as the sum of payments received by client of the control (treatment) group. They are classified as "Formal under-reporting" in case that sum is superior to self-declared revenue, or "Informal" if they are not tax-registered. By construction, the rest of Senegalese suppliers are reported with no valid ID in their clients suppliers lists, and are thus considered "unidentified". Foreign suppliers form a separate group because we cannot cross-check payment received against their self-declaration.

Figure 1.5 – Number of Civil Servants per Taxpayers - by Tax Center and Civil Servant Category : 2016-2021



Notes: The statistical unit is the year-center. We exploit data from 2016 to 2021 and 10 tax centers : the Large Taxpayers Center (LTC), the Middle-sized Taxpayers Center 1 (MTC1) the Middle-sized Taxpayers Center 2 (MTC2), the Regulated Sectors Center (RSC) and all other centres of Dakar (excluding *Rufisque* - the most off-center unit the agglomeration, due to unreliable data). The x-axis indicates the log of the average annual turnover declared by taxpayers. The y-axis indicates the number of civil servant per 100 taxpayers within each tax center. Civil servants at the tax administration are split across four categories : A, B, C or D. Only agents of category A or B are required to have a training in fiscal law and administration, and only them can perform audits, establish fines and force tax recovery. Panel (A) considers civil servants of all four categories, while panel (B) restrict the analysis to most qualified civil servants. Example : over 2016-2021, there was consistently more civil servants of category A or B per taxpayers in the Large Taxpayers Center than in any other centers.

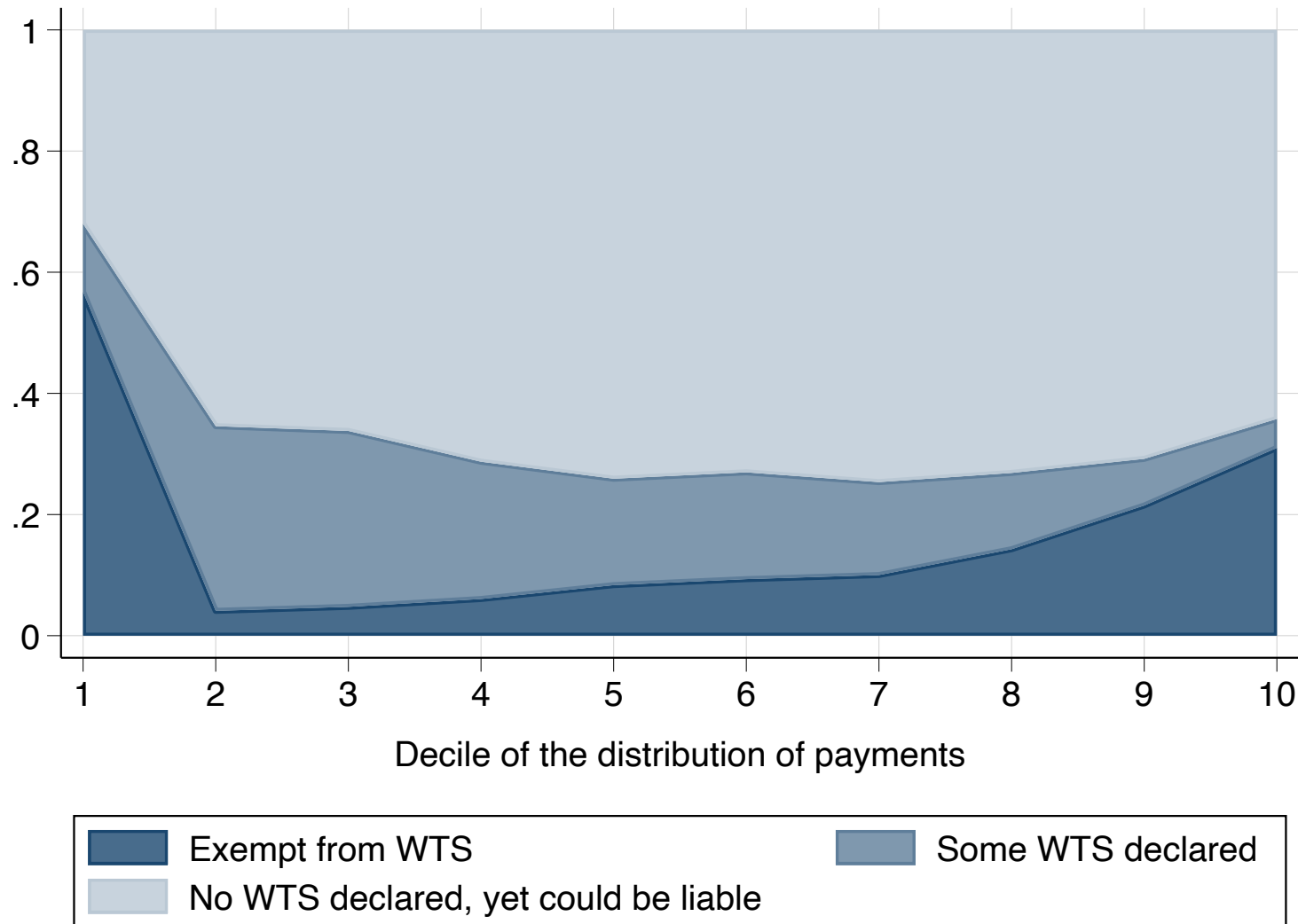
Figure 1.6 – Distribution of the 2,975 Under-reporting Suppliers of the Treatment Group and Revenue Share



Notes: The population for this figure is the full sample of taxpayers active in 2022. On the x-axis are percentiles of the distribution of declared revenue. The blue line (left y-axis), displays the number of under-reporting suppliers of firms from the treatment group in each percentile. The red line (right y-axis) displays the revenue share (in %) of each percentile in aggregate revenue declared. We observe only 3 under-reporting taxpayers within the top 1%, yet that sub-group alone accounts for almost 75% of all revenue. We decide to discard these extremely large taxpayers from all selection methods in our simulations, including our benchmark method.

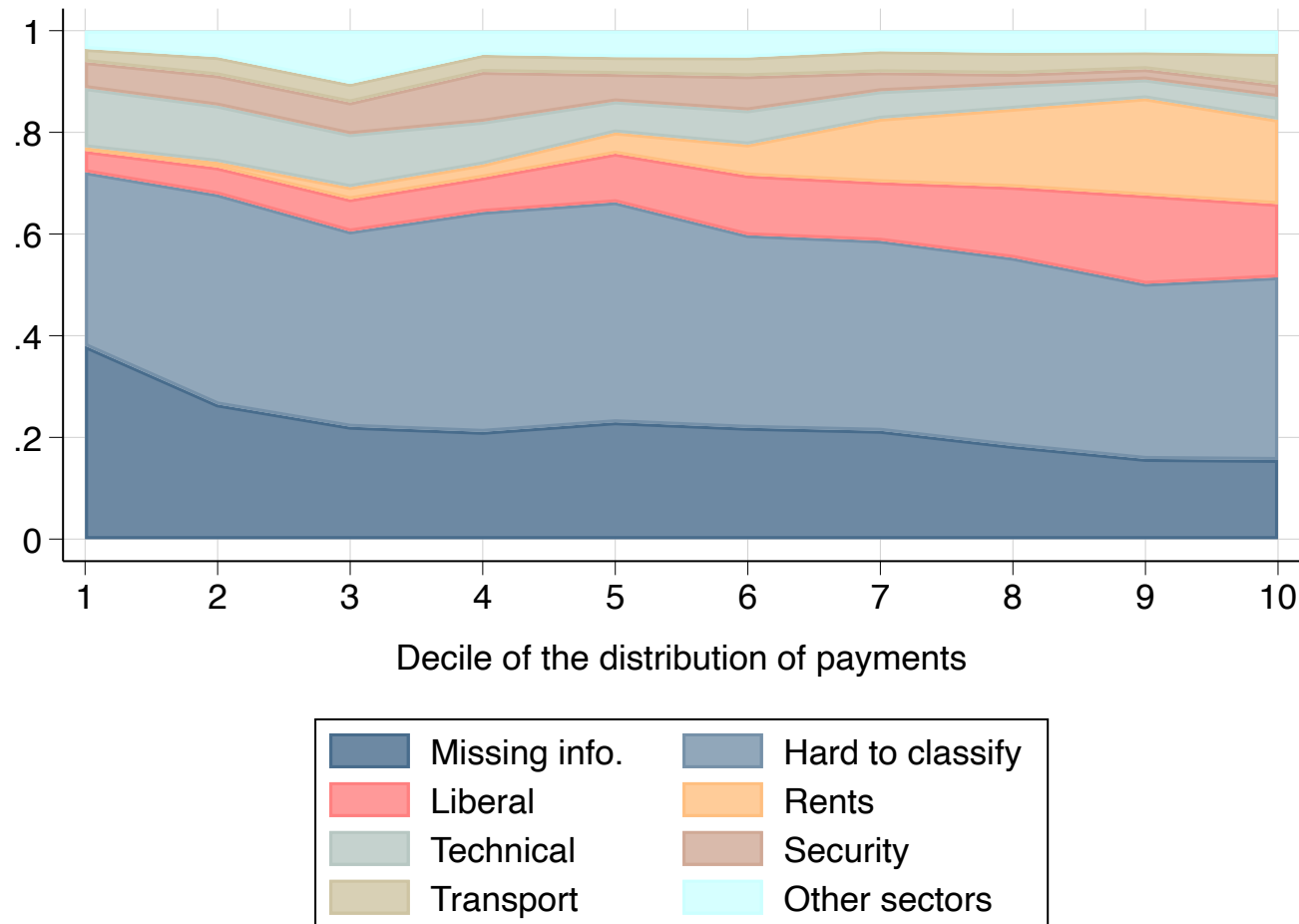
Figure 1.7 – Baseline decomposition of the distribution of payment according WTS liability status - 2018

59



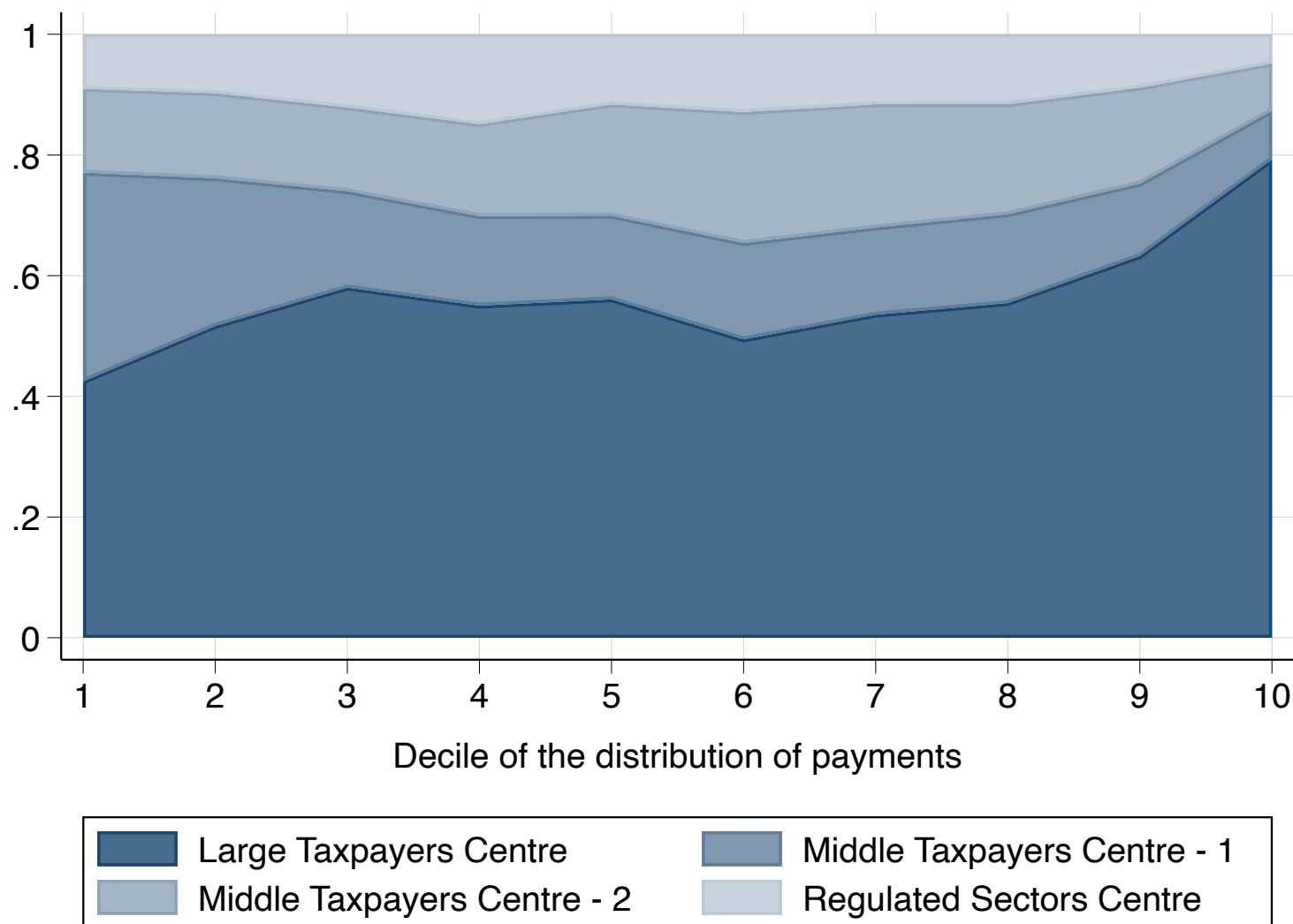
Notes: This Figure decomposes the distribution of all (pooled) payments (client-supplier pairs) declared by taxpayers of our experimental sample at baseline. We consider only the year 2018 as this is the year for which we have the best coverage ex-ante. Suppliers are considered exempt of the WTS if their client reported a correct firm-ID, if they are registered outside of Senegal, if the transaction is below 25,000 XOF (hence the pike for the first decile), or in case they operate in the hospitality sector. Example : among the 10% largest payments declared in 2018 supplier lists about 30% were not liable of the tax and close to 5% were subject to it. However, for the remaining 65%, we do not observe any tax withheld in supplier lists while the complementary information is insufficient to ascertain that the supplier was eligible for exemption.

Figure 1.8 – Baseline decomposition of the distribution of payment according to suppliers' sector - 2018



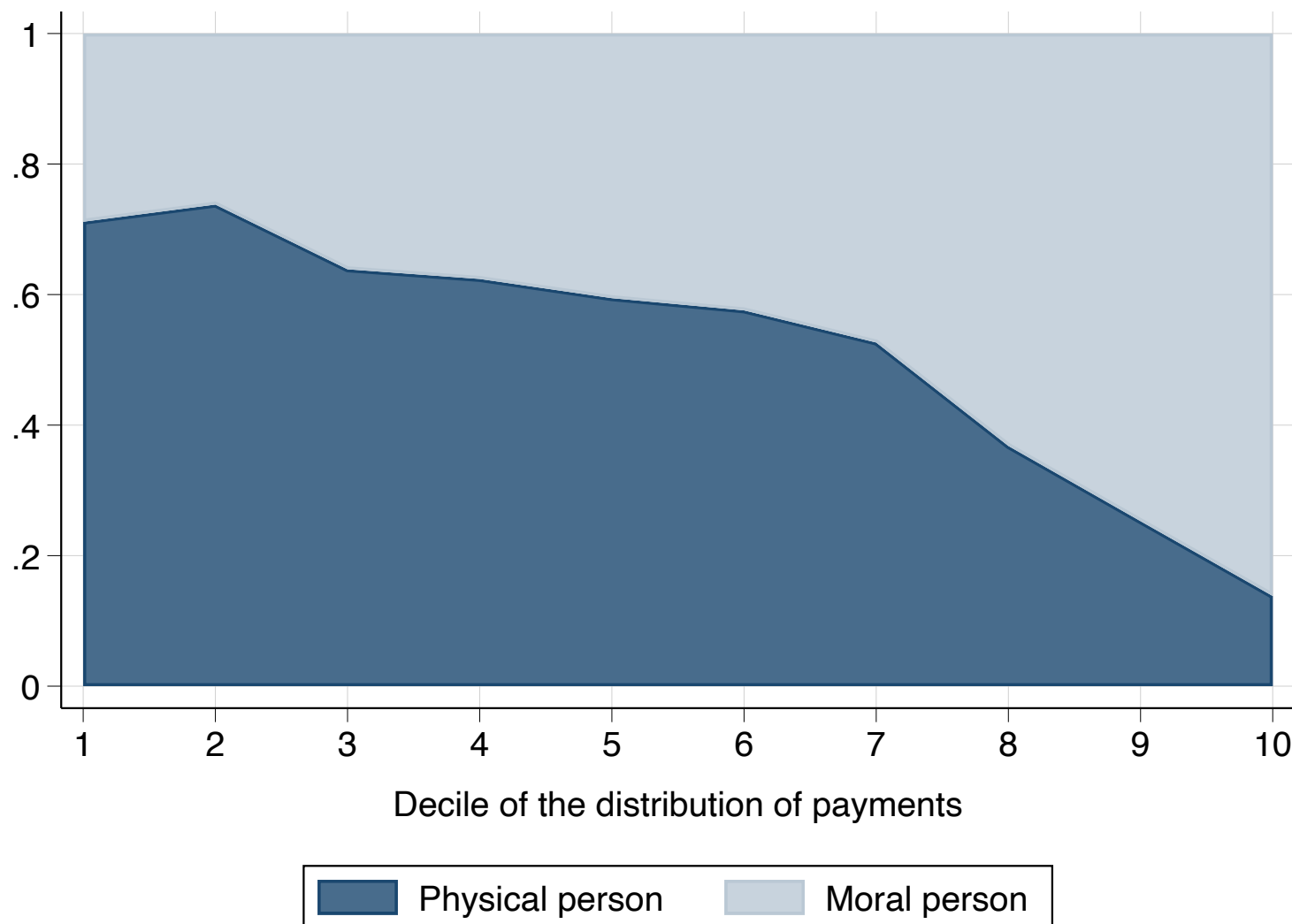
Notes: This Figure decomposes the distribution of all (pooled) payments (client-supplier pairs) declared by taxpayers of our experimental sample at baseline. We consider only the year 2018 as this is the year for which we have the best coverage ex-ante. For about 20% of all payments, the information about the sector is missing. For an additional 40% almost, the information is unfortunately hard to classify: unspecific for the vast majority (e.g. "services") or sometimes too specific (no key words, thus requires case by base inspection). The rest are classified based on key words, in an attempt to capture homogeneous groups. "Liberal": doctors, lawyers, consultants, architects, teachers, accountant etc. "Technical": electricians, plumbers, workers in the construction sectors, technical maintenance, etc. "Other sectors": interns, hotels, restaurants, communication, advertisement and distribution.

Figure 1.9 – Baseline decomposition of the distribution of payment according to clients' tax centre



Notes: This Figure decomposes the distribution of all (pooled) payments (client-supplier pairs) declared by taxpayers of our experimental sample at baseline. We consider only the year 2018 as this is the year for which we have the best coverage ex-ante. We classify each payment based on the tax center of the client. Example: About 60% of all payments come from firms registered at the Large Taxpayers Centre.

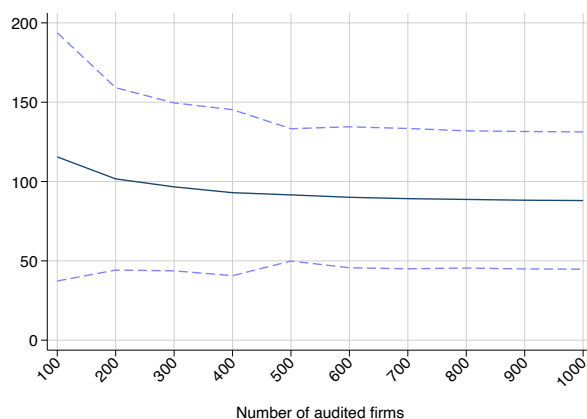
Figure 1.10 – Baseline decomposition of the distribution of payment according the supplier' type of person



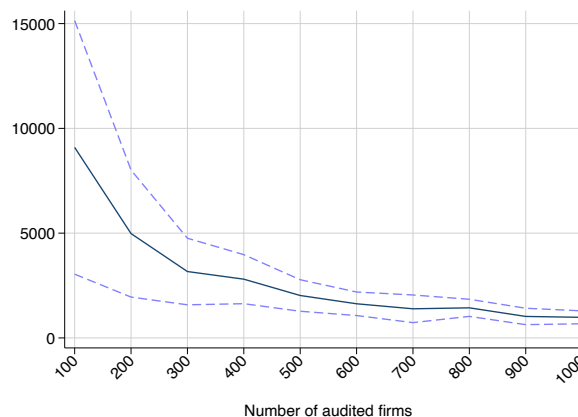
Notes: This Figure decomposes the distribution of all (pooled) payments (client-supplier pairs) declared by taxpayers of our experimental sample at baseline. We consider only payments made in 2020 by taxpayers of the Large Taxpayers Centre. Indeed, we have information on the suppliers' type of person (either moral or physical person) mostly for that sub-sample as it is required only in the online form, which the LTC fully adopts in 2020.

Figure 1.11 – Returns expected from auditing the largest under-reporting formal suppliers according to the treatment group compared to estimated returns when targeting ...

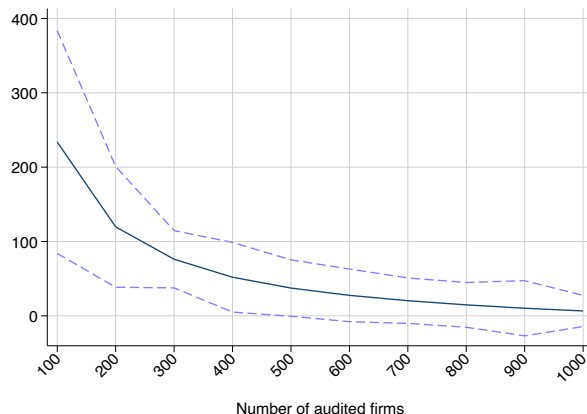
(A) ... under-reporting formal suppliers of the control group



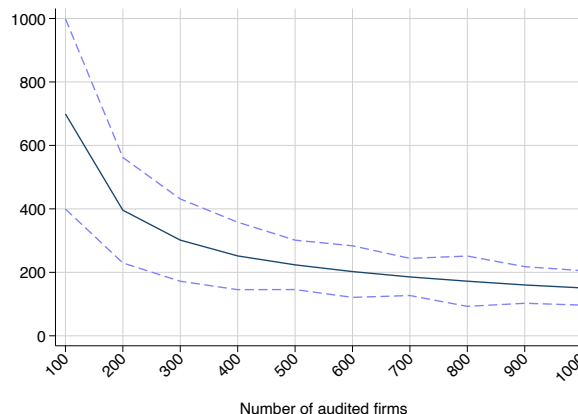
(B) ... randomly selected taxpayers



(C) ... the largest taxpayers



(D) ... the largest taxpayers by tax centre

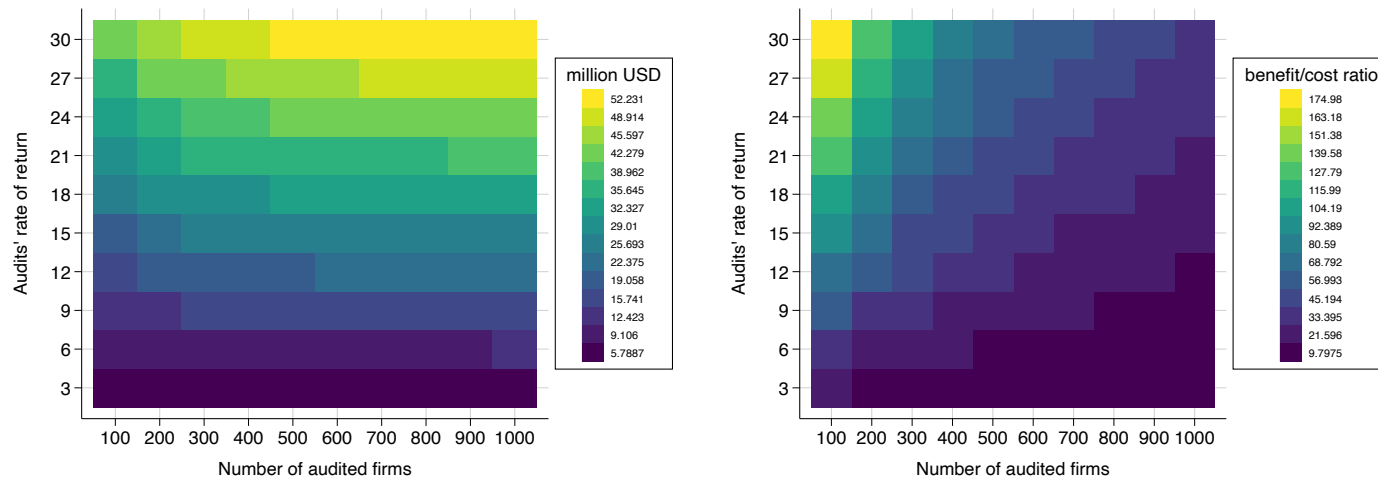


Notes: This Figure plots the *percentage difference* in expected audit returns (tax collected upon hypothetical audits) obtained when comparing audits of the largest under-reporting formal suppliers (according to information declared by the treated group in their 2022 suppliers list), to audits of other formal firms, selected through four alternative methods. In Figure 1.11.(A), the alternative method consists in selecting the largest under-reporting firms according to the *control* group. In Figure 1.11.(B) the alternative method is a random draw from the (active) tax-registry. In Figure 1.11.(C) the alternative method consists in selecting the largest taxpayers (in terms of gross revenue), and for Figure 1.11.(D) we do the same but by tax centre, selecting the same proportion of taxpayers in each tax centre. Dashed lines represent 95% confidence interval estimated by bootstrapping our calculation to allow for some potential measurement errors. *Example:* We estimate that auditing the 200 largest under-reporting suppliers according to the treatment group would multiply audit returns by 4, compared to auditing the largest 200 firms by tax centres. See Section 1.6.3 for a detailed explanation.

Figure 1.12 – Hypothetical Audit Returns and Short-Term Cost-Benefit Analysis

(A) Tax Revenue Collected upon Audits

(B)... and Compared to Costs (Intervention + Audits)



Notes: Figure 1.12.(A) displays the difference, in nominal amounts, in expected audit returns (tax collected upon hypothetical audits) obtained when comparing audits of the largest under-reporting formal suppliers according to the *treatment* group, versus audits of the largest under-reporting formal suppliers according to the *control* group (see Figure 1.11 for a comparison with alternative selection procedure). We estimate 10x10 scenarios by varying the number of taxpayers to be audited and the hypothetical rate of return per audit. Figure 1.12.(B) displays the ratio between the audit surplus and the total cost of our intervention plus the estimated cost of carrying out the follow-up audits.

1.9. Tables

Table 1.1 – Reporting rules for suppliers list - examples for each type of lines

WTS Exemption based on information from (1), (6), (7)	Is the line correct?	Suppliers' type		(1) <i>Amount</i> (‘000 XOF)	(2) <i>Name of supplier</i>	(3) <i>Firm-ID</i>	(4) <i>5% WTS</i>	(5) <i>National-ID</i>	(6) <i>nature</i>	(7) <i>country</i>	Mean Prevalence at baseline (%) <i>in total av. per client</i>	
(A) Not Exempt	Yes	Formal	(a)	78,000	Securitek SA	000987518			surveillance	Senegal	12.3	14.0
	No	Informal	(b)	8,500	Legal Terranga			1385196500..	Service	Senegal	0.7	0.5
	No	Unidentified	(c)	9,000	Dieye Pharmacie					Senegal	54.9	71.5
	Yes	Informal	(d)	200	Abdoulaye Diop		10	1267199080...	consulting	Senegal	2.6	1.6
	No	Unidentified	(e)	8,000	Mouhamed Faye		400			Senegal	19.8	9.5
(B) Exempt	Yes	Informal	(f)	20	Dieynaba Sall			2648200187...		Senegal	0.2	0.1
	Yes	Foreign	(g)	750	S&Q200			BE08.9689.97...		Belgium	0.4	0.1
	No	Unidentified	(h)	24,000	Dakar Beach Hotel				Hotel	Senegal	8.5	2.5

On misreporting: This table illustrates in full details how we proceed to detect misreporting. A taxpayers having submitted at baseline a suppliers list which contains at least one of the error listed here, will be considered as misreporting and thus eligible for the intervention. We also explain below how we use the available information to classify suppliers themselves as “Formal”, “Informal”, “Foreign” or “Unidentified”.

WTS Exemption definition: The supplier of a given line is considered “Exempt” if information present in column (1), (6) or (7) is sufficient to grant exemption, i.e if the amount is lower than XOF 25,000 (see cell (f)(1)), or if the supplier operates in the Hospitality sector (see cells (g)(6) and (i)(6)), or if the supplier is foreign (see cell (h)(7)).

Panel A - “Not Exempt”:

(a) A formally correct *Firm-ID* is indicated in column (3), hence suggesting the supplier is tax-registered therefore that, as far as the client could tell, no withholding was due. This is how transaction with tax-registered suppliers should typically be declared. This *Firm-ID* is then what can be used to check whether this supplier is effectively tax-registered and has declared revenue at least as large as the sum of payments third-party reported here. This supplier will be classified as formal only if the *Firm-ID* matches the tax-registry.

(b) Some information is missing: either a formally correct *Firm-ID* if the supplier is tax-registered (column (3)), or some positive WTS amount if it is not (column (4)).

(c) Some information is missing: either a formally correct *Firm-ID* if the supplier is tax-registered (column (3)) or some positive WTS amount and a formally correct *National-ID* if it is not (column (4) and (5)). Pooling all suppliers lists for 2018-2020 we estimate that 54.9% of the suppliers reported at baseline are of that type.

(d) A positive WTS is declared, together with a formally correct *National-ID*. This is how transaction with informal suppliers (not tax-registered) should typically be declared. Note that in some rare case, a *Firm-ID* may be present together with, or instead of, the *National-ID*. In that case the line is still considered correct.

(e) Some information is missing: A positive WTS is declared but some identifier, typically a *National-ID* (column (5)), or more rarely, a *Firm-ID* (column (3)) is missing.

Panel B - “Exempt”:

(f) The transaction is exempt because the amount is inferior to XOF25,000, the WTS amount (column (4)) is thus rightly missing. Some formally correct *National-ID* is provided (it could have been a *Firm-ID*), making the line correct.

(g) The transaction is exempt because the supplier is foreign (column (7)). Some ID has been provided, hence the line is correct. Note that in that case, the ID needs not even to be formally valid (as we would not know which rule to apply) for the line to be considered correct.

(h) The transaction is exempt because the supplier is a hotel (column (6)). Yet some identifier is missing either a formally correct *Firm-ID* if the supplier is tax-registered (column (3)) or a formally correct *National-ID* if it is not.

Note on IDs: Every individual, tax-registered or not, has a *National-ID*. In addition, all tax-registered actors have a *Firm-ID* as the tax administration requires it for all operations. But the reverse is not true. *Firm-IDs* are produced by the National Agency for Statistics and Demographics and used by many institutions other than the tax administration. For simplification and unless specified, throughout the paper whenever we use the term “ID”, it refers to either a *National-ID* or a *Firm-ID*. Unidentified suppliers are those for which no valid ID has been indicated. As we can see from the last two columns of the table, more than 80% of the suppliers are unidentified at baseline.

Note on names: All names are fictional and made precise only to illustrate the diversity of patterns.

Table 1.2 – Rules and examples for the definition of the under-reporting (formal) suppliers

Experimental Group	Transaction Index	Data from suppliers list			Data form self-declared income tax statements	Our classification
		client	suppliers ID	transaction amount	Annual total turnover of the supplier	
		(1)	(2)	(3)	(4)	(5)
Treatment	(a)	A	firm-ID-1	300	1000	under-reporting
	(b)	A	firm-ID-2	700	2000	not under-reporting
	(c)	A		2300	impossible to cross-check	indeterminate
	(d)	B	firm-ID-3	700	400	under-reporting
	(e)	B	firm-ID-1	800	1000	under-reporting
	(f)	B	firm-ID-2	1200	2000	not under-reporting
Control	(g)	C	firm-ID-1	500	1000	not under-reporting
	(h)	C		200	impossible to cross-check	indeterminate
	(i)	D		400	impossible to cross-check	indeterminate
	(j)	D	firm-ID-2	900	2000	not under-reporting
	(k)	D		600	impossible to cross-check	indeterminate
	(l)	D		100	impossible to cross-check	indeterminate

Notes This table displays fictional data to illustrate how we define under-reporting relative to information declared by the treatment (resp. control) group. Columns (1) to (3) are information provided by two treatment firms (A and B) and two control firms (C and D). Columns (4) display (hypothetical) information on revenue self-declared to the tax administration by suppliers for which a firm-ID has been declared. Finally column (5) indicates the result of our classification method. In green, we display information, which, in that example, has been declared as a response to the intervention by treated firms.

Type 1: firm-ID-1 is an identified supplier of client A, B and C. According to self-declared income statement submitted to the tax administration its revenue equal 1000. Thanks to the intervention client, B reveals that transaction (e) goes to *firm-ID-1*. Summing all transactions from treatment clients going to *firm-ID-1*, we have the $300+800 > 1000$, hence *firm-ID-1* is classified as under-reporting relative to the treatment group. It is not, however, classified as under-reporting relative to the control group, as the sum of transactions going from control clients to *firm-ID-1* equal $500 < 1000$.

Type 2: firm-ID-2 is an identified supplier of client A, B and D. According to self-declared income statement submitted to the tax administration its revenue equal 2000. Thanks to the intervention client, A reveals that transaction (b) goes to *firm-ID-2*. Summing all transactions from treatment clients going to *firm-ID-2*, we have the $700+1200 < 2000$, hence *firm-ID-2* is classified as *not* under-reporting relative to the treatment group (idem for the control group). This is despite the fact that if we were to sum all transactions going to *firm-ID-2*, from both treatment and control clients, we would have $700+1200+900 > 2000$.

Other cases: firm-ID-3 is an identified supplier of client B only. It's under-reporting status is unequivocal. Transactions (c), (h), (i), (k) and (l) benefit unidentified suppliers which prevents any cross-checking against self-declared income tax statement, and forces us to leave their under-reporting status as "indeterminate".

Justification: With our definition, the average share of transaction, per client, going to under-reporting taxpayers is $(1/3+2/3)/2 = 50\%$ in the treatment group, and 0 in the control group. But if under-reporting was defined relative to *all* transactions, the average treatment impact on the share of transaction going to under-reporting taxpayers, would be $(2/3+3/3)/2 - (1/2+1/4)/2 = 45.8\%$. Our definition of *relative* under-reporting prevents externalities outside of the treatment group due to network effects. Most importantly, it better captures what to expect if we were to scale up our intervention to all firms, as, by definition, under-reporting firms of type 2 are only detectable when some clients are not treated.

Table 1.3 – Suppliers’ lists Declaration and Sample Selection by Tax Units

	Number of Taxpayers for which we observe Suppliers Lists					Selected Taxpayers		
	2018	2019	2020	2018-2020				av. number of years observed
				N	as % of all	N	as % of (4)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Large Taxpayers Centre	684	709	760	874	86.5	799	91.4	2.5
Middle Taxpayers Centre - 1	515	191	154	567	83.5	439	77.4	1.5
Middle Taxpayers Centre - 2	1104	145	57	1127	69.6	1000	88.6	1.20
Regulator Sectors Centre	930	1121	336	1366	68.8	1249	91.4	1.79
Total	3233	2166	1307	3934	74.4	3487	88.5	1.70

Notes. This table provides information on the distribution of the baseline data we used to select our experimental groups. In total we got access to 6706 suppliers lists, spanning over three years, reported by 3934 different taxpayers, i.e. 74.4% of the total taxpayers population of the four largest taxpayers units. We identified 3,487 taxpayers having misreported at least once over the period, i.e 88.5% of the initial sample. On average, for that selected sub-sample, we observe about 1.7 suppliers list over 2018-2020

Table 1.4 – Sectoral Decomposition of the Selected Sample

	N	Total Sales mean	% in total	Total Employment mean	% in total
	(1)	(2)	(3)	(4)	(5)
Accommodation and food service activities	64	3557	1.23	78	2.85
Activities of extraterritorial organisations and bodies	4	255	0.00	129	0.28
Administrative and support service activities	424	3962	9.05	85	20.62
Agriculture, forestry and fishing	77	5877	2.44	81	3.54
Arts, entertainment and recreation	9	7240	0.34	92	0.46
Construction	206	6233	6.92	80	9.35
Education	48	918	0.23	21	0.56
Electricity, gas, steam and air conditioning supply	23	43874	5.44	144	1.88
Financial and insurance activities	104	13657	7.65	98	5.80
General public administration activities	18	380	0.03	23	0.23
Human health and social work activities	390	220	0.46	10	2.25
Information and communication	103	9470	5.26	39	2.27
Manufacturing	243	14432	18.88	144	19.96
Mining and quarrying	20	55962	6.03	64	0.73
Other business support service activities n.e.c.	28	2936	0.43	24	0.38
Professional, scientific and technical activities	323	1425	2.48	10	1.88
Real estate activities	85	2325	1.05	14	0.66
Retail sale via stalls and markets	1099	4232	25.05	31	19.73
Transportation and storage	160	5203	4.48	42	3.83
Water supply; sewerage; waste management and remediation activities	5	49040	1.32	309	0.87
<i>unassigned</i>	54	3944	1.14	55	1.70
Total	3487	18560124	100	175265	100

Notes. The classification used is the official NAEMA classification, used by countries the UEMOA, which, for that level of analysis, is consistent with the NACE classification used by European countries. Total sales and total employment are averaged over 2018-2020 per firm, then averaged per sector to obtain figures in column (2) and (4). Example : 206 firms of our selected sample operate in the Construction sector, their average yearly turnover during 2018-2020 was equal to USD 6,233,000, and they employed 80 individuals on average. This sector represents 6.92% of all sales and 9.35% of all employment in our sample.

Table 1.5 – Baseline patterns of declarations and payments of WTS

	Conditional on declaring a supplier list:				Does not declare a supplier list
	pays and 0 declares 0 % (1)	pays > 0 and declares 0 % (2)	pay 0 and declares > 0 % (3)	pay > 0 and declares > 0 % (4)	but pay > 0 % (5)
Large Taxpayers Unit	14.1	47.2	2.0	36.5	6.6
Middle Taxpayers Unit - 1	16.3	59.2	1.5	22.8	4.5
Middle Taxpayers Unit - 2	25.5	63.9	0.6	9.8	1
Regulator Sectors Unit	25.2	61.5	0.6	12.3	11.3
Total	20.6	57	1.2	21.1	6.4

Notes. This table compares total amounts of the WTS tax that clients declare having withheld from their suppliers in a given year (according to yearly supplier lists) to the total amount of the monthly remitted WTS they have effectively paid to the administration for that same period. For column 1-4 we exploit all collected supplier lists - figures add to 100% for each centre. In column 5 we restrict our sample to the years that are best covered by our data collection of the supplier lists, (i.e 2020 for LTC, 2018 for MTC-1 and MTC-2 and 2019 for RSC) and compute the percentage of firms that have remitted a positive amount and yet did not submit any supplier list. Example: in column (2) we can read that in the , 47.2 % of the supplier list we collected from the Large Taxpayers Centre have no information about any withheld amount, although some positive amount has been remitted during the year. Figures of column (3) suggest the contrary is much more rare.

Table 1.6 – Summary Statistics and Balance tests - Full Sample (N = 3,487)

Source	Covariates/Outcomes	Summary Statistics			Balance tests	
		Mean	Std. dev.	Median	β	Std. err.
(A) Suppliers lists)	Number of suppliers	39.663	152.777	10	-6.773	4.635
	Share suppliers without any ID	0.867	0.215	1	-0.008	0.006
	... with a positive WTS	0.115	0.239	0	-0.002	0.007
	... in the Hospitality sector	0.001	0.013	0	-0.000	0.000
	... Senegalese	0.989	0.047	1	0.001	0.001
	... with transaction bellow taxable threshold	0.017	0.067	0	-0.000	0.002
	N. of y. for which we observe a suppliers list	1.705	0.775	2	-0.001	0.019
	Declares suppliers list online	0.326	0.468	0	0.004	0.011
(B) Other Declarations)	Av. Sum of WTS remitted - 2018-2020 (USD'000)	4.142	22.982	0.620	-0.894	0.784
	Av. turnover - 2018-2020 (USD'000)	5480.939	30678.406	407.394	-1213.954	981.129
	Av. number of employees - 2018-2020	56.564	319.884	6.699	-22.606**	11.135
(C) Tax Registry)	Large Taxpayers Centre (binary)	0.229	0.420	0	-1.810	1.277
	Upper-Middle Taxpayers Centre (binary)	0.126	0.332	0	-6.950	5.998
	Upper-Middle Taxpayers Centre (binary)	0.286	0.452	0	-1.209	3.065
	Regulated Sector Centre (binary)	0.358	0.479	0	-1.054	1.777
	Legal person (binary)	0.708	0.454	1	-0.006	0.011
	Number of years since first digital admin. record	11.345	3.275	13	-0.016	0.109
	Firm created before 1990 (binary)	0.299	0.457	0	0.005	0.015
	... btw 1990 and 2000 (binary)	0.076	0.266	0	0.000	0.009
	... btw 2000 and 2010 (binary)	0.274	0.446	0	-0.005	0.014
	... after 2010 (binary)	0.349	0.476	0	0.000	0.016

Notes. The first three columns provide summary statistics about covariates and outcomes we use throughout the study for the entire sample. The last two columns summarize results from OLS estimation of equation (1) (see Section 1.4.2) without covariates, with each variable as the outcome. Variables of panel (A) are measured using suppliers lists collected at baseline (average over the period 2018-2020). Panel (B) exploits information reported in other declarations such as monthly Withholding Tax on Services (line 1), or Personal Income Tax, Corporate Income Tax, Value Added Tax, Business Property Tax, and the Synthetic Tax for small Businesses (line 2), or Withholding on employees (line 3). Panel (C) exploits all relevant information we could retrieve from the tax registry.

Table 1.7 – Direct Responses - Decomposition by Tax Centre

	by tax centre				Full Sample	
	LTC	MTC-1	MTC-2	RSC	N	%
Rectification submitted for 2018-2020 suppliers list	247	128	285	401	1061	60.8
No rectification submitted but message reception confirmed	144	82	154	177	557	31.9
... ask for more time	95	38	52	101	286	16.4
... cannot get the required information	26	20	48	30	124	7.1
... ask for assistance	16	15	14	17	62	3.5
... ceased activity	3	6	7	2	18	1.0
... other explanations	4	3	33	27	67	3.8
No rectification submitted, no confirmation	10	9	61	45	125	7.1
Total	401	219	500	623	1743	100

Notes: This table decomposes the treatment group by type of direct responses and tax centre in order to assess treatment take-up. Out of the 1,743 treated firms, 60.8% have submitted a rectifying declaration for 2018-2020 suppliers list pinned down by our communication campaign, 31.9% did not but have acknowledged reception of the message, and 7.1% have remained unreachable. Among those who received the message but did not submit a rectifying declaration, 286 (i.e half of that sub-group) have asked for more time in the last communication we had with them. *LTC:* Large Taxpayers Centre - *MTC-1:* Middle Taxpayers Centre 1 - *MTC-2:* Middle Taxpayers Centre 2 - *RSU:* Regulated Sectors Centre

Table 1.8 – Impact on Misreporting, Information, Fraud Detection and Remittance of Withholding Tax on Informal Service Suppliers (WTS)

			β	Std. err.	μ_c	N	Bonferroni p	Rand. inf. p
Conditional on declaring 2022 suppliers list	(A) Misreporting	Misreport no supplier (binary)	0.101***	0.015	0.182	2633	0.00	0
		Misreport all suppliers (binary)	-0.195***	0.014	0.289	2633	0.00	0
		Misreport some suppliers (binary)	0.093***	0.018	0.527	2633	0.00	0
		Share of misreported suppliers	-0.233***	0.014	0.543	2633	0.00	0
	(B) Information Prevalence	Share of suppliers ...						
		... with a formally valid ID	0.247***	0.014	0.469	2633	0.00	0
		... with intentionally wrong ID	-0.003	0.005	0.026	2633	1	0.43
		... with positive WTS declared	0.036***	0.014	0.508	2633	0.20	0.01
		... satisfying cond. to be exempt of WTS	0.001	0.003	0.033	2633	1	0.60
		... with a telephone number	0.092***	0.016	0.410	2633	0.00	0
	(C) Fraud Detection	Share of suppliers identified as ...						
		... informal	0.175***	0.012	0.229	2633	0.00	0
		... formal	0.069***	0.012	0.232	2633	0.00	0
		... formal and compliant	0.043***	0.010	0.175	2633	0.00	0
		... formal and under-reporting	0.025***	0.005	0.056	2633	0.00	0
	(D) Structure of Service Supply	Number of suppliers	4.590	7.203	49.903	2633	1	0.53
		Total (ihs)	0.036	0.056	17.934	2633	1	0.52
		Average (ihs)	-0.021	0.043	15.472	2633	1	0.63
		Median (ihs)	-0.079*	0.043	14.765	2633	1	0.06
		Standard Deviation (ihs)	0.105	0.171	13.928	2633	1	0.54
Full Sample		Declare 2022 suppliers list (binary)	0.017	0.013	0.772	3487	1	0.20
	(E) Remittance of the WTS	Remitted a positive WTS in 2022 (binary)	0.018	0.014	0.671	3487	1	0.23
		WTS amount remitted in 2022 (ihs)	0.236	0.205	9.521	3487	1	0.56
		Remitted a positive WTS in 2023 (binary)	0.008	0.015	0.605	3487	1	0.24
		WTS amount remitted in 2023 (ihs)	0.091	0.210	8.258	3487	1	0.67

Notes. This table displays all main results. We use OLS method to estimate the model specified in 1.4.2, with variables displayed in Table 1.6 as controls. Panels (A) to (D) are restricted to taxpayers that have submitted a 2022 suppliers list, after the full communication campaign (hence N=2633). The Statistical unit is always the taxpayer declaring suppliers list (i.e. the client). The last outcome of Panel (A) as well as outcome variables for panels (B) and (C) all measure the share, at the taxpayer level, of some specific type of suppliers. Outcome variables of panel (D) measure properties of the distribution of payments to suppliers, reported by each taxpayers. Outcome variables of panel (E) aggregate information from 2022, or 2023, monthly WTS remittances to assess the impact of the intervention on WTS remittances at the year level. These variables are measured without using 2022 suppliers lists, and thus estimations can exploit the full sample (N=3487)

Table 1.9 – Identified Suppliers in 2022 - Summary Statistics by Intervention Group and Suppliers Compliance Status

Clients' Group	Suppliers' Compliance Status	Number of Suppliers	Av. N. of clients	Income received from this group of clients					
				Total			Net unreported		
				Mean	Aggregate		Mean	Aggregate	
				USD	M. of USD	% of GDP	USD	M. of USD	% of GDP
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(A) Treatment group	Formal	6,176	2.086	237,040	1463.9	5.28	29,666	183.2	0.66
	... and compliant	3,381	2.648	334,767	1131.8	4.08	0	0	0
	... and under-reporting	2,795	1.405	118,823	332.1	1.19	65,552	183.2	0.66
	Informal	23,141	1.021	4,178	96.6	0.34	4,178	96.6	0.34
	All	29,317	1.245	53,233	1560.6	5.63	9,547	279.9	1.01
(B) Control group	Formal	4,610	1.832	96,387	444.3	1.60	17,232	79.4	0.28
	... and compliant	2,868	2.146	117,303	336.4	1.21	0	0	0
	... and under-reporting	1,742	1.316	61,951	107.9	0.38	45,603	79.4	0.28
	Informal	7,750	1.018	7,100	55.0	0.19	7,100	55.0	0.19
	All	12,360	1.322	40,402	499.3	1.80	10,879	134.4	0.48
(C) % variation	Formal	33.96	13.81	145.92	229.46	229.46	72.15	130.63	130.63
	... and compliant	17.88	23.38	185.38	236.43	236.43	0.	0.	0.
	... and under-reporting	60.44	6.76	91.80	207.73	207.74	43.74	130.63	130.63
	Informal	198.59	0.24	-41.15	75.70	75.70	-41.15	75.70	75.70
	All	137.19	-5.81	31.75	212.52	212.52	-12.24	108.15	108.15

Notes. This table provides basic statistics by type of identified suppliers (i.e reported with a valid ID in 2022 suppliers list), according to intervention groups. Example: information submitted by clients of the treatment group in their 2022 suppliers list help identify 29,317 suppliers in total. Out of this group, 23,141 are informal, 6,176 are formal, and 2,795 are flagged as under-reporting when comparing the sum of payments they received from clients of the treatment group to the total revenue they self-declared in their 2022 income tax statement. On average, these under-reporting formal suppliers have 1.405 from the treatment group, receive a total of USD 118,823 but under-report 65,552. In aggregate, this implies that about USD 183 millions have been under-reported by identified formal suppliers of the treatment group, i.e. 0.66 % of GDP. Note that by definition formal compliant suppliers do not under-report, hence their unreported income is 0. Also, as compliance status are defined *relative* to information reported by each intervention group, the same supplier can appear in both panel (A) and (B)

Table 1.10 – Distribution of Unreported amounts - Suppliers of the Treatment Group - 2022

Quantile	Unreported amount (USD)		Underreporting Formal Suppliers			Informal Suppliers		
	threshold	average	Population		% in tot. unreported	Population		% in tot. unreported
			N	%		N	%	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D-0	40	66	84	3.24	0.002	2509	96.76	0.059
D-10	140	241	89	3.43	0.008	2505	96.56	0.216
D-20	346	474	97	3.74	0.016	2496	96.26	0.424
D-30	607	763	118	4.55	0.031	2476	95.44	0.675
D-40	946	1,125	113	4.36	0.044	2480	95.63	0.998
D-50	1,344	1,599	147	5.67	0.083	2447	94.33	1.399
D-60	1,920	2,248	181	6.98	0.14	2413	93.01	1.943
D-70	2,764	3,449	259	9.98	0.302	2334	90.01	2.893
D-80	4,547	6,642	532	20.51	1.22	2062	79.48	4.935
P-90	10,881	11,137	62	23.94	0.238	197	76.05	0.793
P-91	12,480	12,887	83	31.92	0.36	177	68.08	0.837
P-92	14,382	14,702	83	32.04	0.433	176	67.94	0.927
P-93	16,800	17,373	94	36.29	0.56	165	63.70	1.048
P-94	20,056	20,638	120	46.15	0.838	140	53.84	1.079
P-95	24,397	25,770	95	36.68	0.832	164	63.31	1.552
P-96	31,798	34,033	139	53.66	1.611	120	46.33	1.538
P-97	43,301	47,201	137	52.68	2.196	123	47.31	2.188
P-98	65,974	88,300	154	59.45	4.719	105	40.54	3.451
P-99	140,400	639,570	208	80	51.823	52	20	7.586
All		10,777	2795	10.77	65.458	23141	89.22	34.542

Notes. Suppliers of this sample satisfy the following conditions: a) they supply at least one treated taxpayers, b) at least one of their treated clients has provided their ID in the 2022 supplier list c) they have omitted to report some positive amount. If the ID matches the tax-registry they are classified as “Under-reporting”, otherwise they are classified as “Informal”. The statistical unit is the supplier.

Table 1.11 – Underreporting Rates in 2022 by Decile of Declared Revenue

Decile	Threshold (USD)	Average Declared Revenue (USD)	Under-reporting Taxpayers		Average under-reporting rate as a % of declared revenue	
			N	%	<i>conditional on under-reporting</i>	<i>unconditional</i>
1	0	330	1,395	49.56	99.37	49.27
2	1,120	1,970	138	4.90	88.63	4.34
3	2,734	3,916	171	6.07	82.20	5
4	5,402	7,465	169	6.01	81.27	4.88
5	9,920	13,666	163	5.78	81.91	4.73
6	18,164	25,245	195	6.92	77.84	5.38
7	34,359	49,986	189	6.71	72.22	4.84
8	71,246	112,452	145	5.15	78.51	4.05
9	170,928	301,713	125	4.44	82.66	3.67
10	513,231	26,687,004	102	3.61	74.54	2.70
All		320,375	2,792	9.92	81.91	8.89

Notes. The population consists of all the 28,141 active taxpayers of the tax registry, minus the largest 1% (measured in terms of gross revenue). For 2,792 of them, information coming from the treatment group's 2022 suppliers lists indicate that they have under-reported their revenue in 2022. For each of these under-reporting taxpayers, we measure under-reporting rate as the ratio between net unreported income and total *real* income (declared + unreported). Column (2) displays average revenue declared, by decile of income declared. Column (5) and (6) display average under-reporting rates, conditional and unconditional on under-reporting. We use unconditional under-reporting rates to estimate unreported amounts detected upon hypothetical audits in the alternative scenario of Figure 1.11.B, 1.11.C, 1.11.D

Appendices

A. Baseline data, Sample Selection and Randomisation

To identify mis-reporting firms at baseline, we exploit an unbalanced panel of suppliers lists over the period 2018-2020. We digitised hard copies of suppliers list stored by the largest centres (LTC, MTC1, MTC2, RSC) for the year 2018 and 2019. Due to unexpected complexities related to this data collection we could not systematically find and digitize *all* suppliers lists of all four center by the start of the experiment. The coverage is therefore imperfect, in particular for the following center-year: MTU-1 (2019 and 2020) MTU-2 (2019-2020) and RSC (2020 and to a fewer extent 2018). Importantly though, we prioritized data collection from the LTC, by far the most strategic center for the experiment, and therefore managed to guaranty a coverage as complete as possible for the three years. At the time of this data collection, tax centers also did not systematically produce lists of taxpayers having submitted, or not, a supplier lists. We therefore cannot exclude that, even for well covered center-years, a few additional lists could be missing because we did not find them in taxpayers' physical folders even though they had been submitted - this should only be very marginal though. We then completed this database by all reports submitted online until 2020. An online platform was operational from 2019 on but first made available only to a small sample of large firms. In 2020, most firms from the LTC declared online, while most firms from the other three centres still submitted paper declarations. To summarise, taking into account imperfect compliance, our baseline data on suppliers lists covers LTC's firms for about 2-3 years (the last one typically submitted online) while our panel from the other centres has been mostly digitised and typically ends in 2019. The unbalanced nature of our ex-ante sample is therefore correlated with center's assignation. Fortunately, as we stratified our randomization across center, this should not matter for our intervention. In that regard, balance tests show that supplier list of treated firms are not more/less likely to have been observed at baseline.

Our target sample for the intervention are firms (clients) who have misreported suppliers list over 2018-2020. To identify these firms, we exploit data from our 2018-2020 unbalanced panel. For each client, we identify misreported lines of type (c), (e) or (h) (Table 1) and compute the share of these lines in all reported list since 2018. We then use the median of the distribution of this share to create a binary variable splitting the sample into high and low proportion of misreported lines. We also compute, at the client level, the average number of suppliers per year reported in suppliers lists over the observed period and use the 45th and 90th quantiles of the distribution of this variable to split the sample into 3 groups. For about 15 % of the

sample at this selection stage, we could only collect *quarterly* suppliers list and thus could not satisfyingly estimate the average number of suppliers per year; we assigned these firms to a fourth group. Last, we removed from our sample all firms with no or very few mis-reporting issues, as well as firms that we could not unambiguously match to the tax registry. This left us with a sample of 3487 firms. We eventually randomised by stratifying across centres and the two sets of quantile groups defined above, which leads to 32 strata ($4 \times 2 \times 4$), with the least represented cell accounting for 14 firms.

B. Baseline enforcement of supplier lists' declarative rules

Our intervention is rather inexpensive, simple, and yet yields significant changes in a very short period of time. In that regard, one may ask: what was the administration doing before to enforce declarative rules of the supplier lists? and why hadn't it implemented a protocol similar to our intervention before?

According to the information we gathered before the intervention, tax centers did invest some resources to make sure taxpayers would declare the lists, but rarely to enforce completeness, accuracy or consistency of the information they contained. Our interpretation is that this initial state is the result of several constraints.

First, before 2020, the vast majority of the firms submitted hard copy version of the lists (the online platform was made available only to a small percentage of large firms in 2019). This implies that tax inspectors could not run a code to simultaneously detect all anomalies of all submitted supplier lists in a couple of seconds. Detecting misreporting on a comparable scale without computer assistance would have been extremely costly. Now that online declaration of supplier lists has become the norm, this should be less of a constraint. Nevertheless, human resources trained to perform the data analysis necessary to rigorously detect misreporting is still scarce at the tax administration ([Czajka et al., 2022](#)).

Second, inspectors may have given priorities to other enforcement strategies. While supplier lists are a valuable source of information, especially when it is complete and accurate, their submission is not directly tight to a payment, as it is simply a recap of all transactions to service suppliers over the past year. In a situation where over 5% of taxpayers can still avoid submitting both their VAT and PAYE declarations (which are the most significant tax instruments), tax inspectors might have deemed too high the opportunity cost of enforcing better reporting of supplier list.

Third, such a low level of enforcement is probably also due to the fact that the inspectors in charge of monitoring the clients who declare the supplier lists work in a tax centers which are

different from the ones to which are assigned the possibly under-reporting suppliers. In other words, by enforcing a better declaration of the supplier lists, tax inspectors of the LTC, MTC-1, MTC-2 and RSC are contributing to the collection of information which will mostly benefit tax collection by other centers. This hypothesis is confirmed by our analysis of the distribution of formal yet under-reporting taxpayers suppliers across tax centers. Out of the 2795 suppliers identified as under-reporting according to information declared by the treatment group, 1.54% are registered in the LTC, 0.5% in the MTC-1, 2.4% in the MTC-2 and 6 % in the RSC. As for the control group, the figures are 1%, 0.7%, 3% and 7.4% respectively.

Finally, our intervention generalized the use of mail merge technology - to send multiple personalised emails at once - which only one tax center had started using before, and still at a relatively small scale. In that regard, the intervention contributed to speed up the adoption of a digital tool which the administration would have most probably generalised in the following years.

C. Results on the Withholding Tax on Service (WTS)

Our intervention does not specifically alert taxpayers about unlawful past behaviors related to the monthly declaration and/or remittance of the WTS. Yet, by singling out misreported suppliers lists on a case by case basis, the message demonstrates to taxpayers that the TA scrutinises declarations whose content should be consistent with monthly declarations and remittance of the WTS. In addition, the message explicitly recalls all fiscal rules related to the WTS, thus possibly informing some taxpayers about those rules.³⁰ Considering this, and keeping everything else constant,³¹ we should expect a positive impact on WTS declarations and remittances on the sub-sample of under-withholding taxpayers, may they be strategic or ignorant.

³⁰We did collect anecdotal evidence from a few treated taxpayers which seemed to ignore the necessity for them to withhold 5% on transactions made to informal service providers.

³¹Other factors could also play a role but cannot be explored in our setting. First the expected cost of non-complying to the WTS rules may exceed its expected returns. Our intervention essentially consists in reminding firms about a few fiscal and declarative rules and warning them to do better in the future but no audit was conducted on the basis of past misreporting. In this regard, the increase in expected non-compliance cost may be too mild to elicit a significant change in firms behavior, especially for firms not duly withholding the WTS. Also, even though it does not capture entirely the expected cost of non-compliance, the fine for misreporting suppliers lists might be too small (XOF200,000, i.e USD 330) for our intervention to be sufficiently deterrent on that margin. Second, firms previously under-withholding may avoid remitting additional WTS by simply ending contracts with informal suppliers, and possibly replacing them by formal ones. This unfortunately is impossible to verify with our data since identifying formal and informal suppliers requires information whose presence is affected by our intervention. For instance, we are not able to determine whether the ex-post appearance of a new ID among suppliers of the treated group indicates a new trading relationship or simply results from the incomplete information ex-ante.

Unfortunately, the information available in 2018-2020 suppliers list does not allow to identify under-withholding firms ex-ante, due to the widespread misreporting our intervention precisely helps correcting. Because of this, we also cannot assess the prevalence of under-withholding, hence the relative size of the group of taxpayers whose withholding behavior should be affected. Nevertheless, we can proxy under-withholding by *low levels* of withholding ex-ante. We compute the share of remitted WTS in total revenue, before the intervention, and use this variable to split our sample into four quartiles. By construction, lower quartiles will capture two types of taxpayers: (a) those who duly comply with WTS rules but resort to few informal suppliers (relative to their total revenue), and (b) those that have a large informal service supply but withhold less than what they ought to. As a result, exploring treatment heterogeneity by quartile allows to better isolate the group of taxpayers most likely to evade the WTS. Table 1.917 displays estimates of average treatment impact on monthly WTS declaration and remittances, aggregated at the year level, both at the extensive margin (whether some positive amount is declared and/or remitted at least once during the year) and intensive margin (how much is declared and/or remitted in total). We find no significant effect for the full sample and significant effects only for the lowest quartile. Firms of this group are indeed more likely to submit at least one monthly WTS declaration during the year (+6.0 p.p in 2022) and remit some positive amount (+4.6 p.p in 2022). At the intensive margin, they also declare a greater WTS tax base and remit larger WTS amounts on average (Table 1.917 column (3) and (4)). The magnitude of the impacts at the intensive margin is however difficult to interpret given significant response at the extensive margin and large share of zeros (Chen and Roth, 2023). All effects are more pronounced for 2023, most likely because the full communication campaign extends, and thus reinforces treatment, until December 2022. Our interpretation is that, given that the intervention is unlikely to have such effects on firms that duly comply with WTS rules but have a small informal service supply, these results must derive from the subgroup of firms previously under-withholding.

Additional evidence shows that the intervention contributed to increase the consistency between yearly suppliers lists and monthly remittances of the WTS. The absence of results, for the full sample, on total WTS remitted contrasts with the fact that the share of suppliers for which a positive WTS is declared in 2022 suppliers list increases by 3.6 percentage points (i.e 7%). Table 1.916 further highlights the fact that total WTS (ihs) declared in 2022 suppliers list also increases on average. If, according to results of the full sample specification of Table 1.8, firms overall did not declare nor remit more on a monthly basis during 2022, they have at least declared more WTS on average in the 2022 yearly suppliers list. The question is whether these newly declared amounts correspond to remittances that have already been processed or that are still pending. Evidence from Table 1.916 suggests that the first option is more likely. Indeed,

conditional on declaring 2022 suppliers list, the proportion of firms that declare some positive WTS amount in their in suppliers list increases by 6.1 percentage points but the proportion of firms actually remitting some during 2022 is not significantly affected. At the same time, the proportion of firms that remit some, but declare none, decreases by 4.4 percentage points and the proportion of firms that declare *and* remit some increases by 5.4 percentage points, while the share of those that declare without remitting remains stable. Overall, the vast majority firms newly declaring WTS in the their 2022 suppliers list had already remitted the tax throughout the year.

D. Impact on suppliers

In a second model, we estimate the average (indirect) impact for suppliers j on outcome y_j of being a supplier of some treated client ($IndirectTreat_j$). We restrict our estimation to firms that could be identified with a *firm-ID* as suppliers of a treated client *before* the intervention (in 2018, 2019 and/or 2020).

$$y_j = \alpha + \beta IndirectTreat_j + \gamma X_j + \delta_s + \epsilon_j \quad (1.2)$$

Outcome y_j must be measured without using any element of ex-post suppliers list data which otherwise could lead to intervention-induced selection of suppliers. X_j are control variables defined at the suppliers level. A complication here comes from the fact that some suppliers might have made some transactions with clients from both the treatment and control group. Larger, more connected, firms will typically fall into that class. To estimate equation 2.2, a first solution consists in simply removing those suppliers. Unless “more connected” firms are on average more connected to one of the intervention group specifically (which should have been avoided via the randomisation) this selection should not bias our estimation, yet it would imply that our results do not apply to all suppliers. An alternative method would be to measure the share of supplier’s j revenue paid by treated clients as a proportion of j ’s total profit according to the all transactions in suppliers lists, and then either use a 50% cut-off to assign suppliers to the indirect treatment group.

We test both selection strategies to estimate the impact on revenue declared in 2022 and 2023 by indirectly treated suppliers. Table 1.912 displays results from OLS estimation of equation (2). We do not find any significant impact neither at the extensive margin, nor at the intensive margin. We further segment our estimation by decile of average revenue declared over 2018-2020, and still find no significant impact (1.913. On outcome measuring the extensive margin (whether a supplier declares or not), our standard error are reasonably low, which suggest the

absence of detected effect is not a matter of statistical power. Overall this strongly suggests that tax registered suppliers of treated firms did not become more compliant by fear being targeted by follow-up deterrent measures. This contrasts with results obtained in high income countries, where third-party reporting is considered an efficient enforcement tool ([Kleven et al., 2011](#)).

Table 1.912 – Impact on Revenue declared by Suppliers identified ex-ante

	2022		2023	
	Submits a declaration	Revenue declared	Submits a declaration	Revenue declared
	(binary)	(ihs)	(binary)	(ihs)
	(1)	(2)	(3)	(4)
Restricted Sample	0.006 (0.013) [0.594] 5645	0.056 (0.256) [9.849] 5645	0.004 (0.011) [0.226] 5645	0.086 (0.195) [3.559] 5645
Full Sample	0.005 (0.011) [0.631] 6853	0.014 (0.235) [10.822] 6853	0.007 (0.010) [0.272] 6853	0.094 (0.194) [4.478] 6853

Notes. This table displays the average treatment impact, at the suppliers level, on the probability to declare revenue ((1) and (3)) and total revenue declared ((2) and (4)) after the intervention. Outcome of column (1) and (2) are measured based on self-declared 2022 income tax (PIT or CIT) due in April 2023. For 2023, such statements will be submitted in April 2024, hence we measure total revenue until June 2023 by taking the sum of revenue declared in monthly value added tax statements. The statistical unit here is the supplier level. Our sample consists of taxpayers that have been identified as formal suppliers of clients over the period 2018-2020, based on information reported in suppliers lists submitted by clients from the control or the treatment group. For the restricted sample we keep only suppliers that have traded either exclusively with the treatment group, or exclusively with the control group. For the full sample specification, we add suppliers that have traded with both group by assigning each of them to the intervention group form which they received the largest amount. For each quadruplets after the intervention, the first figure is the coefficient of interest, the last one is the number of observations, standard errors are in parenthesis and control group averages between brackets.

E. The Senegalese Tax System

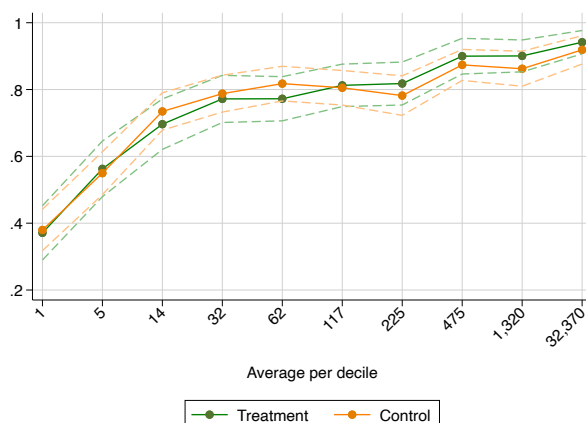
Senegal has three types of income tax: the corporate income tax (CIT), the personal income tax (PIT) and a presumptive tax on small business ("Contribution Global Unique" - CGU). The corporate tax is a 30% tax on business net profit - it applies only to moral persons.

The PIT has a progressive schedule with marginal rates ranging from 18% to 40% and that, in theory, should apply to the sum of all types of personal income (business income, rents, dividends, wages etc.). In practice however, most taxpayers submitting a PIT only declare either business or rental income. Wages, dividends and other source of personal income are never declared directly by their recipients but subject to withholding taxes collected and remitted by firms. Individuals with total taxable income below XOF 630,000 (USD 1008) are exempt from PIT. A strict interpretation of the tax code would imply that all physical persons, including those with taxable income below the threshold, should submit a yearly PIT statement with all their income. But the administration does not enforce that rule as it would not have the capacity to absorb that many declarations. Therefore, in practice, the default rule as a physical person in Senegal is not to declare ([Bachas et al., 2023](#)). Also, even though one needs to be classified as a physical person to be subject to the personal income tax, several taxpayers falling into this category are firms in the economic sense, with employees, registered under the identity of their owner.

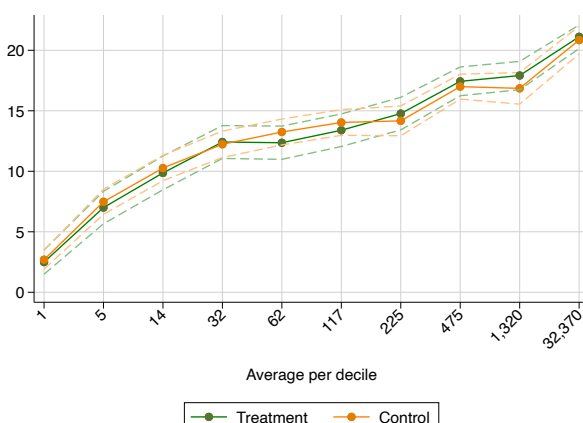
Taxpayers contributing to the PIT or the CIT, must also declare the VAT and two local business taxes. The CGU is a simplified tax regime that allows firms with total turnover below USD 80,000 to be discharged of income tax, VAT and local business taxes all at once. It is equal to 5% of total sales if the taxpayer is a service provider, or 2% if she is doing trade or manufacturing. At first, the withholding tax on service supplier declared in the yearly supplier lists does seem to correspond to the CGU for service supplier. But it is not labelled this way, and there is no corresponding withholding tax on suppliers operating in trade or manufacturing. Interestingly also, service providers operating in the hospitality sector are exempt from the withholding tax on services, but not from the CGU. Transaction below XOF 25,000 are exempt from the WTS, yet there is no exemption threshold for the CGU. The WTS was introduced as a way to tax unregistered individuals working for firms without being employed by them. The CGU was introduced after the WTS, to simplify administrative procedures.

Figure 1.913 – Spillover Effects on Suppliers’ Fiscal Behavior - Heterogeneity by Decile of Average Revenue over 2018-2020

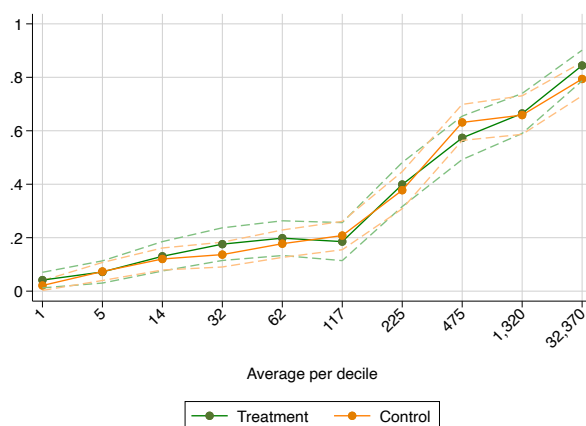
(A) Submits a 2022 Income Tax Declaration



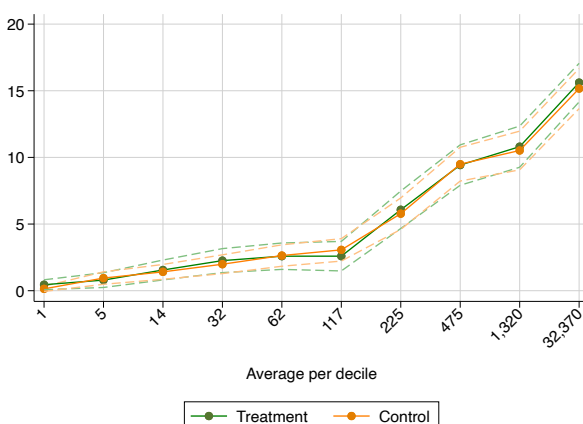
(B) Revenue Declared in 2022 Income Tax Declaration (ihs)



(C) Submits a 2023 VAT Declaration



(D) Revenue Declared in 2023 VAT Declaration (ihs)



Notes: This Figure decompose estimates of Table 1.912 by experimental group and decile of revenue declared. The statistical unit is the supplier. Treatment assignment of suppliers is based on treatment assignment of their client(s). Here we opt for the restricted sample, i.e. removing suppliers of both the treatment and the control group (see panel (A) of Table 1.912). For each decile, averages for the control group are the sample mean, averages for the treatment group are obtained by adding our average treatment estimates (using model specified in Appendix D) to the control group’s average. Dashed lines indicate 95% intervals. Outcome for Figures 1.913.(A) and 1.913.(B) are measured based on self-declared 2022 income tax (PIT or CIT) due in April 2023. Outcomes for Figures 1.913.(C) and 1.913.(D), are measured using value added tax statements declared in the first six months of 2023.

F Couldn't the client be strategic as well?

Throughout the paper, we assume that if the sum of all payments received by a supplier is superior to the total sales she has declared, it must be that the supplier has under-reported. Yet, it could be that clients are also strategic, either by giving wrong information or by over-declaring expenses.

Giving wrong information on the ID could happen accidentally, by making a mistake in one digit for instance. However firm-IDs must follow a strict structure which a random mistake is unlikely to catch. Therefore such mistakes will only make us classify some formal firms as *unidentified* because their firm-ID will most probably not match any of the registries. A more common source of misinformation is when clients simply indicate the *same* firm-ID for many of their suppliers - probably to simplify themselves the task of filling the list while hoping the administration would not bother look into the details. We systematically inspect those cases, which fortunately remain very rare, and classify them as "intentionally wrong ID". Interestingly, the intervention has no significant impact, neither positive nor negative, on the probability to adopt this reporting behavior. Finally, we could also imagine that clients claim to have traded with some formal firms, while in fact they have traded with another one (formal or informal). We cannot detect such cases. By doing so, the client takes the risk that the administration uncovers the fraud when auditing the supplier. If so, then this expenses would not be deductible anymore in the calculation of the CIT, therefore the client would have to pay 30% of that amount. Tax inspectors report never having found such cases, and believe they should be very rare in any case. It would be much simpler and safer to give a wrong ID, and later plead absent-mindedness. We can also imagine that some of the mismatch come from mistakes made in the numbers (adding one digit for instance). It is unlikely however that our intervention increase the prevalence of such mistakes, if anything we would expect the contrary. To limit this as much as possible we could explore how our results vary when restricting our analysis to payments not exceeding a certain percentage of client's total sales.

