

Como Usar Reportes de Terceros para Aumentar la Recaudación Fiscal: Evidencia Experimental en Paraguay

How to Use Third Party Reports to Collect Taxes: Experimental Evidence from Paraguay*

Michael Carlos Best Florian Grosset-Touba Gastón Pierri Evan Sadler
Panos Toulis

22 Feb 2025

Abstract

Estudiamos la efectividad de estrategias de fiscalización que aprovechan datos de terceros para detectar y abordar la subdeclaración fiscal. En colaboración con la administración tributaria de Paraguay, desarrollamos una herramienta escalable para utilizar informes de terceros y así identificar discrepancias en las ventas declaradas por contribuyentes sujetos al IVA. A través de un ensayo controlado aleatorizado, encontramos que notificar a los contribuyentes sobre discrepancias entre sus informes y los de sus clientes es altamente efectivo. El 65% de los contribuyentes notificados corrigieron sus declaraciones dentro de los nueve meses, con la mayoría de los cambios ocurriendo en las primeras dos semanas. Las notificaciones llevaron a un aumento promedio del 900% en las ventas declaradas y a un 55% más de probabilidades de enmiendas. A diferencia de estudios previos, esta intervención generó incrementos significativos en las obligaciones fiscales declaradas y en los pagos de impuestos. Cada notificación produjo beneficios netos por un valor de PYG 936.535 (USD 127) por contribuyente, logrando una tasa de retorno del 37% sobre los costos administrativos en 12 semanas. El estudio también resalta los posibles efectos indirectos, donde las empresas notificadas influyen en el comportamiento de sus socios comerciales, y subraya la necesidad de estrategias de focalización óptimas debido a las limitaciones de recursos.

*Best: Columbia University, BREAD, CEPR, IFS & NBER. michael.best@columbia.edu. Grosset-Touba: CREST, ENSAE Paris, Institut Polytechnique de Paris. florian.grosset@ensae.fr. Pierri: BID/IADB. gpierri@iadb.org. Sadler: Columbia University. es3668@columbia.edu. Toulis: University of Chicago. panos.toulis@chicagobooth.edu. We are grateful to Catalina Gomez Colomer, Daniel Diaz, Krzysztof Zaremba, Veronica Perez, Jack Rosetti, and Akanksha Vardani for excellent research assistance. This paper is based on a partnership with the *Dirección Nacional de Ingresos Tributarios* (DNIT), who implemented the experiment and provided access to the administrative tax data. We are particularly grateful to Juan Vallena, Robert Blanco, and José Monreale for their support. This study has been funded by the International Growth Centre and the National Science Foundation. All remaining errors are ours.

Investigaciones futuras explorarán estos efectos indirectos y métodos de focalización para maximizar el cumplimiento tributario y la eficiencia de las políticas.

We study the effectiveness of enforcement strategies that leverage third-party data to detect and address tax underreporting. Partnering with Paraguay's tax administration, we developed a scalable tool to use third-party reports to identify discrepancies in VAT-liable businesses' reported sales. Through a randomized control trial, we find that notifying taxpayers of discrepancies between their reports and their clients' is highly effective. 65% of notified taxpayers amended their returns within nine months, with most changes occurring in the first two weeks. Overall, the notifications increased firms' likelihood of filing an amendment by 55%, and reported sales by 900% on average. Unlike previous studies, this intervention led to meaningful increases in taxpayers' reported liabilities and tax remittances. Each notification delivered net benefits worth PYG 936,535 (USD 127) per taxpayer, achieving a 37% rate of return on administrative costs in 12 weeks. The study also highlights the potential spillover effects, where notified taxpayers influence their trading partners' behavior, and emphasizes the need for optimal targeting strategies due to resource constraints. Future research will explore these spillovers and targeting methods to maximize tax compliance and policy efficiency.

Contents

A	Como Usar Reportes de Terceros para Aumentar la Recaudación Fiscal: Evidencia Experimental en Paraguay	5
1	Introducción	5
2	Contexto y Diseño Experimental	7
2.1	Colaboración de Investigación	7
2.2	Descubriendo Discrepancias en las Declaraciones	8
2.3	Intervención	9
2.4	Cronograma	11
2.5	Oleadas 1-2: Aleatorización por Clúster	11
2.6	Oleadas 3-4: Diseño <i>Split-Graph</i>	12
2.7	Balance	14
3	Evidencia Descriptiva	16
4	Efectos Causales	18
4.1	Efectos en las Rectificaciones (estimación basada en el diseño)	18
4.2	Heterogeneidad de los Efectos entre Contribuyentes (estimación basada en modelos)	22
5	Discusión y Próximos Pasos	24
5.1	Análisis de Costos y Beneficios	24
5.2	Efectos Indirectos del Control Tributario	26
5.3	Implicaciones de Política para la Focalización de los Controles Masivos	27
6	Conclusión	29
B	How to Use Third Party Reports to Collect Taxes: Experimental Evidence from Paraguay	31
I	Introduction	31
II	Context & Experimental Design	33
II.i	Project Partnership	33
II.ii	Discovering Reporting Discrepancies	34
II.iii	Intervention	35
II.iv	Timeline	36
II.v	Waves 1-2: Cluster Randomization	36

II.vi Waves 3-4: <i>Split-Graph</i> Design	37
II.vii Balance	39
III Descriptive Evidence	40
IV Causal Impacts	43
IV.i Effects on Amendments (design-based approach)	43
IV.ii Heterogeneity of Effects Across Taxpayers (model-based approach)	47
V Discussion and Next Steps	48
V.i Cost Benefit Analysis	48
V.ii Spillovers in Tax Enforcement	50
V.iii Policy Implications for Targeting	51
VI Conclusion	53

Part A

Como Usar Reportes de Terceros para Aumentar la Recaudación Fiscal: Evidencia Experimental en Paraguay

1 Introducción

Un estado funcional constituye la base del desarrollo económico (Page & Pande, 2018). La función principal del estado es proporcionar bienes y servicios públicos, lo cual debe financiarse mediante la recaudación de impuestos. Por lo tanto, comprender cómo pasar de recaudar el 10-15% del PIB en impuestos como es común en los países de ingresos bajos y medios a recaudar el 30-40% del PIB puede considerarse una de las preguntas centrales en el desarrollo económico (Besley & Persson, 2009, 2013).

Los códigos tributarios no difieren drásticamente entre países. Sin embargo, las economías de los países de ingresos bajos y medios se caracterizan por grandes sectores informales y una evasión fiscal generalizada. Esto sugiere que encontrar formas de lograr un mayor cumplimiento tributario es una vía clave para aumentar la capacidad estatal en estos países (Gordon & Li, 2009; Besley & Persson, 2013). Un mecanismo ampliamente destacado para lograr un mayor cumplimiento es poner a disposición de la administración tributaria información de terceros sobre las obligaciones de los contribuyentes (Kleven *et al.*, 2011; Pomeranz, 2015). Sin embargo, se han planteado dudas sobre la eficacia de la información de terceros por sí sola para lograr un mayor cumplimiento (Carrillo *et al.*, 2017; Naritomi, 2019; Best *et al.*, 2023).

En este artículo, presentamos algunos resultados preliminares de una colaboración en curso con la administración tributaria de Paraguay para explorar el potencial de aprovechar los informes de terceros proporcionados por los clientes y proveedores de los contribuyentes para aumentar el cumplimiento del Impuesto al Valor Agregado (IVA) de Paraguay. El IVA presenta un mecanismo de crédito por facturación mediante el cual los contribuyentes informan tanto las ventas como las compras y pagan impuestos sobre la diferencia (valor agregado), creando incentivos opuestos para falsear la información: los compradores querrían sobrestimar sus compras, mientras que los vendedores querrían subestimar sus ventas Pomeranz (2015). Sin embargo, para que este mecanismo conduzca a una mayor fiscalización, la administración tributaria debe usar estos informes de terceros para verificar las declaraciones de los contribuyentes y actuar sobre cualquier discrepancia. Este proyecto explora la eficacia y el diseño de estas acciones. El objetivo general de la colaboración es desarrollar herramientas que aprovechen al máximo la gran cantidad de datos de terceros que pueden ser recolectados por la administración tributaria de un país de ingresos medianos como Paraguay, que cuenta con recursos administrativos y humanos limitados.

Nos basamos en trabajos existentes para desarrollar una herramienta fiscal simple y escalable que utilice los informes de terceros para detectar la posible subdeclaración de ventas gravadas por los contribuyentes sujetos al IVA (Carrillo *et al.*, 2017; Almunia *et al.*, 2024). Combinamos la posible subdeclaración detectada con una notificación de la autoridad tributaria, respaldada por una amenaza creíble de bloquear el número de identificación de los contribuyentes si no toman medidas para generar aumentos significativos en la percepción de los contribuyentes sobre la intensidad del escrutinio y la fiscalización de la autoridad tributaria. Seleccionamos aleatoriamente a los contribuyentes para recibir la notificación informándoles sobre las ventas que se sospecha han sido subdeclaradas y solicitándoles que enmienden las declaraciones fiscales relevantes para corregir la discrepancia con las compras reportadas por sus clientes en un plazo de diez días hábiles.

Estimar los efectos causales de estas notificaciones en los contribuyentes objetivo no es sencillo, debido al potencial de efectos indirectos. Enviar una notificación a una empresa puede afectar el comportamiento de esta, pero también el de sus socios comerciales, por ejemplo, al aumentar su percepción del escrutinio al que están sometidos o al cambiar sus decisiones estratégicas de evasión. Si este es el caso, las herramientas estándar de la microeconomía aplicada probablemente no funcionen bien. Por lo tanto, desarrollamos nuevos métodos para diseñar y analizar el experimento, mejor adaptados al contexto específico.

Presentamos seis hallazgos principales. Primero, nuestras notificaciones son altamente efectivas para inducir a los contribuyentes a enmendar sus declaraciones fiscales. El 65% de los contribuyentes notificados realizó una enmienda dentro de los nueve meses posteriores a la notificación. Estimamos que esto se debe casi en su totalidad a los efectos causales de las notificaciones, con efectos concentrados en las dos primeras semanas después de la notificación, pero que persisten hasta al menos dos meses después.

Segundo, las enmiendas inducidas por las notificaciones implicaron aumentos significativos en las ventas declaradas. Los contribuyentes tuvieron un 55% más de probabilidades de realizar una enmienda que aumentara las ventas declaradas. Estas ventas aumentaron en promedio un 900%, nuevamente concentradas en las primeras dos semanas después de las notificaciones.

Tercero, en consonancia con los hallazgos de Carrillo *et al.* (2017), las compras declaradas también aumentan. Sin embargo, a diferencia de Carrillo *et al.* (2017), observamos aumentos estadísticamente significativos y económicamente relevantes en las obligaciones fiscales declaradas por los contribuyentes.

Cuarto, los aumentos en las obligaciones fiscales declaradas no se compensan con mayores retenciones declaradas u otras deducciones fiscales. Más bien, hay aumentos estadísticamente y económicamente significativos en los impuestos remitidos a la autoridad tributaria. Quinto, para aquellos que no aumentan las remesas de impuestos, hay disminuciones estadísticamente y económicamente significativas en los créditos fiscales que la autoridad tributaria debe al contribuyente.

Sexto, combinando los resultados anteriores, estimamos que la intervención de notificaciones

masivas es altamente efectiva. Genera beneficios en forma de pagos adicionales de impuestos y créditos reducidos por un valor total de PYG 936.535 (USD 127) por contribuyente notificado.¹ Además, una estimación detallada de los costos administrativos de la intervención sugiere que las notificaciones son extremadamente rentables, con una tasa de retorno de aproximadamente el 37% en 12 semanas.

Este informe preliminar presenta solo una parte de los resultados de nuestra colaboración. En particular, nuestro trabajo en curso explora dos direcciones importantes y relacionadas: los efectos indirectos y la focalización de políticas. Enviar notificaciones a los contribuyentes tiene el potencial de influir en el comportamiento de los socios comerciales de los contribuyentes objetivo. Probaremos empíricamente si este es el caso, estimaremos el tamaño de estos efectos indirectos y proporcionaremos evidencia sobre qué los impulsa. Esto es necesario para evaluar con precisión los impactos generales de las auditorías fiscales. Nuestros resultados también ayudarán a mejorar la focalización de las auditorías: una cuestión clave en contextos con capacidad de implementación limitada. En nuestro caso, preparar y enviar las notificaciones es relativamente barato, pero su eficacia depende crucialmente de un seguimiento individual intenso por parte de los agentes fiscales. Por lo tanto, el número de notificaciones que se pueden enviar está directamente limitado por la disponibilidad de agentes fiscales. Saber cómo focalizar mejor las notificaciones es, por lo tanto, importante para maximizar sus beneficios. Propondremos diferentes tipos de reglas de focalización que tienen el potencial de mejorar los beneficios fiscales y evaluaremos empíricamente sus efectos.

Nuestro trabajo sobre efectos indirectos y focalización óptima de políticas contribuirá a la literatura académica activa en ambos temas. También será directamente relevante para los responsables de políticas, tanto en las autoridades fiscales como más allá: comprender los efectos indirectos de las intervenciones políticas es una cuestión central para políticas de salud, intervenciones activas en el mercado laboral, etc. Nuestros métodos y resultados serán el tema de futuros informes y sus correspondientes artículos académicos.

2 Contexto y Diseño Experimental

2.1 Colaboración de Investigación

Nuestro proyecto es una asociación con la administración tributaria de Paraguay. La informalidad y la evasión fiscal son barreras fundamentales para el desarrollo de la economía: el 65% de los trabajadores están empleados de manera informal y los ingresos tributarios representan el 10% del PIB (Banco Mundial, 2017). Por lo tanto, movilizar recursos adicionales para el gobierno es un desafío clave para el desarrollo. Como es común en todo el mundo, los principales retos para aumentar los ingresos fiscales son la existencia de un gran sector informal y la evasión fiscal generalizada (Besley & Persson, 2013). Esto hace que las acciones de cumplimiento por parte

¹Este cálculo valora los aumentos en los pagos de impuestos a su valor nominal y las reducciones en los créditos fiscales al 70% de su valor nominal.

de la administración tributaria para aumentar el cumplimiento fiscal sean un mecanismo clave para abordar este desafío.

Junto con la administración tributaria, desarrollamos un programa de fiscalizaciones masivas de escritorio para aprovechar los datos informados por terceros, los cuales se cruzan con las declaraciones juradas del impuesto al valor agregado (IVA) de los contribuyentes. Los contribuyentes presentan declaraciones mensuales en las que informan sobre sus ventas y compras totales, su obligación tributaria, los impuestos ya pagados/retenidos y el monto de impuesto adeudado o crédito a trasladar. Los grandes, medianos y algunos pequeños contribuyentes también presentan declaraciones informativas mensuales sobre todas las transacciones individuales que realizan con sus clientes y proveedores. Estos informes de terceros constituyen la base de nuestra intervención. El programa se alineó con el programa existente de fiscalizaciones masivas de escritorio de la administración tributaria, pero trabajamos en conjunto para enfocarlo en la verificación cruzada de los informes de terceros.

Para evaluar rigurosamente el impacto de las notificaciones, incorporamos una serie de ensayos controlados aleatorizados (RCTs, por sus siglas en inglés) dentro del programa de fiscalización, como se describe en las siguientes secciones.

2.2 Descubriendo Discrepancias en las Declaraciones

Nuestra intervención se basa en las discrepancias entre la declaración de impuestos de los contribuyentes y las declaraciones informativas de sus socios comerciales.

Para cada contribuyente focal i en cada mes m , calculamos el monto total que los contribuyentes que presentan declaraciones informativas sobre sus compras reportan haber adquirido del contribuyente focal. Este es su monto de ventas reportadas por terceros y_{im}^r . Luego, lo comparamos con lo que el contribuyente focal declara en su declaración de impuestos y_{im}^d . Dado que no todos los contribuyentes están obligados a presentar declaraciones informativas sobre sus compras, las ventas declaradas por el contribuyente focal siempre deberían ser al menos tan grandes como las ventas reportadas por terceros. Es decir, la *discrepancia* de declaración $d_{im} = y_{im}^r - y_{im}^d$ debería cumplir que $d_{im} \leq 0$. Siempre que esto no se cumpla, debe haber un error en la declaración, ya sea porque el contribuyente focal ha subdeclarado sus ventas, porque uno o más clientes han sobredeclarado sus compras o porque ha habido un error en una declaración.²

Al calcular las discrepancias de esta manera, es natural que existan diferencias. Dado que solo los contribuyentes más importantes (principalmente grandes y medianos) presentan informes de terceros, solo una parte de las ventas de los contribuyentes está cubierta por estos reportes. Como resultado, esperamos que la mayoría de las discrepancias sean negativas. La Figura 2 muestra la distribución de las discrepancias normalizadas (dado que las discrepancias

²Los errores comunes descubiertos durante el experimento incluyen importaciones/exportaciones declaradas con una tasa de cambio incorrecta, facturas que luego son canceladas debido a devoluciones de productos, ventas a crédito reportadas como liquidadas y ventas asignadas al mes incorrecto.

varían enormemente en tamaño, las normalizamos calculando $\tilde{d}_{im} = d_{im} / \frac{1}{2} (y_{im}^r + y_{im}^d)$.

El Panel A muestra la distribución general de las discrepancias normalizadas, eliminando los casos en los que las discrepancias son exactamente cero. Se observa que la mayoría de las discrepancias son negativas, pero también hay una cantidad significativa de discrepancias positivas, que podrían ser indicativas de errores en la declaración de impuestos. En el Panel B, se representan por separado las distribuciones de discrepancias estrictamente positivas para las ventas gravadas al 10% (la tasa estándar) y las ventas gravadas al 5% (la tasa reducida aplicada a ciertos bienes). En línea con la hipótesis de subdeclaración, se observa que la distribución de discrepancias en ventas gravadas al 10% presenta discrepancias medianas y grandes con mayor frecuencia, mientras que las gravadas al 5% tienden a ser menores y menos propensas a mostrar ventas completamente no declaradas (una menor concentración en 2).

En la Figura 3, exploramos la incidencia de discrepancias positivas entre los contribuyentes y a lo largo del tiempo. Para cada contribuyente, calculamos la proporción de meses en los que presentan una declaración de impuestos con discrepancias positivas en comparación con los reportes de terceros sobre sus ventas.

El Panel A muestra la distribución de esta propensión entre los contribuyentes, excluyendo el 59% de contribuyentes que nunca presentan discrepancias positivas.³ Se observa que hay muchos contribuyentes cuyas declaraciones siempre generan discrepancias positivas, pero también muchos que solo ocasionalmente presentan discrepancias. En el Panel B, se sigue el porcentaje de declaraciones con discrepancias positivas a lo largo del tiempo. La línea verde muestra el porcentaje de declaraciones con cualquier discrepancia, mientras que la línea roja representa solo discrepancias en ventas gravadas al 10% y la línea azul muestra discrepancias al 5%. La incidencia de discrepancias se mantiene notablemente estable en el tiempo, con aproximadamente un 7% de declaraciones que presentan alguna discrepancia. Sin embargo, no se observa una estacionalidad clara, aunque podría haber una leve reducción a partir del inicio de la pandemia.

No todas las discrepancias son lo suficientemente grandes como para justificar la atención de la autoridad tributaria, por lo que solo se consideraron elegibles los casos en los que la discrepancia positiva en un mes determinado fue de al menos 10 millones de PYG (aproximadamente 1.400 USD).

2.3 Intervención

Nuestra intervención es una versión modificada de la intervención estudiada por Carrillo *et al.* (2017) en Ecuador. Entre los contribuyentes elegibles, se seleccionaron aleatoriamente algunos (siguiendo los protocolos descritos en la sección 2.5) para recibir una notificación. La notificación presenta a los contribuyentes un resumen mensual de sus ventas declaradas, las compras reportadas por sus clientes y las discrepancias entre ambos. Como respaldo de las discrepancias, se adjuntan las facturas subyacentes a las compras reportadas por los clientes. Luego, la

³Esto puede ocurrir porque ninguno de sus clientes es un reportante de terceros o porque declaran ventas totales mayores a las reportadas por terceros.

notificación solicita que el contribuyente presente declaraciones rectificativas de los períodos fiscales pertinentes para corregir las inconsistencias. Un ejemplo de la notificación se muestra en la Figura 1.

Carrillo *et al.* (2017) presentan resultados algo desalentadores en la implementación de este tipo de intervención en Ecuador. Específicamente, surgieron dos problemas. Primero, la tasa de respuesta general en Ecuador fue bastante baja: solo entre el 10 y el 20% de los contribuyentes notificados presentaron una rectificativa. Segundo, aquellos contribuyentes que respondieron no mostraron una mayor conformidad fiscal: por cada dólar de aumento en las ventas declaradas, se registraba un incremento en las compras declaradas de hasta 96 centavos, lo que resultaba en un cambio mínimo o nulo en el valor agregado declarado y, por ende, en las obligaciones tributarias o pagos de impuestos detectables.

Para abordar estos problemas, realizamos dos modificaciones en las notificaciones. Primero, fortalecimos el lenguaje utilizado. En particular, las notificaciones enfatizaban que la solicitud era que el contribuyente presentara una rectificativa en la que aumentara sus ventas declaradas para reflejar la discrepancia, pero sin realizar otras modificaciones en su declaración. Esto tenía como objetivo evitar que los contribuyentes aumentaran simultáneamente sus compras declaradas.⁴

Segundo, las notificaciones fueron acompañadas de un mayor seguimiento y sanciones en caso de falta de respuesta. En particular, se asignaron auditores para dar seguimiento a cada caso. Si el contribuyente respondía pero no rectificaba o lo hacía solo parcialmente, los auditores revisaban el caso y determinaban si aceptaban la respuesta parcial. Si el contribuyente no respondía dentro del plazo de 10 días o respondía de manera insatisfactoria, el auditor bloqueaba el número de identificación del contribuyente (RUC). Los RUC bloqueados no pueden solicitar nuevas facturas fiscales ni realizar diversos trámites con la administración tributaria (aunque aún pueden presentar y modificar declaraciones), lo que puede afectar las actividades comerciales de los contribuyentes incumplidores, especialmente de aquellos más grandes que dependen de la emisión de facturas fiscales y tienen interacciones más complejas con la administración tributaria.

Si bien esta intervención puede ser más efectiva que una estrategia de menor rigor (de hecho, nuestros resultados son mucho más prometedores que los de la intervención de menor rigor en Ecuador (Carrillo *et al.*, 2017)), este diseño requiere que los auditores revisen cada expediente y se comuniquen con los contribuyentes si es necesario. Esto demanda una cantidad significativa de tiempo (ver sección 5.1 para una estimación detallada), por lo que el número de auditores disponibles limita directamente la cantidad de notificaciones que pueden enviarse. Esto convierte la focalización de las notificaciones en un desafío de primer orden, un tema al que regresaremos en la sección 5.3.

⁴Naturalmente, la administración tributaria no puede impedir que los contribuyentes presenten rectificativas que modifiquen múltiples líneas en sus declaraciones. Sin embargo, el mensaje estaba diseñado para ser altamente sugereente.

2.4 Cronograma

Las notificaciones se enviaron en oleadas debido a la limitada capacidad administrativa para dar seguimiento al estado de cada notificación. Había una gran incertidumbre sobre cuánta atención de seguimiento requeriría cada notificación, ya que no se sabía cuántos contribuyentes responderían ni cuántos intentarían justificar las discrepancias. Las tasas de respuesta fueron mejores de lo esperado y el número de contribuyentes que intentaron explicar las discrepancias fue menor de lo previsto, por lo que pudimos hacer que las oleadas posteriores fueran más grandes que la primera.

Realizamos un piloto operativo con 50 notificaciones en enero de 2022. Nuestra primera oleada de 491 notificaciones se envió en agosto de 2022. Una segunda oleada de 998 notificaciones se envió en diciembre de 2022. Luego enviamos oleadas de 1.992 y 1.998 notificaciones en agosto y octubre de 2023, respectivamente.⁵ Esto da un total de 5.529 notificaciones experimentales enviadas hasta la fecha.⁶ En la segunda fase del experimento, planeamos enviar 5.000 notificaciones en noviembre/diciembre de 2024 y 5.000 más en marzo/abril de 2025.

Nuestro diseño experimental evolucionó a lo largo de cuatro oleadas. Las dos primeras oleadas emplearon un diseño estándar de saturación por clúster (Baird *et al.*, 2018). Sin embargo, en las oleadas posteriores implementamos un novedoso diseño experimental en redes, optimizado específicamente para aplicar inferencia basada en aleatorización.

2.5 Oleadas 1-2: Aleatorización por Clúster

El diseño de saturación por clúster (Baird *et al.*, 2018) es un proceso de aleatorización en dos etapas diseñado para permitir la detección y medición tanto de los efectos directos del tratamiento como de los efectos indirectos sobre los contribuyentes vecinos. El diseño de saturación por clúster se lleva a cabo en los siguientes pasos:

1. Creación de clústeres de proveedores en función de la identidad de sus clientes. La idea es que tratar a un proveedor común puede generar efectos indirectos en todos sus clientes, incluidos tanto aquellos mencionados en la notificación como los que no lo están. Para ello, primero construimos una red de clientes donde los enlaces se forman entre pares de clientes que comparten un proveedor. Luego, particionamos esta red de 185.119 clientes en 6.486 clústeres utilizando el algoritmo Leiden (Traag *et al.*, 2011). Posteriormente, asignamos cada proveedor a un clúster en función del grupo de clientes al que más suministra.
2. Asignación aleatoria de clústeres a tasas de saturación del 0%, 50% o 100% de los contribuyentes elegibles.

⁵Se seleccionaron aleatoriamente 2.000 contribuyentes para recibir una notificación, pero un pequeño número no pudo ser contactado o ya había presentado una rectificativa cuando se generaron las notificaciones.

⁶En el análisis descriptivo de la sección 3 incluimos el piloto, pero en nuestra estimación de los efectos del tratamiento en la sección 4 no lo hacemos, lo que da una muestra de 5.429 contribuyentes tratados para ese análisis.

3. Selección de la muestra para la Oleada 1. Se eligió una muestra aleatoria de 500 contribuyentes siguiendo las tasas de saturación asignadas a los clústeres.
4. Selección de la muestra para la Oleada 2. Se eligió una muestra aleatoria de 1.000 contribuyentes de entre los elegibles restantes que no fueron tratados en la Oleada 1, siguiendo las tasas de saturación asignadas a los clústeres.

Esta construcción da lugar a 3.097 clústeres distintos. El tamaño de estos clústeres varía y se puede aproximar mediante una distribución logarítmico-exponencial con un parámetro de tasa de 1,27. La conectividad intra-clúster es del 1-2% y la conectividad entre clústeres es del 0,05–0,1%, ambas aproximadas mediante muestreo de Monte Carlo, lo que sugiere que hemos identificado clústeres significativos dentro de la red de compradores y vendedores.

2.6 Oleadas 3-4: Diseño *Split-Graph*

Para las oleadas de 2023, el protocolo de aleatorización se refinó para maximizar el poder estadístico en la detección tanto de *efectos directos* como de *efectos de contagio* mediados por la red observable de transacciones comerciales entre contribuyentes. Nuestra elección del diseño experimental y sus parámetros ha sido guiada por trabajos econométricos en inferencia por aleatorización, que se describirán en mayor detalle en un artículo complementario. A alto nivel, el diseño experimental, al que llamamos *Split-Graph*, estratifica a los contribuyentes según sus conexiones con otros contribuyentes elegibles y su grado (número de socios comerciales), y luego asigna el tratamiento aleatoriamente dentro de cada tipo.

Aleatorización *Split-Graph* Para seleccionar 2.000 contribuyentes para tratamiento en las Oleadas 3 y 4, seguimos cuatro pasos. Una representación visual del diseño *Split-Graph* se muestra en la Figura 4.

1. Particionar la red de contribuyentes en tres conjuntos según sus conexiones** (consideramos tanto las conexiones de compradores como las de vendedores por igual). Definimos **NE** como el conjunto de contribuyentes no elegibles en el momento del tratamiento. **E₁** como el conjunto de contribuyentes elegibles que *no* están conectados a ningún otro contribuyente elegible (es decir, solo están conectados a contribuyentes en **NE**); y **E₀** como el conjunto de contribuyentes elegibles conectados a al menos un contribuyente elegible.
2. Asignar a cada contribuyente a uno de los 12 tipos $x = (\text{grado}, \text{orientación})$. Los grados de los contribuyentes se clasifican como *pequeño*, *mediano*, *alto* o *muy alto*, según los cuartiles de la distribución del grado de los contribuyentes. La *orientación* de un contribuyente es *externa*, *interna* o *neutral*, dependiendo de si la mayoría de sus conexiones son con contribuyentes no elegibles ($> 60\%$ de sus conexiones están en **NE**), con contribuyentes elegibles ($> 60\%$ de sus conexiones están en **E₀** o **E₁**), o con ambos en proporciones similares

(entre 40% – 60%). Dado que los contribuyentes elegibles son los objetivos de las notificaciones, aquellos con orientación interna podrían tener una mayor incidencia de discrepancias en comparación con los de orientación externa.