Now, clients could also be strategic by over-claiming expenses, this way reducing their liability both for the VAT and the CIT. By doing so, they would expose themselves to the above mentioned risk of non-deductible charges. Yet, if they believe the probability to be detected is low enough, they might still take that risk. While we must acknowledge that this could happen, we believe it should be less frequent than under-reporting by the supplier. In fact, tax inspectors report that as they find a discrepancy themselves, they systematically assume that the fraud comes from the supplier. The main justification for this is that the largest firms tend to be much more scrutinized by the administration. We document this by showing that the number of civil servants per taxpayers is significantly higher for the largest centers (see Figure 1.5). This pattern is even

more pronounced when restricting to civil servants of category A or B - the only ones entitled to perform audits and apply legal sanctions: over the period 2016-2021, small tax centers had less than two high-skilled civil servants per 100 taxpayers, while the Large Taxpayers Center had 6 or more. Based on this, we choose to systematically assume that the fraud is caused by the supplier. Eventually, it is crucial to keep in mind that even if in some cases the fraud may come from the client, this would still count as fraud detected thanks to the intervention, not as under-reported revenue but as over-claimed expenditures. Therefore it would not decrease the total fraud detected. If anything, the amount of the fraud would increase. Indeed, assume client A pays supplier B 200, but claims it paid 1000 - while supplier B declares total sales equal to 500. The discrepancy we observe is $(1000 - 500)$, that is how we compute under-reported amounts, but if a follow-up audit would reveal that supplier B received only 200 from A, then the over-claimed amount would become $(1000 - 200)$.

G. Detailed Message Content

English Translation

Subject: Request for information: rectification of summary statements

Dear Sir or Madam,

After systematic verification by our information teams, it has been established that the summary statements (annual and/or quarterly) of sums paid to Third Parties and as Rents that your company named **COMPANY NAME**, and registered under firm-ID n°XXXXXXXXXX filed for the financial years **XXXX, XXXX, XXXX, XXXX**, do not comply with the provisions of articles 200 and 201 of law 2012-31 of 31-12-2012, as amended, relating to the General Tax Code (GTC).

Indeed, for some of your service providers or lessors, no withholding tax has been declared and the identification information provided (national-ID and/or firm-ID) does not allow their fiscal status to be verified.

Therefore, under penalty of the sanctions provided for by the GTC, we invite you, in accordance with the provisions of article 626 of the aforementioned code, to submit, **within twenty (20) days, rectifying declarations of the summary statements** (annual and/or quarterly) for the aforementioned financial years, indicating in particular the amount of the withholding tax (put "0" if nil) and providing **reliable information** (firm-ID and/or national-ID) which will enable our services to verify their fiscal status.

We would ask you to take account of these shortcomings and omissions so that you do not make the same mistakes again, particularly with regard to the next deadline (the declaration of summary statements of sums paid to third parties/as rents for the 2021 financial year, due on 31 January 2022).

You will find more details on the legislation and reporting procedures in the appendix to this e-mail.

Yours faithfully

***** ADDITIONAL INFORMATION *****

Legislation :

- Under article 200 of the General Tax Code (GTC), a 5% withholding tax must be levied on all sums paid by a debtor established in Senegal to *natural persons* residing in Senegal as remuneration for the provision of services.
- As a reminder, the term "natural persons" is used here to refer to any individual carrying on an economic activity who is not effectively subject to a system of taxation based on actual profits (personal income tax), as well as any legal entity or group that has not opted for corporate income tax (article 200, paragraph 3 of the GTC).
- Under the provisions of article 201 of the GTC, the withholding tax is also applicable to rental income from premises leased by a debtor as defined in point 2 of article 200, to rental income paid by an individual tenant, and to rental income collected on behalf of third parties who are natural persons, by estate agencies, property managers and non-trading property companies.
- In addition, the summary statements of sums paid to third parties must include, among other information, the **firm-ID** or, failing that, the **national-ID** of each service provider (article 200, paragraph 8.b of the CGI). Please note that all persons (legal entities or individuals) actually subject to income tax have a firm-ID.

Submitting your amending tax returns:

- To make things easier for you, you can **use the e-services portal** to submit your corrective declaration:
 - If you have already submitted a declaration via e-services for the financial year in question: ask your tax centre for assistance so that you can be authorised to resubmit a declaration for the same financial year.
 - If you only filed a paper declaration for the financial year in question, you can now submit an amended declaration via e-services.
- As far as possible, we invite you to **use your old declaration** (already completed) **and to make the necessary changes** to bring it into line with the above-mentioned tax aspects. Please note that any declaration indicating a false or incorrectly allocated identifier (firm-ID/national-ID) will be systematically detected and a new corrective declaration will be requested.

Finally, we would like to draw your attention⁸⁹ to the fact that, as part of the Yataal programme for a broader and fairer distribution of the tax burden, it is crucial that companies such as your service providers contribute in line with their income. As the client of these service providers,

MINISTRE DES FINANCES ET DU BUDGET

Le Chef de Centre



DIRECTION GENERALE
DES IMPOTS ET DES DOMAINES

A

DIRECTION DES MOYENNES ENTREPRISES

CENTRE DES MOYENNES ENTREPRISES DAKAR 1

BUREAU DE GESTION ET DES SERVICES AUX
CONTRIBUABLES

~~005782400~~

~~005782400~~

Objet : Demande de renseignements : rectifications d'états récapitulatifs

Madame, Monsieur,

Après vérification systématique par nos équipes du renseignement, il a été établi que les états récapitulatifs (annuels et/ou trimestriels) des sommes versées à des Tiers et/ou au titre des Loyers que votre société dénommée ~~SA ZAHARA~~ et immatriculée sous le NINEA n° ~~00097406~~ a déposé pour les exercices 2018 ne sont pas conformes aux dispositions des articles 200 et 201 de la loi 2012-31 du 31-12-2012, modifiée, portant Code Général des Impôts (CGI).

En effet, pour certains de vos prestataires ou bailleurs, aucune retenue à la source n'a été déclarée et les informations d'identification fournies (CNI et/ou NINEA) ne permettent pas de vérifier leur régime fiscal.

Aussi, sous peine des sanctions prévues par le CGI, nous vous invitons, conformément aux dispositions de l'article 626 du code sus visé, à soumettre, dans un délai de **vingt (20) jours, des déclarations rectificatives des états récapitulatifs** (annuels et/ou trimestriels) pour les exercices sus visés, en indiquant notamment, le montant de la retenue à la source (mettre « 0 » si nulle) et en fournissant des **informations fiables** (NINEA et/ou CNI) qui permettront à nos services de vérifier leur régime fiscal.

Nous vous invitons à prendre en compte ces manquements et omissions pour ne plus reproduire les mêmes erreurs. Notamment en ce qui concerne l'échéance actuelle (la déclaration des états récapitulatifs des sommes versées à des Tiers/ au titre des loyers pour l'exercice 2021, due le 31 janvier 2022), nous vous invitons à soumettre une déclaration rectificative, le cas échéant.

Vous trouverez en annexe de ce mail plus de détails sur la législation et les modalités déclaratives. Nous vous prions d'agréer, Madame, Monsieur, l'expression de mes salutations distinguées



Éléments de législation :

- En vertu de l'article 200 du Code Général des Impôts (CGI), une retenue de 5% doit être prélevée sur toutes les sommes versées par un débiteur établi au Sénégal, à des *personnes physiques* résidant au Sénégal en rémunération de prestations de service.
- Pour rappel, on désigne par « personnes physiques » ici tout exploitant individuel exerçant une activité économique non effectivement soumis à un régime d'imposition d'après les bénéfices réels réalisés (l'impôt sur le revenu des personnes physiques), ainsi que toutes les personnes morales ou groupement n'ayant pas opté pour l'Impôt sur les Sociétés (article 200, alinéa 3 du CGI).
- Aux termes des dispositions de l'articles 201 du CGI, la retenue est également applicable aux loyers des locaux pris à bail par un débiteur tel que défini au point 2 de l'article 200, aux loyers payés par un locataire individuel, ainsi qu'aux loyers encaissés pour le compte de tiers personnes physiques, par des agences immobilières, des gérants de biens et des sociétés civiles immobilières.
- Par ailleurs, les états récapitulatifs des sommes versées à des Tiers doivent mentionner, entre autres informations, le **NINEA**, ou à défaut le **numéro de la pièce d'identité** de chaque prestataire (article 200, alinéa 8.b du CGI). Notez à ce titre que toutes les personnes (morales ou physiques) effectivement assujetties au régime du réel ont un NINEA.

Soumettre vos déclarations rectificatives :

- Afin de vous faciliter la démarche, vous pouvez **utiliser le portail eservices** pour soumettre vos déclarations rectificatives :
 - Si vous avez déjà soumis une déclaration via eservices pour l'exercice en question : demandez de l'aide à votre centre fiscal pour être autorisé à soumettre de nouveau une déclaration pour le *même* exercice.
 - Si pour l'exercice en question n'aviez fait la déclaration que sur papier, vous pouvez dès à présent soumettre une déclaration rectificative via eservices.
- Nous vous invitons autant que possible à **utiliser votre ancienne déclaration** (déjà remplie) **et d'y ajouter les modifications nécessaires** afin de la rendre conforme aux éléments de fiscalité sus visés. Notez que toute déclaration indiquant un identifiant (NINEA/CNI) faux ou mal attribué sera systématiquement détectée et une nouvelle déclaration rectificative vous sera exigée.



Enfin, nous attirons votre attention sur le fait que, dans le cadre du **programme Yataal** pour une répartition plus large et juste de l'effort fiscal, il est crucial que les sociétés comme vos prestataires contribuent à hauteur de leurs revenus. En tant que débiteurs de ces prestataires, votre société a un rôle à jouer, non seulement en prélevant et en déclarant la retenue à la source lorsqu'elle est due et en remplissant correctement ses états récapitulatifs, mais aussi en rappelant à ses prestataires qu'un assujettissement au réel les exonère de la retenue à la source de 5%.

Le chef de centre



04/02/2022



H. Complementary Figures and Tables

Table 1.913 – Summary Statistics and Balance tests - Conditional on Declaring 2022 Suppliers List (N = 2,633)

Source	Covariates/Outcomes	Summary Statistics			Balance tests	
		Mean	Std. dev.	Median	β	Std. err.
(A) Suppliers lists	Number of suppliers	46.211	171.930	12	-7.786	6.030
	Share suppliers without any ID	0.867	0.215	1	-0.008	0.006
	... with a positive WTS	0.123	0.238	0	-0.000	0.008
	... in the Hospitality sector	0.001	0.012	0	-0.000	0.000
	... Senegalese	0.987	0.050	1	0.000	0.001
	... with transaction bellow taxable threshold	0.018	0.068	0	-0.000	0.002
	N. of y. for which we observe a suppliers list	1.838	0.798	2	-0.010	0.023
	Declares suppliers list online	0.384	0.486	0	0.001	0.013
(B) Other Declarations	Av. Sum of WTS remitted - 2018-2020 (USD'000)	4.263	16.494	0.782	-0.748	0.610
	Av. turnover - 2018-2020 (USD'000)	6800.234	34836.098	547.211	-1674.191	1275.619
	Av. number of employees - 2018-2020	66.930	358.904	8.222	-29.915**	14.335
(C) Tax Registry	Large Taxpayers Unit (binary)	0.269	0.443	0	-2.242	2.504
	Upper-Middle Taxpayers Unit (binary)	0.132	0.339	0	4.111	2.780
	Upper-Middle Taxpayers Unit (binary)	0.255	0.436	0	-2.087	1.799
	Regulated Sector Unit (binary)	0.342	0.474	0	1.988	1.488
	Legal person (binary)	0.733	0.442	1	0.002	0.013
	Number of years since first digital admin. record	11.600	3.189	13	-0.091	0.122
	Firm created before 1990 (binary)	0.315	0.464	0	-0.005	0.017
	... btw 1990 and 2000 (binary)	0.083	0.277	0	0.001	0.010
	... btw 2000 and 2010 (binary)	0.281	0.449	0	-0.000	0.017
	... after 2010 (binary)	0.318	0.466	0	0.004	0.018

Notes. The first three columns provide summary statistics about covariates and outcomes we use throughout the study for the entire sample. The last two columns summarize results from OLS estimation of equation (1) (see Section 1.4.2) without covariates, with each variable as the outcome. Variables of panel (A) are measured using suppliers lists collected at baseline (average over the period 2018-2020). Panel (B) exploits information reported in other declarations such as monthly Withholding Tax on Services (line 1), or Personal Income Tax, Corporate Income Tax, Value Added Tax, Business Property Tax, and the Synthetic Tax for small Businesses (line 2), or Withholding on employees (line 3). Panel (C) exploits all relevant information we could retrieve from the tax registry.

Table 1.914 – Summary Statistics and Balance tests - Conditional on Declaring 2021 Suppliers List (N = 2,608)

Source	Covariates/Outcomes	Summary Statistics			Balance tests	
		Mean	Std. dev.	Median	β	Std. err.
(A) Suppliers lists	Number of suppliers	45.997	172.999	11.166	-7.709	6.116
	Share suppliers without any ID	0.867	0.215	1	-0.008	0.006
	... with a positive WTS	0.128	0.241	0	-0.004	0.008
	... in the Hospitality sector	0.002	0.014	0	-0.000	0.000
	... Senegalese	0.987	0.051	1	0.000	0.001
	... with transaction bellow taxable threshold	0.018	0.065	0	-0.000	0.002
	N. of y. for which we observe a suppliers list	1.878	0.793	2	-0.017	0.024
	Declares suppliers list online	0.406	0.491	0	0.000	0.013
(B) Other Declarations	Av. Sum of WTS remitted - 2018-2020 (USD'000)	4.796	25.873	0.751	-1.197	1.014
	Av. turnover - 2018-2020 (USD'000)	6978.232	35199.960	543.427	-1439.344	1293.853
	Av. number of employees - 2018-2020	68.479	363.883	8.116	-28.367*	14.694
(C) Tax Registry	Large Taxpayers Unit (binary)	0.274	0.446	0	1.006	1.778
	Upper-Middle Taxpayers Unit (binary)	0.141	0.348	0	3.641	9.668
	Upper-Middle Taxpayers Unit (binary)	0.205	0.404	0	-1.195*	6.478
	Regulated Sector Unit (binary)	0.378	0.485	0	1.404	9.982
	Legal person (binary)	0.712	0.452	1	-0.001	0.013
	Number of years since first digital admin. record	11.547	3.253	13	-0.086	0.126
	Firm created before 1990 (binary)	0.316	0.465	0	0.002	0.017
	... btw 1990 and 2000 (binary)	0.083	0.277	0	-0.002	0.010
	... btw 2000 and 2010 (binary)	0.277	0.447	0	-0.009	0.017
	... after 2010 (binary)	0.322	0.467	0	0.010	0.018

Notes. The first three columns provide summary statistics about covariates and outcomes we use throughout the study for the entire sample. The last two columns summarize results from OLS estimation of equation (1) (see Section 1.4.2) without covariates, with each variable as the outcome. Variables of panel (A) are measured using suppliers lists collected at baseline (average over the period 2018-2020). Panel (B) exploits information reported in other declarations such as monthly Withholding Tax on Services (line 1), or Personal Income Tax, Corporate Income Tax, Value Added Tax, Business Property Tax, and the Synthetic Tax for small Businesses (line 2), or Withholding on employees (line 3). Panel (C) exploits all relevant information we could retrieve from the tax registry.

Table 1.915 – Conditional Impact on Misreporting, Information, Fraud Detection - comparing 2021 and 2022 suppliers lists

		2021						2022					
		β	Std. err.	Lee Bounds		μ_c	N	β	Std. err.	Lee Bounds		μ_c	N
				lower	upper					lower	upper		
(A) Misreporting	Misreports no supplier (binary)	0.059***	0.014	0.012	0.074	0.158	2608	0.101***	0.015	0.089	0.109	0.182	2633
	Misreports all suppliers (binary)	-0.143***	0.015	-0.197	-0.135	0.304	2608	-0.195***	0.014	-0.214	-0.194	0.289	2633
	Misreports some suppliers (binary)	0.084***	0.018	0.060	0.122	0.536	2608	0.093***	0.018	0.085	0.105	0.527	2633
	Share of suppliers misreported suppliers	-0.163***	0.015	-0.201	-0.138	0.587	2608	-0.233***	0.014	-0.248	-0.228	0.543	2633
(B) Information prevalence	Share of suppliers ...												
	... with a formally valid ID	0.186***	0.015	0.163	0.226	0.415	2608	0.247***	0.014	0.243	0.263	0.469	2633
	... with intentionally wrong ID	-0.002	0.005	-0.031	-0.001	0.031	2608	-0.003	0.005	-0.019	-0.003	0.026	2633
	... with positive WTS	0.018	0.014	-0.013	0.049	0.485	2608	0.036***	0.014	0.026	0.046	0.508	2633
	... satisfying cond. to be exempt of WTS	-0.000	0.003	-0.019	0.001	0.033	2608	0.001	0.003	-0.007	0.002	0.033	2633
	... with a telephone number	0.068***	0.016	0.035	0.098	0.383	2608	0.092***	0.016	0.084	0.104	0.410	2633
(C) Fraud Detection	Share of suppliers identified as ...												
	... informal	0.117***	0.012	0.075	0.137	0.194	2608	0.175***	0.012	0.163	0.182	0.229	2633
	... formal	0.068***	0.011	0.025	0.087	0.214	2608	0.069***	0.012	0.057	0.077	0.232	2633
	... formal and compliant	0.042***	0.010	-0.003	0.056	0.170	2608	0.043***	0.010	0.029	0.049	0.175	2633
	... formal and under-reporting	0.026***	0.005	-0.004	0.030	0.044	2608	0.025***	0.005	0.008	0.027	0.056	2633
(C) Structure of service supply	Number of suppliers	-0.662	5.615	-26.567	0.402	47.944	2608	4.590	7.203	-19.075	4.853	49.903	2633
	Total (ihs)	0.043	0.055	-0.303	0.230	17.841	2608	0.036	0.056	-0.095	0.111	17.934	2633
	Average (ihs)	0.029	0.043	-0.165	0.153	15.409	2608	-0.021	0.043	-0.098	0.033	15.472	2633
	Median (ihs)	0.011	0.043	-0.142	0.151	14.672	2608	-0.079*	0.043	-0.133	-0.018	14.765	2633
	Standard Deviation (ihs)	0.052	0.174	-0.284	0.936	13.783	2608	0.105	0.171	0.014	0.425	13.928	2633
	Declares suppliers list	0.049***	0.013			0.750	3487	0.017	0.013			0.772	3487

Notes. This table displays results on outcome measured in 2021 and 2022 suppliers lists, submitted in January 2022 and January 2023, respectively. We use OLS method to estimate the model specified in 1.4.2, with variables displayed in Table 1.6 as controls. Besides of the last line, which estimates treatment impact on the probability to declare a suppliers list, all results are conditional on declaring a suppliers list, for each year separately. Outcome variable of panel (A) capture impact on misreporting of suppliers list. Outcome variables for panel (B) and (C) measure the share, at the client level, of some specific type of line/suppliers. For each client, i , outcome variables of panel (D) measure properties of the distribution of payments made by client i . We provide conservative lower and upper bounds for our estimates to account potential biases due to selection.

Table 1.916 – Comparing WTS declared in Yearly Suppliers List with monthly WTS Remittances - 2022

		β	Std. err.	Lee Bounds		μ_c	N
				lower	upper		
(A)	Share of suppliers with WTS > 0	0.036***	0.014	0.026	0.046	0.508	2633
	Share of suppliers exempt of WTS	0.001	0.003	-0.007	0.002	0.033	2633
(B)	Total WTS declared (ihs)	0.870***	0.221	0.711	1.079	10.422	2633
	Total WTS remitted (ihs)	0.116	0.209	-0.041	0.320	11.288	2633
(C)	Declares some WTS	0.061***	0.015	0.057	0.077	0.729	2633
	Remits some WTS	0.009	0.014	0.005	0.025	0.791	2633
	Remits some but declares none	-0.044***	0.011	-0.062	-0.042	0.111	2633
	Declares some but remits none	0.007	0.008	-0.010	0.009	0.049	2633
	Declares <i>and</i> remits some	0.054***	0.016	0.048	0.067	0.680	2633

Notes. This table displays results on outcome measured in 2022 suppliers lists, and aggregated data from monthly WTS remittances. We use OLS method to estimate the model specified in 1.4.2, with variables displayed in Table 1.6 as controls. Outcome variable of panel (A) measure the share, at the client level, of some specific type of line/suppliers. Panel (B) allows to compare the total of WTS declared to have been withheld according to yearly suppliers list with the total of WTS effectively remitted on a monthly basis throughout 2022. Outcome variables for panel (C) are all binary. We provide conservative lower and upper bounds for our estimates to account potential biases due to selection.

Table 1.917 – Treatment Heterogeneity on Monthly Declaration and Remittance of the Withholding Tax on Informal Service Suppliers (WTS)

	Quartile of WTS share in revenue (av. 2018-2020)	Has submitted at least once a monthly WTS (binary) (1)	Yearly Total paid to informal service suppliers (ihs) (2)	Has remitted at least once some WTS (binary) (3)	Yearly Total remitted (ihs) (4)
2022	1	0.060* (0.031) [0.527] {872}	1.104*** (0.463) [5.876] {872}	0.046 (0.028) [0.327] {872}	0.694* (0.388) [4.472] {872}
	2	-0.003 (0.017) [0.933] {872}	-0.326 (0.379) [14.870] {872}	-0.002 (0.024) [0.834] {872}	-0.225 (0.339) [11.812] {872}
	3	0.010 (0.015) [0.939] {872}	-0.138 (0.366) [15.299] {872}	0.000 (0.022) [0.868] {872}	0.091 (0.325) [12.212] {872}
	4	-0.006 (0.026) [0.790] {871}	0.296 (0.519) [11.968] {871}	0.009 (0.030) [0.666] {871}	0.206 (0.440) [9.738] {871}
2023	1	0.079*** (0.031) [0.472] {872}	1.508*** (0.447) [4.900] {872}	0.075*** (0.028) [0.281] {872}	0.984*** (0.366) [3.731] {872}
	2	-0.027 (0.022) [0.885] {872}	-0.510 (0.436) [13.124] {872}	-0.011 (0.028) [0.753] {872}	-0.399 (0.380) [10.336] {872}
	3	0.000 (0.021) [0.886] {872}	-0.198 (0.429) [13.563] {872}	-0.028 (0.027) [0.799] {872}	-0.334 (0.378) [10.767] {872}
	4	-0.007 (0.029) [0.719] {871}	-0.027 (0.533) [10.526] {871}	-0.016 (0.032) [0.598] {871}	-0.052 (0.453) [8.338] {871}

Notes. This table displays results from OLS estimation of equation (1) (see Section 1.4.2) with variables displayed in Table 1.6 as controls. on four different outcome variables, across quartiles of WTS share. The statistical unit is the taxpayer (treatment or control). We measure WTS share at baseline as the ratio between the sum of WTS effectively remitted to the tax administration throughout the year and total revenue. We average this measure over 2018-2020 and split our full sample in four accordingly. All outcome variables synthesize data at the year level. We exploit only the first 6 months of 2023. For each quadruplet we have : β -coefficient, (std. err), [control group's average], {N}

Chapter 2

Does Electronic Filing and Payment Increase Tax Compliance?

with Awa DIOUF and Fabrizio SANTORO

Abstract Tax administrations in low-income countries progressively electronic filing and payment technologies with the aim of reducing both enforcement costs for the administration and compliance costs for taxpayers. This shift also holds the potential to decrease opportunities for collusion or extortion by minimizing in-person interactions with tax officials. However, the available evidence on the impact of such technology remains scarce. Using a fuzzy dynamic double-difference method, we investigate the causal impact of the adoption of e-filing and e-payment by the largest taxpayers in Senegal. Our findings indicate either no, or very limited, impact on the probability to declare or pay. Nevertheless, we observe a significant decrease in the prevalence of missing values in digitized declarations. We also find a positive, and possibly large, impact on tax payments, but essentially concentrated on the largest taxpayers. Later in time, additional evidence finally suggests that by reducing compliance costs, online services enabled previously enrolled taxpayers to continue submitting tax statements even amidst the containment measures implemented in response to the COVID pandemic.

2.1. Introduction

Technology and good institutions are considered pillars of economic development. In that regard, the recent acceleration in the production and worldwide diffusion of innovations in information technology could contribute to foster the efficiency and transparency of low-income countries' (LICs) administrations, thereby hastening their convergence toward higher living standards. Accordingly, recent evidence shows that governments from low- and middle-income countries have in fact largely invested in the adoption of IT infrastructure. In 2000, only 31 countries had four or more of the nine main governmental management information systems identified under the World Bank's GovTech classification,¹ but by 2022, 163 countries had at least eight of them. More recently, governments have started to increasingly develop online services, building from existing management information systems to allow remote interactions between citizens and various state departments. This includes solutions to declare (*e-filling*) and pay (*e-payment*) taxes online. As of 2016, 32% of LICs had already adopted e-filling ([World Bank, 2016](#)). In the context of LICs, where revenues are too low to satisfyingly fund various under-provided public goods ([OCDE, 2022](#)), these tools are expected to significantly contribute to increase tax revenue. The underlying assumptions behind such expectations are that e-filling and e-payment reduce both enforcement costs for the administration and compliance costs for taxpayers, and limit opportunities for collusion or extortion by minimizing in-person interactions with tax officials. Relatedly, it is also expected to increase transparency and therefore trust in the fiscal institution. However, such positive impacts for taxation need not to materialise. Indeed, saved costs may be allocated to other objectives than becoming more compliant (for the taxpayers) or strengthening tax enforcement (for the administration). Collusion/extortion related impacts are conditional on the prevalence of such practices ex-ante among taxpayers now having access to online solutions. Finally, increased trust in the tax administration is not necessarily correlated to improved beliefs regarding state departments in charge of *spending* the collected revenues. Considering this, the consequences of e-filling and e-payment on tax compliance remain an empirical question.

This article provides causal evidence on the impact of jointly adopting e-filling and e-payment services in the context of Senegal. After years of development, this technology (*Etax*) was made mandatory by ministerial decree in July 2017 for all taxpayers monitored by the Large Taxpayers Center. This has induced 80% of the targeted taxpayers to effectively register and use the platform by January 2018. Combining newly collected administrative data from three different monthly tax instruments (value added tax (VAT), withholding tax on wages (PAYE)

¹See [GovTech Maturity Index \(GTMI\) Data Dashboard](#)

and withholding on service suppliers (WTS)) with a fuzzy, dynamic, double-difference method, we exploit exogenous variations in e-filing and e-payment adoption generated by the reform to investigate its causal impact on various indicators capturing tax compliance. To improve the comparability of the two groups, we restrict our control group to the subset of its largest taxpayers. While our setting presents some challenges for the identification strategy, it has the advantage of being very representative of how tax administrations would typically deploy new IT technology: starting by the largest taxpayers and gradually expanding to smaller, more loosely monitored ones (Ali et al., 2015). In our main specification, our sample is a balanced panel of 1,691 taxpayers observed during the 36 months of 2017-2019. We opt for a strictly balanced panel to avoid possible confusions between real and pure compositional effects.

First we document a short-lived but clear 2.5 percentage point negative impact on the probability to declare just after the ministerial decree. We interpret this as evidence of the transition costs faced both by the taxpayers and the administration to swiftly register hundreds of taxpayers into the new system and adapt previously entrenched practices. However in the following months, we do not find any significant results on the probability to declare. This is despite recovering standard errors that would have allowed detecting a minimum impact of about 2 percentage points.² However, the divide seems to widen overtime: point estimates become significant one year and a half after the reform. We find a similar, yet slightly more salient pattern for the probability to declare a positive tax liability. In light of this, we cannot reject that the reform may have had a lagged impact on such probabilities, but we believe this interpretation is rather speculative given the available evidence.

We do find sharp and very consistent 5-7 percentage points impact on the probability to submit a declaration containing missing values. Given that we do not find any impact on other measures of information quality (such as inconsistencies of implausible values), we would find it hard to interpret this as resulting from a change in the taxpayers behavior. Rather, we believe it most likely reflect the fact that e-filing contributes to make visible some values which data entry agents, needed before e-filing, were not taking the time to type in order to perform their task more rapidly.

Finally, the reform seems to have had no impact on average liability declared, nor on the probability to pay (the minimum detectable effect for the later being 3.3 percentage points). Nevertheless our evidence suggests that it did have a significant impact on average payment. Keeping the full sample of treated taxpayers, this impact could be as large as USD 20,000 per month on average. Yet sensitivity tests demonstrate that the magnitude of the estimates largely depend on the presence of extremely large taxpayers. Indeed, removing from the treatment

²Assuming a 5% significance level and a statistical power of 80%.

group the largest 5 taxpayers alone divides the estimate by 2. As additional observations get removed, the coefficient eventually converges toward USD 5,000. Under this specification, such an average monthly impact among treated firms would account for about 1.2% of all tax revenues collected in 2018. Unfortunately, available information does not allow to determinate whether this increase is driven by a drop in collusion or an increased capacity to enforce payment recovery.

Overall, our evidence demonstrate that the adoption of e-filing has no or very modest impact on taxpayers' declarative behavior in the short and mid term. This could be due to the fact that, before the reform, larger taxpayers were comparably more compliant along this dimension because they have been historically more scrutinized by the administration. In an alternative application of our dynamic diff-in-diff model, we exploit COVID related containment measures enforced in early 2020 as an exogenous source of variation in the cost of coming in person to administration. We find evidence that taxpayers who had not registered to *Etax* prior to the COVID shock were significantly less likely to submit a declaration afterwards. The timing of the most severe negative responses (- 4 percentage points) corresponds exactly to the months during which restrictions were the strictest, and full recovery takes about 18 month to materialise. By construction of our sample (a balanced panel of taxpayers active beyond our time window) and given that tax declarations can be submitted with delays, such trends cannot be explained by taxpayers permanently shutting down because of the economic downturn. This last result suggest that by lowering the cost of declaring, e-filing did contributed to sustain high rates of declaration even amidst the pandemic.

This paper contributes to the still very scarce the literature the role e-filing and e-payment technology on taxation in LICs. A few studies document the causal implications of electronic billing machines and electronic payment system on tax compliance (Bellon et al., 2022 for Peru, Mascagni, Mengistu, and Woldeyes, 2021 for Ethiopia, Brockmeyer and Sáenz Somarriba, 2022 in Uruguay and Das et al., 2023 for India). However lessons learned from that strand of the literature apply to a technology that digitizes transactions between taxpayers themselves (thus generating a third-party reported paper trail), not between taxpayers and the administration. To our knowledge, there exists only two studies exploring the causal impact e-filing and e-payment on tax compliance in a low/middle income country (Okunogbe and Pouliquen, 2022 for Tajikistan and Amine and Santoro, 2023 for Eswatini). Our study innovates from the later first by investigating the behavior of the largest taxpayers - by far the most strategic sub-sample of taxpayers when considering aggregate tax revenue (Milanez, 2017 Slemrod and Velayudhan, 2018). Second, our use of high frequency monthly data that allows us to document the evolution of treatment impact in the short and mid term. We provide new causal evidence that in the

very short term, the impact can indeed be negative due to the transition costs as hypothesized by [Scarpini, Okunogbe, and Santoro, 2023](#). This paper is the first to provide evidence on the positive impact for information quality - a result consistent with findings related to the adoption of electronic billing machines ([Mascagni, Mengistu, and Woldeyes, 2021](#)). Finally, despite null or very modest impacts on declarative compliance, our results do suggest that e-payment solutions could be instrumental to increase average effective payment by the largest taxpayers, and thus generate possibly very significant revenue gains.

The rest of the paper is structured as follows: section [2.2](#) documents the context, section [2.3](#) specifies our assumptions on taxpayers and administrators' behavior as well as the main expected channels, section [2.4](#) describes our data and empirical strategy, section [2.5](#) discusses the results and section [4.6](#) concludes.

2.2. Institutional Context

The Senegalese Tax Administration (TA) directly monitors about 30,000 taxpayers, one third of which are legal person (including incorporated firms, associations, regulating agencies etc.), others being self-employed individuals. These taxpayers are distributed across 20 tax centres according to business size, sector, and location. Taxpayers with a turnover exceeding USD 3,200,000 for three consecutive years are managed by the Large Taxpayer Centre (LTC), irrespective of their headquarters' location. Smaller taxpayers operating within the region of Dakar are distributed across two medium-sized taxpayers centres (MTC1, MTC2), one unit for firms operating in regulated sectors (RSC) and seven municipal tax centres.³ The rest of the taxpayers are spread across 11 regional centres.

Since 2006, the TA has adopted and progressively deployed an integrated digital tax management system. This multi-module software allows to store tax declarations, compute tax liability, process payments and produce receipts. Under this system, tax declarations are submitted in paper format, mostly by taxpayers themselves and not by mail, to their respective tax centre, and then digitized into the system by data entry agents, and possibly corrected afterwards in case data entry mistakes had been identified. Taxpayers would also have to come to their tax centre in order to pay, mostly in cash or cheque.

Starting in early 2017, the TA began to implement and test a new software (Etax), geared to

³Regulated sectors are medical and legal professions, accountants, architects, state agencies and N.G.O.. Firms within the MTC1 (MTC2) have a turnover between USD 1,600,000 and USD 3,200,000 (USD 800,000. USD 1,600,000). Sector prevails over size for fiscal assignment. Dakar's seven municipal centres are Dakar-Plateau, Ngor-Almadies, Parcelles Assainies, Dakar Liberté, Grand Dakar, Pikine-Guédiawaye and Rufisque.

the existing IT infrastructure, allowing taxpayers to declare online (e-filing) and pay online (e-payment). In July 2017, the Senegalese Minister of Economics and Finance released a decree making mandatory for taxpayers monitored by LTC to use Etax. Most of them adopted it in early January 2018, the rest gradually joining by the end of 2018. During the Fall of 2018, firms from the upper-middle taxpayers centre (MTC1) were also embarked into the new system. About 80% joined, even though there was no legal obligation to do so. The adoption of Etax was subsequently encouraged among taxpayers from two other centres (MTC2 and RSC), but with very limited take-up. The reform had multiple objectives: i) spare taxpayers the cost of moving to their centre to declare and/or pay, ii) spare the administration the cost of handling all paper declarations, iii) improving the quality of the information by removing mistakes introduced by data entry agents. Tax officials report that, especially in the first years of implementation, crashes of the platform due to network saturation close to fiscal deadlines sometimes prevented taxpayers to declare and or pay on time. Nevertheless, as of 2022, the new infrastructure was considered to be robust enough to start systematically embarking more and more tax centres. In order to strictly conform to the declarative nature of the tax law, e-filing through Etax mimics as closely as possible paper declaration: only identification information are pre-filled, inconsistencies as well as missing values are allowed and not automatically flagged. There is no automatic reminder for missing declaration. With a few minor exceptions, all type of taxes can be declared through Etax. Tax liability is then automatically computed by the system based on the information reported by the taxpayer. In case a taxpayers realises it has made a mistake, it must address a correction request which is then handled by the tax centre (taxpayers cannot correct their declaration themselves). Taxpayers can consult their data at all time. Then e-payment occurs through the banking system and as such requires taxpayers to have and share a bank account number. In that regard, tax officials report that some taxpayers of the LTU were initially reluctant to adopt Etax due to concerns about sharing their bank account details and that this was part of the motivation for releasing a ministerial decree.

2.3. Assumptions and Expected Impacts

The introduction of e-filing and e-payment is likely to affect tax compliance through multiple mechanisms, both positively, or negatively, with immediate effect or only in the long run. In this section we synthesize what mechanisms we believe should play a role in this study, based on our understanding of the context and the technology's specifics, theoretical assumptions on the behavior of the various agents involved as well as available evidence from the previous empirical evidence.

2.31. General behavioral assumptions

On the behavior of taxpayers: In accordance to the now classic deterrence model for tax evasion first developed by [Allingham and Sandmo, 1972](#), we assume that taxpayers mostly act as free-riders. In other words, absent any risk associated with tax evasion (bad reputation or legal sanctions), economic actors would simply not pay taxes. This assumption is broadly aligned also with the existing empirical literature whose result mostly converge to the conclusion that, at least in the short-term, deterrent measures are significantly more effective at inducing compliance than those interventions appealing to the intrinsic value of tax morale, or the need to fund public goods (see [Luttmer and Singhal, 2014](#) and [Slemrod, 2019](#)).

On the behavior of tax collectors: Following a now rich corpus of work that builds from the seminal work of [Weber, 1922](#) we assume that the principal-agent model is an adequate framework to reflect on the behavior of bureaucrats. In our specific setting, the principal will typically be a higher-tier bureaucrat (or a minister up the chain), and the agents will be civil servants under her supervision, in charge of performing tasks contributing to tax collection, the final objective of the fiscal administration. The agents decide upon the level of effort they invest to perform the tasks they have been assigned to, and in most cases it is reasonable to assume that this level is not observed by the principal. The latter therefore needs to set up the right incentives (promotion/relegation, bonus/sanctions) based on some measurable aspects of the agents' work, either input (e.g. number of hours worked) or output (e.g. number of reminders sent to late taxpayers) - see [Besley et al., 2022](#) for a formal exposition of this general model. The impact of such incentives on the accomplishment of the administration's mission will vary depending on the principal's capacity to accurately measure relevant inputs/outputs and on the agents' intrinsic motivation to fulfill the tax administration's objective - and mis-designed incentive schemes can even become detrimental to the objectives ([Rasul, Rogger, and Williams, 2018](#)).

2.32. Expected impacts

Saving paper processing related costs: As detailed in section [2.2](#), before the introduction of Etax, all declarations were submitted on paper format. Every month, several agents of the LTC had to collect such declarations, digitize them into the central system and classify them into each taxpayers' physical folder. Beyond a possible transition period, the reduction in the cost to process tax declaration should inevitably free-up some time for agents of the LTC who previously participated to this activity. Unlike for minor organisational change, such gain will inevitably be observed by these agents' principal and we should expect the agents to be given replacing tasks.

Such reallocation could result in increased monitoring of LTC's taxpayers, which in turn may improve tax compliance (see section 2.41 for a discussion on how to measure tax compliance). Yet this is not straightforward, in particular because reallocation could take place across tax centres, especially if the personnel that has acquired expertise in processing paper is needed in other centers where e-filing has not yet been adopted.

Improving information quality and availability: Paper based declarative system requires a lot of man power, but also carries important informational risks: declarations could be lost (before or after having been digitized) and incorrect values could be entered, intentionally or not, during the digitization process. In light of this, we should expect the introduction of e-filing to eradicate at least mistakes introduced during data entry. It may also reduce the share of seemingly missing declarations, as the later cannot be lost anymore. But maybe the most important implication of e-filing is that taxpayers and administrators share, and know that they share, the same information about what has been declared. This implies that a taxpayers knows which inputs had to be used to compute its liability. The impact of that change on observed indicators is hard to predict as it depends on whether taxpayers were previously over paying or not because of incorrect values introduced during the digitisation but which we cannot identify in the data. We would expect it to at least increase taxpayers' consent to pay on average. Another implication of the sharing of information is that tax inspectors now know that what they read in the declaration corresponds to what the taxpayer has effectively declared. Given this, they can more confidently hold taxpayers accountable for any errors found in the declaration, or any contradictions with other sources (as third-party reported revenues for instance). In other words, it facilitates tax inspectors' job. Depending on their efforts to perform their tasks correctly (either through intrinsic motivation of though incentives provided by their principal), this could contribute to increase tax enforcement, and therefore tax compliance. As for the previous mechanism, e-payment is unlikely to play a significantly role here.

Saving transport related cost: Form the point of view of the taxpayer, the possibility to declare and pay online obviously allows to save time and transport related expenditures. This is a desirable outcome per se, it constitutes one the main rationale to adopt such technology. Given our assumptions on taxpayers' motivations, we expect this could induce some taxpayers, who otherwise would have not declared, to declare - but only to the extent that i) they perceive a positive risk for not declaring, and ii) previously faced transport related costs were large enough for them to take that risk. In other words, this mechanism should only impact these previously *almost declaring* taxpayers, and is unlikely for instance to contribute to embark taxpayers who perceive very little risk associated with not declaring. In our setting, given that treated taxpayers consist of the largest firms of the country, it is reasonable to assume that such costs represent a

very small share of revenue and therefore that this mechanism should play little role.

Sharing bank ID: Online payment requires taxpayers to share their bank identity to the administration, a step they were reluctant to take (see 2.2). Indeed, with their bank identify at hand, the tax administration can more easily activate a legal procedure (administrative seizure to third party holder) that forces banks to pay the tax administration any outstanding tax liability a given taxpayer owes - provided there is sufficient provision in the account. The availability of banking information should therefore limit taxpayers' leeway to resist tax recovery - a phenomenon that can be pervasive in low-income countries (Best, Shah, and Waseem, 2021b).⁴ We expect tax payments to increase through this channel for taxpayers resisting tax recovery.

End of cheques: Before e-payment, taxpayers paid in cash or cheque, cheque being more frequent for larger amounts. In both cases the official payment receipt is provided by the tax administration (as a proof of tax payment), upon the reception of the cash or the cheque, which are then transferred to the treasury. An issue with cheques however is that they mostly do not come with a guaranty for sufficient provision in the taxpayer' bank account. According to tax officials, it is not uncommon (even though we do not have data on the exact frequency), for cheques to be returned, by the treasury to the tax administration, because they are (partially) unfunded. When this happens, the tax administration needs to invalidate the initial payment receipt, reach out to the taxpayer and have it pay again. Exploiting this weakness of the payment system is a way for taxpayers to at best defer payment, at worst resist payment to the point they may negotiate collusion-based rebate. As this cannot happen anymore with e-payment, we expect tax payments to increase through this channel for taxpayers that most used such strategy before the reform.

Limiting in-person interactions: In person interactions between bureaucrats and taxpayers may affect tax collection in several ways. In the worst case scenario, they can collude to the detriment of the government: taxpayers may reduce their effective liability in exchange of paying the bureaucrat a share of the implied gain. In that case, in-person interaction reduces tax compliance as well as tax revenue. Inversely, tax officials may also force taxpayers to pay more than they owe to the state in order to meet their own tax revenue target, and perhaps avoid a sanction or receive a bonus. In that case, tax compliance, understood as distance from what should be, is reduced as well but through over-payment. Suggestive evidence for both channels have been found by Okunogbe and Pouliquen, 2022. Considering this, it is hard a priori to predict in what direction should declared liability and payment go following the adoption of online services. However, in the context of this study, given that treated taxpayers are extremely

⁴Note that this procedure is strictly codified and requires the agreement of several officials to be binding, therefore limiting opportunities for tax official to abuse their power. Similar institutional constraints are in place for fines and sanctions as well.

large we expect the bargaining power to be in their favor when interacting in person with a tax official. Hence we suspect over-payment to be rare and expect that the adoption of Etax would increase tax compliance and payment among previously colluding firms.

Saving paper processing costs or improving information will essentially result from e-filing, while the specific consequences of sharing bank ID or stopping using cheques will materialise only because of e-payment. However, both online services will play a role in saving transport costs and reducing in-person interactions.

2.4. Data and Empirical Strategy

2.41. Data

We assemble four datasets, all of which are produced by the tax administration along its operational routine. In addition to the source specific information listed below, each declaration/payment comes with an identification number, the name of the centre it was recorded by, and whether it was processed through Etax or not. We use this information to identify taxpayers' fiscal centre and trace their effective use of the e-filing and e-payment. We do not have data on the date of the declaration.

Value added tax (VAT): includes information on total sales, sales subject to VAT, sales exempt from VAT, exports, gross VAT due, deductible VAT (excluding past VAT credit)

Withholding taxes on employees (PAYE): includes information about the number of employees, total wage bill and total liability declared.

Withholding taxes on informal service suppliers (WTS): includes information on the total amount spent in exchange of services from informal suppliers, and the total liability declared (which should always be 5 % of the amount).

Payment data: this data set was extracted separately and includes information on the amount paid, the type of tax and the fiscal period it applies to, the payment date and the mode of payment.

Tax compliance is inversely proportional to the difference between what a taxpayer *should do* (according to fiscal law) and what it *actually does*. As such, it is captured by a collection of indicators: the probability to declare, the probability to pay when a liability is due, the probability to perform these acts on time, the difference between *declared* and *real* tax base/liability, and finally

the difference between tax *effectively paid* and tax *really due*. Such indicators can be hard to measure directly because most of what we observe is *reported*, therefore subject to manipulation or mistakes, and not a direct measure of the reality. Considering this, the probability to declare is probably the indicator that can most unequivocally capture tax compliance, because we know a declaration is due every month. It is still imperfect however as some declaration might be missing only because they were not digitized ex-ante (e.g. submitted but lost). Assuming taxpayers prefer to not pay tax (see section 2.3) all observed increase in indicators implying that taxpayers pay more taxes, or could be subject to higher taxes is indicative of improvement in tax compliance as long as it does not stem from an increase in real activity or extortion by tax officials. Under these conditions, increases in declared tax base/liability, tax rate, or effective payment should reflect increase in tax compliance.

In addition, for each declaration, we can compute binary indicators encoding whether values are missing, inconsistent or implausible. We consider a declaration has a *missing* value if a positive liability is associated with the declaration but at least one variable which should be positive whenever the liability is positive (e.g. total wage bill for PAYE) is missing or equal to 0. We consider a declaration to have *inconsistent* values if some of its *positive* values are inconsistent. Last we consider a declaration to have *implausible* values, if some its *positive* values is above some variable specific extreme value.

2.42. Selection

We select taxpayers that have started declaring and/or paying before January 2017 and that have continued doing so beyond December 2019. In other words, we restrict to a set of taxpayers which we know existed before and after this time window⁵. Unobserved declarations within that period are considered missing - we therefore “fill the gaps” to impute the corresponding missing observations in our dataset. By construction, this allows to produce a balanced panel of taxpayers from the beginning of 2017 until the end of 2019.

We cannot satisfyingly exploit data declared much before 2017 because a major update of the IT system in late 2016 caused a breach in the use several fiscal ID, thus preventing us from identifying firms across the update. When studying the impact of the introduction of Etax, we also prefer to not exploit data beyond 2019, as the COVID pandemic spreads to Senegal in early 2020, affecting taxpayers differently depending on whether or not they had already adopted

⁵Start and end date are not measured well enough by the administrations to identify *active* taxpayers in a less indirect way. This is essentially because of the prevalence of informality: economic actors can register to the tax administration years after having started being active, and inversely, they may become inactive from the point of view of the TA, and continue to do business, sometimes relocating.

Etax before (a phenomenon we document in the second part of our results). We define the treated group as all firms that have consistently been monitored by the LTC during the reference period. Figure 2.71a and 2.71b show that first registrations and use took place in September 2017 and sharply increased afterwards. By February 2018, 80% of the LTC's taxpayers had used Etax at least once, and each month 85% of taxpayers submitting a monthly declarations used Etax for it, and 80% of taxpayers paying some positive amount of taxes did so through Etax. The adoption took place quickly enough to allow for a credible impact evaluation, and by the end of the period nearly all LTC taxpayers of this selected sample were registered, effective, users of the platform. Yet, a non trivial share of the treatment group did not register in the first half of 2018. As a control group, we use all firms monitored by tax centers in the region of Dakar that did not adopt etax - including MTC2, and RSC where a small proportion of taxpayers did adopt Etax during the second half of 2018, probably on a voluntary basis (see Figure 2.71a and 2.71b). We account for such imperfect compliance to the treatment in our empirical strategy by instrumenting Etax registration on pre-reform centre affiliation (see section 2.43).

After this preliminary selection, the full sample consists 5,694 taxpayers, 831 (15.6%) of which are treated. By construction, treated and control taxpayers are very different. The main forcing variable which the administration uses to allocate firms to the large taxpayers centre is the turnover. Figure 2.72b, displays the empirical distribution of taxpayers in each group with respect to their turnover, indicating the threshold (XOF 1,000,000,000) above which taxpayers ought to declare to the LTC. Graphical evidence suffice to discard the possibility to apply any regression discontinuity design, threshold is not strictly binding and the density around the threshold is too low. To make our sample as comparable as possible without losing too much statistical power, we restrict our control group to its largest 860 taxpayers (measured by average pre-reform turnover). This generates our default dataset, with a total of 1,691 taxpayers. Despite significant trimming of the initial control group, average taxpayers size still differ significantly across groups as show in figure 2.73. For all of our main indicators, we address this issue by running sensitivity tests on 40 different sub samples obtained by incrementally removing the largest 5 taxpayers from the treatment group.

2.43. Empirical Strategy

We estimate the impact of the reform by comparing taxpayers monitored by the Large Taxpayers Centre (LTC) to other taxpayers of Dakar, before and after the use of Etax was made legally mandatory for the LTC (see section 2.42 for a precise definition of treatment and control group). Treatment adoption is set at the centre level, therefore the first treated period is the same across all taxpayers and after this month we consider *effectively treated* any taxpayer monitored by the

LTC, irrespective of whether it actually uses e-filling and or e-payment technology. Considering this, our first estimates will capture the *intention-to-treat* of adopting e-filling and or e-payment at the centre level onto its taxpayers. By construction of the sample, treatment status is also an absorbing state. Given these initial specifications, and following [Roth et al., 2023](#) and [Miller, 2023](#), we can estimate the impact of the reform using an event-study model (or dynamic diff-in-diff model) with this functional form :

$$y_{it} = \alpha + \sum_{k=1}^K \beta^k \cdot Lead_{it}^k + \sum_{j=-2}^J \gamma^j \cdot Lag_{it}^j + \delta_i + \lambda_t + \mu_c + \epsilon_{it} \quad (2.1)$$

The statistical unit is the taxpayer (i) at the month level (t), with $J \leq t \leq -1$, or $1 \leq t \leq K$. Variables Lag_{it}^j and $Lead_{it}^k$ are binary indicators equal to one for an observation of a treatment group taking place respectively j months *before* the adoption of Etax and k months *after*. $Lead_{it}^1$ indicates the first treated month, while Lag_{it}^{-1} (the last untreated month) is omitted from the equation 2.1 to serve as baseline. Each β^k captures the impact of the reform exactly k -months after its introduction compared to baseline ($t = -1$) and is equivalent to the β coefficient one can retrieve by estimating a simple diff-in-diff model with only two periods, either before treatment ($t = -1$) or after ($t = k$). The coefficient γ^j 's are similar but for the period *before* the reform, and act as placebo tests. The coefficients δ_i , λ_t and μ_c are individual, time and centre fixed-effect, and ϵ_{it} is the unobserved error term.

With this specification, we can provide graphical evidence on the evolution of the impact by plotting γ^j 's and β^k 's coefficients across time. Then to summarize the overall impact for the entire post-reform period observed into one single coefficient, we group all observations taking place after the reform and estimate the following 2x2 diff-in-diff model with fixed effect :

$$y_{it} = \alpha + \beta \cdot Treat_{it} \cdot Post_{it} + \delta_i + \lambda_t + \mu_c + \epsilon_{it} \quad (2.2)$$

where $Treat_{it}$ and $Post_{it}$ are indicators for being in treated centre, and observed after the reform respectively. When estimating 2.2 we restrict our sample so that $-1 \leq t \leq K$, thus keeping the same baseline as in our dynamic specification. Similarly, a backward version of model 2.2 allows to estimate a general γ which provides a summary test for the parallel pre-trends.

As for any diff-in-diff model, the underlying identification assumption requires that, absent the reform, indicators y_{it} would have evolved similarly in the treatment and the control group. This assumption cannot be proven and in fact taxpayers from the treatment and the control group are arguably very different (see section 2.42). To address this issue we show evidence that the parallel trend assumption is at least not refuted by placebo tests run for our main outcome

of interest on the pre-reform period. Furthermore, we estimate equation 2.2 from multiple alternative samples of the treatment and the control group to discuss how treatment impact varies as the difference across group widens.

Unfortunately compliance to the treatment is imperfect (see section 2.42): some treated taxpayers do not adopt Etax (“never takers”) and some control taxpayers from the MTC2 and RSC do adopt it (“always takers”). To account for this, we estimate fuzzy versions of model 2.1 and 2.2 above, using $Lead_{it}^k$ and $(Treat_{it} \cdot Post_{it})$ as instruments for effective adoption of Etax. Let m_i^0 be the first month when taxpayer i uses Etax (either by declaring or paying through Etax), we define the effective use of Etax relative to the post-period k as a binary variable $Etax_{it}^k$ equal to one if and only if $m_0 \leq t = k$. Similarly, we define the effective use of Etax relative to the full post-period $Etax_{it}$ as a binary variable equal to one if and only if $m_0 \leq t$. Note that these definitions do not depend anymore on centre assignation. We estimate these fuzzy diff-in-diff versions of our models using two stages least square method. By definition, lag’s coefficient γ^j ’s in model 2.1 will be unaffected, while lead’s coefficients β^k (model 2.1), or β (model 2.2) will become the Wald-DID estimates, capturing the local average treatment effect on the compliers.

Provided that no other reform was implemented at the same time differently affecting treatment and control groups, this strategy allows us to attribute any significantly changes observed around late 2017 as resulting from the Etax reform. However, by anticipation of the reform, or as an immediate response to it, the tax administration could have decided to relocate staff across quasi-experimental groups. This may be a natural strategy considering that online services are adopted at least partly to save the administration the cost of managing paper. In that regard, the staff occupied with digitizing and classifying declaration ex-ante at the LTC may have been relocated to other (control) centers. Movements in the other direction are also possible, whereby high skilled staff is being relocated from control centers to the LTC. Compiling data on the yearly evolution of the distribution of all civil servants working at the tax administration, we show that none of these changes have occurred. Table 2.81 shows that changes between 2017 and 2018 in the number of civil servants affected to the treatment center were very marginal: +2 among civil servants of rank A (highest qualification), -1 for rank B, +1 for rank C and no change for the lowest rank D. Simultaneously, changes observed for civil servants affected to centers of the control group mostly affected the most qualified sub-group: +13 and +19 for rank A and B respectively (i.e 20% and 12% increase respectively), while rank C and D civil servants remain virtually constant. Center from the control group received 13 (ie. 70%) of the 19 newly recruited rank A civil servants. Overall, we observe that changes in the staffing structure have been minor and are unlikely to drive our results. If anything, we would only

expect the increase in rank A civil servants among centers of the control group to make our estimates slightly downward biased.

2.5. Results

Figure 2.74a shows that just in between the release of the ministerial decree in July 2017 and the effective adoption of Etax observed in December 2017 / January 2018, the probability to submit any of the tree monthly declarations sharply drops among taxpayers of the LTC, but not for other taxpayers. Consistently with reports by tax officials working within the LTC, we interpret this short-lived negative impact as a result of the the transition cost to the new system. Taxpayers' accounts had to be created with secure login credentials and both administrators and taxpayers had to learn how to use the platform. For this reason, rather than defining our reference period (i.e $t = -1$ in model 2.1 and 2.2) as the first month *before* we observe any declaration or payment through Etax by any taxpayers (that would be November 2017), we assume September and October were already affected by the reform and therefore set August 2017 as our reference period. We refer to the last four months of 2017 as the *transition period*. Our overall results are robust to the inclusion/exclusion of these four months (see Tables 2.82 and 2.83 for a comparison).

We do not find any significant impact during the first year on the probability to submit at least one of the tree monthly declaration (see Figure 2.74b). A few months do come out significant in the following year, which could suggest that the adoption of Etax increases the probability to declare later in time by about 4 percentage points - before Etax, average declaration rate in the treatment and the control group were 94 % and 96 % (see Figure 2.74a). In case these results are indeed the lagged consequences of the introduction of Etax, they likely do not reflect a response to saving transport costs as we would typically expect this mechanism to be effective in the short-term. Rather, we would interpret it as most likely resulting from an improved monitoring of the taxpayers by the LTC. As highlighted in section 2.32, improved monitoring may stem from better information or reallocation of time previously spent by the bureaucrats on processing paper declarations. Interestingly, we find a similar pattern for the probability to declare a positive total liability (Figure 2.78b): no impact in the the short-term but a significant discrepancy by the very end of the period - with orders of magnitude around 5 percentage points. Trends in monthly averages across groups show that these probabilities are roughly similar in level ex-ante, then the treatment group see its probability increase by 1-2 percentage points and plateau, while that of the control group progressively loses almost 5 percentages points. Note however that, as we cannot completely discard the possibility for an impact on

the probability to submit a declaration, any average impact on outcomes whose measurement depends on information present in the declaration can become harder to interpret because of compositional effect.

We find a sharp and strong negative impact on the submission of declarations with missing values. The probability to submit at least one declaration with a missing value decreases by 5-7 percentages point (see Figure 2.75b). The trend in Figure 2.75a are consistent with what we would expect: missing values drop in the treated group (while remaining relatively stable in the control group). Relative to pre-trends level, this implies a more than 50% change. Nevertheless, we do not find any clear impact on other outcomes capturing the quality of the information available in the declaration - neither on inconsistencies (Figures 2.76a and 2.76b), nor on implausible values (Figures 2.77a and 2.77b). Had the improvement in information come from the taxpayer, we would have expected to observe an impact on other dimensions as well. Our interpretation here is that the drop in missing values results from the eradication of voluntary omissions made by data entry agents. By construction of the indicators for the missing values, most of them relate to fields which data entry agents do not have to fill to validate a given declaration as long as they enter value for the tax liability. As tax payments can still be processed as long as there is information about the tax liability, some level of tolerance for missing secondary information may have been granted to accelerate data entry. Incidentally, these results imply that any average impact whose measurement depends on information present in the declaration may become even harder to interpret as there are now two sources of compositional effect: the composition of those who declares, and, conditional on declaring, that of those who do not have missing value.⁶

Average total liability declared clearly seems unaffected by the reform (Figure 2.79a), estimated treatment effects vary around zero through the entire period (Figure 2.79b). Nevertheless, graphical evidence for the trends in average payments strongly suggest that effective payment increase (Figure 2.711c): average monthly payment among treated taxpayers jumps from roughly USD 40,000 on average in 2017 to at least USD 50,000 on average in 2018, and USD 60,000 in 2019, while it remains flat for the control group. Monthly average treatment impact are significant at the 5% level for 12 out of 24 months in the post-period and increase as well overtime, from approximately USD 20,000 to USD 25,000. If accurate, this impact is arguably large: a net average monthly impact of USD 20,000 across the 831 treated firms would imply that about 6.5 % all tax revenues collected in 2018 resulted from the reform, i.e the equivalent of 1% of Senegal's 2018 GDP. However, the magnitude of the results appears to be largely driven by extremely large taxpayers (measured in terms of turnover declared prior the reform). Figure

⁶This is the reason why we do not include here impacts related to tax-base or tax rates.

2.711g indeed shows that the overall impact decreases from a little more than USD 20,000 per month on average to a little less than USD 10,000, just by removing the 5 largest firms. As a few more firms get removed, it converges to about USD 5,000. The payment data we exploit is a direct trace of payment effectively executed by the tax administration information software for the benefit of the treasury, this source is the cornerstone of the administration's performance results and is its most scrutinized source of information. For that matter, we have very good reasons to believe the source is accurate. Furthermore, a case by case inspection carried out by tax administrators themselves confirmed amounts measured in the database for the largest 20 taxpayers. If we can most safely assume that these extreme levels are not due to measurement errors, our identification strategy obviously does not guaranty that *for these very firms* something else than the reform might have played a role. Given their leverage on the average impact we must acknowledge that the true magnitude of impact on payments will remain uncertain. Our sensitivity tests on 40 alternative samples obtained by incrementally trimming the 5 largest taxpayers from the treatment group do nevertheless suggest that the reform had a positive impact on payment of about USD 5,000 per month on average (Figure 2.711g). Available information does not allow to determinate whether this increase is driven by a drop in collusion or an increased capacity to enforce payment recovery.

On the COVID-shock: Later in time, as the COVID pandemic spread to Senegal and containment measures were put in place, we observe that taxpayers who had not adopted Etax by then were less likely to submit a declaration. Figure 2.715a show that the probability to declare drops from 95% to 91% among these taxpayers, while it remains very stable for taxpayers who could declare online. Applying model 2.1 over 2019-2021 for a balance sample of 2,778 taxpayers, redefining treatment and control according past use of Etax and using March 2020 as the new time reference, we estimate that the COVID-shock decreased the probability to declare for about 4 percentage points for 3 consecutive months, and that it took about one year and a half for this negative shock to fully resorb. By construction of our sample, all taxpayers in the sample, including those negatively affected, do declare and or pay of beyond 2021. In that regard, missing declarations here could and should have been declared later in time, after containment measures had been lifted and thus should have filled the gap we observe.⁷ In other words, the effects we estimate here are not driven by real economic downturn leading firms to close: all of them survived and delayed declaration are possible. The most plausible channel to explain this negative shock is therefore an increase in the cost of submitting a declaration. The state of emergency was proclaimed on the 23rd of March, and maintained until the 29th of June, it

⁷As for all estimation of the paper, we always use *fiscal* time and not *real* time, as we do not have this information. For example: missing declaration for April 2020 in Figure 2.715a could have been declared in July 2021.

implied movement restrictions and a curfew. These 3 months correspond exactly to the period for which we observe the most severe impact. Yet, under this interpretation, these findings must then be reconciled with the fact the introduction of e-filing did not lead to any increase in the probability to declare among taxpayers of the LTC, in particular as ex-ante levels are very similar for treated firms in both settings. A simple explanation would be that the 2018 decrease in the declarative cost induced by the introduction of e-filing was not as large as its increase due to COVID related containment measures.

2.6. Conclusion

This article investigates the causal impact of the adoption of electronic filing (e-filing) and electronic payment (e-payment) technologies by the largest taxpayers in Senegal. The analysis employs a fuzzy dynamic double-difference method, focusing on the period following the mandatory implementation of the Etax system in July 2017. The findings reveal intriguing insights into the complex relationship between technology adoption and tax compliance.

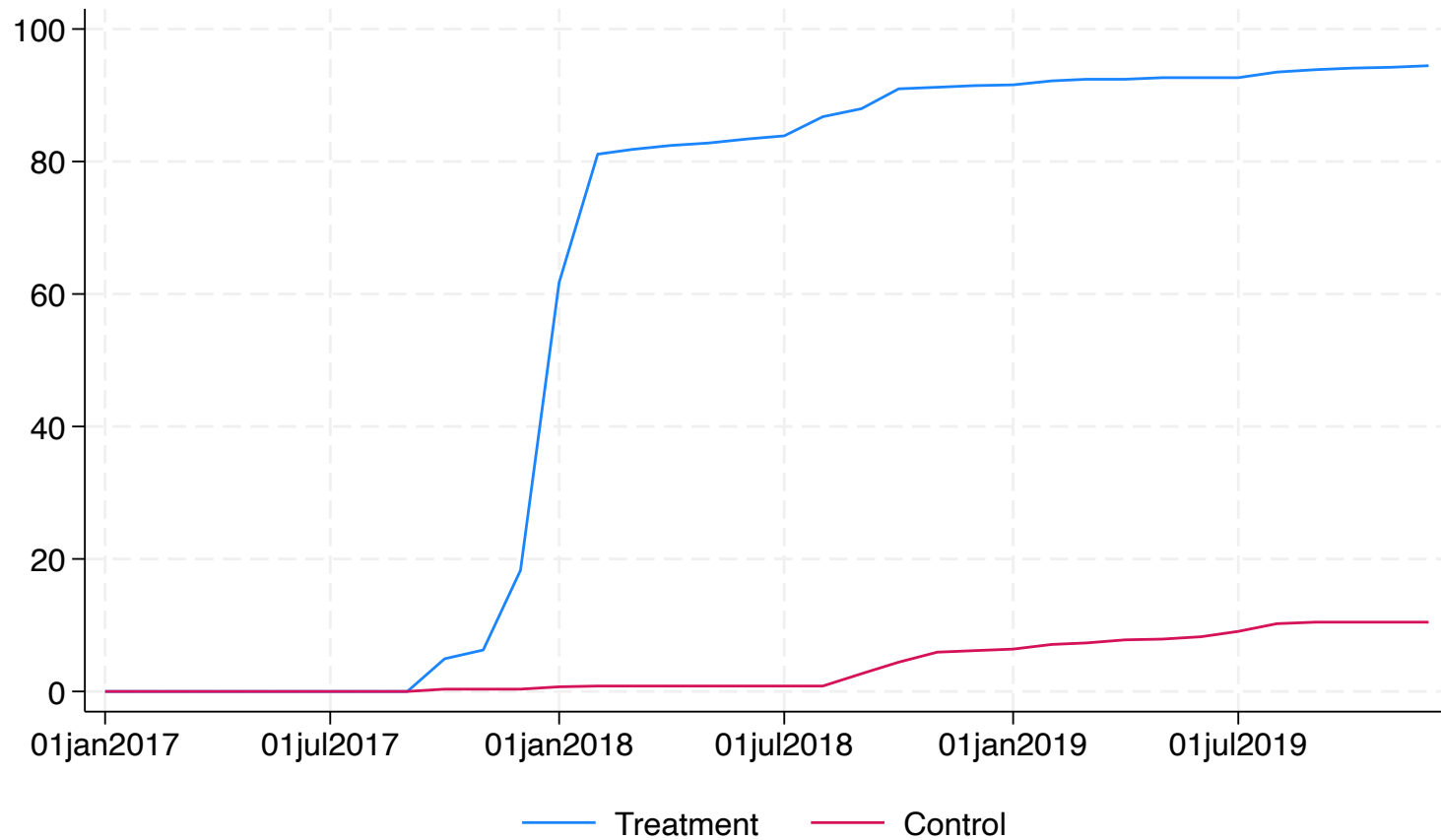
The results indicate a short-lived negative impact on the probability to declare immediately after the ministerial decree mandating Etax adoption which suggest that taxpayers and the administration faced a small transition cost. However, in subsequent months, there is no significant impact on the probability to declare. Interestingly, we observe a clear decrease in the prevalence of missing values in digitized declarations, most probably reflecting the fact that under the previous system, data entry agents were omitting to type values unnecessary to compute tax liability. While there is no significant impact on the probability to pay nor on average liability declared, we find a positive and possibly large impact on tax payments, particularly concentrated among the largest taxpayers. This could be due to reduced opportunities for collusion or increased capacity to enforce payment recovery. Further research would be needed to disentangle the two.

In summary, the article provides valuable causal evidence on the impact of e-filing and e-payment adoption on tax compliance in Senegal, shedding light on both the short and medium-term effects on various indicators. Despite no or modest impacts on declarative tax compliance, the results suggest that e-payment solutions could play a crucial role in increasing average effective payments by the largest taxpayers, potentially leading to significant revenue gains. This research contributes to the limited literature on the role of e-filing and e-payment technologies in low-income countries, offering nuanced insights into their effects on tax compliance and revenue collection.

2.7. Figures

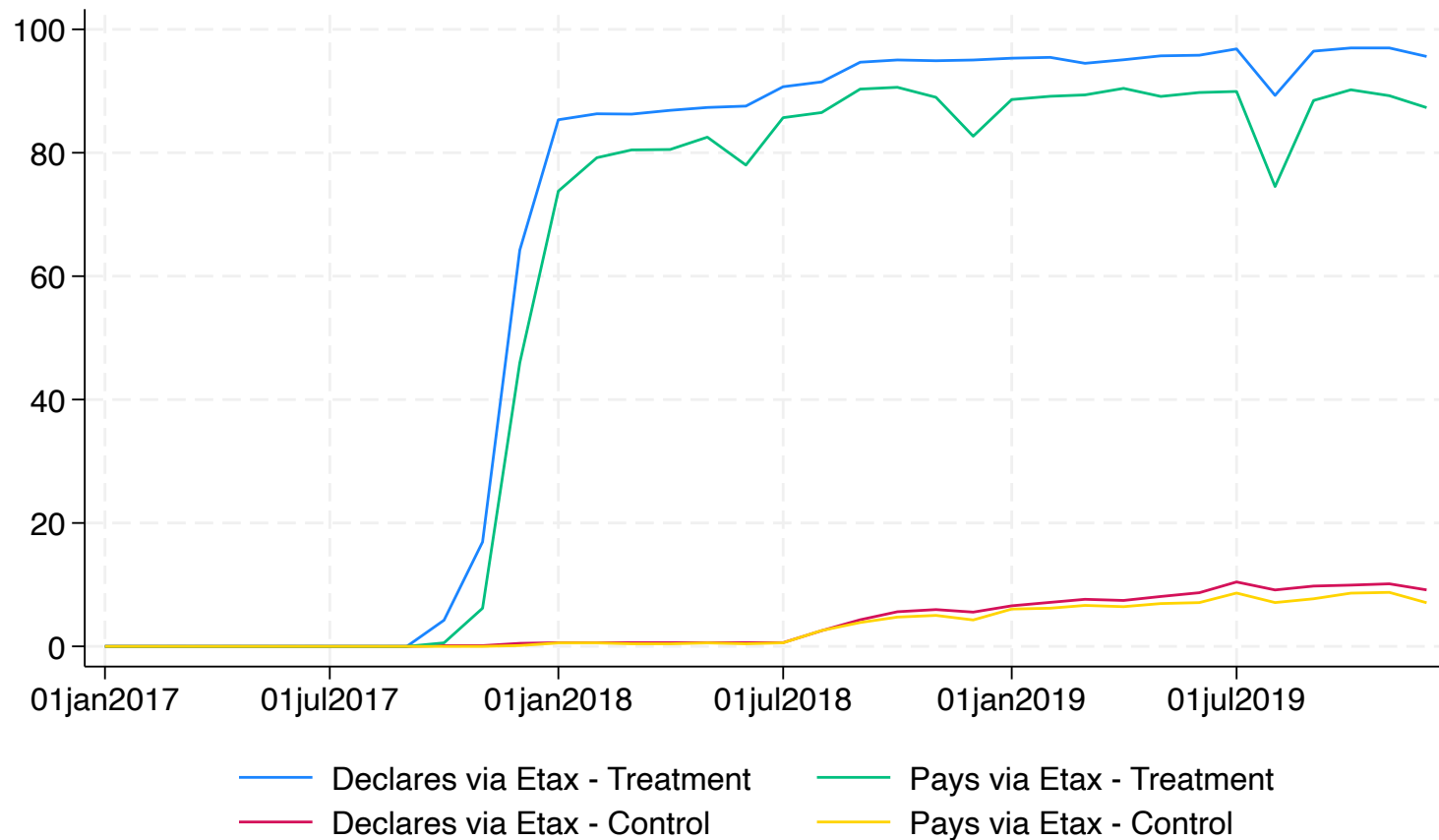
Figure 2.71 – Adoption of Etax

(a) Share of taxpayers having used Etax at least once



Notes. The figure displays the share of taxpayers having used Etax at least once (either through e-filing or e-payment) within each group. The sample here is the sample we use for all our main estimation, with 831 treated taxpayers and 860 taxpayers in the control group.

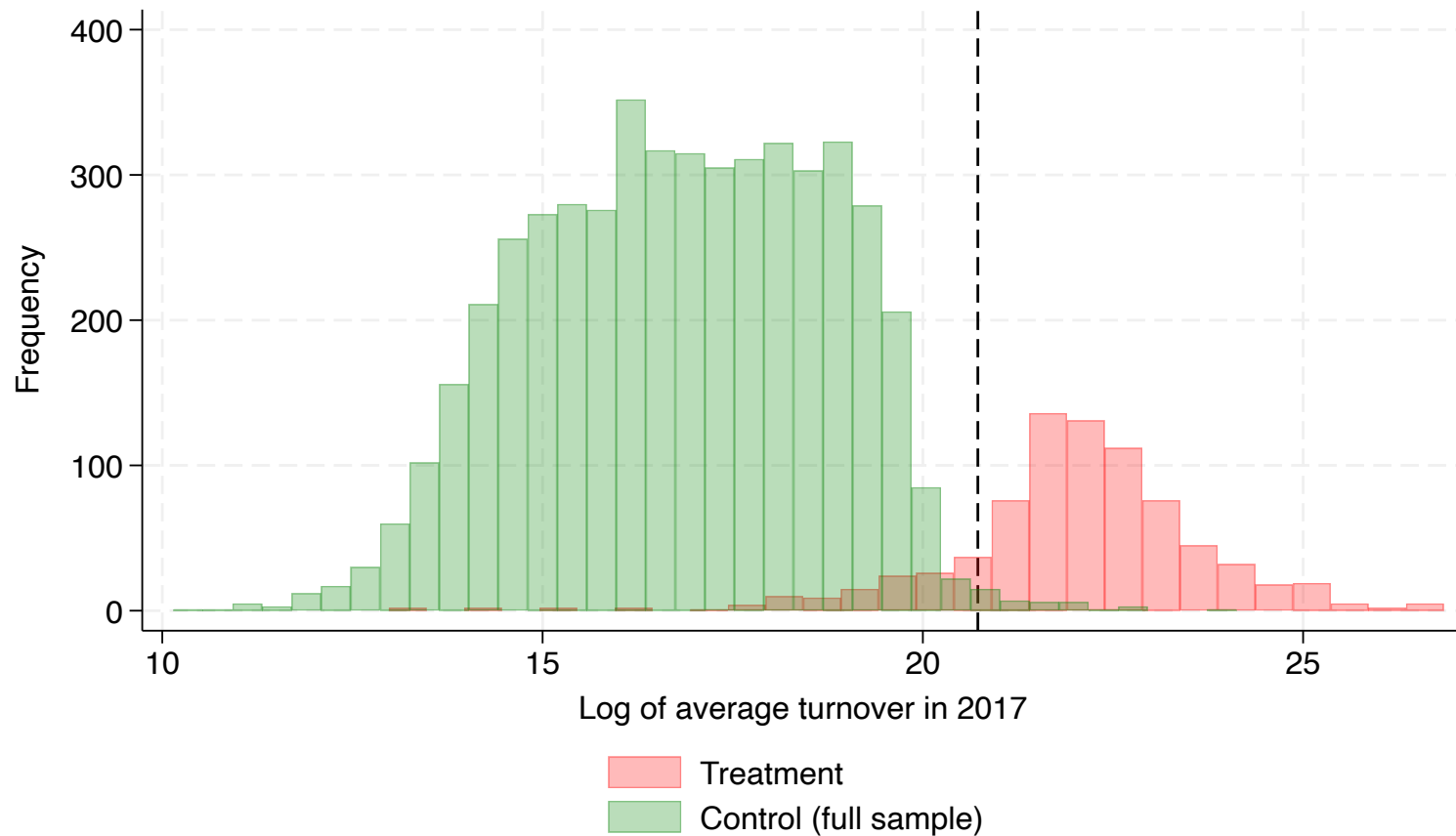
Figure 2.71 – Adoption of Etax
(b) Monthly share of effective users



Notes. The figure displays the share of taxpayers that effectively use Etax each month, either through e-filing or e-payment, within each group. The sample here is the sample we use for all our main estimation, with 831 treated taxpayers and 860 taxpayers in the control group.

Figure 2.72 – Probability distribution with respect to taxpayers size

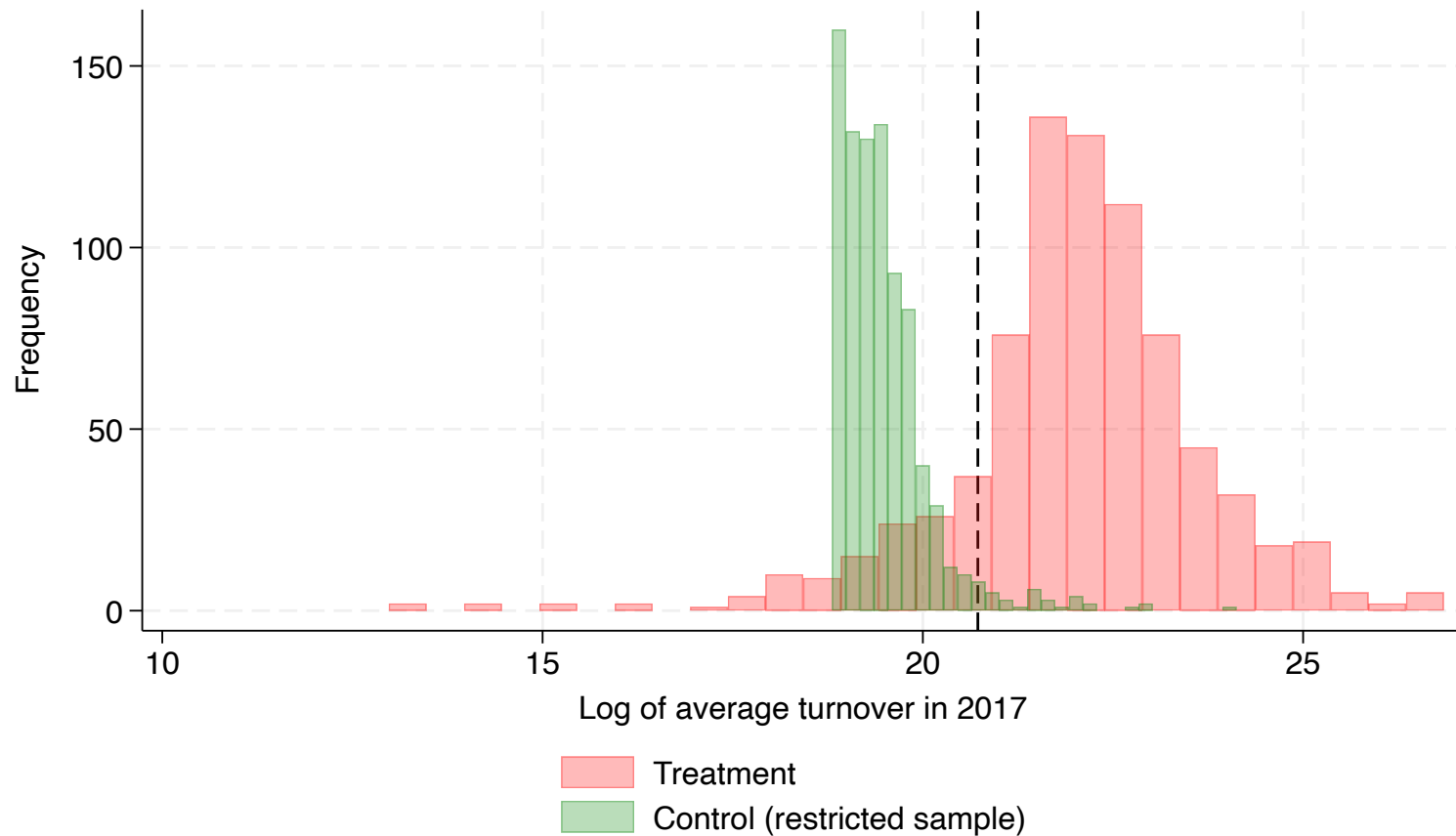
(a) Full (initial) Sample



Notes. The figure displays histograms of the density distribution of the treatment and control group by size, measured as average turnover in 2017 before any trimming of the control group. The 831 treated taxpayers are all those monitored by the Large Taxpayers Center. The dashed line indicates the threshold above which a firms becomes eligible to the LTC.

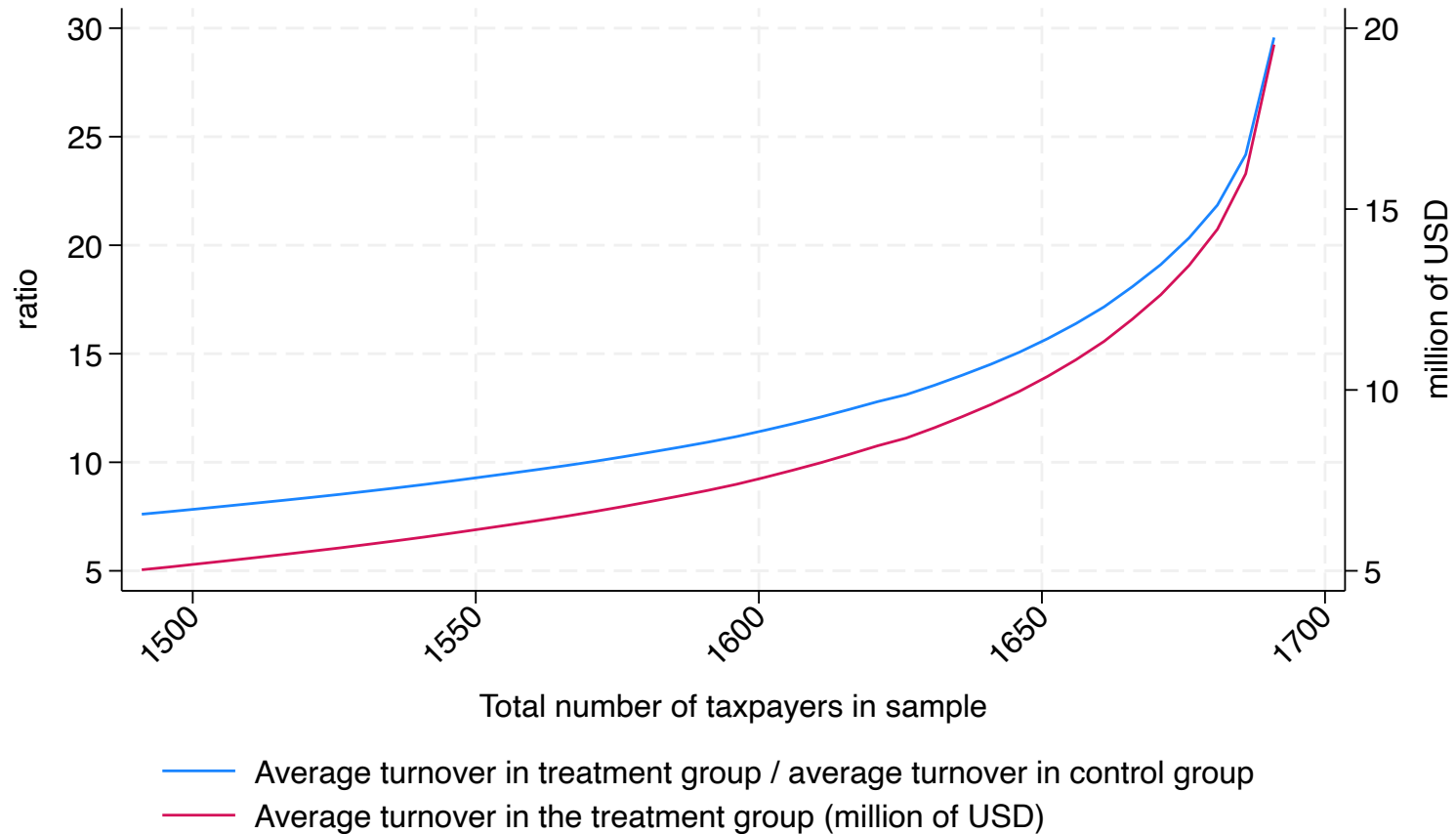
Figure 2.72 – Probability distribution with respect to taxpayers size

(b) Restricted (main) Sample



Notes. The figure displays histograms of the density distribution of the treatment and control group by size, measured as average turnover in 2017, after restricting the control group to its 860 largest taxpayers. The 831 treated taxpayers are all those monitored by the Large Taxpayers Center. The dashed line indicates the threshold above which a firms becomes eligible to the LTC.

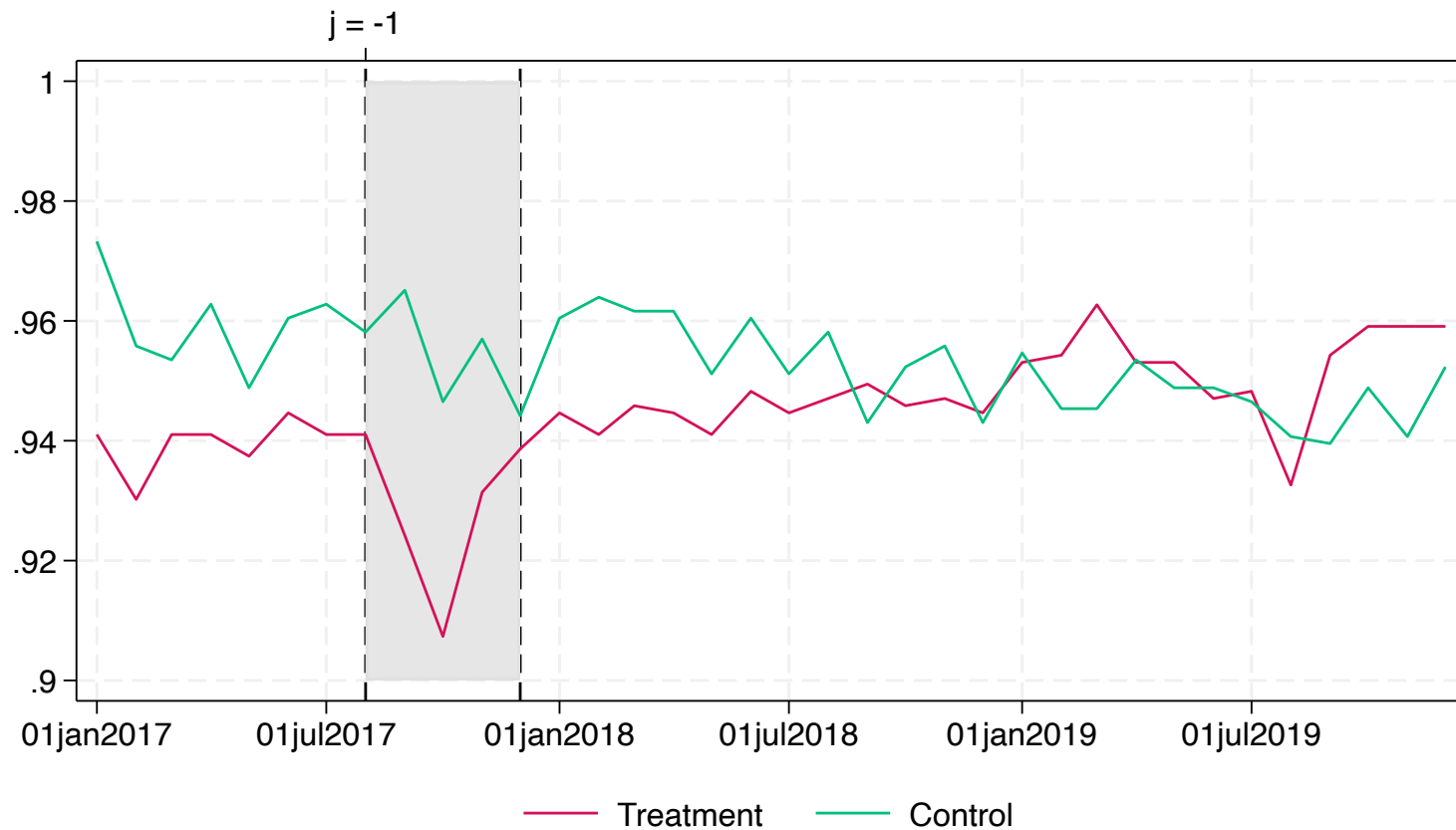
Figure 2.73 – Sensitivity to sample selection



Notes. The figure displays variation in the average turnover among treated taxpayers, and in the ratio between this average and the average turnover among control taxpayers, as we incrementally remove the 5 largest taxpayers from the treatment group. When keeping all 831 treated firms, their average turnover in the pre-reform period is about USD 20 million, which is about 30 times more than in the control group. Trimming the largest 200 taxpayers from the treatment group (top 24%) would bring down this ratio to 7.5.

Figure 2.74 – Submits at least one declaration

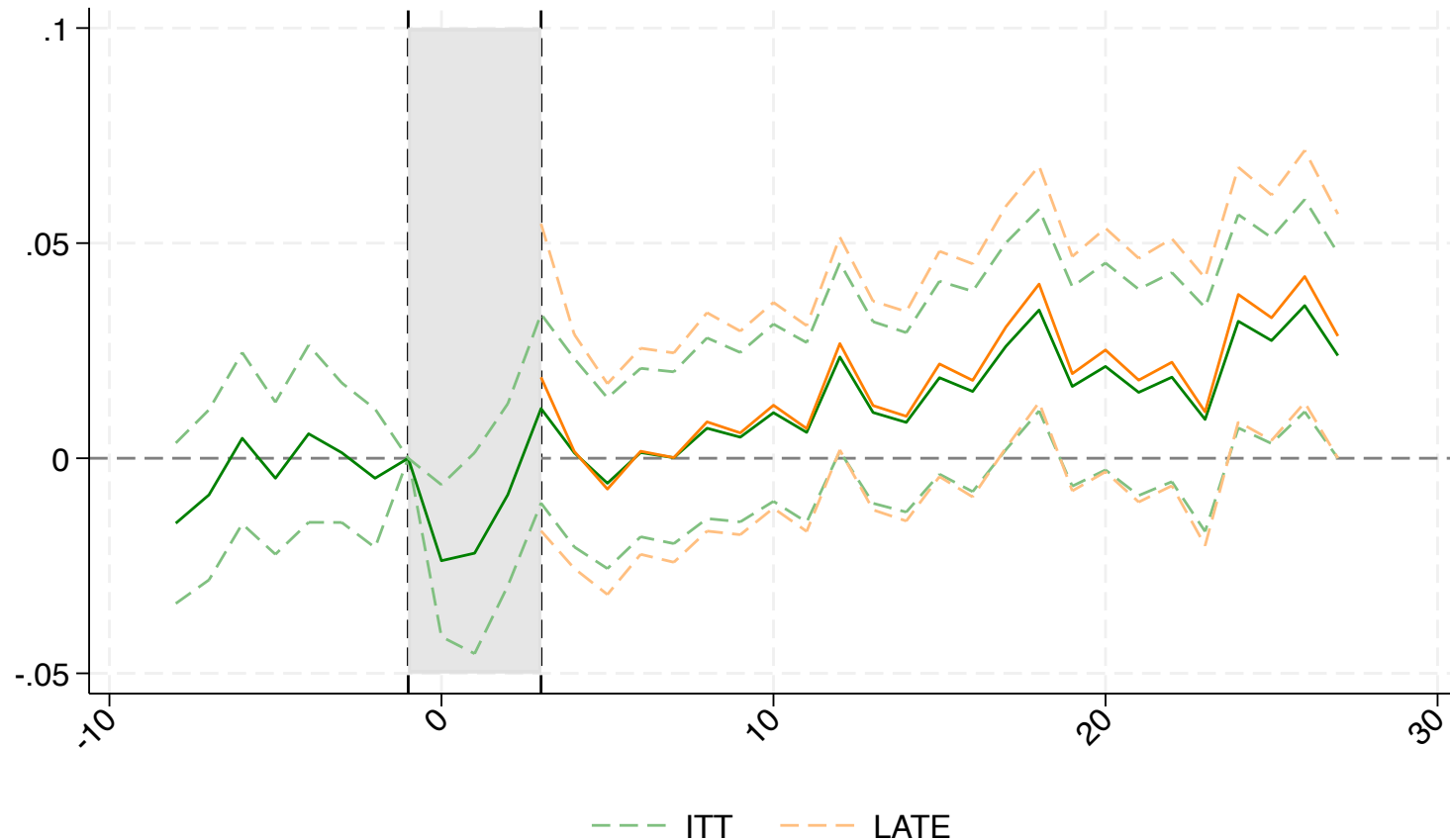
(a) Time trends



The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.74 – Submits at least one declaration

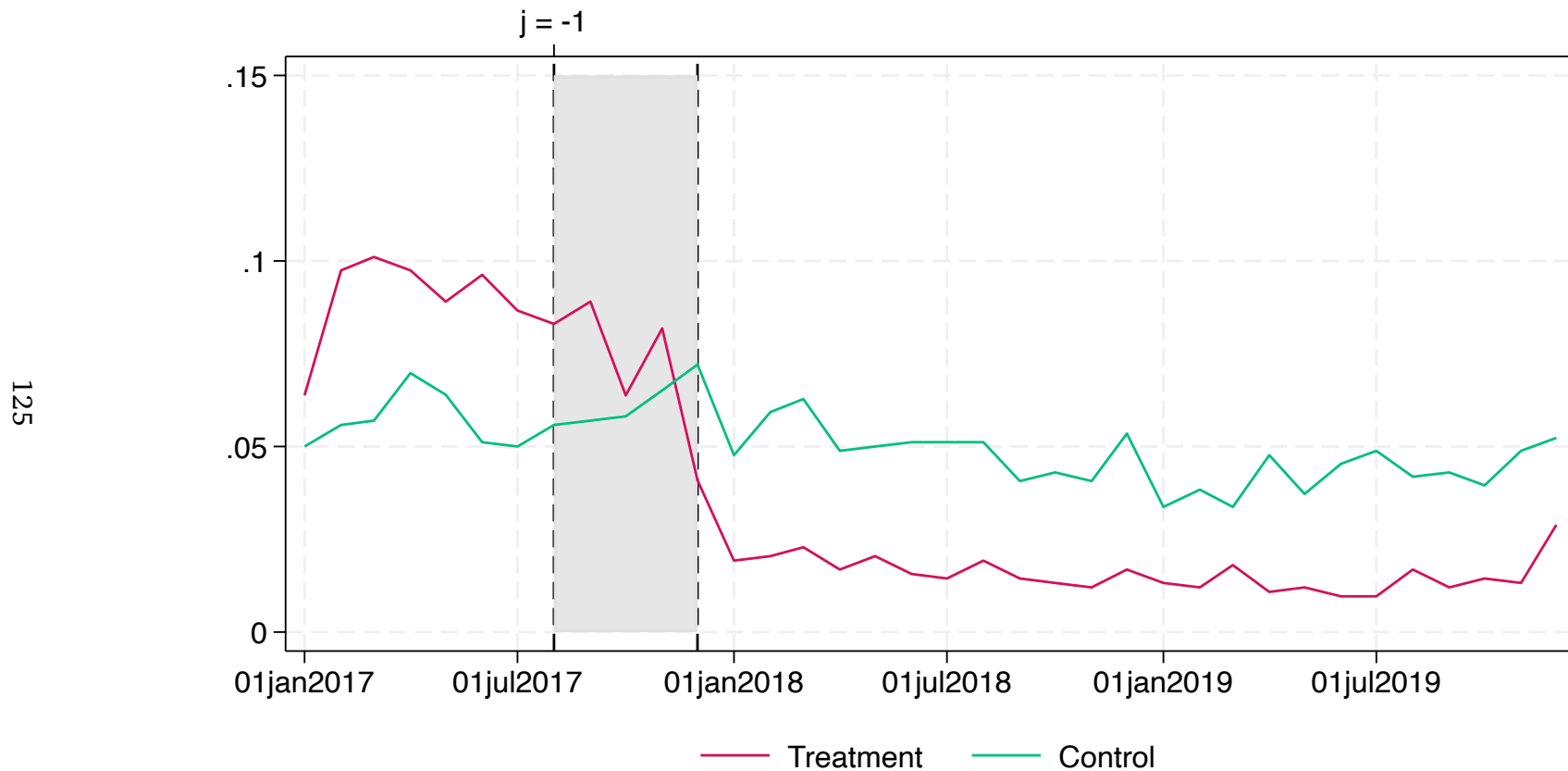
(b) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.75 – Some declaration has missing information

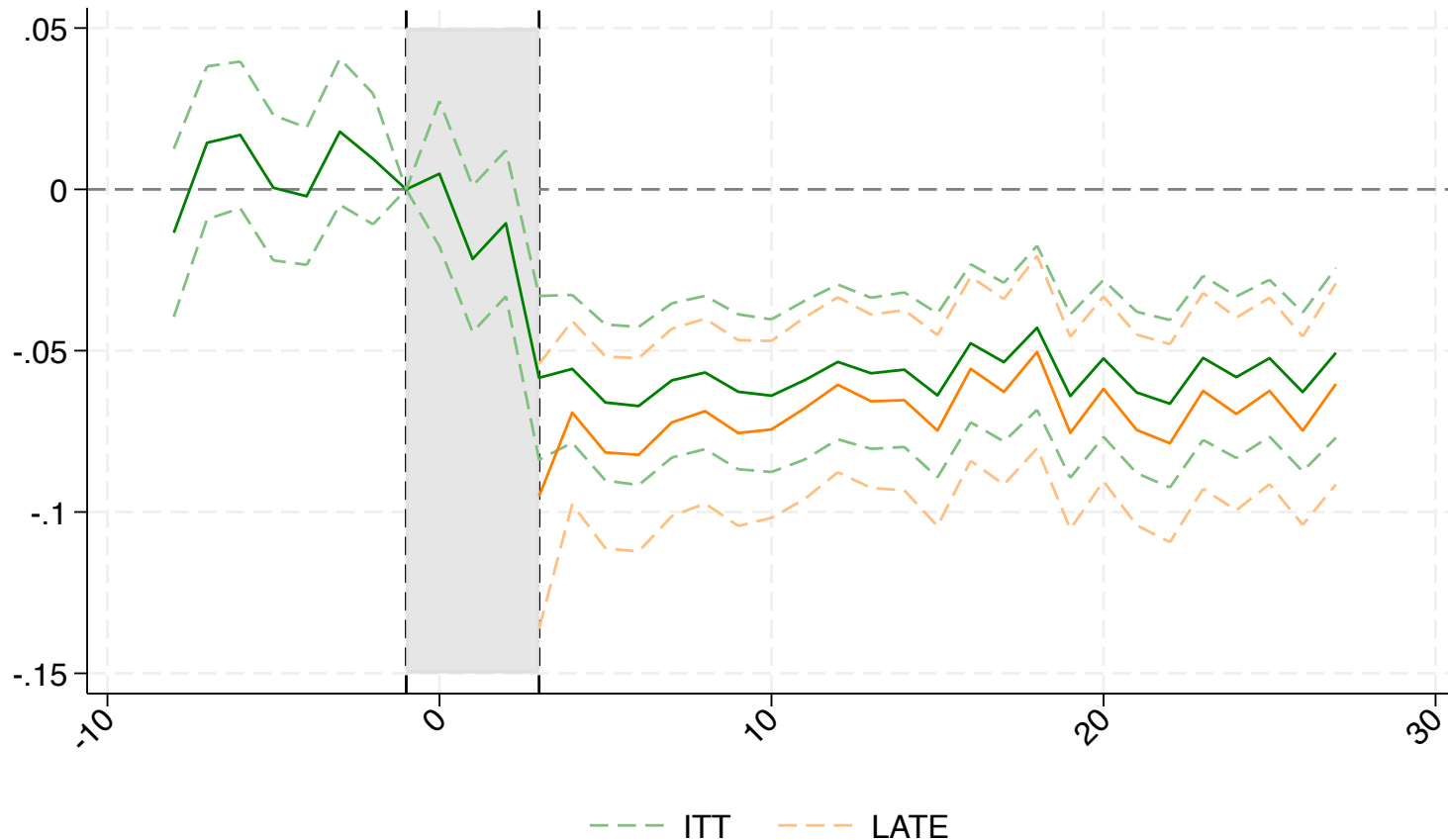
(a) Time trends



The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.75 – Some declaration has missing information

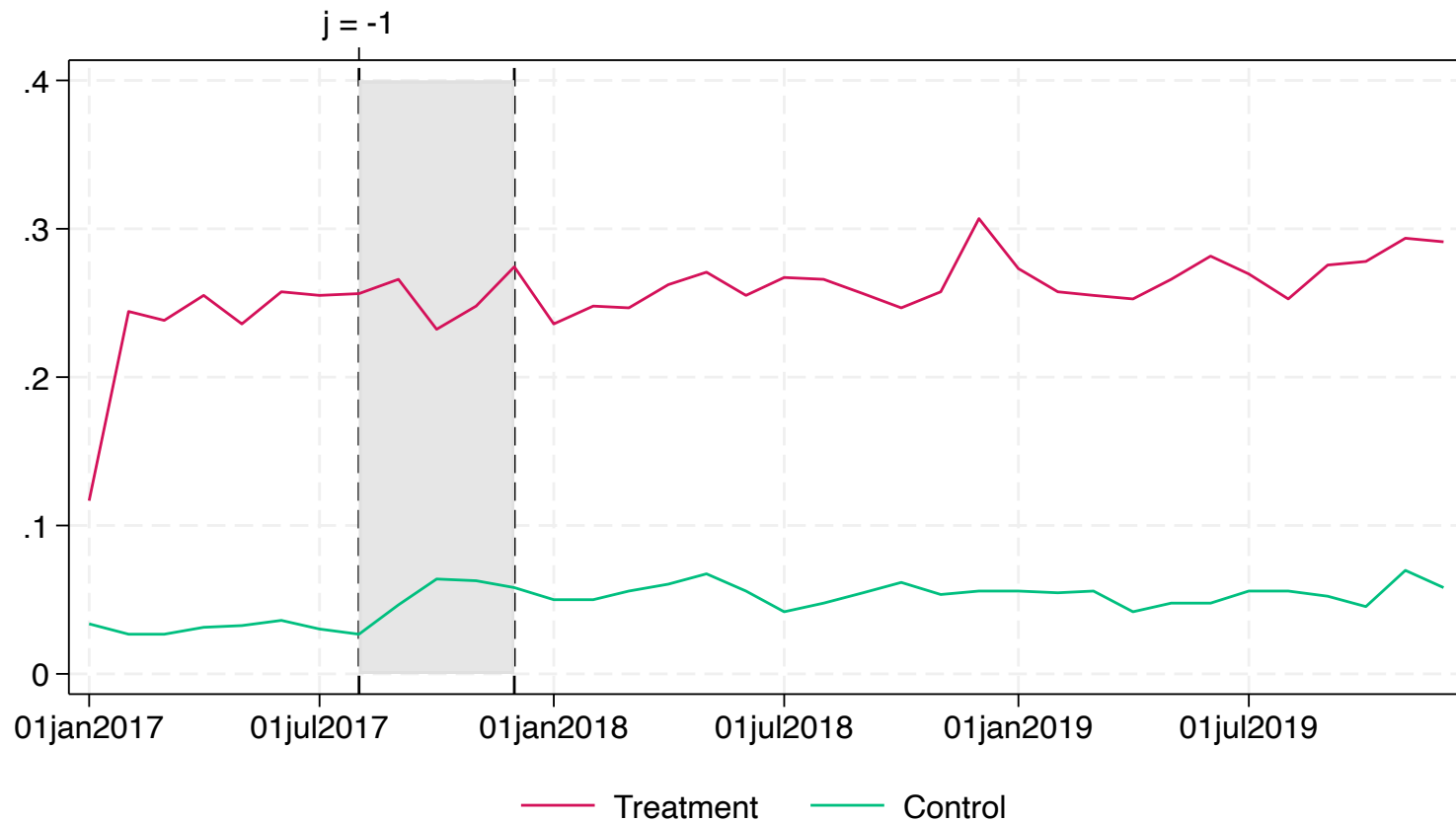
(b) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.76 – Some declaration has inconsistent information

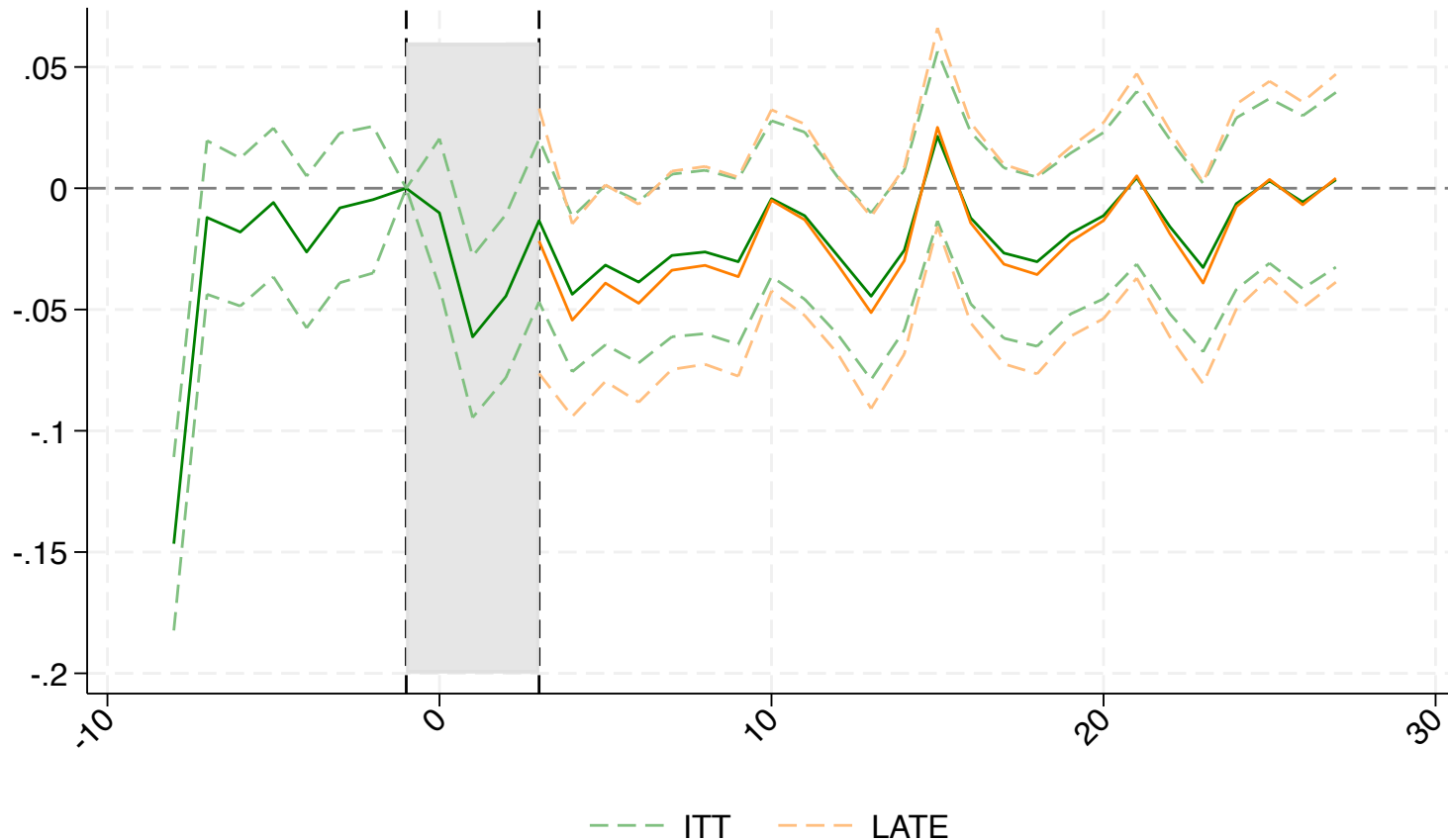
(a) Time trends



The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.76 – Some declaration has inconsistent information

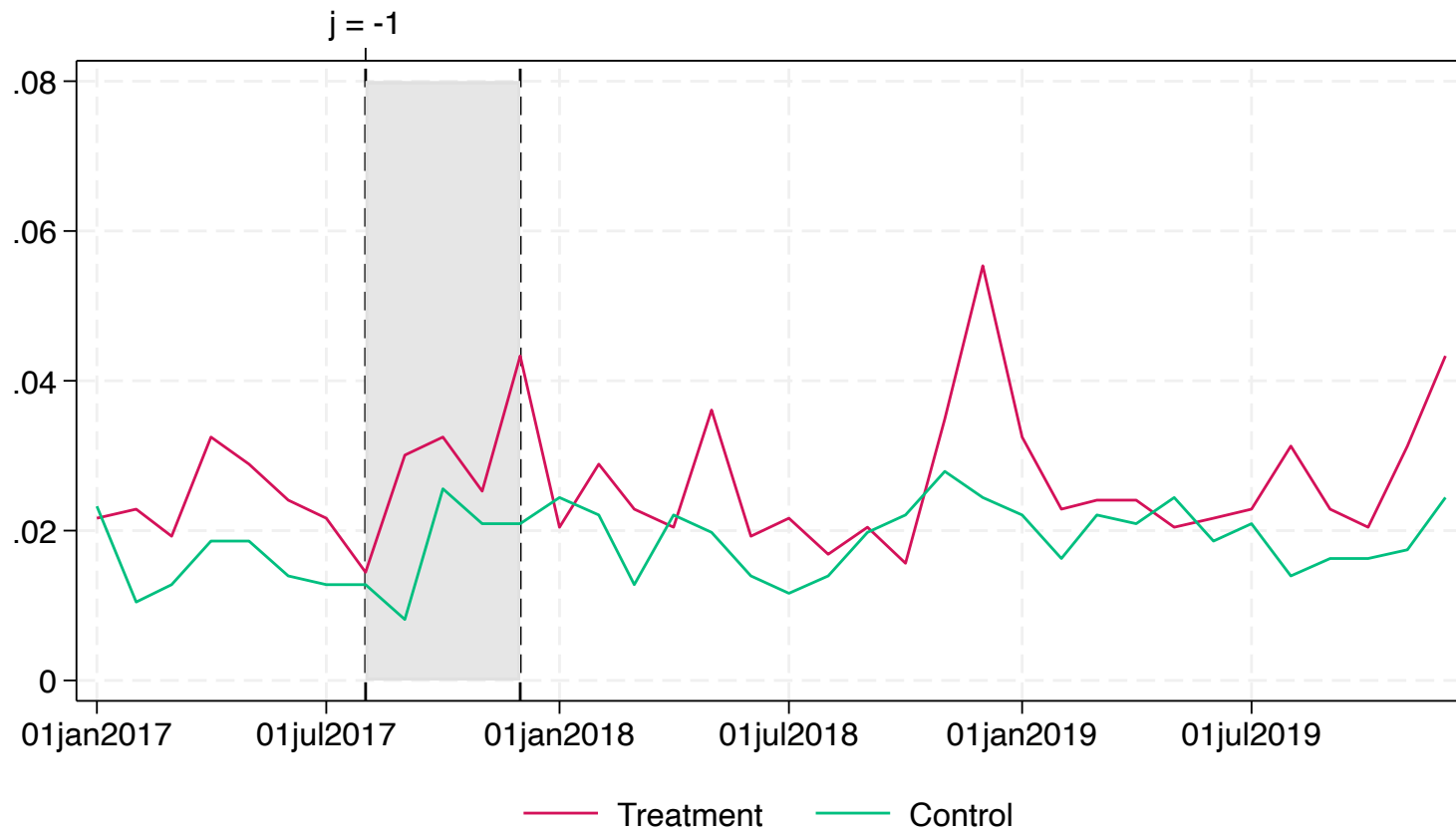
(b) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.77 – Some declaration has implausible information

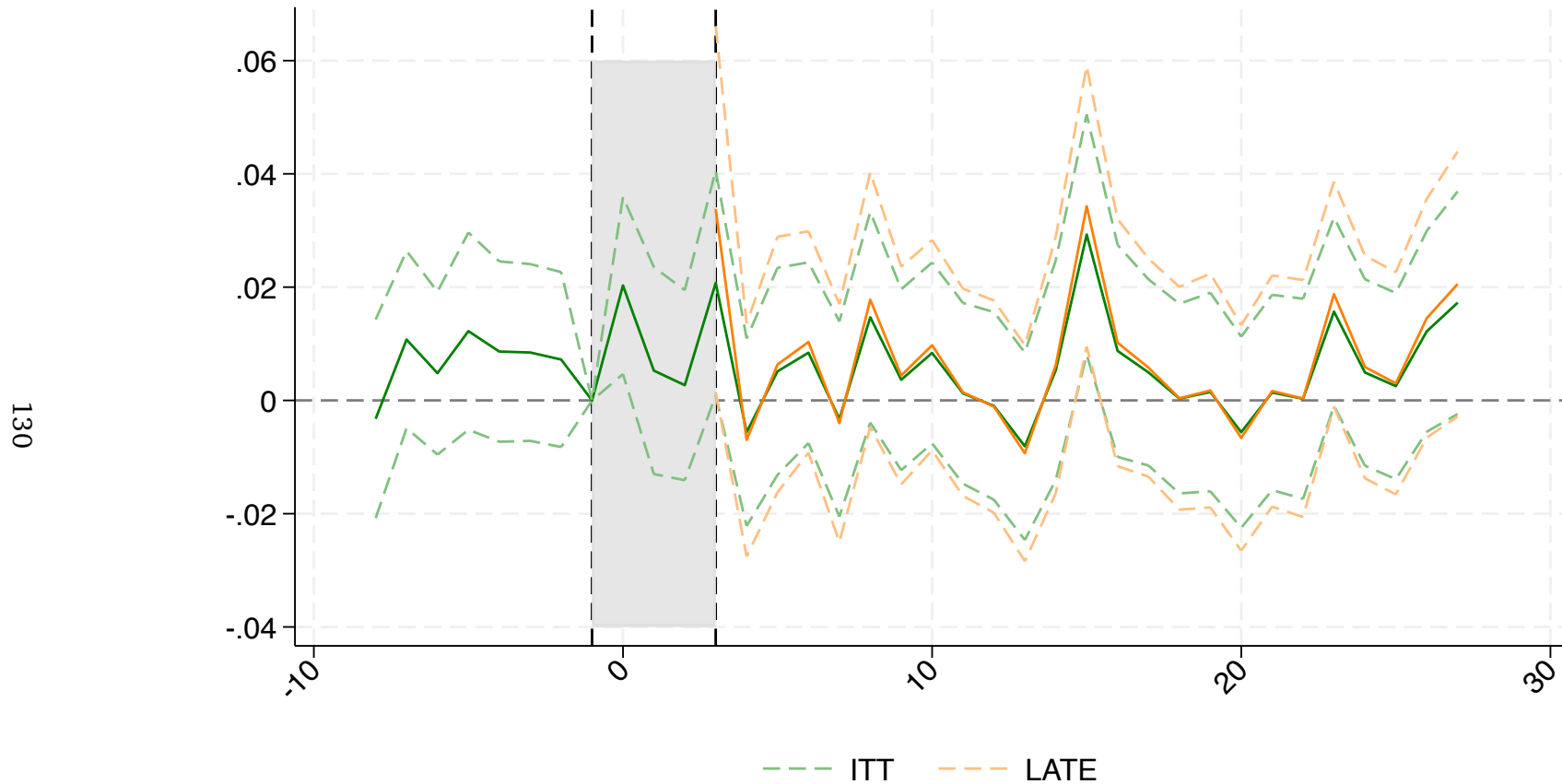
(a) Time trends



The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.77 – Some declaration has implausible information

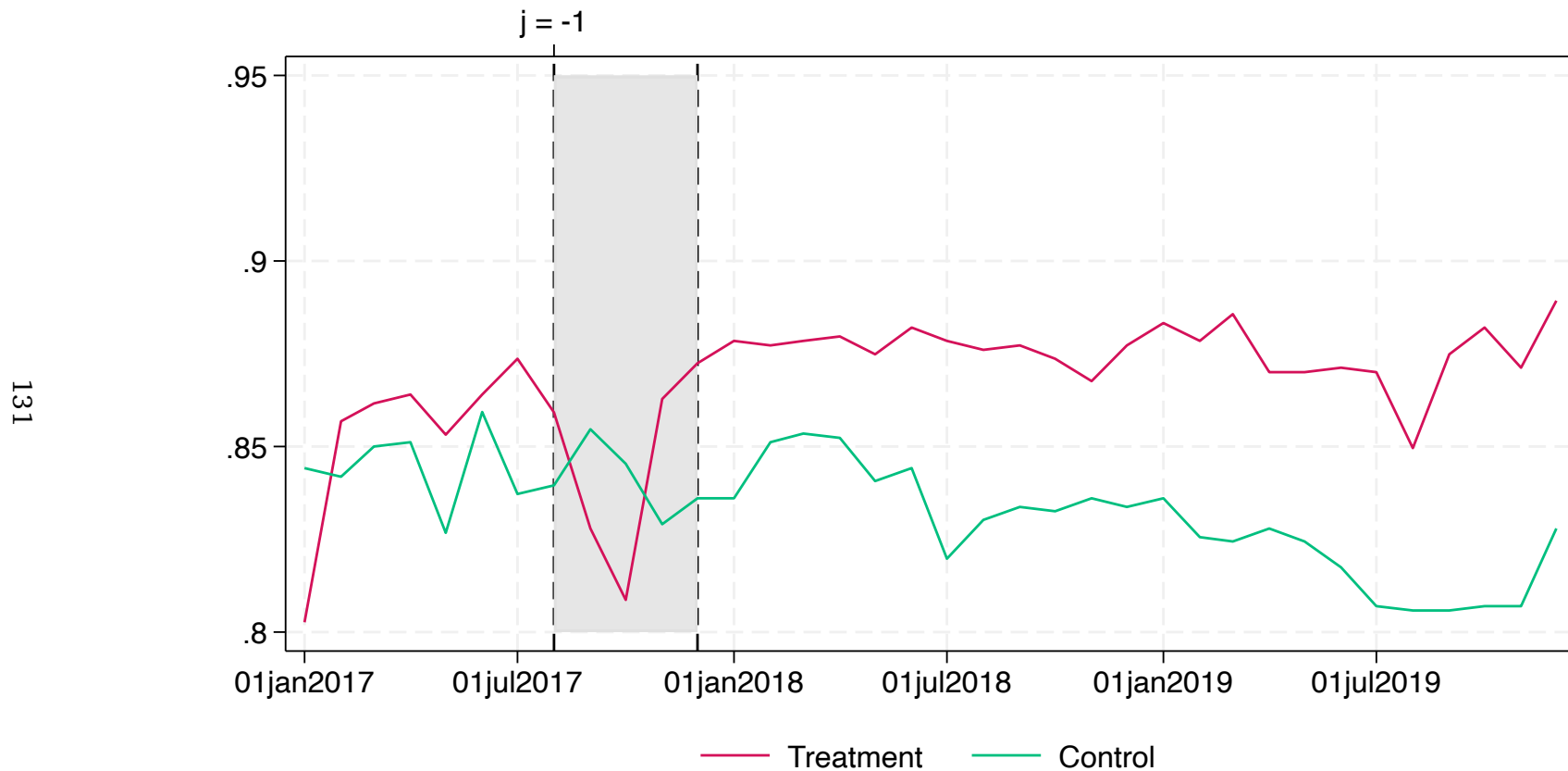
(b) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.78 – Liability declared is positive

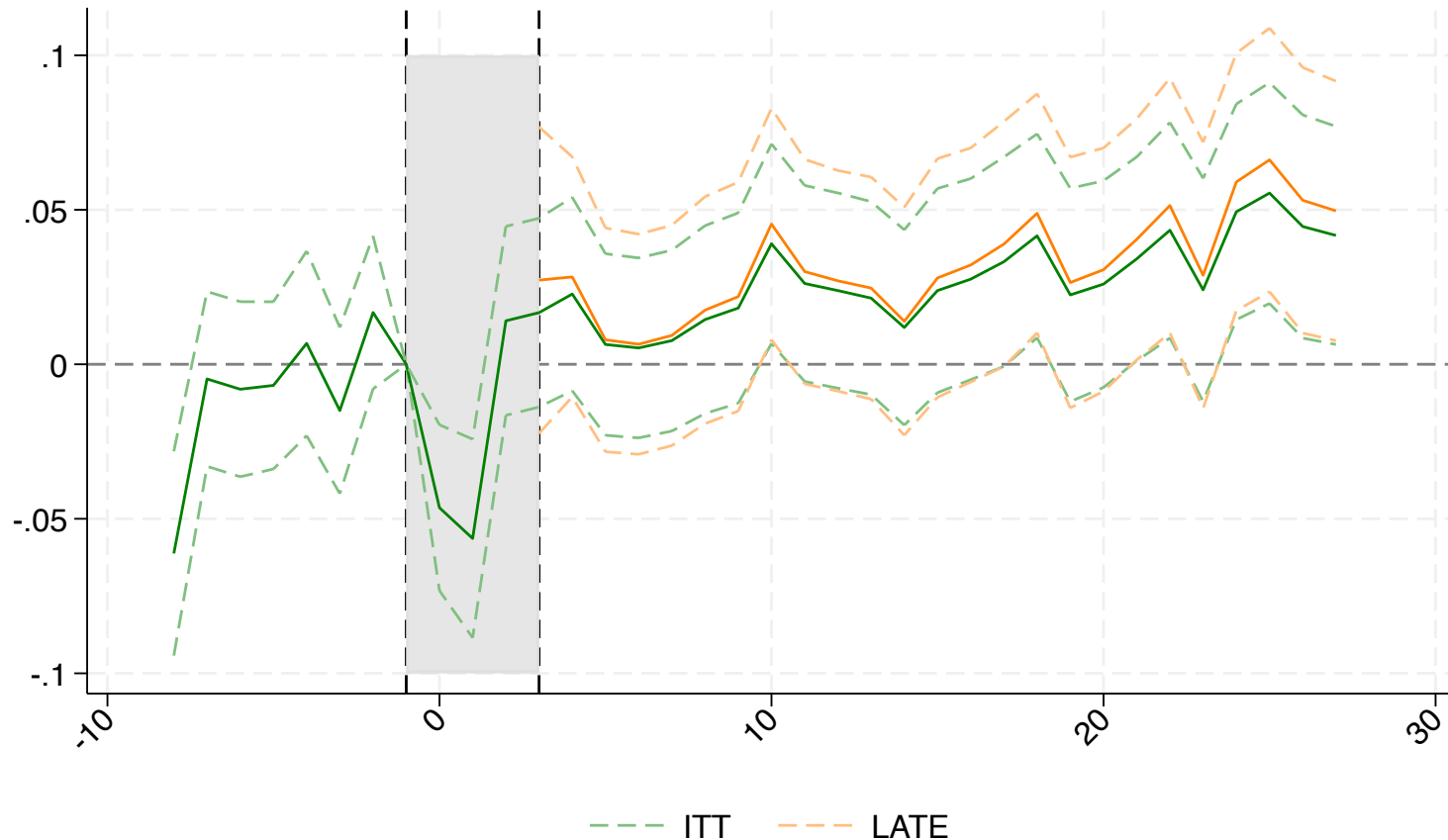
(a) Time trends



The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.78 – Liability declared is positive

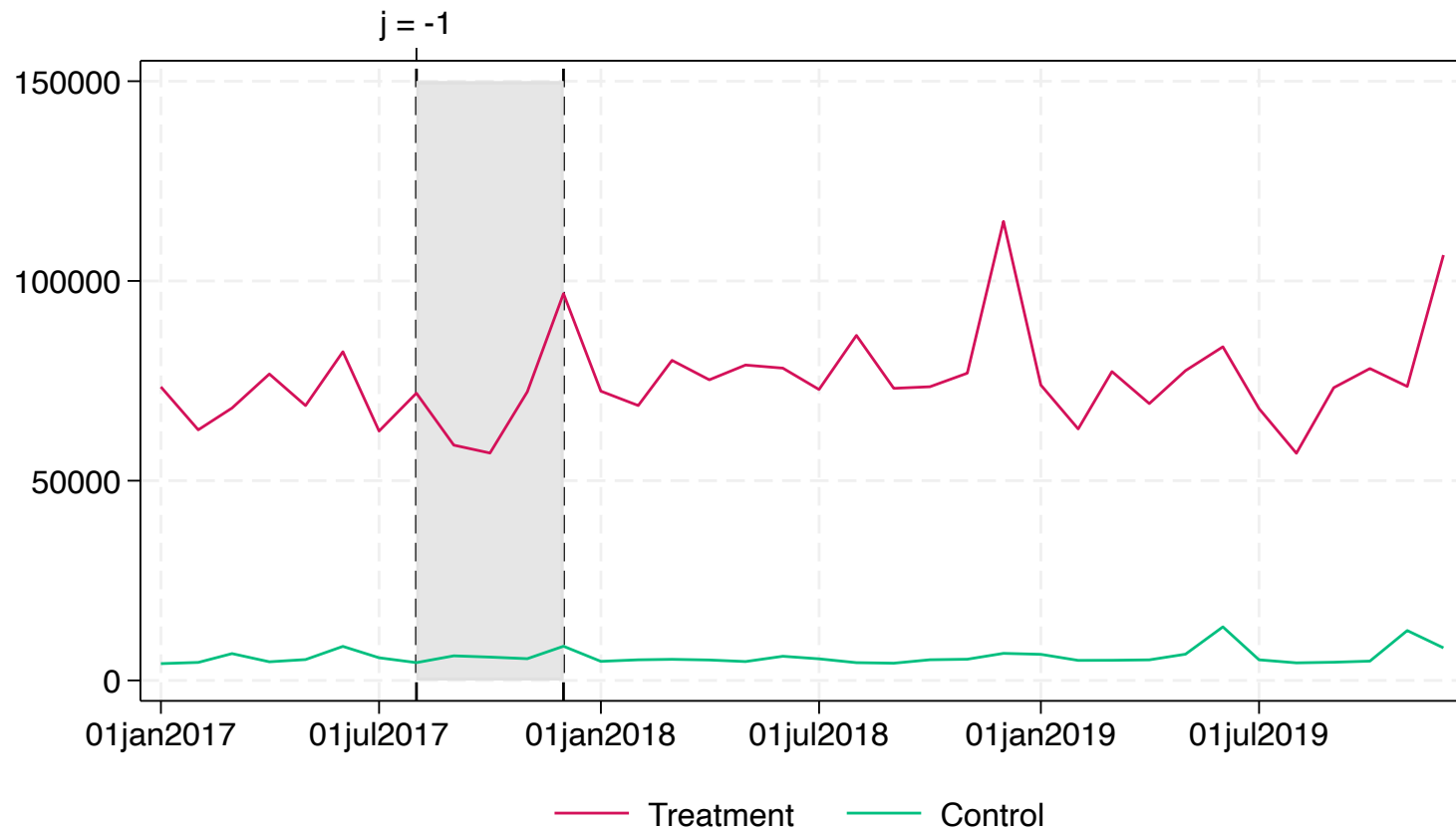
(b) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.79 – Liability declared (USD)

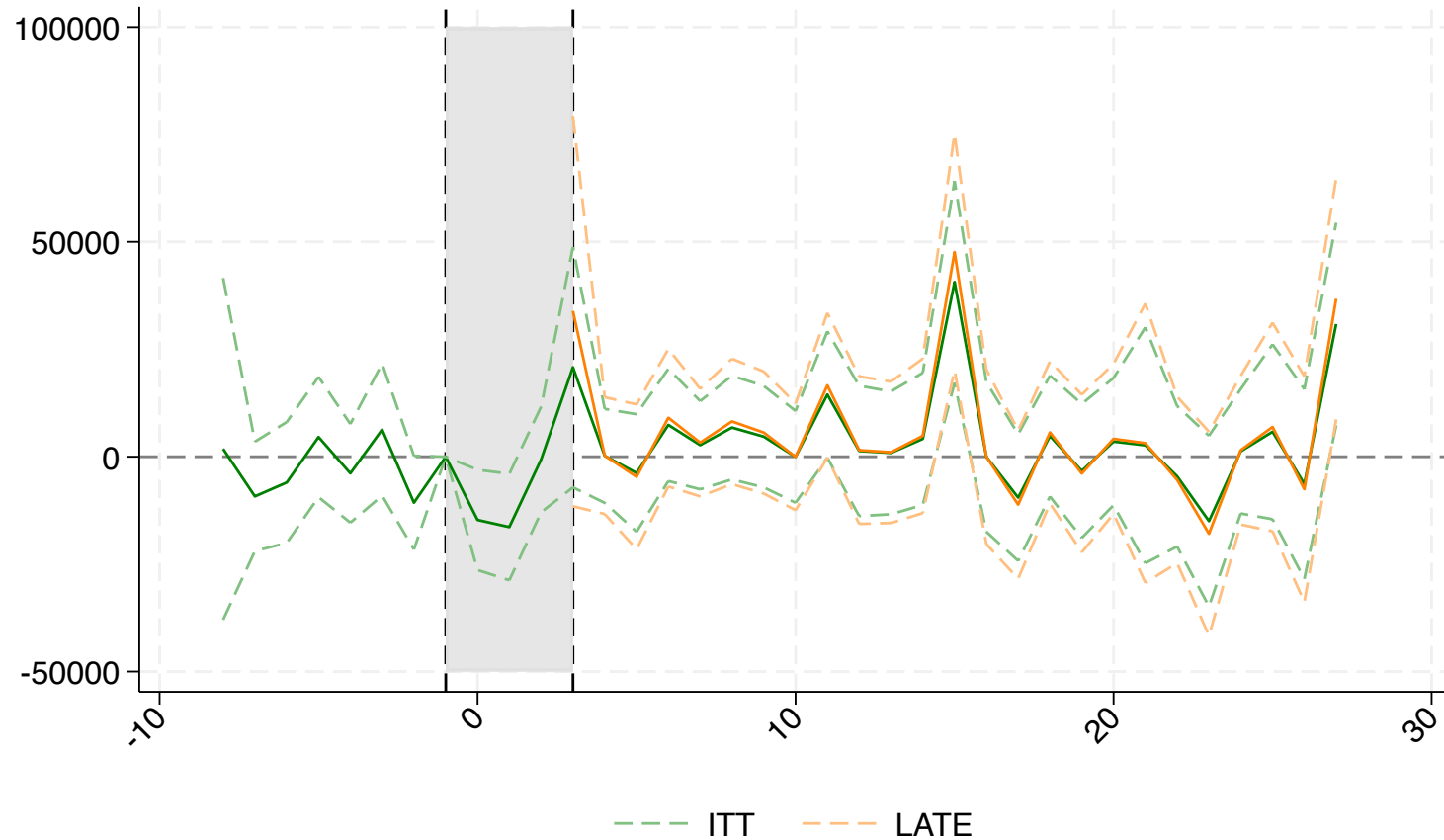
(a) Time trends



The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

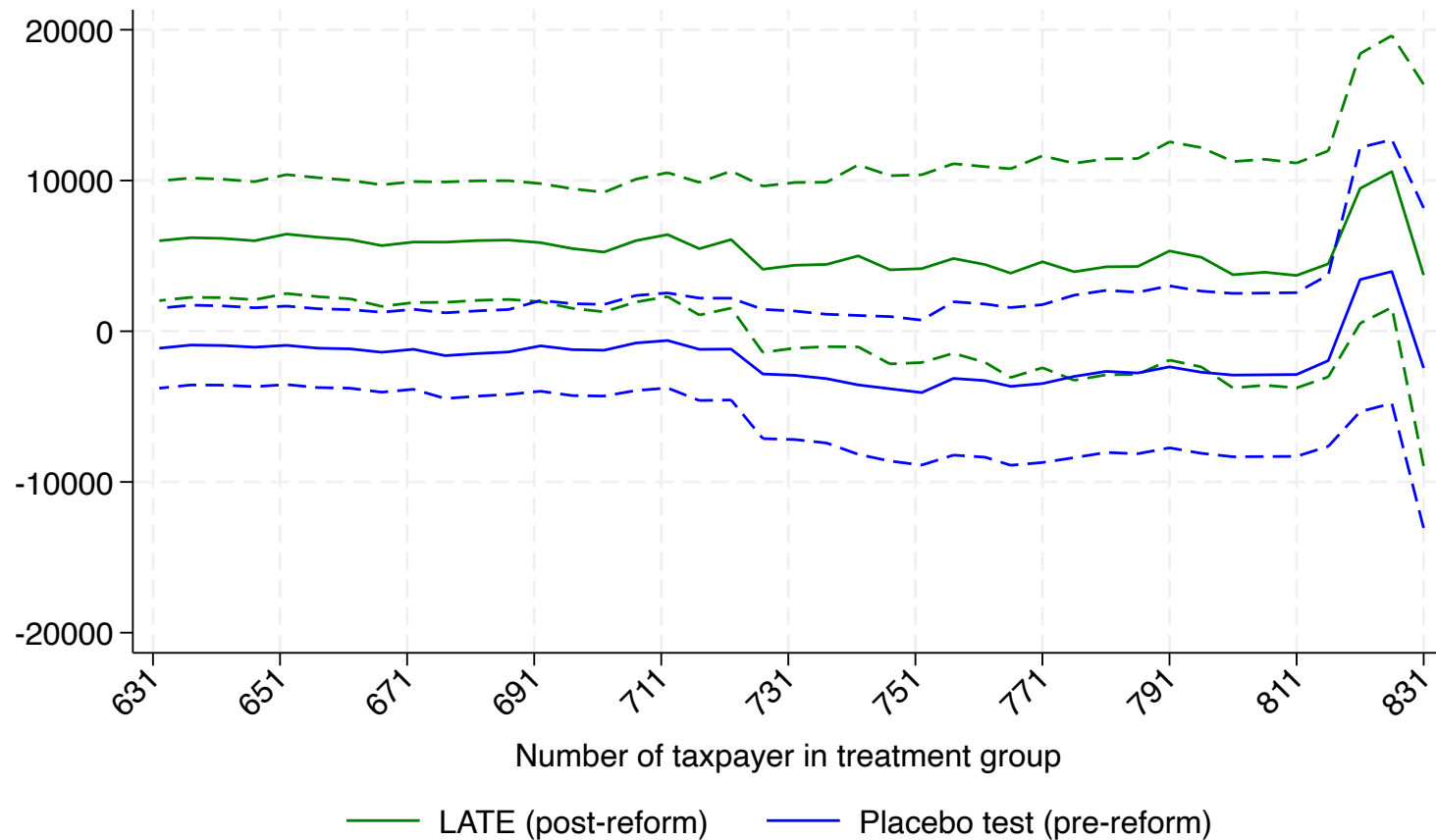
Figure 2.79 – Liability declared (USD)

(b) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

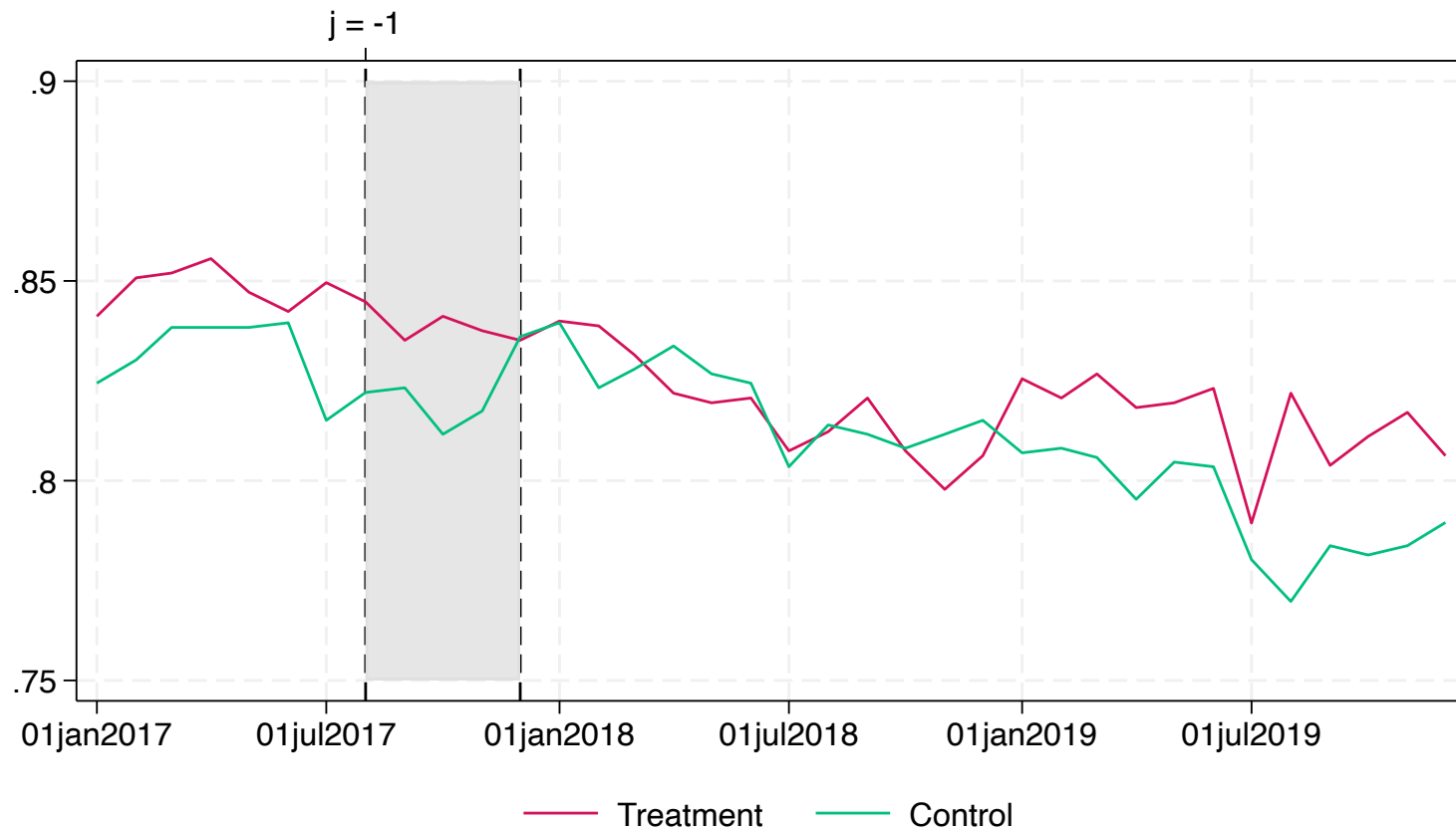
Figure 2.79 – Liability declared (USD)
(c) Sensitivity to extremely large taxpayers



Notes. The Figure displays variations in treatment impact (post-reform) and placebo test (pre-reform) as we trim the treated group (taxpayers monitored by the LTC) by incrementally removing the 5 largest taxpayers. LATE coefficients (green line) are obtained by estimating model 2.2 for the post-reform period, using two stages least square method and instrumenting $(Treat_{it} \cdot Post_{it})$ by an indicator $Etax_{it}$ equal to one if and only if taxpayer i has started using Etax by t . Coefficients of placebo tests (blue line) are obtained by estimating model 2.2 for the pre-reform period replacing $(Treat_{it} \cdot Post_{it})$ by $(Treat_{it} \cdot Pre_{it})$ and using Ordinary Least Square method. The reference period is always August 2017. All samples are balanced, and all estimations use time and taxpayer fixed-effect. The axis indicates the number of taxpayers in the treatment group after trimming. The initial sample size of the treatment group is 831. The size of the control group is 860 and remains constant through all specifications.

Figure 2.710 – Tax paid is positive

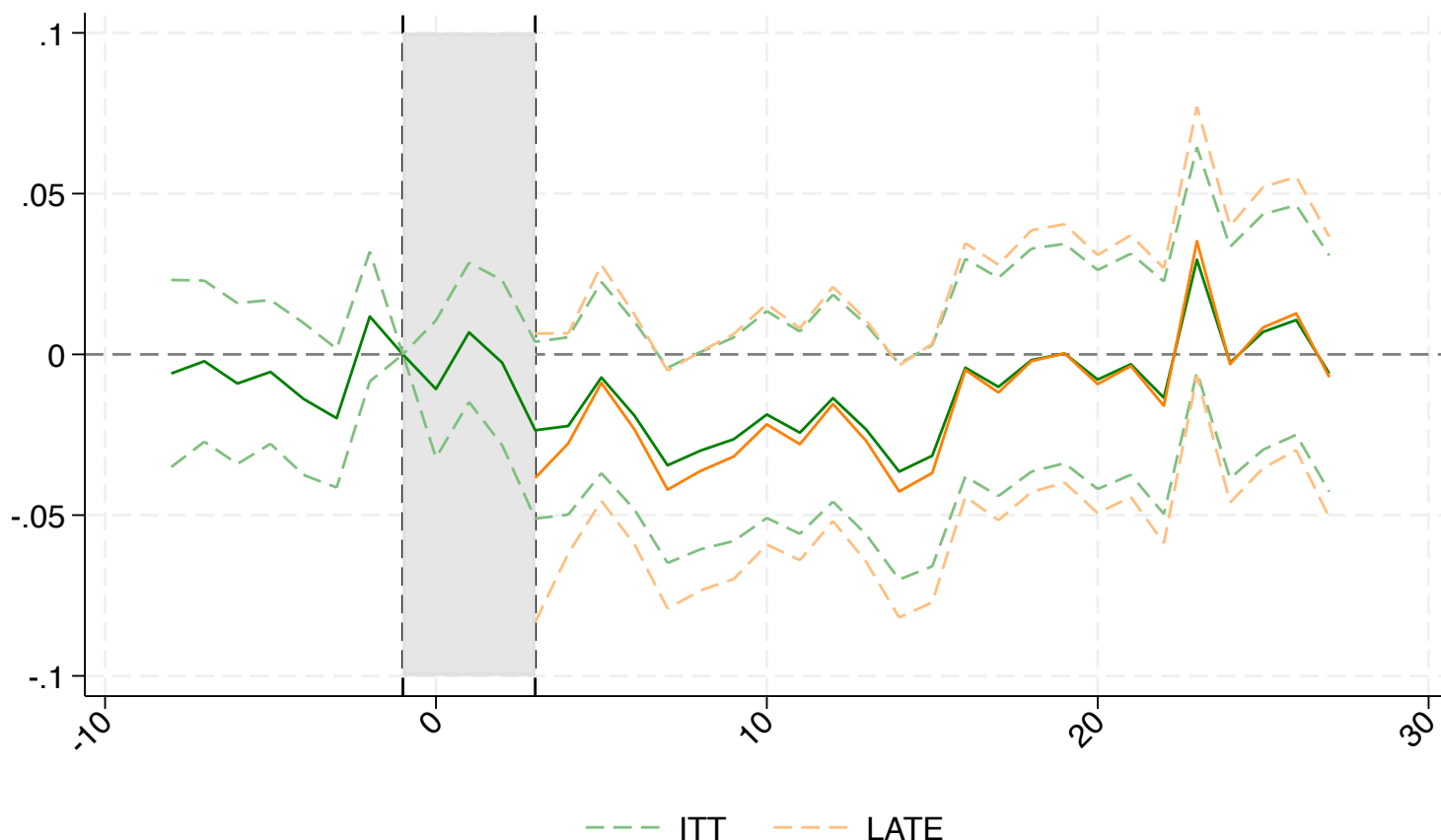
(a) Time trends



The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.710 – Tax paid is positive

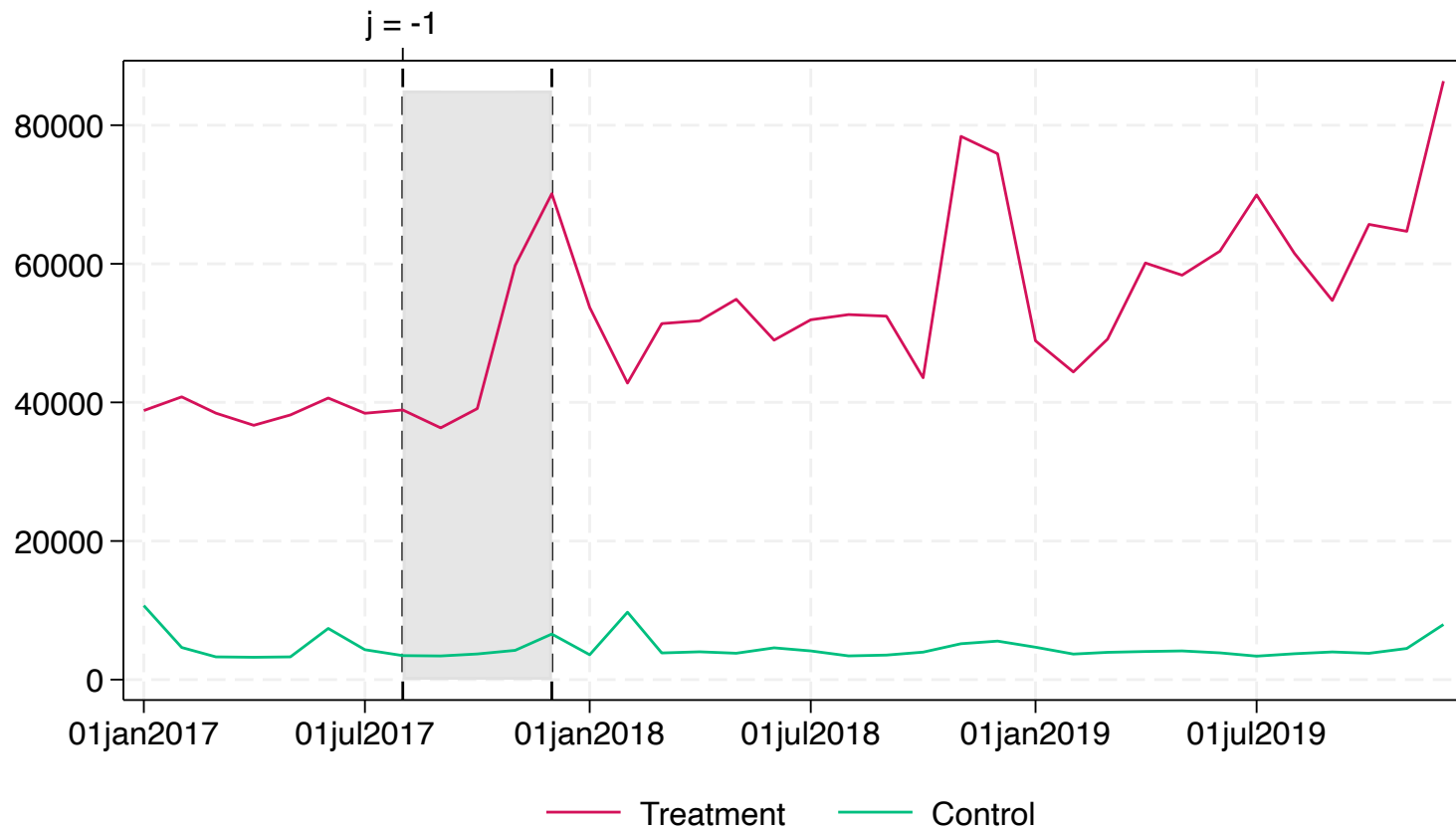
(b) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

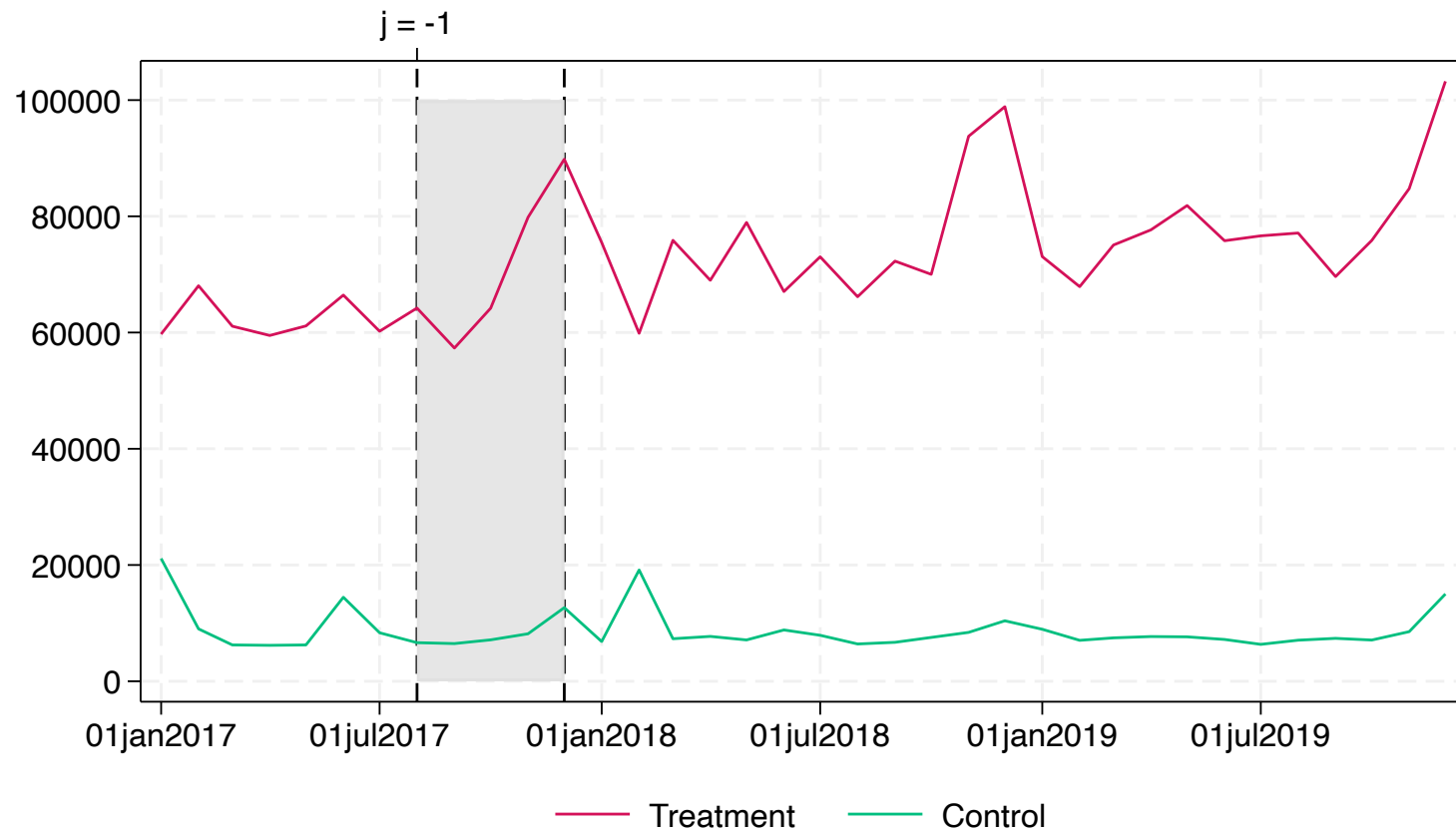
Figure 2.711 – Amount paid (USD)