Para los contribuyentes en \mathbf{E}_0 , calculamos los anticliques $F(x; \mathbf{E}_0)$ para cada tipo x . Un anticlique (o conjunto independiente) es un conjunto de contribuyentes que no están conectados entre sí dentro de la red de contribuyentes \mathbf{G} . Luego, buscamos los conjuntos máximos bajo la restricción $|F(x; \mathbf{E}_0)| \leq mN_x$, donde N_x es el número de contribuyentes tratados en el tipo x según el esquema de tratamiento \mathbf{N}_{E_0} presentado en el paso 4 a continuación, y m es un parámetro de intensidad de tratamiento.

3. Tratar a 100 contribuyentes en \mathbf{E}_1 (de aproximadamente 6.300, lo que equivale a una intensidad de tratamiento del 1,6%), seleccionados completamente aleatoriamente y estratificados por tipo de la siguiente manera:

	externa	neutral	interna
pequeño	90 (1,6%)	0	0
mediano	10 (1,6%)	0	0
alto	0	0	0
muy alto	0	0	0

Los números entre paréntesis indican el porcentaje de contribuyentes tratados en cada tipo (según su grado y orientación). Nótese que, por definición, todos los contribuyentes en \mathbf{E}_1 tienen orientación "externa".

4. Tratar a 1.900 contribuyentes en \mathbf{E}_0 según el siguiente esquema de tratamiento por tipo:

	externa	neutral	interna
pequeño	870 (8,5%)	110 (6,9%)	99 (10,1%)
mediano	720 (3,5%)	50 (9%)	15 (28,3%)
alto	32 (1,9%)	0	0
muy alto	4 (5,1%)	0	0

Esto nos da N_x , el número total de contribuyentes a tratar para cualquier tipo x . El tratamiento de estos N_x contribuyentes se asigna de la siguiente manera: Para cualquier tipo $x = (\text{grado}, \text{orientación})$, con $Q(x) = \{j \in \mathbf{E}_0 : x_j = x\}$ como el conjunto total de contribuyentes de ese tipo, realizamos la siguiente aleatorización:

- Tratamos a $N_{x,F} = pN_x$, para algún p fijo en el intervalo $(0, 1)$, seleccionados completamente al azar dentro del conjunto de anticlique $F(x; \mathbf{E}_0)$.
- Tratamos a $N_x - N_{x,F}$ contribuyentes seleccionados al azar dentro del conjunto restante $Q(x) \setminus F(x; \mathbf{E}_0)$.

A continuación, discutimos los parámetros de este diseño experimental y cómo seleccionamos sus valores óptimos.

Diseño Óptimo El diseño *Split-Graph* descrito anteriormente está controlado por los siguientes cuatro parámetros:

- Matrices de programación del tratamiento: \mathbf{N}_{E_0} y \mathbf{N}_{E_1} .
- Parámetros de intensidad del tratamiento: $m \in \mathbb{N}^+$, $p \in (0, 1)$.

Las matrices de programación del tratamiento se presentan arriba. Los parámetros de intensidad se establecieron en $m = 5$ y $p = 0.7$. Estos valores de los parámetros se obtuvieron después de estudios intensivos de potencia basados en simulaciones, que describimos a continuación.

Inicialmente, seleccionamos valores base para los esquemas de tratamiento, digamos $\mathbf{N}_{E_0}^b$ y $\mathbf{N}_{E_1}^b$. Luego, realizamos simulaciones de potencia sobre perturbaciones, $\mathbf{N}^b + \varepsilon$, de estos diseños base y exploramos una cuadrícula permutada de valores para los parámetros de intensidad (m, p) siguiendo un diseño de llenado de espacio (Box *et al.*, 1978; Joseph, 2016). Los parámetros de diseño sirvieron luego como entrada para un modelo de bosque aleatorio, mientras que los resultados experimentales de potencia correspondientes se utilizaron como salida. Con el modelo de bosque aleatorio ajustado, pudimos determinar los valores de los parámetros de diseño que, según el modelo, producirían una potencia óptima. Los valores óptimos revelaron una interesante compensación entre la estimación de efectos directos y efectos de contagio: Para los efectos directos, es mejor tratar conjuntos de anticlique grandes con una intensidad media (m grande, p pequeño). Para los efectos de contagio, es mejor tratar anticliques pequeños con alta intensidad (m pequeño, p grande).

2.7 Balance

Para verificar que la aleatorización se implementó correctamente y garantizar que los atributos clave de los contribuyentes tratados y de control sean similares (para reducir preocupaciones sobre variables de confusión que podrían comprometer la interpretación de los efectos del tratamiento), verificamos el balance de varias covariables. Dado que nuestros protocolos de aleatorización variaron entre las diferentes olas, presentamos los resultados de balance por separado para las Olas 1–2 y las Olas 3–4.

Las olas 1–2 siguieron el diseño de aleatorización por clúster saturado descrito en la Sección 2.5. En los clústeres asignados a 0 o 100% de saturación, no podemos comparar a los contribuyentes tratados y de control. En su lugar, resumimos las covariables de los contribuyentes dentro del clúster en *covariables a nivel de clúster* y luego evaluamos el balance a nivel de clúster. Específicamente, realizamos los siguientes pasos para diagnosticar el balance:

1. Para una covariable Z con niveles $k = 1, \dots, K$, calculamos la covariable a nivel de clúster para cada clúster J :

$$C_J = \text{entropy}(\{c_k\}_{k=1}^K), \quad c_k = \sum_{i \in J} 1\{Z_i = k\}/|J|.$$

Aquí, $\{c_k\}_{k=1}^K$ representa la distribución empírica de X dentro del clúster J , mientras que C_J es la entropía de esta distribución, que sirve como la covariable a nivel de clúster.

2. Regresamos C_J sobre p_J , donde p_J es la tasa de saturación del clúster J . Reportamos el valor- p de la prueba- t de la regresión.

Los resultados de este cálculo se presentan en la Tabla 1. Mostramos que la forma legal (incorporado/no incorporado), el tamaño (pequeño/mediano/grande), la región, la actividad económica y el grupo de edad de los contribuyentes están bien balanceados entre clústeres (los valores- p están lejos de 0).

En las Olas 3–4, la aleatorización cambió a un experimento de red estratificado, como se describe en la Sección 2.5. Como se muestra en la Figura 4, la aleatorización del tratamiento consistió en múltiples experimentos aleatorizados simples dentro de conjuntos de contribuyentes con el mismo tipo x y estado de membresía en una anti-clique. Como resultado, podemos evaluar el balance a nivel de contribuyente de la siguiente manera:

1. Dentro de cada ola, calcular el tipo x_i de cada contribuyente i (grado, orientación); su tratamiento $D_i \in \{0, 1\}$; y su estado de anti-clique $A_i \in \{0, 1\}$.
2. Para cada covariable Z , ajustar dos modelos de regresión multinomial:

$$M_0 : Z_i \sim x_i + A_i, \text{ y } M_1 : Z_i \sim D_i + x_i + A_i,$$

y reportar el valor- p de la comparación ANOVA entre M_0 y M_1 . Si el tratamiento es predictivo de la covariable, entonces dicho valor- p debería ser pequeño.

Los resultados de este cálculo para las Olas 3 y 4 se presentan en la Tabla 1 para las mismas covariables consideradas en las Olas 1 y 2. Los resultados indican un buen balance en ambas olas, lo que sugiere que la aleatorización del experimento se implementó con éxito y que la comparación entre los contribuyentes tratados y de control proporcionará estimaciones confiables del impacto causal de las notificaciones. Antes de analizar los efectos causales, comenzamos con algunas pruebas descriptivas sobre cómo los contribuyentes notificados respondieron a las notificaciones que recibieron.

3 Evidencia Descriptiva

Comenzamos nuestra exposición de los resultados del experimento estudiando lo que sucedió con los contribuyentes que fueron notificados y, en la sección 4, incorporamos a los contribuyentes del grupo control para estimar los impactos causales de las notificaciones.

La Figura 5 muestra los resultados de cada caso hasta agosto de 2024, dividiendo a los contribuyentes notificados en cuatro grupos: aquellos que no respondieron; aquellos que respondieron pero buscaron explicar las discrepancias sin modificar sus declaraciones; aquellos que realizaron modificaciones que solo abordaron parcialmente las discrepancias; y aquellos que realizaron modificaciones que corrigieron completamente las discrepancias. En general, la tasa de respuesta es de aproximadamente un 70%.⁷ Comparando esto con el 10–20% de contribuyentes que respondieron en Ecuador (según lo reportado en Carrillo *et al.*, 2017), las modificaciones en la intervención para hacerla más efectiva parecen haber dado el resultado deseado. Curiosamente, aunque la tasa general de respuesta es similar en todas las olas, la tasa de modificaciones parciales es algo mayor en las olas de 2022 que en las de 2023.

Por normativa, los contribuyentes notificados tienen 10 días hábiles para responder a los avisos. La Figura 6 traza las tasas de respuesta en cada ola a lo largo del tiempo. Muestra la proporción total de contribuyentes notificados que han respondido a medida que transcurren los días hábiles desde el día en que fueron notificados. Las líneas verticales punteadas indican la fecha límite para cada ola.⁸ Se observa un aumento significativo en las respuestas alrededor del plazo límite: la tasa de respuesta es del 15–20% justo antes de la fecha límite, pero aumenta al 25–30% cuando el plazo expira. Quizás más interesante aún, la tasa de respuesta sigue aumentando. Para cuando terminan nuestros datos, la tasa de respuesta de las primeras olas parece haberse estabilizado en torno al 65–70%, más del doble de lo que era cuando venció el plazo, y sigue aumentando lentamente. Esto es, muy probablemente, el efecto de que los contribuyentes ven bloqueados sus identificadores fiscales y responden a las notificaciones cuando esta restricción empieza a afectar sus actividades comerciales.

En la Figura 7, analizamos el contenido de las modificaciones realizadas por los contribuyentes notificados. Para cada ola de contribuyentes notificados, rastreamos los cambios acumulativos en los resultados implícitos en sus modificaciones (por ejemplo, el aumento acumulado en las ventas gravables reportadas). Luego, los dividimos por el número de contribuyentes notificados para generar medidas de la respuesta promedio por contribuyente notificado.⁹ Los paneles A–C presentan los aumentos acumulados en ventas, compras y obligaciones tributarias, respectivamente. El Panel A muestra aumentos claros en las ventas reportadas, con las cuatro olas informando un promedio de más de 200 millones de PYG (28.000 USD) al final de nuestros datos. Curiosamente, las olas más recientes parecen estar aumentando a un ritmo más rápido.

⁷Cabe señalar que la fracción de contribuyentes que no respondieron es mayor en la ola de octubre de 2023, pero como muestra la Figura 6, esto se debe a que las respuestas continúan llegando muchos meses después de la notificación, y nuestros datos se detienen nueve meses después de la ola de octubre de 2023.

⁸Estas pueden desviarse de los 10 días hábiles debido a feriados y un error de cálculo en la ola de agosto de 2023.

⁹Esta normalización nos permite comparar olas de diferentes tamaños.

Aunque no era un requisito de las notificaciones (que pedían explícitamente a los contribuyentes que modificaran *únicamente* sus ventas reportadas), muchos contribuyentes notificados optaron también por modificar sus compras reportadas. Como muestra el Panel B, al final de nuestros datos, los contribuyentes notificados habían reportado en promedio 30–35 millones de PYG adicionales en compras. Sin embargo, dado que el aumento en las compras reportadas fue un orden de magnitud menor que el aumento en las ventas reportadas, las obligaciones tributarias promedio también aumentaron considerablemente, como se observa en el Panel C, con un incremento promedio de 17–25 millones de PYG (2.300–3.400 USD) en las obligaciones fiscales de los contribuyentes notificados al final de nuestros datos.

Los paneles D–F analizan los cambios en la posición neta de los contribuyentes notificados con respecto a la autoridad tributaria. El Panel D documenta grandes aumentos en los pagos de los contribuyentes notificados al fisco, con un promedio de 1 millón de PYG por contribuyente al vencimiento del plazo legal para responder, pero que siguen aumentando después de este. Veinte meses después, la ola de agosto de 2022 había alcanzado un promedio de 3,9 millones de PYG por contribuyente notificado. Un patrón muy similar se observa en el Panel E, que estudia los créditos que los contribuyentes tienen con la autoridad tributaria. Estos disminuyen considerablemente, alcanzando una reducción de aproximadamente 4 millones de PYG al final de nuestros datos. Estas reducciones en créditos representan pagos de impuestos más altos en el futuro, por lo que, para combinarlos con el aumento en los pagos inmediatos de impuestos en el Panel D, descontamos las reducciones en créditos por un factor de 0.7 para crear una medida del valor neto de la notificación.¹⁰ Esta medida del valor neto de las modificaciones de los contribuyentes notificados aumenta de manera constante y pronunciada en el Panel F, con la ola de agosto de 2022 alcanzando los 6,5 millones de PYG al final de nuestros datos.

Finalmente, analizamos la heterogeneidad de estas respuestas entre diferentes tipos de contribuyentes notificados, centrándonos en su clasificación de tamaño (pequeño/mediano/grande) y su forma jurídica (incorporado/no incorporado). En la muestra de modificaciones realizadas por los contribuyentes notificados, realizamos una regresión del resultado del contribuyente i sobre indicadores para los diferentes grupos $g(i) \in \{\text{Físico}, \text{Jurídico}\} \times \{\text{Pequeño}, \text{Mediano}, \text{Grande}\}$ a los que pertenece el contribuyente i :

$$y_i = \beta_h \mathbf{1}[g(i) = h] + \varepsilon_i, \quad (1)$$

y reportamos los coeficientes β_h .¹¹

La Tabla 2 muestra las diferencias en las tasas de respuesta entre los distintos tipos de contribuyentes. Vemos que los contribuyentes pequeños no incorporados son los más propensos a responder, con un 76% de ellos realizando modificaciones al final de nuestros datos, mientras que, en el extremo inferior, solo el 28% de los contribuyentes medianos no incorporados

¹⁰Específicamente, el valor neto se define como el cambio en los pagos menos $0.7 \times$ el cambio en los créditos.

¹¹No hay contribuyentes de tamaño grande y de carácter legal físico en nuestra muestra, por lo que omitimos esta categoría.

respondieron. Sin embargo, las modificaciones de los contribuyentes pequeños tienden a ser menores.

En general, los resultados presentan un panorama prometedor. A mediano plazo, entre el 65–70% de los contribuyentes notificados responden, generando beneficios para la autoridad tributaria de alrededor de 5.000.000 PYG por notificación enviada. Sin embargo, estas estimaciones no representan efectos causales de la política: algunas de las modificaciones de los contribuyentes notificados podrían haber ocurrido incluso sin que la autoridad tributaria les enviara una notificación. Por lo tanto, en la siguiente sección incorporamos a los contribuyentes de control y aprovechamos la asignación aleatoria de los avisos para estimar los efectos causales sobre las rectificaciones.

4 Efectos Causales

Queremos (i) probar si las auditorías aumentan la probabilidad de que los contribuyentes presenten una enmienda a su declaración de IVA, (ii) evaluar cuándo se materializan estos efectos, (iii) estimar cómo estas enmiendas influyen en la obligación tributaria y en los pagos de los contribuyentes a la autoridad tributaria, y (iv) verificar qué empresas impulsan estos efectos.

La subsección 4.1 aborda los primeros tres de estos objetivos. Para ello, construimos intervalos de confianza basados en la aleatorización para un efecto de tratamiento constante hipotetizado. Este análisis basado en la aleatorización es válido para muestras finitas bajo un supuesto de estabilidad en los resultados potenciales, el cual describimos a continuación. Favorecemos este enfoque porque puede extenderse de manera natural para proporcionar pruebas válidas de efectos indirectos, y lo utilizamos para guiar nuestro diseño experimental en las Olas 3 y 4.

La subsección 4.2 replica nuestro análisis principal sobre los efectos de las auditorías fiscales en las rectificaciones, pero emplea la inferencia basada en modelos que es comúnmente utilizada en la microeconomía aplicada. Luego de demostrar que nuestros resultados principales también se mantienen con este método de inferencia alternativa, los ampliamos estimando la heterogeneidad en los efectos del tratamiento según el tipo de contribuyente.

4.1 Efectos en las Rectificaciones (estimación basada en el diseño)

Preliminares. Sea N el número total de contribuyentes observados durante un período $\{0, 1, \dots, T\}$. En nuestro caso, $T = 1.546$ días, con la fecha más temprana siendo el “2019-12-06” y la más reciente el “2024-02-29”.

La asignación del tratamiento está representada por la matriz $N \times T$, $\mathbf{D} = (D_{it})$, para $i = 1, \dots, N$ y $t = 0, \dots, T$. La dimensión de \mathbf{D} se omite para simplificar la notación. El indicador $D_{it} = 1$ representa el evento en el que el contribuyente i recibe una notificación en la fecha t .¹² El diseño experimental, $\text{pr} : \{0, 1\}^{N \times T} \rightarrow [0, 1]$, es conocido y se describe en detalle en la siguiente sección. Sea $\mathbb{D} = \{\mathbf{d} \in \{0, 1\}^{N \times T} : \text{pr}(\mathbf{d}) > 0\}$ el soporte de esta distribución.

¹²No requerimos que el tratamiento sea un “estado absorbente”.

Durante el experimento, en conjunto con la autoridad tributaria, modificamos los criterios que determinaban qué contribuyentes podían ser notificados, lo que hizo que la elegibilidad variara entre distintas olas del experimento. Tomamos esto en cuenta en nuestro análisis.

Así, un contribuyente i se considera *elegible* en el momento t si puede recibir tratamiento en ese momento, es decir, $\text{pr}(d_{it} = 1) > 0$. Un contribuyente *no elegible* en t no puede recibir tratamiento en ese instante. En cualquier momento t , el conjunto de contribuyentes no elegibles se denota como NE_t , y el conjunto de elegibles como E_t . La mayoría de los contribuyentes no eran elegibles, con una proporción aproximada de $|\text{NE}_t|/|\text{E}_t| \approx 10$.

Los contribuyentes están conectados a través de una red de compradores y vendedores medida en 2021, antes de la asignación de cualquier tratamiento. Sea $\mathbf{C} = (c_{ij})$, para $i, j \in \{1, \dots, N\}$, la matriz de adyacencia ponderada $N \times N$ de esta red. El valor c_{ij} en el vínculo entre el contribuyente i y el contribuyente j es no negativo y representa el valor monetario total que i (comprador) adquirió de j (vendedor), según lo informado por el contribuyente i .

A partir de esta matriz, definimos: - $w_{ij}^b = c_{ij} / \sum_j c_{ij}$, $w_{ij}^b \in [0, 1]$, como el valor ponderado del vendedor j para el comprador i . - $w_{ij}^s = c_{ij} / \sum_i c_{ij}$, $w_{ij}^s \in [0, 1]$, como el valor ponderado del comprador i para el vendedor j .

La matriz de adyacencia binaria $N \times N$, $\mathbf{G} = (g_{ij})$, representa todas las conexiones posibles entre contribuyentes en la red, tal que $g_{ij} = 1\{w_{ij} + w_{ji} > 0\}$, para todos los i, j . Utilizaremos tanto la matriz de adyacencia binaria como la matriz ponderada para estudiar efectos de contagio en la siguiente etapa de nuestro proyecto. Destacamos que estas matrices fueron observadas antes de cualquier asignación de tratamiento y, por lo tanto, pueden utilizarse como covariables fijas en nuestro análisis.

Sea $w(t)$ la función escalón que asigna a t la fecha de la ola más reciente que no es posterior a t . Por ejemplo, si $t = \text{"2022-12-25"}$, se tiene que $w(t) = \text{"2022-12-19"}$, que es la fecha de tratamiento de la Ola 2. Si $t' = \text{"2023-09-10"}$, se tiene que $w(t') = \text{"2023-07-31"}$, que es la fecha de tratamiento de la Ola 3, y así sucesivamente.

Además, definimos $\text{week}_{t'-t} \in \mathbb{Z}$ como la distancia en semanas entre las fechas t' y t , permitiendo valores negativos si t es posterior a t' . Por ejemplo, con los valores anteriores, obtenemos $\text{week}_{t'-t} = 37$ y $\text{week}_{t-t'} = -37$.

Una variable clave de interés es $A_{it}(\mathbf{d}) \in \{0, 1\}$, que indica si el contribuyente i presenta una enmienda en la fecha $t \in [0, T]$ bajo el esquema de tratamiento $\mathbf{d} \in \mathbb{D}$. Para simplificar, asumimos la condición de "no anticipación", donde cualquier resultado medido en t depende solo de los tratamientos recibidos hasta t :

$$A_{it}(\mathbf{d}) = A_{it}(\mathbf{d}_{:t}), \quad (2)$$

donde $\mathbf{d}_{:t}$ denota la submatriz de \mathbf{d} que solo incluye tiempos desde 1 hasta t . Esta suposición es razonable en nuestro contexto, ya que el día del tratamiento se define como el día en que el contribuyente recibe la notificación, siendo esta la primera vez que el contribuyente sabe que está siendo controlado.

Por lo tanto, A_{it} es un resultado intermedio que "media" otros resultados de interés, que describiremos a continuación. Medir el impacto de nuestra intervención en el aumento de la tasa de enmiendas será un objetivo clave de nuestro análisis.

Cuando un contribuyente presenta una rectificación, debe especificar qué casillas de la declaración está modificando, el valor previo a la rectificación y el valor corregido que desea declarar. Agregamos estas rectificaciones en categorías más amplias, como cambios en el valor monetario de ventas gravadas, compras gravadas, pagos de impuestos, créditos fiscales o responsabilidad tributaria del período reportado.

Sea $Y_{it}(\mathbf{d}_{:t})$ el cambio en el resultado Y , donde $Y \in \{\text{ventas, compras, responsabilidad, créditos, pagos}\}$, reportado por la empresa i en el tiempo t . Notamos que solo observamos Y cuando la empresa realmente presenta una rectificación, es decir, $A_{it}(\mathbf{d}_{:t}) = 1$.

Adoptamos la convención de que $Y_{it}(\mathbf{d}) = 0$ para cualquier \mathbf{d} en la que $A_{it}(\mathbf{d}) = 0$. Aunque esta práctica es común para manejar datos no observados de manera uniforme, puede tener implicaciones analíticas. En nuestro contexto, justificamos este enfoque señalando que un contribuyente que no presenta una rectificación implica un valor de cero para la autoridad tributaria, mientras que aquellos que presentan una rectificación pueden reportar un cambio positivo o negativo.

Esta convención impide diferenciar entre contribuyentes que presentan una rectificación sin cambios ($A = 1$ y $Y = 0$) y aquellos que no presentan ninguna rectificación ($A = 0$). En consecuencia, esta estrategia supone que todos los contribuyentes son "always takers" en términos de cumplimiento, donde $A_{it}(\mathbf{d}_{:t}) = 1$ para todos los i, t y \mathbf{d} , con los contribuyentes reportando un cambio $Y \neq 0$ o sin cambio $Y = 0$. Si bien es útil aquí, esta suposición no es estrictamente necesaria, y exploraremos alternativas en trabajos futuros.

Hipótesis nulas. Ahora nos enfocamos en las hipótesis nulas de interés. En este informe, consideramos la hipótesis nula de un efecto de tratamiento directo constante, que requiere la siguiente suposición:

Assumption 1. *El resultado potencial de cada contribuyente, cuando presenta una rectificación, depende solo de su estado de tratamiento individual y de la semana en la que se presentó la rectificación. Es decir, existen funciones $f_i : \{0, 1\} \times \mathbb{Z} \rightarrow \mathbb{R}$ tales que:*

$$Y_{it}(\mathbf{d}_{:t}) = f_i(d_{iw(t)}, \text{week}_{t-w(t)}).$$

Bajo esta suposición, definimos la hipótesis nula de efecto constante como:

$$H_0^{te}(\tau) : f_i(1, w) = f_i(0, w) + \tau, \text{ para todos } i \in [N], w \in \mathbb{Z}. \quad (3)$$

Inferencia. Nuestro objetivo es construir intervalos de confianza basados en la aleatorización para los efectos constantes del tratamiento hipotetizados, a través de la inversión de hipótesis

nulas como H_0^{te} definida en la Ecuación (3). El análisis basado en la aleatorización es válido para muestras finitas bajo la Suposición 1.

Para evaluar los resultados en distintas semanas posteriores a la recepción de las notificaciones, nuestro diseño sigue una *especificación de datos apilados*. Esta especificación se muestra en la Figura 8. La figura ilustra que, para una semana de valor positivo (por ejemplo, 1, 2, ...), nuestras pruebas de aleatorización comparan los resultados de los contribuyentes elegibles observados en esa semana con los resultados de los mismos contribuyentes en los 30 días previos al evento de tratamiento (marcados como “controles” en la figura). Para una semana de valor no positivo (0, -1, -2, ...), la prueba de aleatorización compara los resultados de los contribuyentes elegibles en esa semana con sus resultados en los 30 días anteriores a esa semana.

Para cada resultado y una semana fija $w \in \{-4, \dots, 0, \dots, 4\}$, probamos la hipótesis $H_0^{te}(\tau)$, definida en la Ecuación (3), para varios valores de τ . Luego, invertimos la prueba conservando aquellos valores de τ que corresponden a un valor p de aleatorización no mayor a 0.05, lo que produce el intervalo de confianza basado en la aleatorización. Destacamos que la validez de estos intervalos no depende de la correcta especificación del modelo, sino únicamente del procedimiento de aleatorización conocido y la Asunción 1.

Resultados. Los intervalos de confianza (IC) del 95% basados en la aleatorización para el efecto directo del tratamiento —suponiendo ausencia de interferencia— se presentan en las Figuras 9 (donde el resultado es la presentación de rectificaciones) y 10, donde cada panel corresponde a resultados relacionados con la responsabilidad tributaria, los pagos, los créditos, las compras y las ventas.

A partir de este análisis, obtenemos cuatro hallazgos empíricos. Primero, las notificaciones tuvieron un efecto cuantitativo fuerte sobre las tasas de rectificación. La Figura 9 muestra que los grupos de tratamiento y control se comportaron de manera similar en las semanas previas a las notificaciones, pero el grupo tratado mostró un aumento inmediato y pronunciado después de recibirlos. Nuestras estimaciones implican que 800 contribuyentes notificados ($0.145 \cdot 5.000$) presentaron enmiendas a su declaración de impuestos en la segunda semana después de la notificación *debido* a haber sido notificados.

Segundo, los efectos del tratamiento disminuyen con el tiempo, pero son persistentes. Observamos un pico en todos los resultados dos semanas después de la notificación, lo que coincide con el plazo que tienen los contribuyentes para responder. Sin embargo, analizamos los efectos hasta cuatro semanas después de la notificación y encontramos impactos significativos incluso en la cuarta semana.

Tercero, las rectificaciones fueron beneficiosas para la autoridad fiscal. Los contribuyentes reportaron más ventas y más compras, pero las compras adicionales no compensaron completamente las ventas adicionales, lo que resultó en mayores responsabilidades tributarias, mayores pagos a la autoridad fiscal y una reducción en los créditos adeudados a los contribuyentes.

Cuarto, el patrón de los efectos del tratamiento en otros resultados en la Figura 10 refleja

la forma de los efectos sobre la presentación de enmiendas. Esto es consistente con la hipótesis de un efecto constante del tratamiento: las enmiendas que ocurrieron debido a las notificaciones fueron, en promedio, mayores (reportando más ventas, un ligero aumento en las compras, mayores responsabilidades tributarias, mayores pagos al fisco y una reducción en los créditos), mientras que el patrón temporal de los efectos está impulsado por la evolución de la tasa de enmienda.

En general, los resultados sugieren que las notificaciones fueron altamente efectivas en múltiples dimensiones. Para resumir, el panel F de la Figura 10 muestra los efectos del tratamiento sobre la medida agregada del valor neto de una enmienda, descrita en la sección 3.¹³ Una vez más, observamos efectos marcados inmediatamente después de las notificaciones, con un pico en la segunda semana, pero con persistencia hasta la cuarta semana, que es el final de nuestro análisis.

4.2 Heterogeneidad de los Efectos entre Contribuyentes (estimación basada en modelos)

Replicación de los resultados principales. Primero buscamos replicar nuestros resultados principales de la sección anterior sobre los efectos de las auditorías fiscales en la probabilidad de presentar enmiendas y el tipo de rectificación, utilizando un enfoque estadístico diferente: una estimación basada en modelos.

Específicamente, empleamos el método de diferencias en diferencias apiladas introducido en Cengiz *et al.* (2019). Este método se describe y se compara con otros enfoques recientes en la literatura de diferencias en diferencias con efectos fijos bidireccionales en Baker *et al.* (2022). La idea consiste en crear conjuntos de datos específicos para cada ola de notificación, denominados “2×2 limpios”. Cada uno de estos conjuntos de datos incluye contribuyentes que no habían sido tratados antes de la ola de notificación (es decir, están “limpios” en la línea de base), con algunos seleccionados aleatoriamente para ser tratados durante la ola de notificación, mientras que los demás permanecen como grupo de control. Cada conjunto de datos incluye un panel balanceado con la misma ventana temporal antes y después de la notificación.