(a) Time trends - full sample



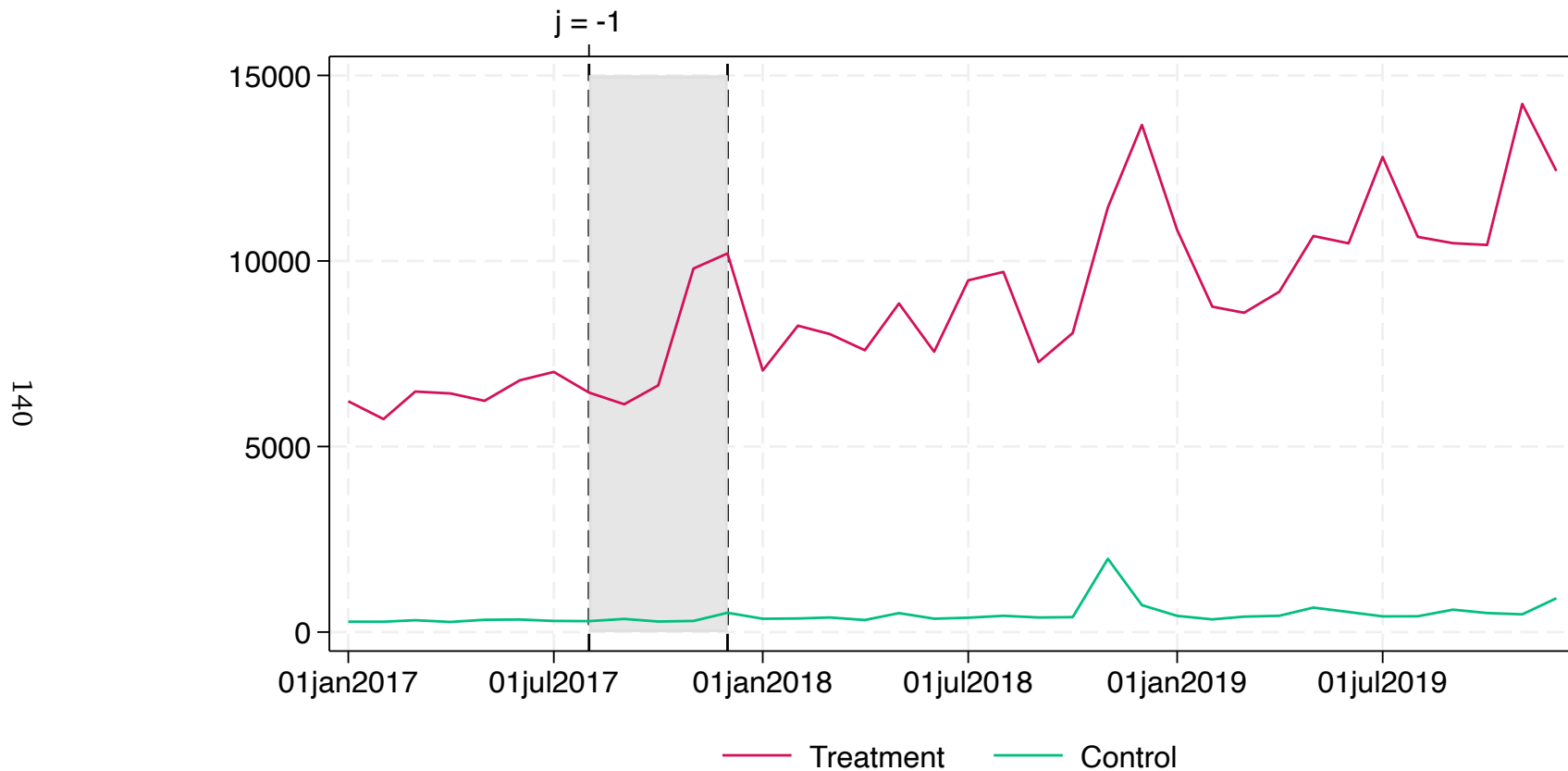
Notes. The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.711 – Amount paid (USD)
(b) Time trends - high compliance group



Notes. The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

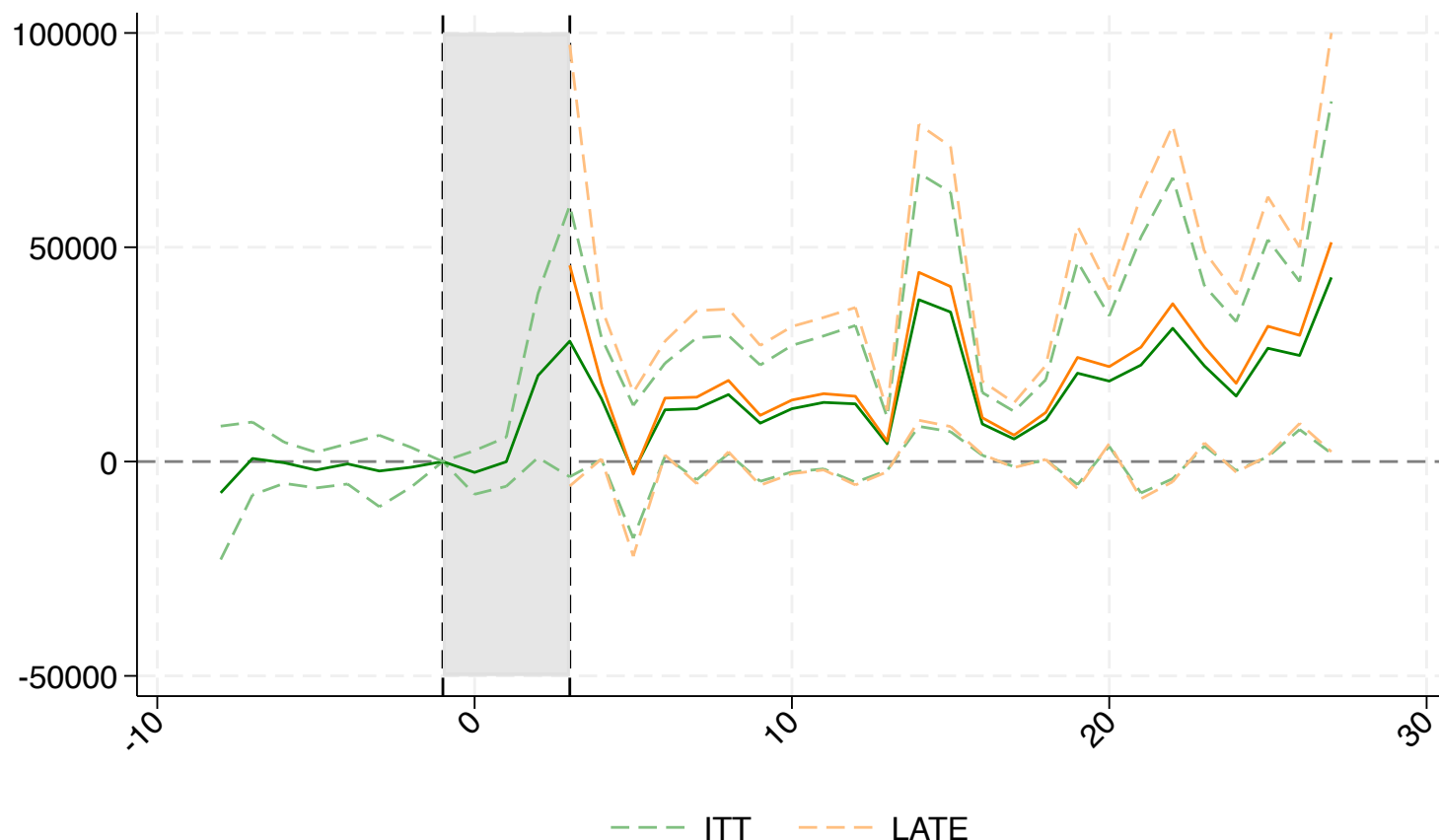
Figure 2.711 – Amount paid (USD)
(c) Time trends - low compliance group



Notes. The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.711 – Amount paid (USD)

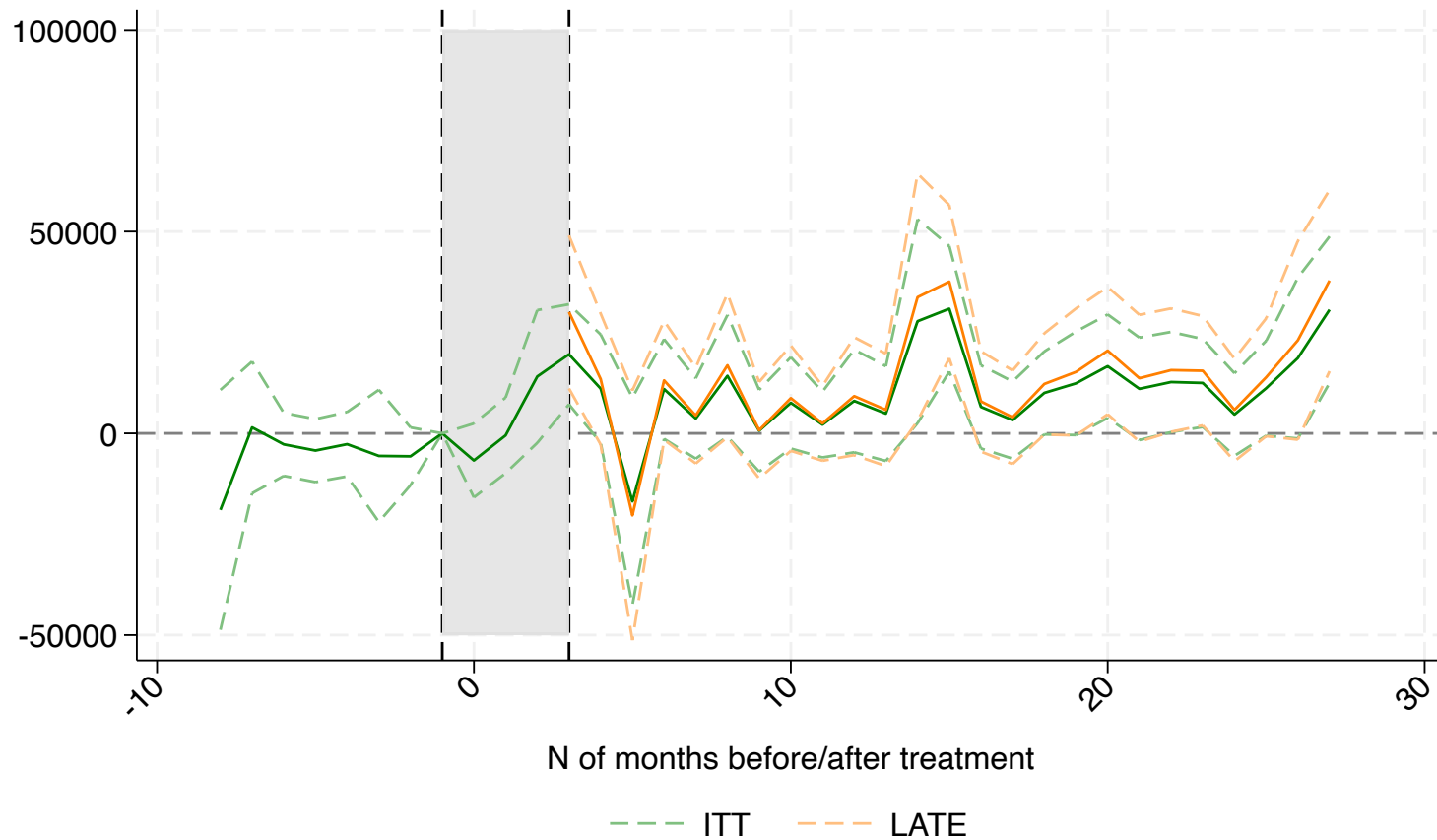
(d) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.711 – Amount paid (USD)

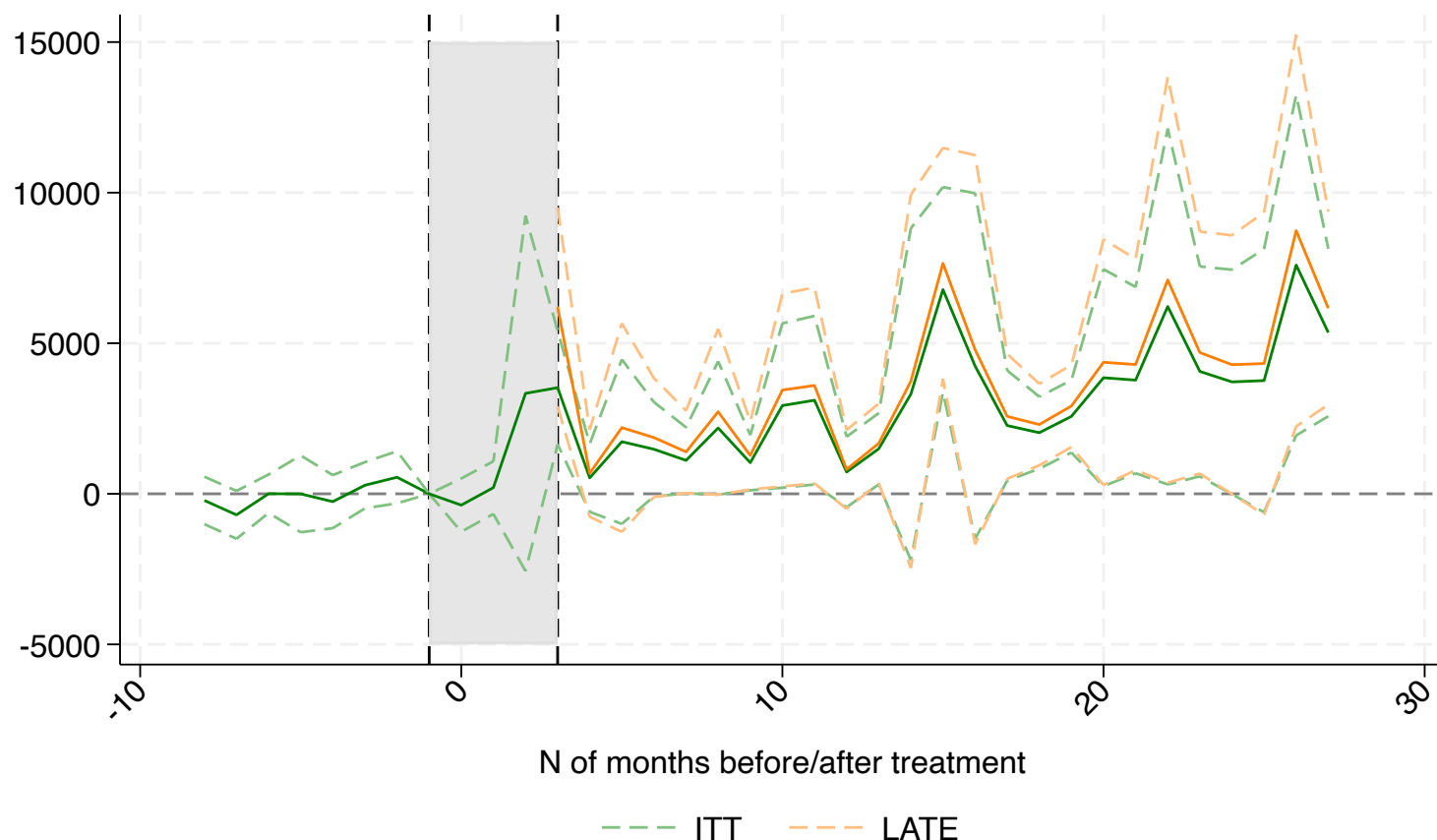
(e) Time-varying impact - high compliance group



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 862 taxpayers observed for exactly 36 months. Total sample size is 31,032. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.711 – Amount paid (USD)

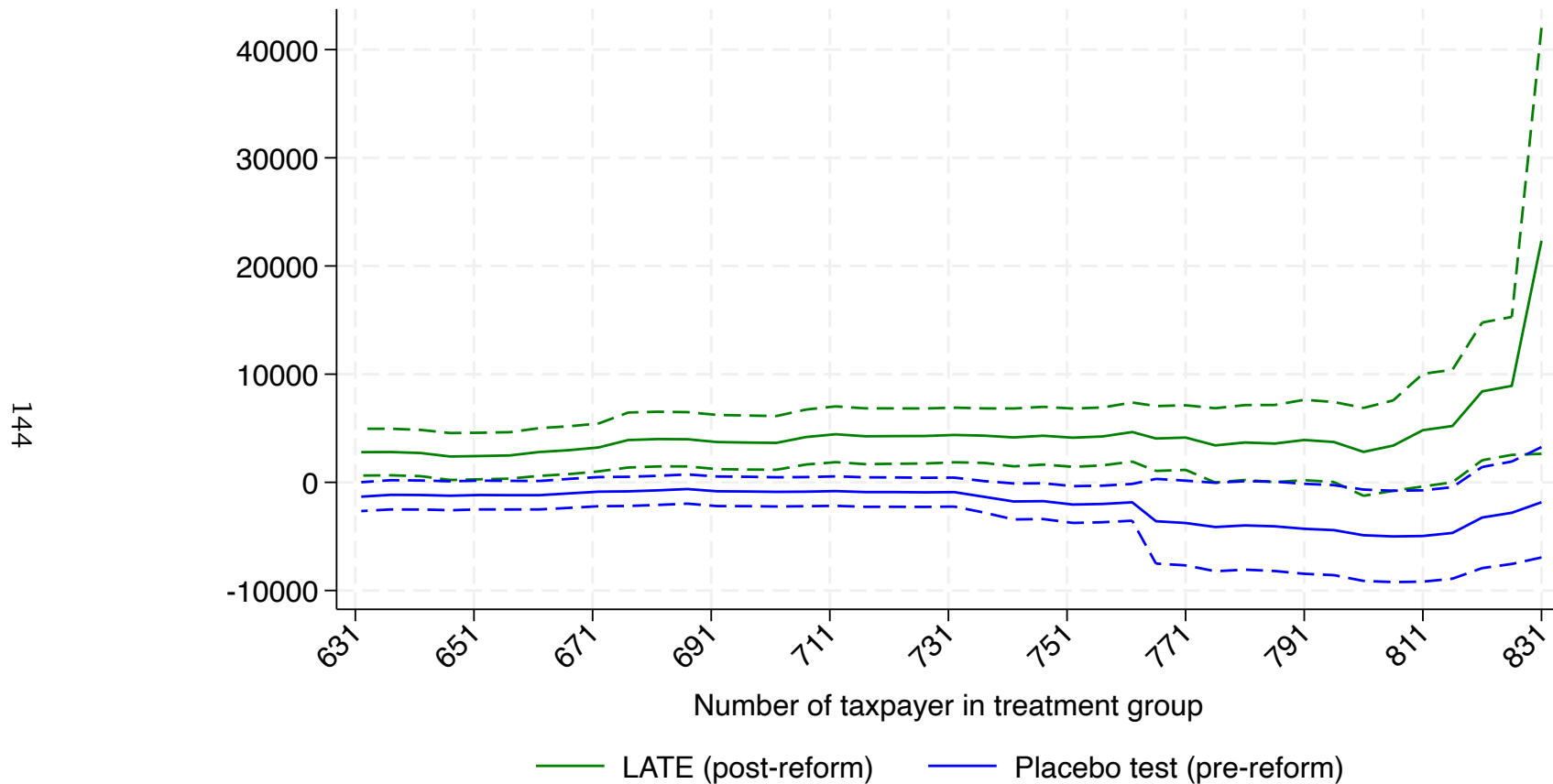
(f) Time-varying impact - low compliance group



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 862 taxpayers observed for exactly 36 months. Total sample size is 31,032. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.711 – Amount paid (USD)

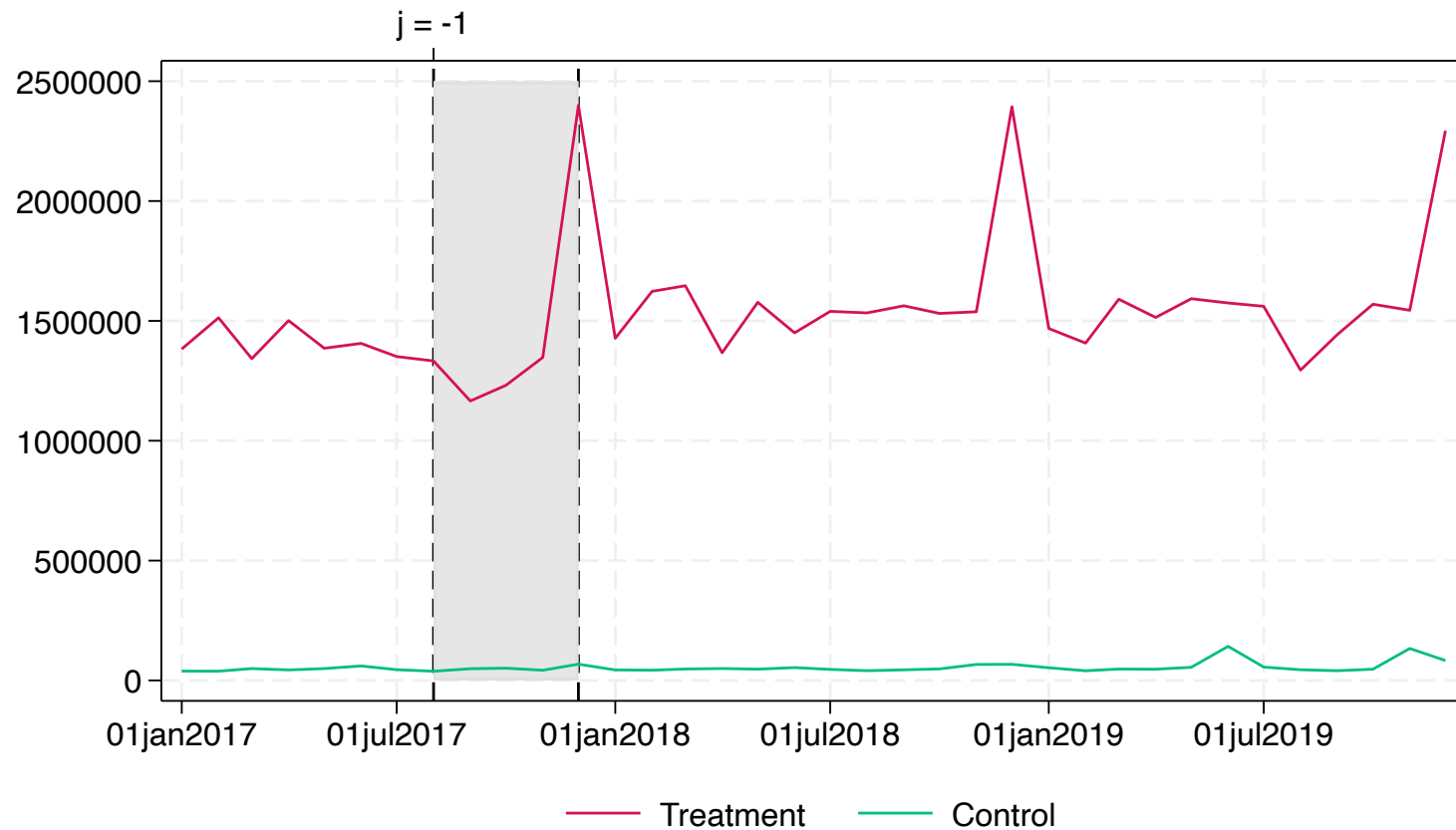
(g) Sensitivity to extremely large taxpayers



Notes. The Figure displays variations in treatment impact (post-reform) and placebo test (pre-reform) as we trim the treated group (taxpayers monitored by the LTC) by incrementally removing the 5 largest taxpayers. LATE coefficients (green line) are obtained by estimating model 2.2 for the post-reform period, using two stages least square method and instrumenting $(Treat_{it} \cdot Post_{it})$ by an indicator $Etax_{it}$ equal to one if and only if taxpayer i has started using Etax by t . Coefficients of placebo tests (blue line) are obtained by estimating model 2.2 for the pre-reform period replacing $(Treat_{it} \cdot Post_{it})$ by $(Treat_{it} \cdot Pre_{it})$ and using Ordinary Least Square method. The reference period is always August 2017. All samples are balanced, and all estimations use time and taxpayer fixed-effect. The axis indicates the number of taxpayers in the treatment group after trimming. The initial sample size of the treatment group is 831. The size of the control group is 860 and remains constant through all specifications.

Figure 2.712 – Turnover (USD)

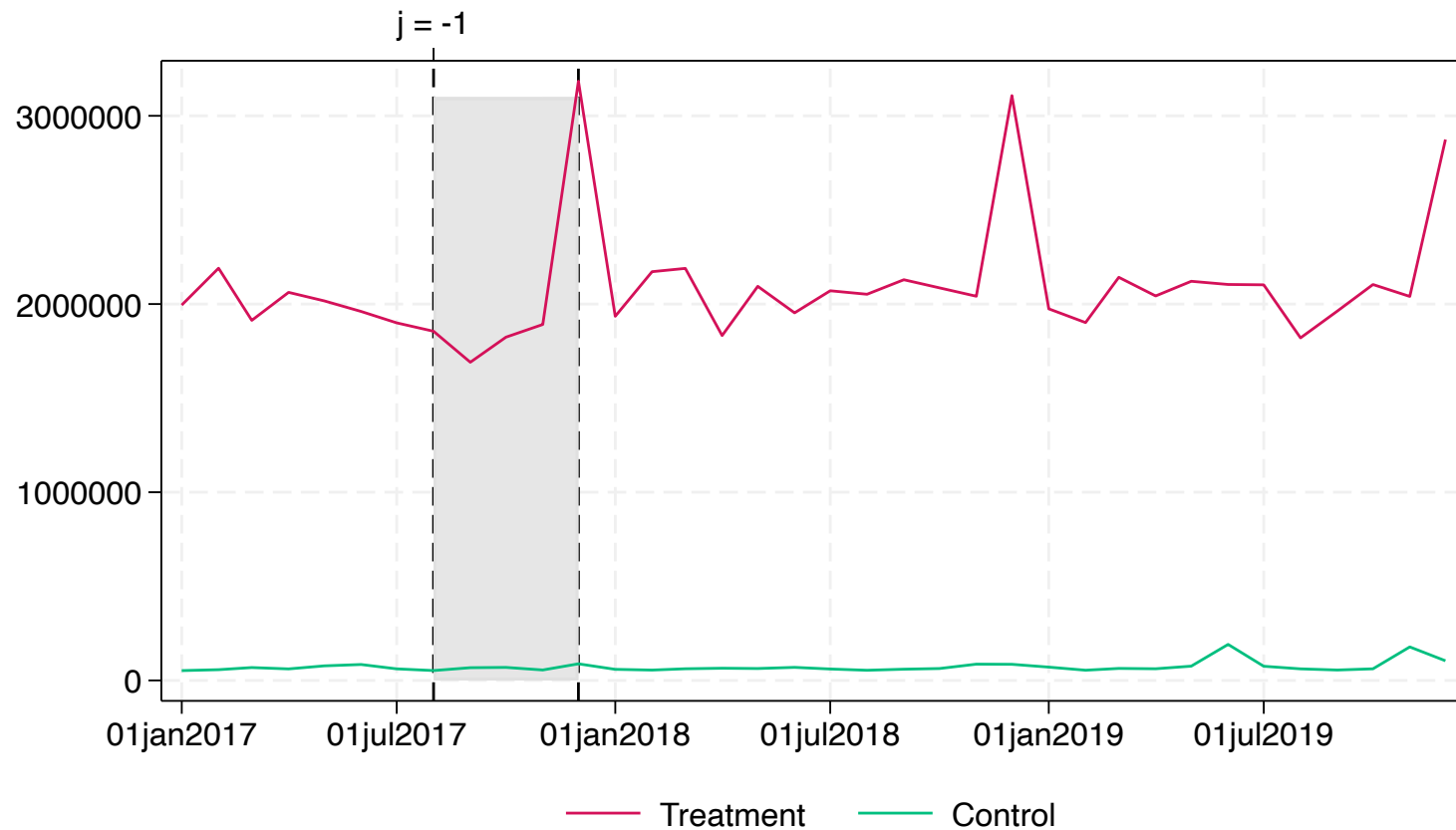
(a) Time trends



Notes. The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.712 – Turnover (USD) - conditional on being positive

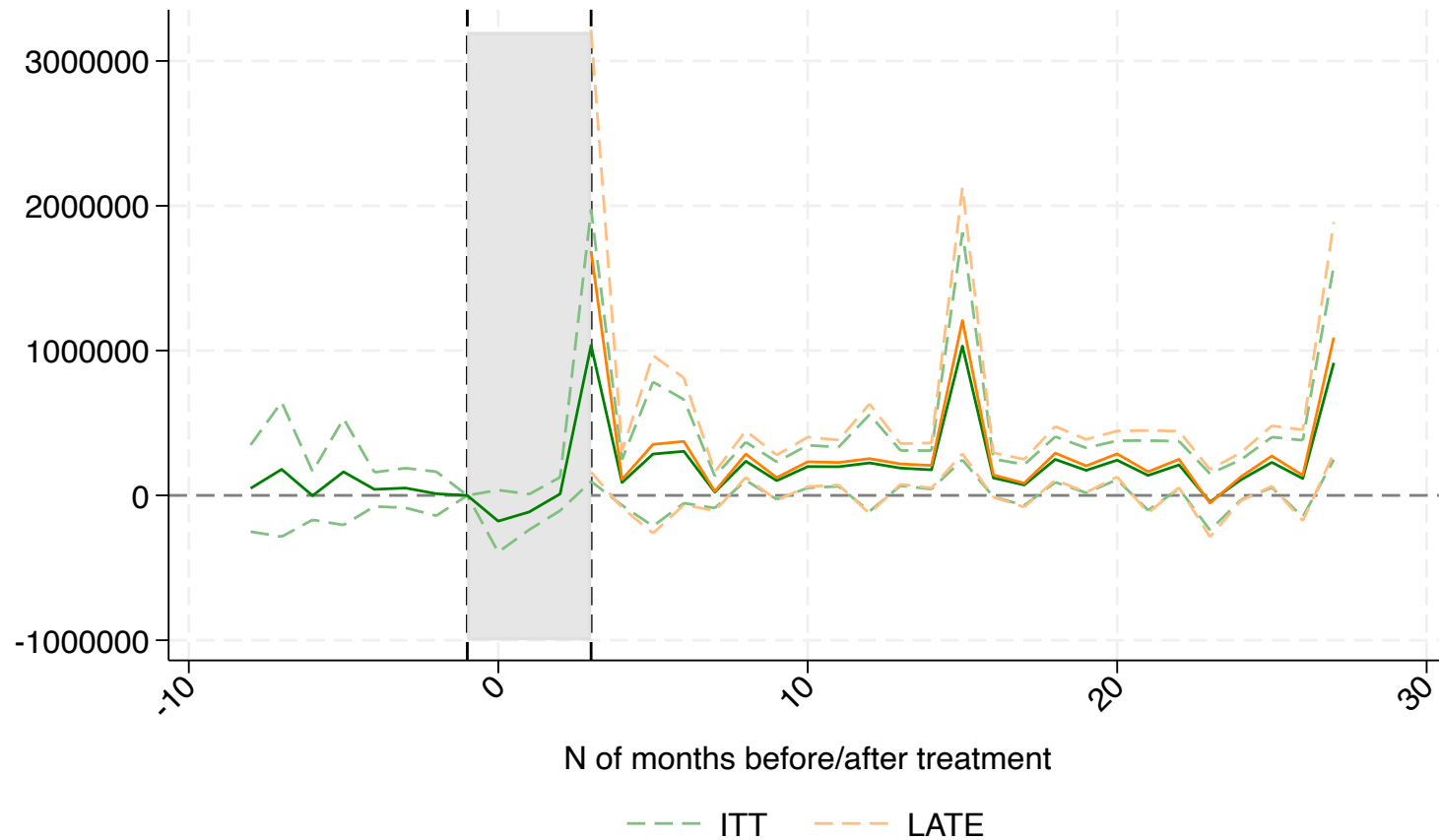
(b) Time trends



Notes. The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is an *unbalanced* sample of 1,549 taxpayers.

Figure 2.712 – Turnover (USD)

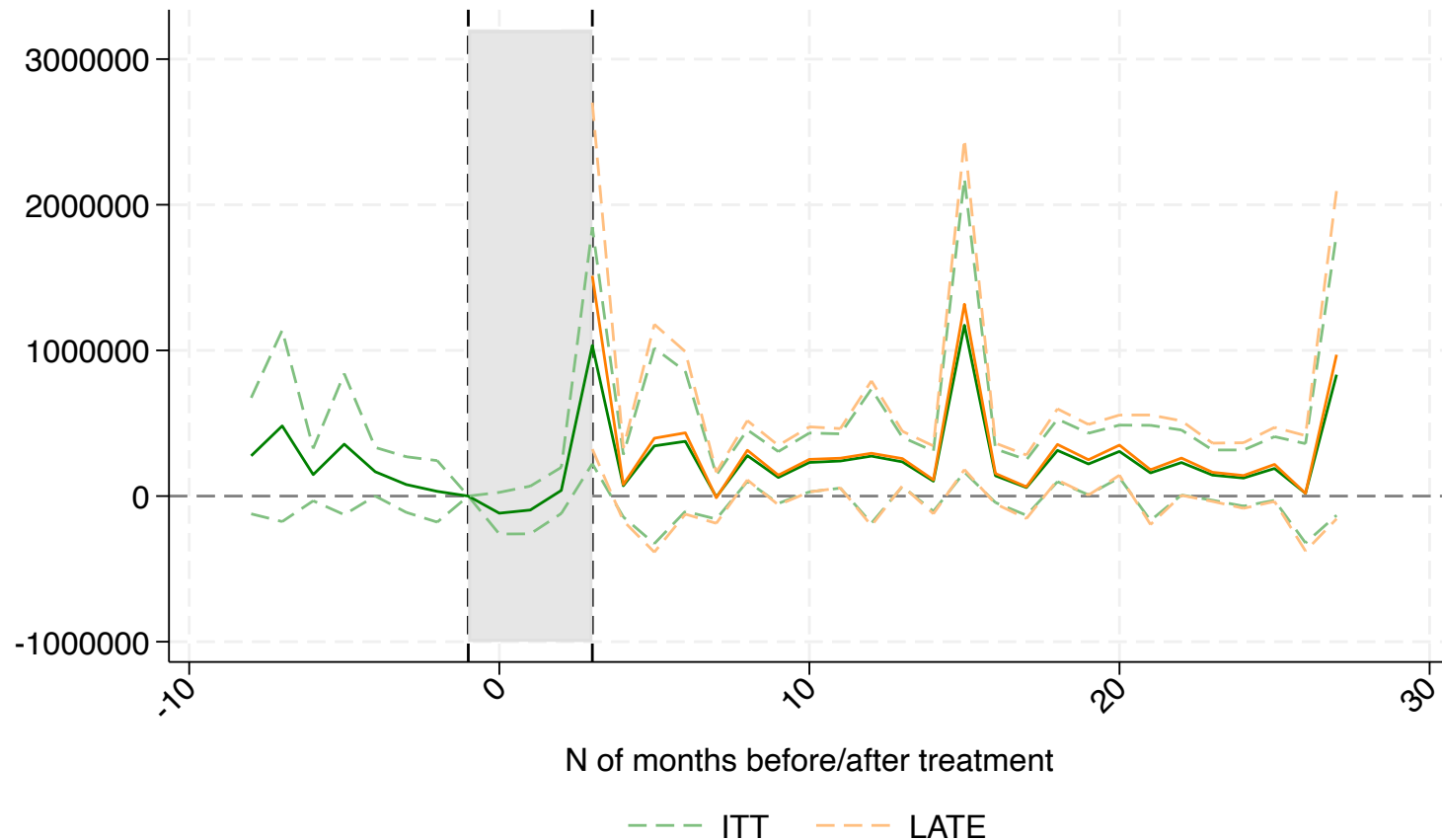
(c) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.712 – Turnover (USD) - conditional on being positive

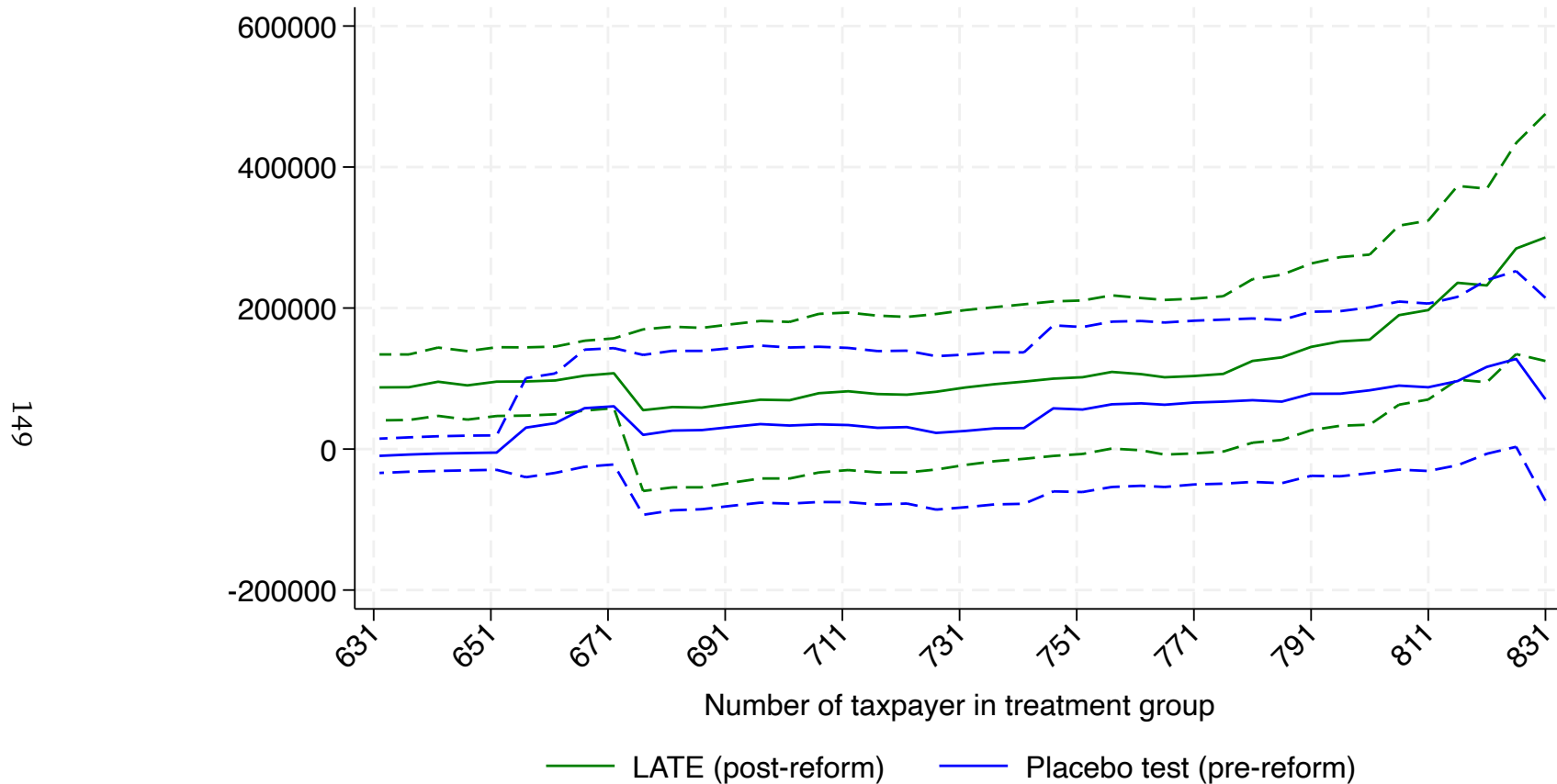
(d) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is an unbalanced sample of 1,549 taxpayers. Total sample size is 44,958. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.712 – Turnover (USD)

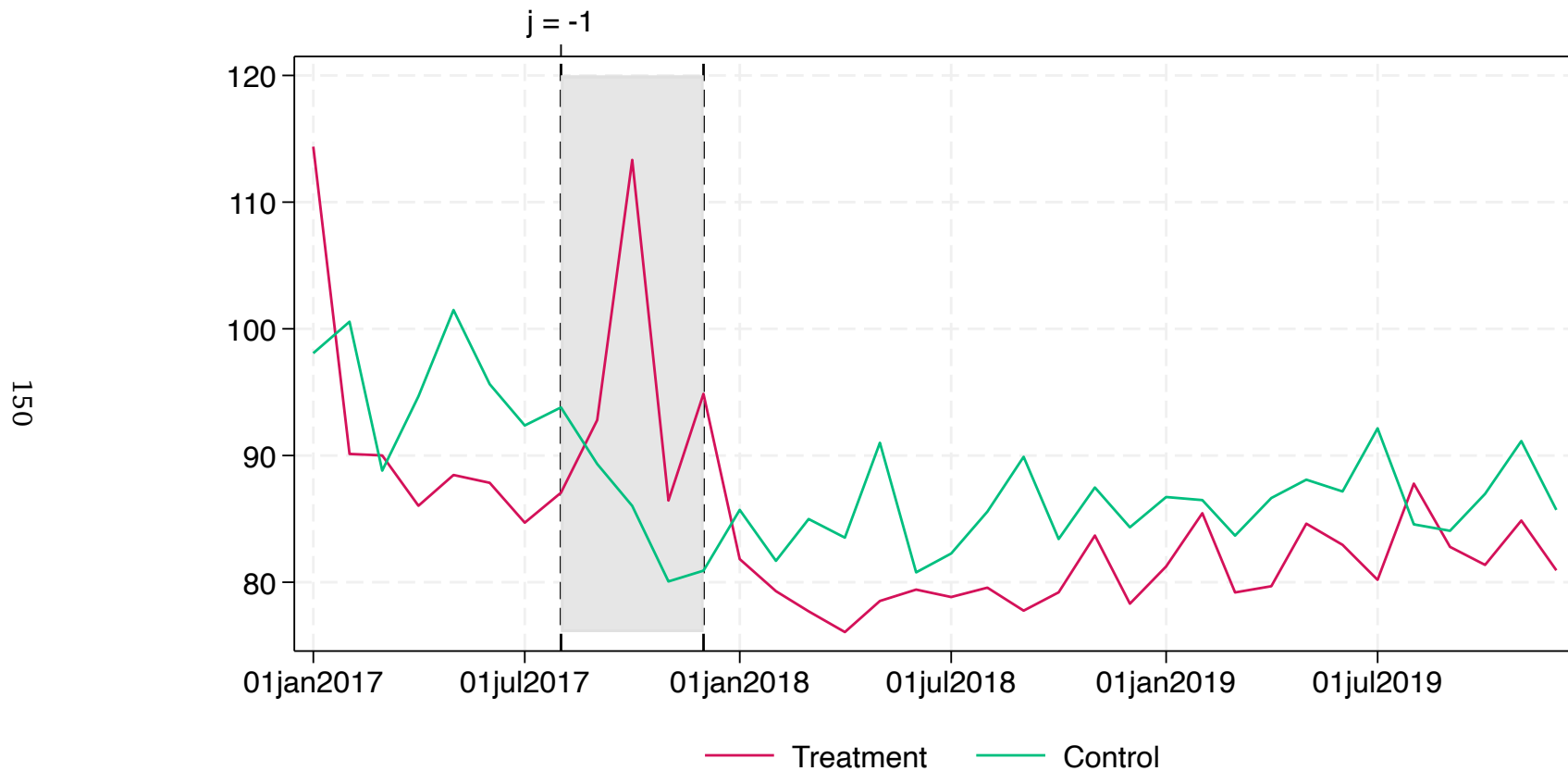
(e) Sensitivity to extremely large taxpayers



Notes. The Figure displays variations in treatment impact (post-reform) and placebo test (pre-reform) as we trim the treated group (taxpayers monitored by the LTC) by incrementally removing the 5 largest taxpayers. LATE coefficients (green line) are obtained by estimating model 2.2 for the post-reform period, using two stages least square method and instrumenting $(Treat_{it} \cdot Post_{it})$ by an indicator $Etax_{it}$ equal to one if and only if taxpayer i has started using Etax by t . Coefficients of placebo tests (blue line) are obtained by estimating model 2.2 for the pre-reform period replacing $(Treat_{it} \cdot Post_{it})$ by $(Treat_{it} \cdot Pre_{it})$ and using Ordinary Least Square method. The reference period is always August 2017. All samples are balanced, and all estimations use time and taxpayer fixed-effect. The axis indicates the number of taxpayers in the treatment group after trimming. The initial sample size of the treatment group is 831. The size of the control group is 860 and remains constant through all specifications.

Figure 2.713 – Payment as a percentage of total liability paid

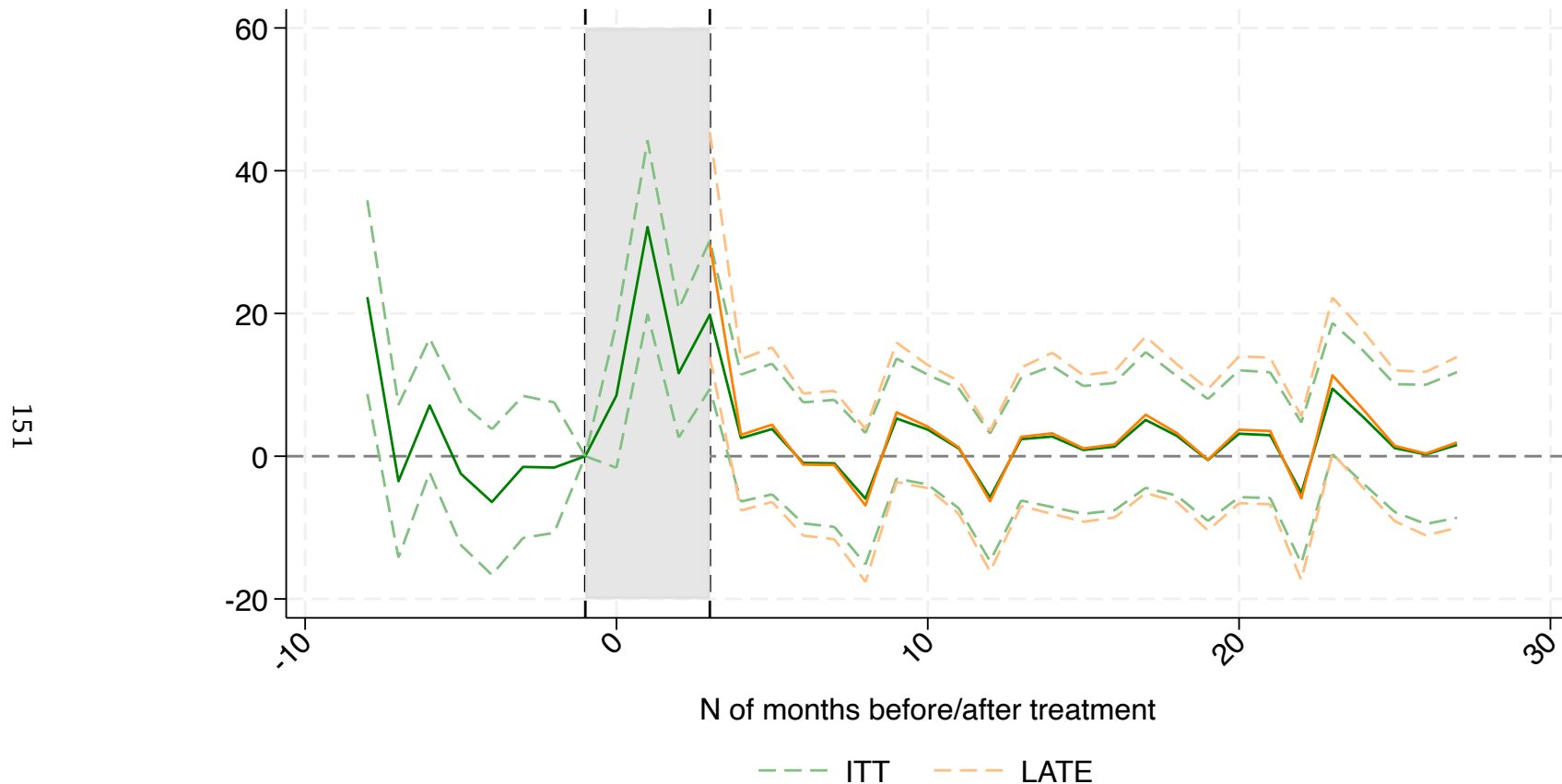
(a) Time trends



Notes. The figure displays average values of the outcome by group and across time. Treated taxpayers taxpayers monitored by the Large Taxpayers Center (LTC), where the adoption of Etax was made mandatory by a ministerial decree issued in July 2017. Controlled taxpayers are the largest 860 taxpayers monitored by other centers of Dakar. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876.

Figure 2.713 – Payment as a percentage of total liability paid

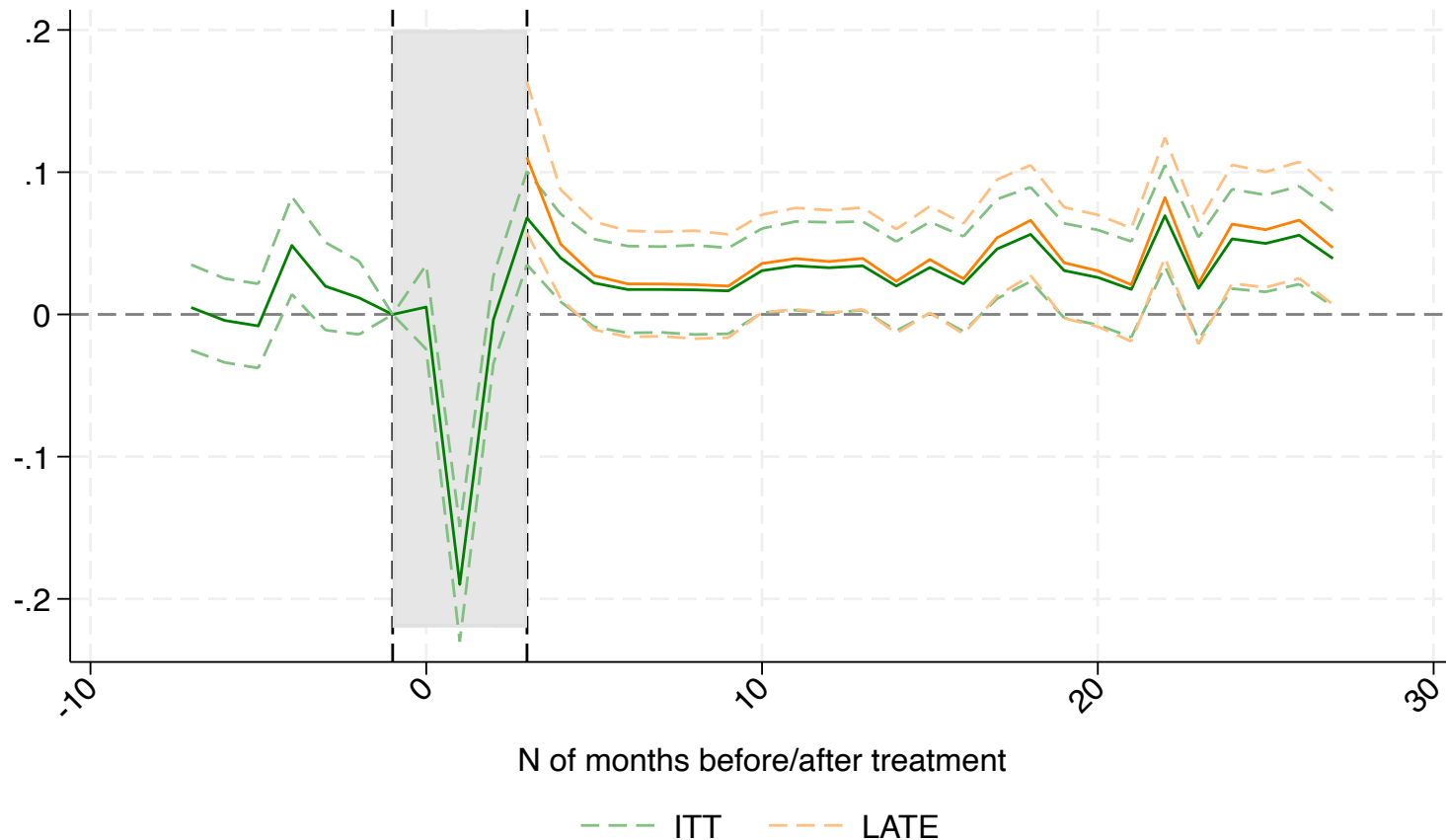
(b) Time-varying impact



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.714 – On measurement and missing information

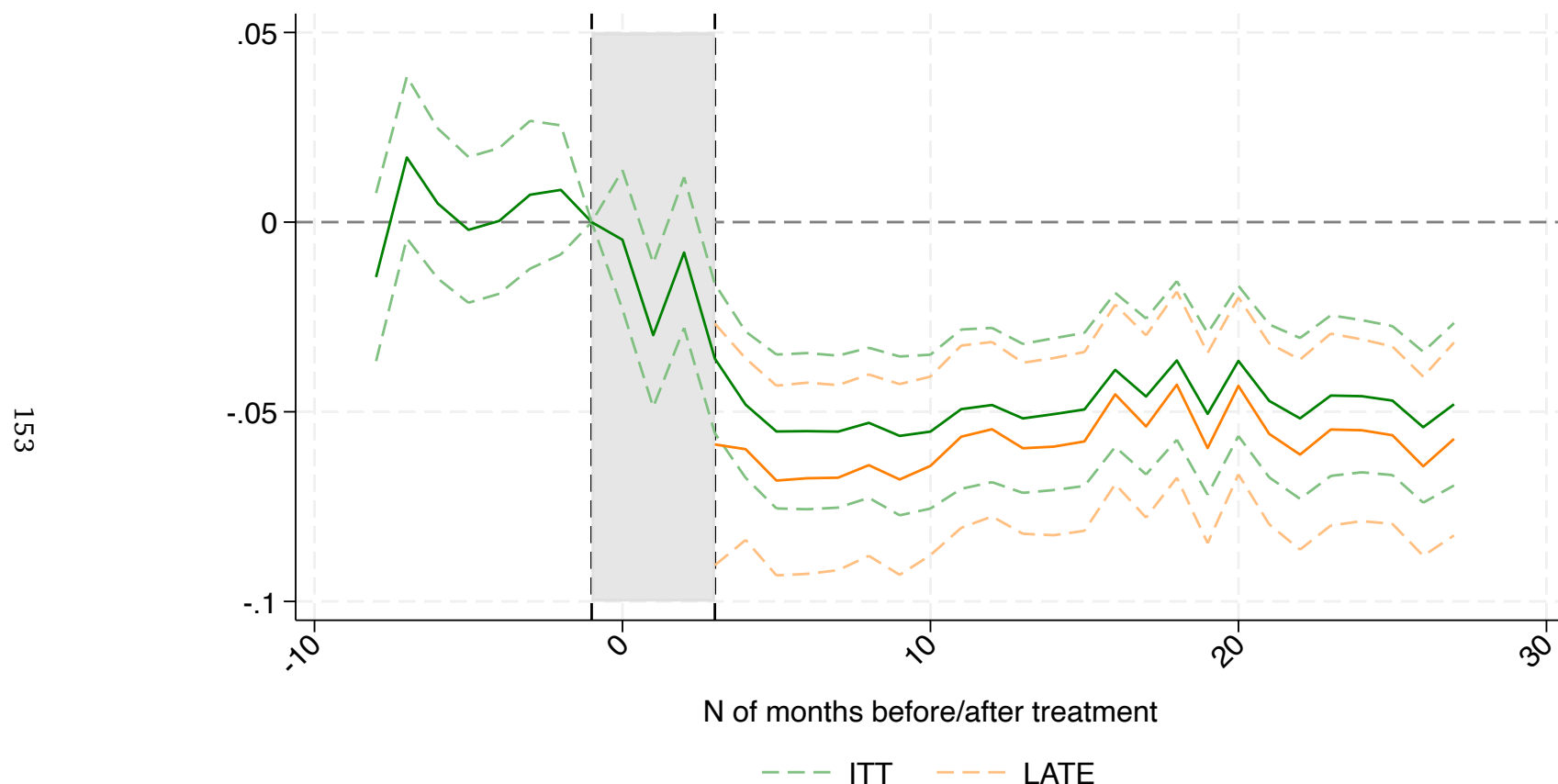
(a) Outcome : Declares withholding on employees



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.714 – On measurement and missing information

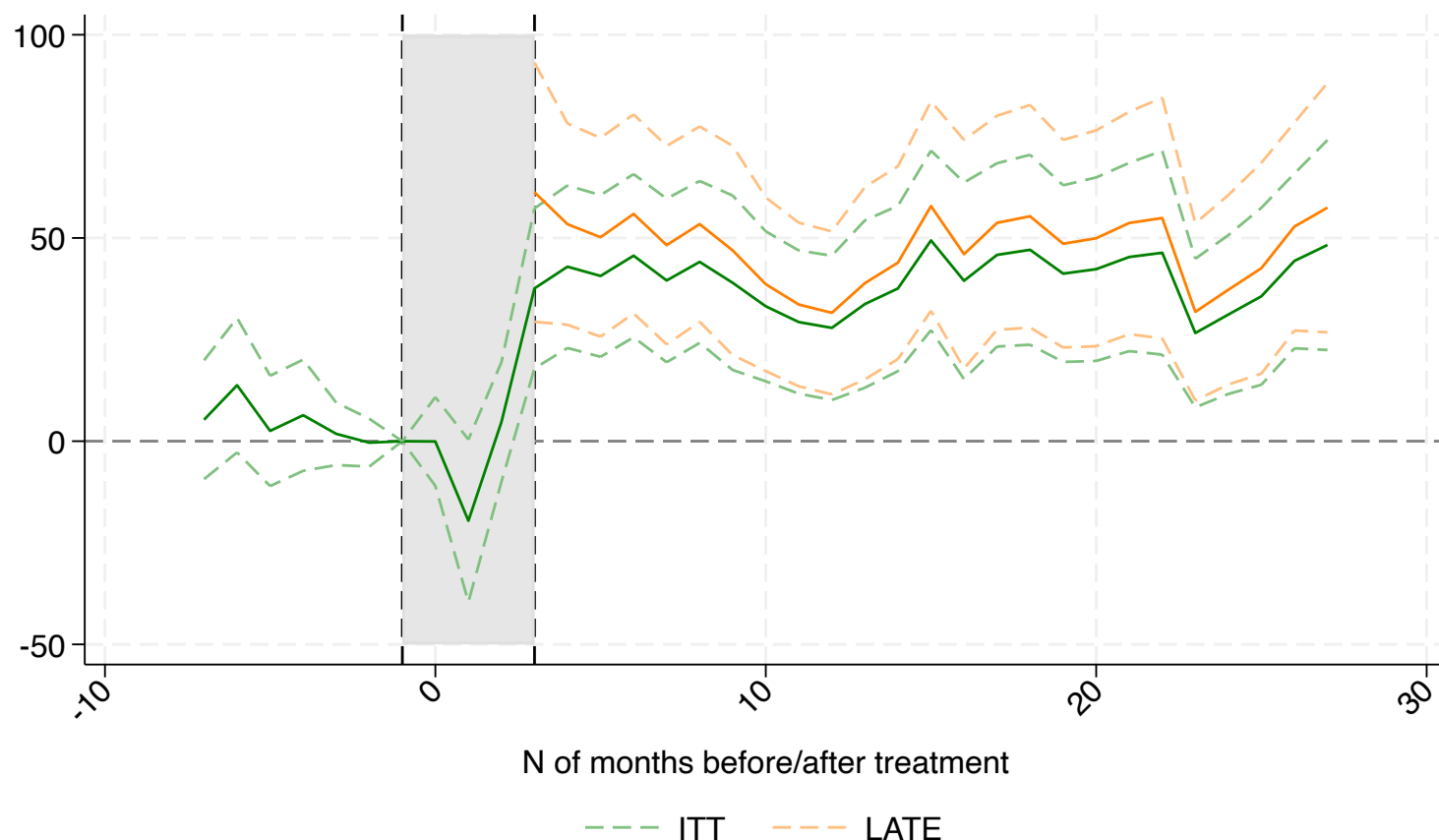
(b) Outcome : The information on the number of employees is missing



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.714 – On measurement and missing information

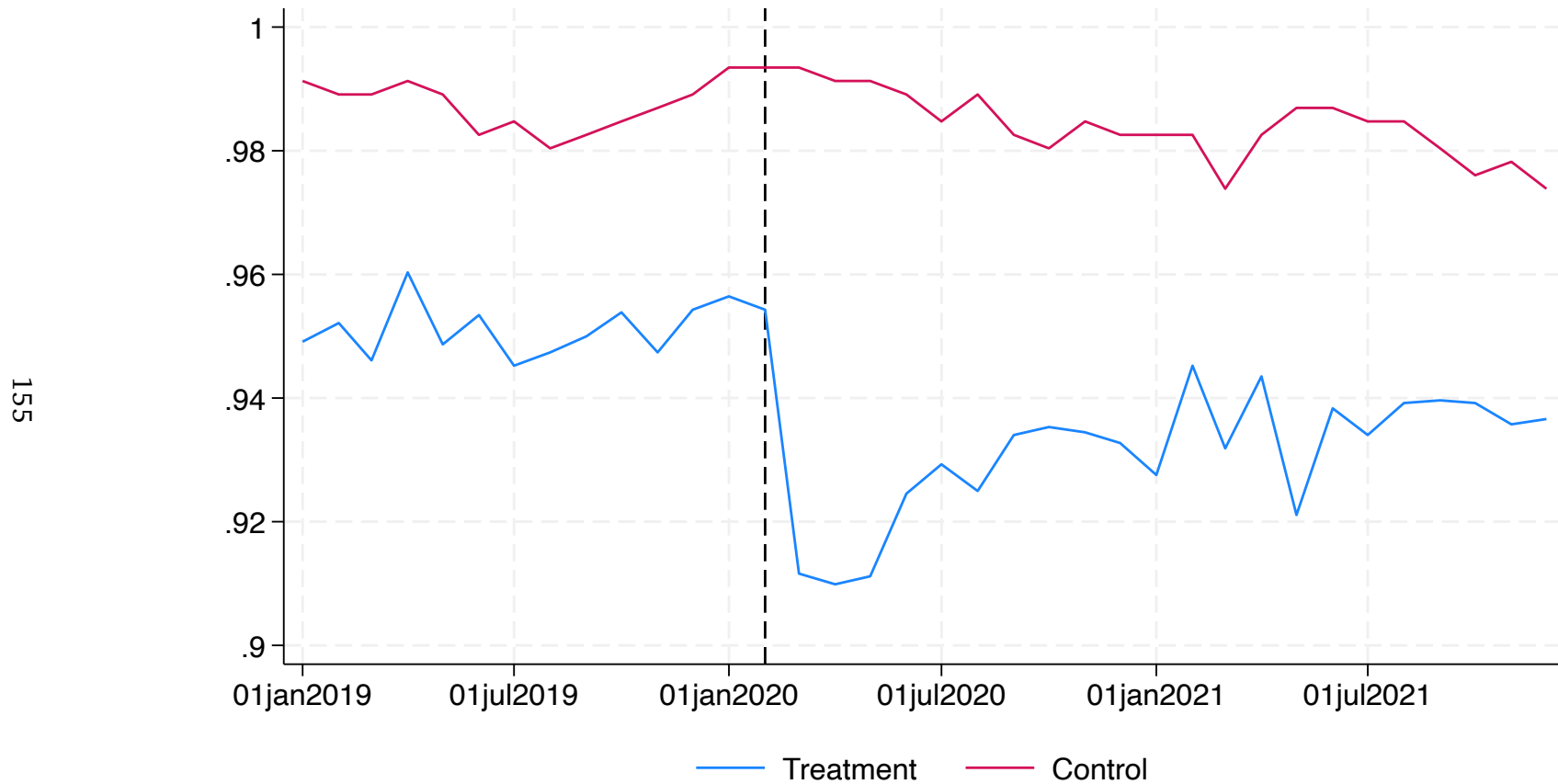
(c) Outcome : Number of employees



Notes. The figure displays coefficients of lags, and leads (ITT or LATE) indicators of model 2.1 using the mandatory adoption of Etax for the Largest Taxpayers Center, set by ministerial decree in July 2017, as a source of exogenous variation. The reference period ($t = -1$) is August 2017. A transition period is highlighted in grey, during which adoption progressively took place. In the LATE specification, we instrument each leads of model 2.1 by an indicator $Etax_{it}^k$ equal to one if and only if $t = k$ and taxpayer i has effectively adopted etax by t . All estimations use time and taxpayer fixed-effects. Data is a balanced sample of 1,691 taxpayers observed for exactly 36 months. Total sample size is 60,876. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.715 – COVID shock on the Probability to submit at least one declaration

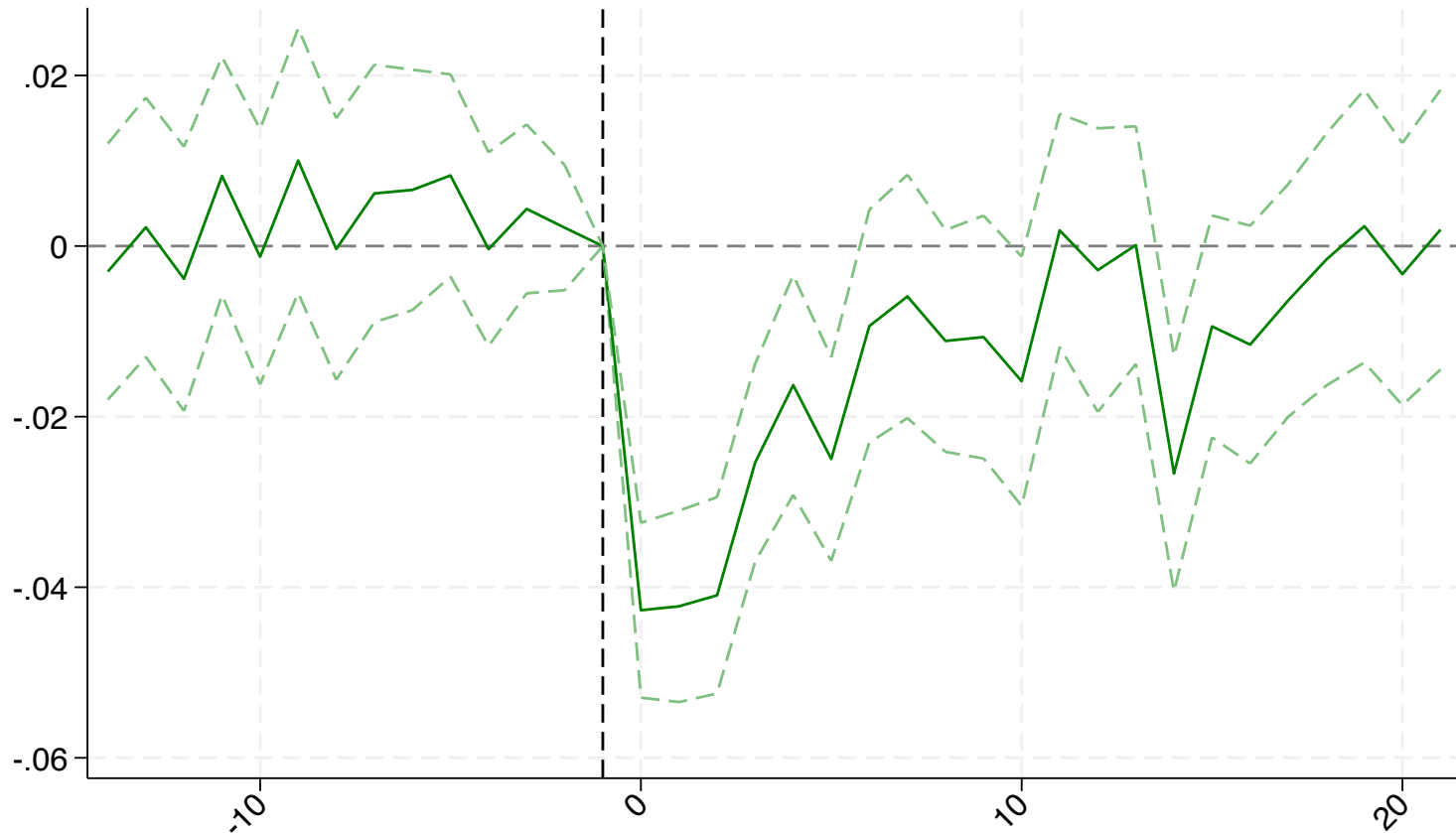
(a) Time trends



Notes. The figure displays average values of the outcome by group and across time. Treated taxpayers are those who had not adopted Etax by March 2020, the reference period. Data is a balanced sample of 2,778 taxpayers, of which 2,319 are in the treatment group, observed for exactly 36 months. Total sample size is 100,008.

Figure 2.715 – COVID shock on the probability to submit at least one declaration

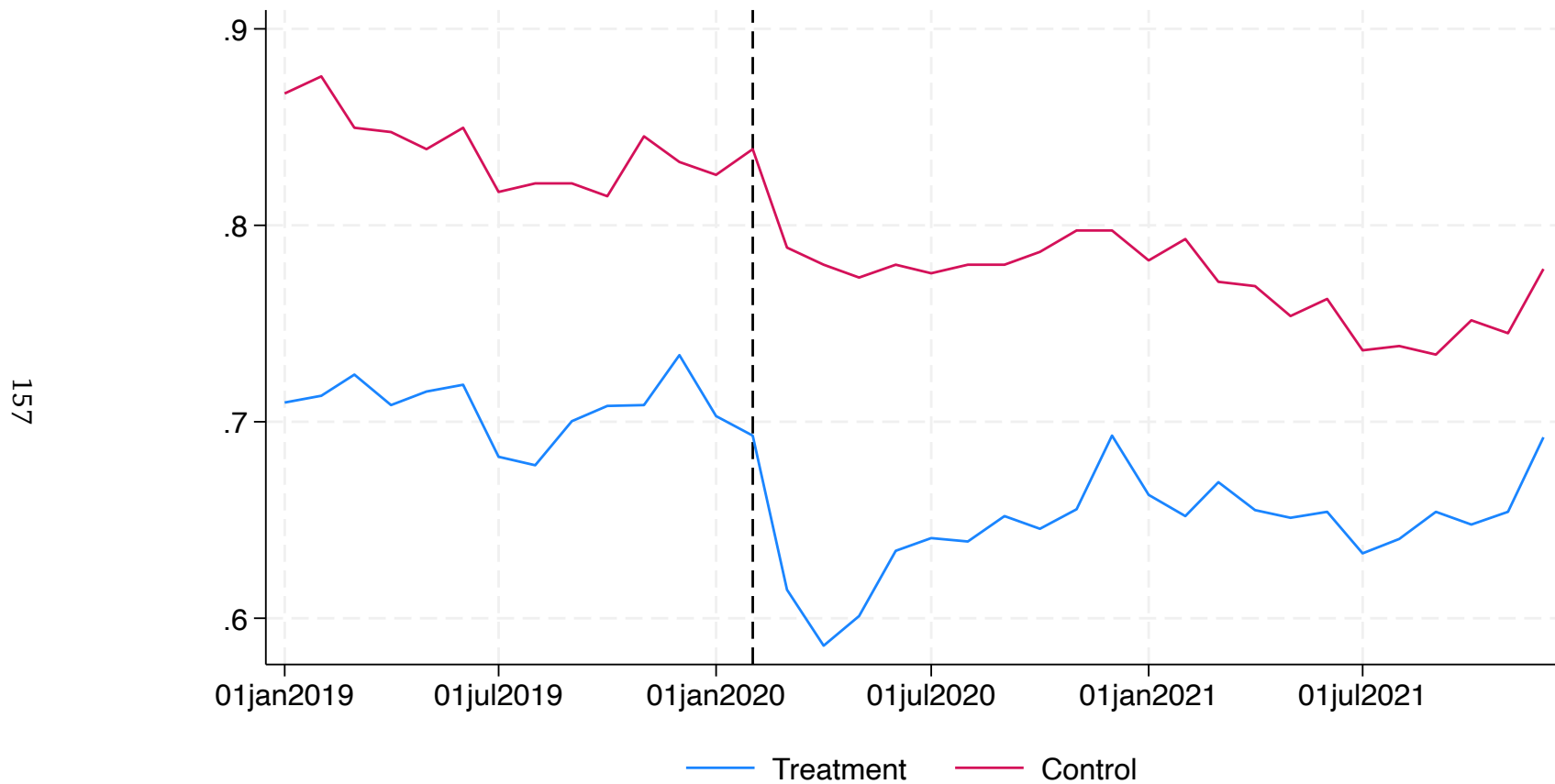
(b) Time-varying impact



Notes. The figure displays coefficients of lags and leads indicators of model 2.1 using the COVID shock as exogenous source of variation and defining treated taxpayers as those who had not adopted Etax by March 2020, the reference period ($t = -1$). Data is a balanced sample of 2,778 taxpayers, of which 2,319 are in the treatment group, observed for exactly 36 months. Total sample size is 100,008. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.716 – COVID shock on the probability to pay a positive amount

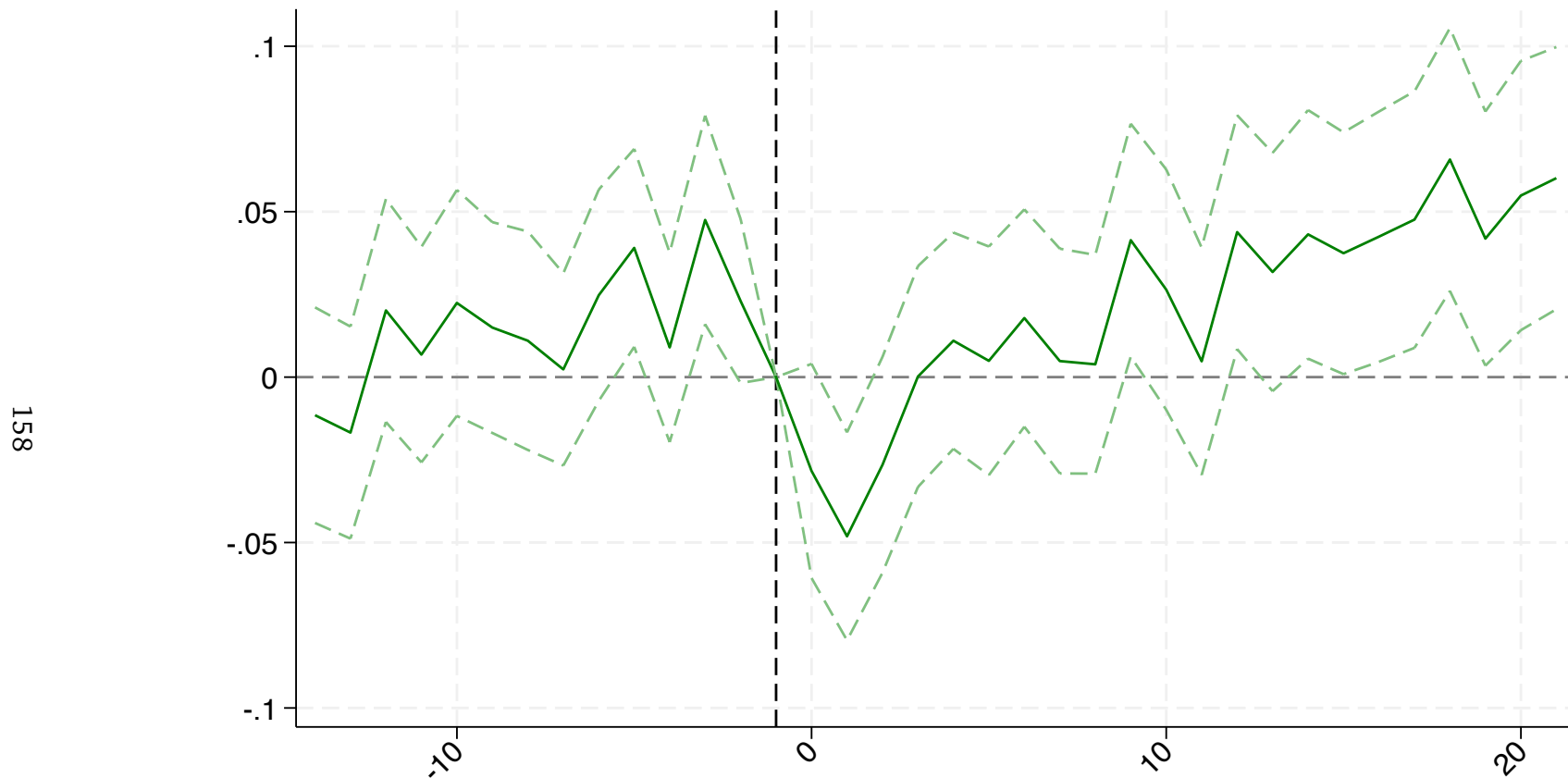
(a) Time trends



Notes. The figure displays average values of the outcome by group and across time. Treated taxpayers are those who had not adopted Etax by March 2020, the reference period. Data is a balanced sample of 2,778 taxpayers, of which 2,319 are in the treatment group, observed for exactly 36 months. Total sample size is 100,008.