Implementamos esta estrategia en nuestro contexto construyendo un conjunto de datos para cada una de las cuatro olas de notificaciones, con datos de 60 días antes y 60 días después de la notificación para cada empresa elegible, excluyendo a los contribuyentes que fueron tratados en olas anteriores. Luego, combinamos estos cuatro conjuntos de datos específicos de auditoría en un solo conjunto de datos para estimar la siguiente ecuación:

$$y_{kit} = \beta D_{kit} + \omega p_{kit} + \mu_{ki} + \delta_{kt} + \varepsilon_{kit} \quad (4)$$

donde y_{kit} es el resultado para el contribuyente i en la fecha t dentro de la ola de notificación k (por ejemplo, un indicador de presentación de rectificación; presentación de rectificación que au-

¹³Específicamente, el valor neto se calcula como el cambio en pagos menos $0.7 \times$ el cambio en créditos.

mentó las ventas reportadas, etc.); D_{kit} es un indicador que toma el valor de 1 si el contribuyente i fue notificado dentro de la ola k y si la fecha t es posterior a la notificación; p_{kit} toma el valor de 0 antes de la notificación y, posteriormente, es igual a la probabilidad de que el contribuyente i haya sido seleccionado aleatoriamente para recibir una notificación en la ola k ; μ_{ki} representa efectos fijos por contribuyente y ola de notificación; y δ_{kt} representa efectos fijos por tiempo y ola de notificación.

La inclusión de p_{kit} permite tener en cuenta el hecho de que, debido a nuestro procedimiento de aleatorización, los contribuyentes con diferentes características pueden tener distintas probabilidades de ser tratados, lo que nos permite identificar el efecto causal directo de las notificaciones a través de β . Calculamos p_{kit} empíricamente mediante la simulación de la asignación del tratamiento 10.000 veces. Reportamos el valor estimado de β y los valores p asociados para la prueba $H_0 : \beta = 0$ tanto a partir de inferencia asintótica estándar como mediante inferencia basada en aleatorización.

Encontramos que los contribuyentes tienen 4,5 veces más probabilidades de presentar una rectificación a su declaración jurada cuando han sido notificados sobre discrepancias. La Tabla 4, Col. 1 muestra que los contribuyentes notificados tienen 0,6 puntos porcentuales más de probabilidad de realizar una enmienda en un día determinado, en comparación con una tasa de enmienda base de 0,14%. En línea con el enfoque de las notificaciones sobre las ventas reportadas, encontramos que la probabilidad de presentar una rectificación que aumente las ventas reportadas aumenta significativamente.

Curiosamente, también encontramos que ser controlado aumenta la probabilidad de presentar una enmienda que incremente las compras reportadas (lo que, manteniendo todo lo demás constante, reduce el valor agregado reportado). Aunque este efecto es mucho menor que el de las ventas, podría indicar una respuesta estratégica por parte de los contribuyentes para mitigar el impacto de las notificaciones en su carga tributaria total.

En general, encontramos que los contribuyentes notificados tienen 6,6 veces más probabilidades de presentar una enmienda que aumente su responsabilidad tributaria en comparación con los contribuyentes no notificados (Tabla 4, Col. 4). Esto se traduce en un aumento de 9 veces en la probabilidad de que una enmienda reduzca los créditos pendientes a favor del contribuyente ante la autoridad fiscal (Col. 5) y un aumento de 12 veces en la probabilidad de que una enmienda incremente los pagos adeudados (Col. 6). Estos resultados son coherentes con la evidencia descriptiva presentada en la sección 3, pero ahora representan estimaciones causales del impacto de las notificaciones.

Heterogeneidad. El análisis previo confirma los efectos significativos de las notificaciones. Ahora buscamos explorar cómo varían los efectos causales de las notificaciones según el tipo de contribuyente.

En la Tabla 5, presentamos los resultados de la estimación de efectos heterogéneos del tratamiento, enfocándonos en la clasificación del tamaño de los contribuyentes (pequeño / mediano / grande)

y su forma jurídica (física / jurídica). Para ello, extendemos la ecuación (4) de la siguiente manera:

$$y_{ijkt} = \sum_j \beta_j D_{kit} \times S_{ij} + Post_t \times S_{ij} + \omega p_{kit} + \mu_{ki} + \delta_{kt} + \varepsilon_{ijkt} \quad (5)$$

La Tabla 5 muestra los resultados. Se destacan tres hallazgos principales. Primero, los efectos estimados del tratamiento para los grandes contribuyentes son indistinguibles de cero. Esto se debe, en parte, al número muy reducido de grandes contribuyentes que fueron notificados (de hecho, en la tercera y cuarta olas, los grandes contribuyentes eran inelegibles), pero principalmente a su mayor complejidad y tamaño.

Segundo, los efectos son más fuertes para los pequeños contribuyentes que para los medianos y más fuertes para los contribuyentes no incorporados que para los incorporados. Esto es consistente con lo que podríamos esperar: los contribuyentes más pequeños y menos formalizados tienen mayor facilidad para reportar incorrectamente de manera deliberada y/o son más propensos a cometer errores en sus declaraciones fiscales.

Tercero, la excepción a este patrón es que los impactos sobre la responsabilidad tributaria son mayores para los contribuyentes medianos no incorporados. Sin embargo, este patrón no se observa en los aumentos en los pagos o en la reducción de créditos. Esto sugiere que estos contribuyentes medianos tienen mayores impuestos retenidos en origen, lo que hace que los cambios en su responsabilidad tributaria sean menos propensos a traducirse en cambios en los pagos de impuestos. Esta interacción con las retenciones es un aspecto interesante para futuras investigaciones.

5 Discusión y Próximos Pasos

5.1 Análisis de Costos y Beneficios

Hasta ahora, nuestro análisis se ha centrado en los beneficios de las notificaciones en términos de resolución de discrepancias y aumento del cumplimiento tributario. Sin embargo, las notificaciones también implican costos para la autoridad tributaria que deben considerarse para evaluar el éxito o no de la campaña de notificaciones.¹⁴ Para ello, trabajamos con la administración tributaria para realizar una contabilidad detallada del tiempo y los costos asociados a la preparación de las notificaciones y las acciones de seguimiento necesarias para procesar las respuestas de los contribuyentes.

Para estimar los costos de la administración tributaria, intentamos calcular el costo total de atender las 7,081 notificaciones enviadas en 2022 y 2023.¹⁵ Dividimos los costos en cinco fases

¹⁴Las notificaciones también generan costos para los contribuyentes notificados, los cuales podrían ser tenidos en cuenta. Sin embargo, no contamos con estimaciones fiables de estos costos, por lo que no los incluimos en nuestros cálculos. Además, no está claro cuánto peso social debería asignarse a los costos soportados por contribuyentes que potencialmente están infringiendo la ley.

¹⁵La autoridad tributaria envió 1,552 notificaciones no experimentales además de las 5,529 notificaciones experimentales analizadas aquí.

correspondientes al ciclo de vida de una notificación.

Generación Esta etapa cubre los costos de planificación de las notificaciones (decidir qué tipos de inconsistencias estudiar y analizar casos de ejemplo antes de ampliar el análisis a toda la población de contribuyentes), los costos de extracción y análisis de datos del universo de contribuyentes y los costos de preparación de las notificaciones para cada contribuyente notificado, incluyendo la verificación de los datos presentados en cada caso. Se estima que esto requirió un total de 5 días de trabajo de un equipo compuesto por tres auditores, un supervisor y dos gerentes, implicando un costo total de USD 1.616.

Transmisión Esta etapa cubre los costos de envío de las notificaciones. Estas deben ser enviadas individualmente desde la cuenta de correo oficial de la autoridad tributaria. En total, esto requirió un día completo de trabajo de un auditor y un gerente, con un costo total de USD 117.

Aceptación de Rectificaciones La tercera fase corresponde al análisis y aceptación de las rectificaciones a las declaraciones juradas realizadas por los contribuyentes en respuesta a las notificaciones. Hasta abril de 2024 (cuando se realizó la estimación de costos), 2.702 contribuyentes habían presentado modificaciones. Estas deben ser verificadas contra las discrepancias iniciales para comprobar si han eliminado o reducido la discrepancia a un nivel insignificante. Para procesarlas, se requirió un día de trabajo de un gerente y un supervisor, con un costo total de USD 123.

Análisis de Objeciones/Explicaciones La fase más costosa es el análisis de las respuestas de los contribuyentes que descargan la notificación o buscan justificar la discrepancia. Las objeciones se envían por correo electrónico y luego se asignan a un auditor para su revisión. El auditor evalúa los argumentos y cualquier evidencia documental proporcionada. También compara la información con otros registros en las bases de datos gubernamentales antes de tomar una decisión sobre la aceptación de la respuesta. La autoridad tributaria estimó que cada caso requirió un promedio de 53 horas de trabajo de los auditores. Hasta abril de 2024, 1.925 contribuyentes habían presentado objeciones, lo que llevó a un costo total de USD 530.010.

Bloqueo de Números de Identificación en el Registro Único de Contribuyentes (RUCs) Otra fase muy costosa es el bloqueo de los RUCs de contribuyentes que no responden o cuyas respuestas son insatisfactorias. Bloquear un RUC implica que un supervisor prepare un formulario indicando la razón del bloqueo y que un gerente lo apruebe. La autoridad tributaria estima que cada caso requiere un promedio de 15 minutos de trabajo de cada funcionario. Hasta abril de 2024, se habían bloqueado o estaban pendientes de bloqueo 4.061 RUCs, lo que generó un costo total de USD 123.994.

Costo-Beneficio Sumando las fases anteriores, estimamos que el costo total para la autoridad tributaria de gestionar 7.081 casos fue de PYG 4.853.366.550 (USD 655.860). Esto implica un costo promedio por caso de PYG 685.407 (USD 93).

Comparando esto con los beneficios estimados en las secciones [III](#) y [IV](#), podemos calcular la tasa de retorno promedio del envío de las notificaciones. Para estimar el beneficio promedio, utilizamos los efectos estimados sobre el beneficio total (pago adicional - $0.7 \times$ reducción de créditos) mostrados en la Figura [X](#), y asumimos que durante las 8 semanas posteriores al final de nuestro período de análisis, el efecto causal es en promedio la mitad del efecto observado 4 semanas después de la notificación, y luego se detiene. Esta es una estimación conservadora, ya que como se observa en la Figura [VII](#), los contribuyentes notificados siguen presentando modificaciones durante al menos 8-9 meses después de ser notificados. Esta estimación conservadora arroja un beneficio promedio de PYG 936.535 (USD 127), lo que implica una tasa de retorno del 37%.

Este resultado es altamente favorable en comparación con otras intervenciones. Por ejemplo, la versión más beneficiosa de los incentivos salariales para inspectores fiscales en el contexto estudiado por [Khan et al. \(2016\)](#) tuvo una tasa de retorno del 35%. Asimismo, una evaluación integral del bienestar de las notificaciones de cumplimiento en la Ciudad de México mostró tasas de retorno negativas ([Brockmeyer et al., 2023](#)).

5.2 Efectos Indirectos del Control Tributario

El análisis anterior presenta solo una imagen parcial del impacto total de notificar a los contribuyentes. Las notificaciones resaltan *discrepancias* entre los reportes de los contribuyentes y los reportes de sus clientes al gobierno. Es natural esperar que las notificaciones no solo afecten a los contribuyentes notificados, sino también a sus clientes y, potencialmente, a sus proveedores.

Estos efectos indirectos pueden manifestarse de diversas maneras. Si los contribuyentes notificados perciben que las discrepancias entre sus ventas declaradas y las compras declaradas por sus clientes se han vuelto más costosas y transmiten esta percepción a sus clientes, esperaríamos que los contribuyentes notificados aumenten sus ventas reportadas y que sus clientes reduzcan las compras reportadas a los contribuyentes notificados, reduciendo así la discrepancia entre sus reportes. También podríamos esperar que los contribuyentes notificados reduzcan sus compras reportadas para que coincidan mejor con las ventas reportadas por sus proveedores.

Alternativamente, si los contribuyentes notificados perciben que están bajo un escrutinio especial dirigido específicamente hacia ellos y menos hacia sus clientes, podríamos observar que los contribuyentes notificados enmienden sus declaraciones para corregir la discrepancia, reduciendo así la necesidad de que sus clientes realicen modificaciones en sus propias declaraciones, lo que llevaría a menos y menores correcciones por parte de los clientes de los contribuyentes notificados.

Finalmente, podríamos esperar que las notificaciones tengan efectos indirectos en las futuras declaraciones fiscales de los contribuyentes notificados (y posiblemente también en las de sus

clientes). Si las notificaciones llevan a los contribuyentes a actualizar sus creencias sobre la capacidad de la administración tributaria para detectar reportes incorrectos, podrían aumentar sus obligaciones tributarias declaradas en los meses posteriores a la recepción de las notificaciones.

Todos estos efectos son objeto de un análisis en curso sobre los datos experimentales. Este análisis se ve facilitado por el diseño del experimento descrito en la sección 2.5 y será reportado en un documento complementario a este.

5.3 Implicaciones de Política para la Focalización de los Controles Masivos

Los resultados de nuestra intervención indican que enviar notificaciones a los contribuyentes sospechados de declarar incorrectamente sus ventas es altamente rentable. Permite a la autoridad tributaria aumentar los pagos de impuestos recaudados y reducir los créditos fiscales pendientes, con un costo de implementación muy bajo.

A pesar de esta rentabilidad, la posibilidad de ampliar la intervención e incrementar el número de notificaciones enviadas a los contribuyentes está limitada por la capacidad de la autoridad tributaria: cada notificación requiere tiempo de un agente tributario para procesarla y garantizar una respuesta adecuada por parte del contribuyente notificado. Nuestros resultados indican que relajar estas limitaciones de capacidad, contratando más agentes tributarios para implementar esta intervención a mayor escala, sería una inversión muy valiosa.

Sin embargo, es muy probable que siempre existan restricciones de capacidad: no será posible dirigir las notificaciones a todos los contribuyentes elegibles al mismo tiempo. Por ello, mejorar la focalización de las notificaciones es clave para maximizar los beneficios de la intervención. Dado un número determinado de notificaciones que pueden enviarse, ¿a qué contribuyentes se les debe enviar para maximizar los beneficios fiscales?

Para responder a esta importante pregunta, planeamos combinar la teoría económica con una segunda fase del experimento. Crearemos cuatro posibles reglas de focalización y evaluaremos su efectividad, como se describe a continuación.

1. *Focalización aleatoria.* La forma más sencilla de decidir qué contribuyentes deben recibir una notificación es seleccionarlos aleatoriamente entre todos los elegibles. Este es un procedimiento fácil de implementar, pero es poco probable que elija a los contribuyentes que maximicen los beneficios para la autoridad tributaria.
2. *Perfil de riesgo.* Una mejora sobre la focalización aleatoria es dirigir las notificaciones a los contribuyentes que tengan más probabilidades de responder al control. Una forma de hacerlo es aprovechar la información y experiencia de los auditores tributarios, permitiéndoles indicar qué contribuyentes creen que actualizarán sus declaraciones y obligaciones fiscales en mayor medida tras recibir una notificación. Otra forma, basada en datos, es estimar cómo varía el efecto directo de la intervención según las características de los contribuyentes y utilizar estas estimaciones para predecir cuáles responderían más. Ya hemos

realizado dicho análisis de heterogeneidad, cuyos resultados se presentan en la sección 4.2. Estos sugieren que se debe priorizar a los contribuyentes pequeños y no incorporados.

Aunque es una mejora respecto a la focalización aleatoria, este enfoque de perfil de riesgo no es completamente satisfactorio, ya que no considera los efectos indirectos y, por lo tanto, es poco probable que maximice los beneficios fiscales de la intervención. En su lugar, proponemos dos reglas de focalización que sí toman en cuenta los efectos indirectos.

3. *Enfoque flexible basado en datos.* Un primer enfoque que considera tanto los efectos directos como los indirectos de las auditorías fiscales se basa en inteligencia artificial (IA). Con este enfoque, aprenderemos de manera flexible a partir de las rondas previas del experimento y los datos disponibles cómo el comportamiento de los contribuyentes se ve influenciado por su tratamiento y el de los miembros de su red, y prediciremos los ingresos fiscales bajo diferentes reglas de focalización. De este modo, podremos elegir, de manera completamente basada en datos, la regla de focalización que se espera que maximice los ingresos fiscales. [Viviano \(2024\)](#) proporciona un método posible para hacerlo.

Un enfoque basado en datos tiene dos beneficios principales: puede utilizar toda la información disponible de manera flexible y requiere suposiciones mínimas sobre la forma de los efectos indirectos. Sin embargo, también tiene desventajas. Primero, los enfoques de aprendizaje automático para la heterogeneidad en los efectos del tratamiento pueden ser inestables: ejecutar el mismo algoritmo varias veces en los mismos datos puede arrojar resultados muy diferentes, como se ha demostrado en [Engl et al. \(2021\)](#). Segundo, el aprendizaje automático solo puede aprender de lo que ha sucedido en los datos en los que ha sido entrenado. En nuestro contexto, esto significa que puede predecir bien los efectos de las reglas de focalización sobre los tipos de contribuyentes que han sido elegibles en nuestro experimento, pero podría fallar al predecir los efectos de cambiar los criterios de elegibilidad o si ahora otros tipos de contribuyentes son elegibles.

4. *Enfoque basado en un modelo teórico económico.* Un segundo enfoque que considera tanto los efectos directos como los indirectos de las auditorías fiscales se basa en la modelización económica. Crearemos un modelo de las decisiones de reporte de impuestos de los contribuyentes, que capturará los mecanismos clave de los efectos directos e indirectos de las auditorías fiscales. Al ajustar este modelo a nuestros datos existentes, podremos evaluar la relevancia empírica de varios mecanismos posibles aprendiendo así mejor cómo las empresas responden a la fiscalización y derivar implicaciones sobre la focalización óptima.

Este enfoque proporcionará predicciones más estables que el enfoque flexible basado en datos, podrá extrapolar predicciones para tipos de contribuyentes que no hayan sido observados en el experimento anterior y mostrará claramente qué mecanismos están impulsando las recomendaciones de focalización. No obstante, este enfoque requiere hacer suposiciones específicas sobre el comportamiento de los contribuyentes y tratar de capturar

los mecanismos clave: omitir un factor importante en el modelo puede reducir significativamente la calidad de la regla de focalización derivada.

Dado los distintos costos y beneficios potenciales de estos enfoques, es necesario evaluar empíricamente su eficacia. Lo haremos en una segunda fase del experimento, que se lanzará en marzo o abril del 2025.

En esta segunda fase, asignaremos 5.000 notificaciones según las cuatro reglas de focalización descritas y compararemos sus efectos sobre los ingresos fiscales. Para ello, dividiremos la red completa de contribuyentes en clústeres, creando grupos de contribuyentes con relaciones comerciales intensas dentro del grupo pero conexiones limitadas con otros grupos. Luego, asignaremos aleatoriamente diferentes clústeres a diferentes reglas de focalización y compararemos los ingresos fiscales recaudados en los clústeres asignados a distintas reglas. Esto nos permitirá evaluar, teniendo en cuenta tanto los efectos directos como los indirectos de la intervención, qué regla de focalización es más efectiva.

Con esta segunda fase del experimento, podremos proporcionar recomendaciones de política relevantes a la autoridad tributaria sobre la mejor manera de focalizar futuras auditorías en un contexto de capacidad limitada.

6 Conclusión

Obtener información de terceros sobre la actividad económica de los contribuyentes cuando se combina con acciones efectivas de fiscalización puede ser muy valioso para las autoridades fiscales que buscan reducir la evasión fiscal y aumentar la recaudación de impuestos. Mostramos que hacer coincidir lo que los compradores y vendedores informan sobre sus transacciones y notificar a los contribuyentes sobre las discrepancias puede ser un método simple pero efectivo para reducir el fraude. Si bien evaluaciones anteriores de esta idea en otros contextos tuvieron un éxito limitado, mostramos que una versión en la que la notificación de la discrepancia está respaldada por una amenaza creíble en nuestro caso, bloquear el número de identificación fiscal de los contribuyentes si no toman medidas puede funcionar muy bien.

En la primera fase de nuestra colaboración con la autoridad tributaria de Paraguay, encontramos que la intervención lleva a una alta tasa de respuesta entre los contribuyentes seleccionados (65% dentro de los 9 meses posteriores a la notificación) y a un aumento correspondiente en las ventas declaradas. Si bien algunos contribuyentes también responden estratégicamente aumentando sus compras reportadas (para reducir el aumento de la obligación tributaria inducido por sus enmiendas), encontramos que, en general, nuestra intervención condujo a un aumento de la obligación tributaria. Este aumento en la obligación tributaria corresponde tanto a un aumento en los pagos realizados a la autoridad tributaria como a una disminución de los créditos adeudados por la autoridad tributaria. Por lo tanto, la autoridad tributaria se beneficia de un aumento inmediato en los ingresos fiscales recaudados y un aumento esperado en

la recaudación futura de impuestos. En general, evaluamos que el efecto directo de la intervención en los contribuyentes seleccionados genera un beneficio de PYG 936.535 (USD 127) por contribuyente notificado y lo hace a bajo costo. Las notificaciones tienen una tasa de retorno estimada de 37% después de 12 semanas, lo que las hace muy competitivas frente a otros tipos de actividades de fiscalización.

Hasta ahora, nuestro análisis se ha centrado en los efectos directos de la intervención en los contribuyentes seleccionados. Sin embargo, dada la naturaleza de nuestro contexto e intervención (comparar los informes entre compradores y vendedores, e implementar la intervención entre los contribuyentes que comercializan regularmente entre sí), podemos esperar que la intervención genere efectos indirectos y también influencie el comportamiento de los contribuyentes no directamente seleccionados. En última instancia, esto puede tener implicaciones importantes para la focalización óptima de las auditorías. Este será el enfoque de futuras fases de nuestro experimento y colaboración con la autoridad tributaria.

Part B

How to Use Third Party Reports to Collect Taxes: Experimental Evidence from Paraguay

I Introduction

A functioning state forms the bedrock of economic development (Page & Pande, 2018). The core function of the state is to provide public goods and services, which needs to be financed by the collection of taxes. Thus, understanding how to move from collecting 10–15% of GDP in taxes — as is common in low- and middle-income countries — to collecting 30–40% of GDP in taxes can be viewed as one of the central questions in economic development (Besley & Persson, 2009, 2013).

Tax codes do not differ drastically across countries. However, low- and middle-income country economies are characterized by large informal sectors and widespread tax evasion. This suggests that finding ways to achieve higher tax compliance is the key potential avenue to increase state capacity in low- and middle-income countries (Gordon & Li, 2009; Besley & Persson, 2013). A much-vaunted mechanism through which higher compliance might be achieved is to make third-party information about taxpayers’ obligations available to the tax administration (Kleven *et al.*, 2011; Pomeranz, 2015). However, doubts have been raised about the efficacy of collecting third party information alone to achieve higher compliance. The available third-party information needs to be connected to concrete enforcement actions to meaningfully increase tax compliance (Carrillo *et al.*, 2017; Naritomi, 2019; Best *et al.*, 2023).

In this paper, we report some preliminary results from an ongoing partnership with the tax administration in Paraguay to explore the potential to leverage third-party reports by taxpayers’ clients and suppliers to increase compliance with Paraguay’s Value Added Tax (VAT). The VAT features an invoice-credit mechanism through which taxpayers report both sales and purchases and pay taxes on the difference (value-added), creating opposing incentives to misreport: buyers would like to over-report their purchases, while sellers would like to under-report their sales Pomeranz (2015). However, for this mechanism to lead to higher enforcement, the tax administration needs to use these third-party reports to cross-check taxpayers’ declarations and act on any discrepancies. This project explores the efficacy and design of these actions. The overarching goal of the partnership is to develop tools that take maximal advantage of the large quantities of third-party data that can be collected by the tax administration of a middle-income country like Paraguay, which is subject to limited administrative and human resources.

We build on existing work to develop a simple, scalable tax enforcement tool to use third-party reports to detect potential under-reporting of taxable sales by VAT-liable taxpayers (Carrillo *et al.*, 2017; Almunia *et al.*, 2024). We combine the detected potential under-reporting with a notice from the tax authority, backed by a credible threat—blocking taxpayers’ identification

number if they don't take action—to generate meaningful increases in taxpayers' perceived strength of scrutiny and enforcement from the tax authority. We randomly select taxpayers to receive the notice informing them of the suspected under-reported sales and requesting that they amend the relevant tax returns to address the discrepancy with their clients' reported purchases within ten working days.

Estimating the causal effects of these notices on the targeted tax payers is non-trivial due to the potential for spillover effects. Sending a notice to a given firm may affect not only the behavior of the notified taxpayer but also that of its trading partners—for instance, by increasing their perception of scrutiny or by altering strategic evasion decisions. If this is the case, standard tools in applied microeconomics are unlikely to work well. We therefore develop new methods to design and analyze the experiment, which are better suited to the context at hand.

We present six main findings. First, our notifications are highly effective at inducing taxpayers to amend their tax returns. 65% of notified taxpayers had made an amendment within nine months of notification. We estimate that this is almost entirely driven by the causal effects of notifying taxpayers, with effects concentrated in the first two weeks after notification but persisting for at least two months.

Second, the amendments induced by the notifications entailed significant increases in reported sales. Notified taxpayers were 55% more likely to make an amendment that increased reported sales. On average, these amendments increased reported sales by 900% with effects again concentrated in the first two weeks after notification. This combines the fact that several taxpayers make small “business as usual” amendments even without notifications, with the fact that our notifications induced a very large increase in the number of amendments and these experimentally induced amendments increased reported sales much more than a typical “business as usual” amendment.

Third, consonant with the findings in [Carrillo *et al.* \(2017\)](#), reported purchases also increase. However, in contrast with [Carrillo *et al.* \(2017\)](#), we do see statistically significant and economically meaningful increases in taxpayers' reported tax liabilities. This is likely driven by the combination of the use of third-party data to detect likely under-reporting with the credible threat of blocking taxpayers' tax IDs.

Fourth, the increases in reported tax liabilities are not offset by increases in reported withholding or other tax deductions. Rather, there are statistically and economically meaningful increases in taxes remitted to the tax authority. Fifth, for those who do not increase tax remittance, there are statistically and economically meaningful decreases in tax credits owed by the tax authority to the taxpayer.

Sixth, combining the above results, we estimate that the mass notifications intervention is highly effective. It delivers benefits in the form of additional tax payments and reduced credits for a total value of PYG 936,535 (USD 127) per notified taxpayer.¹⁶ Moreover, a detailed estima-

¹⁶This computation values increases in tax payments at their nominal value, and reductions in tax credits at 70 cents on the dollar.

tion of the administrative costs of the intervention suggests that the notifications are extremely cost effective with a rate of return of approximately 37%.

This preliminary report presents only a partial account of the outputs from our partnership. In particular, our ongoing work explores two important and related dimensions: spillovers and policy targeting. Sending notices to taxpayers has the potential to influence the behavior of the targeted taxpayers' trading partners. We will test whether this is the case empirically, estimate how large these spillovers are, and provide evidence on what drives them. This is necessary to accurately assess the overall impacts of tax audits. Our results will also help improve the targeting of audits: a key policy question in settings with limited implementation capacity. In our case, preparing and sending notices is relatively cheap—but their efficacy crucially relies on intense individual follow-up by tax agents. As such, the number of notices that can be sent is directly limited by the availability of tax agents. Knowing how best to target notices is therefore important to maximize their benefits. We will propose different types of targeting rules that have the potential to improve tax benefits, and we will empirically assess their effects.

Our work on spillovers and optimal policy targeting will contribute to the active academic literature on both of these topics. It will also be directly relevant for policy-makers in tax authorities and beyond: understanding the spillover effects of policy interventions is a core question for health policy, active labor market interventions, etc. Our methods and results will be the subject of future reports and their accompanying academic paper(s).

II Context & Experimental Design

II.i Project Partnership

Our project is a partnership with the tax administration in Paraguay. Informality and tax evasion are central barriers to the development of the economy: 65% of workers are informally employed, and tax revenues are 10% of GDP (World Bank, 2017). Hence mobilizing additional resources for the government is a key development challenge. As is common around the world, the key challenges in raising additional revenues are the existence of a large informal sector and widespread tax evasion ([Besley & Persson, 2013](#)). This makes enforcement actions by the tax administration to increase tax compliance a key lever for efforts to address the challenge.

Together with the tax administration, we developed a program of mass desk audits to leverage third-party reported data cross-checked against taxpayers' value-added tax (VAT) declarations. Taxpayers file monthly declarations reporting their total sales and purchases, their tax liability, taxes already paid/withheld, and the amount of tax due or credit to carry forward. Large, medium-sized, and some small taxpayers also make monthly informational declarations on all the individual transactions they make with their clients and their suppliers. These third-party reports form the basis of our intervention. The program dovetailed with the tax administration's existing program of mass desk audits, but we worked together to focus the program on cross-checks against third-party reports. In order to be able to rigorously evaluate the impact of the

notifications, we embedded a series of randomized control trials (RCTs) in the audit program, as described in the following sections.

II.ii Discovering Reporting Discrepancies

Our intervention is based on the discrepancies between taxpayers' tax return declaration and their trading partners' informational declarations.

For each focal taxpayer i in each month m , we compute the total amount that taxpayers who make informational declarations about their purchases report buying from the focal taxpayer. This is their third-party reported sales y_{im}^r . This is then compared with what the focal taxpayer declares in their tax return y_{im}^d . Since not all taxpayers are required to make informational declarations about their purchases, the focal taxpayer's declared sales should always be at least as large as third-party reported sales. That is, it should be the case that the reporting *discrepancy* $d_{im} = y_{im}^r - y_{im}^d \leq 0$. Whenever this is not the case, there must be misreporting, either because the focal taxpayer has under-reported their sales, because one or more clients have over-reported their purchases, or because there has been a mistake in a declaration.¹⁷

When computed this way, it is natural for there to be discrepancies. Since only the most important taxpayers (mostly large- and medium-sized taxpayers) make third-party reports, only part of taxpayers' sales are covered by third-party reporting. As a result, we expect most discrepancies to be negative. Figure II shows the distribution of normalized discrepancies (since discrepancies vary tremendously in size, we normalize them by computing $\tilde{d}_{im} = d_{im} / \frac{1}{2} (y_{im}^r + y_{im}^d)$). Panel A shows the overall distribution of normalized discrepancies, removing cases where discrepancies are exactly zero. We see that the bulk of discrepancies are negative, but there is a meaningful mass of positive discrepancies that are potentially the result of tax misreporting. In Panel B we plot the distributions of strictly positive discrepancies separately for sales taxable at 10% (the standard rate) and sales taxable at 5% (the reduced rate applied to a subset of goods). Consistent with misreporting, we see that the distribution of 10% discrepancies features more medium-sized and large discrepancies while those taxable at 5% tend to be smaller and are less likely to feature completely unreported sales (a smaller mass at 2).

In Figure III, we explore the incidence of positive discrepancies across taxpayers and over time. For each taxpayer, we compute the proportion of months in which they file a tax return that shows positive discrepancies compared to third-party reports of their sales. Panel A shows the distribution of this propensity across taxpayers, removing the 59% of taxpayers who never have positive discrepancies.¹⁸ We see that there are many taxpayers whose returns always generate positive discrepancies, but also many taxpayers who only sometimes generate positive discrepancies. In Panel B we track the percentage of returns that have positive discrepancies

¹⁷Common mistakes discovered during the experiment include imports/exports being declared using the wrong exchange rate, invoices that are later canceled due to product returns, sales on credit reported as liquidated, and sales assigned to the incorrect month.

¹⁸This can happen either because none of their clients are third-party reporters or because they report sales totaling more than the available third-party reports.

over time. The green line shows the percentage of returns with any discrepancy, while the red line breaks out only discrepancies in sales taxable at 10% and the blue line shows discrepancies at 5%. The incidence of discrepancies is remarkably stable over time, with about 7% of returns featuring some discrepancy. However, there is no discernible seasonality, though there may be a small reduction starting at the onset of the pandemic.

Not all discrepancies are large enough to merit the tax authority's attention, so only taxpayer-months in which the positive discrepancy was at least 10M PYG (approximately USD 1,400) were considered eligible.

II.iii Intervention

Our intervention is a modified version of the intervention studied by Carrillo *et al.* (2017) in Ecuador. Among the eligible taxpayers, taxpayers were randomly selected (following the protocols described in section II.v) to be sent a notice. The notice presents taxpayers with a summary by month of their reported sales, the purchases reported by their clients, and the discrepancies between them. The invoices underlying the purchases reported by clients are attached as an annex to back up the discrepancies. The notice then requests that the taxpayer file amendments of the relevant tax returns to address the inconsistencies. A sample of the notice is included below in Figure I.

Carrillo *et al.* (2017) present somewhat disappointing results from the rollout of this type of intervention in Ecuador. Specifically, two issues arose. First, the overall response rate in Ecuador was fairly low: only 10–20% of notified taxpayers made an amendment. Second, those taxpayers that did respond did not become more compliant: each dollar of increases in reported sales tended to be accompanied by increases in reported purchases of up to 96 cents, leading to little or no change in reported value added and hence no detectable increase in tax liabilities or tax payments.

To address these issues we made two changes to the notices. First, we strengthened the language used. Specifically, the notices emphasized that the request was for the taxpayer to file an amendment to their tax return in which they increase their reported sales to account for the discrepancy, but not make any other amendments to their return. This was intended to avoid taxpayers simultaneously increasing their reported purchases.¹⁹ Second, the notices were accompanied by greater follow-up and sanctions for non-response. In particular, auditors were assigned to follow up on each case. If the taxpayer responded but did not amend or amended only partially, the auditors reviewed the case and made a determination about whether to accept the partial response. If the taxpayer did not respond by the 10-day deadline, or responded unsatisfactorily, the auditor blocked the taxpayer's ID (RUC). Blocked IDs are not able to request more tax invoices, and are unable to perform a range of other processes with the tax authority (though they are still able to file and amend returns), potentially disrupting business activities

¹⁹Naturally, the tax authority cannot prevent taxpayers from filing amendments that amend many lines on their returns. Rather, the message was intended to be (highly) suggestive.

for non-compliant taxpayers, especially larger ones that rely more on being able to issue tax invoices and have more complicated interactions with the tax authority.

While potentially more effective than a lighter touch intervention (indeed our results are much more promising than those for the lighter touch intervention in Ecuador ([Carrillo et al., 2017](#))), this design requires auditors to review each file and to correspond with the taxpayers if needed. This requires a significant amount of time (see section [V.i](#) for a detailed estimation), so the number of available auditors directly constrains the number of notices that can be sent. This makes targeting the notifications effectively a first-order challenge, an issue we return to in section [V.iii](#).

II.iv Timeline

The notices were sent out in waves given the limited administrative capacity to follow up on the status of each notice. There was a great deal of uncertainty about how much follow-up attention each notice would require, since there was uncertainty about how many taxpayers would respond and how many would reply seeking to explain away the discrepancies. The response rates were better than expected and the number of taxpayers seeking to explain away the discrepancies smaller than expected, so we were able to make the later waves larger than the early wave.

We performed an operational pilot of 50 notices in January 2022. Our first wave of 491 notices was sent in August 2022. A second wave of 998 notices was sent in December 2022. We then sent waves of 1,992 and 1,998 notices in August and October 2023, respectively.²⁰ This yields a total of 5,529 experimental notices sent so far.²¹ In the second phase of the experiment we anticipate sending 5,000 notices in November/December 2024 and 5,000 in March/April 2025.

Our experimental design evolved across four waves. The first two waves employed a standard cluster saturation design ([Baird et al., 2018](#)). In the subsequent waves, however, we implemented a novel network experimental design optimized specifically for applying randomization inference.

II.v Waves 1-2: Cluster Randomization

The cluster saturation design ([Baird et al., 2018](#)) is a two-stage randomization process designed to permit the detection and measurement of both direct treatment effects and indirect effects on neighboring taxpayers. The saturated cluster design proceeds in the following steps:

1. We create clusters of suppliers based on who their clients are. The idea is that treating a common supplier may have indirect effects on all their clients, including both those mentioned in the notification and those who are not. To do this, we first create a network of

²⁰2,000 taxpayers were randomly selected to receive a notice, but a small number could not be reached, or had already made an amendment when the notices were built.

²¹In the descriptive analysis in section [III](#) we include the pilot, but in our estimation of treatment effects in section [IV](#), we do not, leading to a sample of 5,429 treated taxpayers for that analysis.

clients where links are formed between pairs of clients who share a supplier. We then partition this network of 185,119 clients into 6,486 clusters using the Leiden algorithm (Traag *et al.*, 2011). We then assign each of the suppliers to one cluster based on which cluster of clients they supply to the most.

2. We randomly assign clusters to 0%, 50% or 100% saturation rates of eligible taxpayers.
3. For Wave 1, we chose a random sample of 500 taxpayers following the clusters' assigned saturation rates.
4. For Wave 2, we chose a random sample of 1,000 taxpayers from the remaining eligible taxpayers, not treated in Wave 1, following the clusters' assigned saturation rates.

This construction leads to 3,097 distinct clusters. The sizes of these clusters varies and can be approximated by a logarithmic-exponential distribution with rate parameter 1.27. The intra-cluster connectivity is 1-2% and the between-cluster connectivity is 0.05-0.1%, both approximated via Monte Carlo sampling, suggesting that we have discovered meaningful clusters in the buyer-seller network.

II.vi Waves 3-4: *Split-Graph* Design

For the 2023 waves, the randomization protocol was refined to maximize statistical power for detecting both *direct effects* and *spillover effects* mediated by the observable network of business transactions between taxpayers. Our choice of experimental design and parameters has been guided by econometric work in randomization inference, to be described in greater detail in a companion paper. At a high level, the experimental design, which we refer to as *Split-Graph*, stratifies taxpayers according to their connections to other eligible taxpayers and their degree (number of trading partners), and then randomized treatment within types.

Split-Graph Randomization. To select 2,000 taxpayers for treatment in Waves 3 or 4, we proceed in four steps. A visual representation of the Split-Graph design is shown in Figure IV.

1. Partition the network of taxpayers into three sets based on how they are connected (we consider both buyer and seller connections equally here). Define **NE** as the set of ineligible taxpayers at the time of treatment; **E₁** as the set of all eligible taxpayers that are *not* connected to any other eligible taxpayer (i.e. it is only connected to taxpayers in **NE**); and **E₀** the set of eligible taxpayers that are connected to at least one other eligible taxpayer.
2. Assign each taxpayer to one of 12 types $x = (\text{degree}, \text{orientation})$. Taxpayers' degrees are classified as *small*, *medium*, *high*, or *very-high* based on the quartiles of the distribution of taxpayers' degree. A taxpayer's *orientation* is either *out*, *in*, or *neutral* depending on whether it connects mainly to ineligible taxpayers ($> 60\%$ of its connected taxpayers

are in \mathbf{NE}), or mainly to eligible taxpayers ($> 60\%$ of its connected taxpayers are in \mathbf{E}_0 or \mathbf{E}_1), or to both with roughly equal proportions (between 40%-60%), respectively. Since eligible taxpayers are the targets of the notifications, eligible taxpayers with inward orientation could have a higher incidence of discrepancies as opposed to eligible taxpayers with outwards orientation.

For taxpayers in \mathbf{E}_0 we compute the anticliques $F(x; \mathbf{E}_0)$ for each type x . An anticlique (or independent set) is a set of taxpayers such that no two taxpayers are connected to each other in the inter-taxpayer network \mathbf{G} . Then, we search for the maximal sets under the constraint $|F(x; \mathbf{E}_0)| \leq mN_x$, where N_x is #treated taxpayers in type x as defined in the treatment schedule, \mathbf{N}_{E_0} presented in step 4 below, and m is a treatment intensity parameter.

3. Treat 100 taxpayers in \mathbf{E}_1 (out of $\sim 6,300 \approx 1.6\%$ treatment intensity) completely at random stratified by type as follows:

	out	neutral	in
small	90 (1.6%)	0	0
medium	10 (1.6%)	0	0
high	0	0	0
very-high	0	0	0

The numbers in parentheses indicate the proportion % of treated taxpayers of each type (i.e., degree and orientation). Note that by definition, all taxpayers in \mathbf{E}_1 have an “out” orientation.

4. Treat 1,900 taxpayers in \mathbf{E}_0 . The treatment schedule by type is

	out	neutral	in
small	870 (8.5%)	110 (6.9%)	99 (10.1%)
medium	720 (3.5%)	50 (9%)	15 (28.3%)
high	32 (1.9%)	0	0
very-high	4 (5.1%)	0	0

This gives us N_x , the overall number of taxpayers to treat for any type x . Treatment of these N_x taxpayers is assigned as follows. For any type $x = (\text{degree}, \text{orientation})$, with $Q(x) = \{j \in \mathbf{E}_0 : x_j = x\}$ the total type set, we execute the following randomization:

- Treat $N_{x,F} = pN_x$, for some fixed $p \in (0, 1)$, taxpayers in the anticlique set $F(x; \mathbf{E}_0)$ chosen completely at random.
- Treat $N_x - N_{x,F}$ taxpayers at random in the remainder set $Q(x) \setminus F(x; \mathbf{E}_0)$.

Next, we discuss the parameters of this experimental design, and how we selected their optimal values.

Optimal design The Split-Graph design described above is controlled by the following four parameters:

- Treatment schedule matrices: \mathbf{N}_{E_0} and \mathbf{N}_{E_1} .
- Treatment intensity parameters: $m \in \mathbb{N}^+, p \in (0, 1)$.

The treatment schedule matrices are presented above. The intensity parameters were set at $m = 5$ and $p = 0.7$. These parameter values were obtained after intensive simulation-based power studies, which we outline below.

Initially, we picked baseline values for the treatment schedules, say $\mathbf{N}_{E_0}^b$ and $\mathbf{N}_{E_1}^b$. We then ran power simulations over perturbations, $\mathbf{N}^b + \varepsilon$, of these baseline designs and a permuted grid of values for the intensity parameters (m, p) following a space filling design (Box *et al.*, 1978; Joseph, 2016). The design parameters were then served as input to a random forest model, while the corresponding power experiment results were used as the output. Using the fitted random forest model, we were able to determine the design parameter values that were estimated to yield optimal power according to the model. The optimal values revealed an interesting trade-off between estimating direct and spillover effects: For direct effects, it is best to treat large anticlique sets with medium intensity (large m , small p), while for spillover effects it is best to treat small anticliques with high intensity (small m , large p).

II.vii Balance

To verify that the randomization was implemented correctly, and to ensure that key attributes of the treated and control taxpayers are similar (to reduce concerns about confounding variables that could compromise the interpretation of the treatment effects), we check for balance of a number of covariates. As our randomization protocols varied across waves, we present balance results separately for Waves 1–2 and Waves 3–4.

Waves 1–2 followed the saturated cluster randomized design described in Section II.v. In clusters assigned to 0 or 100% saturation we cannot compare treatment and control taxpayers. Instead, we summarize the covariates of taxpayers in the cluster into *cluster-level covariates*, and then test balance at the cluster level. Specifically, we perform the following steps to diagnose balance:

1. For covariate Z with levels $k = 1, \dots, K$, calculate the cluster covariate for each cluster J :

$$C_J = \text{entropy}(\{c_k\}_{k=1}^K), \quad c_k = \sum_{i \in J} 1\{Z_i = k\} / |J|.$$

Here, $\{c_k\}_{k=1}^K$ is the empirical distribution of X within cluster J , while C_J is the entropy of this distribution, which serves as the cluster-level covariate.

2. Regress C_J on p_J , where p_J is the saturation rate for cluster J . Report the p -value of the t -test from the regression.

The results from this calculation are shown in Table I. We show that the legal form (incorporated/unincorporated), the size (small/medium/large), the region, the economic activity, and the age group of the taxpayers are well balanced across clusters (p-values are well away from 0).

In Waves 3-4, the randomization changed to a stratified network experiment as described in Section II.v. As shown in Figure IV, treatment randomization consisted of multiple simple randomized experiments within sets of taxpayers with the same type x and anticlique membership status. As a result, we can test for balance at the taxpayer level as follows:

1. Within each wave, calculate each taxpayer i 's type x_i (degree, orientation); treatment $D_i \in \{0, 1\}$; and anti-clique status $A_i \in \{0, 1\}$.
2. For each covariate Z , fit two multinomial regression models:

$$M_0 : Z_i \sim x_i + A_i, \text{ and } M_1 : Z_i \sim D_i + x_i + A_i,$$

and report the p -value from the ANOVA comparison between M_0 and M_1 . If treatment is predictive of the covariate, then such p -value should be small.

The results from this calculation for Waves 3 and 4 are shown in Table I for the same covariates as considered for Waves 1 and 2. The results indicate good balance for both waves, indicating that the experiment's randomization was successfully implemented and that comparing treatment taxpayers to control taxpayers will yield reliable estimates of the causal impact of the notifications. Before turning to the causal effects, we begin with some descriptive evidence on how notified taxpayers responded to the notices they received.

III Descriptive Evidence

We begin our exposition of the results of the experiment by studying what happened to the taxpayers who were notified, and in section IV we incorporate the control taxpayers to estimate causal impacts of the notifications.

Figure V shows the outcomes of each case as of August 2024, dividing the notified taxpayers into 4 groups: Those who did not respond; those who responded but sought to explain the discrepancies without amending their returns; those who made amendments that only partially addressed the discrepancies; and those who made amendments that addressed the discrepancies. Overall, the response rate is around 70%.²² Comparing this to the 10–20% of taxpayers who responded in Ecuador (as reported in Carrillo *et al.*, 2017), the modifications to the intervention to make it more effective seem to have delivered the desired effect. Interestingly, while the overall response rate is similar across the waves, the rate of partial amendment is somewhat higher in the 2022 waves than in the 2023 waves.

²²Note that the fraction of taxpayers who did not respond is higher in the October 2023 wave, but as figure VI shows, this is because responses continue to arrive many months after the notification, and our data stops 9 months after the October 2023 wave.

By statute, notified taxpayers are given 10 working days within which to respond to the notices. Figure VI traces out the response rates in each wave over time. They show the total proportion of notified taxpayers who have responded as time (measured in working days) elapses from the day they were notified. The vertical dashed lines denote the deadline for each wave.²³ We see a significant increase in responses around the deadline: the response rate is at 15–20% just before the deadline, but rises to 25–30% as the deadline expires. Perhaps more interestingly, the response rate continues to increase. By the time our data ends, the response rate for the earlier waves appears to have leveled off at around 65–70%, more than double what it was when the deadline expired, and it is still increasing slowly. This is most likely the effect of taxpayers having their tax IDs blocked, and then responding to the notices when this blockage starts to interfere with their business activities.

In figure VII we turn to the content of notified taxpayers’ amendments. For each wave of notified taxpayers we track the cumulative changes in outcomes implied by their amendments (e.g. the cumulative increase in reported taxable sales). We then divide it by the number of notified taxpayers to create measures of the average response per notified taxpayer.²⁴ Panels A–C report cumulative increases in sales, purchases, and tax liabilities, respectively. Panel A shows clear increases in reported sales, with all four waves reporting an average of over PYG 200 million (USD 28,000) by the end of our data. Interestingly, the later waves appear to be trending upwards faster.

Even though it was not part of the requirements of the notifications (the notifications explicitly asked taxpayers *only* to amend their reported sales), many notified taxpayers chose to also amend their reported purchases. As panel B shows, by the end of our data notified taxpayers had reported an average of an additional 30–35 million PYG of sales. However, since the increase in reported purchases was an order of magnitude smaller than the increase in reported sales, average reported tax liabilities increased sharply as well, as seen in panel C, with tax liabilities of notified taxpayers an average of 17–25 million PYG (2,300–3,400 USD) higher by the end of our data.

Panels D–F study changes to the notified taxpayers’ net position with respect to the tax authority. Panel D documents large increases in payments from notified taxpayers to the treasury, averaging PYG 1 million by the statutory deadline to reply, but continuing to grow thereafter. By the end of our data 20 months later, the August 2022 wave had reached PYG 3.9 million per notified taxpayer. We see a very similar pattern in panel E studying credits the taxpayers have from the tax authority. These decrease sharply, also reaching around PYG 4 million less by the end of our data. These reductions in credits represent higher tax payments by these taxpayers in the future, and so to combine them with the increase in immediate tax payments in panel D, we discount reductions in credits by a factor of 0.7 to create a measure of the net value of the notification.²⁵ This measure of the net value of the notified taxpayers’ amendments climbs

²³These can deviate from the 10 working days because of holidays and a miscalculation in the August 2023 wave.

²⁴This normalization allows us to compare across waves of different sizes.

²⁵Specifically, the net value is the change in payments - 0.7 × the change in credits.

sharply and steadily in panel F, with the August 2022 wave reaching PYG 6.5 million by the end of our data.

Finally, we study heterogeneity of these responses across types of notified taxpayers, focusing on their size classification (small/medium/large) and their legal form (incorporated/unincorporated). In the sample of amendments by notified taxpayers, we regress the outcome of taxpayer i on indicators for the different groups $g(i) \in \{\text{Incorporated, Unincorporated}\} \times \{\text{Small, Medium, Large}\}$ that the taxpayers i belong to:

$$y_i = \beta_h \mathbf{1}[g(i) = h] + \varepsilon_i, \quad (6)$$

and we report the β_h coefficients.²⁶

Table II shows how different the response rates of different types of taxpayers were. We see that small, unincorporated taxpayers are the most likely to respond, with 76% of them amending by the end of our data, while at the lower end, medium-sized, unincorporated taxpayers only respond in 28% of cases. However, small taxpayers' amendments tend to be smaller. They report an average of PYG 177 million additional sales, 30 million of additional purchases, and a tax liability that is PYG 13 million higher. As a result, they yield benefits of PYG 3.2 million per taxpayer. By contrast, medium, incorporated taxpayers make far larger amendments, with an average net value to the tax authority of PYG 26 million. Interestingly, among small taxpayers, unincorporated taxpayers are more likely to respond than incorporated taxpayers. However, among medium-sized taxpayers, the opposite is true with unincorporated taxpayers less than half as likely to respond. These findings suggest that factors such as the size and legal status of the taxpayers should be considered when targeting taxpayers for notification.

In Table III, we present baseline summary statistics of key outcomes measured in our regression. We note that our sample only includes 14 large firms, which is consistent with their point estimates being indistinguishable from zero. As well, we see heterogeneity in the effect size by firm type. For instance, small unincorporated and incorporated firms increase their reported sales by 70% and 40%, respectively. This decreases for medium unincorporated and incorporated firms, only changing their reported sales by 17% and 22%, respectively. The magnitude in change in reported purchases is much smaller, with small unincorporated and incorporated firms changing reported purchases by 16% and 8%, respectively. Medium firms regardless of legal classification change their reported purchases by less than 5%. While the results of the change in tax liability are slightly harder to interpret, we note that small unincorporated firms have an increased tax liability of 8583%.

Overall, the results paint a promising picture. In the medium run, 65–70% of notified taxpayers respond, and yield benefits to the tax authority of around PYG 5,000,000 per notification sent. Of course, these estimates are not causal effects of the policy: Some of the amendments by notified taxpayers may have occurred even without the tax authority sending them a notification. Therefore, the next section incorporates the control taxpayers and exploits the random assignment of the notices to estimate the causal effects on amendments.

²⁶There are no unincorporated, large taxpayers in our sample so we omit this category.

IV Causal Impacts

We want to (i) test whether the audits increase the likelihood that taxpayers file an amendment to their VAT tax return, (ii) assess when these effects materialize, (iii) estimate how these amendments influence taxpayers' tax liability and payments to the tax authority, and (iv) verify which firms drive these effects.

Sub-section IV.i addresses the first three of these objectives. We do so by building randomization-based confidence intervals for a hypothesized constant treatment effect. This randomization-based analysis is finite-sample valid under a stability assumption on the potential outcomes, which we describe below. We favor this approach as it can naturally be extended to provide valid tests for spillover effects—and relied on it to guide our experimental design in Waves 3 and 4.

Sub-section IV.ii replicates our main analysis on the effects of tax audits on amendments, but uses the model-based inference commonly used in applied microeconomics. After showing that our main results also hold with this alternative inference method, we expand them by estimating the heterogeneity in treatment effects by taxpayers' type.

IV.i Effects on Amendments (design-based approach)

Preliminaries. Let N denote the total number of taxpayers observed over a period $\{0, 1, \dots, T\}$. In our setting, $T = 1,546$ days with the earliest date being “2019-12-06” and the latest being “2024-02-29”. The treatment assignment is denoted by the $N \times T$ matrix $\mathbf{D} = (D_{it})$ for $i = 1, \dots, N$ and $t = 0, \dots, T$. The dimension of \mathbf{D} is implicit to ease notation. The indicator $D_{it} = 1$ represents the event that taxpayer i receives a notification on date t .²⁷ The experiment design, $\text{pr} : \{0, 1\}^{N \times T} \rightarrow [0, 1)$, is known and is described in detail in the following section. Let $\mathbb{D} = \{\mathbf{d} \in \{0, 1\}^{N \times T} : \text{pr}(\mathbf{d}) > 0\}$ denote its support.

Over the course of the experiment, together with the tax authority, we modified the criteria that determined which taxpayers could be audited, so that eligibility varies across waves. We take this into account in our analysis. Thus, taxpayer i is called *eligible* at time t if they can be treated at time t , i.e. $\text{pr}(d_{it} = 1) > 0$. An *ineligible* taxpayer at time t cannot be treated at that time. At any time t , the set of non-eligible taxpayers will be denoted by \mathbf{NE}_t and the set of eligible taxpayers by \mathbf{E}_t . A large majority of taxpayers were not eligible, such that $|\mathbf{NE}_t|/|\mathbf{E}_t| \approx 10$.

The taxpayers are connected through a buyer-seller network measured in 2021, before any treatment is administered. Let $\mathbf{C} = (c_{ij})$, for $i, j \in \{1, \dots, N\}$ denote the $N \times N$ weighted adjacency matrix of this buyer-seller network. The value c_{ij} on the edge between taxpayer i and taxpayer j is non-negative and represents the total monetary value that i (buyer) purchased from j (seller), as reported by taxpayer i . From this matrix, we can define $w_{ij}^b = c_{ij} / \sum_j c_{ij}$, $w_{ij}^b \in [0, 1]$ as the weighted value of seller j for buyer i . We can also define $w_{ij}^s = c_{ij} / \sum_i c_{ij}$, $w_{ij}^s \in [0, 1]$

²⁷We do not require that treatment is an “absorbing state.”

as the weighted value of buyer i for seller j . The $N \times N$ binary adjacency matrix, $\mathbf{G} = (g_{ij})$, represents all possible connections between taxpayers in the network, such that $g_{ij} = 1\{w_{ij} + w_{ji} > 0\}$, for all i, j . We will use both the binary adjacency matrix and the weighted matrix to study spillover effects in the next step of our project. We emphasize that these matrices are observed before any treatment assignment, and so they can be used as fixed covariates in our analysis.

Let $w(t)$ denote the step function that maps t to the date of the latest wave that is no later than t . For instance, for $t = \text{"2022-12-25"}$ it holds that $w(t) = \text{"2022-12-19"}$, which is the treatment date of Wave 2; for $t' = \text{"2023-09-10"}$ it holds that $w(t') = \text{"2023-07-31"}$, which is the treatment date of Wave 3, and so on. Moreover, let $\text{week}_{t'-t} \in \mathbb{Z}$ denote the distance in weeks as an integer between dates t' and t , allowing for negative values if t is later than t' . For instance, using the example values above, we obtain $\text{week}_{t'-t} = 37$ and $\text{week}_{t-t'} = -37$.

One important outcome of interest is $A_{it}(\mathbf{d}) \in \{0, 1\}$ denoting whether taxpayer i files an amendment at date $t \in [0, T]$ under treatment schedule $\mathbf{d} \in \mathbb{D}$. To simplify, throughout our analysis we make a “no anticipation” assumption, whereby any outcome measured at t depends only on treatments received up to t :

$$A_{it}(\mathbf{d}) = A_{it}(\mathbf{d}_{:t}), \quad (7)$$

where $\mathbf{d}_{:t}$ denotes the sub-matrix of \mathbf{d} that only includes times from 1 to t . This assumption is reasonable in our setting, since the treatment day is defined as the day on which the taxpayer receives the notification and this is the first time the taxpayer learn that they are being scrutinized.

As such, A_{it} is an intermediate outcome that “mediates” other outcomes of interest, which we will describe next. Measuring the impact of our intervention on increasing the amendment rate will be one important goal of our analysis.

Whenever a taxpayer amends a tax return, they specify which boxes on the return they are modifying, the pre-amendment value in that box, and the amended value that they wish to declare in that box. We take the amended boxes and aggregate them to coarser outcomes, such as the change in the monetary value of taxable sales, taxable purchases, tax payments, tax credit, or tax liability for the reported period. The reported change could be zero for some of these outcomes but not all. A taxpayer does not need to receive a notice to initiate an amendment. In fact, the amendment rate in the baseline pre-treatment period is about 1% of taxpayers amending in a week, which is similar to the amendment rate of the *untreated* taxpayers during the treatment periods.

Let $Y_{it}(\mathbf{d}_{:t})$ denote the change in outcome Y , where $Y \in \{\text{sales, purchases, liability, credits, payments}\}$, as reported by firm i at time t . Note that we only observe Y when the firm actually files an amendment, i.e., $A_{it}(\mathbf{d}_{:t}) = 1$.

We will adopt the convention that $Y_{it}(\mathbf{d}) = 0$, for any \mathbf{d} for which $A_{it}(\mathbf{d}) = 0$. Although this type of “padding with zeros” is common in practice to handle unobserved outcomes uniformly, it can have implications for the analysis. Here, we justify this approach by noting that taxpayers not filing an amendment imply a zero value outcome for the tax authority, while those filing an

amendment may present either a positive or negative change.