Figure 2.716 – COVID shock on the probability to pay a positive amount

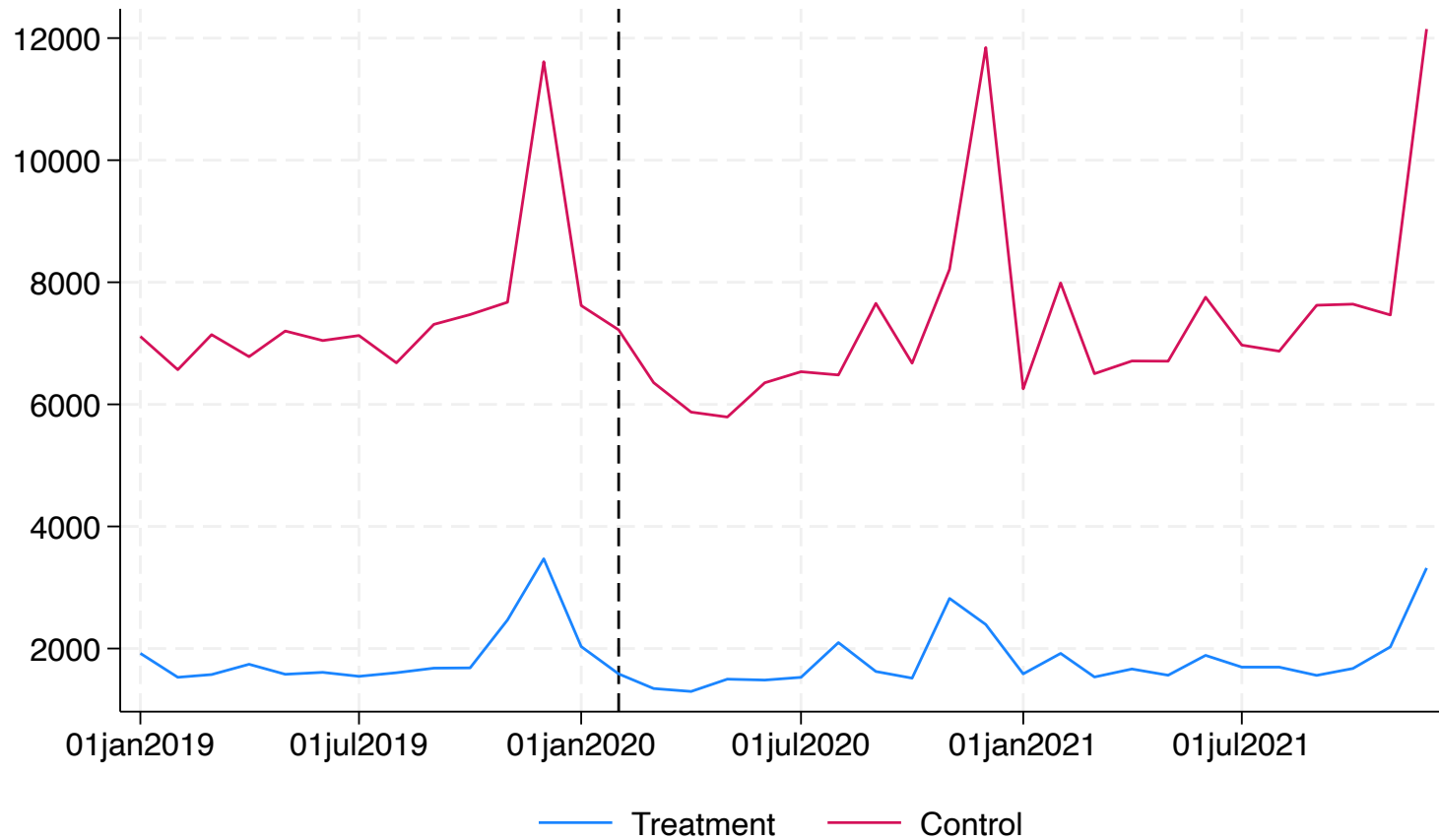
(b) Time-varying impact



Notes. The figure displays coefficients of lags and leads indicators of model 2.1 using the COVID shock as exogenous source of variation and defining treated taxpayers as those who had not adopted Etax by March 2020, the reference period ($t = -1$). Data is a balanced sample of 2,778 taxpayers, of which 2,319 are in the treatment group, observed for exactly 36 months. Total sample size is 100,008. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

Figure 2.717 – COVID shock on the amount of tax paid

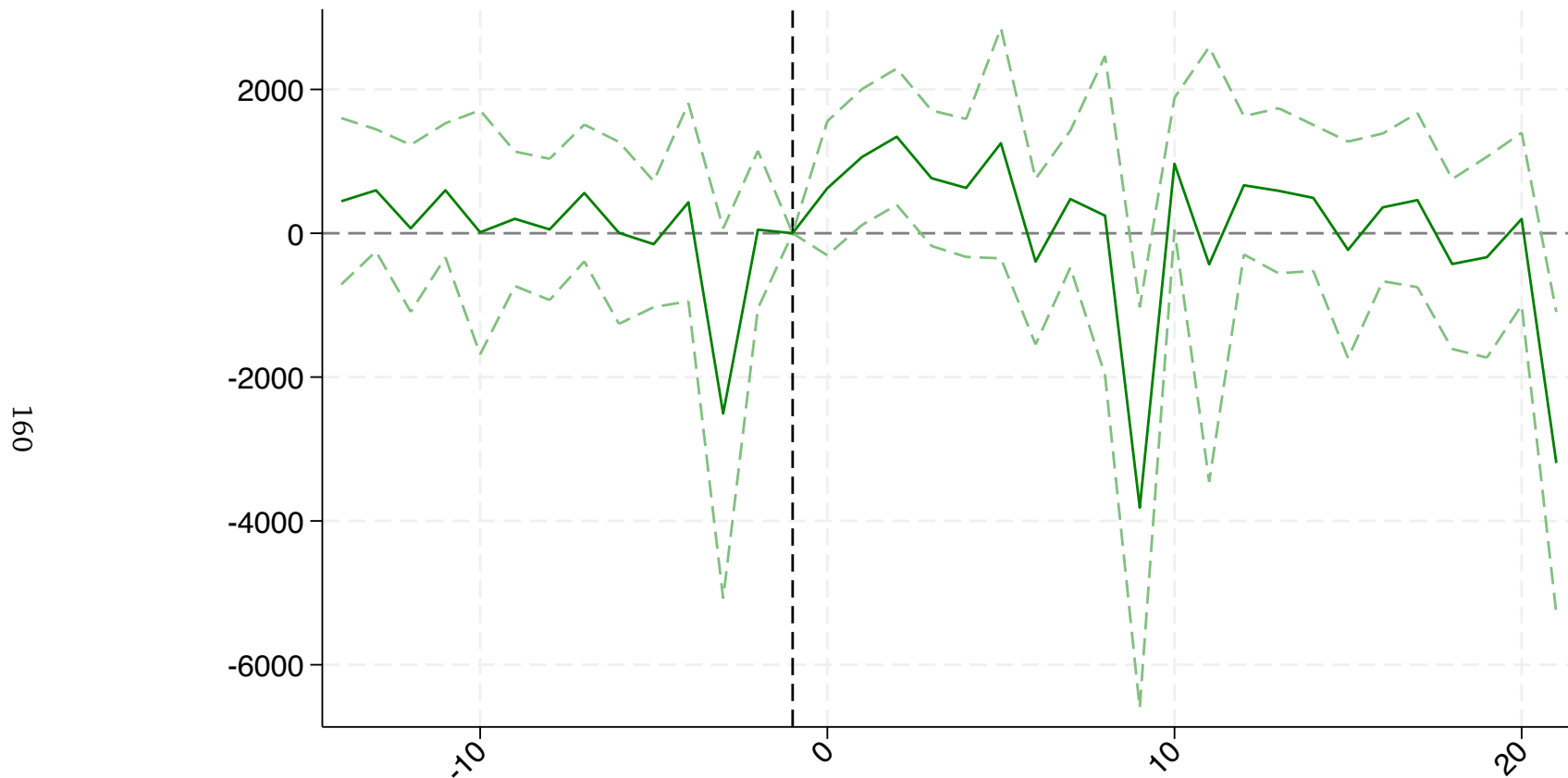
(a) Time trends



Notes. The figure displays average values of the outcome by group and across time. Treated taxpayers are those who had not adopted Etax by March 2020, the reference period. Data is a balanced sample of 2,778 taxpayers, of which 2,319 are in the treatment group, observed for exactly 36 months. Total sample size is 100,008.

Figure 2.717 – COVID shock on the amount of tax paid

(b) Time-varying impact



Notes. The figure displays coefficients of lags and leads indicators of model 2.1 using the COVID shock as exogenous source of variation and defining treated taxpayers as those who had not adopted Etax by March 2020, the reference period ($t = -1$). Data is a balanced sample of 2,778 taxpayers, of which 2,319 are in the treatment group, observed for exactly 36 months. Total sample size is 100,008. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Dashed lines indicate 95% confidence interval.

2.8. Tables

Table 2.81 – Number of Civil Servants by treatment status and qualification - 2016-2022

		2016		2017		2018		2019		2020		2021		2022	
		N.	N.	% change	N.	% change	N.	% change	N.	% change	N.	% change	N.	% change	
Treatment center (LTU)	A	34	34	0	36	5.8	34	-5.5	35	2.9	29	-17.1	30	3.4	
	B	28	31	10.7	30	-3.2	29	-3.3	29	0	37	27.5	38	2.7	
	C	13	13	0	14	7.6	14	0	14	0	13	-7.1	13	0	
	D	4	4	0	4	0	3	-25	3	0	2	-33.3	0	-100	
	Total	79	82	3.7	84	2.4	80	-4.7	81	1.2	81	0	81	0	
Control centers	A	67	66	-1.4	79	19.7	94	18.9	92	-2.1	97	5.4	96	-1.0	
	B	167	165	-1.2	184	11.5	193	4.8	193	0	216	11.9	216	0	
	C	241	253	4.9	252	-0.4	249	-1.1	249	0	266	6.8	257	-3.3	
	D	33	37	12.1	35	-5.4	33	-5.7	33	0	35	6.0	35	0	
	Total	508	521	2.5	550	5.5	569	3.4	567	-0.3	614	8.2	604	-1.6	
Other centers/departments	A	159	148	-6.9	152	2.7	168	10.5	169	0.6	187	10.6	195	4.2	
	B	218	220	0.9	211	-4.0	220	4.2	220	0	218	-0.9	218	0	
	C	166	169	1.8	169	0	170	0.5	170	0	187	10	190	1.6	
	D	71	69	-2.8	65	-5.8	70	7.6	70	0	61	-12.8	61	0	
	Total	614	606	-1.2	597	-1.4	628	5.1	629	0.1	653	3.8	664	1.6	
All	A	260	248	-4.6	267	7.6	296	10.8	296	0	313	5.7	321	2.5	
	B	413	416	0.7	425	2.1	442	4	442	0	471	6.5	472	0.2	
	C	420	435	3.5	435	0	433	-0.4	433	0	466	7.6	460	-1.2	
	D	108	110	1.8	104	-5.4	106	1.9	106	0	98	-7.5	96	-2.0	
	Total	1201	1209	0.6	1231	1.8	1277	3.7	1277	0	1348	5.5	1349	0.0	

Notes. This table displays the number of civil servants - by qualification - affected to each quasi-experimental group and its evolution since 2016. Panel (A) consists of civil servants affected to the Large Taxpayers Center (the treated center in our main specification); panel (B) consists of civil servants affected to centers which are included in our control group (all centers of Dakar, excluding that for upper-middle sized firms); panel (C) gathers all other civil servants - this includes regional centers as well as all staff working at function-specific centralised departments such as the IT, legislation, human resources. Finally panel (D) is the sum of (A),(B) and (C).

Table 2.82 – Main Results

Outcome	β	Std. err.	Specification
Submits at least one declaration	0.011	0.009	ITT
	0.015	0.012	LATE
Some declaration has missing information	-0.052***	0.010	ITT
	-0.069***	0.014	LATE
Some declaration has inconsistent information	-0.020	0.012	ITT
	-0.026	0.016	LATE
Some declaration has implausible information	0.006	0.005	ITT
	0.008	0.007	LATE
Liability declared is positive	0.021*	0.013	ITT
	0.028*	0.017	LATE
Liability declared (USD)	2,812	4,874	ITT
	3,726	6,459	LATE
Tax paid is positive	-0.011	0.012	ITT
	-0.015	0.016	LATE
Amount paid (USD)	16,851**	7,588	ITT
	22,327**	10,045	LATE

Notes. The table reports main coefficients obtained by estimating model 2.2 with Ordinary Least Squares (ITT) or two stages least square method (LATE) when instrumenting $Treat_{it} \cdot Post_{it}$ by a indicator $Etax_{it}$ equal to one if and only if taxpayer i has started using Etax by t (either through e-filing or e-payment). All estimations uses time and taxpayer fixed-effect. Data is a balanced sample of 1,691 taxpayers observed for exactly 29 months, the first of which is the reference period of the model: August 2017. Months of the transition period are included in the post-period. Total sample size is 49,039. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Significance level are indicated by * (10%), ** (5%) and *** (1%).

Table 2.83 – Main Results - excluding the transition period

Outcome	β	Std. err.	Specification
Submits at least one declaration	0.015	0.009	ITT
	0.017	0.011	LATE
Some declaration has missing information	-0.057***	0.011	ITT
	-0.069***	0.013	LATE
Some declaration has inconsistent information	-0.018	0.013	ITT
	-0.021	0.015	LATE
Some declaration has implausible information	0.005	0.006	ITT
	0.006	0.007	LATE
Liability declared is positive	0.028**	0.014	ITT
	0.032**	0.016	LATE
Liability declared (USD)	3,735	5,283	ITT
	4,430	6,267	LATE
Tax paid is positive	-0.012	0.013	ITT
	-0.014	0.016	LATE
Amount paid (USD)	17,760**	7,819	ITT
	21,064**	9,271	LATE

Notes. The table is the same as table 2.82, but for the fact that we drop the four months of the transition period (September, October, November and December 2017). The table reports main coefficients obtained by estimating model 2.2 with Ordinary Least Squares (ITT) or two stages least square method (LATE) when instrumenting $Treat_{it} \cdot Post_{it}$ by a indicator $Etax_{it}$ equal to one if and only if taxpayer i has started using Etax by t (either through e-filing or e-payment). All estimations uses time and taxpayer fixed-effect. Data is a balanced sample of 1,691 taxpayers observed for exactly 29 months, the first of which is the reference period of the model: August 2017. Months of the transition period are included in the post-period. Total sample size is 49,039. Standard errors are adjusted using Huber-White (“robust”) Sandwich estimator. Significance level are indicated by * (10%), ** (5%) and *** (1%).

Chapter 3

Wealth Inequality in South Africa, 1993–2017

with Aroop CHATTERJEE and Amory GETHIN

Abstract This paper estimates the distribution of personal wealth in South Africa by combining microdata covering the universe of income tax returns, household surveys, and macroeconomic balance sheets statistics. We document unparalleled levels of wealth concentration. The top 10% own 86% of aggregate wealth and the top 0.1% close to one third. The top 0.01% of the distribution (3,500 individuals) concentrate 15% of household net worth, more than the bottom 90% as a whole. Such levels of inequality can be accounted for in all forms of assets at the top end, including housing, pension funds and financial assets. We find no sign of decreasing inequality since the end of apartheid.

3.1. Introduction

A growing number of studies have made significant progress in estimating the distributions of income and consumption across countries and over time, yet still little is known on the dynamics of household wealth. This lack of data is particularly acute in the developing world, where available data sources are scarce and tend to be prone to important measurement error. Given the challenges posed by tax evasion, tax-related migration, and the rise of global wealth concentration ([Alvaredo et al., 2018b](#); [Kleven et al., 2020](#); [Zucman, 2019](#)), there is a pressing need to improve our knowledge of the wealth distribution. Such knowledge is key to informing public debate and policymakers on the potential policies that could be implemented to reduce wealth inequality, such as progressive wealth taxes, which have recently received renewed attention in a number of countries (see for instance [Davis Tax Committee, 2018](#) in South Africa and [Saez and Zucman, 2020](#) in the United States).

This paper estimates the wealth distribution in South Africa from 1993 to 2018 by combining household survey data, tax microdata, and macroeconomic balance sheets statistics. A number of results emerge from our analysis.

First, we document unparalleled levels of wealth concentration. The top 10% of South African wealth holders own more than 85% of household wealth, while the top 1% wealth share reaches 55%. The top 0.01% of the wealth distribution, corresponding to about 3,500 adults, own a higher share of wealth than the bottom 90% as a whole (32 million individuals). The average wealth of the bottom 50% is negative: the market value of their assets is lower than their liabilities. Such levels of wealth inequality are higher than in any other country in the world for which high-quality estimates of the wealth distribution are available.

Secondly, we find that these inequalities are reproduced at the level of all asset classes. The top 10% of wealth holders own more than 55% of business assets and housing wealth, and over 99% of bonds and stock. Financial assets constitute the bulk of the assets of the top 0.1%, while owner-occupied housing and pension wealth represent approximately equal proportions of the assets of the bottom 90%. We find significant wealth accumulation over the life cycle, but no significant decrease in net worth at old age, which stabilizes after reaching a peak at age 60.

Thirdly, our estimates do not suggest that wealth inequality has decreased since the end of the apartheid regime. The top 10% wealth share has fluctuated between 80% and 90% between 1993 and 2017, largely as the result of the rise and fall of household debt before and after the 2007-2008 crisis, with no sign of long-run trend. The share of wealth captured by the top 1% and the top 0.01% may even have increased, although limitations in available data imply that evolutions at the very top should be interpreted with care.

This paper contributes to the literature on wealth inequality in South Africa in several ways. Most existing studies have exclusively relied on self-reported wealth in household surveys to study the wealth distribution in South Africa, generally finding much lower levels of wealth concentration than we do (Daniels and Augustine, 2016; Mbewe and Woolard, 2016; Van Tonder, Van Aardt, and De Clercq, 2018).¹ This, as we show in section 4, is largely due to the massive underreporting of financial assets by survey respondents, which leads to strongly underestimating top end inequality, as is now well documented in the inequality measurement literature (Alvaredo et al., 2016; Blanchet, Flores, and Morgan, 2018b; Blanchet, Fournier, and Piketty, 2017b; Korinek, Mistiaen, and Ravallion, 2006). By contrast, our methodology combines survey and tax microdata with macrodata on household wealth totals, in such a way that average wealth and the portfolio composition of assets across the distribution are fully consistent with the household balance sheets statistics published by the South African Reserve Bank. Our approach of combining data on labor and capital incomes with macroeconomic totals also allows us to cover the entire 1993-2017 period, in contrast to approaches relying on wealth surveys (which only cover a few years and did not exist before 2008).

More generally, this paper contributes to the growing literature on the measurement of wealth inequality. Recent studies on the evolution of wealth concentration in developed countries with high-quality data, such as the United States (Saez and Zucman, 2016b) or France (Garbinti, Goupille-Lebret, and Piketty, 2018a), have generally combined exhaustive tax microdata with macroeconomic balance sheets to study the concentration of wealth over long periods of time. Meanwhile, studies focusing on countries with limited or no tax data have mostly relied on self-reported wealth in surveys (Bharti, 2018; Novokmet, Piketty, and Zucman, 2018; Piketty, Yang, and Zucman, 2019b). In this paper, we systematically compare estimates of the wealth distribution obtained from three different methodologies: direct measurement of net worth, rescaling of reported wealth components to balance sheets totals, and capitalization of income flows. In particular, we show that capitalizing income flows in household surveys (that is, rescaling reported income flows to match macroeconomic wealth totals) can yield relatively good estimates of overall wealth inequality, even in the absence of tax microdata. Crucially, these estimates appear to be much more reliable than those relying solely on survey-based reported wealth, which omit the bulk of financial wealth. In other words, bridging the micro-macro gap in wealth measurement appears to be an absolutely essential step to accurately measure the wealth distribution. This opens new avenues for estimating the dynamics of wealth inequality in

¹See Chatterjee, 2019 for a full review of existing studies. A notable exception is Orthofer, 2016, who uses the sum of investment and pension incomes from personal income tax microdata to approximate the wealth distribution. However, this study does not account for owner-occupied housing (28% of household wealth, see below) and does not rescale income flows to match macroeconomic wealth totals. The resulting estimates therefore better correspond to the distribution of financial incomes than to the distribution of household wealth.

low- and middle-income countries, where wealth microdata are not available or unreliable, yet where macroeconomic balance sheet statistics can be usefully combined with surveys collecting data on household income. In that respect, we hope that this paper can serve as a useful guide for future studies aiming to measure wealth inequality in countries with limited data such as South Africa.

The rest of the paper is organized as follows. Section 2 presents the available sources on aggregate wealth and the distribution of income and wealth in South Africa. Section 3 assesses the different methodologies to measure wealth inequality and contrasts the results obtained with different data sources. Section 4 presents our main results and compares our estimates with that of other countries.

3.2. Data sources

This section discusses the data sources available to measure total household wealth and wealth inequality in South Africa. We start by defining aggregate household wealth and presenting existing wealth macrodata. We then turn to a comparison of the microdata sources available to measure the distribution of income and wealth.

3.21. Data sources on aggregate wealth

The household balance sheets

Although the estimation of wealth inequality has a long history ([Atkinson and Harrison, 1974](#); [Clay, 1925](#); [Daniels and Campion, 1936](#); [Langley, 1950](#); [Mallet, 1908](#)), the development of national accounts statistics covering wealth concepts is a relatively recent phenomenon, with international guidelines on the measurement of total wealth only being adopted in 1993 as part of the United Nations System of National Accounts (UN SNA) ([Piketty and Zucman, 2014](#); [United Nations, 2009](#)). In South Africa, the first comprehensive attempt to estimate the value of total household wealth in the economy goes back to [Muellbauer and Aron, 1999](#), who collect and combine a number of data sources to provide figures on the assets and the liabilities of the household sector since 1975. The South African Reserve Bank (SARB) has since then updated and revised these figures on a yearly basis.

Household wealth, as defined by the UN SNA, corresponds to the total market value of the assets and liabilities held by the household sector. The SARB currently publishes decompositions of household wealth into its financial and non-financial components, along with broad

decompositions by asset class and information on household debt, which we use in what follows to match macrodata sources with their corresponding asset or income flow at the micro level. Non-financial assets are divided into two components: residential buildings (the market value of residential properties owned by household, excluding land) and other non-financial assets (including land and business assets). Financial assets are divided into three components: interest in pension funds and long-term insurers, assets with monetary institutions, and other financial assets. Interest in pension funds and long-term insurers corresponds to all pension assets and life insurance holdings of the household sector.² Assets with monetary institutions include all forms of currency and deposits with banks, as well as notes and coins held by households. Other financial assets include investment in government and public entities stock, collective investment schemes, corporate bonds and equities, other long-term deposits and households' investment in foreign assets. Finally, the SARB decomposes household debt into two components: mortgage advances, corresponding to loans provided by the commercial banking sector, and other debt (including trade credit, personal bank loans, credit card debt, instalment sales and lease agreements, and other formal and informal loans).

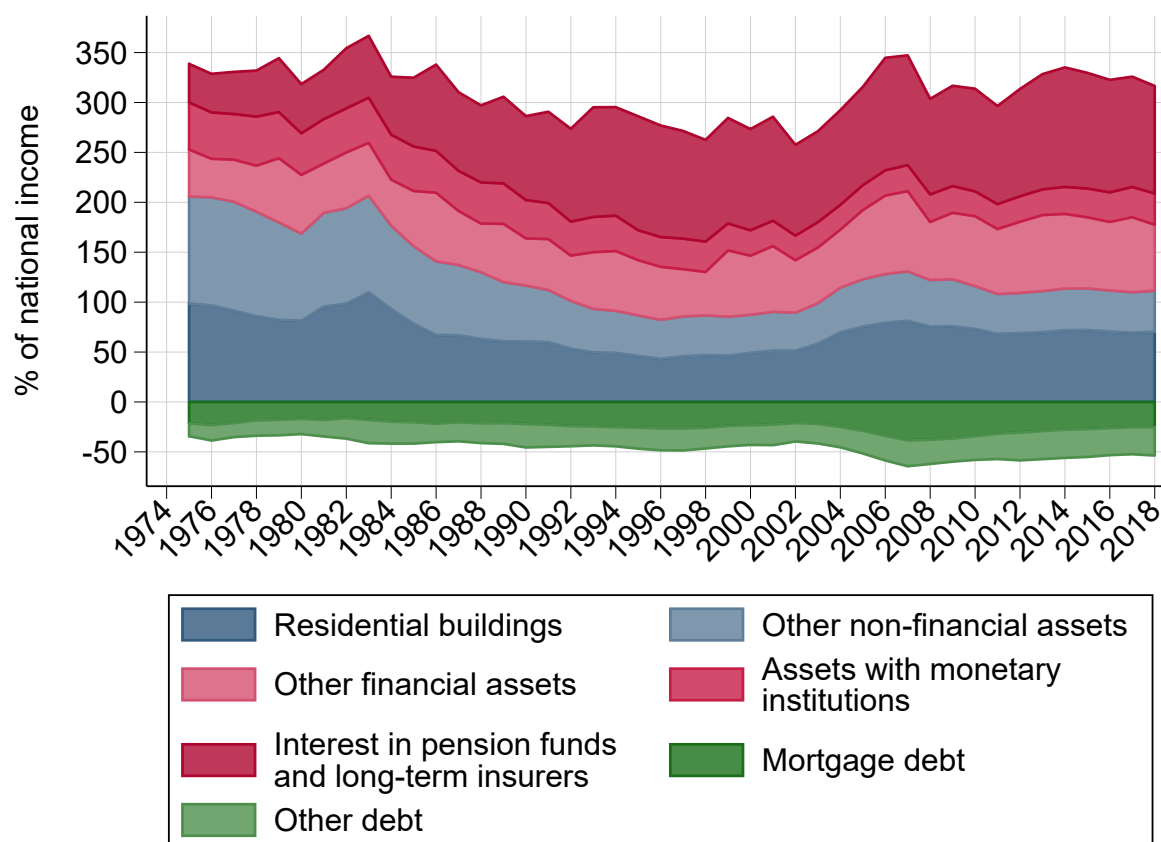
Figure 3.21 plots the evolution and composition of total household wealth between 1975 and 2018, expressed as a share of national income. Aggregate net wealth has followed a U-shaped curve, declining from about 300 per cent of national income in 1975 to 220 per cent at the end of the 1990s, and rising back to more than 260 per cent at the beginning of the twenty-first century. In 2018, financial assets amounted to two years of national income. Within financial assets, pension assets have risen significantly and are now the biggest component of household wealth. Non-financial assets amounted to one year of national income in 2018, declining in importance over the years from just under two years of national income in 1974. Household debt rose significantly between 1975 and 2008, in large part due to a boom in mortgage advances in the early 2000s (see supplementary appendix Figure A4). It has slightly declined since then, amounting to about 55 per cent of the national income today.

The limitations of available balance sheets: from institutions to asset classes

The objective of this paper is to estimate the distribution of household wealth by matching macrodata on wealth with microdata on reported assets and capital income flows. In order to

²This corresponds to the sum of the total assets of official pension and provident funds (series KBP2215 in Capital Markets Statistics), the total liabilities of private self-administered pension and provident funds (KBP2339), and the liabilities of long-term insurers under unmatured policies from the pension business (KBP2215). Notice that the original estimates of the South African household balance sheets done by [Muellbauer and Aron, 1999](#) excluded life insurance assets and all other assets associated with the non-pension business of long-term insurers. However, these items are now included by the SARB in line with the SNA guidelines.

Figure 3.21 – The evolution of household wealth in South Africa, 1975-2018



Notes: This figure shows the level and composition of household wealth in South Africa between 1975 and 2018, expressed as a share of the net national income.

Source: authors' compilation based on data from the South African Reserve Bank.

improve our estimates of the wealth distribution and obtain a better mapping of macrodata and microdata components, we address five shortcomings of available household balance sheets: the decomposition of non-financial assets, the decomposition of housing wealth into tenant-occupied and owner-occupied, the decomposition of financial assets, the decomposition of pension and life insurance assets, and the inclusion of wealth held offshore in tax havens.

Land underlying dwellings The "Other non-financial assets" category provided by the SARB includes both land underlying dwellings and business assets. These two components are arguably distributed very differently. In particular, it is reasonable to assume that land underlying dwellings is distributed similarly to residential buildings (therefore defining total housing assets as the sum of land and residential buildings), while business assets are better approximated

by business income. We assume that 70% of other non-financial assets corresponds to land underlying dwellings, the remaining 30% amounting to the assets held by unincorporated businesses. This implies that total housing wealth (including land) was equal to 38% of net wealth in 2018, while business assets (machinery and equipment, excluding land) amounted to about 5% of net wealth (see supplementary appendix).

Tenant- versus owner-occupied housing Housing wealth can be decomposed into tenant-occupied housing and owner-occupied housing. Available studies combining surveys with tax microdata typically assume that the distribution of tenant-occupied housing can be well approximated by the distribution of rental income, while owner-occupied housing assets are better captured using direct measurement available from surveys or administrative data [Garbinti, Goupille-Lebret, and Piketty, 2017](#); [Saez and Zucman, 2016b](#). Unfortunately, the "Residential buildings" category published by the SARB does not provide this decomposition, so we choose to derive the proportions from survey data (General Household Survey). We estimate that between 22 per cent and 26 per cent of household housing assets are owned by households renting to private individuals over the 2013-2018 period (see supplementary appendix).

Non-pension financial wealth "Assets with monetary institutions" and "other financial assets" gather together very different forms of financial assets. "Assets with monetary institutions" include both non-interest bearing deposits such as cheque accounts, which do not generate any income flow, and interest bearing deposits which generate interest income. "Other financial assets" include both bonds and corporate shares, which generate interest and dividends respectively. We follow [Orthofer, 2015](#) and assume that the composition of other financial assets held by households is similar to that reported by unit trusts. This implies that between 80 per cent and 95 per cent of other financial assets consist in corporate shares over the 1975-2018 period, the remaining being classified as interest-bearing deposits.³ Finally, we separate currency, notes and coins (0.8% of net wealth) from interest-bearing deposits (17% of net wealth) using SARB data.⁴

³More precisely, we estimate the share of corporate shares in other financial assets by comparing the market value of ordinary shares held by unit trusts (KBP 2412) to the sum of the market values of security holdings of public sector entities, stocks and debentures held by unit trusts (KBP 2410 + KBP 2411) in the capital market statistics published by the SARB.

⁴The variable "Monetary sector liabilities: banknotes and coins in circulation" (series KBP1312) corresponds to currency, notes and coins held by all institutions. We assume that 70% of the total can be attributed to households. Given the small share of this component in total wealth, especially at the top of the wealth distribution, our results are not affected by alternative scenarios.

Pension assets and life insurance Pension assets correspond to the assets accumulated by wage earners through contributions to pension funds throughout their career, so they should in large part be distributed to wage earners and pensioners receiving pension income or annuities. Life insurance assets, by contrast, better correspond to a form of savings device, but they do not directly generate interest income, so they cannot be categorised with interest deposits or bonds and have to be distributed differently. Accordingly, we use available SARB data to decompose the “Interest in pension funds and long-term insurers” item into these two components.⁵ In 2018, pension and life insurance assets amounted to about 28% and 13% of net wealth respectively (see supplementary appendix).

Offshore wealth Offshore wealth corresponds to the assets held abroad by South African residents, mainly for tax avoidance purposes. By definition, these assets are not recorded in official records and are therefore not included in the household balance sheets. [Alstadsæter, Johannesen, and Zucman, 2018](#) combine a number of macroeconomic data sources to measure the total amount of financial assets held in offshore tax havens and distribute it to specific countries. They estimate that the equivalent of about 11.8 per cent of South African GDP was held offshore in 2007, corresponding to about 5% of net wealth. We add this quantity to total household wealth in 2007 and extrapolate it to other years by assuming that it has remained a constant fraction of GDP.⁶

3.22. Data sources on the distribution of household wealth

Household surveys

Household surveys provide information either on the market value of household assets, and/or on labor and capital incomes, which can be used to estimate the distribution of the various components of household wealth outlined above. The National Income Dynamics Survey (waves 4 and 5, conducted in 2015 and 2017 respectively) is the only publicly available survey that explicitly collects data on wealth. For the purposes of this paper, we only consider the variables that allow us to build a net wealth concept consistent with the System of National Accounts guidelines [United Nations, 2009](#) for comparability and consistency. The surveys have

⁵The share of interest in pension funds and long-term insurers corresponding to assets held by long-term insurers is recorded in the Capital Markets Statistics published by the SARB under series KBP2215, “liabilities of long-term insurers under unmatured policies from the pension business”.

⁶Given that offshore wealth is known to have grown globally, this is a relatively conservative assumption: if anything, wealth inequality could have increased more since 1993 than what our estimates suggest (see section 5), as offshore wealth is well-known for been concentrated at the very top end of the distribution [Alstadsæter, Johannesen, and Zucman, 2018](#).

Table 3.21 – Ownership rates and coverage of household balance sheets by asset class in NIDS

	% of adults with asset or debt		% of balance sheets covered	
	Wave 4	Wave 5	Wave 4	Wave 5
Household assets				
Owner-occupied housing	72.3 %	65.2 %	151.7 %	220.8 %
Tenant-occupied housing	3.3 %	3.5 %	122.4 %	97.2 %
Business assets	5.6 %	5.0 %	135.4 %	59.6 %
Pension and life insurance	25.7 %	24.4 %	110.0 %	104.3 %
Bonds and stock	1.5 %	1.3 %	3.9 %	3.8 %
Household debts				
Mortgage debt	8.0 %	7.0 %	71.0 %	56.8 %
Other debts	36.3 %	33.7 %	54.5 %	37.0 %

Notes: The table shows the share of South Africans who declare having a particular type of asset or debt, along with the share of the total value of this asset or debt in the economy captured by the NIDS survey.

Source: authors' computations based on data. The unit of observation is the adult individual aged 20 or above. Calculations based on weighted sample using design weights.

information on business wealth, housing properties and land, life insurance, pension and other retirement assets, equity wealth, debt and bonds. However, we faced several challenges to extract reliable wealth estimates from this source.

Firstly, we uncovered issues in coverage and reliability in each of the five components of wealth. Looking at housing first, approximately 40 per cent of debtors do not know the house bond value. With pension and other retirement funds, the information is quite inconsistent. For example, in wave 5 of the survey, 61 per cent of individuals declaring contributions to pensions funds declare having no "pension or retirement annuity", while 77 per cent of individuals declaring income from a pension or provident fund declare no "pension or retirement annuity". We correct for these gaps by imputing all missing values using predictive mean matching. For housing wealth, the market value of the house is modelled by number of rooms, number of adults, province and household income. Similarly, we impute missing business wealth for self-employed individuals, as well as pension funds and corporate shares based on factor income, contributions to pension funds and pension income.

The comparison of household assets and liabilities reported in NIDS to macroeconomic statistics show important inconsistencies (see table 3.21). The market value of owner-occupied housing wealth is between 50 per cent and 120 per cent higher in NIDS than in the balance sheets, while tenant-occupied housing is closer to the macro aggregate. This most likely reflects the

Table 3.22 – The coverage of selected national accounts components in South African surveys

	Gross wages	Mixed income	Rental income	Interest and dividends
PSLSD, 1993	87.7 %	51.7 %	38.4 %	11.5 %
IES, 1995	76.9 %	55.0 %	9.9 %	8.8 %
IES, 2000	70.9 %	37.2 %	23.1 %	3.4 %
IES, 2005	80.5 %	64.2 %	21.7 %	3.8 %
IES, 2010	80.2 %	71.9 %	13.5 %	4.5 %
LCS, 2008	77.7 %	75.8 %	16.3 %	8.4 %
LCS, 2015	74.6 %	86.8 %	21.6 %	12.6 %
NIDS, wave 1	62.7 %	12.0 %	65.4 %	7.3 %
NIDS, wave 2	67.6 %	4.1 %	13.0 %	0.8 %
NIDS, wave 3	65.7 %	20.6 %	20.7 %	7.3 %
NIDS, wave 4	73.5 %	12.9 %	43.9 %	2.5 %
NIDS, wave 5	72.1 %	14.1 %	41.0 %	5.5 %

Notes: The table shows the ratio of total income reported in surveys to the total corresponding income reported in the national accounts published by the SARB. PSLSD: Project for Statistics on Living Standards and Development. IES: Income and Expenditure Survey. LCS: Living Conditions Survey. NIDS: National Income Dynamics Study.

Source: authors' computations based on data. The unit of observation is the adult individual aged 20 or above. Calculations based on weighted samples using weights calibrated by the authors' (see appendix).

different methods in measuring market values.⁷ Business assets are covered very differently in the two waves: they are overestimated in wave 4 and underestimated in wave 5. Pension and life insurance assets, after the correction, seem to be relatively close to balance sheets figures, and they even slightly overestimate them. Other financial assets are extremely badly covered: the total reported in NIDS does not exceed 4 per cent of households' bonds and stock reported in the balance sheets by the SARB. Household debts are slightly better covered, but still fall significantly below macroeconomic statistics.

An alternate method to estimate wealth distribution consists of capitalising incomes (usually the method used for estimating the top end of the distribution from tax data). As more surveys deal with incomes, and generally income reporting is seen as more credible, this provides alternate sources of information for the wealth distribution. The method is discussed in more detail in

⁷It is beyond the scope of this paper to discuss and evaluate these methods. However, this issue is not one specific to South Africa - in the US, survey values have also been found to be higher than in balance sheets figures, and which source of information provides the more accurate estimate of market values is contested [Blanchet, 2016](#); [Dettling et al., 2015](#); [Henriques and Hsu, 2014](#). As a robustness check, we show in appendix Figures A11 and A12 that our estimates of the wealth distribution are only marginally affected if one assumes that the balance sheets underestimate housing assets by a factor of two. Another potential issue is how to treat RDP housing, a government-funded social housing project in South Africa, due to complexities around ownership. However, given the typical low market value of these properties, it is unlikely to affect our distributional estimates.

Table 3.23 – The coverage of owner-occupied housing, mortgage debt and other debt in South African surveys

	Owner-occupied housing	Mortgage debt	Other debt
PSLSD, 1993	143.5 %	86.5 %	37.4 %
IES, 1995	121.7 %	27.2 %	16.5 %
IES, 2000		40.3 %	34.9 %
IES, 2005	105.9 %	67.9 %	41.5 %
IES, 2010	193.9 %	16.4 %	20.5 %
LCS, 2008	145.4 %	13.9 %	18.4 %
LCS, 2015	179.5 %	51.0 %	22.2 %
NIDS, wave 4	122.3 %	74.3 %	57.4 %
NIDS, wave 5	258.8 %	56.8 %	37.0 %

Notes: The table shows the ratio of total assets or debts reported in surveys to the corresponding totals reported in the household balance sheets. PSLSD: Project for Statistics on Living Standards and Development. IES: Income and Expenditure Survey. LCS: Living Conditions Survey. NIDS: National Income Dynamics Study.

Source: authors' computations based on data. The unit of observation is the adult individual aged 20 or above. Calculations based on weighted samples using weights calibrated by the authors' (see appendix).

section 3.31. In this section, we compare incomes from surveys to the corresponding totals recorded in the national accounts. For our purposes, the components we consider are gross wages (to capitalise pension wealth), mixed income (income from unincorporated enterprises, to capitalise non-financial assets), rental income (to capitalise tenant-occupied housing) and interest and dividends (for equity and bonds). The surveys we consider were designed to capture information about consumption, expenditure and earnings: these are the Project for Statistics on Living Standards and Development (PSLSD) conducted in 1993, the Income and Expenditure Surveys (IES) from 1995 to 2010, the Living Conditions Surveys (LCS) of 2008 and 2015, and the NIDS.⁸ The labour force surveys which provide wage data only cover labour incomes, so would not be appropriate for this exercise.

As table 3.22 shows, rental income, interest and dividends are very poorly covered in household surveys. This is due to this sort of income being concentrated by those at the upper end of the income distribution, who are typically underrepresented in surveys due to issues of sampling and non-response. This motivates our use of the tax microdata to better cover top incomes. Gross wages and mixed income are much better covered in the PSLSD, IES, and LCS than in NIDS. Owner-occupied housing seems to be over-stated relative to the balance sheets in these surveys as in NIDS, echoing the earlier discussion (see table 3.23). Debts are always below

⁸There are concerns about whether these surveys are comparable: see for instance Berg and Louw, 2005; Leibbrandt, Woolard, and Woolard, 2009; Pauw and Mncube, 2007; Yu, 2005.

balance sheets totals, but with important fluctuations across surveys. All these limitations justify the need to correct for discrepancies between micro and macro totals. Indeed, the households balance sheets have the advantage of tracking the evolution of wealth consistently, in contrast with surveys which show much greater fluctuations in reported aggregates. By mapping the surveys with macroeconomic statistics, we are at least able to get estimates of the wealth distribution which are consistent with what we know of the level of aggregate wealth and its composition over time, which is what we do in the next section.

Tax data

The tax data, for purposes of this paper, refers to two data sources - the IRP5 data, and the ITR12 data. The IRP5 forms are income tax forms submitted to the South African Revenue Service (SARS) by employers on behalf of their employees, hence covers incomes related to the employment relationship. Specific variables of interest include gross wages as well as contributions to retirement assets (pension, annuities, etc). The ITR12 forms are self-assessment forms, that require taxpayers to disclose income from sources other than employment, so that taxable income can be calculated. Thus, data from this form provides information on business, rental, interest and dividend incomes, which can be capitalised to calculate the asset bases from which the incomes derive. These data sources have been combined into a panel that provides detailed information about all incomes, allowances and deductions (for an overview and discussion of the dataset, see [Ebrahim and Axelson, 2019a](#)).

Due to its administrative nature, this data covers the full tax paying population, including individual observations at the top of the distribution. As it is not a sample, it identifies all individual taxpayers, which greatly increases the granularity of measured income flows. This is an advantage over surveys which often suffer from low sample biases. That being said, there are a number of limitations with tax microdata which should also be emphasised. The fact that the ITR12 forms are self-assessed implies that there may be tax evasion or under-reporting of income flows, especially if the likelihood of being controlled by tax authorities is low. More importantly, tax microdata only covers forms of incomes which are useful for tax collection and deductions purposes, which implies that other forms of non-taxable incomes are not reported in the data. This, as we show below, is particular problematic for the measurement of capital incomes.

In order to combine the tax data with survey data at the bottom of the distribution and capitalise income flows, we categorise the source codes reported in the tax data into seven broad categories: gross wages, business income, pension contributions, pension income, interest income, rental

Table 3.24 – The coverage of capital income
in the tax microdata

	Rental income	Interest income	Dividends
2010	9.5 %	25.4 %	2.4 %
2011	11.7 %	25.0 %	5.3 %
2012	12.3 %	28.3 %	3.9 %
2013	13.4 %	28.8 %	5.2 %
2014	12.1 %	27.8 %	25.1 %
2015	12.3 %	27.8 %	10.6 %
2016	13.7 %	31.0 %	13.1 %
2017	6.9 %	18.3 %	15.8 %

Notes: The table shows the ratio of total income reported in the tax microdata to the corresponding total reported in the national accounts published by the SARB.

Source: authors' computations based on data.

income and dividends (see supplementary appendix table A2).⁹ Table 3.24 shows that when looking specifically at capital incomes in the tax data, the reported totals fall significantly below the national accounts. Interest income is better measured than rental income and dividends, reaching between 25 per cent and 30 per cent of total interest received by households in the national accounts. Rental income and dividends are significantly lower and inconsistent, covering between 2 per cent to 25 per cent of national accounts totals.¹⁰

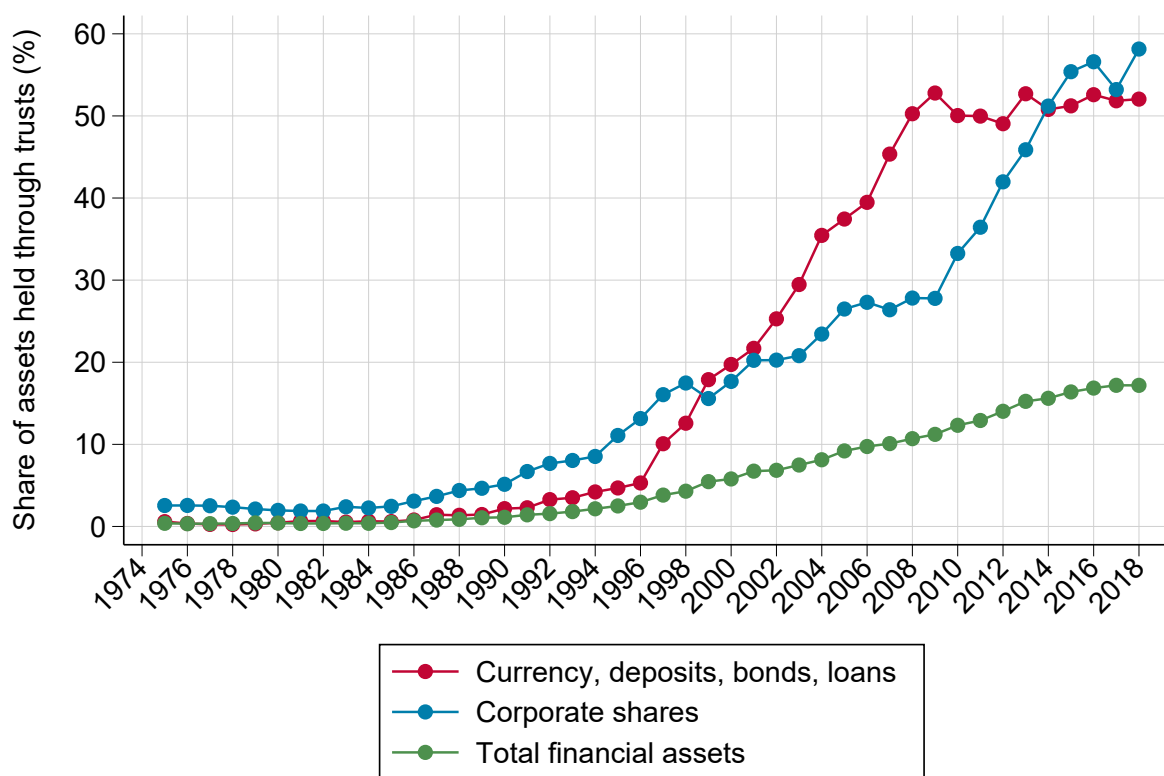
This under-representation of capital incomes in the tax data is due to three main factors. First, the taxable incomes are different from incomes reported in the national accounts, due to filing rules and tax base. This is particularly problematic for dividends, which in the ITR12 relate to dividends from equities that form part of compensation packages, such as equity share plans. These sort of dividends are subject to income tax, and so part of this data set, whereas dividends from regular ownership of equity is subject to a separate dividend tax. Approximately 80 per cent of dividend information would be recorded through this dividend tax returns (DTR01/2 forms), and this information is urgently required to make our estimate more reliable.

Secondly, there may be issues of misreporting of incomes by individual taxpayers. Interest income seems to be poorly covered as a result of incomplete tax filing by taxpayers. In principle, the South African Reserve Bank receives direct information from banks and financial services that they provide about interest. Banks and financial service providers separately supply customers

⁹The IRP5 and ITR12 data are presented in the form of source codes corresponding to specific taxable income concepts, exemptions and deductions. See [Ebrahim and Axelson, 2019a](#) for a more complete discussion.

¹⁰The particularly low figures obtained in 2017 (fiscal year 2018) are mainly due to the fact that assessment was incomplete at the time of writing.

Figure 3.22 – Share of financial assets held through trusts, 1975-2018



Notes: The figure shows the share of total household assets in the economy held by unit trusts.
Source: authors' compilation based on data from the SARB.

with a tax certificate (IT3(b) certificate), which is meant to inform the interest income declared by the taxpayer. At the same time, the bank sends the South African Revenue Service a third-party submission about incomes its customers' receive. However, given that interest income is typically low relative to total taxable income, it is possible that small interest income received go unreported. The misreporting of rental income received by individual taxpayers is likely to be more significant, given that rental income is self-reported and that there may be a significant amount of informal letting of fixed property. ¹¹

Finally, the most important issue regarding the coverage of capital incomes in the tax microdata is likely to be due to the definition of the taxpayer. The tax data covers only individuals and does not account forms of capital incomes received through units trusts or investment funds. This is

¹¹Notice here that total rental income paid to individuals in the economy is estimated by the authors based on data from the PSLSD, the IES and the GHS surveys on total rental income paid by households to individual landlords. Therefore, this includes informal rents paid, which may explain why the rental income the tax data is so low compared to the macro aggregate.

particularly problematic in the case of South Africa, both because wealth is highly concentrated at the top of the distribution and because top wealth groups rely extensively on unit trusts. As shown in figure 3.22, the share of financial assets held through trusts exploded around the time of, politically, the end of apartheid, and economically, liberalisation and financialisation. Over half of specifically interest bearing and dividend earnings financial assets are held in trusts. Trusts in South Africa are used more extensively, including housing mutual funds, as well as tax avoidance vehicles, and one mechanism of several to protect against wealth dilution (wealth loss across generations) Ytterberg and Weller, 2010. There is therefore a clear need to access data on trusts to gain clearer estimates of wealth at the top of the distribution, as well as to understand the mechanisms that results in the persistence of wealth concentration. Trusts are required to submit ITR12T forms, this is discussed in the appendix.

3.3. Bridging the micro-macro gap: the distribution of wealth in South Africa

This section brings together micro and macro data sources on household wealth in South Africa and discusses several main methods available to estimate the distribution of personal assets and liabilities that are harmonised to the national aggregates. We compare the results from what we identify as three broad approaches to measuring wealth inequality: direct measurement of wealth, rescaling of reported wealth, and capitalisation of income flows.

3.31. Methodological approaches: direct measurement, rescaling and capitalisation

For our purpose, it is interesting to compare three different ways of estimating the distribution of household wealth. The first one, henceforth direct measurement, consists in using reported data on the market value of the assets and liabilities of households. In South Africa, the only publicly available data source enabling such measurement for the entire spectrum of household wealth components is the NIDS survey. This approach is likely to suffer from strong under-estimation of wealth inequality due to non-response and under-sampling issues at the top of the distribution. In particular, the direct measurement approach implies that figures are not consistent with macroeconomic statistics, both in terms of levels and composition of household wealth. In the case of the NIDS we showed in section 4 that the direct measurement approach implies overstating the importance of housing assets and understating the significance of non-pension

financial assets.

A second way of measuring the distribution of wealth, which we coin as “rescaling” in what follows, consists in assuming that the distribution of recorded wealth components and their correlation is relatively well measured by the household survey, but that it is mainly the average amounts of each component which are understated or overstated. In this case, one can obtain an estimate of the wealth distribution by effectively blowing up individual-level assets and liabilities to match the totals recorded in the national accounts. The core identifying assumption is that individuals overestimate or underestimate the value of the assets and liabilities that they report, but that this misreporting is uncorrelated to rank within each asset class. This approach, as we show below, is problematic in our case because it tends to create a number of outliers, both at the top and at the bottom ends of the distribution. This is in large part because debts are very badly measured in survey data, so that rescaling reported values leads to giving unrealistic levels of debt at the very bottom of the distribution.

A third approach to measuring wealth inequality is the income capitalisation method. This approach consists in using the capital income flows corresponding to the assets and liabilities of households to approximate the distribution of wealth. In practice, this involves multiplying the income flow of a given asset class by the inverse of the rate of return of this type of asset. Just as in the case of rescaling, the capitalisation of income flows has the advantage of leading to figures which are consistent with aggregate household wealth. The identifying assumption in this case is that of constant rates of return by asset class. If the return to a given asset increases with wealth, for instance, then the income capitalisation method will lead to overestimating wealth concentration.

The capitalisation method is only possible for types of assets and liabilities generating income flows. In this paper, we rather propose a “mixed approach” as our preferred methodology to estimate the distribution of wealth, by combining income capitalisation for available income flows with rescaling when no flow counterpart data is available. As shown in table 3.35, two types of household assets, owner-occupied dwellings and currency, notes and coins cannot be capitalised and have to be measured directly from available household surveys. We choose to capitalise six types of assets: tenant-occupied dwellings from the rental income received by individual landowners; business assets from the business income received by the business owners of unincorporated enterprises; pension assets from the pension contributions and pension income of formal wage earners and pensioners; life insurance assets from factor income; bonds and interest deposits from interest income; and corporate shares and equity from dividends

Table 3.35 – Estimating the distribution of personal wealth in South Africa: a mixed approach

Asset / liability	Variable	Measurement method
Non-financial assets		
Owner-occupied dwellings	Value of home (GHS)	Rescaling
Tenant-occupied dwellings	Rental income	Capitalisation
Business assets	Business income	Capitalisation
Financial assets		
Pension assets	Pension contributions and pension income	Mixed method
Life insurance assets	Factor income	Mixed method
Currency, notes and coins	Bank account balance (NIDS)	Rescaling
Bonds and interest deposits	Interest income	Capitalisation
Corporate shares and equity	Dividends	Capitalisation
Liabilities		
Mortgage debt	Reported debt and house value	Mixed method
Other debts	Reported debts and consumption	Mixed method

Notes: The table shows the methodological approach used to estimate distribution of the different assets and liabilities reported in the household balance sheets. Direct measurement corresponds to reported data on the market value of assets or liabilities. Capitalisation corresponds to assuming that the distribution of an asset follows that of one or several corresponding income flows. GHS: General Household Survey. NIDS: National Income Dynamics Survey.

Source: authors' elaboration.

received.¹²

Mortgage debt and other forms of debts have been recorded consistently in the NIDS and other household surveys, but as we showed in section 2, the coverage of liabilities remains partial and inconsistent. As a result, rescaling debts to balance sheets totals may result in overestimating the number of individuals with negative net worth and extrapolating implausibly high debt values. Instead, we follow a mixed method: we assume that the mortgage debt from the household balance sheet is distributed proportionally to the value of the house of mortgagors in the surveys, and that other forms of debts are distributed proportionally to the consumption of those declaring having contracted debts. These are conservative assumptions, as mortgages and other forms of debt are likely to be more unequally distributed than house values and

¹²In the case of pension assets, we follow the approach proposed by [Saez and Zucman, 2016b](#) in allocating them to wage earners and pensioners so as to match their distribution recorded in the NIDS. In our case, we assume that 75 per cent of pension assets belong to formal wage earners proportionally to pension contributions paid, and 25 per cent belong to pensioners proportionally to pension income received. As we show in the supplementary appendix (figure A5), this capitalisation technique applied to the NIDS data yields results which are very similar to those obtained from direct measurement. For life insurance assets, we assume that 50% belong to wage earners proportionally to factor income – the sum of wages, self-employment income and pension income – and that 50% belong to all other adults proportionally to factor income. This again reproduces well the distribution of life insurance assets reported in NIDS (see appendix figure A6).

Table 3.36 – Shares of household wealth held by groups in South Africa:
survey-based results

	Bottom 50%	Middle 40%	Top 10%	Top 1%	Top 0.1%
Direct measurement					
NIDS, wave 4	-3.3 %	18.4 %	84.9 %	41.3 %	9.7 %
NIDS, wave 5	-0.5 %	16.9 %	83.6 %	40.2 %	8.6 %
Rescaling					
NIDS, wave 4	-8.2 %	10.9 %	97.3 %	58.3 %	24.6 %
NIDS, wave 5	-7.0 %	8.0 %	99.1 %	63.9 %	29.3 %
Mixed approach					
NIDS, wave 4	-4.5 %	14.5 %	90.0 %	58.5 %	25.2 %
NIDS, wave 5	-3.3 %	12.5 %	90.8 %	60.6 %	30.1 %
PSLSD, 1993	-1.3 %	12.0 %	89.3 %	51.7 %	20.6 %
IES, 1995	-5.1 %	15.3 %	89.8 %	50.6 %	23.7 %
IES, 2000	-1.8 %	14.9 %	86.9 %	52.8 %	26.0 %
IES, 2005	-0.2 %	13.6 %	86.6 %	54.2 %	28.6 %
LCS, 2008	-8.0 %	14.0 %	94.0 %	52.3 %	22.4 %
IES, 2010	-7.3 %	14.8 %	92.4 %	60.0 %	31.7 %
LCS, 2015	-3.2 %	14.0 %	89.2 %	51.1 %	20.0 %

Notes: The table compares estimates of the share of household wealth owned by the bottom 50 per cent (p0p50), the middle 40 per cent (p50p90), the top 10 per cent (p90p100), the top 1 per cent (p99p100) and the top 0.1 per cent (p99.9p100) obtained from household surveys using different methodological approaches. The unit of observation is the individual adult aged 20 or above. PSLSD: Project for Statistics on Living Standards and Development. IES: Income and Expenditure Survey. LCS: Living Conditions Survey. NIDS: National Income Dynamics Study.

Source: authors' computations based on data.

consumption respectively.

3.32. Measuring wealth inequality using survey data

We start by looking at the distribution of personal wealth estimated from survey data. For all the following results, we take the individual adult aged 20 or above as the unit of analysis.¹³

¹³We therefore provide “individual” wealth inequality series rather than series where wealth is divided among spouses (narrow equal-split), among adult household members (broad equal-split) or among both children and adult household members (per capita). The main motivation is that the tax microdata is only available at the individual level, so that applying equivalence scales to the survey data but not to the tax data would imply that the results are not comparable. Notice however that there are wealth components which are only measured at the household level – namely owner-occupied housing wealth, mortgage debt and non-mortgage debt. We split equally these components among adult members of the household. This is far from being a perfect solution, but overall wealth inequality is only moderately affected by changes in units of observation, especially at the top of the distribution. We report in the appendix (figure A9) how changes in equivalence scales affect survey-based top and bottom wealth shares.

Table 3.36 compares estimates of the share of wealth held by the bottom 50 per cent (p0p50), the middle 40 per cent (p50p90), the top 10 per cent (p90p100), the top 1 per cent (p99p100) and the top 0.1 per cent (p99.9p100) obtained from direct measurement, rescaling and the mixed approach. The NIDS survey is the only survey collecting direct data on wealth and thus for which the results from the three methodologies can be compared. Other household surveys collect data on the value of owner-occupied housing and household debts, so they can be used to estimate the wealth distribution with the mixed approach.¹⁴

The first result which clearly stands out is that all approaches converge in revealing an extreme degree of wealth concentration. Regardless of the methodology, the bottom 50 per cent of the South African adult population is consistently negative, while the top 10 per cent is higher than 80 per cent in all surveys and methods. According to these results, wealth inequality in South Africa appears to be substantially larger than in any other country for which relatively reliable data is available (see below).

The second result is that there are some important differences in the results obtained from the three different approaches, especially at the top of the distribution. Direct measurement in the NIDS implies a top 0.1 per cent share below 10 per cent, more than twice lower than most of the results obtained from rescaling or the mixed approach. This is mainly due to the very poor coverage of other financial assets in the NIDS, which are particularly concentrated at the top end of the wealth distribution. Rescaling financial assets to balance sheets totals or capitalising income flows corrects for this micro-macro discrepancy. Rescaling the value of assets and liabilities increases wealth inequality significantly compared to the mixed approach. This is mainly due to the fact that blowing up debts to balance sheets totals creates a large number of households with strongly negative net worth (the bottom 50 per cent goes down by several percentage points). Finally, it is interesting to note that the mixed approach yields relatively close results across years and data sources: the top 10 per cent share lies between 85 per cent and 90 per cent and the top 1 per cent is estimated to be between 50 per cent and 60 per cent in most cases. This suggests that despite the fact that these household surveys were conducted using different sampling methods and questionnaires, capitalising reported income flows remains somehow an efficient method to broadly capture the structure of wealth concentration in South Africa. Yet, all these surveys are likely to suffer from misreporting or non-response, which implies a misrepresentation of income and wealth inequality at the top

¹⁴The PSLSD, IES and LCS surveys did not collect data on currency and non-interest deposits, so we impute their value from the NIDS by assuming that their distribution has remained constant, both in terms of overall concentration and conditionally to post-tax income. Given the small share of currency and non-interest deposits in aggregate wealth, this imputation does not affect our results. Also notice that the data on owner-occupied housing wealth in the IES and LCS surveys is very erratic, so we keep the rank of housing wealth reported in these surveys but force its distribution to match that observed in the GHS between 2002 and 2018.

Table 3.37 – Shares of household wealth held by groups in
South Africa: results from tax microdata and survey
combined

	Bottom 50%	Middle 40%	Top 10%	Top 1%	Top 0.1%
2010	-6.8 %	16.6 %	90.2 %	57.3 %	30.0 %
2011	-6.4 %	16.7 %	89.8 %	57.0 %	29.3 %
2012	-5.3 %	16.5 %	88.9 %	57.2 %	33.5 %
2013	-4.0 %	16.0 %	87.9 %	56.3 %	32.1 %
2014	-3.0 %	16.2 %	86.8 %	54.5 %	29.9 %
2015	-2.9 %	16.0 %	86.9 %	55.0 %	29.2 %
2016	-2.9 %	16.2 %	86.7 %	53.5 %	27.5 %
2017	-2.5 %	16.9 %	85.6 %	54.7 %	29.8 %

Notes: The table shows estimates of the share of household wealth owned by the bottom 50 per cent (p0p50), the middle 40 per cent (p50p90), the top 10 per cent (p90p100), the top 1 per cent (p99p100) and the top 0.1 per cent (p99.9p100) obtained from the income capitalisation method combining surveys and tax microdata. The unit of observation is the individual adult aged 20 or above.

Source: authors' computations based on data.

end.

3.33. Measuring wealth inequality using tax data

We now turn to the estimation of the distribution of wealth obtained by combining surveys and tax data. As explained in section 4, tax microdata has the advantage of both better covering capital income flows and capturing with a greater degree of precision the levels and composition of incomes at the top end of the distribution. As a result, one may expect that our mixed method will lead to higher measured wealth inequality levels as compared to the capitalisation of income flows in household surveys. That being said, the important limitations of the tax data itself discussed above do prevent us from considering the tax-based estimates presented below as satisfactory.

Income tax data in South Africa does not cover the full adult population: the matched IRP5-ITR12 panel only covers between 40 per cent and 42 per cent of adults over the 2010-2017 period. In order to get a reliable estimate of wealth inequality, we combine the tax data with household surveys in two steps. In a first step, we derive an income concept which is comparable between the two sources, which we name “merging income”, defined as the sum of gross wages, self-employment income, rental income, interest income and private pension income. We then merge the two data sources based on the exact rank of “merging income” observed at

the individual level. In a second step, we identify the quantile of the South African income distribution q starting from which reported merging incomes are higher in the tax data than in the survey data, and we assume that the tax data is more reliable than the survey data only above q . In practice, this implies keeping all variables from the survey data below q , and replacing all comparable variables from the tax data above q – namely wages, self-employment income, rental income, interest, dividends, private pension income, and contributions to pension funds. Between 2010 and 2017, we find q to be consistently located between the 70th and the 75th percentiles (see supplementary appendix figures A7 and A8).¹⁵

Table 3.37 shows the results obtained from combining the surveys and tax data and applying the mixed approach. Wealth inequality appears to be relatively similar when measured by combining surveys with tax data than when measured solely from the surveys available in similar years (the NIDS, the IES 2010 and the LCS 2015). The top 10 per cent wealth share stands at between 86 per cent and 90 per cent over the 2010-2017 period. The top 0.1 per cent share exceeds 30 per cent, compared to 20 per cent to 30 per cent in household surveys. This is a relatively surprising result, as one would expect the under-representation of top incomes in surveys to imply significantly lower levels of wealth concentration. A careful look at the particular structure of capital income concentration in South Africa can help solve this apparent paradox. The relative consistency between the two sources is mainly due to the fact that both in the surveys and the tax data, financial incomes (interest, dividends and rental income) are extremely concentrated, so that both sources imply attributing a substantial share of wealth – and in particular of tenant-occupied housing, bonds and shares – to the top 0.1 per cent of the distribution.¹⁶ Thus the benefit of tax data in providing more reliability of estimates of wealth at the top end is undermined by the lack of data on capital incomes.

¹⁵Our choice of a merging point based on an income concept differs slightly from the approach of [Hundenborn, Woolard, and Jellema, 2018](#), who rather derive a taxable income concept from survey data, and then keep the tax data above the filing threshold of taxable income. The main reason for merging our two datasets based on a broad income concept is twofold. First, our IRP5-ITR12 panel covers a large number of individuals who are below the filing threshold, given that all employers in South Africa are now required to file an IRP5 tax form for all their employees, regardless of their level of remuneration. However, as is emphasised in the SARS' Tax Statistics, this rule was not followed strictly by all employers, so that the tax data cannot be considered to be representative of the universe of formal wage earners. In other words, our data covers relatively well the top of the distribution up to a certain point, below which it contains a mix of low- and middle-income wage earners. It seems therefore most useful to keep as many individuals as possible from the tax data, while removing those whose location in the distribution of income cannot be identified precisely, which is what our method does in a simple way. Secondly, defining taxable income remains a complex task, and it remains unclear whether this can be done with a sufficient level of precision and consistency, in particular given that surveys tend to not properly capture the top of the distribution.

¹⁶According to our matched survey-tax dataset, about half of rental income and 60 per cent of interest income were received by 0.1 per cent of the South African population in 2017.

3.4. The distribution of wealth in South Africa: key results and comparative perspectives

We now present some key figures on the levels, evolution and structure of wealth inequality in South Africa. Our preferred estimate of the wealth distribution is the one obtained from combining household surveys with the tax microdata and applying the mixed method. For the years preceding 2010, we use household surveys to estimate wealth inequality using the mixed method, and we assume that the under-representation of top wealth groups is similar to that observed during the 2010-2017 period. Finally, we combine the PSLSD, the IES and the LCS with labour force surveys to have a more consistent estimate of wage inequality and business income inequality – and therefore of pension and business assets –, as well as to have yearly estimates over the entire 1993-2018 period. We explain the methodology used to combine these various data sources in the appendix. We stress again that none of the results presented below are fully satisfactory given the lack of proper data available to measure the distribution of wealth and in particular of financial assets in South Africa (see sections 3 and 4).

The distribution of wealth in South Africa in 2017

Table 3.48 provides information on the number of adults, the entry thresholds, the average wealth and the share of wealth of various groups of the wealth distribution in 2017. Average wealth per adult in South Africa amounts to about 326,000 rands, or 52,200 dollars at purchasing power parity. This is three times higher than the national income per adult, which stands at about 110,000 rands (18,000 dollars) per year or 9200 rands (1450 dollars) per month.

Average wealth varies hugely across the distribution. The bottom 50 per cent of the South African population have negative net worth: the levels of the debts that they owe exceeds the market value of the assets they own. The middle 40 per cent of the distribution – individuals located between the median and the 90th percentile – have a net worth more than twice lower than the average wealth per adult. Together, the bottom 90 per cent of the South African population own about 14 per cent of total personal wealth in the economy, while the remaining 86 per cent belong to the top decile. The average wealth of the bottom 90 per cent of the population is about four times lower than the national average, while the top 10 per cent has an average wealth about nine times higher than the average wealth per adult.

Ownership is not only polarised between top and bottom wealth groups, it is also extremely concentrated within the top 10 per cent. The top 1 per cent of the South African adult population (350,000 individuals) own 55 per cent of aggregate personal wealth, and the top 0.1 per cent

Table 3.48 – The distribution of personal wealth in South Africa in 2017

	Number of adults	Wealth threshold	Average (2018 R)	Average (2018 PPP \$)	Wealth Share
Full population	35,400,000		R 326,000	\$ 52,200	100%
Bottom 90% (p0p90)	31,860,000		R 94,100	\$ 15,100	14.4%
Bottom 50% (p0p50)	17,700,000		R -16,000	\$ -2,600	-2.5%
Middle 40% (p50p90)	14,160,000	R 27,700	R 138,000	\$ 22,000	16.9%
Top 10% (p90p100)	3,540,000	R 496,000	R 2,790,000	\$ 447,000	85.6%
Top 1% (p99p100)	354,000	R 3,820,000	R 17,830,000	\$ 2,860,000	54.7%
Top 0.1% (p99.9p100)	35,400	R 30,350,000	R 96,970,000	\$ 15,540,000	29.8%
Top 0.01% (p99.99p100)	3,540	R 146,890,000	R 486,200,000	\$ 77,920,000	14.9%

Notes: The table shows the distribution of household wealth in South Africa in 2017. The unit of observation is the individual adult aged 20 or above. Wealth thresholds are in 2018 Rands.

Source: authors' computations based on data.

Table 3.49 – Share of total assets held by wealth group by asset class, 2017

	Currency	Business assets	Housing	Pensions, life insurance	Bonds & Stock
Bottom 90% (p0p90)	37.3%	40.4%	41.2%	36.2%	0.2%
Bottom 50% (p0p50)	9.7%	1.4%	14.0%	5.3%	0.0%
Middle 40% (p50p90)	27.7%	39.1%	27.2%	30.9%	0.2%
Top 10% (p90p100)	62.7%	59.6%	58.8%	63.8%	99.8%
Top 1% (p99p100)	10.6%	41.9%	27.8%	14.1%	95.2%
Top 0.01% (p99.99p100)	1.5%	13.4%	8.5%	2.1%	62.7%
% of total assets	0.6%	3.6%	28.8%	32.5%	34.6%

Notes: The table shows the shares of different types of assets held by specific wealth groups in 2017. The unit of observation is the individual adult aged 20 or above. In 2017, the top 1 per cent of South Africans in terms of net worth owned 95 per cent of the bonds and corporate shares in the economy. Bonds and shares represented 34.1 per cent of total household assets in the economy at this date. Figures may not add up due to rounding.

Source: authors' computations based on data.

alone (35,000 individuals) own almost a third of wealth. The top 0.01 per cent of the distribution, amounting to some 3,500 individuals, own about 15% of household wealth, greater than the share of wealth owned by the bottom 90 per cent of the population consisting of 32 million individuals. They have an average wealth which is more than 1500 times that of the average South African adult, and 6000 times that of the bottom 90 per cent.

The composition of personal wealth across the distribution

The extreme degree of wealth inequality that we observe is in large part driven by the relative exclusion of poorer wealth groups from any form of wealth accumulation, and by the concentration of all forms of assets at the top end of the distribution. Table 3.49 provides some insights into this polarisation by showing the share of different types of assets held by wealth groups across the distribution. The top 10 per cent own more than 55 per cent of all forms of assets, including pension assets, housing wealth, business assets and currency, notes and coins. They own more than 99 per cent of all bonds and stock held in the economy. The top 1 per cent alone holds more than a tenth of all forms of assets and as much as 90 per cent of bonds and corporate shares. Currency and housing wealth are the least concentrated form of wealth, but low wealth groups only possess a small share of them: the bottom 50 per cent of the wealth distribution own about 10 per cent of currency, notes and coins, and less than 15 per cent of housing assets.

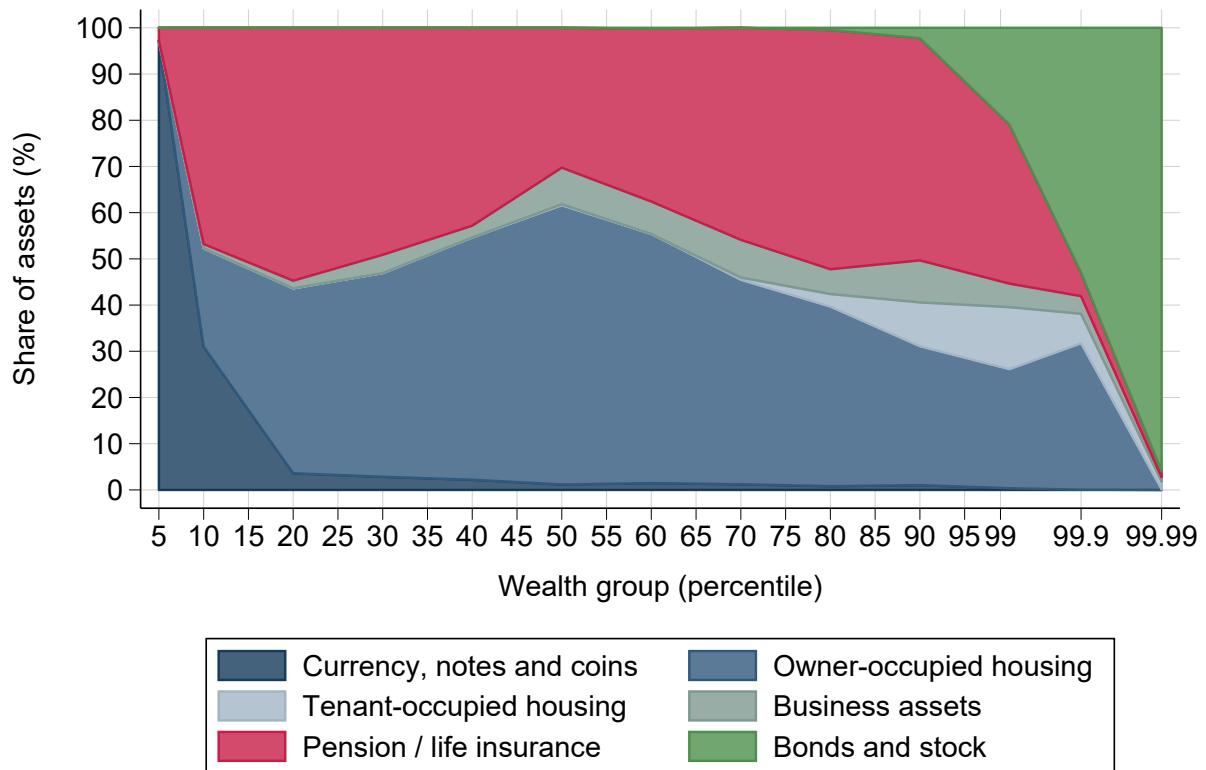
Figure 3.43 gives another view of the link between forms of asset and wealth inequality by showing the portfolio composition of percentiles in the wealth distribution. Currency, notes and coins are the main form of assets held by poorest South African adults, while owner-occupied housing, pensions and life insurance form the majority of assets for most remaining income groups within the bottom 90 per cent. Business assets represent a small share of portfolios for the upper-middle class. Bonds and stock, finally, represent a large share of wealth for the top 1 per cent and the bulk of assets of wealth groups within the top 0.1 per cent.