However, this assumption prevents us from distinguishing taxpayers who file amendments reporting zero change ($A = 1$) and ($Y = 0$) from those not filing an amendment at all ($A = 0$). Consequently, this approach effectively assumes all taxpayers are “always takers” in compliance terms, where $A_{it}(\mathbf{d}_{:t}) = 1$ for all i, t and \mathbf{d} , with taxpayers either reporting a change $Y \neq 0$ or no change $Y = 0$. Although useful here, this assumption is not strictly necessary, and we will explore alternatives in future work.

Null hypotheses. With these definitions, we now turn our attention to the null hypotheses of interest. In this report, we consider the null hypothesis of a constant direct treatment effect. To define this null hypothesis we need an assumption as follows.

Assumption 2. *The potential outcome of each taxpayer, when they file an amendment, depends only on the taxpayer’s individual treatment status and the particular week on which the amendment was filed. That is, there exist functions $f_i : \{0, 1\} \times \mathbb{Z} \rightarrow \mathbb{R}$, all potentially unknown, such that*

$$Y_{it}(\mathbf{d}_{:t}) = f_i(d_{i\mathbf{w}(t)}, \text{week}_{t-\mathbf{w}(t)}).$$

Assumption 2 is a form of the stable unit treatment value assumption (SUTVA), commonly used in causal inference. This assumption has two main implications. First, it asserts that there is *no interference* between units: the potential outcome of a unit depends only on the treatment assigned to that unit, without being affected by treatments assigned to other taxpayers. While this assumption may not always be realistic, we use it here as a first-order approximation for estimating the direct effects of tax audits.

Under Assumption 2, we may define a null hypothesis of constant treatment effect as follows.

$$H_0^{te}(\tau) : f_i(1, w) = f_i(0, w) + \tau, \text{ for all } i \in [N], w \in \mathbb{Z}. \quad (8)$$

This null hypothesis implies that notifying a taxpayer has a treatment effect τ that remains constant across taxpayers and weeks. Note that the functional definition of $f_{it}(0, \cdot)$ may be arbitrary.

Inference. Our goal is to build randomization-based confidence intervals for hypothesized constant treatment effects through inverting null hypotheses like H_0^{te} defined in Equation (8). The randomization-based analysis is finite-sample valid under Assumption 2.

To test outcomes across various weeks following the receipt of the notifications, our design follows a *stacked data specification*. This specification is shown in Figure VIII. The figure shows that for a positive-value week (e.g., 1, 2, ...), our randomization tests compare the outcomes of eligible taxpayers observed in that particular week to the outcomes of the same taxpayers within 30 days leading to the treatment event (marked as “controls” in the figure). For a nonpositive-valued week (0, -1, -2, ...), the randomization test compares the outcomes of eligible taxpayers in that week to their outcomes in the 30 days prior to that week.

For each outcome and a fixed week $w \in \{-4, \dots, 0, \dots, 4\}$, we test the hypothesis $H_0^{te}(\tau)$, defined in Equation (8), for various values of τ . We then invert the test by keeping those values of τ that correspond to a randomization p -value that is no larger than 0.05, producing the randomization-based confidence interval. We emphasize that the validity of these intervals does not rely on correct specification of the model, but relies solely on the known randomization procedure and Assumption 2.

Results. The randomization-based 95% confidence intervals (CIs) for the direct treatment effect —assuming no interference— are shown in Figures IX (where the outcome is amendment) and X where each panel corresponds to outcomes related to Tax Liability, Payments, Credits, Purchases, and Sales.

From this analysis, we glean four empirical findings. First, the notifications had a strong quantitative effect on amendment rates. Figure IX shows that the treatment and control groups behaved similarly in the weeks leading up to the notifications, but the treatment group diverged sharply and immediately following the notifications. Our estimates imply that 800 notified taxpayers ($0.145 \times 5,000$) filed amendments to a tax return in the second week after a notification *because* of their notification.

Second, the treatment effects decay, but are persistent. There is a peak in all outcomes two weeks after the notification, which is taxpayers' deadline to respond. However, we study outcomes up to four weeks after the notifications and we see significant effects even four weeks after the notifications.

Third, the amendments were beneficial for the tax authority. Taxpayers reported more sales and more purchases, but the additional purchases did not fully offset the additional sales, leading to higher tax liabilities, higher payments to the tax authority, and reduced credits owed to taxpayers.

Fourth, the pattern of treatment effects for other outcomes in Figure X all mirror the shape of the treatment effects on amendment. This is consistent with the constant treatment effect assumption that the amendments that occurred because of the notifications were all larger (reporting more sales, slightly more purchases, higher tax liabilities, higher payments to the treasury, reducing credits) by a similar amount, while the pattern over time is driven by the pattern in the amendment rate.

Overall, the results suggest that the notifications were highly successful across a range of outcomes. To summarize them, panel F of Figure X shows treatment effects on the summary measure of the net value of an amendment described in section III.²⁸ Again, we see sharp treatment effects immediately following the notifications that are strongest in week two, but persist until the end of our analysis in week four.

²⁸Specifically, the net value is the change in payments - $0.7 \times$ the change in credits.

IV.ii Heterogeneity of Effects Across Taxpayers (model-based approach)

Replicating our main results. We first seek to replicate our main results from the preceding section on the effects of tax audits on the likelihood of amending and the type of amendment when using a different type of statistical approach: a model-based approach.

Specifically, we use the stacked difference-in-differences method introduced in [Cengiz *et al.* \(2019\)](#). It is described, and compared to other recent approaches in the two-way fixed effects difference-in-differences literature, in [Baker *et al.* \(2022\)](#). The idea is to create notification wave-specific “clean 2×2” datasets. Each of these datasets includes taxpayers that were not treated before the notification wave (i.e., they are all “clean” at baseline), with some randomly selected to be treated during the notification wave and the others remaining in control. Each of these datasets include a balanced panel with the same time window before and after the notifications. We implement this strategy in our setting by constructing one dataset for each of the four waves of notices, with data from 60 days before and 60 days after the notification for each eligible firm—excluding taxpayers that have been treated in earlier waves. We then stack each of these 4 audit-specific datasets into a single dataset to estimate

$$y_{kit} = \beta D_{kit} + \omega p_{kit} + \mu_{ki} + \delta_{kt} + \varepsilon_{kit} \quad (9)$$

where y_{kit} is the outcome for taxpayer i at date t in notification wave k (e.g. indicator for filing an amendment; filing an amendment that increased reported sales; etc.); D_{kit} is an indicator equal to 1 if taxpayer i was notified within wave k and if date t is after the notification date for that wave; p_{kit} takes value 0 before the notification date, and is otherwise equal to the probability that taxpayer i was randomly selected to receive a notice within wave k ; μ_{ki} are notification wave-by-taxpayer fixed effects; and δ_{kt} are notification wave-by-time fixed effects.

The inclusion of p_{kit} accounts for the fact that, due to our randomization procedure, taxpayers with different characteristics might have different likelihoods of being treated, and allows us to identify the direct causal effect of audits with β . We compute p_{kit} empirically, by simulating the treatment assignment 10,000 times. We report the estimated value of β , and the associated p-values for the test $H_0 : \beta = 0$ from both standard asymptotic inference and using randomization inference.

We find that taxpayers are 4.5 times more likely to file an amendment to their tax returns when they have been notified of discrepancies. Table IV, Col. 1 reports that notified taxpayers are 0.6 percentage points more likely to make an amendment on any given day, off a baseline amendment rate of 0.14%. In line with the focus of the notifications on reported sales, we find that the likelihood of filing an amendment that increases reported sales strongly increases. Interestingly, we find that being audited also increases the likelihood of filing an amendment that increases reported purchases (hence reducing reported value-added, all else equal). This effect, albeit much smaller than for sales, potentially indicates a strategic response by taxpayers to mitigate the effects of the notifications on their total tax liability.

Overall, we find that notified taxpayers are 6.6 times more likely to file an amendment that increases their reported tax liability than non-notified taxpayers (Table IV, Col. 4). This translates into both a 9-fold increase in the probability that an amendment leads to a reduction in the outstanding credits the tax authority the taxpayer has (Col. 5) and a 12-fold increase in the probability that an amendment leads to an increase in the payments due (Col. 6). These results are consonant with the descriptive evidence shown in section III, but now represent causal estimates of the impact of the notifications.

Heterogeneity. The preceding analysis confirms the large effects of the notifications. We now seek to explore how the causal effects of the notifications vary across types of taxpayers.

In Table V we present results of the estimation of heterogeneous treatment effects, focusing on taxpayers' size designation (small / medium / large) and their legal form (incorporated / unincorporated). We do this by extending equation (9) as follows:

$$y_{ijkt} = \sum_j \beta_j D_{kit} \times S_{ij} + Post_t \times S_{ij} + \omega p_{kit} + \mu_{ki} + \delta_{kt} + \varepsilon_{ijkt} \quad (10)$$

Table V shows the results. Three main findings emerge. First, the estimated treatment effects for large taxpayers are indistinguishable from zero. This is, in part, due to the very small number of large taxpayers who were notified (indeed in the third and fourth waves, large taxpayers were ineligible), but mostly due to their greater complexity and size. Second, the effects are stronger for small taxpayers than for medium taxpayers, and stronger for unincorporated taxpayers than for incorporated ones. This is consistent with what we might expect: smaller, less formalized taxpayers have an easier time deliberately misreporting and/or are more likely to make mistakes in their tax declarations. Third, the exception to the pattern just mentioned is that the impacts on the tax liability are largest for the medium-sized, unincorporated taxpayers, however this pattern is not present for increased payments or reduced credits. This suggests that these medium-sized taxpayers are more likely to have more withheld taxes such that changes in their tax liability are less likely to translate into changes in tax payments. This interaction with withholding is an interesting avenue for future exploration.

V Discussion and Next Steps

V.i Cost Benefit Analysis

Thus far our analysis has focused on the benefits from notifications in terms of the resolution of discrepancies and increased tax compliance. However, the notifications also involve costs to the tax authority that should be taken into account in order to evaluate the success or otherwise of the notifications campaign.²⁹ To do this, we work with the tax administration to do a compre-

²⁹The notifications also involve costs to notified taxpayers that could also be taken into account. However, we do not have reliable estimates of the costs to notified taxpayers so we do not include them in our calculations here.

hensive accounting of the time and associated costs incurred in the preparation of notices and the follow-up actions required to process responses by taxpayers.

To estimate the tax administration's costs we attempt to total the cost of attending to all 7,081 notices sent in 2022 and 2023.³⁰ We divide the costs into five phases corresponding to the lifecycle of a notification.

Generation This stage covers the costs of planning the notifications (deciding which types of inconsistencies to study and analyzing example cases before scaling up to the population of taxpayers), the costs of extracting and analyzing data on the universe of taxpayers, and the costs of preparing the notices for each notified taxpayer—including verifying the data being presented in each case. It is estimated that this required a total of 5 days of work by a team of three auditors, one supervisor, and two managers, implying a total cost of USD 1,616.

Transmission This stage covers the costs of actually sending the notifications. These need to be sent to the taxpayers one by one from the tax authority's formal email account. In total, this required one full day of work by one auditor and one manager, for a total cost of USD 117.

Acceptance of Amendments The third phase is the analysis and acceptance of taxpayers' amendments of their returns in response to the notices. As of April 2024 (when the cost estimation was carried out), 2,702 taxpayers had made amendments. These amendments then need to be cross-checked against the initial discrepancies to check that the amendments have eliminated the discrepancy or reduced it to a negligible amount. To process these, one day of work was required from one manager and one supervisor, for a total cost of USD 123.

Analysis of Objections/Explanations The costliest phase by far involves the analysis of the responses from taxpayers who object to the notice or seek to explain away the discrepancy. The objections are made by email and then assigned to an auditor to review. The auditor reviews the arguments and any documentary evidence supplied. The auditor also checks any information provided against other reports in the government's databases and then makes a determination about whether to accept the response. The tax authority estimated that each case required an average of 53 hours of auditors' time. 1,925 taxpayers had objected by April 2024 leading to a total cost of USD 530,010.

Blocking Tax IDs The other very costly phase involves blocking the tax IDs of taxpayers who do not respond or whose responses are not satisfactory. Blocking the tax ID involves a supervisor preparing a form indicating why the tax ID is being blocked and then a manager accepting the

Moreover, it is not immediately clear how much social weight should be placed on the costs borne by taxpayers who are potentially breaking the law.

³⁰The tax authority sent 1,552 non-experimental notices in addition to the 5,529 experimental notices analyzed here.

submission. The tax authority estimates that each case requires an average of 15 minutes of each official's time. As of April 2024, 4,061 tax IDs were blocked or pending blockage leading to a total cost of USD 123,994.

Cost Benefit Combining the above phases, we estimate that the total cost to the tax authority of handling 7,081 cases was PYG 4,853,366,550 (USD 655,860). This implies an average cost per case of PYG 685,407 (USD 93).

Placing this next to the estimated benefits detailed in sections III and IV, we can estimate the average rate of return to sending the notifications. To estimate the average benefit, we use the estimated effects on the total benefit (additional payment - $0.7 \times$ reduced credits) shown in figure X and we assume that over the 8 weeks after our analysis ends, the causal effect is an average of half of the effect 4 weeks after notification and then stops. This is conservative since as we see in figure VII, notified taxpayers continue to make amendments for at least 8-9 months after being notified. This conservative calculation yields an estimated average benefit of PYG 936,535 (USD 127), yielding a rate of return of 37%. This compares very favorably to other types of interventions. For example, the most beneficial version of incentive pay for tax inspectors in the context studied by Khan *et al.* (2016) had a rate of return of 35%. A comprehensive welfare assessment of enforcement notices in Mexico City showed negative rates of return (Brockmeyer *et al.*, 2023).

V.ii Spillovers in Tax Enforcement

The foregoing analysis presents only a partial picture of the full impact of notifying taxpayers. The notices highlight *discrepancies* between taxpayers' reports and their clients' reports to the government. As a result, it is natural to expect that the notifications affect not just the taxpayers who are notified themselves, but also their clients and, potentially, their suppliers.

These spillovers may take many forms. If notified taxpayers perceive that discrepancies between their declared sales and their clients' declared purchases have become more costly, and they convey this to their clients, we would expect notified taxpayers to increase their reported sales and their clients to reduce their reported purchases from notified taxpayers, reducing the discrepancy between their reports. We might also expect that notified taxpayers reduce their reported purchases so as to better match their suppliers' reported sales.

Alternatively, if notified taxpayers perceive that they are under special scrutiny that is specifically targeted at them and less at their clients, we may see that the notified taxpayers amend their returns to address the discrepancy, reducing the need for the clients to amend their returns, leading to fewer, smaller amendments by notified taxpayers' clients.

Finally, we might expect spillovers of the notices onto notified taxpayers' future tax declarations (and possibly also onto their clients' future declarations). If the notices lead taxpayers to update their beliefs about the capacity of the tax administration to detect misreporting, they may increase their reported tax liabilities in the months following receipt of the notices.

All of these are the subject of ongoing analysis of the experimental data. They are facilitated by the design of the experiment as described in section II.v, and will be reported in a companion paper to this one.

V.iii Policy Implications for Targeting

The results from our intervention indicate that sending notices to taxpayers suspected of misreporting their sales is highly cost-effective. It allows the tax authority to increase its collected tax payments and reduce its outstanding tax credits, while costing very little to implement.

Despite this cost-effectiveness, the potential to scale the intervention and increase the number of notices sent to taxpayers is limited by the capacity of the tax authority: each notice requires time from a tax agent to process it and ensure an adequate response by the notified taxpayers. Our results indicate that relaxing these capacity constraints, by hiring additional tax agents to implement this notice-based intervention on a larger scale, would be a very worthwhile investment.

In all likelihood, however, capacity constraints will always exist: it won't be possible to target all eligible taxpayers at once. Improving the targeting of the notices is therefore key to maximize the benefits from the intervention. Given a number of notices that can be sent, which taxpayers should receive them to maximize the task benefits?

To answer this important question, we plan on combining economic theory with a second phase of the experiment. We will create four possible targeting rules and assess their respective effectiveness, as described below.

1. *Random targeting.* The easiest way to decide which taxpayers should receive a notice is to randomly select them among all eligible ones. This is an easy procedure to implement, but is unlikely to select the taxpayers that will maximize benefits to the tax authority.
2. *Risk profiling.* An improvement over random targeting is to target the taxpayers that will respond to the audit the most. One way of doing so is to rely on the tax auditors' information and experience, and let them indicate which taxpayers they expect will update their tax returns and liability the most following a notice. Another, data-driven, way is to estimate how the direct effect of the intervention varies with taxpayers' characteristics and use these estimates to predict which taxpayers would respond the most. We have already conducted such heterogeneity analysis, with results presented in section IV.ii. They suggest targeting in priority small, unincorporated taxpayers.

While an improvement over random targeting, such risk profiling approach is not fully satisfactory as it does not take spillovers into account, and thus is unlikely to maximize the tax benefits from the intervention. Instead, we propose two targeting rules that take spillover effects into account.

3. *Flexible data-based approach.* A first possible approach that accounts for both direct and

spillover effects from tax audits is based on machine learning. With such approach, we will flexible learn from the existing rounds of the experiment and available data how taxpayers' behavior is influenced by their treatment and that of their network members, and predict the tax revenues from many different possible targeting rules. We can therefore choose, in a fully data-driven manner, which targeting rule is expected to maximize tax revenues. [Viviano \(2024\)](#) provides a possible method to do so.

A data-based approach has two main benefits: it can flexible use all available data, and only requires minimal assumptions on the shape of the spillover effects. It also has drawbacks. First, machine learning approaches for heterogeneity in treatment effects can be unstable: running the same algorithm multiple times on the same data can provide vastly different results. This has been shown, among others, in [Engl et al. \(2021\)](#). Second, machine learning approaches can only learn about what has happened in the data they have been trained on. In our context, this means that machine learning approaches can perform very well to predict the effects of targeting rules on the types of taxpayers that have been eligible in our experiment; but could perform poorly when trying to predict the effects of changing the eligibility criteria, or if different types of taxpayers are now eligible.

4. *Economic modeling approach.* A second approach that accounts for both direct and spillover effects from tax audits is based on economic modeling. We will create a model of taxpayers' tax reporting decisions, which will capture key mechanisms for the direct and spillover effects of tax audits. Bringing this model to our existing data, we will be able to test the empirical relevance of various possible mechanisms—hence better learn how firm behave with respect to tax reporting—and derive implications in terms of optimal targeting.

This approach will provide more stable predictions than the flexible data-based approach, will be able to extrapolate predictions for types of taxpayers that haven't necessarily been observed in the data from our earlier experiment, and will clearly show which mechanisms are driving the targeting recommendations. Nonetheless, this approach relies on making specific assumptions about taxpayers' behavior and trying to capture key mechanisms: omitting an important factor from the model can strongly reduce the quality of its targeting rule.

Given the various potential costs and benefits of the different approaches, it is necessary to assess empirically their respective efficacy. We will do so with a second phase of the experiment, to be launched in March/April 2025.

For this second phase of the experiment, we will allocate 5,000 notices based on the four targeting rules described above, and compare their effects on tax revenue. To do so, we will partition the full network of taxpayers into clusters—create groups of taxpayers with intense trading relationships with other members of the group but limited relationships with taxpayers from other groups. We will then randomly assign different clusters to different targeting rules—and compare the tax revenue raised in clusters assigned to different rules. This will allow us

to assess, taking both the direct and spillover effects of the intervention into account, which targeting rule performs best.

With this second phase of the experiment, we will be able to provide relevant policy recommendations to the tax authority as to the optimal way to target their future audits in a context of limited capacity.

VI Conclusion

Obtaining third-party information on taxpayers' economic activity—when combined with effective enforcement actions—can be highly valuable for tax authorities seeking to reduce tax evasion and increase tax collection. We show that matching what buyers and sellers report about their transactions and notifying tax payers of the discrepancies, can be a simple yet effective method to reduce fraud. While earlier evaluations of this idea in other settings had limited success, we show that a version in which the discrepancy notification is backed with a credible threat—in our case, blocking taxpayers' tax identification number if they don't take action—can work very well.

From a first phase of our partnership with the tax authority of Paraguay, we find that the intervention leads to a high response rate among targeted taxpayers (65% within 9 months of notification) and corresponding increases in reported sales. While some taxpayers also strategically respond by increasing their reported purchases (to reduce the tax liability increase induced by their amendments), we find that overall, our intervention led to increased tax liability. This increase in tax liability corresponds to both an increase in payments made to the tax authority and a decrease in credits owed by the tax authority. The tax authority therefore benefits from an immediate increase in collected tax revenue, and an expected increase in future tax collection. Overall, we assess that the direct effect of the intervention on the targeted taxpayers generates a benefit of PYG 936,535 (USD 127) per notified taxpayer and does so at low cost. The notifications have an estimated rate of return of 37% after 12 weeks, comparing very favorably to other types of enforcement activities.

So far, our analysis has focused on the direct effects of the intervention on the targeted taxpayers. However, given the nature of our context and intervention (comparing reports between buyers and sellers, and implementing the intervention among taxpayers that are regularly trading with each other), we can expect the intervention to generate spillovers and also influence the behavior of taxpayers not directly targeted. Ultimately, this can have important implications for the optimal targeting of audits. This will be the focus of future phases of our experiment and partnership with the tax authority.


References

- ALMUNIA, MIGUEL, HENNING, DAVID J., KNEBELMANN, JUSTINE, NAKYAMBADDE, DOROTHY, & TIAN, LIN. 2024. *Leveraging Trading Networks to Improve Tax Compliance: Experimental Evidence from Uganda*. mimeo: CUNEF.
- BAIRD, SARAH, BOHREN, J. AISLINN, MCINTOSH, CRAIG, & ÖZLER, BERK. 2018. Optimal Design of Experiments in the Presence of Interference. *Review of Economics and Statistics*, **100**, 844–860.
- BAKER, ANDREW C, LARCKER, DAVID F, & WANG, CHARLES C Y. 2022. How much should we trust staggered difference-in-differences estimates? *J. Financ. Econ.*, **144**(2), 370–395.
- BESLEY, T., & PERSSON, T. 2013. Taxation and Development. In: AUERBACH, A, CHETTY, R, & FELDSTEIN, M (eds), *Handbook of Public Economics volume V*. Elsevier.
- BESLEY, TIMOTHY, & PERSSON, TORSTEN. 2009. The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, **99**, 1218–44.
- BEST, MICHAEL CARLOS, SHAH, JAWAD, & WASEEM, MAZHAR. 2023. *Detection without Deterrence: Tax Audits with Limited Fiscal Capacity*. mimeo: University of Manchester.
- BOX, GEORGE EP, HUNTER, WILLIAM H, HUNTER, STUART, *et al.* 1978. *Statistics for experimenters*. Vol. 664. John Wiley and sons New York.
- BROCKMEYER, ANNE, ESTEFAN, ALEJANDRO, ARRAS, KARINA RAMÍREZ, & SUÁREZ SERATO, JUAN CARLOS. 2023. *Taxing Property in Developing Countries: Theory and Evidence from Mexico*. NBER working paper 28637.
- CARRILLO, PAUL, POMERANZ, DINA, & SINGHAL, MONICA. 2017. Dodging the Taxman: Firm Misreporting and the Limits to Tax Enforcement. *American Economic Journal: Applied Economics*, **9**, 144–164.
- CENGIZ, DORUK, DUBE, ARINDRAJIT, LINDNER, ATTILA, & ZIPPERER, BEN. 2019. The Effect of Minimum Wages on Low-wage Jobs. *The Quarterly Journal of Economics*, **134**(3), 1405–1454.
- ENGL, FLORIAN, RIEDL, ARNO, & WEBER, ROBERTO. 2021. Spillover effects of institutions on cooperative behavior, preferences, and beliefs. *Am. Econ. J. Microecon.*, **13**(4), 261–299.
- GORDON, R, & LI, W. 2009. Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation. *Journal of Public Economics*, **93**, 855–866.
- JOSEPH, V ROSHAN. 2016. Space-filling designs for computer experiments: A review. *Quality Engineering*, **28**(1), 28–35.
- KHAN, A, KHWAJA, A, & OLKEN, B. 2016. Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors. *Quarterly Journal of Economics*, **131**, 219–271.
- KLEVEN, H. J., KNUDSEN, M. B., KREINER, C. T., PEDERSEN, S., & SAEZ, E. 2011. Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, **79**(3), 651–692.
- NARITOMI, J. 2019. Consumers as Tax Auditors. *American Economic Review*, **109**(9), 3031–72.


- PAGE, LUCY, & PANDE, ROHINI. 2018. Ending Global Poverty: Why Money Isn't Enough. *Journal of Economic Perspectives*, **32**, 173–200.
- POMERANZ, D. 2015. No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, **105**(8), 2539–69.
- TRAAG, V.A., DOOREN, P. VAN, & NESTEROV, Y. 2011. Narrow Scope for Resolution-limit-free Community Detection. *Physical Review E*, **84**.
- VIVIANO, DAVIDE. 2024. Policy targeting under network interference. *arXiv*.

Figuras & Tablas


FIGURE 1: EJEMPLO NOTIFICACIÓN MASIVA




TETÁ VIRU
MOHENDAPY
Ministerio de
HACIENDA



SET
TRIBUTACIÓN
Promoviendo confianza



TETÁ REKUÁI
GOBIERNO NACIONAL



Paraguay
de la gente

Asunción, «fechanota».

Nota de Control Masivo DGFT N° «nota»

«nombre»
RUC:
«ruc»
Correo electrónico: «correo1»

De nuestra consideración:

Tenemos el agrado de dirigimos a Usted, con el fin de informarle que la Dirección General de Fiscalización Tributaria, en virtud de sus facultades dispuestas en el artículo 189 de la Ley N° 125/91 modificada por la Ley N° 2421/04, se encuentra ejecutando procesos de controles masivos a contribuyentes que registran inconsistencias originadas en diversos cruces.

En este sentido, se han detectado diferencias entre las ventas o ajustes de débitos fiscales declaradas por Usted en su(s) Declaración(es) Jurada(s) del IVA donde registra montos menores a los informados por sus clientes en sus libros de compras del Hechauka.

Conforme a los resultantes de cruces y análisis tributarios resultantes del Big Data, esta inconsistencia, representa una irregularidad en la cuantía de ventas o ajustes de débito fiscal registrados por Usted. Es importante mencionar que los análisis de Big Data son realizados sobre todas las informaciones disponibles en la Administración Tributaria y el conocimiento de los perfiles de riesgo conforme a las detecciones de irregularidades producto de controles puntuales.

Por ello se solicita la revisión y rectificación de sus declaraciones de ventas y/o ajustes de débitos fiscales **exclusivamente de los rubros indicados**, en un plazo no mayor a 10 (diez) días hábiles, a partir de la notificación de la presente, conforme a los detalles expuestos en el siguiente cuadro:

PERÍODO FISCAL		VENTAS S/ DDJJ DE IVA		VENTAS S/ LIBRO DE COMPRAS HECHAUKA DE CLIENTES INFORMANTES (*)		VENTAS O AJUSTES NO DECLARADOS		
AÑO	MES	GRAV 10	GRAV 5	GRAV 10	GRAV 5	GRAV 10	GRAV 5	TOTAL
«año»	«mes»	«grav10iva»	«grav5iva»	«grav10hechauka»	«grav5hechauka»	«difgrav10»	«difgrav5»	«diftotal»

(*) EL DETALLE DE LOS DATOS DE LAS FACTURAS QUE RESPALDAN LAS COMPRAS INFORMADAS POR EL CLIENTE SE ENCUENTRA EN ANEXO ADJUNTO A LA PRESENTE NOTA


En caso de constatar alguna irregularidad en las exposiciones, remitir sus observaciones únicamente desde la dirección de correo declarado en el RUC a la dirección «correo2».

Atentamente,

DEPARTAMENTO DE PLANEAMIENTO OPERATIVO
DIRECCIÓN GENERAL DE FISCALIZACIÓN TRIBUTARIA

DECLARACIONES IMPORTANTES:

- En caso de dudas o consultas, comuníquese a los teléfonos «telef1» o «telef2», o recurra a las oficinas de la Dirección General de Fiscalización Tributaria, Departamento «Óptimo», sito en Eduardo Víctor Haedo 128 esq. Independencia Nacional, Asunción.
- La fecha de remisión del presente es el «fecharemisión», por lo que en virtud a la RG 102/13 art 10, la notificación efectiva se considera el «fechanotif».
- El vencimiento del plazo otorgado es el «fechavenc».
- Los controles masivos no admiten solicitudes de prórroga.
- Las rectificaciones solicitadas no requieren Opinión del Auditor Externo Fiscal, por lo que en el caso de que el ejercicio fiscal a rectificar se vea afectado por la RG 67/2015, deberá solicitar la autorización para rectificar mediante correo electrónico a «correo2». Cualquier mensaje de retorno que reciba, no implica el rechazo del correo remitido. Por tanto, tenga la seguridad que su solicitud de habilitación para rectificar fue efectivamente recibida.
- La aprobación de la solicitud de autorización para rectificar llegará a su Buzón del Marandú en el Sistema de Gestión Tributaria de Marandú dentro de las 24 horas siguientes a la solicitud realizada.
- En caso de que las Declaraciones Juradas queden en estado de "NO APLICAR", el trámite para solicitar la aplicación deberá ser realizado por el contribuyente en las respectivas áreas según la jurisdicción a la que pertenece.
- En el descargo efectuado podrá anexar las documentaciones que considere pertinente para respaldar sus argumentos.
- Se recomienda el apoyo técnico en su profesional contable o auditor externo impositivo en el análisis de las transacciones eventualmente sospechadas de irregular.
- Big Data es una tecnología que permite el almacenamiento, procesamiento y análisis masivo de datos. En la SET se da uso de esta herramienta entre otros fines, para detectar posibles casos de contribuyentes que emiten o utilizan facturas de contenido falso.

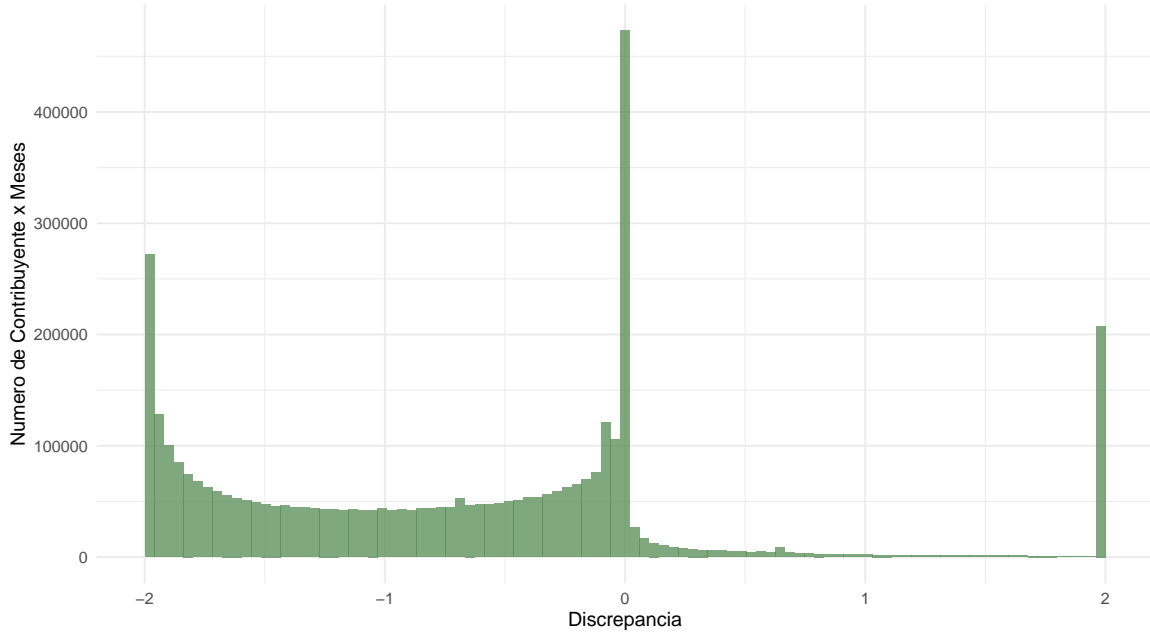


Eduardo V. Haedo 128 esq. Independencia Nacional

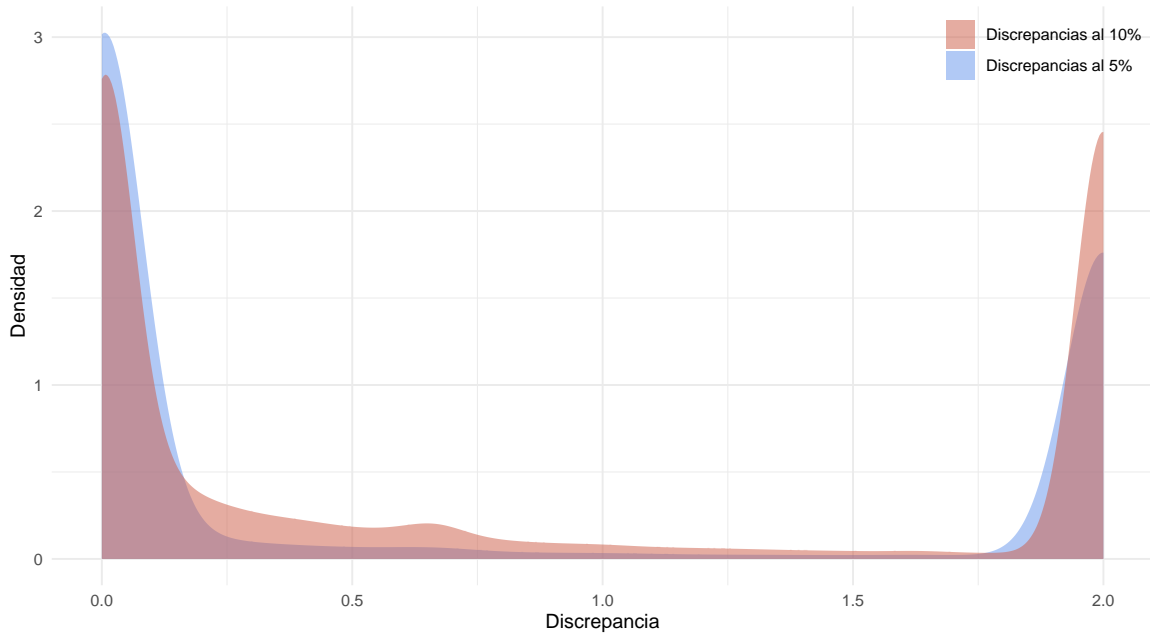
56

FIGURE 2: DISTRIBUCIONES DE DISCREPANCIAS 2019–2021

PANEL A: DISTRIBUCIÓN DE DISCREPANCIAS



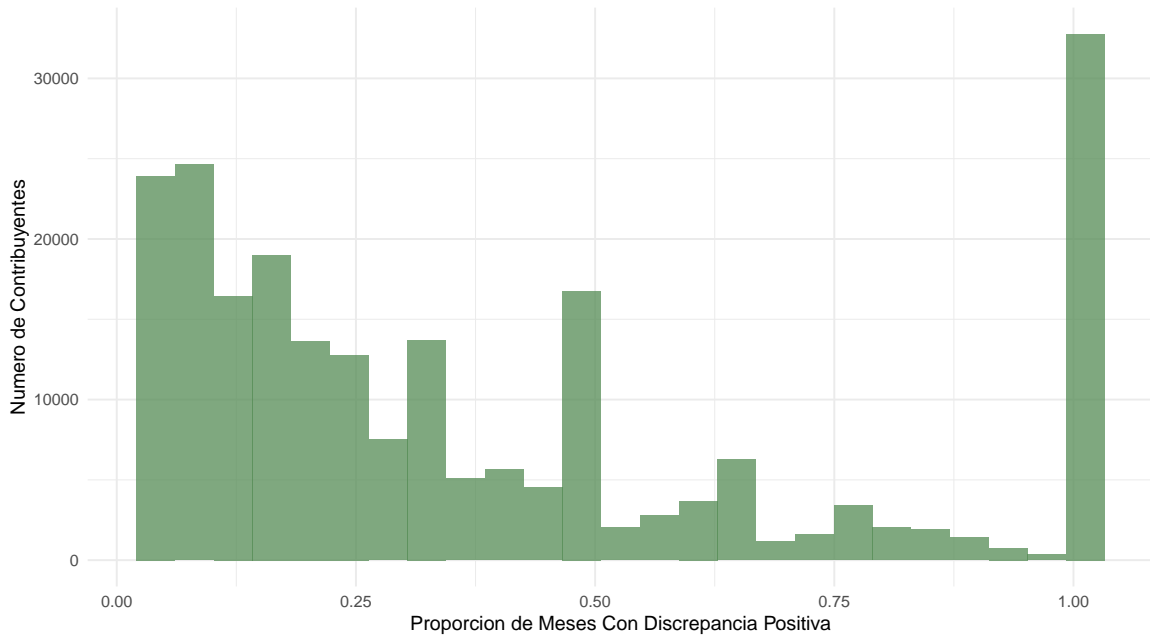
PANEL B: DISTRIBUCIÓN DE DISCREPANCIAS POSITIVAS POR TASA



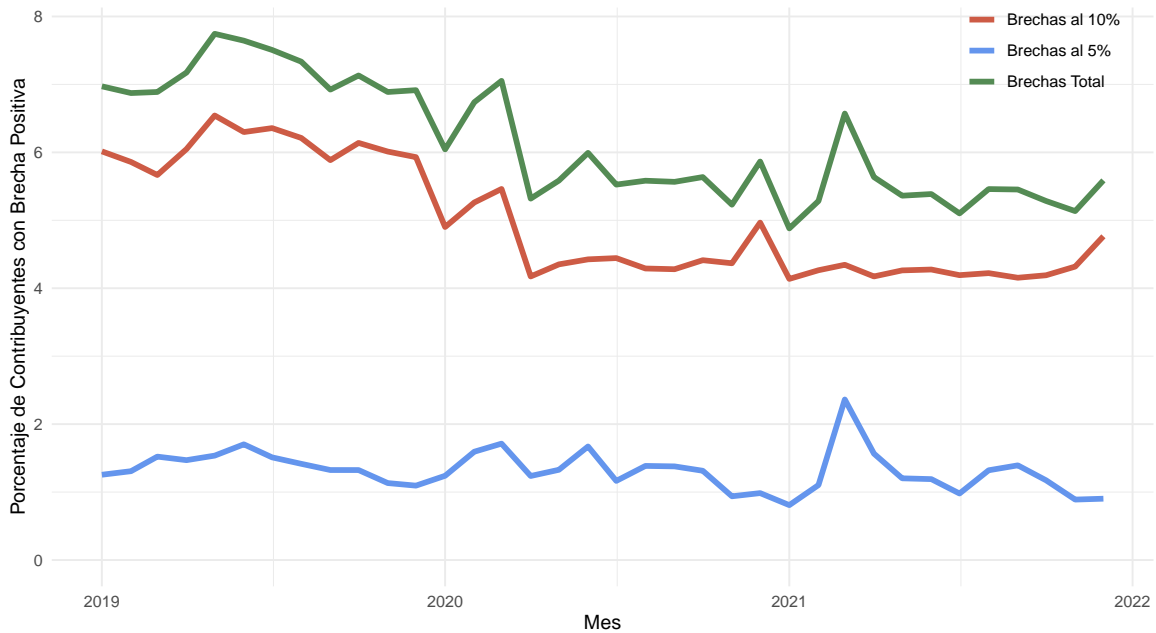
Notas: La figura muestra la distribución de las discrepancias entre las ventas declaradas por el contribuyente i en el mes m en sus declaraciones de IVA, y_{im}^d , y las compras totales del contribuyente reportadas por sus clientes en sus declaraciones informativas, y_{im}^r . Normalizamos estas discrepancias mediante el promedio de los dos reportes para tener en cuenta las grandes diferencias en el tamaño de las empresas: $\tilde{d}_{im} = y_{im}^r - y_{im}^d / \frac{1}{2} (y_{im}^r + y_{im}^d)$. Cuando $\tilde{d}_{im} > 0$, es indicativo de una posible subdeclaración. El Panel A muestra la distribución general de las discrepancias normalizadas, eliminando los casos donde las discrepancias son exactamente cero. Observamos que la mayoría de las discrepancias son negativas, pero existe una masa significativa de discrepancias positivas. En el Panel B, graficamos por separado las distribuciones de discrepancias estrictamente positivas para las ventas gravadas al 10% (la tasa estándar) y las ventas gravadas al 5% (la tasa reducida aplicada a un subconjunto de bienes).

FIGURE 3: FRECUENCIA DE DISCREPANCIAS POSITIVAS 2019–2021

PANEL A: FRECUENCIA DE DISCREPANCIAS POSITIVAS POR CONTRIBUYENTE

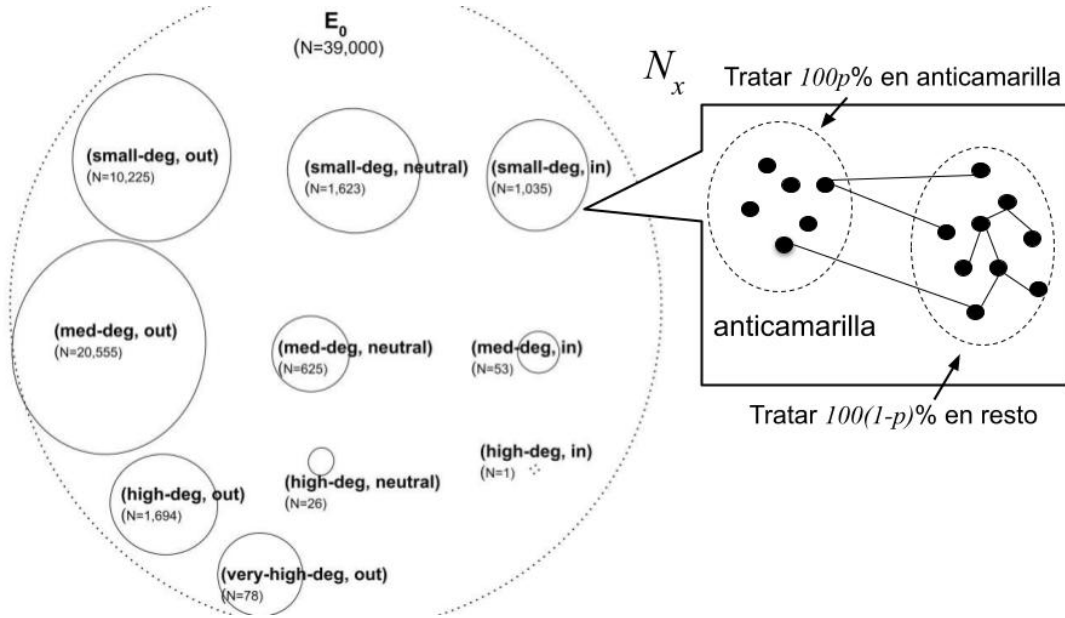


PANEL B: PORCENTAJE DE DECLARACIONES CON DISCREPANCIAS POSITIVAS



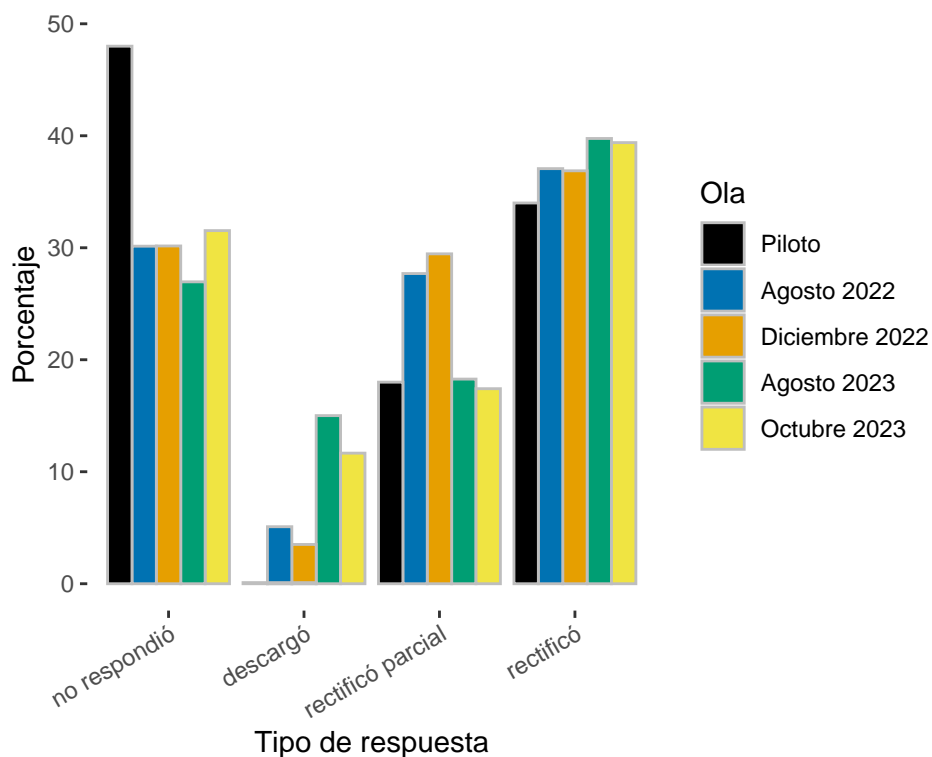
Notas: La figura muestra la incidencia de discrepancias positivas entre los contribuyentes y sobre el tiempo. Para cada contribuyente, calculamos la fracción de meses en los que presentan una declaración jurada que genera discrepancias positivas en comparación con los reportes de ventas de terceros. El Panel A muestra la distribución de esta propensión entre los contribuyentes, excluyendo al 59% de los contribuyentes que nunca tienen discrepancias. El Panel B sigue el porcentaje de declaraciones con discrepancias positivas a lo largo del tiempo. La línea verde muestra el porcentaje de declaraciones con cualquier discrepancia, mientras que la línea roja se refiere únicamente a las discrepancias en ventas gravadas al 10% y la línea azul muestra las discrepancias al 5%.

FIGURE 4: ILUSTRACIÓN DEL DISEÑO SPLIT-GRAPH



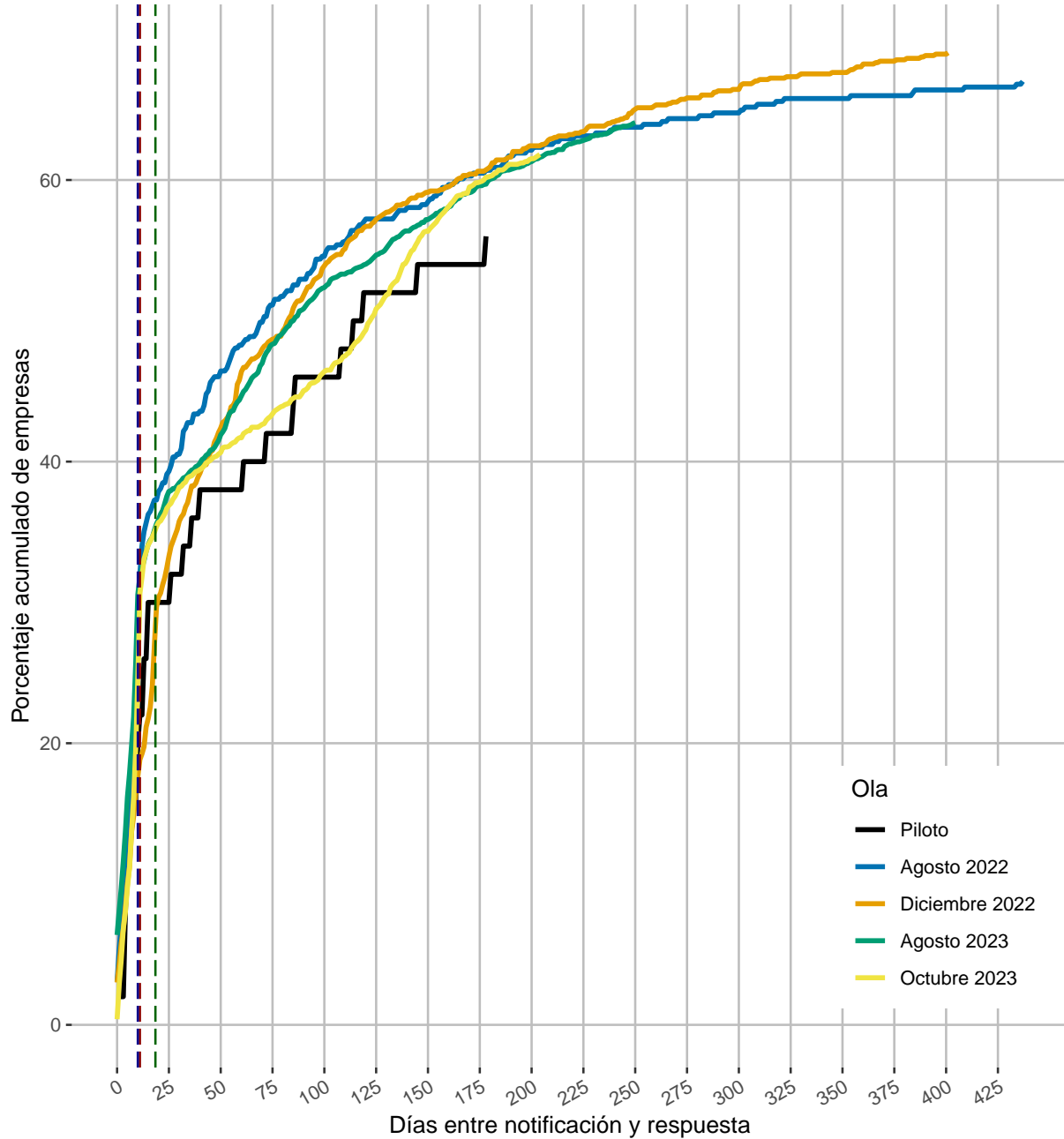
Notas: Izquierda: Contribuyentes se dividen segun su tipo = (degree, orientation) como se explica en la sección 2.6. El programa de tratamiento determina el número total de contribuyentes, N_x , a ser tratado, por tipo x . Derecha: Por cada tipo, calculamos el conjunto independiente (anticamarilla) y el resto. Tratamos pN_x contribuyentes dentro de la anticamarilla aleatoriamente, y $(1 - p)N_x$ en el resto, tambien aleatoriamente. Los contribuyentes de la anticamarilla se usan para analizar los efectos tratamiento directos via pruebas de permutación validas en muestras finitas.

FIGURE 5: TASAS DE RESPUESTA



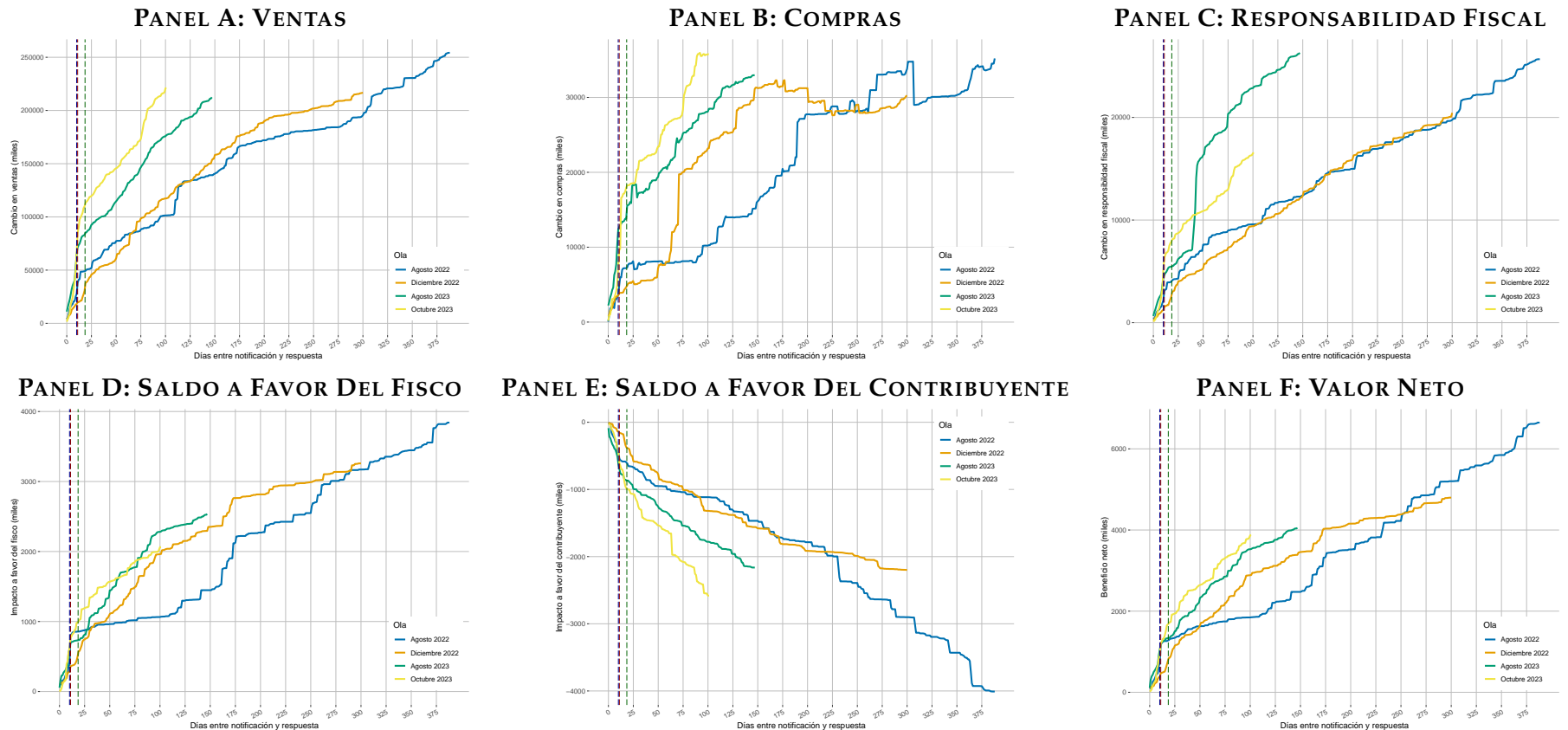
Notas: La figura muestra los resultados de cada caso de los contribuyentes notificados hasta agosto de 2024, dividiendo a los contribuyentes notificados en 4 grupos: aquellos que no respondieron; aquellos que respondieron pero buscaron explicar las discrepancias sin enmendar sus declaraciones; aquellos que realizaron enmiendas que solo abordaron parcialmente las discrepancias; y aquellos que realizaron enmiendas que resolvieron las discrepancias. En general, la tasa de respuesta es alrededor del 70%. La fracción de contribuyentes que no respondieron es mayor en la ola de notificaciones de octubre de 2023, pero como muestra la figura 6, esto se debe a que las respuestas siguen llegando muchos meses después de la notificación, y nuestros datos se detienen 9 meses después de la ola de octubre de 2023.

FIGURE 6: RESPUESTAS SOBRE EL TIEMPO



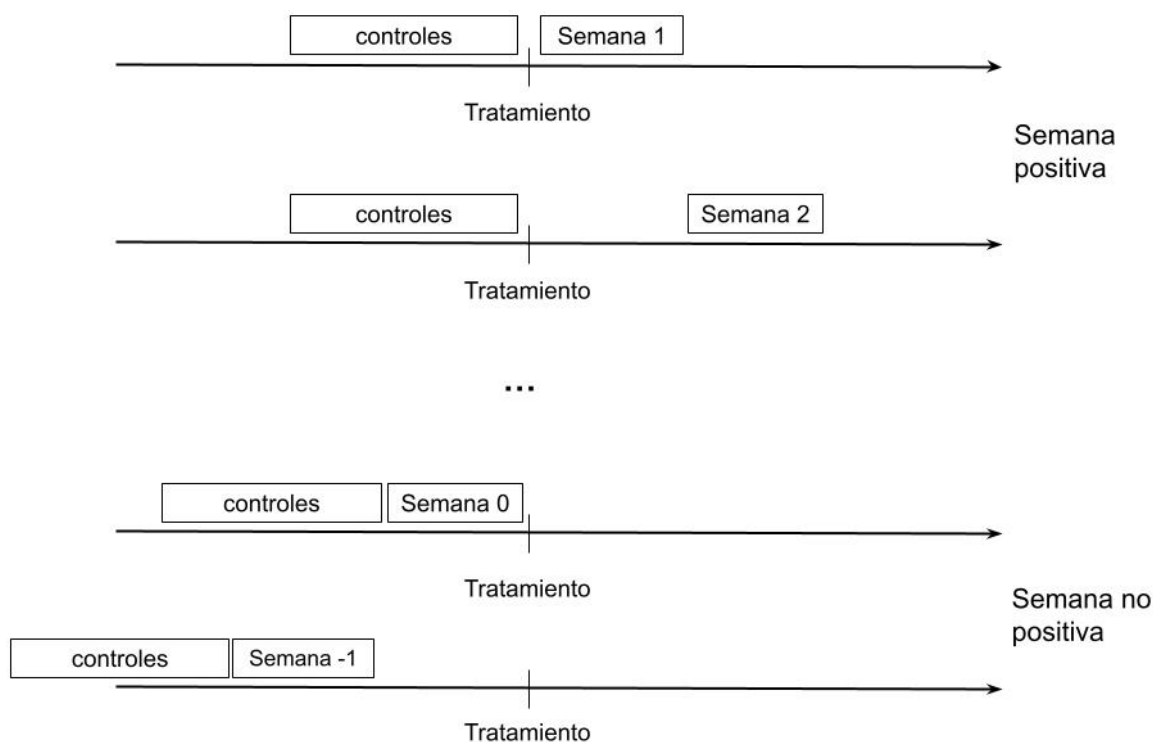
Notas: La figura muestra las tasas de respuesta en cada ola sobre el tiempo. Indican la proporción total de contribuyentes notificados que han respondido conforme transcurren los días hábiles desde la fecha en que fueron notificados. Las líneas verticales señalan el final del plazo para cada ola.