Wealth and age¹⁷

How does wealth change during the life cycle, and to what extent wealth accumulation and reduction throughout the lifetime account for wealth inequalities? Figure 3.44 shows a stable relationship between age and average wealth over the 2012-2017 period. Average net worth rises significantly and linearly between ages 20 and 55: individuals aged between 20 and 25 have an average net worth lower than 25 per cent of the national average, while those aged between 50 and 55 are between 50 per cent and two times wealthier than the average adult. Average wealth then stabilises between ages 50 and 65 and decreases slightly for older individuals, but still remains more than 50 per cent higher than the national average for individuals older than 75. Interestingly, this pattern is almost perfectly similar to that found in the case of France [Garbinti, Goupille-Lebret, and Piketty, 2017](#).

¹⁷There are other important categories to investigate in the context of wealth inequality in South Africa. Although the tax data is more complete, it has less covariates than the surveys, therefore, given our methodology, we are restricting our decomposition to age. We leave gender, race and other related categories for future work.

Figure 3.43 – The composition of assets by wealth group in 2017

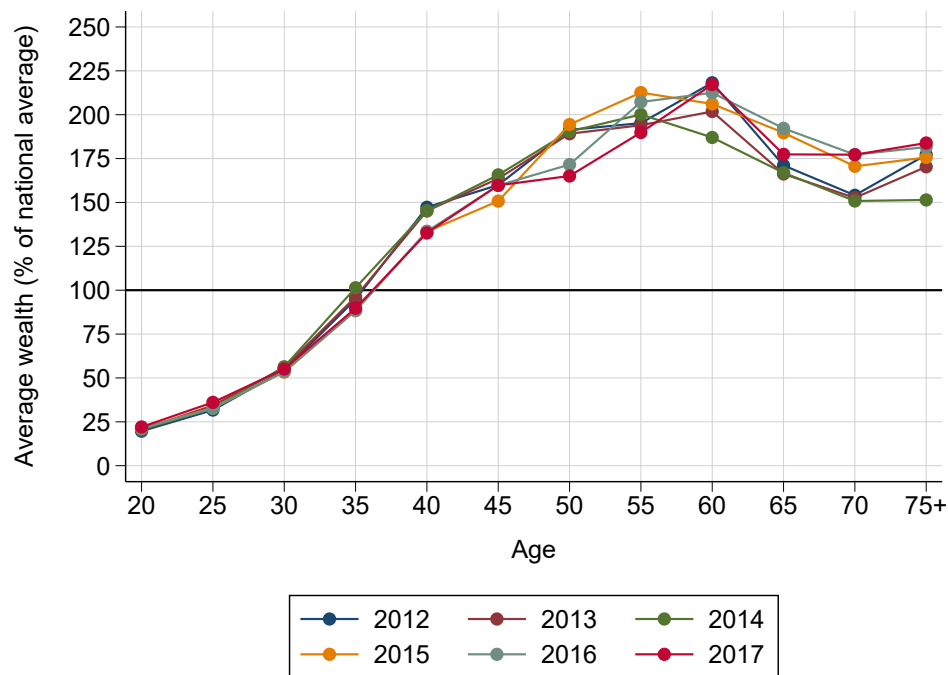


Notes: The figure shows the composition of assets of various groups in the distribution of household assets in South Africa in 2017. The unit of observation is the adult aged 20 or above. The results come from the harmonised survey data file, and wealth is split equally among adult members of the household, except for the top 1 per cent and above for which the individual data built from the combined survey and tax microdata are used.

Source: authors' computations based on data.

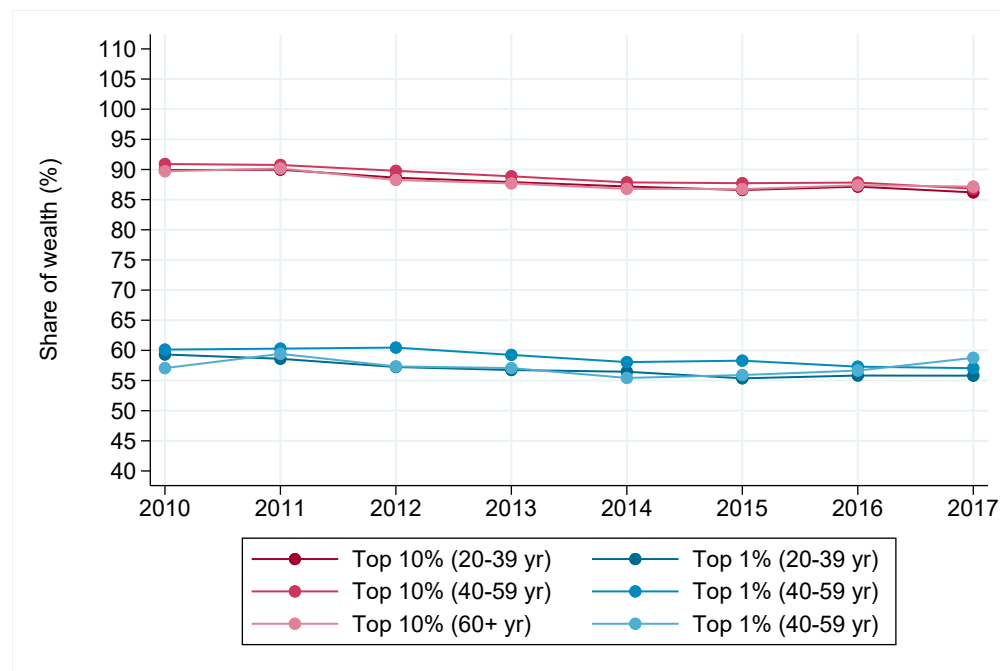
While average wealth does vary significantly across age groups, age does not explain the observed levels of wealth concentration. Top wealth shares are almost perfectly similar within each age group than in South Africa as a whole: the share of wealth held by the top 10 per cent exceeds 85 per cent, and the top 1 per cent share is higher than 55 per cent, whether one restricts the analysis to those aged between 20 and 39, between 40 and 59, or older than 60 (figure 3.45). This apparent paradox can be better understood when directly comparing differences across age groups to differences across wealth groups in the distribution. The average wealth of those aged 20 to 25 was about 6.5 times lower than that of those aged 75 or above in 2017. In comparison, the average wealth of the top 10 per cent was about 30 times higher than the average net worth of the bottom 90 per cent of the distribution. Our results therefore point to inequalities in access to wealth accumulation across the life cycle – via income

Figure 3.44 – Average wealth by age relative to average wealth per adult, 2010-2017



Notes: The figure shows the mean net worth of South African adults by age group relative to the national average. The unit of observation is the individual adult aged 20 or above. Authors' computations based on data.

Figure 3.45 – Wealth inequality within age groups, 2010-2017



Notes: The figure shows top 10 per cent wealth share and the top 1 per cent wealth share estimated when splitting the South African population into three age groups (20-39 years old, 40-59 years old, and 60+ years old). The unit of observation is the individual adult aged 20 or above. Authors' computations based on data.

inequality, debt and savings patterns – as well as inequalities in access to inheritance as being the primary drivers of the high levels of wealth concentration observed in South Africa.¹⁸

Long-run trends and comparative perspectives

We conclude this paper by bringing together our South African wealth inequality series with comparable data on other countries. For all the countries outlined below, corresponding studies followed the Distributional National Accounts methodology by combining all available micro and macro data sources to distribute household wealth [Alvaredo et al., 2016](#). They are therefore directly comparable to our estimates. Figure 3.46 plots the evolution of the share of wealth accruing to the top 10 per cent in South Africa, China, Russia, India and the United States.¹⁹ The top 10 per cent wealth share has risen in all these countries, while in the long run it has remained more stable in South Africa, increasing between 2005 and 2010 but gradually moving back to its level of the early 2000s since then. Wealth concentration has remained substantially higher in South Africa since the beginning of the 1990s than in any other country for which comparable data is available. The South African top 10 per cent wealth share has fluctuated between 80 per cent and 90 per cent during the 1993-2018 period, while it has remained below 75 per cent in the US, 70 per cent in Russia and China, 65 per cent in India and 55 per cent in France or the United Kingdom. The same result holds for the top end of the distribution: the top 1 per cent wealth share was 55 per cent in South Africa in 2017, compared to 43 per cent in Russia, 39 per cent in the United States, 31 per cent in India, 30 per cent in China and less than 25 per cent in France and the UK (figure 3.47).

In terms of trends, our results suggest that wealth inequalities in South Africa have remained stable at very high levels since 1993. Two facts are however worth noticing. First, wealth concentration seems to have rapidly increased between 2005 and 2008 before slowly coming

¹⁸Notice that the estimates presented here correspond to individual series, rather than to “equal-split” series where wealth would be split equally among household adult members. In practice, splitting wealth among household members would imply redistributing wealth to younger individuals, thereby making the wealth-age profile less steep. If anything, this reinforces our argument that age is not a primary determinant of wealth inequality in South Africa.

¹⁹Notice that the wealth shares presented here for South Africa are based on individualisation of assets, given the difficulty to split wealth equally among household members or spouses after combining surveys and tax data. In the capitalised survey series, moving from individual series to broad equal-split series decreases slightly the top 10 per cent share by 4-5 percentage points, and has a more limited effect on top 1 per cent and top 0.1 per cent wealth groups (see figure A9). This implies that our series are not perfectly comparable to that of other countries, which generally split wealth equally among spouses (narrow equal-split). From available evidence on the differences between individual, broad equal-split and narrow equal-split series, we can expect top shares in narrow equal-split series for South Africa to be lying between the broad equal-split and the individual-based results, so the top 10 per cent share would be lower by between 1 and 3 percentage points [Blanchet, Chancel, and Gethin, 2019](#). Changing equivalence scale will therefore have no consequence on our main conclusions.

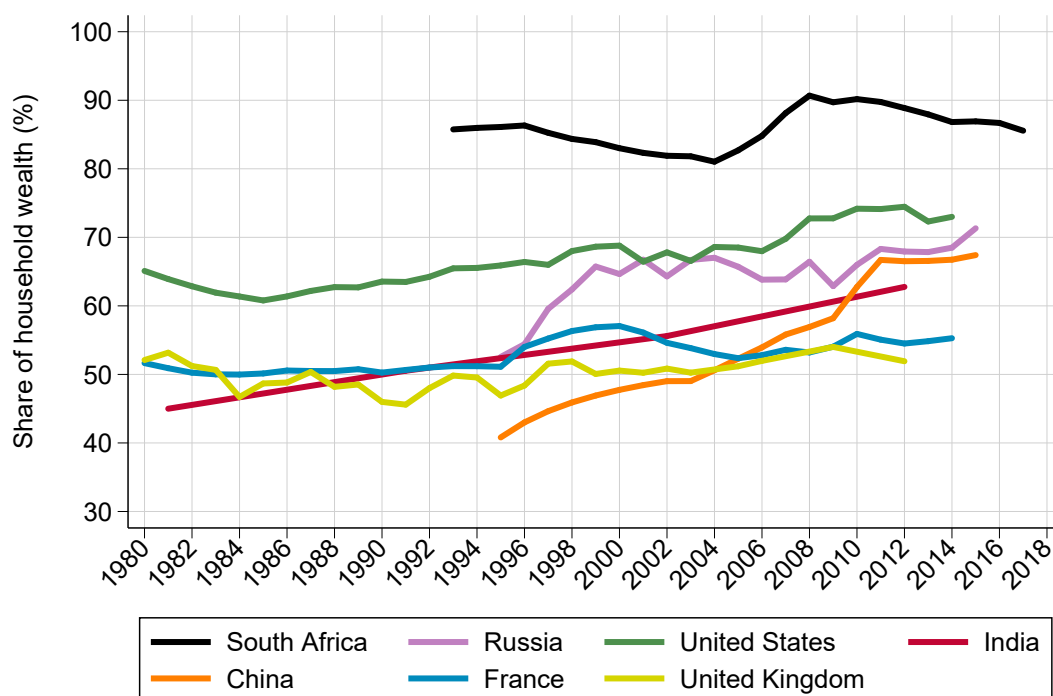
back to its long-run level between 2009 and 2017. This short-run dynamic was in large part due to the strong fall in the bottom 90 per cent share driven by the boom and bust in mortgage advances in the 2000s, which temporarily drove a higher share of households into negative net worth. Between 2004 and 2008, in particular, mortgage debt increased from 9 per cent of net household wealth to almost 15 per cent, and decreased back to 9 per cent in 2018 (see appendix figure A4). This temporary fall in bottom wealth shares driven by expanding debts mirrors that observed in the US at about the same period (see appendix figure A3). In the appendix, we discuss in greater detail the importance of household debt in South Africa and how it explains why bottom wealth shares have remained significantly negative throughout our period of interest. We also show that the concentration of household assets has remained remarkably constant (i.e. excluding debts from the analysis removes virtually all fluctuations across the period): the top 10 per cent share of assets has remained at about 80 per cent between 1993 and 2018 (see appendix figure A13).

A second result which appears from our long-run series is that while the top 10 per cent share has remained broadly stable, there seems to have been a slight increase in inequality within the top 10 per cent. Between 1993 and 2017, the top 1 per cent share increased from 54 per cent to 57 per cent and the top 0.1 per cent share from 22 per cent to 31 per cent (see supplementary appendix figure A2). This is likely to be due to two factors: the increase in the share of non-pension financial assets from 19 per cent to 24 per cent of net household wealth between 1992 and 2018, and the increase in wage inequality in South Africa during this period – which indirectly affected the distribution of pension assets. That being said, we should stress that the low quality and the important issues regarding the comparability of the household surveys conducted during this period do not allow us to conclude to an increase in wealth inequality with a high degree of certainty. The best we can say with a certain level of confidence is that there is no evidence that wealth inequality in South Africa has decreased since the end of apartheid, and that South Africa remains significantly more unequal than any emerging or developed country for which good-quality data on the distribution of wealth is available throughout the world.

3.5. Conclusion

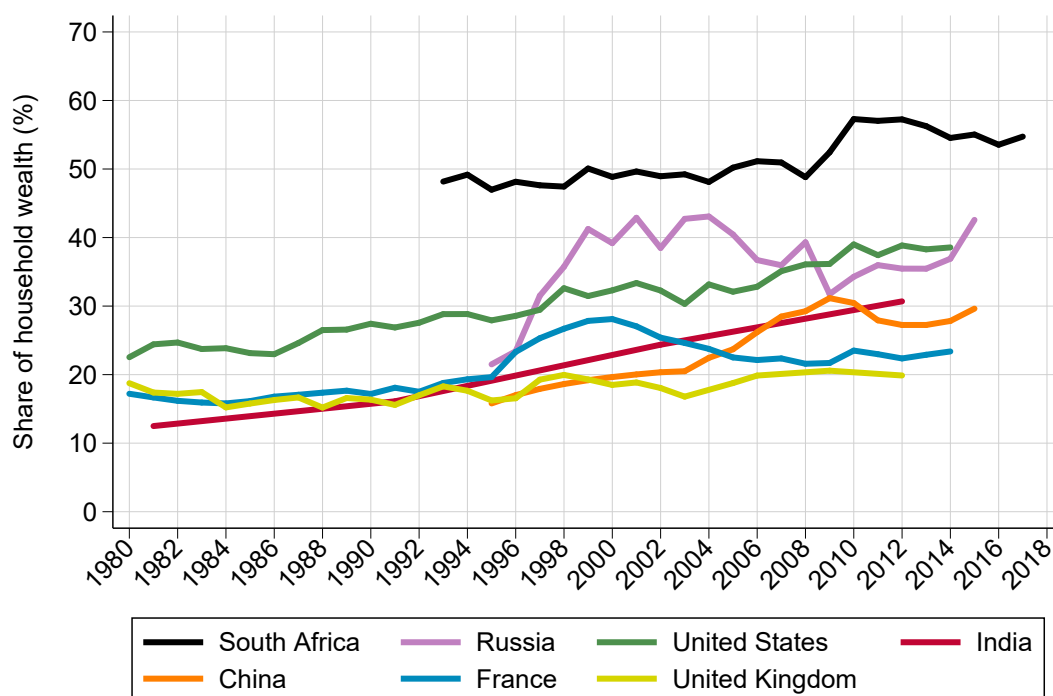
This paper presented a first attempt to systematically compare income and wealth reported in surveys and tax data to official macroeconomic statistics in South Africa, and to assess several methods to correct the micro-macro gap in the measurement of households' net worth. Our analysis has revealed two main findings.

Figure 3.46 – South African wealth inequality in comparative perspective: Top 10 per cent wealth share



Notes: The figure compares the top 10 per cent wealth share in South Africa to that of other countries. The unit of observation is the individual adult aged 20 or above. Wealth is individualised (South Africa) or split equally among adult household members (other countries). Authors' computations based on data for South Africa; World Inequality Database (<http://wid.world>) for other countries.

Figure 3.47 – South African wealth inequality in comparative perspective: Top 1 per cent wealth share



Notes: The figure compares the top 1 per cent wealth share in South Africa to that of other countries. The unit of observation is the individual adult aged 20 or above. Wealth is individualised (South Africa) or split equally among adult household members (other countries). Authors' computations based on data for South Africa; World Inequality Database (<http://wid.world>) for other countries.

Firstly, the data sources available to measure wealth in South Africa remain largely unsatisfactory. Reported housing wealth is substantially higher in household surveys than in balance sheets statistics, while most surveys cover very poorly business assets, financial assets and household debts. More importantly, two major limitations will have to be addressed in future research: the absence of any reliable source of the distribution of dividends received by households, and the lack of distributional data on the wealth held through unit trusts, which represents a substantial part of the net worth of top end groups in recent years. While such data exist – through the trust forms and dividends tax forms reported by taxpayers to the South African Revenue Service –, they have unfortunately not yet been made available to academic researchers. Access to such data will be crucial to understand not only the distribution of wealth, but also the processes by which wealth gets accumulated, transmitted and redistributed in the economy.

Secondly, all data sources do suggest that wealth inequality in South Africa is the highest among all countries for which data is available. The top 10 per cent own more than 85 per cent of wealth, and the top 0.1 per cent at least 25 per cent. Drawing on household surveys, we find no evidence that wealth concentration has decreased since apartheid. This finding calls for further research on the dynamics of asset transmission and on the factors accounting for the persistence of wealth inequality in the post-apartheid era.

Chapter 4

Redistribution without Inclusion? Inequality in South Africa Since the End of Apartheid

with Aroop CHATTERJEE and Amory GETHIN

Abstract This article sheds new light on the evolution of income inequality and government redistribution in post-apartheid South Africa. We combine survey, tax, and historical budget data to construct a new microdatabase on the distribution of labor and capital incomes, taxes, cash transfers, and public services since 1993. Pretax income inequality has increased, but this rise has been overcompensated by major expansions in government redistribution. After accounting for taxes and transfers, low-income households have benefited from the greatest real income gains. However, South Africa still stands out as one of the most unequal countries in the world. In 2019, the top 1% received almost 20% of posttax income, more than the bottom 50% as a whole. Racial inequalities have declined, but this decline has been entirely driven by the boom of top Black income groups. We highlight the role of taxes and transfers as powerful levers of inclusive growth yet insufficient tools to curb South Africa's extreme inequalities.

4.1. Introduction

Numerous studies have provided new insights into the determinants of economic deprivation in recent years, yet considerable challenges remain when it comes to accurately understanding the link between poverty, inequality, and growth. How inclusive has economic growth been in the developing world in the past decades? To what extent have cash transfers and government investments in health, education, and infrastructure development accrued to low-income groups, and what fraction of these benefits has been mitigated by an increased tax burden? Because of a critical lack of data on the joint distributions of income, consumption, taxes, and transfers, answering these questions has until today proved to be a remarkably challenging task. At the heart of this difficulty lies major differences in data sources, methods, and research communities. At the micro level, studies investigating poverty and inequality have almost exclusively relied on household surveys, often the only source at our disposal to observe the distributions of income, consumption, and wealth. At the macro level, researchers studying the determinants of growth have mostly worked with national accounts, which provide crucial information on key macroeconomic aggregates and the size of government intervention in the economy. The sometimes inconsistent and conflicting stories arising from these two sources have made it particularly difficult to understand how economic growth is shared over time and to what extent government redistribution in its various forms effectively benefits the poor.

This paper attempts to make progress in this direction by constructing a new micro dataset on the distribution of macroeconomic growth in South Africa from 1993 to 2019. Combining available data sources—surveys of various kinds, income tax data, national accounts, and historical administrative data on government taxes and expenditure from budget reports—we systematically allocate all components of the net national income, all government taxes, and all government expenditure to individuals. The resulting dataset is consistent with income, expenditure, and wealth aggregates reported in the national accounts. It is also consistent with what we know from administrative reports on various key parameters of government intervention, including the number of recipients of social grants, the total spending on each of these grants, the distribution of top taxable incomes reported in personal income tax data, and other key statistics on the size and targeting of taxes and transfers. Importantly, it covers in-kind transfers and public goods, which are particularly large and progressive in South Africa, drawing on recent related work ([Gethin, 2023](#)). It also incorporates all these key parameters while keeping the richness of the information reported in household surveys, allowing us to decompose the evolution of inequality, redistribution, and growth according to various economic variables such as consumption, labor income, capital income, and wealth, or sociodemographic

variables such as age, gender, race, and geography.

The case of South Africa is particularly revealing of the limitations we face in our understanding of the links between inequality, redistribution, and growth. On the one hand, the country is widely acknowledged as standing at the upper frontier of contemporary inequality today (Alvaredo et al., 2018a). The richest 10% own a striking 85% of total household wealth, with an average net worth exceeding \$440,000 at purchasing power parity (Chatterjee, Czajka, and Gethin, 2022), while 57% of the population lives with less than \$5.5 per day (World Bank, 2020). These extreme disparities, despite the end of the apartheid regime of racial segregation and exclusion at the beginning of the 1990s, have been found to have increased significantly in the past decades, driven by the boom of top incomes, chronic unemployment, and persisting household indebtedness (Bassier and Woolard, 2018; Leibbrandt et al., 2010).

On the other hand, South Africa is often regarded as displaying one of the most ambitious and efficient welfare states of the developing world. It has developed a highly progressive personal income tax, which collects substantial revenue in comparison to the majority of other emerging and developing economies. It has invested growing resources in education, health, and social protection, and its relatively well-targeted social grants system has provided critical social relief to the poor and the elderly (Bassier et al., 2020; Duflo, 2000; Dutronc-Postel and Tondini, 2023; Maboshe and Woolard, 2018; Tondini, 2022). The reductions in inequality and poverty operated by South Africa's tax-and-transfer system have even been found to be the largest achieved among all emerging economies with comparable data (Inchauste et al., 2015).

This contrasting trajectory, mixing rising pretax income inequality, low growth, and large and increasing government redistribution leaves us with a puzzling and unclear track record of South Africa's success in improving the living conditions of the poor since the end of apartheid. We do not know, for instance, whether the decline in absolute poverty observed in the 2000s, as measured by consumption expenditure reported in household surveys, was driven by higher market incomes, improved access to credit, or social transfers (such as the Child Support Grant introduced in 1998). Commonly used consumption or income aggregates do not account for in-kind transfers, such as education and healthcare, hence leaving aside crucial elements of government redistributive policy. We know even less of the distributional incidence of taxes, in particular indirect taxes (such as VAT or excise duties) and the corporate income tax, which are generally excluded from studies tracking the evolution of inequality and poverty over time. The objective of this paper is to make advances in filling these gaps by making the best of all the available data sources at our disposal (surveys, tax data, national accounts, and historical data on the structure of taxes and transfers) to get a more complete picture of the distribution of growth and redistribution over time. While we still face considerable challenges in measuring

these various components and our analysis is not devoid of limitations and uncertainties, we hope that it can contribute to improving our knowledge of how inequality, poverty, and redistribution interact in the long run.

Our analysis reveals a number of striking findings. National income per capita grew by about 35% from 1993 to 2019, yet this figure masks large heterogeneity across income groups. The average pretax income of the top 1% increased by almost 80%, while that of the poorest 20% declined. The share of pretax income accruing to the top 10% of the population thus shifted from about 64% to 69%, putting South African inequality levels much higher than anywhere else in the world, including countries such as Brazil (57%), India (57%), or the United States (45%). This dramatic rise of top incomes was driven by both capital and labor income, although labor income played a more decisive role after the 2007-2008 crisis.

Turning to the impact of taxes and transfers, we find that major increases in government redistribution more than compensated the rise of pretax inequality. This transformation was driven by both cash and in-kind transfers, which increased in size and became more progressive over time. The rise of redistribution was in part financed by higher taxes on the top 1%, who saw their effective tax rate shift from about 25% to 40%, mainly through expansions in the personal income tax and the corporate income tax. A significant part of this redistribution, however, was annulled by increases in taxes paid by the poorest 50% in the form of VAT, excise duties, and local taxes. In 2019, the profile of taxes paid by pretax income group was thus distinctly U-shaped, with higher tax rates paid by the bottom and the top of the distribution than by middle income groups, who saw their tax burden remain nearly unchanged in the past twenty-five years.

All in all, we find that growing redistribution generated substantial improvements in the living standards of South Africa's poorest individuals, but only had small effects on overall inequality. After accounting for taxes and transfers, the top 1% income share stood at nearly 20% in 2019, almost the exact same level as in 1993. The rise of South Africa's welfare state has thus succeeded at redistributing the fruits of economic growth, but it has been insufficient to curb the extreme inequalities inherited from apartheid.

Finally, we decompose inequality and redistribution into two key historical determinants of South Africa's extreme economic disparities: race and geography. We document a significant decline in the income gap between White and Black South Africans since the mid-2000s: the ratio of average White income to average Black income fell from 14 in 2005-2009 to 8 in 2015-2019. However, much of this decrease can be accounted for by the top 10% of Black earners, who witnessed exceptional income gains. When we exclude this group from the analysis, the racial income gap appears to have stood at about the same level in 2019 as since 1993. Racial

inequalities are substantially larger in terms of wealth than in terms of income or consumption and are only moderately reduced by the tax-and-transfer system. Turning to geography, we find that the South African state operates significant transfers from the two richest provinces, Gauteng and Western Cape, to the rest of the country, although posttax spatial inequalities remain large. Another dimension of inequality for which redistribution seems to have succeeded at fully absorbing the rise of pretax inequalities is the rural-urban gap, which grew dramatically in terms of pretax income but remained stable after accounting for taxes and transfers. This striking rise of redistribution from urban to rural areas is mostly attributable to higher in-kind transfers and public goods, which disproportionately improved in rural areas over the period considered.

This paper contributes to the growing literature attempting to bridge the micro-macro gap in poverty and inequality studies. [Piketty, Saez, and Zucman, 2018](#) combine surveys, tax, and national accounts data to create Distributional National Accounts (DINA) allocating the entirety of national income growth to individuals in the United States since 1913. A number of studies following the DINA framework (see [Alvaredo et al., 2018a](#)) have been conducted since then on other countries or regions of the world, with the objective of constructing comparable, yearly statistics on the long-run distribution of income and wealth.¹ The major innovation of DINA studies is their consistency with macroeconomic figures reported in the national accounts and their allocation of all taxes and transfers (including indirect taxes, in-kind transfers, and collective government expenditure) to individuals. One of their limitations, however, has been the degree of precision with which taxes and transfers are distributed. [Piketty, Saez, and Zucman \(2018\)](#), for instance, distribute education spending as a lump sum per child, leaving aside variations in expenditure across space and level of education. [Blanchet, Chancel, and Gethin \(2022\)](#) distribute health expenditure on a lump sum basis in the context of Europe, assuming that all adults benefit from the same amount of health investment regardless of age, location, or socioeconomic status. Similarly, [Bozio et al. \(2020\)](#) allocate all consumption taxes on value added, energy, or tobacco proportionally to overall consumption expenditure, regardless on the type of goods on which these taxes fall.

Several studies have made significant efforts to refine our understanding of the distribution of indirect taxes and in-kind transfers, but they have typically not done so in a way that is consistent with the national accounts. In the context of South Africa, [Inchauste et al. \(2015\)](#) exploit data from the Living Conditions Survey to allocate government taxes and social spending to individuals in a particularly granular way, combining for instance microdata on educational

¹See for instance [Blanchet, Chancel, and Gethin \(2022\)](#) on Europe, [Garbinti, Goupille-Lebret, and Piketty \(2018b\)](#) and [Bozio et al. \(2020\)](#) on France, or [Piketty, Yang, and Zucman \(2019a\)](#) on China. The results of these studies have been compiled in the World Inequality Database (see <http://wid.world>).

attendance by program with figures on total expenditure on each of these programs by province to allocate total education spending. This allows the authors to derive a much more precise estimate of the distributional incidence of some of the elements of the tax-and-transfer system. However, this estimate covers only one year, is not consistent with national accounts, relies exclusively on surveys (which tend to underestimate income at the top end), and excludes key components of government revenue and spending (e.g., the corporate income tax).

In this paper, we attempt to take the best from all of these contributions to derive a comprehensive picture of the distribution of growth, taxes, and transfers in South Africa given the data at our disposal. We directly follow the Distributional National Accounts framework and distribute, component by component, the national income between 1993 and 2019. Our estimates account for incomes that are never directly received by individuals, such as imputed rents or corporate undistributed profits, yet are key components of macroeconomic growth figures. We allocate all taxes to individuals, accounting for key features of the tax system such as VAT-exempt goods, the types of expenditure facing excise taxes, the heterogeneous effects of trade duties through variations in import densities by type of good, expenditure made in the informal sector, and personal income tax exemptions. We distribute all government expenditure as precisely as possible, incorporating information on the value and the number of recipients of each social grant from historical budget reports, excluding individuals relying on private health insurance or going to private schools from public spending, and decomposing education and health transfers by province and function following recent work by [Gethin \(2023\)](#). Stated in a general way, our methodological contribution compared to previous DINA studies is to further refine the allocation method so as to more closely mimic policy rules and better take into account the heterogeneity across individuals. Although our estimates are far from being perfect and could be improved as better data becomes available, we hope that these methodological insights can contribute to make new steps towards the much needed reconciliation between macro and micro sources in economics research.

Section [4.2](#) covers data sources and methodology. Section [4.3](#) presents results on the distribution of income before accounting for taxes and transfers. Section [4.4](#) studies the impact of taxes and transfers on inequality and the distribution of growth since 1993. Section [4.5](#) decomposes inequality and redistribution by race and geography.

4.2. Data and Methodology

This section presents the data sources and methodology used to estimate the distribution of pretax income, posttax income, consumption, and wealth in South Africa between 1993 and

2019. Section [4.21](#) outlines our conceptual framework. Section [4.22](#) explains how we distribute factor national income combining surveys, tax, and national accounts data. Section [4.23](#) details how we move from factor income to pretax income. Sections [4.24](#) and [4.25](#) cover the allocation of taxes and transfers. Section [4.26](#) describes how we estimate the distribution of household final consumption expenditure and household wealth.

4.21. Conceptual Framework: Distributional National Accounts

We are interested in distributing the consumption, income, and wealth aggregates codified in the United Nations' System of National Accounts (UN SNA), which are routinely estimated by statistical institutes and used to estimate and decompose macroeconomic growth. These include net national income, household final consumption expenditure, and household net worth.

Net National Income. Our benchmark income concept is net national income. National income equals GDP minus capital depreciation plus net foreign income. It is the sum of the primary incomes of the different sectors of the economy: households, corporations, and the government (see Table [4.81](#)). The primary income of households can itself be decomposed into four main components: compensation of employees, mixed income, net property income, and the imputed rents of owner-occupiers. The primary income of corporations corresponds to the net benefit that companies retain after having paid suppliers, employees, shareholders, and taxes, and that we refer to interchangeably as “retained earnings” or “undistributed profits”. The primary income of the general government is the sum of taxes less subsidies on production and imports (i.e., indirect taxes collected during the production process) and of its net property income.

Distributional Income Concepts. Following the DINA framework ([Alvaredo et al., 2020](#)), we consider three main income concepts to distribute national income at the individual level. Factor national income is the sum of all income flows accruing to individuals before any tax or transfer. Pretax national income equals factor income after the operation of unemployment and pension systems, that is, after payment of social contributions and distribution of pension and unemployment benefits. Posttax national income equals pretax income after deduction of all taxes (including indirect taxes and the corporate income tax), payment of all kinds of transfers (including collective government expenditure in health, education, defense etc.), and allocation of the general government deficit or surplus. By definition, individual factor incomes, pretax incomes, and posttax incomes all add up to the net national income.

Distributional Consumption and Wealth Concepts. In addition to income, we also distribute consumption and wealth concepts consistent with national accounts definitions. Household final consumption expenditure (HFCE) is the sum of all purchases made by resident households. The net saving of households is the difference between net disposable income (posttax income excluding collective government expenditure) and HFCE. Personal wealth is the net wealth of the households sector, that is, the sum of all financial and non-financial assets held by households, minus their financial liabilities.

4.22. From Reported Household Income to Factor National Income

We now outline our methodology to distribute factor national income. We first combine survey and tax data to measure the distribution of reported household income (wages, property income, and mixed income). We then allocate unreported income components (imputed rents, property income attributed to policy insurance holders, undistributed profits, and government primary income) to individuals. Table 4.81 outlines the methodology used to distribute each of these subcomponents of factor national income. We discuss in greater detail these methodological steps in appendix 4.8.

Harmonization of Survey Data. Household surveys represent our main data source to distribute income at the individual level. Seven surveys collecting detailed information on all components of household income and expenditure have been conducted in South Africa since 1993. We combine these “income surveys” with labor force surveys, which provide more detailed data on wages and self-employment income on an annual basis, to build a microfile covering the distribution of “reported household income” every year since 1993.

Combination of Survey and Tax Data. Surveys can be well-suited to measure income and expenditure at the bottom of the distribution, yet they are well-known to underestimate inequality at the top end (e.g., [Blanchet, Chancel, and Gethin, 2022](#)). To better capture the levels and dynamics of top incomes, we combine our survey microfile with tabulated income tax returns available from the South African Revenue Service. The available tabulations report the number of taxpayers and total taxable income by income tax bracket every year since 2002. We correct the survey data with the tax data in four steps. First, we approximate full distributions from the tax tabulations using Generalized Pareto Interpolation ([Blanchet, Fournier, and Piketty, 2017a](#)). Secondly, we define a “taxable income” concept in the survey data that is comparable to that observed in the tax data (excluding in particular dividends, which are not subject to personal income tax in South Africa). Thirdly, we calibrate the survey microdata on the tax tabulations

using the algorithm developed by [Blanchet, Flores, and Morgan \(2018a\)](#), which reweighs survey observations so as to match the distribution of top taxable incomes reported in the tax data. This method has the major advantage of preserving the survey microdata and the dependency between its different variables (such as income components and sociodemographics), while enforcing that the survey becomes fully representative of top taxable incomes, in the same way that statistical institutes routinely adjust survey weights to make them more representative in terms of age or gender. Finally, we extrapolate the correction to the 1993-2001 period, for which no tax data is available, assuming that top incomes were underrepresented during this period to the same extent as in 2002.

Rescaling of Household Income Components to National Accounts Totals. Having combined survey and tax data, we now have a microfile covering reported household income for the full South African population since 1993. However, for various reasons linked to sampling, mismeasurement of income flows, and non-response, income aggregates reported in this microfile do not necessarily match those recorded in the national accounts. Following other DINA studies, we rescale proportionally each of the five income flows reported in survey and tax data—compensation of employees, mixed income, rental income, interest, and dividends—to their corresponding national accounts totals. This step only has minor distributional implications at the bottom of the distribution, but it leads to significantly increasing the income share of the top 1%. This is because capital incomes, in particular interest and dividends, are both massively underreported in household surveys and by construction mostly absent from South African income tax data ([Chatterjee, Czajka, and Gethin, 2022](#)).

Imputed Rents. The imputed rents of owner-occupiers represent about 3% of national income. Imputed rents are not recorded consistently in South African surveys as such, but income surveys have asked households to give an approximate value of the value of their home since 1993. We use this information to distribute imputed rents proportionally to the market value of owner-occupied housing wealth.

Other property income. Other property income, also referred to as property income attributed to insurance holders and pension entitlements, corresponds to investment income indirectly received by individuals through their ownership of unmatured insurance and pension assets. Accordingly, we assume that it is distributed proportionally to pension and life insurance assets, estimated by combining data on wages, social contributions, and self-reported wealth data from the National Income Dynamics Study (see [Chatterjee, Czajka, and Gethin, 2022](#)). This component represents a significant share of national income in South Africa (6% in 2018),

where private pensions, life insurance policies, and investment funds are widespread and have been growing in the past decades.

Interest Paid by Households. Household debts in the form of mortgages and other loans are significant in South Africa (53% of national income in 2018), and particularly widespread at the bottom of the wealth distribution ([Chatterjee, Czajka, and Gethin, 2022](#)). Accordingly, interest paid by households represents a sizable component of national income, reaching 5% of NNI in 2018. Data on debt balances have been recorded in income surveys since 1993, but debt repayments are only partially and inconsistently measured. To avoid artificially creating too many households with negative income, we therefore choose to distribute interest paid proportionally to factor income among individuals who declare having unpaid debts.

Corporate Undistributed Profits. Undistributed profits correspond to profits that are kept within the company rather than distributed to shareholders as dividends. These income flows ultimately increase the wealth of shareholders and therefore represent a source of income to them. Accordingly, we allocate retained earnings proportionally to stock ownership, including both directly held shares and shares held indirectly through pension funds. We only distribute the share of retained earnings attributable to the private domestic sector, hence excluding that held by the government.

Taxes less Subsidies on Production and Imports. We allocate the primary income of the government proportionally to factor income, assuming this component of national income is distributionally neutral. This assumption is meaningful to the extent that one could replicate our entire analysis by relying on a definition of net national income at factor cost (instead of market prices), excluding indirect taxes and subsidies from the final measure of output. Our inequality series is thus insensitive to adopting one or the other of these approaches to national accounting.

Remaining components of factor national income. The remaining components of national income (3 % of NNI) mainly include government and foreign shares of corporate retained earnings, as well as other small income flows such as miscellaneous government transfers. In the absence of better information on the incidence in these items, we assume for simplicity that they are distributionally neutral and allocate them proportionally to factor income.

4.23. From Factor National Income to Pretax National Income

To recover pretax income from factor income, we remove all pension and unemployment contributions from individual income and we add all corresponding pension and unemployment benefits. This has only minor distributional incidence in South Africa, given that private pension benefits are received by a small share of the population and that the unemployment insurance system only redistributes a tiny fraction of national income (see Table 4.81).

Pension Contributions. Contributions to private pension plans (6% of national income) are recorded in income surveys, so we directly deduct them from individual factor incomes.

Pension Benefits. Private pension benefits (3% of national income) are also recorded in income surveys. However, these surveys tend to significantly underestimate the share of adults receiving private pension income (2-3% in income surveys vs. 5-6% according to administrative data). We use predictive mean matching to impute incomes to individuals declaring no pension income but with characteristics similar to those who do, in such a way that the total number of pension income recipients becomes exactly equal to that observed in administrative data sources. This ensures that our microfile is representative of what we know about the actual number of recipients of pension benefits in South Africa, while preserving the observed relationships between pension income and the other covariates recorded in the surveys.

Unemployment Insurance Contributions. Unemployment insurance contributions are set at a fixed rate of 2% of gross wage in South Africa and capped at a maximum amount in Rand. About 25% of adults contribute to the Unemployment Insurance Fund (UIF), collecting some 0.4% of national income in 2018. UIF contributors are well identified in labor force surveys, so we directly impute contributions based on statutory rules.

Unemployment Insurance Benefits. Unemployment insurance benefits are only available to adults having previously made monthly contributions to the UIF. This explains why they only cover a small fraction of the population (1.9% in 2018) and represent only 0.4% of national income. Unemployment benefits and beneficiaries are recorded in income surveys but are typically underrepresented. As in the case of private pension income, we therefore impute UIF benefits to additional recipients and we proportionally rescale the value of these benefits, so as to perfectly match both the official number of recipients and total UIF expenditure recorded in administrative data sources.

Pension and Unemployment Deficits or Surpluses. To ensure that pretax national income equals factor national income, we have to distribute the surpluses or deficits of the pension and unemployment systems. Following other DINA studies, we distribute 50% of the gap between contributions and benefits to contributors proportionally to contributions paid, and 50% to recipients proportionally to benefits received. This corresponds to assuming that the burden of the deficit (or the benefits of the surplus) will eventually be shared 50/50 by contributors and recipients.

4.24. From Pretax National Income to Posttax National Income: Taxes

To move from pretax income to posttax income, we start by deducting all taxes paid (see Table 2). These include all direct taxes (including the personal income tax and the corporate income tax) and all indirect taxes (including the Value Added Tax and excise duties).

Personal Income Tax. The personal income tax (PIT) is the tax collecting the highest share of government revenue in South Africa, amounting to 11% of national income in 2018. We microsimulate the income tax at the individual level, for each year since 1993, exploiting information on statutory rules, thresholds and marginal tax rates collected from historical administrative sources. As our microfile is calibrated on tabulated income tax returns, it is perfectly representative of taxable incomes at the top. It is therefore fully consistent with administrative data, both in terms of the number of taxpayers and total income tax receipts.

Corporate Income Tax. The corporate income tax (CIT) is the second biggest direct tax on income in South Africa (6% of national income in 2018). The CIT is paid on corporations' profits, so we distribute it similarly to retained earnings, that is, proportionally to directly and indirectly held shares.

Other Direct Taxes on Income and Wealth. Other direct taxes on individual income and wealth only represent a small fraction of national income (1.7% in 2018). We distribute the dividends tax (0.8% of NNI), a flat tax of 20% paid by individuals on dividends received from South African companies, proportionally to dividends received. The Skills Development Levy (0.4% of NNI) is a flat tax of 1% paid on the wages of employees registered with the UIF, so we impute it directly based on rules. We allocate the remaining direct taxes to their corresponding tax bases: transfer duties to housing wealth (0.2% of national income), the securities transfer tax to equity ownership (0.1%), the estate duty and the donations tax to net wealth (0.1%),

and the remaining taxes on income to pretax income (0.1%).

Value Added Tax. The value added tax (VAT) is the largest indirect tax in South Africa, enforced at a standard rate of 15% and collecting 8% of national income in 2018. In line with DINA studies and with standard tax incidence analyses, we assume that the VAT is paid by consumers. However, we refine our VAT tax incidence model in two ways. First, we exclude 19 “basic food goods”, which are zero-rated and therefore not subject to VAT, as well as all other VAT-exempt goods and services (including housing rents, transport services, petrol products, educational expenditure, and financial services: see [South African Reserve Bank, 2019](#)). Household expenditure on each of these items has been recorded in all income surveys, so we can directly remove them from our consumption aggregate. Secondly, following [Bachas, Gadenne, and Jensen \(2020a\)](#), we exclude goods and services bought on the informal market, approximated by the type of store in which purchases occur. These two steps significantly mitigate the regressive impact of VAT, although not sufficiently to make it progressive, given the particularly high gap between consumption and income at the bottom of the distribution and the small size of the informal sector in South Africa.

General Fuel Levy and Excise Duties. Other indirect taxes on domestic products include the general fuel levy (1.8% of NNI), other excise duties (1.1%), and other taxes on goods and services (0.3%). The general fuel levy is a tax on fuel consumption, so we distribute it proportionally to fuel and transport expenditure. Other excise duties correspond to taxes on tobacco and alcohol, paid at production, so we distribute them proportionally to spending on these two goods. Other taxes on goods and services include a number of other minor indirect taxes, which we distribute proportionally to overall household expenditure.

Taxes on International Trade. Import duties and other taxes on international trade together represent about 1.4% of national income. A simplified way to distribute these taxes would be to assume that they are borne by consumers as VAT. However, the nature of imported goods might differ significantly from that bought by a typical consumer, leading to biased estimates. To correct for heterogeneity in consumption of domestic vs. imported goods, we use input-output tables published by Statistics South Africa to derive an estimate of import density by COICOP category of household expenditure. We then distribute taxes on international trade proportionally to import-density-corrected consumption.

Local Taxes Local government revenue in South Africa consists mainly in property rates, service charges for the provision of electricity, water, and other services such as refuse removal, and transfers received from the central government. Since the latter are financed by central government revenue, we do not allocate them to individuals (doing so would lead to double counting, as transfers to municipalities are indirectly financed by national taxes). Property rates, electricity charges, and water charges are directly reported by households in income surveys, so we allocate budget totals proportionally to these reported values. We distribute the remaining components of municipal operating revenue proportionally to the total municipal tax burden of each individual, so as to match total revenue reported in municipal budgets.

Other Tax and Non-Tax Revenue. To reach total consolidated government revenue, we distribute the remaining tax and non-tax revenue proportionally to pretax income (i.e., in a distributionally neutral way). These include all other taxes not previously mentioned (less than 0.1% of national income), payments to the Southern African Customs Union (-1.2%), non-tax revenue (0.8%), and revenue collected by provinces and other public entities (2.4%).

4.25. From Pretax National Income to Posttax National Income: Transfers

Having removed all taxes from pretax income, we now allocate all government expenditure—including both direct and in-kind transfers, as well as the government deficit—to individual incomes to reach posttax national income.

Direct Social Transfers. Social protection spending represents about 5% of the national income in South Africa, the majority of which consists in three social grants: the old age grant (1.8%), the child support grant (1.5%), and the disability grant (0.6%). The old age grant is a means-tested monthly benefit available to South Africans older than 60. The child support grant is granted to a child's primary caregiver whose income falls below a specific threshold. The disability grant is provided to workers suffering from a permanent disability. As in the case of pension and unemployment benefits, data on social grants is available in income surveys, but the number of self-reported recipients tends to be lower than in administrative data (although only slightly). For consistency, we attribute social grants to additional eligible beneficiaries using a linear probability model, and we impute the value of grants received based on statutory rules, each year since 1993 (see appendix 4.8). This ensures that our microfile is fully consistent with both the number of grant beneficiaries and total government expenditure on grants.

Other In-Kind Transfers and Public Goods. We take the distribution of other government expenditure, including education, healthcare, transport infrastructure, police services, and other public goods from [Gethin \(2023\)](#). This related paper combines various surveys and historical budget data to identify beneficiaries of public services and the corresponding distribution of public spending by function of government from 1993 to 2019.

Government Deficit. As in the case of the deficits of pension or unemployment systems, we assume that 50% of the general government deficit (6% of NNI in 2019) is borne by taxpayers proportionally to total taxes paid, and 50% proportionally to total transfers received.

4.26. Household Expenditure and Household Wealth

In addition to factor income, pretax income, and posttax income, we also distribute consumption and wealth concepts consistent with national accounts definitions.

Household Expenditure. Following our approach for income, we distribute household final consumption expenditure by proportionally rescaling subcomponents of consumption reported in income surveys to their corresponding totals recorded in the national accounts, for each of the 12 COICOP categories available in both micro and macro data. This allows us to document the joint dynamics of consumption and income at the individual level, as well as to derive estimates of net saving (net household disposable income minus HFCE) by income group that are consistent with macroeconomic figures.

Household Wealth. Finally, we combine survey data on income and wealth with households balance sheets statistics published by the South African Reserve Bank to add an estimate of household net worth and its composition to our microfile since 1993. We do so by applying a “mixed method” combining rescaling and income capitalization (following [Saez and Zucman, 2016a](#)), whereby specific household wealth components from the balance sheets are distributed proportionally either to the corresponding income flows they generate, or to the market values of assets or liabilities reported by survey respondents. For more information on this methodology, we refer to our companion paper dedicated to the estimation of wealth inequality in South Africa ([Chatterjee, Czajka, and Gethin, 2022](#)).

4.3. The Distribution of Factor Income

In this section, we present results on the distribution of factor national income, that is, income arising from the use of production factors (capital and labor) before any form of government redistribution. This analysis serves as a basis for understanding the evolution and structure of income inequality in South Africa, which play a key role in determining the allocation of taxes and transfers. Section 4.31 provides background information on macroeconomic growth in South Africa. Section 4.32 describes on the dynamics of factor income inequality. Section 4.33 decomposes the distribution of factor income into its labor and capital components. For all results the statistical unit is the equivalized household per capita.

4.31. National Income Growth in South Africa since 1993

Figure 4.71 plots the evolution and composition of real national income per capita in South Africa since 1993. Average national income was equal to \$12,800 (80,000 Rand) at purchasing power parity in 2019, up by 37% from its 1993 level. Macroeconomic growth can be decomposed into three main phases: a phase of economic stagnation between 1993 and 2000, during which real national income remained relatively stable; a phase of fast growth between 2000 and 2011; and a phase of decline since 2012, characterized by a negative average annual growth rate.

Following the income approach to national accounting, the national income can be decomposed into its different income components. These components can be grouped into four main aggregates: compensation of employees (57% of NNI in 2019), mixed income and imputed rents (14%), household property income and corporate undistributed profits (17%), and government primary income (12%). Gross wages have followed a U-shaped curve over the period, dropping from about 60% of the national income to 50% in 2006, before bouncing back since then. The share of mixed income and imputed rents in national income has fluctuated with no clear trend. Conversely to wages, property income received by household and corporate undistributed profits have followed an inverted U-shaped curve, growing from 20% of NNI to about 25% from 1993 to the mid-2000s, before falling back to 17% in 2019.

4.32. The Distribution of Factor National Income

We now turn to documenting changes in the distribution of national income, before the operation of the pension, unemployment insurance, and tax-and-transfer systems. Table 4.83 reports data on the distribution of factor national income in 2019 across selected income groups, revealing

extreme income disparities. About one third of total income accrued to the poorest 90% of the population in 2019, compared to two-thirds for the richest 10% and over 28% for the top 1% alone. The top 0.01% of the population (5,860 individuals) received about 2% of factor national income, almost as much as the poorest 50% as a whole (29 million individuals). The bottom 50% have an average income of \$600 per year at purchasing power parity, which is 20 times lower than the national income per adult. Meanwhile, the top 10% received \$80,000 (7 times the national average) and the top 0.1% almost 1 million PPP dollars (80 times the national average).

In Figure 4.72a, we represent the evolution of income inequality in South Africa over time and compare it to that observed in other countries for which comparable distributional national accounts studies have been conducted. South Africa stands at the upper frontier of global income inequality today: the share of income accruing to the top 10% exceeded 65% in 2019, compared to 55-60% in Brazil and India, 40-45% in China and the United States, and below 35% in France. Furthermore, the top 10% share has increased significantly since 1993, moving up by almost ten percentage points between 1993 and the early 2000s, before stabilizing thereafter. The 2007-2008 crisis has been associated with a slight drop in top income inequality, as observed for instance in a number of European countries (see [Blanchet, Chancel, and Gethin, 2022](#)).

Figure 4.72b plots the cumulative income growth of the top 1%, the top 10%, the middle 40%, and the bottom 50%. A striking divergence in real factor incomes has taken place between the bottom and the top of the distribution since the end of apartheid. Between 1993 and 2019, the average national income grew by 37%, yet it rose by almost 50% for the top decile and by over 70% for the richest percentile. Meanwhile, the average factor income of the middle 40% grew by 20% and that of the bottom 50% almost stagnated. Coming back to the three phases of national income growth outlined above, we see that the stagnant decade of the 1990s was associated with dramatically different trajectories across income groups, as the boom of top incomes was almost perfectly compensated by income losses among the bottom 90%. Economic growth in the early 2000s benefited both the top and the bottom of the distribution, albeit significantly more the former than the latter. Finally, the drop in real incomes after 2011 was mostly driven by the top of the distribution, but this decline was insufficient to bring back inequality to its 1993 level.

4.33. Decomposing Factor Income Inequality: Labor versus Capital

To shed light on some of the factors behind the rise of income inequality, one can start by decomposing income into its labor and capital components. This decomposition also directly informs the tax incidence analysis conducted in section 4.4, given that the distribution of taxes is highly dependent on the distribution of the various income components on which taxes fall.

Figure 4.73a represents the evolution of the top 1% capital income, labor income, and total factor income shares since 1993. Three results stand out. First, in line with what we observe in the majority of countries with available data (e.g., [Garbinti, Goupille-Lebret, and Piketty, 2018b](#); [Piketty, Saez, and Zucman, 2018](#)), capital income inequality has remained substantially higher than labor income inequality throughout the entire period. Second, the boom of top 1% incomes of the 1990s and 2000s was driven by both labor and capital income: the top 1% capital income share grew from 55% to 60-65% from 1993 to 2007 and the top 1% labor income share from 15% to 25%. Third, the stabilization of top incomes after 2007 masks a divergence between a continued increase in labor income concentration and a decline in top capital income inequality. This decline partly mirrors dynamics at the macro level, in particular the fall of household property income and the growing share of wages in the national income since the 2007-2008 crisis that we documented in section 4.31.

As shown in 4.73b, a direct consequence of these differential dynamics has been a significant increase in the labor share of income at the top. In 1993, investment income (interest, dividends, and rental income) and undistributed profits represented 60% of factor income of the top 1%, while labor income amounted to less than 30%. The share of wages in top 1% incomes has grown to nearly 50% in 2019, due in large part to the shrinking size of investment income since the early 2010s.

4.4. The Distribution of Taxes and Transfers

To what extent have taxes and transfers curbed the rise of inequality in South Africa? To answer this question, we now analyze the distribution of taxes and transfers and its impact on inequality and the real incomes of income groups throughout the distribution since 1993. Section 4.41 discusses the impact of the pension and unemployment system. Sections 4.42 and 4.43 respectively analyze the distributional incidences of taxes and transfers. Section 4.44 documents how the tax-and-transfer system as a whole has shaped the distribution of macroeconomic growth since the end of apartheid.

4.41. From Factor to Pretax Income

In most advanced economies, the pension and unemployment insurance systems redistribute a substantial share of the national income the elderly and the unemployed every year, leading to very large reductions in inequality when moving from factor income to pretax income (Blanchet, Chancel, and Gethin, 2022). This is not the case in South Africa, where the private pension system and unemployment insurance through the Unemployment Insurance Fund (UIF) only benefit to a small fraction of the population.

In 2019, about 20% of the population contributed a total of about 6% of the net national income to private pension and provident funds. Meanwhile, on the income side, about 6% of adults received private pension income (about 3% of NNI). The size of the unemployment insurance system was even smaller: in 2019, about 25% of adults contributed a total of 0.5% of NNI to the fund, enabling some 2% of the population to benefit from unemployment benefits. This share stood in sharp contrast with the 29% unemployment rate, a gap that can be explained in large part by the relative stringent conditions required to benefit from unemployment insurance in South Africa (Bhorat, Goga, and Tseng, 2013).

Figure 4.74 represents the net transfers operated by the pension and UIF systems between income deciles, expressed as a share of NNI, in 2019. The private pension system appears to mostly redistribute small fractions of the national income from the top decile to the middle class. The top 1% are the main contributors to the system, losing 0.5% of NNI more in contributions than they receive in benefits, while the ninth decile (p80p90) receive a net transfer of about 0.2% of NNI. Redistribution operated through the UIF system is even smaller, with no group receiving more than 0.1% of NNI in unemployment benefits. Furthermore, while it is overall progressive, it is notably regressive at the top end of the distribution, with the top 1% contributing less than 0.02% of NNI in unemployment contributions. This can be explained both by the relatively lower share of labor income among top incomes and by the maximum cap set on UIF contributions (R178,464, or about \$28,500 at PPP, in 2019), which effectively turn the contribution into a regressive tax at high wage levels.

In summary, moving from factor national income to pretax national income has almost no impact on inequality at all in South Africa (see also figure 4.710a below), given that the sums transferred are very small and mostly redistribute income from top to middle income groups.

4.42. The Distribution of Taxes

We now consider the distributional incidence of all taxes collected in South Africa throughout our period of interest. Figure 4.75 plots the level and composition of general government revenue, expressed as a share of national income, from 1993 to 2019. Total revenue has grown significantly in the past quarter of century, from 35% of national income in 1993 to 43% in 2019. This represents a 23% increase in the share of national income extracted from economic output every year by the general government. Of the three most important taxes in South Africa—the Personal Income Tax (PIT), the Corporate Income Tax (CIT), and the Value-Added Tax (VAT)—the CIT is the one whose revenue has grown most rapidly in relative terms, from 3.6% of NNI in 1993 to 5.8% in 2019, followed by VAT (6.3% to 7.8%) and finally by the PIT (9.7% to 11.6%). If one groups taxes in South Africa into three broad categories, direct taxes (including the PIT, the CIT, and other taxes on income and wealth), indirect taxes (including VAT, other taxes on goods and services, and taxes on international trade), and other government revenue (including other taxes, non-tax revenue, and local government revenue), direct taxes appear to represent the largest and most rapidly growing component of government revenue. Direct taxes rose from 14% to 19% of NNI between 1993 and 2019, while indirect taxes expanded from 11% to 13% and other revenue from 10% to 12%.

How have these changes in the magnitude and structure of taxation affected the distribution of taxes paid by income group? Figures 4.76a and 4.76b provides a first answer to this question by decomposing total taxes paid by pretax national income group in 1993 and 2019. In 1993, the profile of taxation was relatively flat, except for the upper-middle of the income distribution, where effective taxation was slightly higher. Nearly all deciles transferred between 20% and 35% of their pretax incomes in taxes to the government. Bottom income groups paid almost all of their taxes in indirect and local taxes. Meanwhile, the personal income tax and the corporate income tax represented the bulk of the tax burden of the top decile. It is also interesting to note that the personal income tax was regressive at the very top, which is directly due to the fact that top income groups relied heavily on non-taxable capital incomes, in particular corporate undistributed profits held through stock ownership. The corporate income tax did not compensate sufficiently for this regressive aspect of the tax system, leading the top 0.1% to pay lower taxes than the rest of top 10% earners.

Moving to 2019, we see that the increase in taxation has been almost entirely concentrated at two parts of the distribution: the very bottom and the very top. At the top, taxation is no longer regressive, mostly due to the rise in the share of corporate income tax paid by the top 0.1% (from about 9% of its pretax income in 1993 to 19% in 2019). At the bottom, the share of income paid by low-income groups in indirect and local taxes has grown substantially, with the

tax burden of the third decile more than doubling. This can be explained both by the increase in total revenue collected from indirect and local taxes and by the rising gap between income and consumption among the poor in the past decades, to which we come back below. Meanwhile, the effective tax rate faced by middle income groups has barely changed, with individuals located between the median and the 90th percentile still paying less than 35% of their pretax income in taxes.

Figures 4.77a and 4.77b provide another perspective on this transformation by representing the yearly evolution of total taxes paid by top 1% and bottom 50% pretax income earners since 1993. In 1993, the top 1% faced a slightly higher effective tax rate than the bottom 50%. By 2019, the tax burden of the top 1% had increased to 45%, while that of the bottom 50% had surged to 60% of their total pretax income. The increase in top income taxation has been driven by the corporate income tax (from 9% to 13% of pretax income), but also by the personal income tax (from 9% to 16%). This latter evolution reflects both the fact that top taxable incomes have grown faster than the threshold required to enter top marginal income tax rates and the declining share of non-taxable capital income (dividends and undistributed profits) in top 1% pretax incomes. On the contrary, we see that the bottom 50% pay almost no personal income tax or corporate income tax at all, while local taxes, VAT, and excise duties have driven nearly all of the increase in their tax burden.

At this stage, let us discuss a bit further our results on the very high tax rates faced by bottom pretax income groups. It might look surprising and even unrealistic at first sight to observe such extremely high effective tax rates, given in particular that some of these rates are higher than the statutory rates of the taxes considered (for instance, the bottom 50% pay 20% of their pretax income in VAT while the statutory VAT rate is 15%). This is a mechanical result of our allocation strategy, which implies distributing indirect taxes proportionally to consumption (excluding exempted goods and the informal sector). Given that low-income groups have consumption levels that can greatly exceed their pretax incomes, the tax base on which these taxes are applied (consumption) may be substantially higher than the denominator considered for tax incidence analysis (pretax income). The presence of such a large discrepancy between the consumption and income distribution profiles, leading to extreme negative (respectively positive) savings at the bottom (respectively top) is not new (see [Chancel et al., 2023](#); [Czajka, 2017](#); [Deaton, 1997](#)), yet it is not fully understood.

If a large fraction of the poor are effectively consuming from their savings or from consumer debt, such tax rates may then not seem extraordinary. On the one hand, one cannot exclude that some measurement issues in household surveys (underreporting of income at the bottom of the distribution, overreporting of consumption at the bottom, or alternatively underreporting

of consumption at the top) may lead to biased estimates of savings across income groups, implying an overestimation of the regressivity of indirect taxes. On the other hand, there is suggestive evidence of strongly negative and deteriorating savings rates among the poor in South Africa. According to national accounts published by the South Africa Reserve Bank, the ratio of households' saving to their disposable income has remained systematically negative since the mid-2000s, fluctuating between 0 and -2% after a sharp decline in the 1990s, so that households have, in aggregate, consumed more goods and services than their disposable income allows alone. In 2019, as much as 5.7% of the entire national income (or 8% of household disposable income) was absorbed in interest repayments by households on previously contracted loans (authors' computations using national accounts data). [Chatterjee, Czajka, and Gethin \(2022\)](#), combining microdata on income, assets, and debts with macrodata on households' balance sheets, estimate that the total net worth of the poorest 50% is negative, that is, the total market value of the assets they own is lower than the debts they owe. This is consistent with data from the 2008 Living Conditions Survey, in which 72% of adults, and an overwhelming share of respondents at the bottom of the income distribution, declared having “no regular savings for emergencies.”

4.43. The Distribution of Transfers

We now analyze how government expenditure has been distributed since 1993. We focus on the main stylized facts; an extended analysis of changes in the size and progressivity of government transfers, with a particular focus on in-kind transfers, can be found in [Gethin \(2023\)](#). As shown in Figure 4.78, the rise of public spending has mirrored that of revenue in the past decades: total consolidated government expenditure grew from 36% to 42% of NNI between 1993 and 2019. Even more so than in the case of taxes, this transformation has been accompanied by significant changes in the nature of government intervention. General public services and defense are the two only types of spending that have declined as a share of NNI, from 8.7% to 7.0% and from 2.6% to 1.2% respectively. Meanwhile, spending on social protection is the item that has grown the fastest, nearly doubling from 3.3% to 6.5%, followed by health, local government expenditure, public order and safety, and education.

Figure 4.79a plots the share of total transfers in grants, education, healthcare, and other public goods received by the top 10% and bottom 50% as a share of national income. Consistently with the fact that the South African government has invested a rising share of NNI in individualized transfers that primarily benefit the poor, the share of national income transferred to the bottom 50% has grown much faster than that received by the top 10%. The bottom 50% received almost 19% of national income in the form of cash and in-kind transfers in 2019, representing

an increase of over 50% since 1993. Meanwhile, the share of national income redistributed to the top 10% has declined, from 14% of NNI in 1993 to 11% in 2019.

Figure 4.79b plots the cumulative growth rate of the bottom 50% before and after transfers. The rise of redistribution has generated substantial real income gains for low-income households. Bottom 50% average income growth is barely affected by the inclusion of the old age grant and the disability grant, mainly because these grants already existed in 1993 and have not increased significantly in real terms since then. In contrast, the introduction of the child support grant in 2002 and its progressive deployment over the course of the 2000s has strongly benefited the bottom 50%, whose total growth rate shifts from below 20% to over 50% when accounting for child support grants received. Finally, substantial increases in the size and progressivity of public services have further contributed to improvements in the living standards of low-income households (see [Gethin, 2023](#)). Accounting for in-kind transfers and public goods—including in-kind social protection, education, healthcare, local government services, and other public services—bring bottom 50% real income growth to almost 100%.

4.44. The Overall Impact of the Tax-and-Transfer System

Our analysis of taxes and transfers has shown mixed results. On one hand, in-kind transfers have grown substantially since 1993, and this rise has primarily benefitted bottom income groups. On the other hand, the bottom 50% have faced increasing effective tax rates, driven by the rise of indirect and local taxes. Combining these two pictures, who has benefited most from the rise of South Africa's welfare state since the end of apartheid?

Figure 4.710a compares the top 1% and bottom 50% shares in terms of factor, pretax, posttax disposable, and posttax national income since 1993. South Africa's tax-and-transfer system is progressive overall and has become significantly more progressive over time. Between 1993 and 2019, the top 1% factor income share grew from 22% to 28%, while the top 1% posttax national income share first rose but then came back to its 1993 level, at about 18%. This result directly mirrors the rising tax burden of the top 1%, which has not come with greater transfers.

Turning to the bottom 50%, redistribution appears to have increasingly benefited this group, due in particular to the rising role of in-kind transfers and public goods. In 2019, moving from pretax to posttax disposable income (that is, removing all taxes but only adding back cash transfers) increases the bottom 50% share from about 3% to 5%, while moving from posttax disposable income to posttax national income (that is, adding in-kind transfers and all other government expenditure) raises it to 15%. In terms of pretax and posttax disposable income, the bottom 50% share first dropped from 1993 to the mid-2000s, before coming back

to approximately the same level. In terms of posttax national income, in contrast, it declined from 11% in 1993 to 10% in 2005, but then rose steadily until reaching 15% in 2019. Rising redistribution in the form of education, healthcare, local government services, and other public goods has thus acted as a powerful equalizer since the end of apartheid, although inequality remains high even after accounting for taxes and transfers.

Figure 4.710b provides a more granular picture of redistribution in South Africa by representing the share of national income transferred by the tax-and-transfer system between income deciles in 1993 and 2019. Two results stand out. First, in 2019, all deciles within the bottom 80% were net beneficiaries, while the top 10% saw its pretax income reduced by a net total of 20% of national income. Redistribution in South Africa thus appears substantial, transferring about a fifth of the entire national income from the top decile to the rest of the population. Second, redistribution operated by the tax-and-transfer system has intensified over time. At the top, the net transfer of the top 10% grew by over 50%. Meanwhile, all deciles within the bottom 80% received significantly higher net transfers in 2019 than in 1993. The net transfer received by the bottom 50% grew from 10% to 15% of national income.

Having considered the impact of taxes and transfers on overall inequality, let us focus more specifically on the evolution of real incomes. Figure 4.711 provides a granular picture of the distribution of growth throughout the South Africa population by representing the cumulative evolution of real income by percentile between 1993 and 2019. The dramatic rise of pretax income inequality, combined with low macroeconomic growth, have implied drastically different trajectories at the top and bottom of the distribution. The top 1% has grown at the fastest pace, experiencing an almost 80% increase in average pretax income, compared to about 20% for most percentiles at the middle of the income distribution and strongly negative growth rates within the bottom 25%. The rise of redistribution, however, has more than compensated increases in pretax inequality. Removing all taxes and adding all cash and in-kind transfers from individual incomes reduces top 1% real income growth to 40%, while it raises median growth to about 45% and growth for the poorest 25% to over 100%. The rise of the South African welfare state has thus turned the distribution of economic growth since the end of apartheid from very regressive to unambiguously progressive.

In summary, our analysis of inequality and growth has revealed a striking surge in both pretax income inequality and government redistribution in South Africa since the end of the apartheid regime. This “chase between inequality and redistribution” has to some extent been won by the latter, as substantial improvements in tax progressivity at the top of the distribution and rising cash and in-kind transfers have made the final distribution of growth strongly progressive. This positive assessment should not be exaggerated, however. Even after taxes and transfers, income

inequality remains exceptionally high in comparative perspective, with the bottom 50% as a whole still receiving less income than the top 1% in 2019.

4.5. The Evolution of Racial and Spatial Inequality

Our new dataset does not only cover income, taxes and transfers, it also preserves all the richness of household surveys and thus allows us to decompose inequality and redistribution by a number of sociodemographic variables. In this section, we study the evolution of income concentration along two key dimensions of South African inequality: race and geography.

4.51. Racial Inequality

Race has always been at the heart of economic and political conflicts since the making of the South African state. Throughout the twentieth century, inequalities between racial groups stood at unparalleled levels. These inequalities were institutionalized through the political domination of the White minority, which culminated in the apartheid regime of strict racial segregation established in 1948. Between the early twentieth century and the late 1980s, the per capita income of African South Africans thus remained stable at a level reaching less than 10% of that of the White population ([Leibbrandt et al., 2010](#)). This represents some of the most extreme inequalities between racial or sociocultural groups observed in contemporary history. By comparison, the White-Black income gap has fluctuated between 50% and 60% in the United States between the 1950s and today ([Piketty, 2019](#)).

How have the end of apartheid and the transition to democracy in the mid-1990s, rising inequality, and enhanced redistribution reshaped South Africa's historical legacy of extreme racial disparities? To answer this question, we first provide a long-run view on racial inequality in figure 4.713a by representing the evolution of the share of White and Black South Africans in top income groups since 1955. The figure combines historical tabulated tax returns collected by [Alvaredo and Atkinson \(2022\)](#), census data (1970, 1980, 1990), and our distributional national accounts data after 1993. Under apartheid, Whites represented over 95% of top 1% earners and over 90% of the top 10%, while the share of Black South Africans in upper income groups was nearly zero. A remarkable transformation in the composition of top incomes has taken place since the early 1990s: the share of Africans in the top 10% jumped from 2% in 1980 to 15% in 1990-1994, and then increased monotonically until reaching about 45% in 2019. A similar evolution occurred within the top 1%, although racial inequalities continue to be higher in the top 1% than in the top 10%.

Figure 4.713b turns to the evolution of the overall White-to-Black income ratio, focusing on the role played by changes at the top of the distribution in the decline of racial inequality. As shown by the bottom line of the figure, White South Africans' average factor income was about 14 times higher than Black South Africans' in the early 1990s. This ratio remained stable until the 2001, then increased to 18 (+28%) in 2005 as the bottom 50% hit its lowest point (Figure 4.72b), to progressively decline to 8 (-44%) in 2015-2019. The picture looks very different if one excludes top Black earners from the analysis: excluding Black earners belonging to the top 1% leads to a decline in the gap from 14 to 11, while removing all those in the top 5% of the factor income distribution leads to an even smaller change, from 15 to 13. If one excludes completely all Black South Africans belonging to the top 10% from the analysis, then the White-Black income ratio appears to have remained constant, at about 17. These results imply that if the racial gap has decreased by 42% over the entire period, it is in part because of the emergence of a new Black elite, who has occupied a growing share of the top 10% of the income distribution.

Relatedly, figure 4.713c shows how factor income growth has been distributed within each population group from 1993 to 2019. Two results stand out. First, inequality has risen dramatically within each group as growth rates increase with income. The top 10% of Asian, Black, Coloured, and White earners saw their average factor income grow by 70 to 220%, compared to growth rates of approximately 0 to 40% for the bottom 50%. Second, the average factor income of Black South Africans grew substantially faster than that of other groups: it rose by 150% over the period considered, compared to 50-80% for Asians, Coloureds, and Whites. Much of this dynamic was driven by differential trajectories at the top of the distribution: the average factor income of top 10% Black earners increased much faster than that of the top 10% of other groups, while growth rates of the middle 40% and bottom 50% of each group are of the same order of magnitude.

Figure 4.714a provides more detail on the contemporary structure of racial inequality in South Africa by decomposing the White-Black gap by economic concept. Two results stand out. First, racial inequality remains substantially larger in terms of wealth than in terms of income or consumption: the White-Black income ratio reaches almost 14 in terms of personal net wealth versus 7-9 in terms of consumption, factor income, and pretax income. Second, the tax-and-transfer system strongly reduces racial inequalities, in particular in-kind transfers, yet posttax income gaps remain high. The ratio decreases from 9 to 8 when moving from pretax to posttax disposable income, and drops to 5 in terms of posttax national income. Taxes and transfers therefore significantly reduce racial inequalities, yet they remain high. As shown in figure 4.714b, which represents the racial composition of posttax national income groups in 2019, White earners continue to be massively overrepresented at the top end of the distribution

even after accounting for taxes and transfers. In 2019, they represented over 70% of the top 1% compared to less than 5% of all posttax income percentiles within the poorest half of the population.

4.52. Spatial Inequality

To conclude this paper, we consider another dimension of inequality: geography. How large are spatial inequalities in South Africa and how are they affected by the tax-and-transfer system?

Figure 4.715a compares the relative average incomes of South Africa's provinces before and after accounting for government taxes and transfers. Regional inequalities are significant in South Africa, and clearly separate the country into two groups: that of the richer provinces of Western Cape and Gauteng, whose average factor incomes exceed the average national income by 60-80%, and the rest of the country, with incomes falling between 40% and 60% of the national average. These regional disparities are larger, for instance, than inequalities between European countries, and substantially wider than differences in average incomes across US States (see [Blanchet, Chancel, and Gethin \(2022\)](#)). In line with our finding on the overall progressivity of the tax-and-transfer system, we find that the government also operates redistribution between provinces, although only to a moderate extent. Western Cape and Gauteng are net contributors, while all other provinces are net beneficiaries. The provinces that benefit most from the tax-and-transfer system are Limpopo, KwaZulu-Natal, and Eastern Cape, whose relative income increases by 30-50% after accounting for taxes and transfers. Meanwhile, Gauteng sees its relative average income decrease by over 15% between factor and posttax income.