FIGURE 7: RESPUESTAS ACUMULADAS SOBRE EL TIEMPO



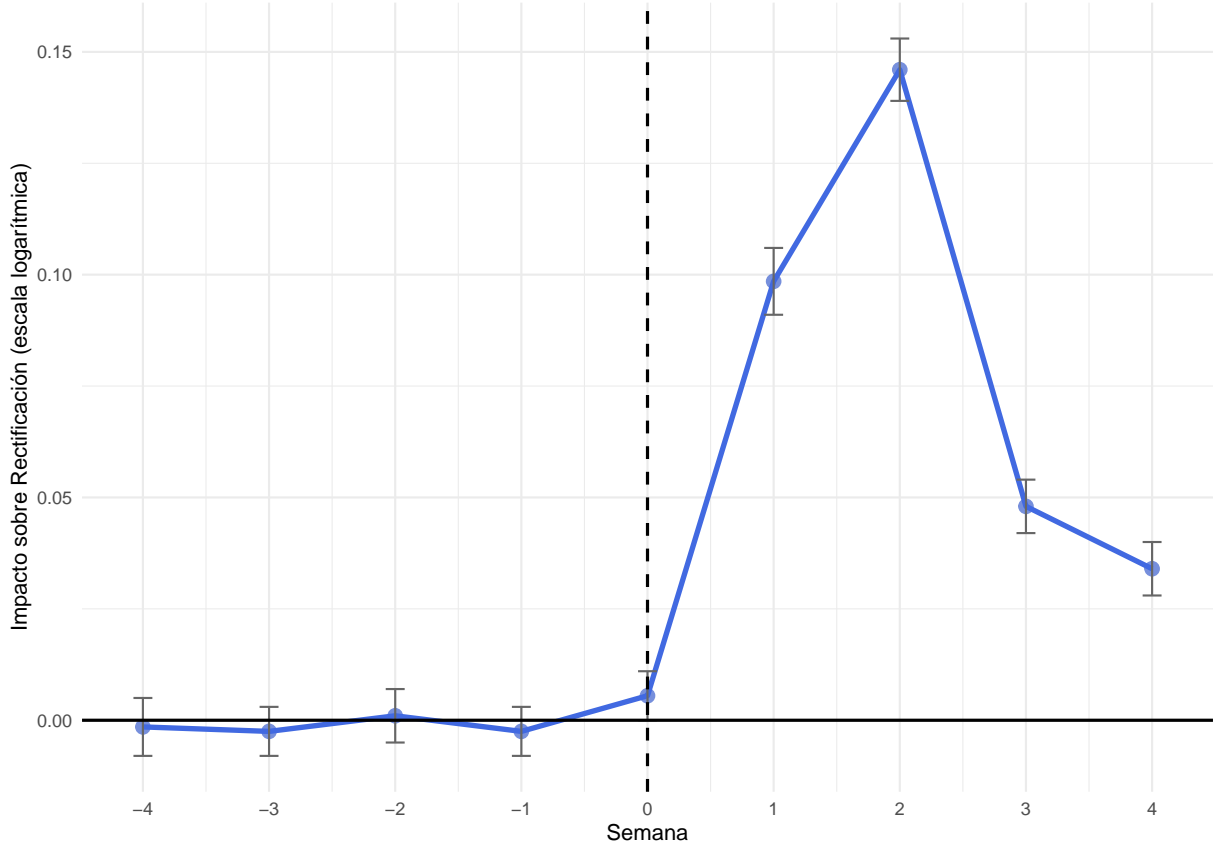
Notas: La figura rastrea el contenido de las rectificaciones realizadas por los contribuyentes notificados. Para cada ola de notificación, seguimos los cambios acumulados en los resultados derivados de sus rectificaciones (por ejemplo, el aumento acumulado en las ventas gravadas declaradas). Luego dividimos esto por el número de contribuyentes notificados para crear medidas de la respuesta promedio por contribuyente notificado, permitiendo así comparar entre olas de diferentes tamaños. Los Paneles A–C reportan aumentos acumulados en ventas, compras y obligaciones tributarias, respectivamente. Los Paneles D–F analizan los cambios en la posición neta de los contribuyentes notificados respecto a la autoridad tributaria. El Panel D documenta grandes aumentos en los pagos de los contribuyentes notificados al tesoro, mientras que el Panel E muestra disminuciones en los créditos que los contribuyentes tienen con la autoridad tributaria. El Panel F combina los aumentos en los pagos y la reducción de los créditos para crear una medida del valor neto de la notificación. Específicamente, el valor neto es el cambio en pagos menos $0.7 \times$ el cambio en créditos.

FIGURE 8: ESPECIFICACIÓN DE DATA APILADA USADA PARA INFERENCIA DE RANDOMIZACIÓN



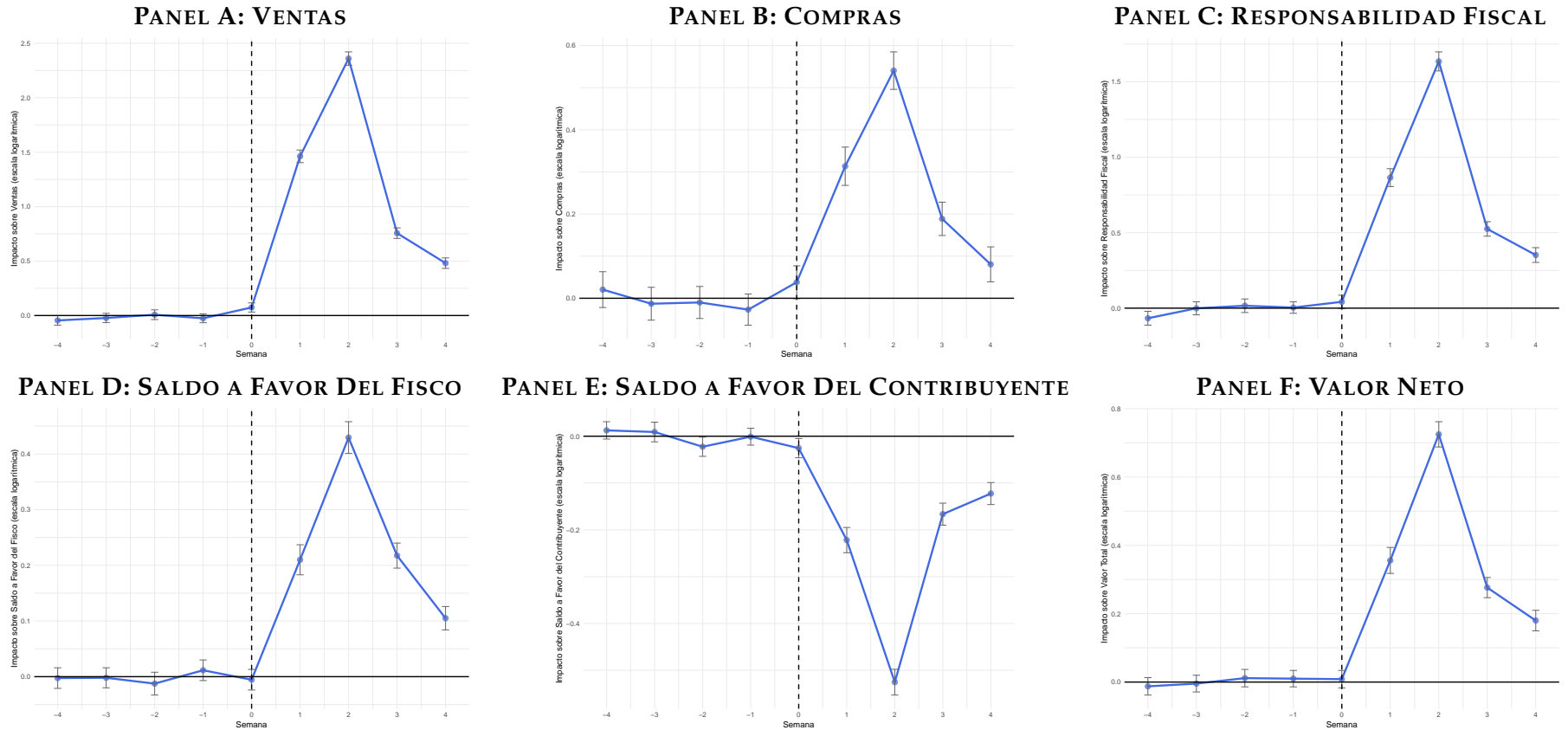
Notas: La figura ilustra nuestra *especificación de datos apilados*. La figura muestra que, para una semana de valor positivo (por ejemplo, 1, 2, . . .), nuestras pruebas de aleatorización comparan los resultados de los contribuyentes elegibles observados en esa semana específica con sus resultados en los 30 días previos al evento de tratamiento (marcados como "controles" en la figura). Para una semana de valor no positivo (0, -1, -2, . . .), la prueba de aleatorización compara los resultados de los contribuyentes elegibles en esa semana con sus resultados en los 30 días previos a dicha semana.

FIGURE 9: CONTROLES AUMENTAN LA TASA DE RECTIFICACIÓN



Notas: La figura muestra los intervalos de confianza (IC) del 95% basados en aleatorización para el efecto directo del tratamiento suponiendo ausencia de interferencia sobre la tasa de rectificación en cada semana $w \in -4, \dots, 0, \dots, 4$. En cada semana probamos la hipótesis $H_0^{te}(\tau)$, definida en la Ecuación (3): $H_0^{te}(\tau) : f_i(1, w) = f_i(0, w) + \tau$, para todo $i \in [N], w \in \mathbb{Z}$. Luego invertimos la prueba conservando aquellos valores de τ que corresponden a un valor p de aleatorización no mayor a 0.05, generando el intervalo de confianza basado en aleatorización.

FIGURE 10: EFECTOS TRATAMIENTO SOBRE CONTENIDO DE RECTIFICACIONES



Notas: La figura muestra los intervalos de confianza (IC) del 95% basados en aleatorización para el efecto directo del tratamiento suponiendo ausencia de interferencia sobre el contenido de las rectificaciones en cada semana $w \in -4, \dots, 0, \dots, 4$. En cada semana probamos la hipótesis $H_0^{te}(\tau)$, definida en la Ecuación (3): $H_0^{te}(\tau) : f_i(1, w) = f_i(0, w) + \tau$, para todo $i \in [N], w \in \mathbb{Z}$. Luego invertimos la prueba conservando aquellos valores de τ que corresponden a un valor p de aleatorización no mayor a 0.05, generando el intervalo de confianza basado en aleatorización. El Panel A muestra efectos tratamiento sobre las ventas declaradas, el Panel B sobre compras, y el panel C sobre la responsabilidad fiscal total. Los Paneles D–F analizan los cambios en la posición neta de los contribuyentes notificados respecto a la autoridad tributaria. El Panel D documenta grandes aumentos en los pagos de los contribuyentes notificados al tesoro, mientras que el Panel E muestra disminuciones en los créditos que los contribuyentes tienen con la autoridad tributaria. El Panel F combina los aumentos en los pagos y la reducción de los créditos para crear una medida del valor neto de la notificación. Específicamente, el valor neto es el cambio en pagos menos $0.7 \times$ el cambio en créditos.

TABLE 1: BALANCE DE LA ALEATORIZACION

Variable	Diseño Agrupado		Diseño Split-Graph	
	Ola 1	Ola 2	Ola 3	Ola 4
1 Tipo	0.43	0.22	0.97	0.75
2 Tamaño	0.05	0.93	0.49	0.18
3 Región	0.94	0.54	0.13	0.10
4 Actividad Economica	0.57	0.99	0.27	0.12
5 Grupo Edad	0.35	0.60	0.80	0.58

Notas: La tabla muestra los resultados de las pruebas de balance en la aleatorización. Dado que nuestros protocolos de aleatorización variaron entre olas, presentamos los resultados de balance por separado para las olas 1–2 y las olas 3–4. Las olas 1–2 siguieron el diseño saturado de aleatorización por conglomerados descrito en la Sección ?? . En los conglomerados asignados a una saturación del 0% o 100%, no podemos comparar contribuyentes tratados y de control. En su lugar, resumimos las covariables de los contribuyentes en el conglomerado como *covariables a nivel de conglomerado* y probamos el balance a nivel de conglomerado. Específicamente, realizamos los siguientes pasos para diagnosticar el balance: 1: Para la covariable Z con niveles $k = 1, \dots, K$, calculamos la covariable del conglomerado para cada conglomerado J : $C_J = \text{entropy}(\{c_k\}_{k=1}^K)$, $c_k = \sum_{i \in J} 1\{Z_i = k\}/|J|$. Aquí, $\{c_k\}_{k=1}^K$ es la distribución empírica de X dentro del conglomerado J , mientras que C_J es la entropía de esta distribución, que sirve como covariable a nivel de conglomerado; 2: Regresamos C_J sobre p_J , donde p_J es la tasa de saturación para el conglomerado J . Reportamos el valor p de la prueba t de la regresión. Los resultados de este cálculo se muestran en las columnas 1 y 2.

En las olas 3–4, la aleatorización cambió a un experimento de red estratificado, como se describe en la Sección 2.5. La aleatorización consistió en múltiples experimentos simples aleatorizados dentro de conjuntos de empresas del mismo tipo x y con el mismo estatus de anticlique. Como resultado, podemos probar el balance a nivel de contribuyente de la siguiente manera: 1: Dentro de cada ola, calcular el tipo x_i (grado, orientación) de cada contribuyente i ; el tratamiento $D_i \in \{0, 1\}$; y el estatus de anticlique $A_i \in \{0, 1\}$. 2: Para cada covariable Z , ajustar dos modelos de regresión multinomial: $M_0 : Z_i \sim x_i + A_i$, y $M_1 : Z_i \sim D_i + x_i + A_i$, y reportar el valor p de la comparación ANOVA entre M_0 y M_1 . Si el tratamiento predice la covariable, el valor p debería ser pequeño. Los resultados de este cálculo se muestran en las columnas 3 y 4.

TABLE 2: RESPUESTAS POR TIPO DE CONTRIBUYENTE

	1[Rectificación]	Δ Ventas	Δ Compras	Δ Impuesto	Δ Saldo Fisco	Δ Saldo Contribuyente	Beneficio neto
Contribuyente Física X Pequeña	0.7554*** (0.0066)	177322411*** (6397185)	29696102*** (3275113)	12749234*** (536528)	1971559*** (158306)	-1782806*** (110613)	3219523*** (183296)
Contribuyente Jurídica X Pequeña	0.6014*** (0.0202)	553409959*** (49299112)	81129622*** (21659882)	62550316*** (5193608)	7553385*** (984461)	-7226746*** (838371)	12612107*** (1177654)
Contribuyente Física X Mediana	0.2778*** (0.0757)	249001324*** (36097221)	26623084** (9132347)	22250785*** (2965695)	9452988*** (1493819)	-6710725*** (1408340)	14150496*** (2453741)
Contribuyente Jurídica X Mediana	0.5354*** (0.0504)	529787981*** (79043820)	23758595* (9582732)	208518286*** (52699566)	14450181*** (3258257)	-15841085*** (4785661)	25538941*** (5056076)
Contribuyente Jurídica X Grande	0 (0)	97898871*** (27152261)	-584180055* (289308763)	300551875*** (71452081)	0 (0)	-2141232** (685743)	1498863** (480020)
Num.Obs.	4975	4975	4975	4975	4975	4975	4975

Notas: La tabla muestra estimaciones de la heterogeneidad de las respuestas según los tipos de contribuyentes notificados, centrándose en su clasificación por tamaño (pequeño/mediano/grande) y su forma jurídica (jurídica/física). En la muestra de rectificaciones realizadas por los contribuyentes notificados, realizamos una regresión del resultado del contribuyente i sobre indicadores para los diferentes grupos $g(i) \in \text{Constituido, No constituido} \times \text{Pequeño, Mediano, Grande}$ a los que pertenece el contribuyente i : $y_i = \beta_h \mathbf{1}[g(i) = h] + \varepsilon_i$, y reportamos los coeficientes β_h .

TABLE 3: VENTAS, COMPRAS, Y SALDO CONTRIBUYENTE REPORTADO POR TIPO DE CONTRIBUYENTE

(A) VALORES MEDIANAS DECLARADOS

Tipo	N	Ventas Reportadas	Compras Reportadas	Saldo Contribuyente Reportado
FÍSICA X PEQUENO	4239	30934198	27536364	0
JURÍDICA X PEQUENO	587	181956567	132041760	0
FÍSICA X MEDIANO	36	1088716775	571590510	12671677
JURÍDICA X MEDIANO	99	613819719	211464037	4510789
JURÍDICA X GRANDE	14	3658694296	2405293225	-45408485
Total	4975	4975	4975	4975

(B) VALORES PROMEDIOS DECLARADOS

Tipo	N	Ventas Reportadas	Compras Reportadas	Saldo Contribuyente Reportado
FÍSICA X PEQUENO	4239	254892186	189985567	148540
JURÍDICA X PEQUENO	587	1329616126	1048966522	-13537108
FÍSICA X MEDIANO	36	1435296620	742567195	62006771
JURÍDICA X MEDIANO	99	2412371339	1836405125	-3675766
JURÍDICA X GRANDE	14	9361276024	5777092862	-12856772
Total	4975	4975	4975	4975

Notas: La tabla muestra estimaciones de la heterogeneidad de las respuestas según los tipos de contribuyentes notificados, centrándose en su clasificación por tamaño (pequeño/mediano/grande) y su naturaleza jurídica (jurídica/física).

TABLE 4: EFECTOS CAUSALES DE NOTIFICACIONES

	Rectificación	Ventas	Compras	Impuesto	Saldo Fisco	Saldo Contribuyente	Valor
Notificado	0.6103 (0.0181) [0.010]	0.5462 (0.0152) [0.006]	0.1192 (0.0101) [0.003]	0.5161 (0.0149) [0.002]	0.164 26 (0.008 22) [0.004]	0.1698 (0.0089) [0.005]	0.2452 (0.0106) [0.006]
Promedio control	0.141	0.0794	0.0718	0.0783	0.014	0.0184	0.0268
Num.Obs.	14 918 211	14 918 211	14 918 211	14 918 211	14 918 211	14 918 211	14 918 211

Notas: Esta tabla presenta los efectos causales de las notificaciones sobre la probabilidad de rectificar las declaraciones de impuestos. Cada observación corresponde a una empresa en un día específico. Cada resultado toma el valor 100 para una empresa en un día dado si la empresa presentó un determinado tipo de rectificación ese día, y 0 en caso contrario. La columna 1 muestra los impactos sobre la probabilidad de presentar cualquier rectificación; Col. 2: presentar una rectificación que aumente las ventas declaradas por la empresa; presentar una rectificación que aumente las compras declaradas; Col. 4: presentar una rectificación que aumente la obligación tributaria de la empresa; Col. 5: presentar una rectificación que aumente los pagos del contribuyente a la autoridad tributaria; Col. 6: presentar una rectificación que reduzca los créditos fiscales pendientes del contribuyente con la autoridad tributaria; Col. 7: presentar una rectificación que aumente el valor neto de la declaración del contribuyente ante la autoridad tributaria, donde el valor neto es el cambio en el pago de impuestos menos $0.7 \times$ el cambio en el crédito fiscal. Los resultados provienen de la estimación de la ecuación (4): $y_{kit} = \beta D_{kit} + \omega p_{kit} + \mu_{ki} + \delta_{kt} + \epsilon_{kit}$ donde y_{kit} es el resultado para el contribuyente i en la fecha t , dentro de la ola de notificación k ; D_{kit} es un indicador que toma el valor 1 si el contribuyente i fue notificado en la ola k y si la fecha t es posterior a la fecha de notificación para esa ola; p_{kit} toma el valor 0 antes de la fecha de notificación, y en caso contrario es igual a la probabilidad de que el contribuyente i haya sido seleccionado aleatoriamente para recibir una notificación en la ola k ; μ_{ki} son efectos fijos por ola de notificación y contribuyente; y δ_{kt} son efectos fijos por ola de notificación y tiempo. Los errores estándar agrupados a nivel de contribuyente se reportan entre paréntesis, y los valores p de inferencia por aleatorización se presentan entre corchetes.


TABLE 5: EFECTOS CAUSALES DE NOTIFICACIONES: EFECTOS POR TIPO DE CONTRIBUYENTE

	Rectificación	Ventas	Compras	Impuesto	Saldo Fisco	Saldo Contribuyente	Valor
Notificado:Grande,Jurídico	-2.41 (2.36) [0.186]	-0.663 (0.581) [0.576]	-0.940 (0.585) [0.733]	-1.52 (1.83) [0.011]	0.0659 (0.0547) [0.043]	-1.92 (1.72) [0.473]	-1.87 (1.71) [0.078]
Notificado:Mediano,Físico	0.521 (0.203) [0.299]	0.615 (0.147) [0.312]	-0.124 (0.139) [0.122]	0.653 (0.159) [0.060]	0.1346 (0.0674) [0.887]	0.0419 (0.0349) [0.769]	0.142 (0.067) [0.582]
Notificado:Mediano,Jurídico	0.341 (0.149) [0.455]	0.307 (0.101) [0.639]	-0.0406 (0.0824) [0.967]	0.325 (0.127) [0.494]	0.0428 (0.0366) [0.509]	0.0191 (0.0557) [0.279]	0.0559 (0.0620) [0.685]
Notificado:Pequeño,Físico	0.6432 (0.0193) [0.784]	0.5730 (0.0167) [0.875]	0.1369 (0.0109) [0.131]	0.5373 (0.0161) [0.534]	0.172 31 (0.009 06) [0.644]	0.180 08 (0.009 52) [0.708]	0.2606 (0.0114) [0.586]
Notificado:Pequeño,Jurídico	0.496 (0.056) [0.666]	0.4422 (0.0433) [0.910]	0.0562 (0.0306) [0.380]	0.4492 (0.0428) [0.184]	0.1449 (0.0241) [0.554]	0.1656 (0.0263) [0.444]	0.2158 (0.0313) [0.593]
Promedio control	0.142	0.0804	0.0727	0.0792	0.0142	0.0187	0.0271
Num.Obs.	14 714 447	14 714 447	14 714 447	14 714 447	14 714 447	14 714 447	14 714 447


Notas: Esta tabla presenta las estimaciones de la heterogeneidad de los efectos de las notificaciones sobre la probabilidad de enmendar las declaraciones de impuestos según el tipo de contribuyente. Cada observación corresponde a una empresa en un día específico. Cada resultado toma el valor 100 para una empresa en un día dado si la empresa presentó un determinado tipo de enmienda ese día, y 0 en caso contrario. La columna muestra impactos sobre la probabilidad de presentar cualquier rectificación; Col. 2: presentar una rectificación que aumente las ventas declaradas por la empresa; Col. 3: presentar una rectificación que aumente las compras declaradas; Col. 4: presentar una rectificación que aumente la obligación tributaria de la empresa; Col. 5: presentar una rectificación que aumente los pagos del contribuyente a la autoridad tributaria; Col. 6: presentar una rectificación que reduzca los créditos fiscales pendientes del contribuyente con la autoridad tributaria; Col. 7: presentar una rectificación que aumente el valor neto de la declaración del contribuyente ante la autoridad tributaria, donde el valor neto es el cambio en el pago de impuestos menos $0.7 \times$ el cambio en el crédito fiscal. Los resultados provienen de la estimación de la ecuación (4): $y_{ijkt} = \sum_j \beta_j D_{kit} \times S_{ij} + Post_t \times S_{ij} + \omega p_{kit} + \mu_{ki} + \delta_{kt} + \varepsilon_{ijkt}$ donde los términos están definidos como en las notas de la tabla 4; S_{ij} indica a los contribuyentes del tipo j , donde j indexa los subgrupos; y $Post_t$ es un indicador para los períodos posteriores a la notificación. Los errores estándar agrupados a nivel de empresa se reportan entre paréntesis, y los valores p de inferencia por aleatorización se presentan entre corchetes.

Figures & Tables


FIGURE I: EXAMPLE NOTIFICATION




TETĀ VIRU MOHENDAPY
Ministerio de HACIENDA



SET TRIBUTACIÓN
Promoviendo confianza



TETĀ REKUÁI GOBIERNO NACIONAL



Paraguay de la gente

Asunción, «fechanota».

Mass Audit Notice DGFT N° «nota»

«nombre»
RUC:
«ruc»
email: «correo1»

To whom it may concern:

We are pleased to address you to inform you that the General Directorate of Tax Auditing, by virtue of its powers granted by Article 189 of Law No. 125/91, as amended by Law No. 2421/04, is conducting mass control processes on taxpayers who have inconsistencies arising from various cross-checks.

In this regard, differences have been detected between the sales or fiscal debit adjustments declared by you in your VAT returns, where you have reported amounts lower than those provided by your clients in their Hechauka purchase books.

According to the results of cross-checks and tax analyses derived from Big Data, this inconsistency represents an irregularity in the amount of sales or fiscal debit adjustments you have reported. It is important to mention that the Big Data analyses are carried out on all the information available to the Tax Administration and the knowledge of risk profiles based on irregularities detected during specific controls.

Therefore, you are requested to review and rectify your sales declarations and/or fiscal debit adjustments **exclusively for the indicated items**, within a period not exceeding 10 (ten) business days from the notification of this letter, in accordance with the details provided in the following table:

FISCAL PERIOD		SALES IN VAT DECLARATION		SALES IN HECHAUKA PURCHASE BOOKS OF INFORMING CLIENTS (*)		SALES OR ADJUSTMENTS NOT DECLARED		
YEAR	MONTH	TAXED AT 10%	TAXED AT 5%	TAXED AT 10%	TAXED AT 5%	TAXED AT 10%	TAXED AT 5%	TOTAL
«año»	«mes»	«grav10va»	«grav5iva»	«grav10hechauka»	«grav5hechauka»	«difgrav10»	«difgrav5»	«diftotal»

(*) THE DETAILS OF THE INVOICE DATA THAT SUPPORTS THE PURCHASES DECLARED BY YOUR CLIENTS IS PRESENTED IN THE ANNEX ATTACHED TO THIS NOTIFICATION

In case you disagree with any of these statements, please send your observations **exclusively** from the email address registered with your RUC to the address «correo2».

Sincerely,

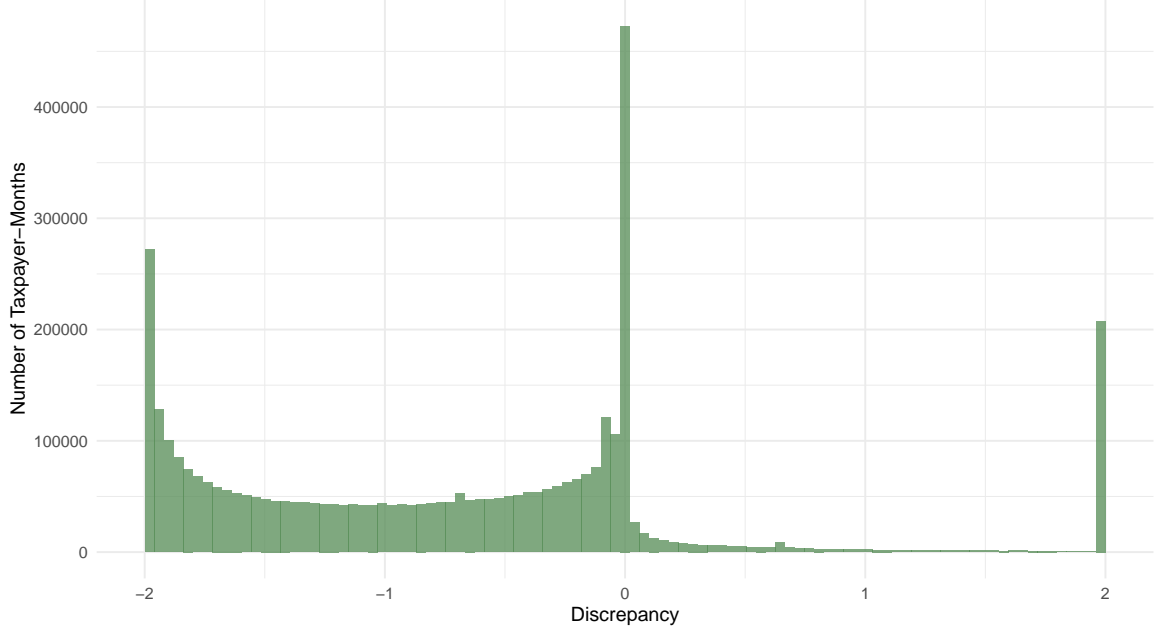
OPERATIONAL PLANNING DEPARTMENT
DIRECTORATE GENERAL OF TAX AUDITING

IMPORTANT CLARIFICATIONS:

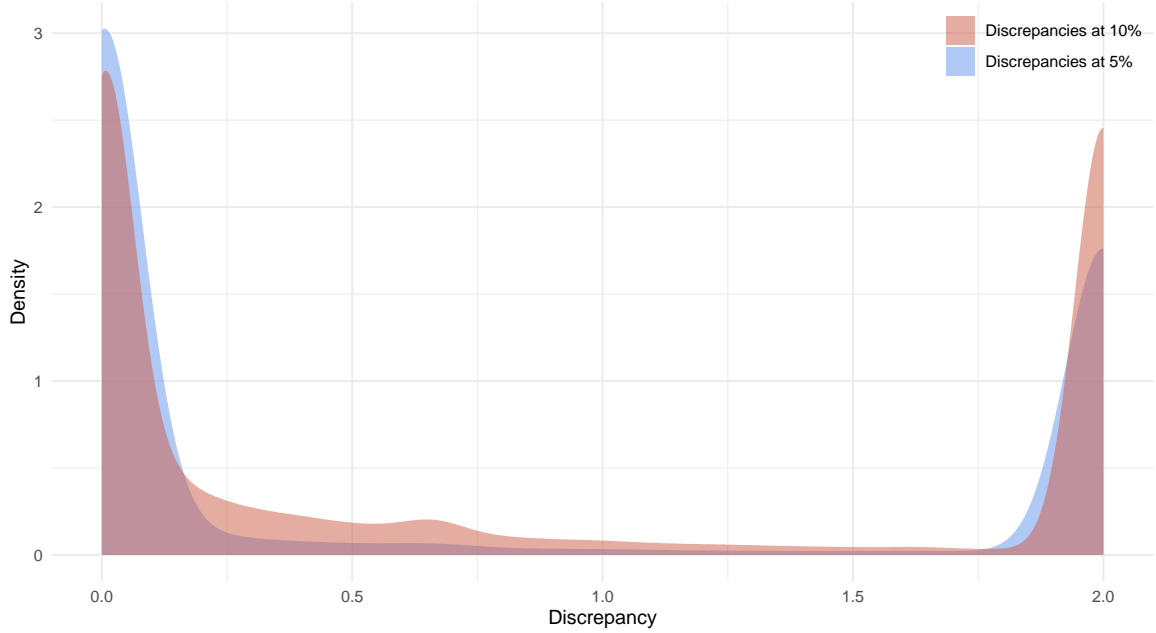
- In case of any doubts or inquiries, please contact the phone numbers «telef1» or «telef2», or visit the offices of the General Directorate of Tax Auditing, Department «Dptoop», located at Eduardo Víctor Haedo 128, corner of Independencia Nacional, Asunción.
- The date of issuance of this letter is «fecharemisión», and according to RG 102/13, Article 10, the effective notification is considered to be on «fechanotif».
- The deadline for your response is «fechavenc».
- Mass audits do not allow for extension requests.
- The requested amendments do not require the Opinion of an External Tax Auditor. However, if the fiscal year to be amended is affected by RG 67/2015, you must request authorization to amend it via email to «correo2». Any return message received does not imply the rejection of the submitted email. Therefore, be assured that your request for authorization to amend was effectively received.
- The approval for the amendment authorization request will arrive in your Marandu Inbox in the Marangatu Tax Management System within 24 hours of submitting the request.
- If the Tax Returns are in the status of "NOT APPLY," the process to request application must be carried out by the taxpayer in the relevant areas according to their jurisdiction.
- In the submitted explanation, you may attach any documents you deem pertinent to support your arguments.
- It is recommended to seek technical support from your accounting professional or external tax auditor when analyzing transactions potentially suspected of irregularity.
- Big Data is a technology that enables the storage, processing, and mass analysis of data. The SET uses this tool, among other purposes, to detect potential cases of taxpayers issuing or using invoices with false content.