In addition to regional inequality, the rural-urban income gap has been found to be significant in many countries throughout the world, often determining a substantial share of overall income inequality, migration patterns, and human capital accumulation ([Young, 2013](#)). South Africa is no exception to this general pattern, yet we find that rural-urban disparities have risen significantly since 1993. The average factor income of urban earners was almost 6 times higher than that of rural areas in 2015-2019, compared to about 4 in 1993-1994 (see figure 4.715b). However, the rise of government redistribution has prevented posttax inequality from increasing: the rural-urban gap grew from 3.6 to 4.8 in terms of posttax disposable income, while it stagnated at about 2.8 in terms of posttax national income. Growing spending on in-kind transfers and public services has thus disproportionately benefited rural areas since the end of apartheid, fully compensating the rise of rural-urban pretax income inequality.

4.6. Conclusion

By most contemporary measures, South Africa continues to stand out as the most unequal country in the world, yet this paper has documented dramatic changes in the structure of these inequalities since the end of the apartheid regime in the 1990s. The surge of pretax income inequality has implied radically different growth trajectories across income groups. The top 1% experienced an 80% increase in their average pretax income, while that of the bottom 20% declined. However, increasing government redistribution in the form of progressive taxation, cash grants, and public services has overcompensated the rise of pretax inequality, generating large real income gains for low-income households mostly at the expense of the richest decile. That being said, the expansion of South Africa's welfare state has been largely insufficient to significantly alter the extreme disparities inherited from a century of racial discrimination and oppression. The share of posttax income accruing to the richest 1% was about the same in 2019 as in 1993, while the bottom 50% only received 15% of national income, even after accounting for all cash transfers, in-kind transfers, and public goods received. While racial inequalities have declined, this decrease has been entirely driven by the income gains of a few Black earners at the top end of the distribution, thereby excluding the majority of the poor. These inequalities continue to be much larger in terms of wealth than in terms of income and have not been substantially affected by the growing progressivity of the tax-and-transfer system.

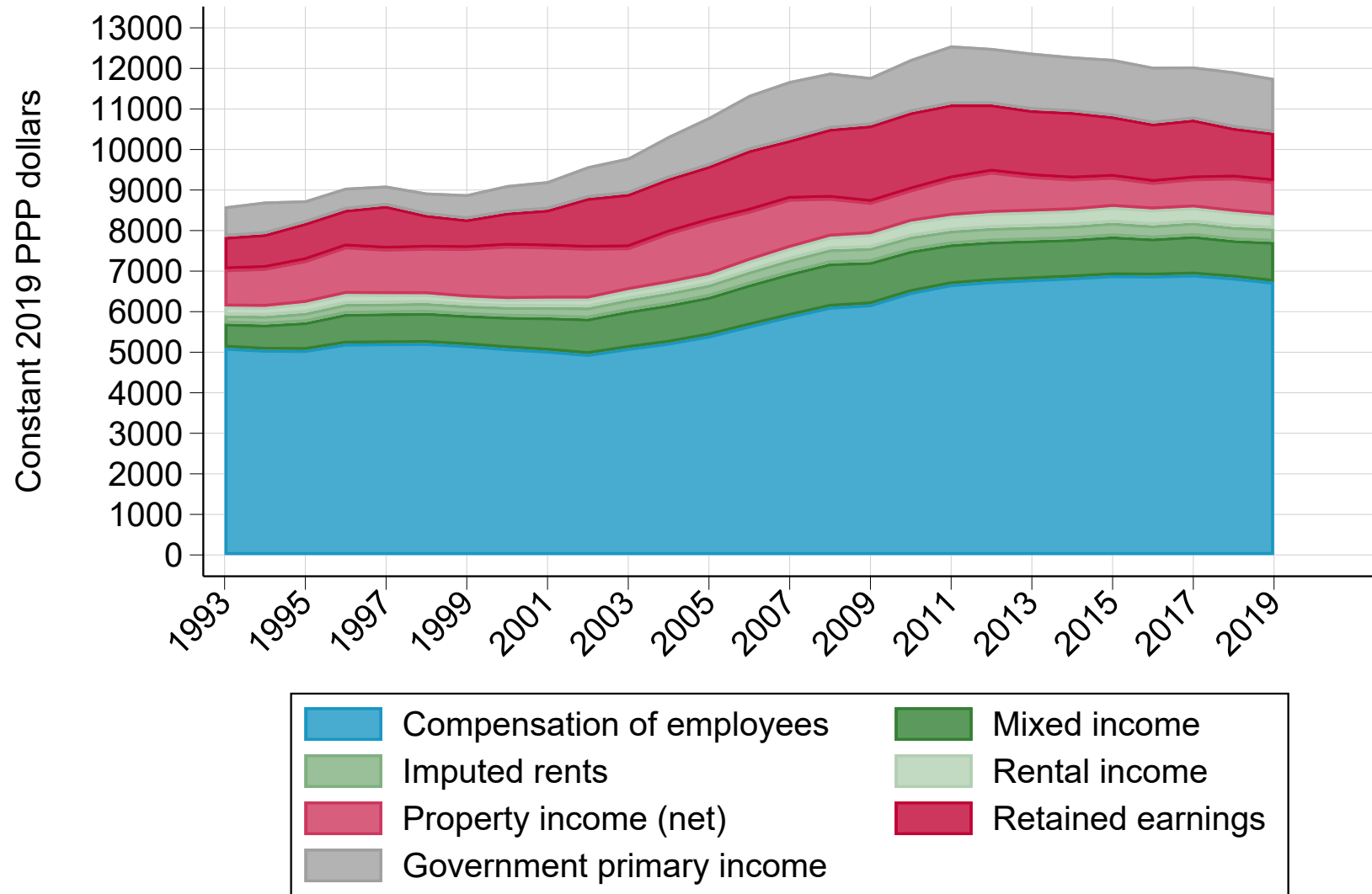
We see at least two avenues for future research. First, this paper has demonstrated the crucial importance of allocating indirect taxes and in-kind transfers when estimating the impact of taxes and transfers on poverty, inequality, and the distribution of economic growth. Yet, while we believe we have made significant advances in facing this challenge, the data sources at our disposal to properly understand who pays government taxes, and who gains from spending in health, education, and other collective expenditure remain largely unsatisfactory. Who benefits from investments in infrastructure development, industrial policy, or housing programs at the macro level, and how has this changed over time? What kinds of government spending most effectively accrue to low-income groups and how? These are important questions on which our knowledge remains all too limited.

Second, while our results have shed new light on the interactions between taxes, transfers, and the distribution of growth, much remains to be understood when it comes to the behavioral and general equilibrium mechanisms underlying the persistence of extreme economic inequalities and the ability of government redistribution to reduce these inequalities in the long run. To what extent can progressive taxation contribute to limiting income and wealth concentration beyond their immediate impact on top pretax incomes? Can cash and in-kind transfers truly

reduce poverty and inequality beyond the short-term relief they provide, especially in countries where the poor are highly leveraged and vulnerable to transitory income shocks as in South Africa? To what extent taxes and transfers shape future pretax incomes? Answering these questions requires going beyond the descriptive analysis conducted in this paper and modelling the joint relationships between income, wealth, savings, and household debt (for recent fruitful attempts, see for instance [Blanchet, 2023](#); [Mian, Straub, and Sufi, 2021](#)). We hope that our new database and the stylized facts presented in this paper will contribute to research in these multiple directions.

4.7. Figures

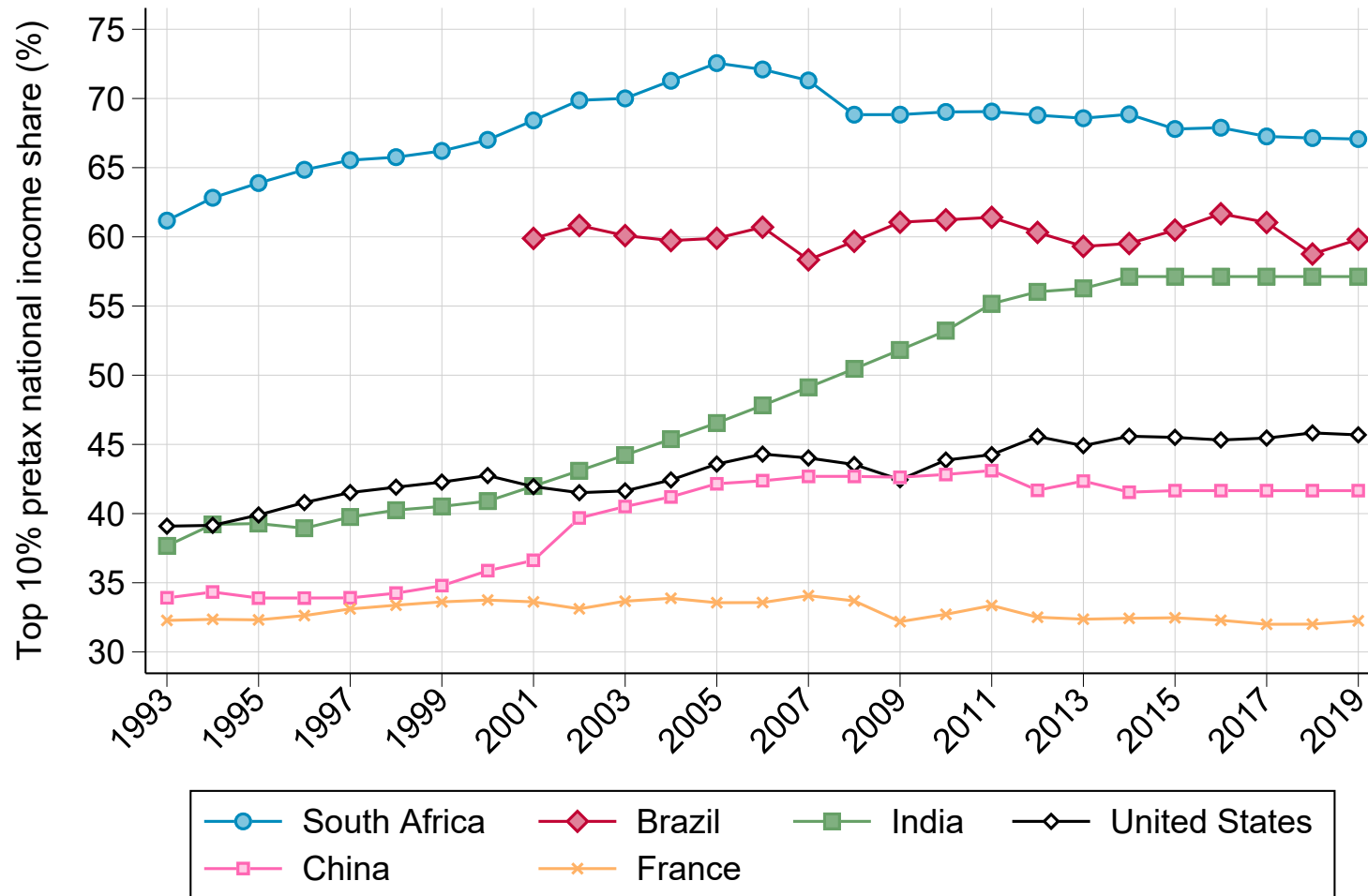
Figure 4.71 – Average national income per capita, 1993-2019



Notes. Authors' computations using national accounts series from the South African Reserve Bank Quarterly Bulletin.

Figure 4.72 – The distribution of factor national income, 1993-2019

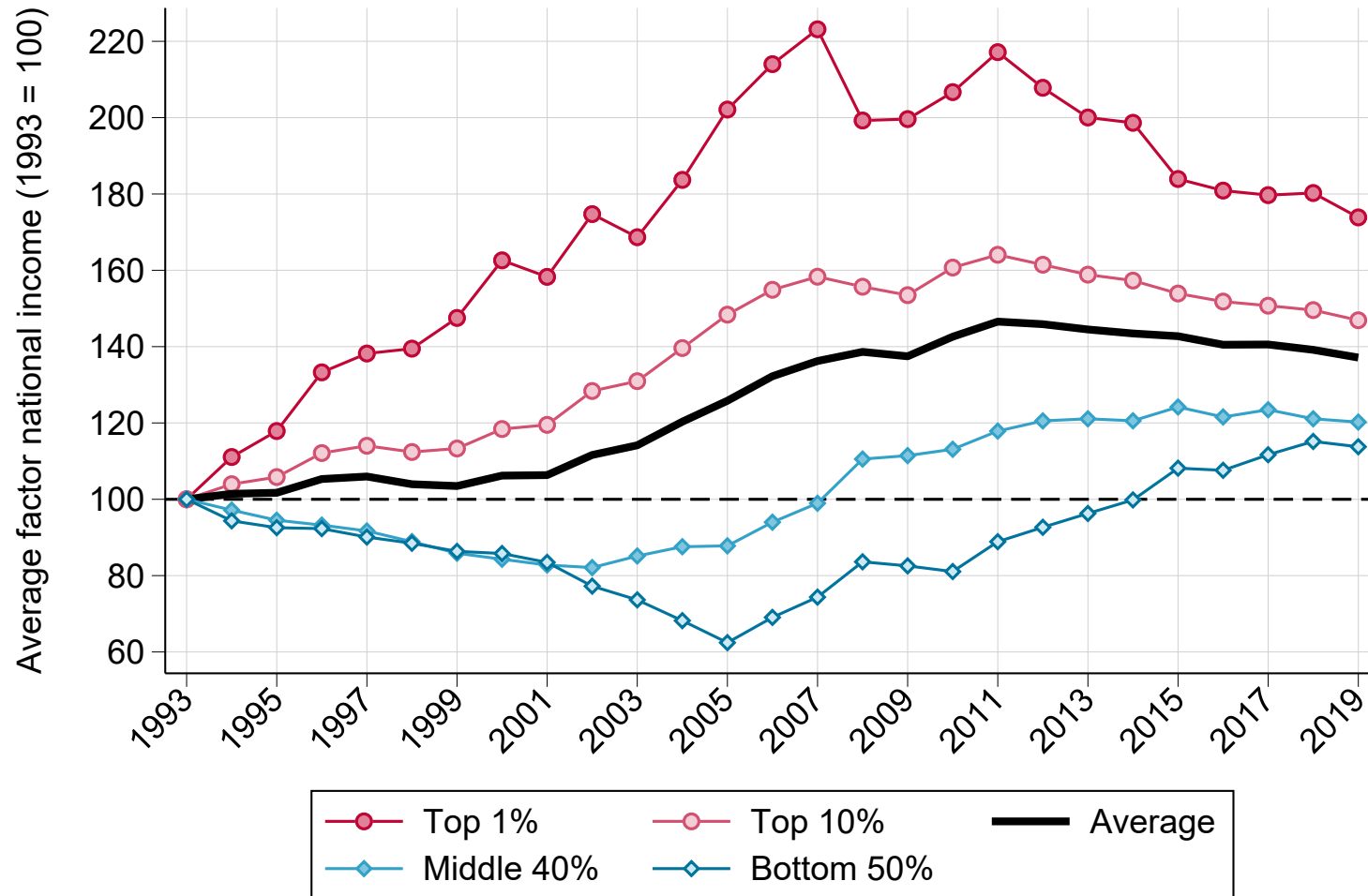
(a) South Africa in comparative perspective: top 10% pretax income share



Notes. Authors' computations combining survey, tax, and national accounts data (South Africa); World Inequality Database (other countries). Equivalised household factor income per adult.

Figure 4.72 – The distribution of factor national income, 1993-2019

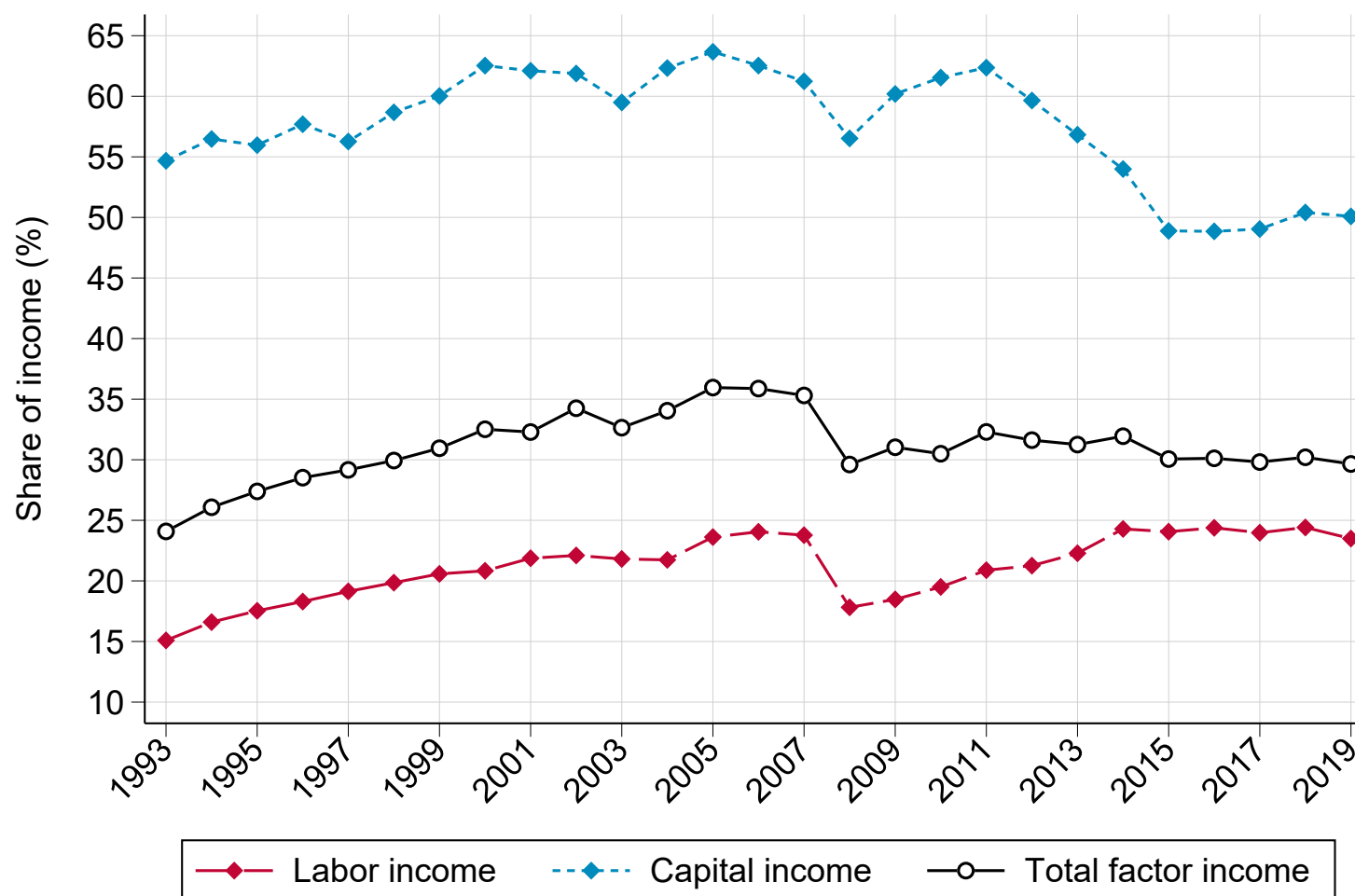
(b) Cumulated income growth by factor income group



Notes. Authors' computations combining survey, tax, and national accounts data. Equivalised household factor income per capita.

Figure 4.73 – Decomposing top factor income inequality

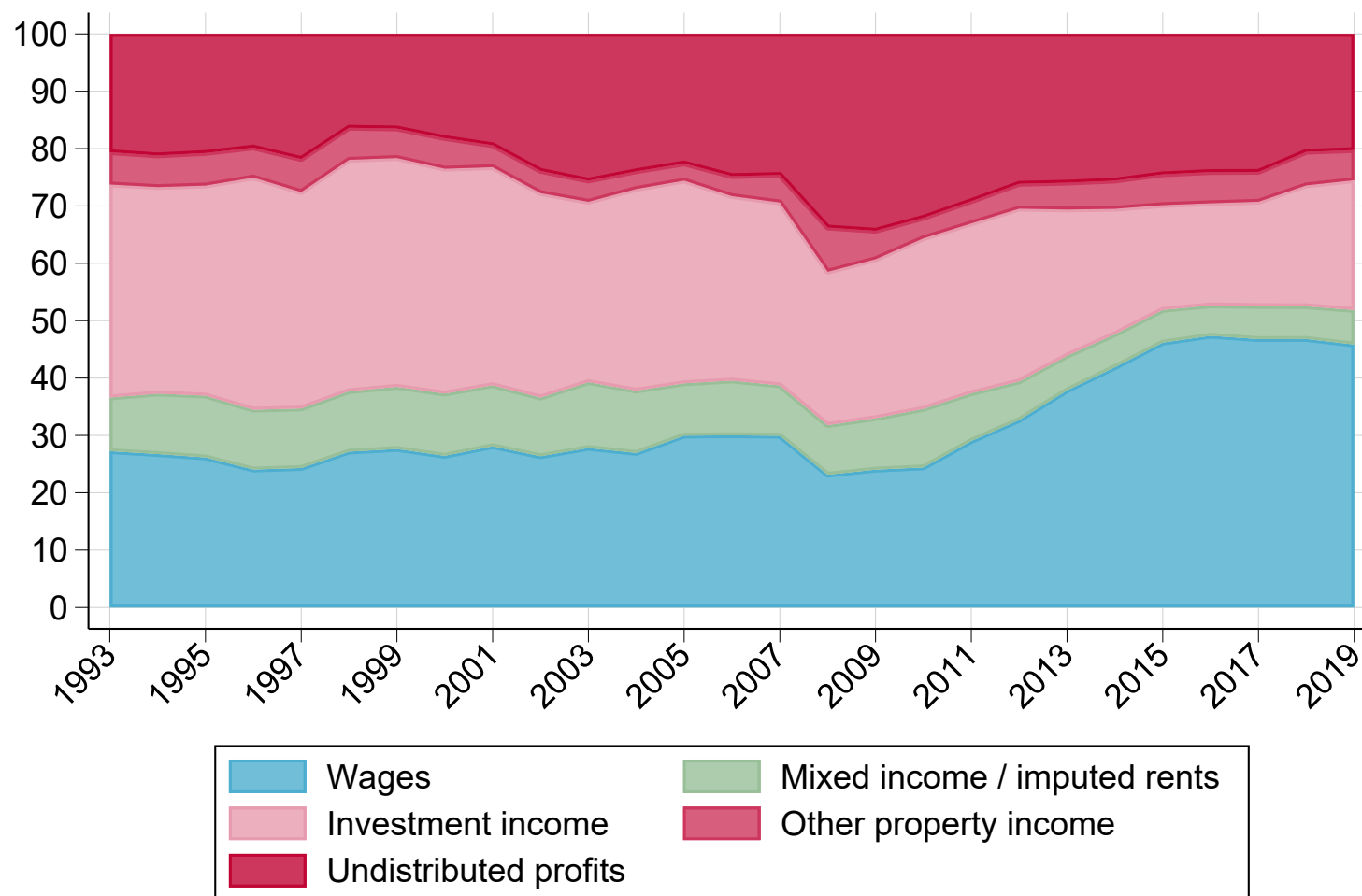
(a) Top 1% income share: labor versus capital



Notes. Authors' computations combining survey, tax, and national accounts data. Labor income is defined as the sum of compensation of employees and 70% of mixed income. Capital income is defined as the sum of 30% of mixed income, property income (rental income, interest, dividends, and other property income), and the private share of corporate undistributed profits. Equivalised household factor income per capita.

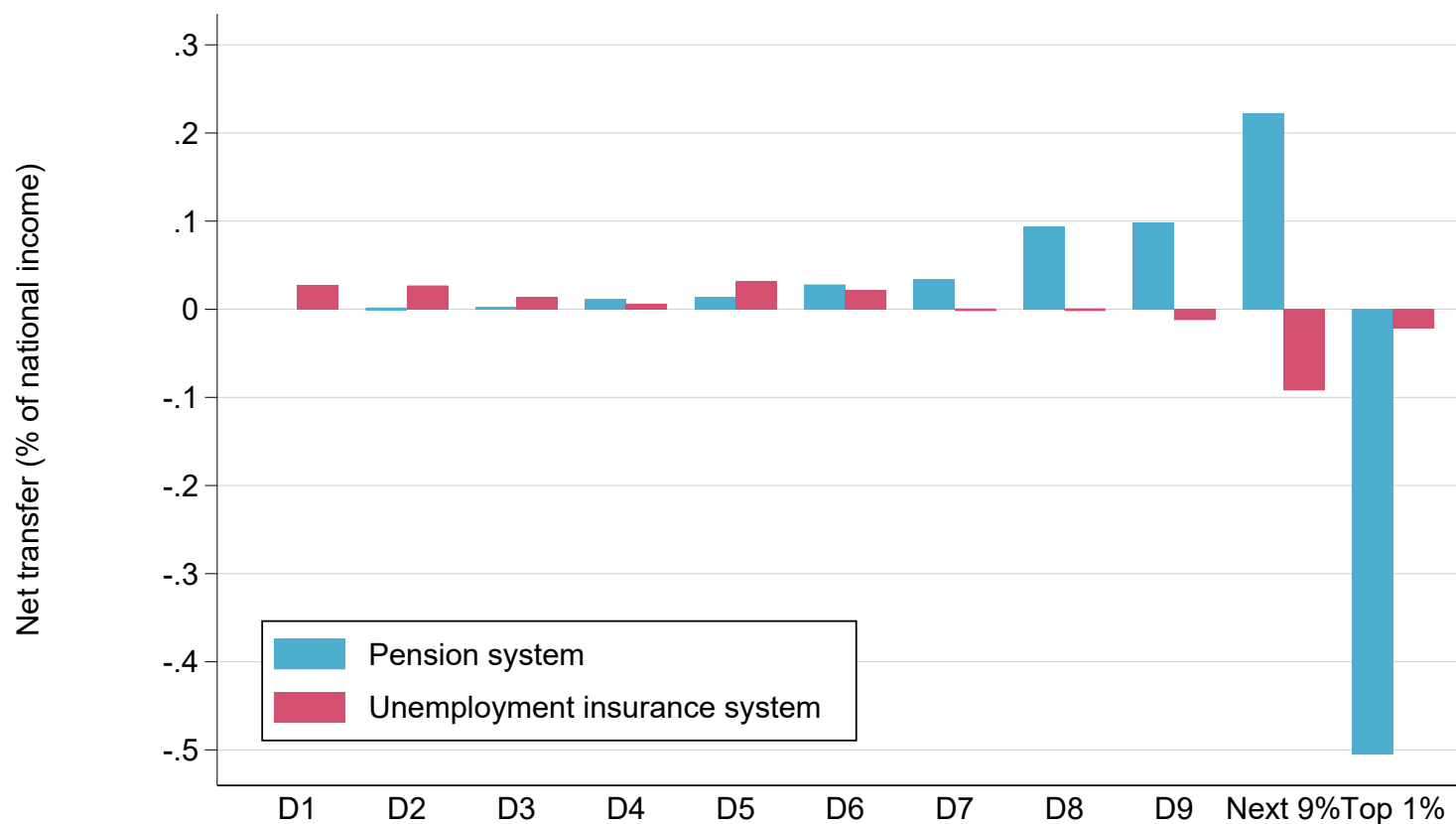
Figure 4.73 – Decomposing top factor income inequality

(b) Composition of top 1% factor income



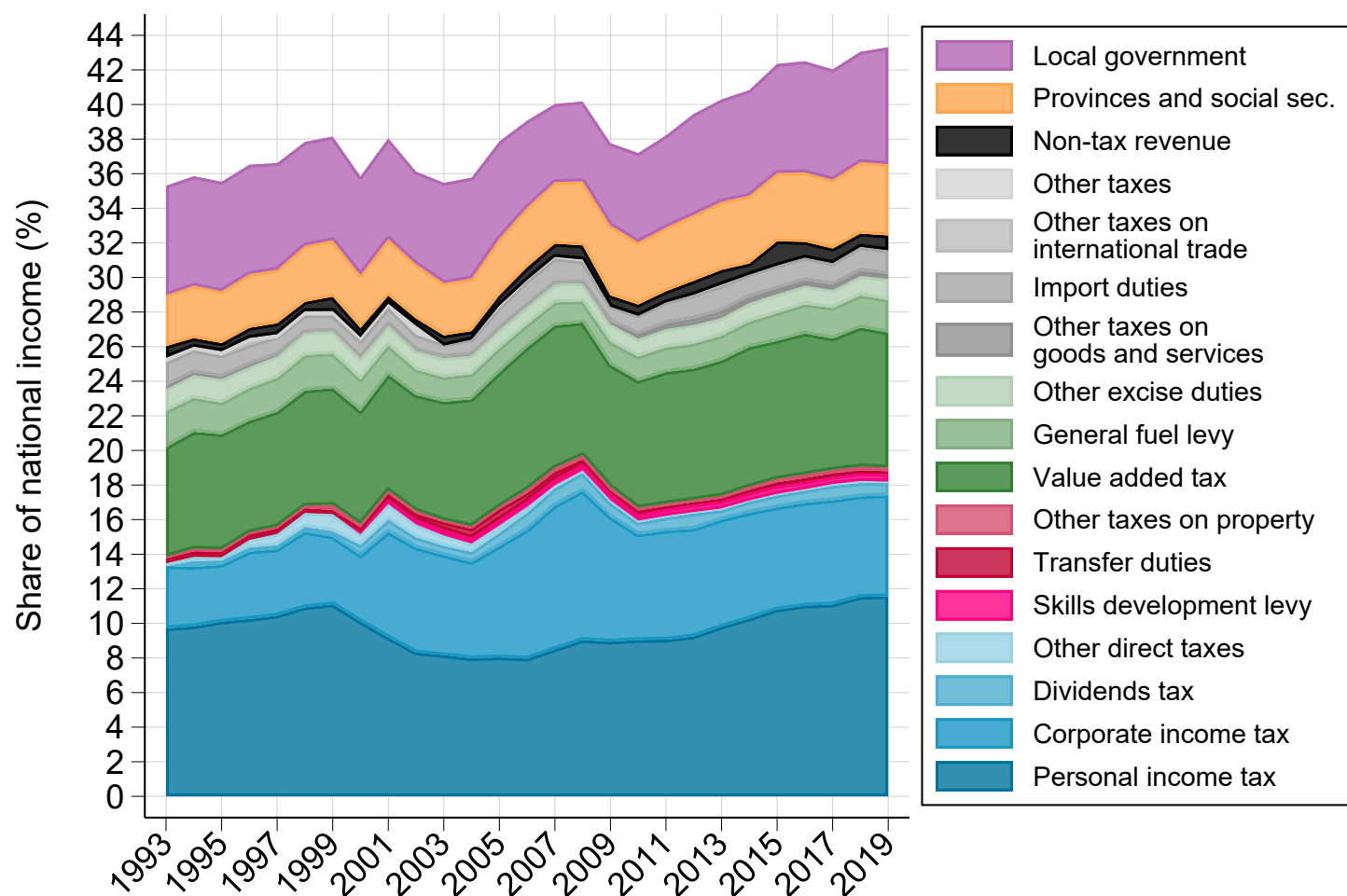
Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the composition of the factor income of top 1% earners. Equivalised household factor income per capita.

Figure 4.74 – From factor to pretax income: net transfers operated between factor income groups by the pension and unemployment insurance systems



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the net transfers received or paid by factor income group through the pension and unemployment insurance systems (that is, the difference between total benefits received and total contributions paid), expressed as a share of national income. Equivalised household income per capita.

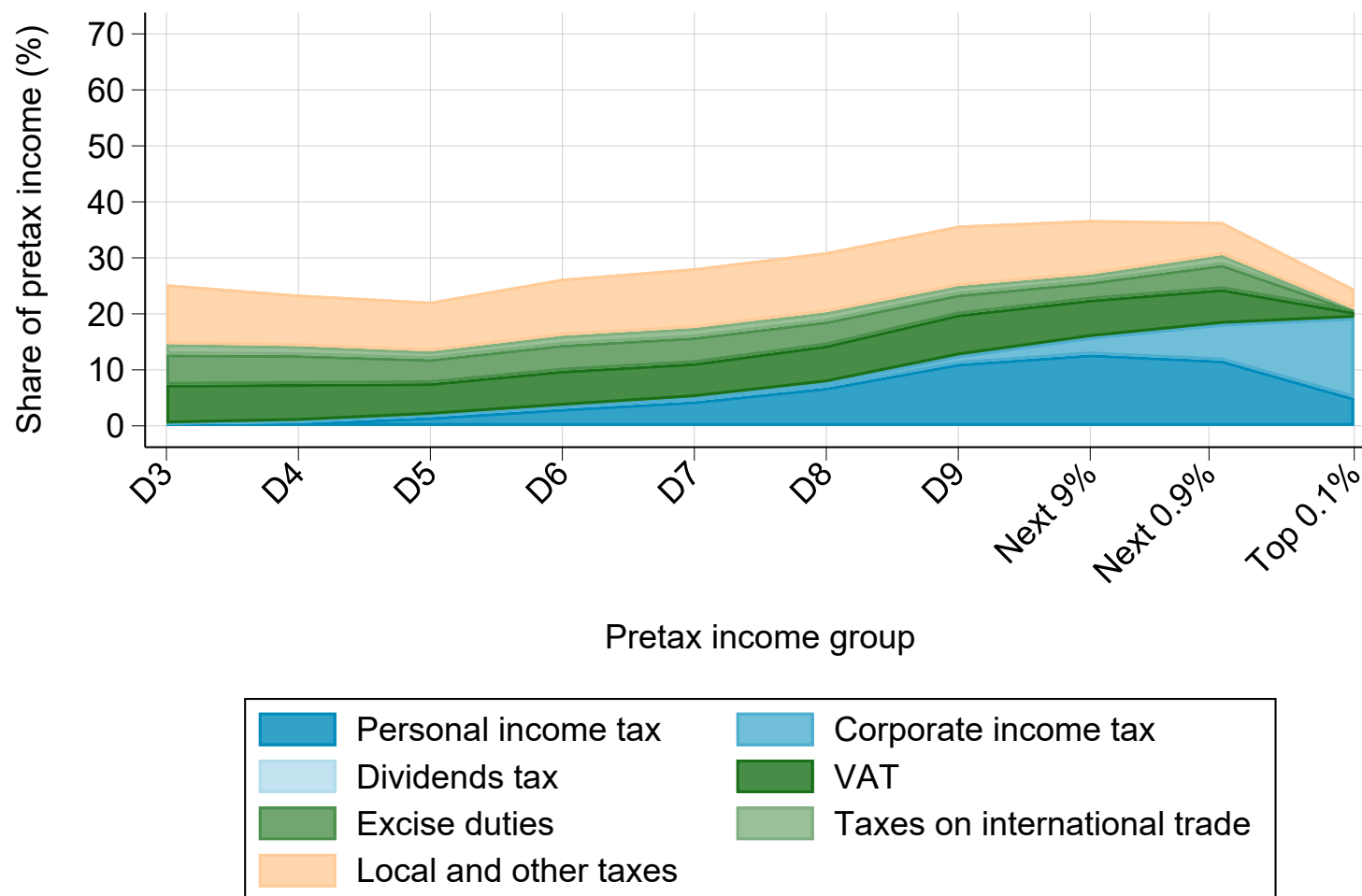
Figure 4.75 – Government revenue in South Africa



Notes. Authors' computations combining national accounts series from the South African Reserve Bank Quarterly Bulletin with government budget data collected from Treasury National Budget Reports.

Figure 4.76 – Taxes paid by pretax national income group: 1993 versus 2019

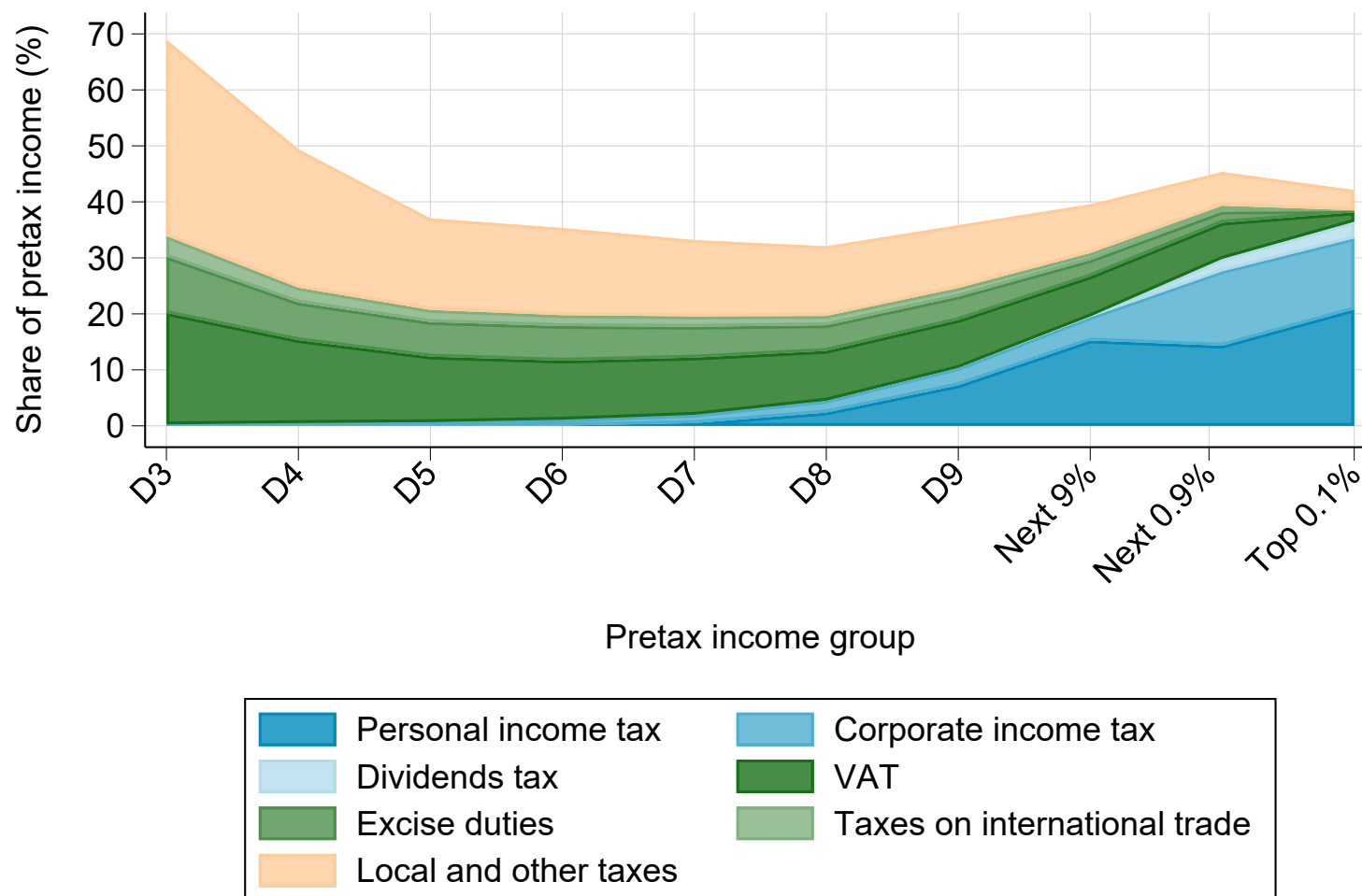
(a) 1993



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the effective tax rate faced by pretax income group in South Africa in 1993, expressed as a share of pretax income. Equivalised household pretax income per capita.

Figure 4.76 – Taxes paid by pretax national income group: 1993 versus 2019

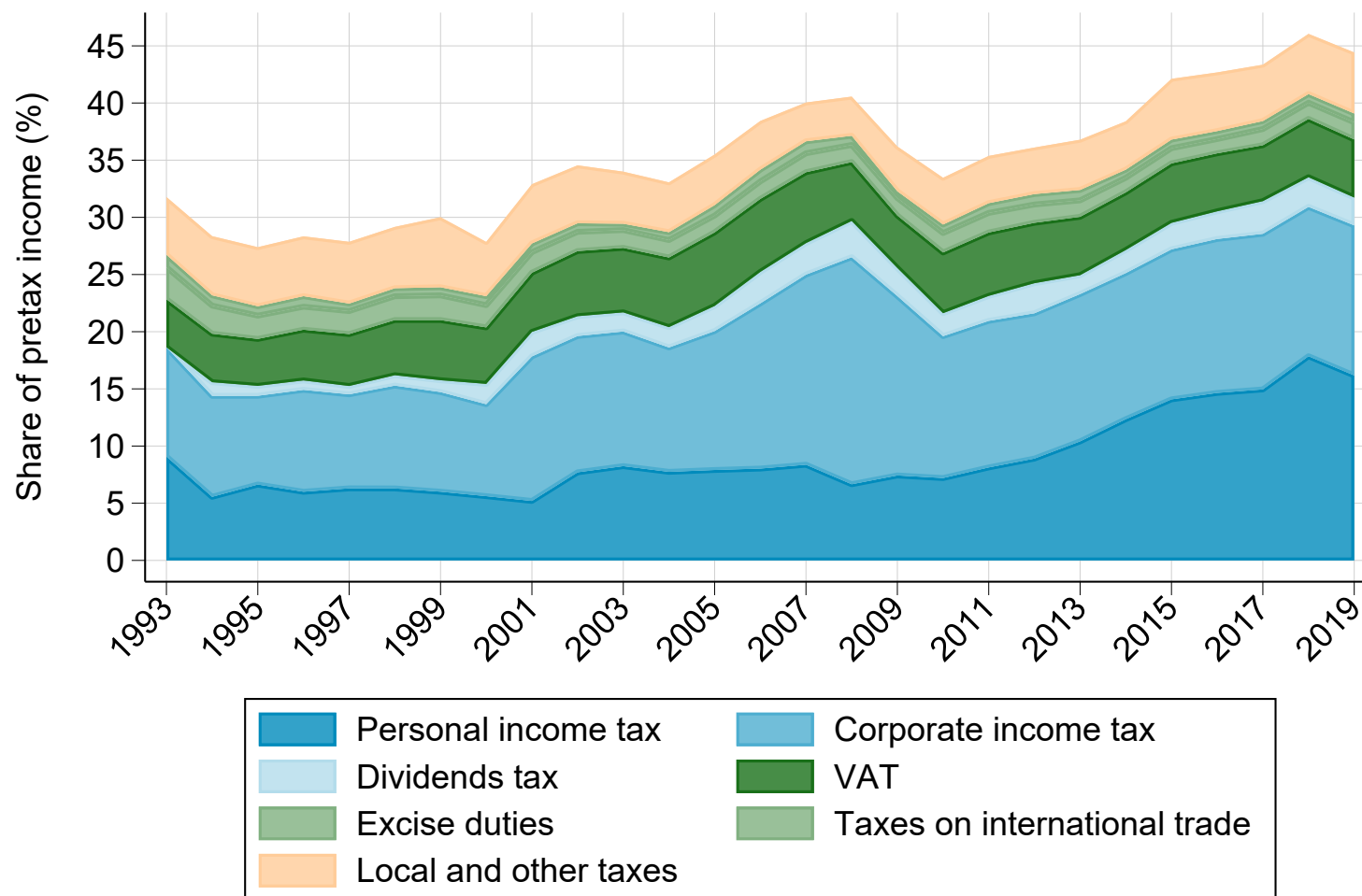
(b) 2019



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the effective tax rate faced by pretax income group in South Africa in 2019, expressed as a share of pretax income. Equivalised household pretax income per capita.

Figure 4.77 – Taxes paid as a percentage of pretax income

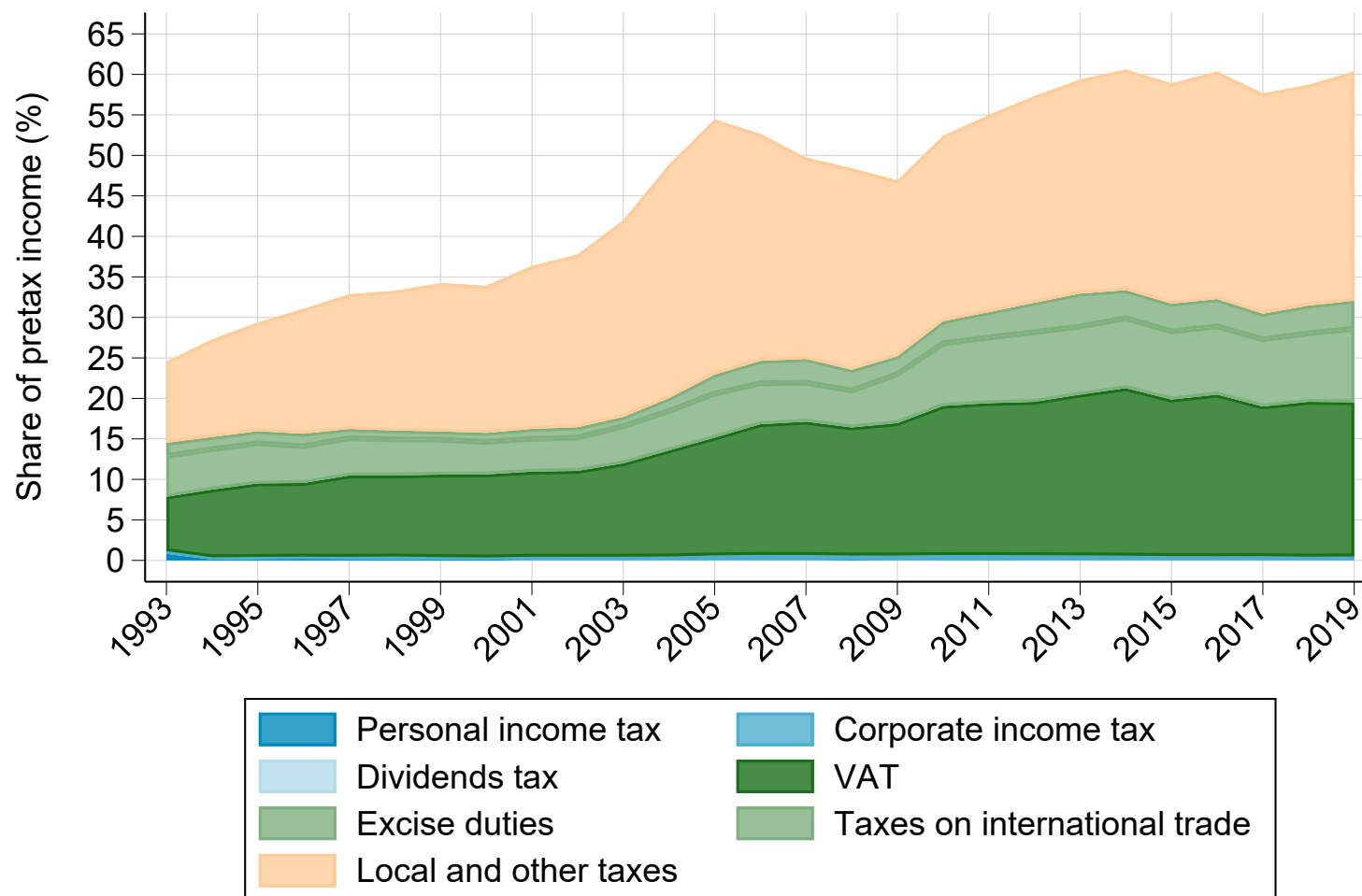
(a) Top 1%



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the effective tax rates faced by top 1% pretax income earners (panel A) and bottom 50% pretax income earners (panel B), expressed as a share of pretax income. Equivalised household pretax income per capita.

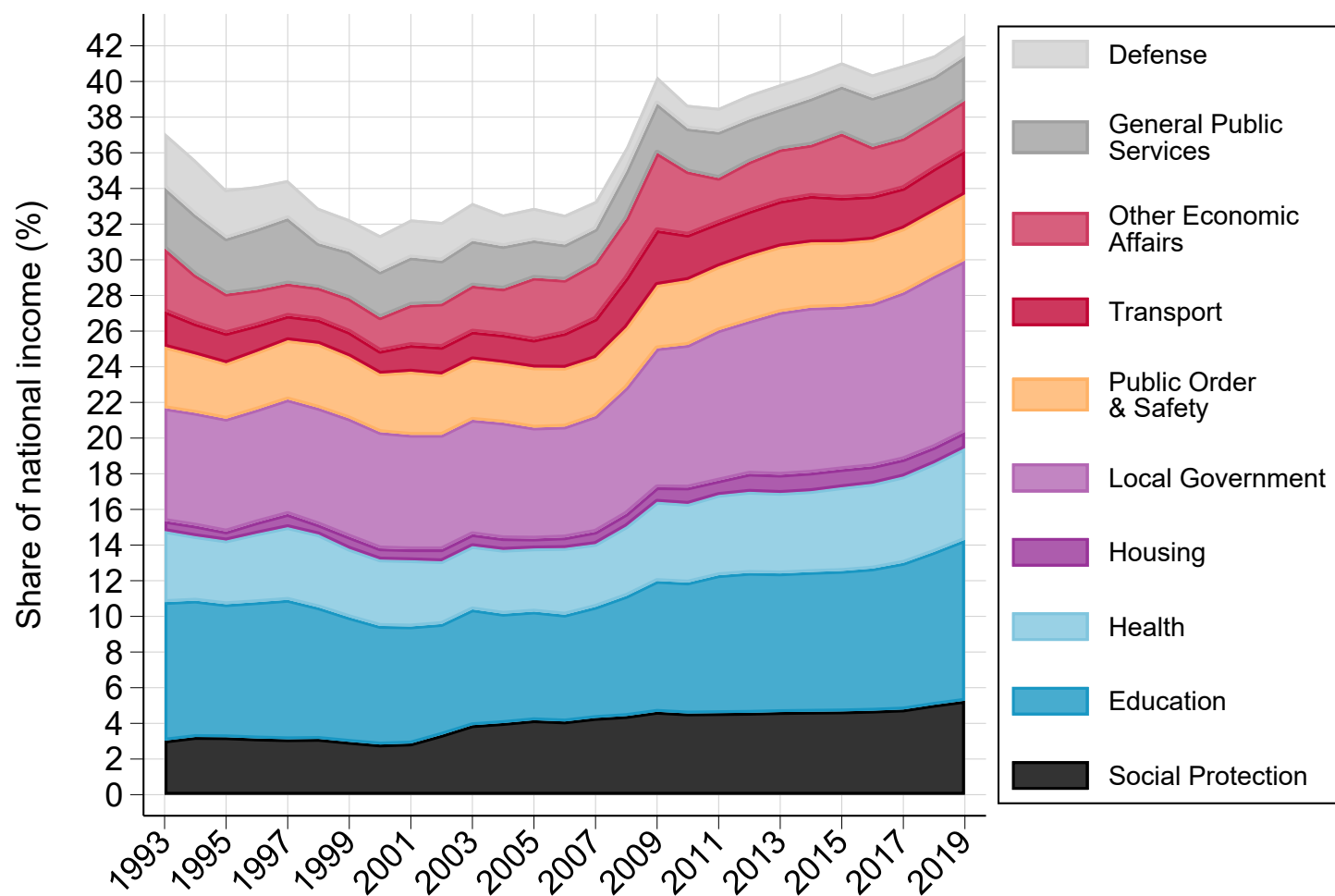
Figure 4.77 – Taxes paid as a percentage of pretax income

(b) Bottom 50%



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the effective tax rates faced by top 1% pretax income earners (panel A) and bottom 50% pretax income earners (panel B), expressed as a share of pretax income. Equivalised household pretax income per capita.

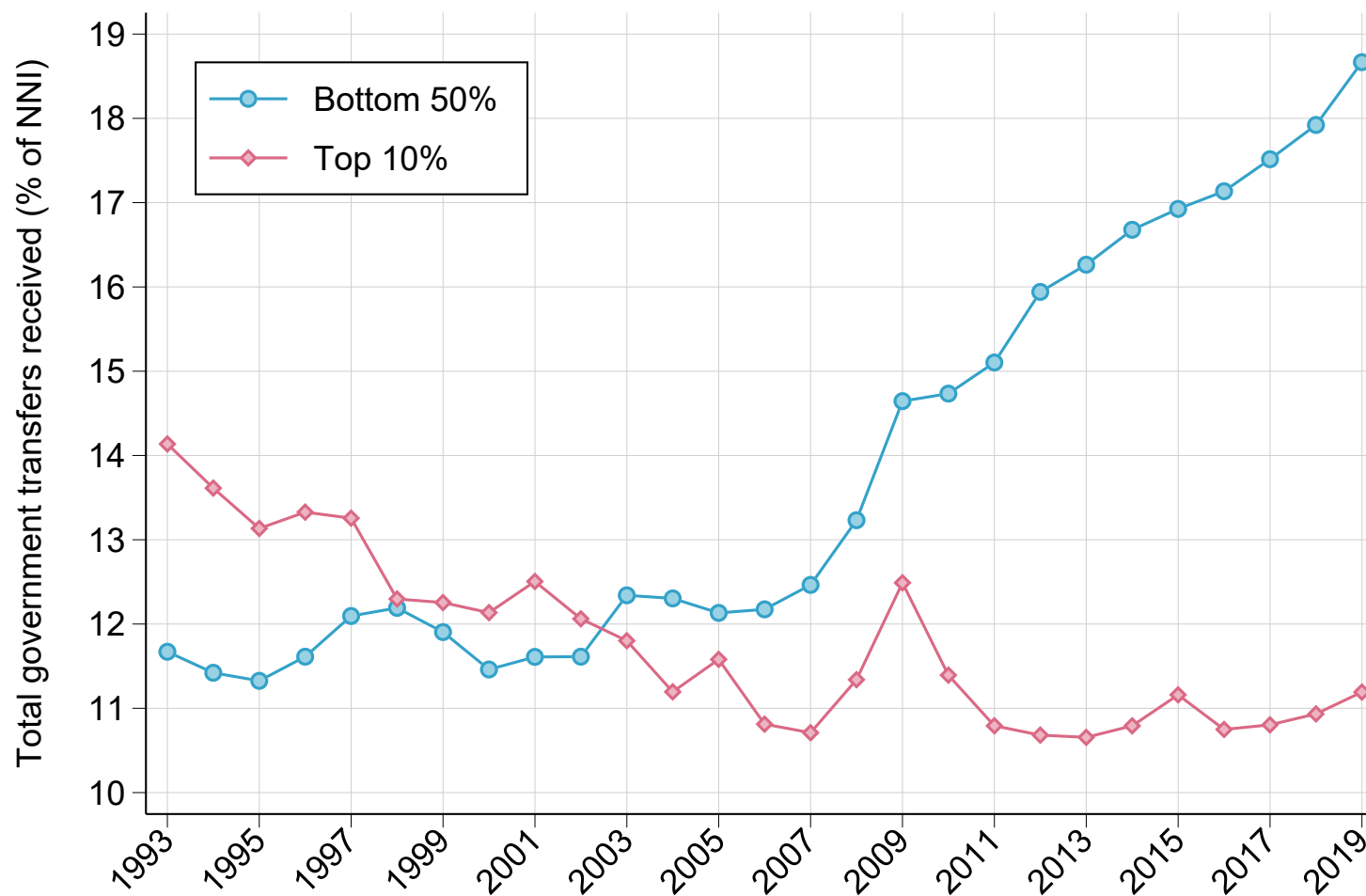
Figure 4.78 – Government expenditure in South Africa



Notes. Authors' computations combining national accounts series from the South African Reserve Bank Quarterly Bulletin with government budget data collected from Treasury National Budget Reports.

Figure 4.79 – The rise of social transfers

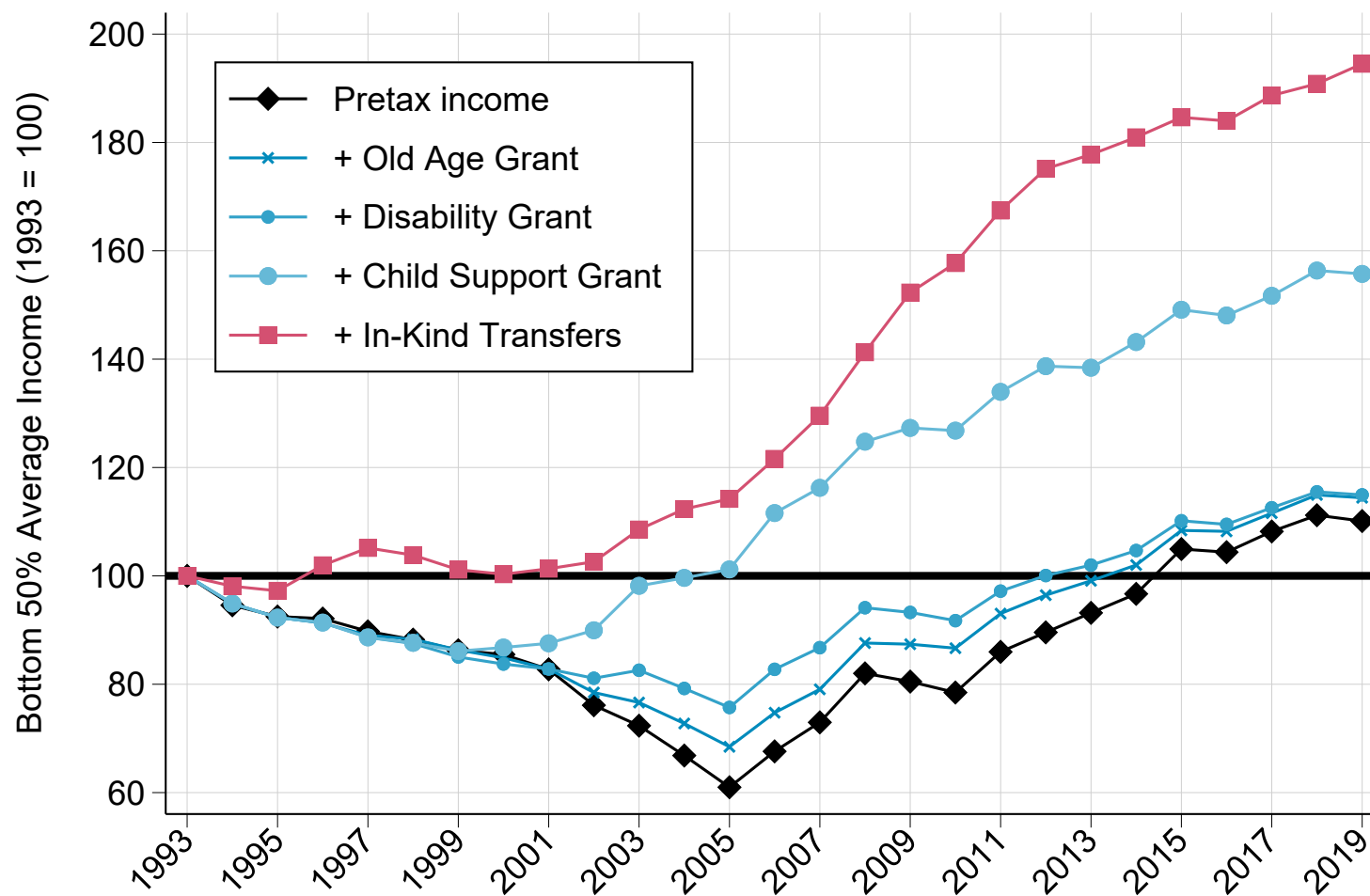
(a) Total individualized transfers received by pretax income group



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the total individualized transfers (social protection, education, and health transfers) received by bottom 50%, middle 40%, and top 10% pretax income earners, expressed as a share of national income. Equivalised household pretax income per capita.

Figure 4.79 – The rise of social transfers

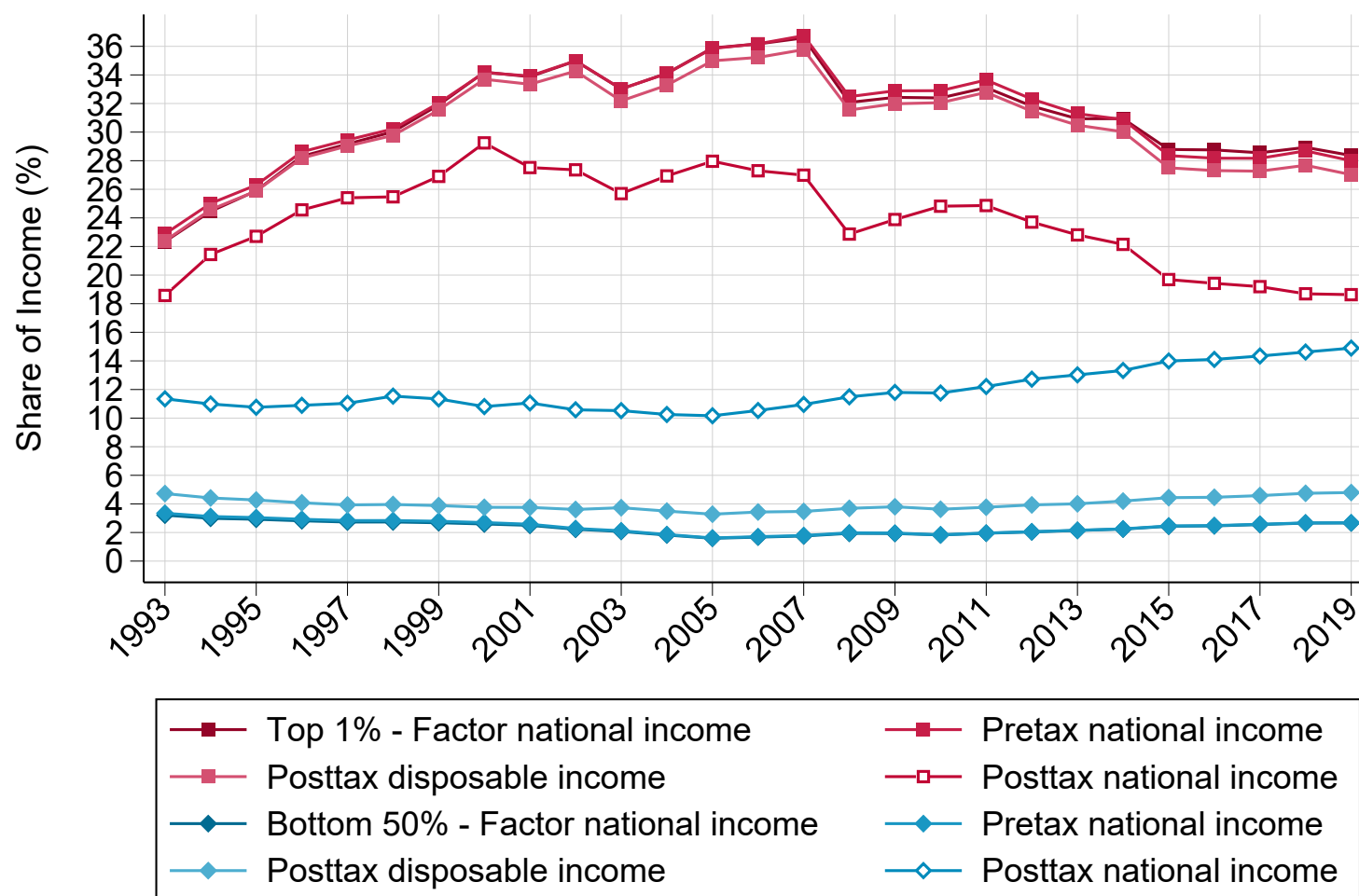
(b) Bottom 50% average income growth, before and after transfers



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the cumulated income growth rate of bottom 50% earners between 1993 and 2019, before and after adding specific social transfers to individual pretax incomes. Equivalised household income per capita.

Figure 4.710 – The overall impact of taxes and transfers on inequality

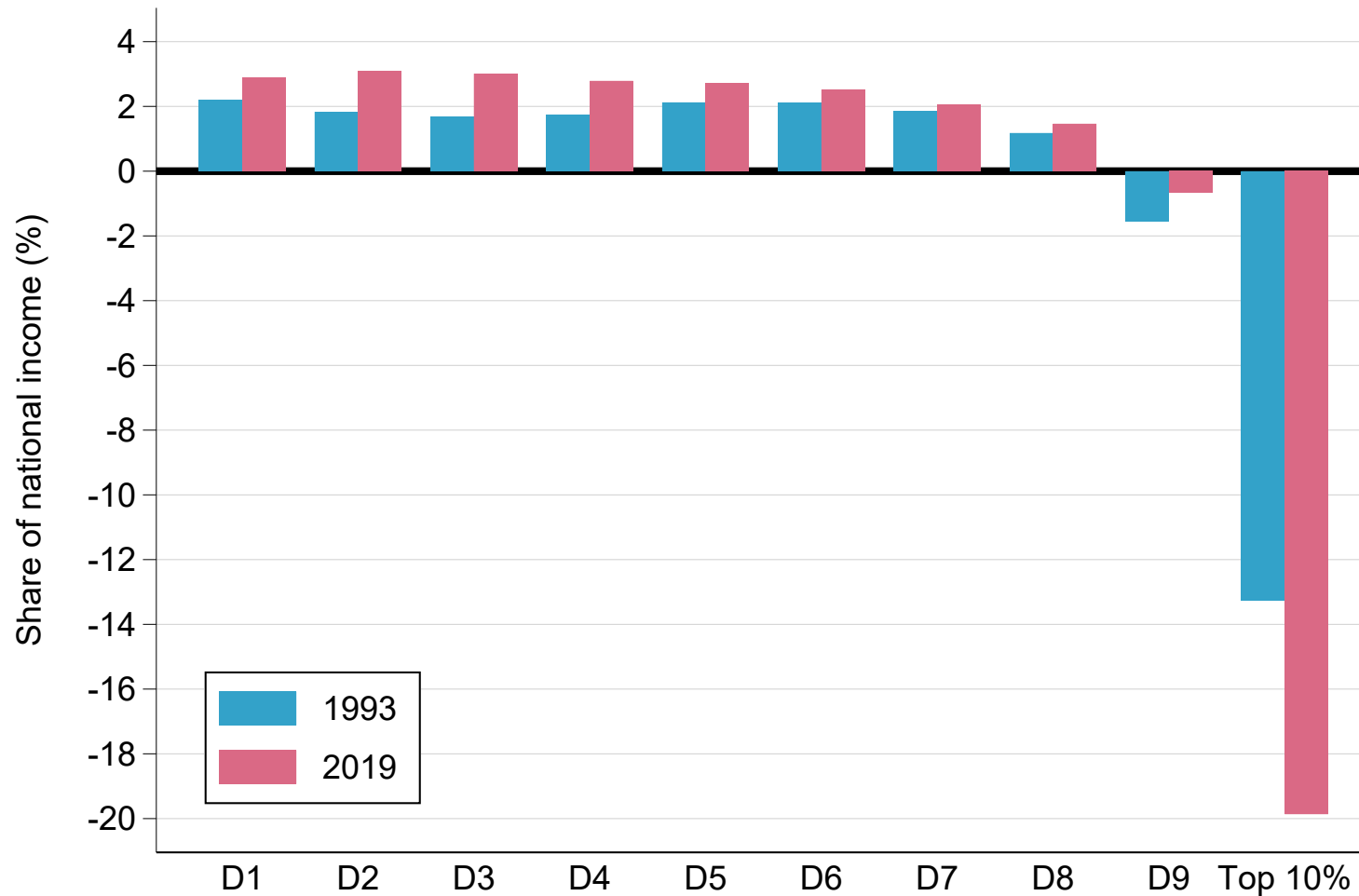
(a) Top 1% versus bottom 50%: from factor to posttax national income



Notes. Authors' computations combining survey, tax, and national accounts data. Equivalised household income per capita.

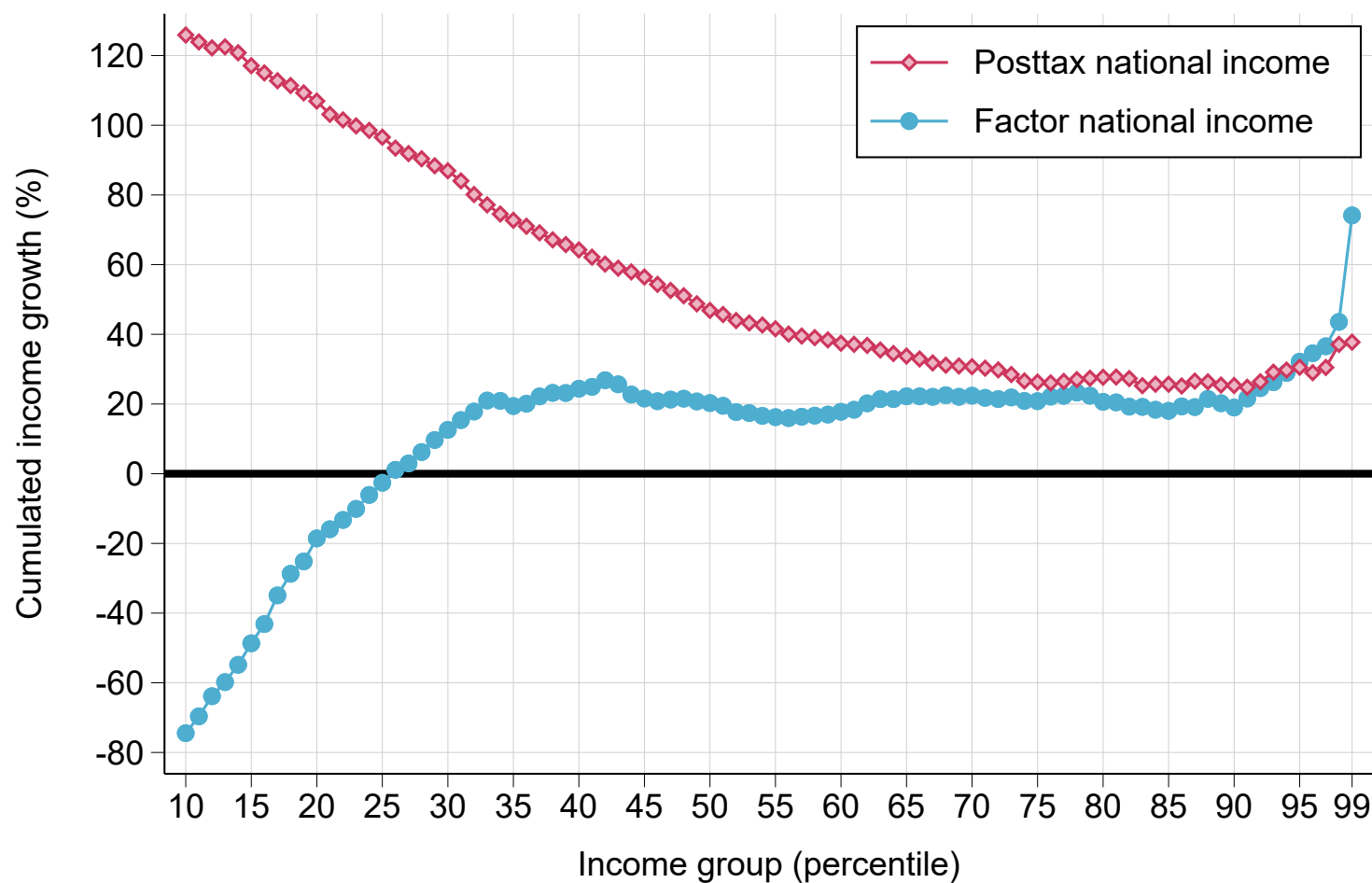
Figure 4.710 – The overall impact of taxes and transfers on inequality

(b) Net transfers operated by the tax-and-transfer system by factor income group



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the net transfers operated between factor income deciles by the tax-and-transfer system, that is, the difference between total transfers received and total taxes paid, expressed as a share of national income. Equivalised household income per capita.

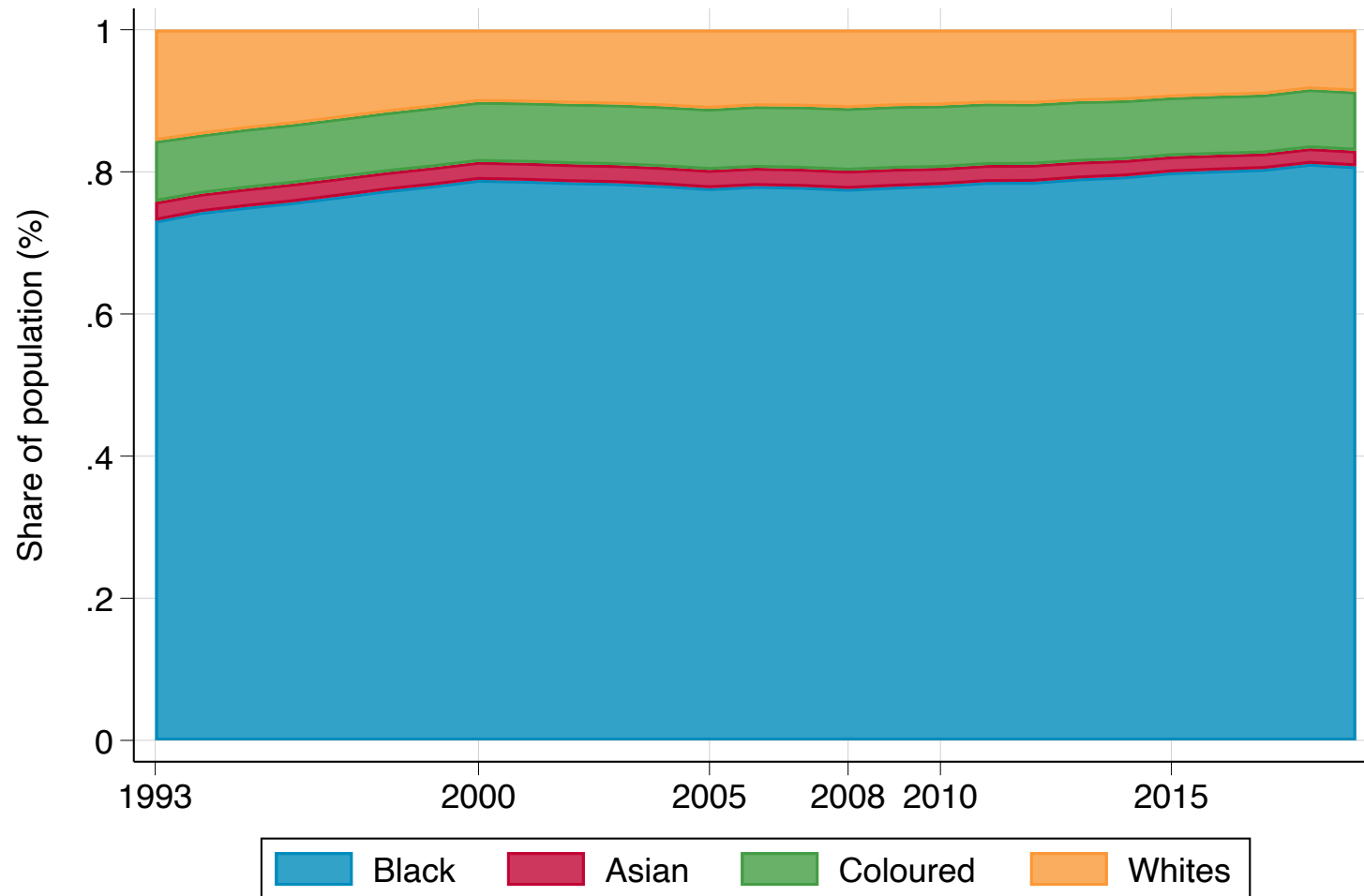
Figure 4.711 – Redistribution, inequality, and growth: cumulated income growth by percentile, 1993-2019



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the cumulated income growth rate by percentile between 1993 and 2019 in terms of factor national income and posttax national income. Equivalised household income per capita.

Figure 4.712 – Demographic Decomposition by Racial Group

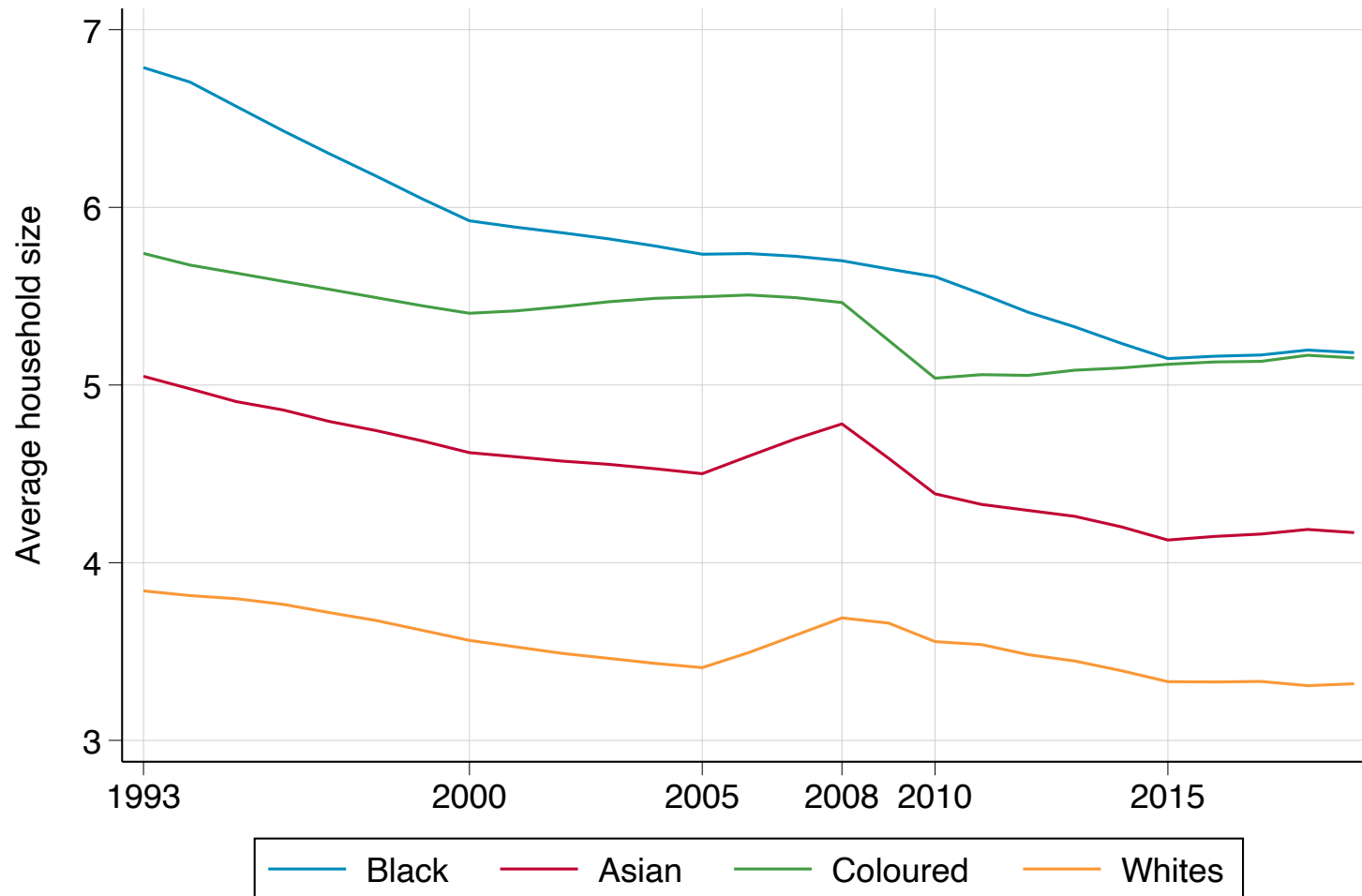
(a) Population size



Notes. Authors' computations using our six main income surveys.

Figure 4.712 – Demographic Decomposition by Racial Group

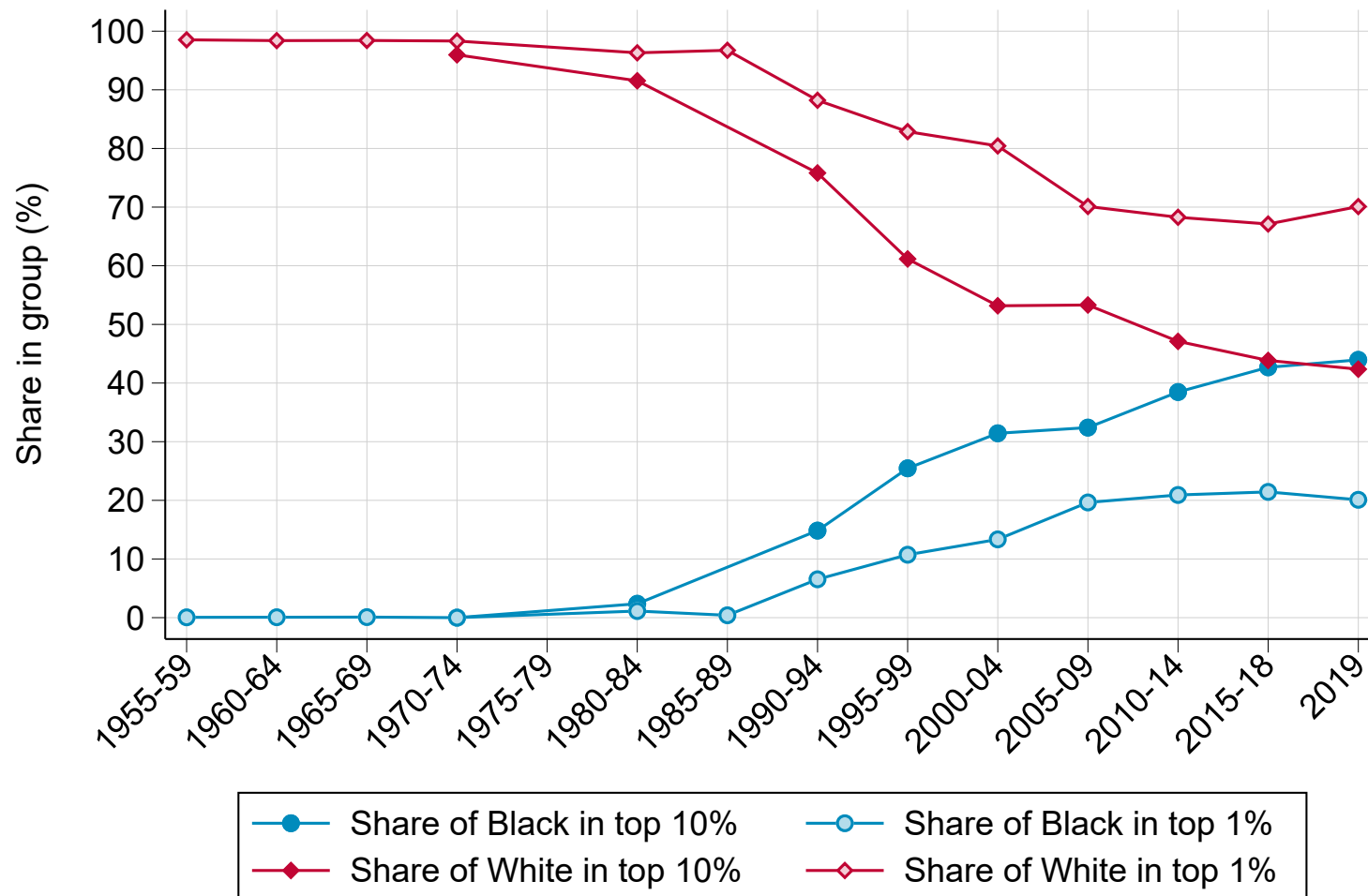
(b) Household size



Notes. Authors' computations using our six main income surveys.

Figure 4.713 – Racial inequality and top incomes

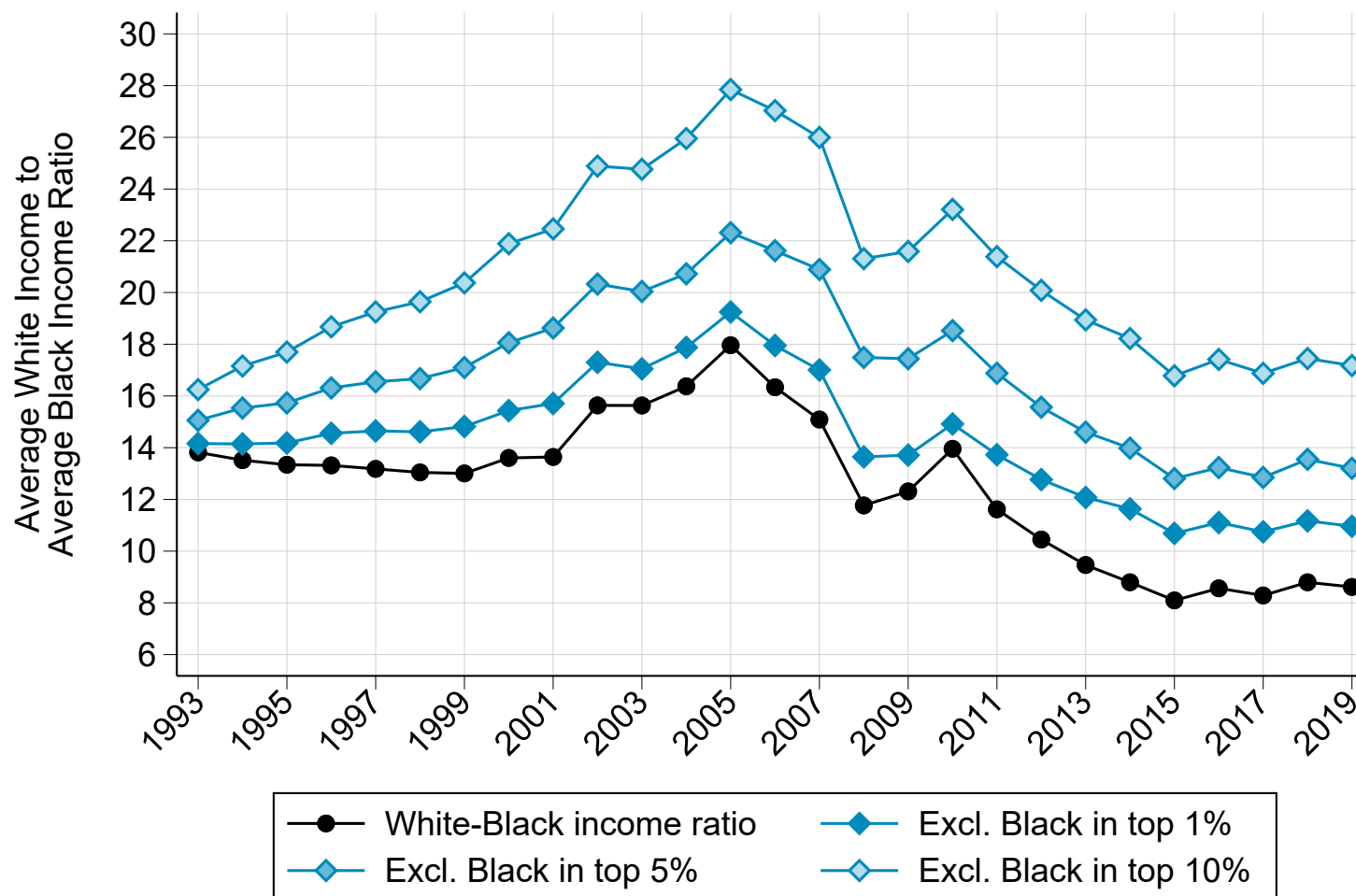
(a) Share of Black versus White earners in top factor income groups, 1955-2019



Notes. Authors' computations using from Alvaredo and Atkinson (2010) for the top 1% before 1993; census data for 1970, 1980, and 1990 (top 10%); distributional national accounts for 1993-2019. The figure represents the share of Black earners and White earners in top factor income groups between 1955 and 2019. Equivalised household income per adults.

Figure 4.713 – Racial inequality and top incomes

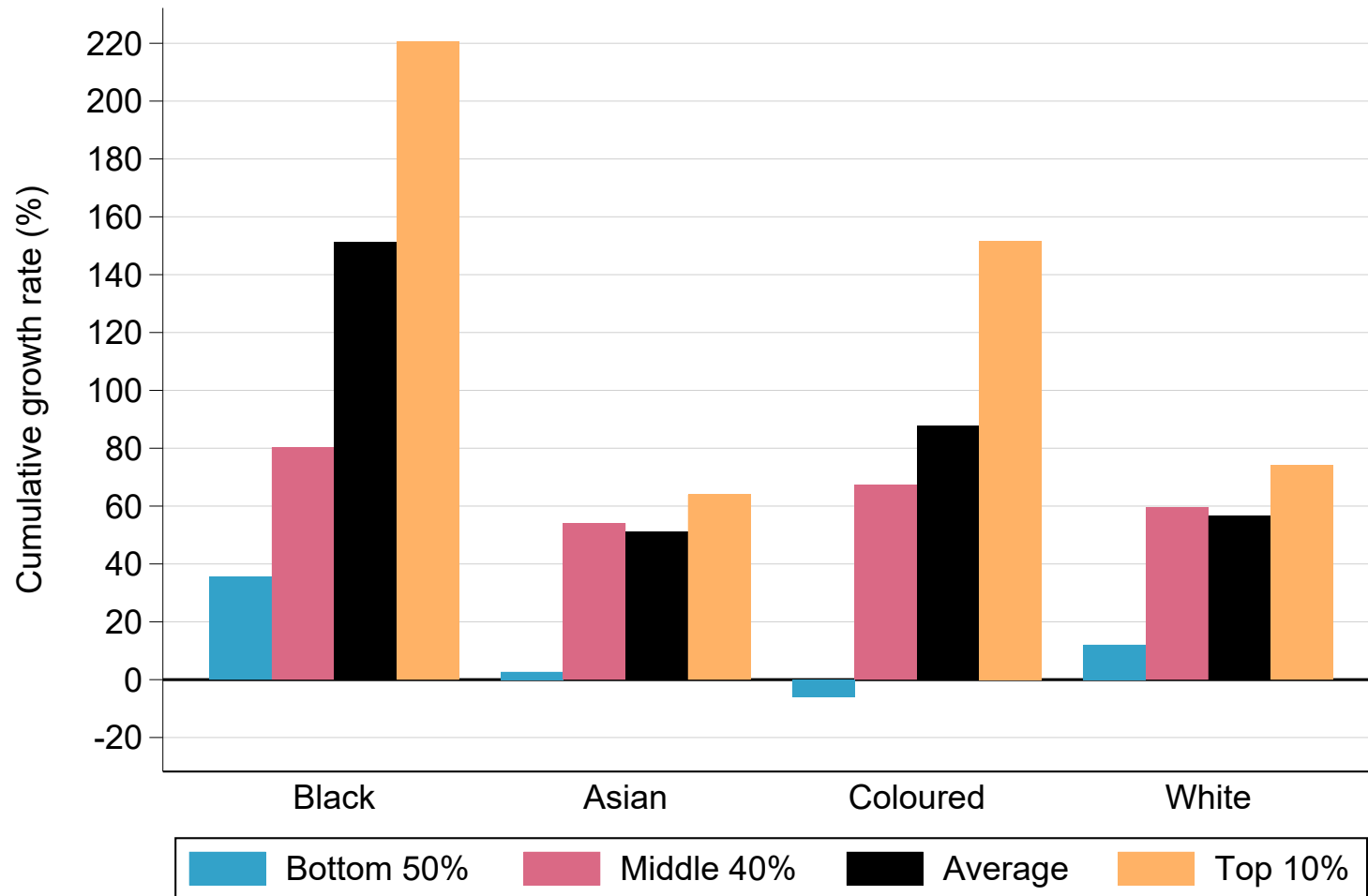
(b) Top Black incomes and the decline in the racial factor income gap



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the ratio of the average factor income of White earners and the average factor income of Black earners between 1993 and 2019, before after excluding all Black earners located in the top 10%, the top 5%, and the top 1% of the overall factor income distribution. Equivalised household income per capita.

Figure 4.713 – Racial inequality and top incomes

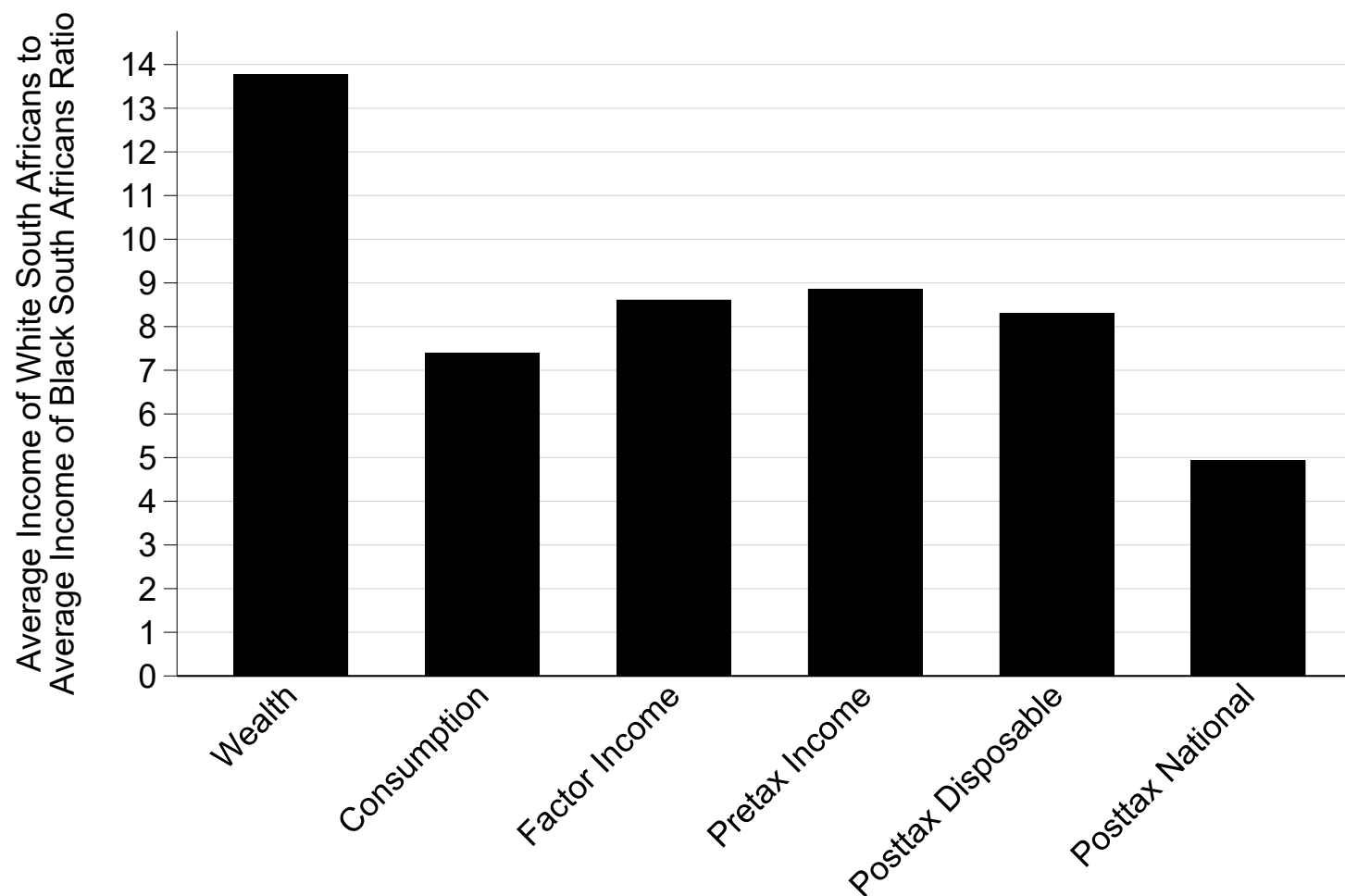
(c) The distribution of factor income growth within population groups, 1993-2019



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the total growth rate of selected factor income groups within each population group from 1993 to 2019. The top 10% of Black earners saw their average income grow by 200% during this period. Equivalised household income per capita.

Figure 4.714 – The structure of racial inequality in 2019

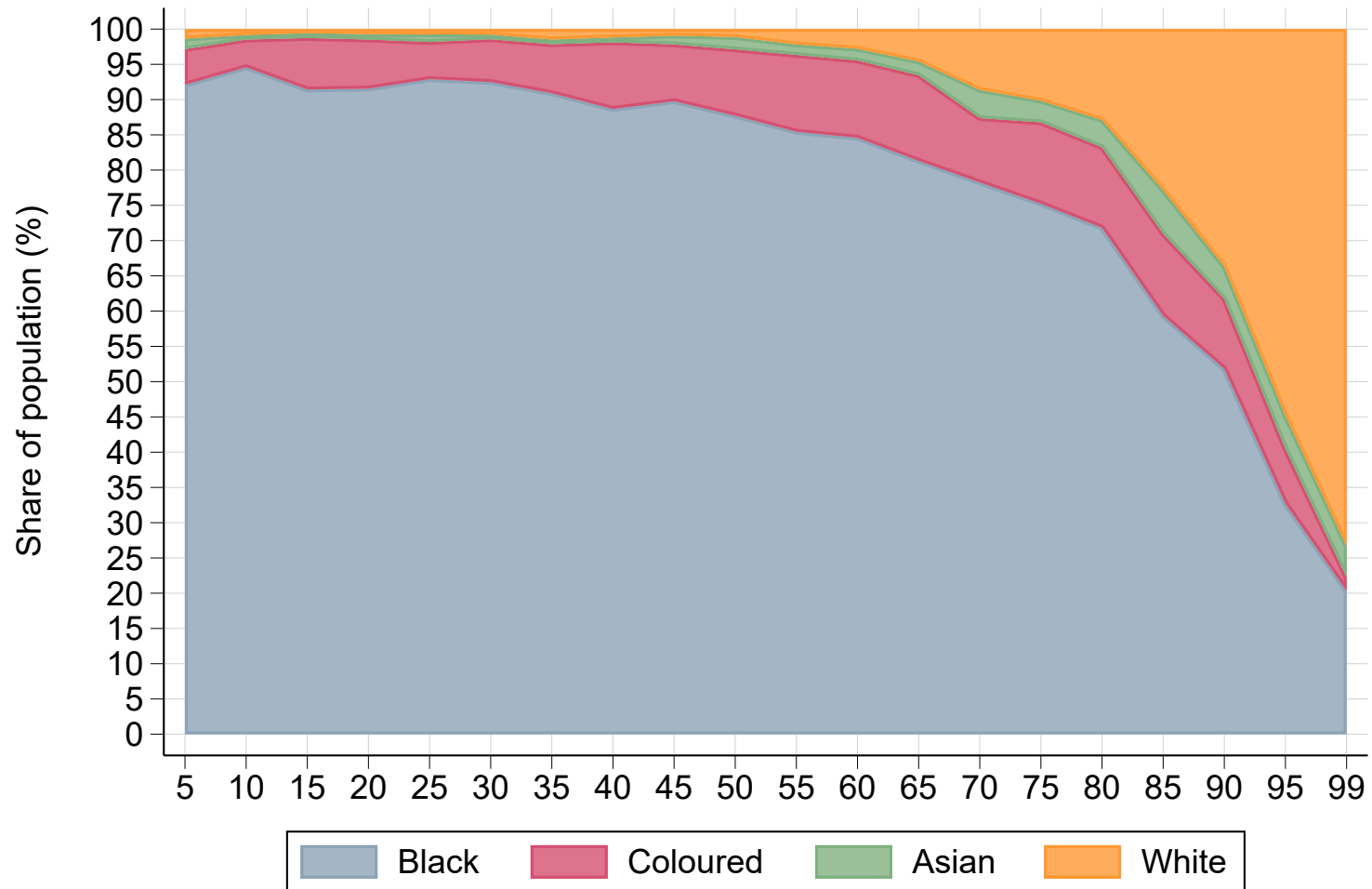
(a) The White-Black gap in income, consumption, and wealth



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the ratio of White to Black average income, consumption, and wealth in 2019. Equivalised household income per capita.

Figure 4.714 – The structure of racial inequality in 2019

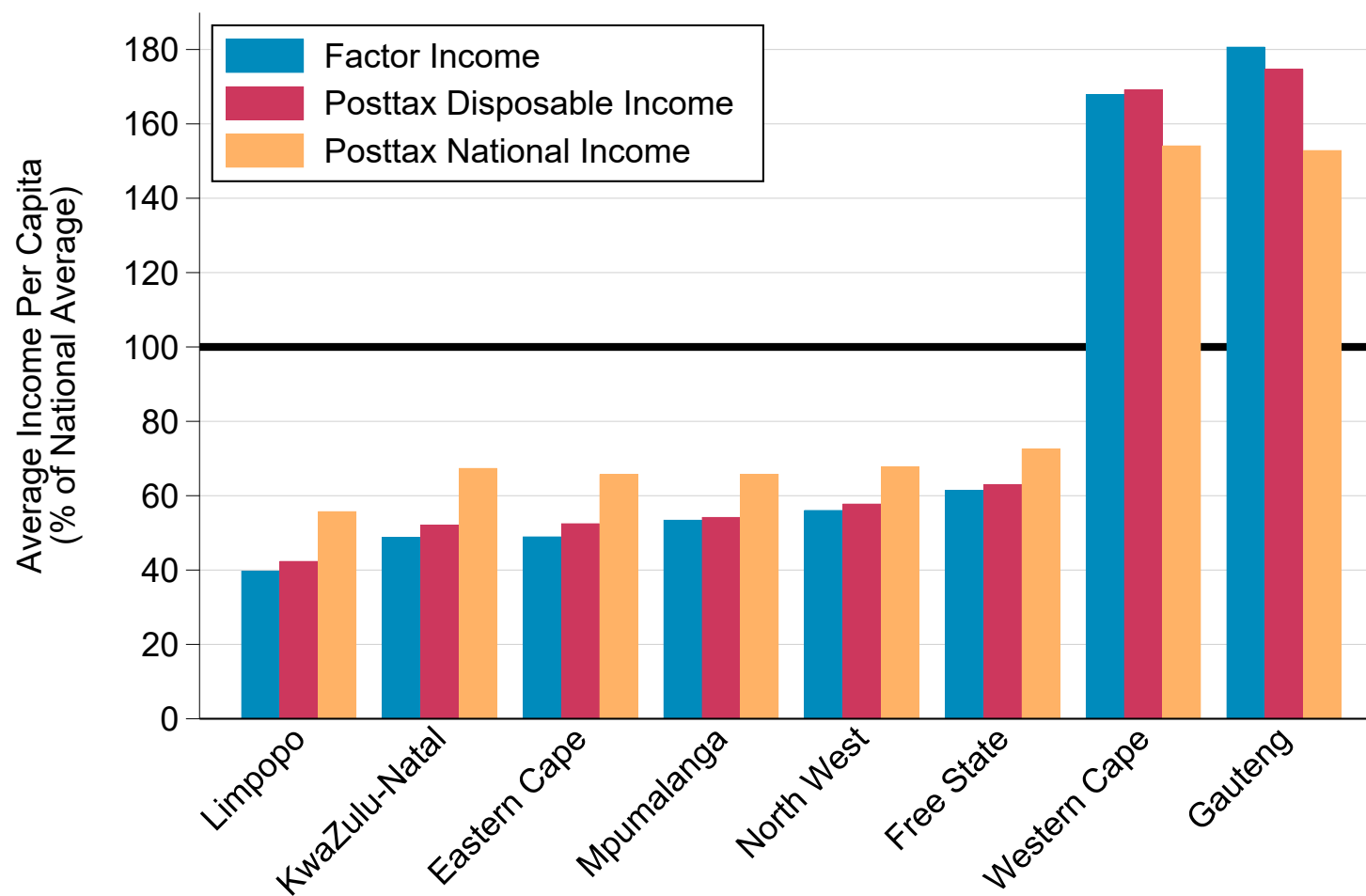
(b) Racial composition of posttax national income groups



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the composition of posttax national income groups (ventiles) by population group in 2019. Equivalised household posttax income per capita.

Figure 4.715 – Spatial inequality and redistribution

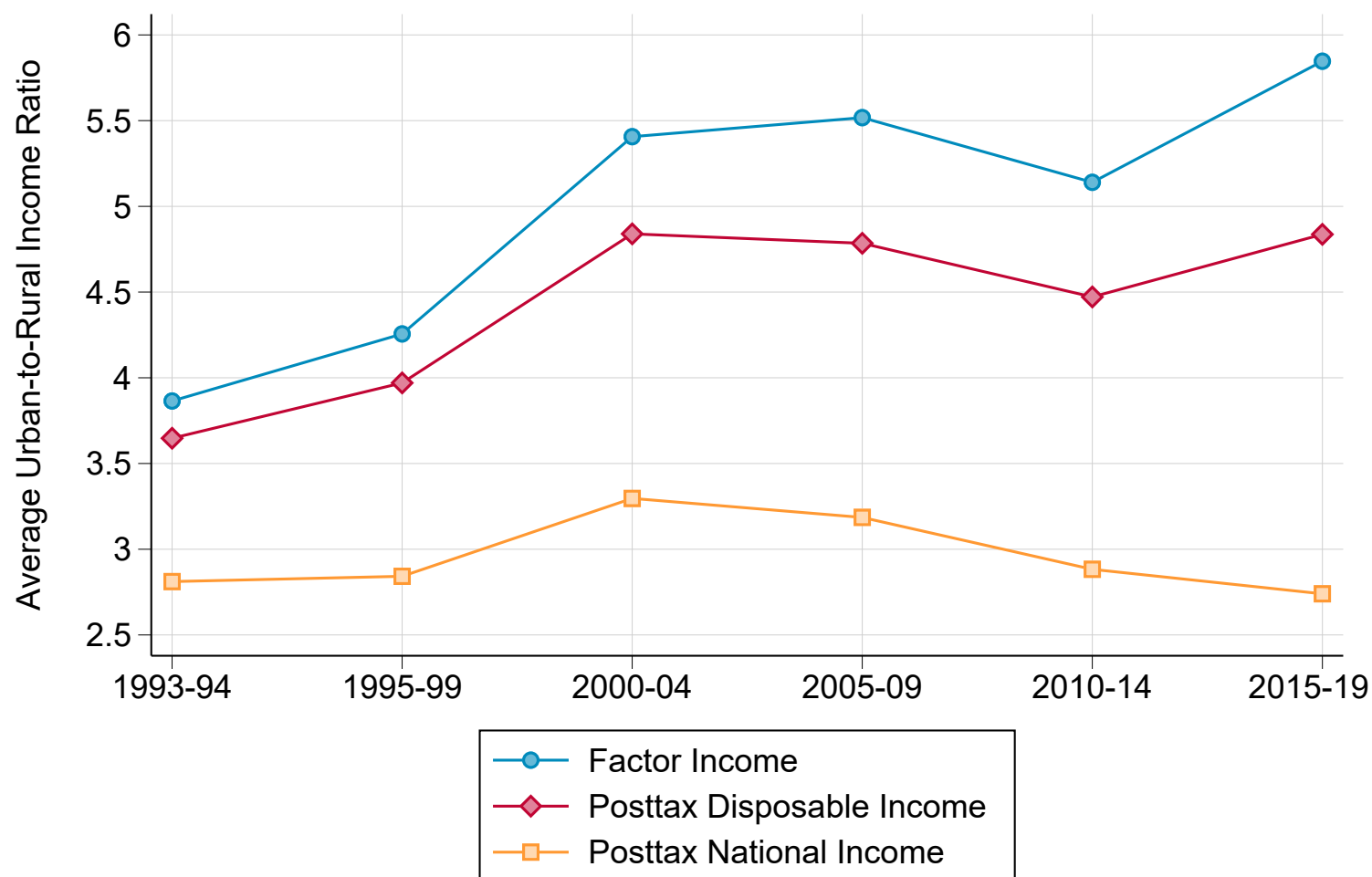
(a) Average income by province relative to national income, 2019



Notes. Authors' computations combining survey, tax, and national accounts data. Limpopo includes the North West province. The figure represents the average income of South African provinces, before and after taxes and transfers, relative to the national average in 2019. Equivalised household income per capita.

Figure 4.715 – Spatial inequality and redistribution

(b) Social transfers and the rural-urban income gap



Notes. Authors' computations combining survey, tax, and national accounts data. The figure represents the ratio of the average income of urban areas to the average income of rural areas, before and after accounting for taxes and transfers. Equivalised household income per capita.

4.8. Tables

Table 4.81 – The distribution of factor national income and pretax national income

Item	Distribution method	% of NNI (2019)
Factor national income		100%
Compensation of employees	Proportional rescaling	57%
Mixed income	Proportional rescaling	9%
Property income, net		9%
Rents	Proportional rescaling	2%
Interest	Proportional rescaling	2%
Dividends	Proportional rescaling	4%
Other property income	Proportionally to pension and life insurance wealth	6%
Interest paid by households	Proportionally to factor income of debtors	-5%
Imputed rents of owner-occupiers	Proportionally to housing wealth of owner-occupiers	3%
Corporate undistributed profits	Proportionally to equity	8%
Taxes less subsidies on production and imports	Proportionally to factor income	11%
Remaining national income components	Proportionally to factor income	3%
Pretax national income		100%
Pension contributions	Observed	6%
Pension benefits	Observed	3%
Pension deficit or surplus	50% prop. to contributions, 50% prop. to benefits	3%
Unemployment insurance contributions	Rule-based imputation	0.5%
Unemployment insurance benefits	Observed	0.4%
Unemployment insurance fund deficit or surplus	50% prop. to contributions, 50% prop. to benefits	0.1%

Notes. The table reports the methodology used to distribute the various components of factor national income and pretax national income (for more details, see sections II.B and II.C), along with the size of each component expressed as a share of net national income (NNI) in 2019. Factor national income is the sum of all income flows accruing directly or indirectly to individuals, before accounting for the operation of the tax-and-transfer system, and before accounting for the operation of the pension and unemployment systems. Pretax national income is equal to factor income after the operation of the pension and unemployment systems. Both factor national income and pretax national income sum to the net national income.

Table 4.82 – The distribution of taxes

Item	Distribution method	% of NNI (2019)
Direct taxes		19.0%
Personal income tax	Rule-based imputation	11.2%
Corporate income tax	Proportionally to equity	6.1%
Dividends tax	Proportionally to dividends	0.8%
Skills development levy	Rule-based imputation	0.4%
Transfer duties	Proportionally to housing wealth	0.2%
Securities transfer tax	Proportionally to equity	0.1%
Estate duty	Proportionally to net wealth	0.1%
Donations tax	Proportionally to net wealth	0.0%
Other taxes on income	Proportionally to pretax income	0.1%
Indirect taxes		12.6%
Value added tax	Proportionally to expenditure (excl. zero-rated / informal market)	8.0%
General Fuel Levy	Proportionally to fuel and transport expenditure	1.8%
Other excise duties	Proportionally to tobacco and alcohol expenditure	1.1%
Other taxes on goods and services	Proportionally to total expenditure	0.3%
Taxes on international trade	Proportionally to import-density-corrected expenditure	1.4%
Other government revenue	Proportionally to pretax income	2.0%
Total consolidated revenue		33.6%

Notes. The table reports the methodology used to distribute all taxes in South Africa at the individual level (for more details, see section II.D), along with the size of each component, expressed as a share of net national income (NNI), in 2019.

Table 4.83 – The distribution of factor income in South Africa in 2019

	Number of individuals	Income threshold	Average income	Income share
Full population	58,600,000	\$ 0	\$ 11,700	100%
Bottom 90% (p0p90)	52,740,000	\$ 0	\$ 4,100	31.3%
Bottom 50% (p0p50)	29,300,000	\$ 0	\$ 600	2.7%
Middle 40% (p50p90)	23,440,000	\$ 2,200	\$ 8,400	28.7%
Top 10% (p90p100)	5,860,000	\$ 26,200	\$ 80,600	68.7%
Top 1% (p99p100)	586,000	\$ 129,000	\$ 332,000	28.3%
Top 0.1% (p99.9p100)	58,600	\$ 662,000	\$ 973,000	8.3%
Top 0.01% (p99.99p100)	5,860	\$ 1,370,000	\$ 2,400,000	2.0%

Notes. The table reports the distribution of factor national income in 2019, providing information for each income group on the number of adults belonging to this group, the minimum income required to belong to this group, the average income of this group expressed in 2019 PPP US dollars (\$1 = R6.3), and the share of factor national income received. Factor national income is the sum of all income flows accruing directly or indirectly to individuals, before accounting for the operation of the tax-and-transfer system, and before accounting for the operation of the pension and unemployment systems. Income is split equally among all adults members of the household (aged 20 or above).

Appendices

This section provides additional details on the methodology used to build South African Distributional National Accounts. Section A lists the data sources used to estimate macroeconomic aggregates, including national accounts, population estimates, and other government budget and administrative data. Section B describes the combination of available survey and tax data to build a microfile covering the distributions of factor national income every year from 1993 to 2019. Section C explains how taxes and transfers are allocated to reach posttax national income.

A. Harmonization of Macroeconomic Aggregates

National Accounts Data

Main Aggregates Estimates of national income, wealth, and expenditure aggregates come from the South African Reserve Bank (SARB) quarterly bulletin.² The published files provide detailed decompositions of national accounts components, which we directly match with the microfile to estimate distributional national accounts. The exceptions are mixed income and corporate undistributed profits, which we decompose further to refine the imputation.

Decomposition of Mixed Income and Imputed Rents The SARB data does not publish separate series for mixed income, rental income, and imputed rents, instead providing a single aggregate for [B2N + B3N, S14]. To derive an estimate of total rental income received by households, we combine all income surveys (1993, 1995, 2000, 2005, 2008, 2010, 2015: see section B) and General Household Surveys (GHS, 2016-2019), which have collected information on rents paid by South African tenants.³ The resulting total rental income represented 1.9% of national income (14% of [B2N + B3N, S14]) in 2019, up from 1.4% (12%) in 1993. Following recommendations by the South Africa Reserve Bank, we assume that imputed rents represent a fixed 20% of the total, and we compute mixed income (*i.e.*, self-employment income excluding rental income) as the residual of these two categories.

²See <https://www.resbank.co.za/en/home/publications/quarterly-bulletin1/download-information-from-xlsx-data-files>.

³I first aggregate all rent payments recorded in income surveys. We then interpolate the series linearly between years to cover the entire 1993-2015 period. Finally, we use GHS growth rates in rent payments to extrapolate series forward to 2019.

Decomposition of Corporate Undistributed Profits To allocate corporate retained earnings to individuals, one has to decompose them between the part that belongs to households (distributed proportionally to equity ownership) and the part that belongs to the government (distributed proportionally to factor income). We do so by relying on a preliminary estimate published by the SARB on the equity assets and liabilities of the household and government sectors in 2011 (see [Beer and Kock, 2017](#)). Dividing the sum of the equity assets held by the government by the total equity liabilities of the corporate sector, we estimate that about 93% of retained earnings can be attributed to households. In the absence of better data, we assume that this share has remained stable over the 1993-2019 period.

General Government Revenue and Expenditure Data

To move from factor income to pretax income and then posttax income, we collect data on general government revenue and expenditure from three main sources: the SARB, the OECD, and the South African National Treasury.

Government Revenue Yearly data on consolidated government revenue and its decomposition are available from the public finance series published in the SARB Quarterly Bulletin. We complement these harmonized series with OECD public revenue data to further decompose revenue from direct taxes into the personal income tax, the corporate income tax, and other taxes on income and wealth.⁴

Government Expenditure Data on the composition of general government expenditure by function are available from the Treasury Budget Reviews.⁵

Social Security Data To make the DINA microfile more representative of Unemployment Insurance Fund (UIF) and private pension contributions and benefits, we collect data on total contributions/benefits and number of contributors/recipients to the UIF and private pension funds in South Africa. Data on total UIF revenue and expenditure (2001-2019) and on the number of UIF recipients (2008-2012) are reported in various issues of the Treasury Budget Review. The number of individuals earning private pension income is estimated from the income tax panel microdata (2011-2017) available from the South African Revenue Service (see [Ebrahim and Axelson, 2019b](#)), and extrapolated to 1993 assuming that it has remained

⁴See <https://stats.oecd.org/Index.aspx?DataSetCode=REVZAF>.

⁵See <http://www.treasury.gov.za/documents/national%20budget/default.aspx>.

a constant share of the adult population.⁶ Total contributions to private pension funds and total private pension income are also estimated from the income tax panel, and extrapolated to 1993 using the growth rates of social contributions received by financial corporations and social benefits paid by financial corporations, respectively (both available from SARB national accounts data).

Social Protection Data We also collect data on the number of recipients and the monthly values of social grants from various issues of the Treasury Budget Reviews. Data on grant values are available every year since 1993 (or since the year the grant was implemented) for all major cash transfers in South Africa (including the old age grant, the disability grant, the child support grant, the foster care grant, and the care dependency grant). Data on the number of recipients of each grant are available since 1996.

B. Construction of DINA Microfile

Combination of Survey Data Sources

The main data source used to estimate the distributions of income, consumption, and wealth at the micro level are household surveys that have collected detailed information on the earnings and expenditure of households in South Africa. Seven such surveys, which we refer to as “income surveys” in what follows, have been conducted since 1993: the Project for Statistics on Living Standards and Development (1993), the Income and Expenditure Surveys (1995, 2000, 2005, 2010), and the Living Conditions Surveys (2008, 2015). Drawing on representative samples of households, they ask individuals to report earnings from various sources (such as wages, self-employment income, and property income), as well as other information such as contributions to private pension funds, taxes and transfers received, the market value of the home individuals live in, or expenditure on specific goods and services.

I create a harmonized microfile covering the entire 1993-2019 period by combining all available surveys (1993, 1995, 2000, 2005, 2008, 2010, and 2015) and filling missing years in the following way. For a given missing year (for instance 1997), we create a new dataset by appending all observations from the two surveys available in surrounding years (1995 and 2000), and then reweight observations so as to give a weight to each survey that is proportional to the distance from the year considered. To approximate the distribution of income in 1997, for instance, we append the 1995 and 2000 IES surveys, and then multiply existing sample

⁶This is a reasonable assumption to the extent that the number of pension recipients has also remained stable in income surveys, although at a lower level than in the tax microdata.

weights by 1/2 in the former and 1/3 in the latter. This is similar to a linear interpolation strategy: it amounts to considering that in 1997 the distribution of income was somewhere between that of 1995 and that of 2000, and was closer to that of 1995. The resulting microfile thus combines all available surveys to cover individual-level data every year from 1993 to 2019.

Combination of Surveys with Tax Data

I correct surveys for misreporting of income at the top of the distribution by combining them with tabulated income tax returns. This correction is performed in three steps, following the methodology developed by [Blanchet, Flores, and Morgan \(2018a\)](#).

First, we define an income concept, “merging income”, that can be consistently measured in both survey data and the income tax panel microdata (2011-2017). This income concept is equal to the sum of gross wages, business income, interest, rental income, and private pension income.

Second, we generate a “taxable income” variable in the survey microfile by multiplying merging income by the ratio of taxable income to merging income by percentile observed in the tax microdata. This effectively amounts to incorporating deductions (that is, the gap between merging income and taxable income) in the survey microdata.⁷

Third, we calibrate the survey microfile on the tabulated income tax returns available from SARS, which report the number of taxpayers and total taxable income by income tax bracket every year since 2002 (as well as in 1993). We first recover full distributions from the tax tabulations using Generalized Pareto Interpolation ([Blanchet, Fournier, and Piketty, 2017a](#)).⁸ We then calibrate the survey microdata on the tax tabulations using the algorithm developed by [Blanchet, Flores, and Morgan \(2018a\)](#), which reweights survey observations so as to match the distribution of top taxable incomes reported in the tax data. The resulting survey microfile is thus perfectly representative of the distribution of taxable income reported in the income tax tabulations.

⁷For simplicity, we take the overall average of this ratio by percentile observed in 2011-2017 and apply it to the entire period. This corresponds to assuming that the profile of deductions has remained relatively stable between 1993 and 2019.

⁸For missing years (1994-2001), we assume that the extent of the under-representation of top incomes in survey data has evolved linearly, that is, we create synthetic income tax tabulations by linearly interpolating the correction by percentile observed in 1993 and 2002.

Combination of Survey and Tax Data with Macroeconomic Aggregates

After combining surveys with tax data, we rescale reported household income components to macro totals, and distribute components of the net national income that are not directly received by individuals.

First, we proportionally scale up household income components to their corresponding totals reported in the national accounts:

- Gross wages proportionally to compensation of employees.
- Self-employment/business income proportionally to mixed income (excluding rental income, see section [A](#))
- Rental income proportionally to total rents paid by households
- Interest income proportionally to total interest received by households
- Dividends proportionally to total dividends received by households

Second, we distribute unreported income components proportionally to proxy variables available in surveys:

- Imputed rents proportionally to the reported market value of the home of owner-occupiers
- Property income attributed to insurance holders and pension entitlements proportionally to the value of pension and life insurance assets
- Interest paid by households proportionally to the factor income of debtors
- Private corporate undistributed profits proportionally to directly and indirectly held stock ownership
- Government primary income and other remaining national income components proportionally to factor income

C. Distribution of Taxes and Transfers

Pension and Unemployment Systems

Pension and unemployment contributions and benefits are recorded in income surveys, so we distribute macro aggregates proportionally to values reported by respondents. In order to reach

pretax national income, we distribute 50% of the deficit or surplus of each system proportionally to contributions paid, and 50% proportionally to benefits received.

Taxes

Personal Income Tax We microsimulate the personal income tax every year from 1993 to 2019. To do so, we first collect data on taxable income thresholds, marginal tax rates, and rebates at each income level from various reports published by the South African Revenue Service. We then apply the corresponding rules in the microdata to calculate the tax burden of each individual. Because we have calibrated top taxable incomes directly on income tax tabulations (see section B), the estimates of total personal income tax revenue derived from microsimulation match almost perfectly actual revenue statistics. We close the residual gap between micro and macro estimates by proportionally rescaling the income tax burden of each individual.

Dividends Tax we distribute the dividends tax proportionally to dividends reported in income surveys.

Corporate Income Tax we distribute the corporate income tax proportionally to equity ownership, including both directly held equity and equity held indirectly through pension funds (see [Chatterjee, Czajka, and Gethin, 2022](#)).

Skills Development Levy The Skills Development Levy (SDL) is a 1% additional levy paid by wage earners who already contribute to the Unemployment Insurance Fund. We simulate it following this rule, and proportionally rescale the total to match total SDL revenue throughout the period.

Other Direct Taxes Other direct taxes include a number of minor taxes and levies, which have represented less than 1% of national income from 1993 to 2019. We distribute them proportionally to pretax income.

Transfer Duties The Transfer Duty is a tax levied on the value of properties acquired by individuals in South Africa. In the absence of information on property transactions, we distribute it proportionally to housing wealth (including both owner-occupied and tenant-occupied housing: see [Chatterjee, Czajka, and Gethin, 2022](#)).

Securities Transfer Tax The Securities Transfer Tax is a small tax that applies to the purchase and transfers of listed and unlisted securities. We distribute it proportionally to equity ownership.

Estate Duty and Donations Tax The Estate Duty and the Donations Tax are taxes on inheritance. In the absence of data on these transactions, we distribute them proportionally to total household wealth.

Value Added Tax we distribute total VAT revenue proportionally to household consumption expenditure, excluding both VAT-exempt goods and goods purchased on the informal market. Following the tax legislation, we directly identify VAT-exempt goods in income surveys and exclude them from taxable consumption. To identify goods purchased on the informal market, we derive a profile of informal consumption by income rank using the 2010 Income and Expenditure Survey, which reports the type of store at which the household purchased different kinds of goods. We extrapolate this profile to all years, assuming it has remained constant over the period. Expenditure in the informal sector is very small in South Africa, so that accounting for informality only has a negligible impact on the estimated distributional incidence of indirect taxation.

General Fuel Levy The General Fuel Levy is an excise tax charged on petroleum products. We distribute it proportionally to total transport expenditure reported by households in income surveys.

Other Excise Taxes Other excise duties mainly consist in excises applied to alcohol and tobacco products. In the absence of data on the decomposition of these taxes category by type of product, we distribute total revenue from non-GFL excises proportionally to combined alcohol and tobacco expenditure, as reported in income surveys.

Other Taxes on Goods and Services Other taxes on goods and services include a number of other small taxes, which have represented less than 0.5% of national income from 1993 to 2019. We distribute them proportionally to overall consumption expenditure.

Taxes on International Trade Import duties are effectively paid by households consuming a greater proportion of goods imported from abroad. Accordingly, we distribute taxes on international trade proportionally to import-intensive household expenditure, which we estimate using input-output tables available from the OECD (2005-2015).

Other Taxes Other taxes consist in a number of other small taxes and levies such as stamp duties. They have represented less than 0.5% of national income since 1993. We distribute them proportionally to pretax income.

Other Government Revenue we distribute all other government revenue, including non-tax revenue, proportionally to pretax income, so as to match total consolidated general government revenue in South Africa throughout the 1993-2019 period.

Social Protection

Social protection expenditure in South Africa mainly consists in the old age grant, the disability grant, the child support grant, other small cash transfers, and other social protection expenditure.

Old Age Grant The old age grant is a means-tested benefit paid to South African citizens who are 60 years or older. Old age grant beneficiaries are directly reported in income surveys, but their number is slightly below that reported in administrative data sources, suggesting a tendency to under-report. To correct this bias and ensure that my microfile matches both the true number of beneficiaries and total expenditure on the grant as reported in government budgets, we impute additional beneficiaries in two steps. First, we estimate the probability of surveyed individuals to receive the grant using a saturated linear probability model with the following explanatory variables: pretax income percentile, household expenditure percentile, gender, age, race, province or residence, and rural-urban location. Second, we rank individuals according to the predicted probability to receive the grant, and recursively allocate additional grants to those individuals most likely to receive it, until reaching the true number of beneficiaries every year from 1993 to 2019.

Disability Grant The disability grant is a means-tested benefit given to South African citizens who have a physical or mental disability that makes them unfit to work for a period of longer than six months. As in the case of the old age grant, it is reported in income surveys. We follow the same two-step strategy to impute additional beneficiaries when necessary, so as to match administrative statistics on both number of beneficiaries and total grant expenditure.

Child Support Grant The child support grant is a means-tested benefit given to low-income South African families to assist parents with the costs of the basic needs of their children. As in the case of the old age and disability grants, it is reported in income surveys. We follow the

same imputation strategy as for these two grants, so as to match administrative statistics on both number of beneficiaries and total grant expenditure. The child support grant was first implemented in 1998, so we set grant expenditure and beneficiaries to zero before that year.

Other Social Grants Other small cash grants in South Africa include the foster care grant, the care dependency grant, the grant-in-aid, and social relief. We distribute them proportionally to their values reported in income surveys.⁹

Other Social Protection Expenditure Other social protection expenditure mainly consists in “provincial social development” expenditure, which brings together a large number of heterogeneous subnational policies targeted to poor households. These include, for instance, projects dedicated to reducing HIV prevalence, supporting disabled persons, providing centers for the treatment and prevention of drug abuse, or developing services aimed to prevent violence against women and children. In the absence of precise information on who benefits from each of these policies, we distribute other social protection expenditure proportionally to total social grants received.

Other Government Transfers

See [Gethin \(2023\)](#).

⁹Most income surveys do not report receipts from these grants separately, so we derive an aggregate for “other social grants” in each survey and distribute total expenditure on these grants proportionally to this aggregate.

Bibliography

- Ali, Merima, Abdulaziz Shifa, Abebe Shimeles, and Firew B Woldeyes (2015). “Information technology and fiscal capacity in a developing country: evidence from Ethiopia”. In.
- Allingham, Michael G and Agnar Sandmo (1972). “Income Tax Evasion: A Theoretical Analysis”. In: *Journal of public economics* 1.3-4, pp. 323–338.
- Almunia, Miguel, David J Henning, Justine Knebelmann, Dorothy Nakymbadde, and Lin Tian (2023). *Leveraging Trading Networks to Improve Tax Compliance: Experimental Evidence from Uganda*. Centre for Economic Policy Research.
- Almunia, Miguel, Jonas Hjort, Justine Knebelmann, and Lin Tian (2022). “Strategic or Confused Firms? Evidence From “Missing” Transactions in Uganda”. In: *Review of Economics and Statistics*, pp. 1–35.
- Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman (2019). “Tax Evasion and Inequality”. In: *American Economic Review* 109.6, pp. 2073–2103.
- Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman (2018). “Who owns the wealth in tax havens? Macro evidence and implications for global inequality”. In: *Journal of Public Economics* 162.1, pp. 89–100.
- Alvaredo, A., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2018a). *World Inequality Report 2018*. Harvard University Press.
- Alvaredo, F., A. B. Atkinson, T. Blanchet, L. Chancel, L. Bauluz, M. Fisher-Post, I. Flores, B. Garbinti, J. Goupille-Lebret, C. Martínez-Toledano, M. Morgan, T. Neef, T. Piketty, A.-S. Robillard, E. Saez, L. Yang, and G. Zucman (2020). “Distributional National Accounts Guidelines: Methods and Concepts Used in the World Inequality Database”. WID.world.
- Alvaredo, F. and Anthony B. Atkinson (2022). “Top Incomes in South Africa in the Twentieth Century”. In: *Cliometrica* 16.1, pp. 477–546.
- Alvaredo, Facundo, Anthony Atkinson, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman (2016). “Distributional National Accounts (DINA) guidelines: Concepts and methods used in WID.world”. WID.world Working Paper 2016/2.
- Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman (2018b). *World inequality report 2018*. Belknap Press.

- Amine, Razan and Fabrizio Santoro (2023). *Mandating Digital Tax Tools as a Response to Covid: Evidence from Eswatini*. Tech. rep.
- Antinyan, Armenak and Zareh Asatryan (2019). “Nudging for Tax Compliance: A Meta-Analysis”. In: *ZEW-Centre for European Economic Research Discussion Paper* 19-055.
- Atkinson, Anthony Barnes and Allan James Harrison (1974). “Wealth distribution and investment income in Britain”. In: *Review of Income and Wealth* 20.2, pp. 125–142.
- Bachas, P, L. Gadenne, and A. Jensen (2020a). “Informality, Consumption Taxes and Redistribution”. NBER Working Paper No. 27429.
- Bachas, Pierre, Anne Brockmeyer, Alipio Ferreira, and Bassirou Sarr (2021). *How to Target Enforcement at Scale? Evidence From Tax Audits in Senegal*. Tech. rep. Working Paper, July.
- Bachas, Pierre, Lucie Gadenne, and Anders Jensen (2020b). *Informality, Consumption Taxes and Redistribution*. Tech. rep. National Bureau of Economic Research.
- Bachas, Pierre Jean, Leo Czajka, Aissatou Diallo, Justine Simone Cecile Knebelmann, and Anaëlle Binta Toure (2023). “Strengthening Personal Income Taxation in Senegal-Case Study”. In.
- Bassier, I., J. Budlender, R. Zizzamia, M. Leibbrandt, and V. Ranchhod (2020). “Locked Down and Locked Out: Repurposing Social Assistance as Emergency Relief to Informal Workers”. SALDRU Working Paper 261.
- Bassier, I. and I. Woolard (2018). “Exclusive Growth: Rapidly Increasing Top Incomes Amidst Low National Growth in South Africa”. REDI3x3 Working Paper 47.
- Beer, B. de and M. Kock (2017). “The further development of sectoral financial balance sheets for South Africa’s integrated economic accounts”. In: *South African Reserve Bank Quarterly Bulletin*.
- Bellon, Matthieu, Era Dabla-Norris, Salma Khalid, and Frederico Lima (2022). “Digitalization to improve tax compliance: Evidence from VAT e-Invoicing in Peru”. In: *Journal of Public Economics* 210, p. 104661.
- Berg, Servaas and Megan Louw (2005). “Changing patterns of South African income distribution: towards time series estimates of distribution and poverty”. In: *South African Journal of Economics* 72.3, pp. 546–572.
- Besley, Timothy (2020). “State Capacity, Reciprocity, and the Social Contract”. In: *Econometrica* 88.4, pp. 1307–1335.
- Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu (2022). “Bureaucracy and development”. In: *Annual Review of Economics* 14, pp. 397–424.
- Besley, Timothy and Torsten Persson (2014). “Why Do Developing Countries Tax So Little?” In: *Journal of economic perspectives* 28.4, pp. 99–120.

- Best, Michael, Jawad Shah, and Mazhar Waseem (2021a). “Detection Without Deterrence: Long-Run Effects of Tax Audit on Firm Behavior”. In.
- (2021b). *The Deterrence Value of Tax Audit: Estimates from a Randomised Audit Program*. Tech. rep. mimeo.
- Bharti, Nitin Kumar (2018). “Wealth Inequality, Class and Caste in India, 1961-2012”. In: *WID.world Working Paper* 2018/14.
- Bhorat, H., S. Goga, and D. Tseng (2013). “Unemployment Insurance in South Africa: A Descriptive Overview of Claimants and Claims”. University of Cape Town Working Paper.
- Blanchet, T. (2023). “Uncovering the Dynamics of the Wealth Distribution”. Working Paper.
- Blanchet, T., I. Flores, and M. Morgan (2018a). “The Weight of the Rich: Improving Surveys Using Tax Data”. *WID.world Working Paper*.
- Blanchet, T., J. Fournier, and T. Piketty (2017a). “Generalized Pareto Curves: Theory and Applications”. *WID.world Working Paper*.
- Blanchet, Thomas (2016). “Wealth inequality in Europe and in the United States: estimations from surveys, national accounts and wealth rankings”. Paris School of Economics Master Thesis.
- Blanchet, Thomas, Lucas Chancel, and Amory Gethin (2019). “How Unequal Is Europe? Evidence From Distributional National Accounts, 1980-2017”. *WID.world Working Paper* 2019/6.
- (2022). “Why Is Europe More Equal Than the United States?” In: *American Economic Journal: Applied Economics* 14.4, pp. 1–40.
- Blanchet, Thomas, Ignacio Flores, and Marc Morgan (2018b). “The Weight of the Rich: Improving Surveys Using Tax Data”.
- Blanchet, Thomas, Juliette Fournier, and Thomas Piketty (2017b). “Generalized Pareto Curves: Theory and Applications”.
- Bonnet, Florence, Joann Vanek, Martha Chen, et al. (2019). “Women and Men in the Informal Economy: A Statistical Brief”. In: *International Labour Office, Geneva* 20.
- Bozio, A., J. Goupille-Lebret, G. Malka, and T. Piketty (2020). “Predistribution vs. Redistribution: Evidence from France and the U.S.” *CEPR Working Paper*.
- Brockmeyer, Anne and Marco Hernandez (2022). “Taxation, Information and Withholding: Evidence from Costa Rica”. In.
- Brockmeyer, Anne and Magaly Sáenz Somarriba (2022). “Electronic Payment Technology and Tax Compliance: Evidence from Uruguay’s Financial Inclusion Reform”. In.
- Chancel, Lucas, Denis Cogneau, Amory Gethin, Alix Myczkowski, and Anne-Sophie Robilliard (2023). “Income Inequality in Africa, 1990-2019: Measurement, Patterns, Determinants”. In: *World Development* 163.1, pp. 1061–1084.

- Chatterjee, Aroop (2019). “Measuring wealth inequality in South Africa: An agenda”. In: *Development Southern Africa* 36.6, pp. 839–859.
- Chatterjee, Aroop, Léo Czajka, and Amory Gethin (2022). “Wealth Inequality in South Africa, 1993-2017”. In: *World Bank Economic Review* 36.1, pp. 19–36.
- Chen, Jiafeng and Jonathan Roth (2023). “Logs With Zeros? Some Problems and Solutions”. In: *arXiv preprint arXiv:2212.06080*.
- Clark, Julia, Anna Metz, and Claire Casher (2022). “ID4D Global Dataset 2021, Volume 1”. In.
- Clay, Sir Henry (1925). *The distribution of capital in England and Wales*. John Heywood.
- Czajka, Leo, Florence Kondylis, Bassirou Sarr, and Mattea Stein (2022). “Data Management at the Senegalese Tax Authority: Insights from a Long-term Research Collaboration”. In.
- Czajka, Leo and Bassirou Sarr (2023). “Firm Size and Mobility: Descriptive Evidence From the Senegalese Formal Sector, 2008-2020”. In: *Research Note*.
- Czajka, Léo (2017). “Income Inequality in Côte d’Ivoire: 1985-2014”. WID.world Working Papers n°2017/8.
- Daniels, George William and Harry Campion (1936). *The distribution of national capital*. Manchester University Press.
- Daniels, R. and T. Augustine (2016). “The measurement and distribution of household wealth in South Africa using the National Income Dynamics Study (NIDS) Wave 4”. NIDS Discussion Paper 2016/10.
- Das, Satadru, Lucie Gadenne, Tushar Nandi, and Ross Warwick (2023). “Does going cashless make you tax-rich? Evidence from India’s demonetization experiment”. In: *Journal of Public Economics* 224, p. 104907.
- Davis Tax Committee (2018). *Report on the Feasibility of a Wealth Tax in South Africa for the Minister of Finance*. Report for the Minister of Finance of South Africa.
- De Neve, Jan-Emmanuel, Clement Imbert, Johannes Spinnewijn, Teodora Tsankova, and Maarten Luts (2021). “How to Improve Tax Compliance? Evidence From Population-Wide Experiments in Belgium”. In: *Journal of Political Economy* 129.5, pp. 1425–1463.
- Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. World Bank Publications.
- Dettling, L. J., J. Devlin-Foltz, J. Krimmel, S. J. Pack, and J. P. Thompson (2015). “Comparing Micro and Macro Sources for Household Accounts in the United States: Evidence from the Survey of Consumer Finances”. Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs of the Federal Reserve Board.
- Duflo, E. (2000). “Child Health and Household Resources in South Africa: Evidence from the Old Age Pension Program”. In: *American Economic Review* 90.2, pp. 393–398.

- Dutronc-Postel, Paul and Alessandro Tondini (2023). “Large means-tested pensions with informal labor markets: Evidence from south africa”. In: *Economic Development and Cultural Change* 71.3, pp. 1149–1178.
- Ebrahim, Amina and Chris Axelson (2019a). “The creation of an individual level panel using administrative tax microdata in South Africa”. UNU-WIDER Working Paper 2019/27.
- (2019b). “The creation of an individual level panel using administrative tax microdata in South Africa”. UNU-WIDER Working Paper 2019/27.
- Garbinti, B., J. Goupille-Lebret, and T. Piketty (2018a). “Income inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA)”. In: *Journal of Public Economics* 162.1, pp. 63–77.
- (2018b). “Income Inequality in France, 1900-2014: Evidence from Distributional National Accounts (DINA)”. In: *Journal of Public Economics* 162, pp. 63–77.
- Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty (2017). “Accounting for wealth inequality dynamics: Methods, estimates and simulations for France (1800-2014)”. WID.world Working Paper 2016/5.
- Garriga, Pablo and Dario Tortarolo (2022). *Firms as Tax Collectors*. Institute for Fiscal Studies.
- Gethin, Amory (2023). “Who Benefits from Public Goods? Evidence from South Africa”. Working Paper.
- Hallsworth, Michael (2014). “The Use of Field Experiments to Increase Tax Compliance”. In: *Oxford Review of Economic Policy* 30.4, pp. 658–679.
- Henriques, A. M. and J. W. Hsu (2014). “Analysis of Wealth Using Micro- and Macrodata: A Comparison of the Survey of Consumer Finances and Flow of Funds Accounts”. In: *Measuring Economic Sustainability and Progress*. Ed. by D. W. Jorgenson, J. S. Landefeld, and P. Schreyer. University of Chicago Press. Chap. 9, pp. 245 –274.
- Hundenborn, Janina, Ingrid Woolard, and Jon Jellema (2018). “The effect of top incomes on inequality in South Africa”. UNU-WIDER Working Paper 2018/90.
- IMF (2023). “Requests for an Extended Arrangement Under The Extended Fund Facility, an Arrangement Under the Extended Credit Facility, and An Arrangement Under the Resilience And Sustainability Facility”. In: *IMF Country Report* 23/250.
- Inchauste, G., N. Lustig, M. Maboshe, C. Purfield, and I. Woolard (2015). “The Distributional Impact of Fiscal Policy in South Africa”. World Bank Policy Research Working Paper.
- Jensen, Anders (2022). “Employment Structure and the Rise of the Modern Tax System”. In: *American Economic Review* 112.1, pp. 213–34.
- Kasper, Matthias and Matthew D Rablen (2023). “Tax compliance after an audit: Higher or lower?” In: *Journal of Economic Behavior & Organization* 207, pp. 157–171.

- Kleven, Henrik, Camille Landais, Mathilde Muñoz, and Stefanie Stantcheva (2020). “Taxation and Migration: Evidence and Policy Implications”. In: *Journal of Economic Perspectives* 34.2, pp. 119–142.
- Kleven, Henrik Jacobsen, Martin B Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez (2011). “Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark”. In: *Econometrica* 79.3, pp. 651–692.
- Korinek, Anton, Johan A. Mistiaen, and Martin Ravallion (2006). “Survey nonresponse and the distribution of income”. In: *Journal of Economic Inequality* 4.1, pp. 33–55.
- Kotsogiannis, Christos, Luca Salvadori, John Karangwa, and Theonille Mukamana (2024). “Do tax audits have a dynamic impact? Evidence from corporate income tax administrative data”. In: *Journal of Development Economics* 170, p. 103292.
- Langley, Kathleen M. (1950). “The distribution of capital in private hands in 1936–1938 and 1946–1947”. In: *Bulletin of the Oxford University Institute of Economics & Statistics* 12.12, pp. 339–350.
- Lee, David S et al. (2009). “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects”. In: *Review of Economic Studies* 76.3, pp. 1071–1102.
- Leibbrandt, M., I. Woolard, A. Finn, and J. Argent (2010). *Trends in South African Income Distribution and Poverty since the Fall of Apartheid*. OECD Social, Employment and Migration Working Papers 101.
- Leibbrandt, Murray, Ingrid Woolard, and Christopher Woolard (2009). “Poverty and inequality dynamics in South Africa: Post-apartheid developments in the light of the long-run legacy”. In: *South African Economic Policy under Democracy*. Ed. by J. Aron, B. Kahn, and G. Kingdon. Oxford: Oxford University Press. Chap. 10, pp. 270–300.
- Londoño-Vélez, Juliana and Javier Ávila-Mahecha (2021). “Enforcing Wealth Taxes in the Developing World: Quasi-Experimental Evidence From Colombia”. In: *American Economic Review: Insights* 3.2, pp. 131–148.
- Luttmer, Erzo FP and Monica Singhal (2014). “Tax morale”. In: *Journal of economic perspectives* 28.4, pp. 149–168.
- Maboshe, M. and I. Woolard (2018). “Revisiting the Impact of Direct Taxes and Transfers on Poverty and Inequality in South Africa”. WIDER Working Paper 2018/79.
- Mallet, Bernard (1908). “A Method of Estimating Capital Wealth from the Estate Duty Statistics”. In: *Journal of the Royal Statistical Society* 71.1, pp. 65–101.
- Mascagni, Giulia, Roel Dom, Fabrizio Santoro, and Denis Mukama (2023). “The VAT in Practice: Equity, Enforcement, and Complexity”. In: *International Tax and Public Finance* 30.2, pp. 525–563.

- Mascagni, Giulia, Andualem T Mengistu, and Firew B Woldeyes (2021). “Can ICTs increase tax compliance? Evidence on taxpayer responses to technological innovation in Ethiopia”. In: *Journal of Economic Behavior & Organization* 189, pp. 172–193.
- Mascagni, Giulia and Christopher Nell (2022). “Tax Compliance in Rwanda: Evidence From a Message Field Experiment”. In: *Economic Development and Cultural Change* 70.2, pp. 587–623.
- Mascagni, Giulia, Fabrizio Santoro, Denis Mukama, John Karangwa, and Napthal Hakizimana (2022). “Active Ghosts: Nil-Filing in Rwanda”. In: *World Development* 152, p. 105806.
- Mbewe, Samson and Ingrid Woolard (2016). “Cross-sectional features of wealth inequality in South Africa: Evidence from the National Income Dynamics Study”. NIDS Discussion Paper 2016/12.
- Mian, A., L. Straub, and A. Sufi (2021). “Indebted Demand”. In: *The Quarterly Journal of Economics*.
- Milanez, Anna (2017). “Legal Tax Liability, Legal Remittance Responsibility and Tax Incidence: Three Dimensions of Business Taxation”. In.
- Miller, Douglas L (2023). “An Introductory Guide to Event Study Models”. In: *Journal of Economic Perspectives* 37.2, pp. 203–230.
- Muellbauer, John and Janine Aron (1999). “Estimates of personal sector wealth for South Africa”. CEPR Discussion Paper No. 4646.
- Novokmet, Filip, Thomas Piketty, and Gabriel Zucman (2018). “From Soviets to oligarchs: inequality and property in Russia 1905-2016”. In: *The Journal of Economic Inequality* 16.2. Publisher: Springer, pp. 189–223.
- OCDE (2020). “How to Construct a Global Business Register?” In.
- (2022). “Revenue Statistics in Africa 1990-2020”. In.
- Ohnsorge, Franziska and Shu Yu (2022). *The Long Shadow of Informality: Challenges and Policies*. World Bank Publications.
- Okunogbe, Oyebola and Victor Pouliquen (2022). “Technology, taxation, and corruption: evidence from the introduction of electronic tax filing”. In: *American Economic Journal: Economic Policy* 14.1, pp. 341–372.
- Orthofer, Anna (2015). “Private wealth in a developing country: A South African perspective on Piketty”. Economic Research Southern Africa Working Paper no. 564.
- (2016). “Wealth inequality in South Africa: Evidence from survey and tax data”. REDI3x3 Working Paper 15.
- Pauw, Kalie and Liberty Mncube (2007). “The Impact of Growth and Redistribution on Poverty and Inequality in South Africa”. DPRU Working Paper 07/126.
- Piketty, T. (2019). *Capital and Ideology*. Harvard University Press.

- Piketty, T., E. Saez, and G. Zucman (2018). “Distributional National Accounts: Methods and Estimates for the United States”. In: *The Quarterly Journal of Economics* 133.2, pp. 553–609.
- Piketty, T., L. Yang, and G. Zucman (2019a). “Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015”. In: *American Economic Review* 109.7, pp. 2469–2496.
- Piketty, Thomas, Li Yang, and Gabriel Zucman (2019b). “Capital accumulation, private property, and rising inequality in China, 1978–2015”. In: *American Economic Review* 109.7, pp. 2469–96.
- Piketty, Thomas and Gabriel Zucman (2014). “Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010”. In: *The Quarterly Journal of Economics* 129.3, pp. 1255–1310.
- Rasul, Imran, Daniel Rogger, and Martin J Williams (2018). “Management and bureaucratic effectiveness: Evidence from the ghanaian civil service”. In: *World Bank Policy Research Working Paper* 8595.
- Roth, Jonathan, Pedro HC Sant’Anna, Alyssa Bilinski, and John Poe (2023). “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature”. In: *Journal of Econometrics*.
- Saez, E. and G. Zucman (2016a). “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data”. In: *The Quarterly Journal of Economics* 131.2, pp. 519–578.
- Saez, Emmanuel and Gabriel Zucman (2016b). “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data”. In: *The Quarterly Journal of Economics* 131.2, pp. 519–578.
- (2020). *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*. WW Norton and Co.
- Santoro, Fabrizio, Edward Groening, Winnie Mdluli, and Mbongeni Shongwe (2020). “How to best nudge taxpayers? The impact of a tailored letter experiment in Eswatini”. In.
- Scarpini, Celeste, Oyebola Okunogbe, and Fabrizio Santoro (2023). “The Promise and Limitations of Information Technology for Tax Mobilisation”. In: *Institute of Development Studies, International Centre for Tax and Development Working Papers* 17871.
- Shimeles, Abebe, Daniel Zerfu Gurara, and Firew Woldeyes (2017). “Taxman’s Dilemma: Coercion or Persuasion? Evidence From a Randomized Field Experiment in Ethiopia”. In: *American Economic Review* 107.5, pp. 420–424.
- Slemrod, Joel (2019). “Tax Compliance and Enforcement”. In: *Journal of Economic Literature* 57.4, pp. 904–954.
- Slemrod, Joel and Tejaswi Velayudhan (2018). “Do Firms Remit at Least 85% Of Tax Everywhere? New Evidence From India”. In: *Journal of Tax Administration* 4.1, pp. 24–37.

- South African Reserve Bank (2019). *Tax Chronology of South Africa: 1979–2019*. Technical Report.
- Tondini, Alessandro (2022). “The Lasting Labor-Market Effects of Cash Transfers: Evidence from South Africa’s Child Support Grant”. In: *The World Bank Economic Review* 36.4, pp. 934–954.
- United Nations (2009). *System of National Accounts 2008*. United Nations, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, and World Bank joint publication.
- Van Tonder, Johann, Carel Van Aardt, and Bernadene De Clercq (2018). “South African Household Wealth Index 2017”. Momentum/Unisa Report.
- Weber, Max (1922). *Economy and Society: An Outline of Interpretive Sociology*. Vol. 1. University of California Press.
- World Bank (2020). *South Africa. Poverty and Equity Brief*.
- World-Bank (2021). *World Development Report 2021: Data for Better Lives*.
- World Bank, World (2016). *World development report 2016: Digital dividends*. World Bank Publications.
- Young, A. (2013). “Inequality, the Urban-Rural Gap, and Migration”. In: *The Quarterly Journal of Economics* 128.4, pp. 1727–1785.
- Ytterberg, Alan V and James P Weller (2010). “Managing Family Wealth Through a Private Trust Company”. In: *ACTEC Law Journal* 36.1, pp. 501–518.
- Yu, Derek (2005). “The comparability of Income and Expenditure Surveys 1995, 2000 and 2005/2006”. Working Papers 11/2008, Stellenbosch University, Department of Economics.
- Zucman, Gabriel (2019). “Global Wealth Inequality”. In: *Annual Review of Economics* 11.1, pp. 109–138.