FIGURE II: DISTRIBUTION OF DISCREPANCIES: 2019–2021

PANEL A: DISTRIBUTION OF DISCREPANCIES



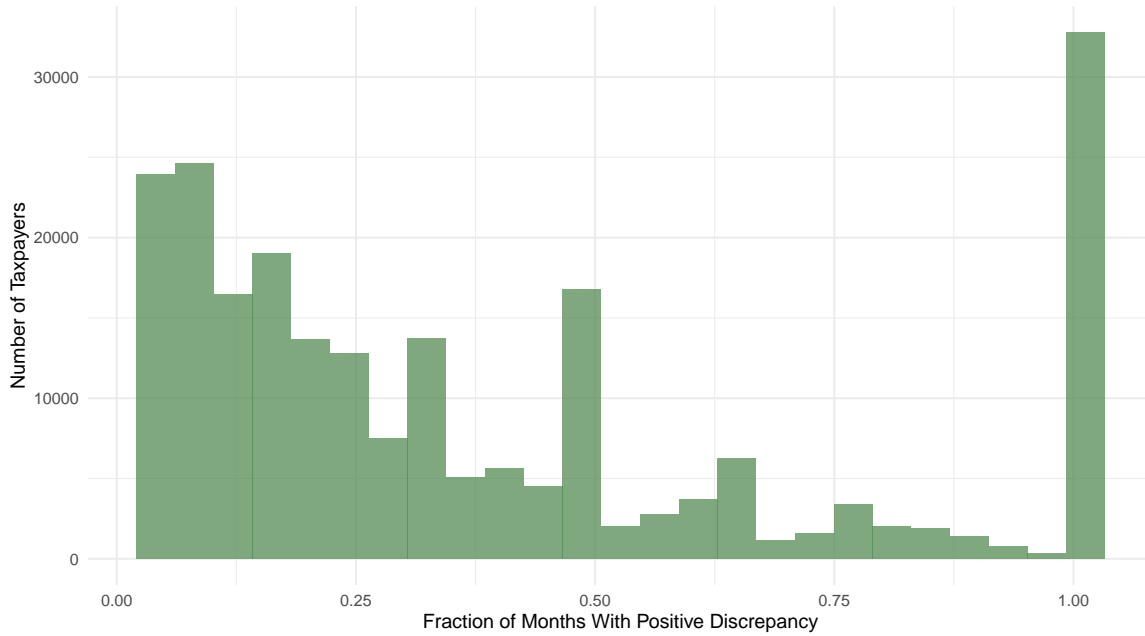
PANEL B: DISTRIBUTION OF DISCREPANCIES BY TAX RATE



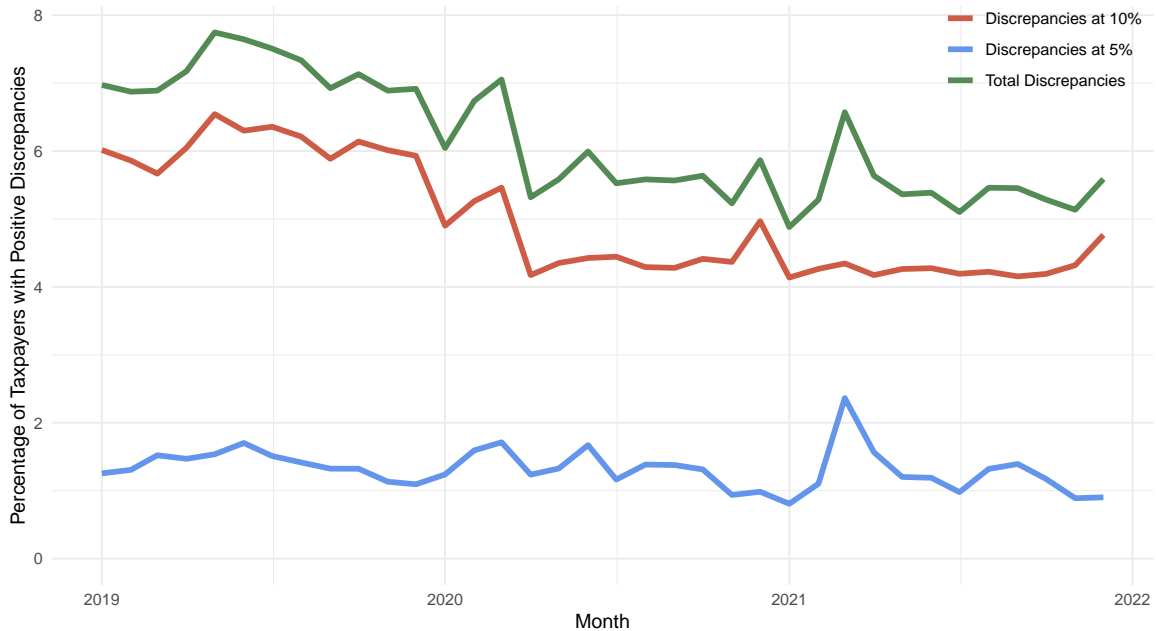
Notes: The figure shows the distribution of discrepancies between the sales declared by taxpayer i in month m in their VAT returns, y_{im}^d , and the total purchases from the taxpayer reported by their clients in their informational declarations y_{im}^r . We normalize these by the average of the two reports to account for the large differences in firm size: $\tilde{d}_{im} = y_{im}^r - y_{im}^d / \frac{1}{2} (y_{im}^r + y_{im}^d)$. Whenever $\tilde{d}_{im} > 0$, it is indicative of potential misreporting. Panel A shows the overall distribution of normalized discrepancies, removing cases where discrepancies are exactly zero. We see that the bulk of discrepancies are negative, but there is a meaningful mass of positive discrepancies. In Panel B we plot the distributions of strictly positive discrepancies separately for sales taxable at 10% (the standard rate) and sales taxable at 5% (the reduced rate applied to a subset of goods).

FIGURE III: FREQUENCY OF POSITIVE DISCREPANCIES 2019–2021

PANEL A: INCIDENCE OF POSITIVE DISCREPANCIES ACROSS TAXPAYERS

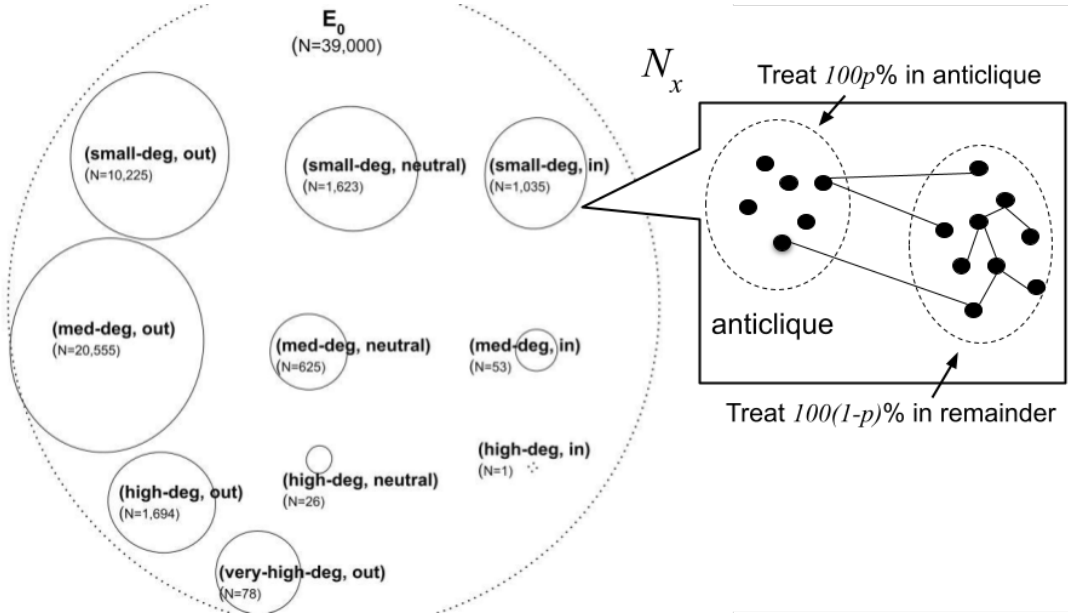


PANEL B: PERCENTAGE OF DECLARATIONS WITH POSITIVE DISCREPANCIES



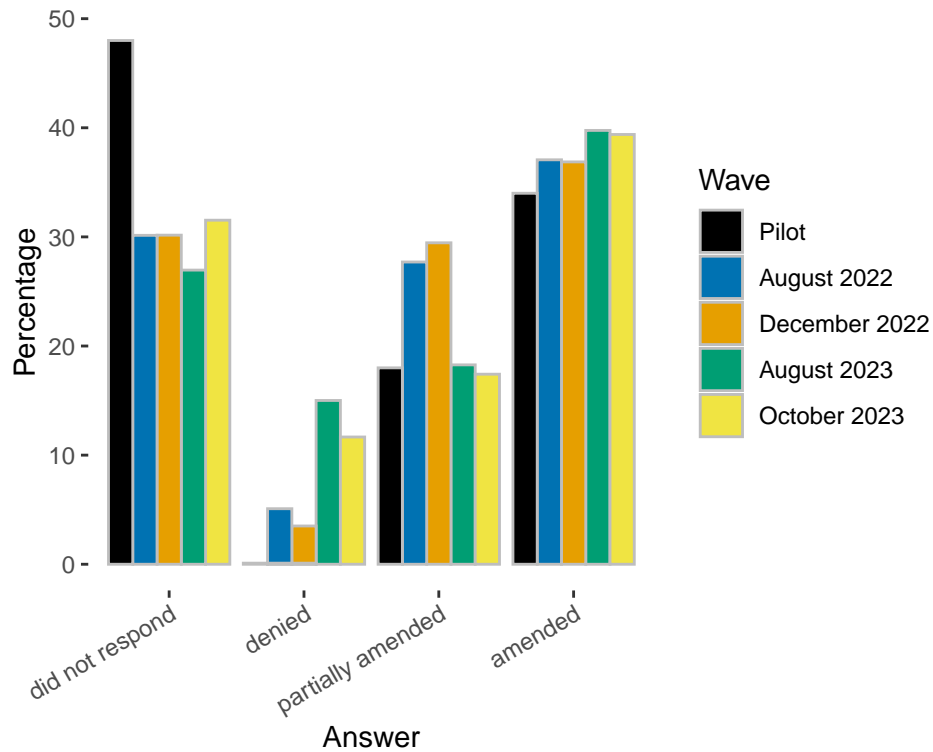
Notes: The figure shows the incidence of positive discrepancies across taxpayers and over time. For each taxpayer, we compute the fraction of months in which they file a tax return but it generates positive discrepancies when compared to third-party reports of their sales. Panel A shows the distribution of this propensity across taxpayers, removing the 59% of taxpayers who never have discrepancies. Panel B tracks the percentage of returns that have positive discrepancies over time. The green line shows the percentage of returns with any discrepancy, while the red line breaks out only discrepancies in sales taxable at 10% and the blue line shows discrepancies at 5%.

FIGURE IV: ILLUSTRATION OF SPLIT-GRAPH DESIGN.



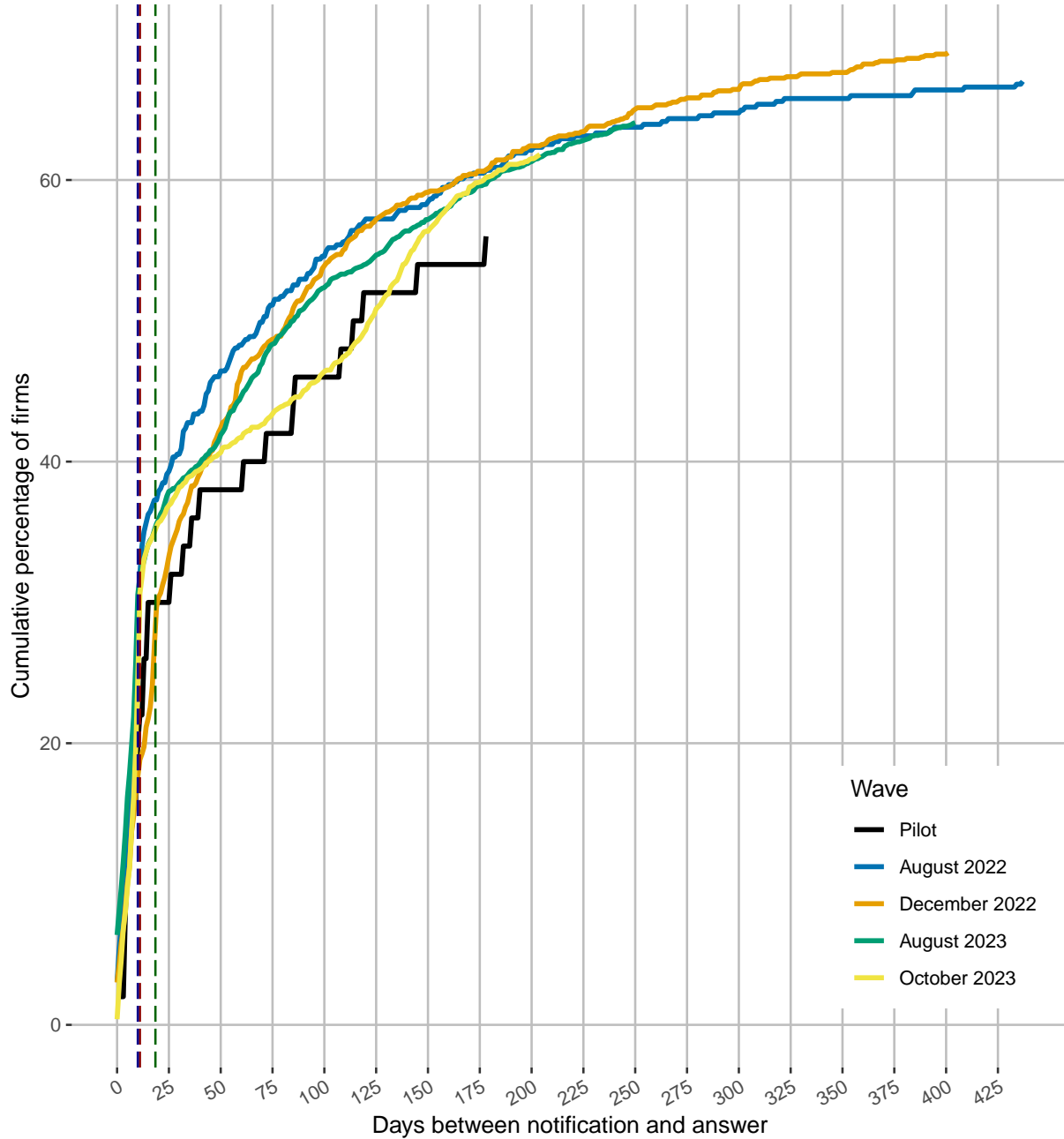
Notes: *Left*: taxpayers are split according to type = (degree, orientation) as explained in section II.vi. The treatment schedule determines the total number, N_x , of taxpayers to be treated per type x . *Right*: Within each type, we calculate an independent set (anticlique) and the remainder set. We treat pN_x of taxpayers in the anticlique at random, and $(1 - p)N_x$ in the remainder set, also at random. Anticlique taxpayers are used to test for direct treatment effects via finite-sample valid permutation tests.

FIGURE V: RESPONSE RATES



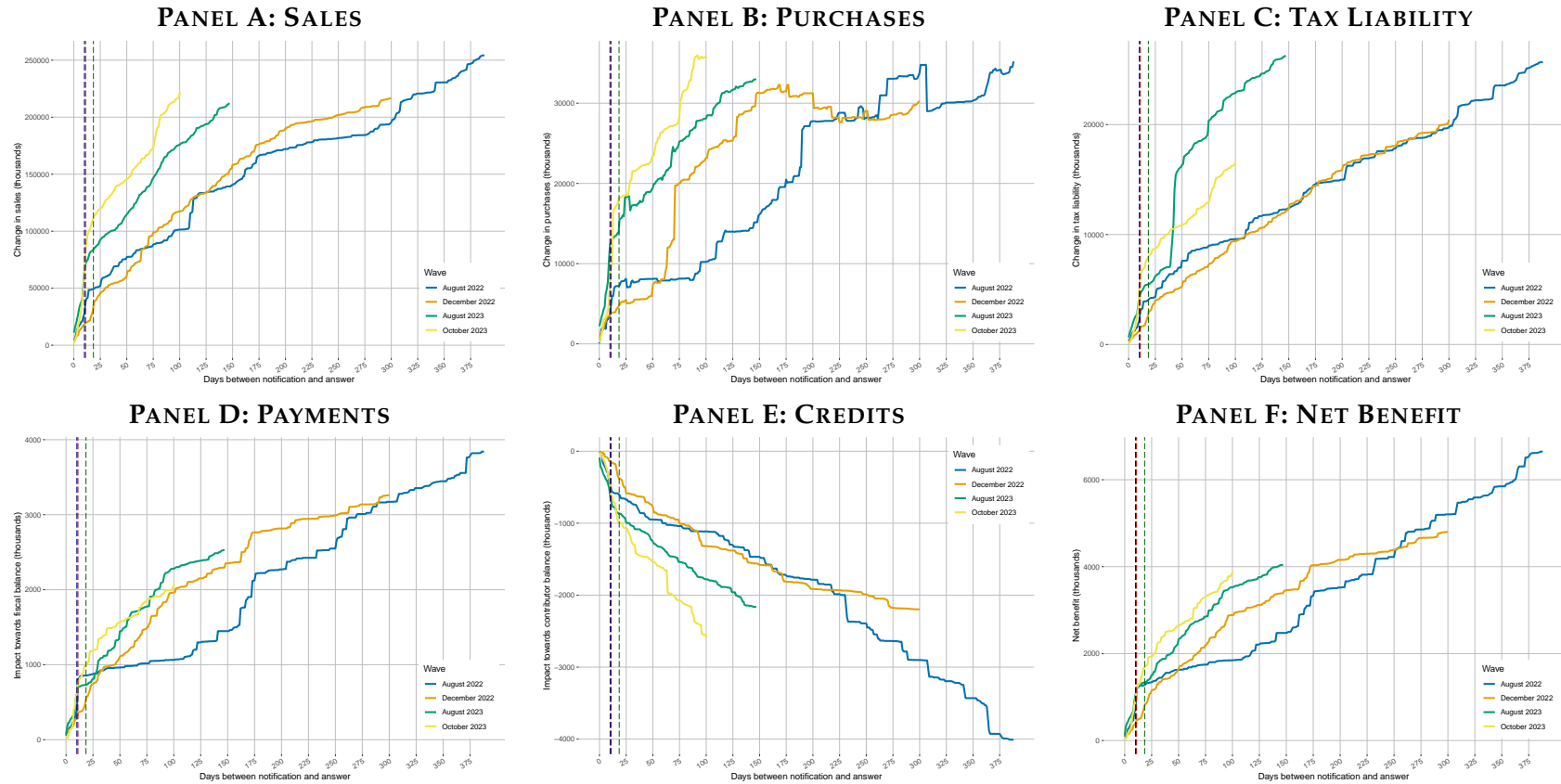
Notes: The figure shows the outcomes of each notified taxpayers' case as of August 2024, dividing the notified taxpayers into 4 groups: Those who did not respond; those who responded but sought to explain the discrepancies without amending their returns; those who made amendments that only partially addressed the discrepancies; and those who made amendments that addressed the discrepancies. Overall, the response rate is around 70%. The fraction of taxpayers who did not respond is higher in the October 2023 wave, but as figure VI shows, this is because responses continue to arrive many months after the notification, and our data stops 9 months after the October 2023 wave.

FIGURE VI: RESPONSE RATES OVER TIME



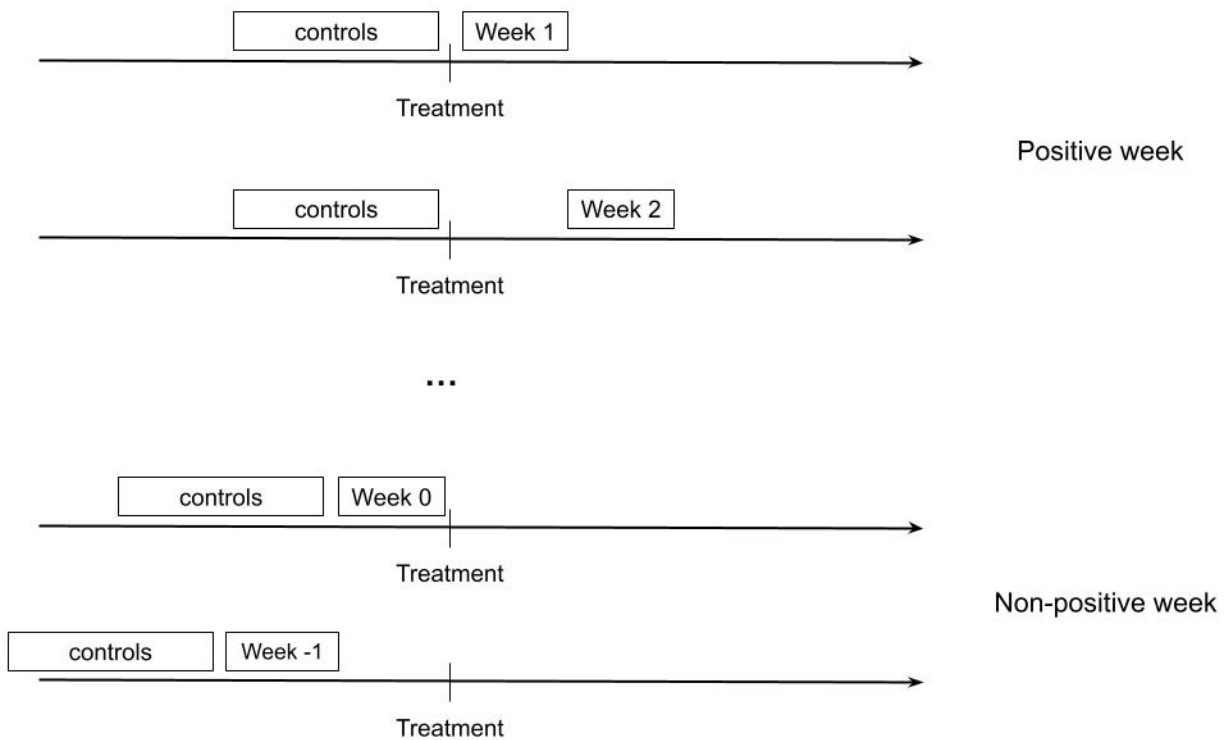
Notes: The figure shows the response rates in each wave over time. They show the total proportion of notified taxpayers who have responded as time (measured in working days) elapses from the day they were notified. The vertical dashed lines denote the deadline for each wave.

FIGURE VII: CUMULATIVE RESPONSES OVER TIME



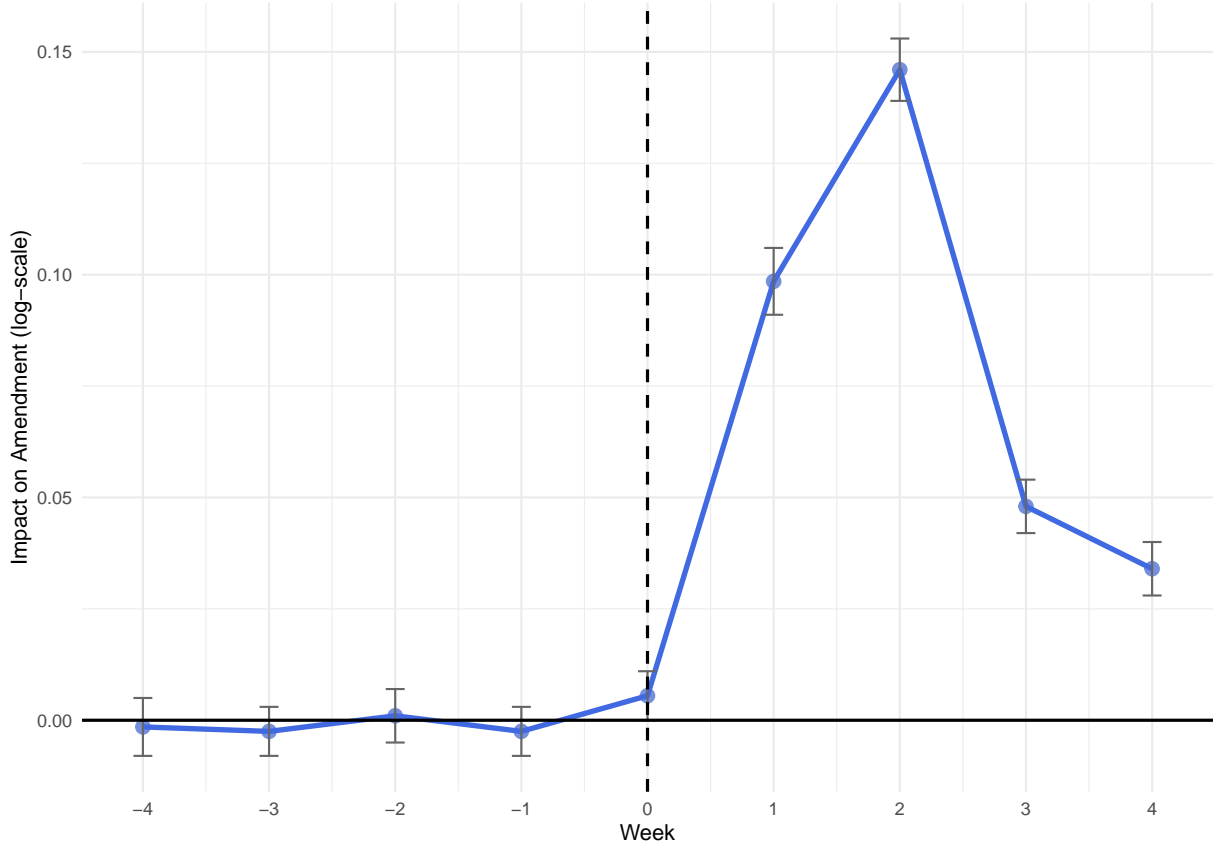
Notes: The figure tracks the content of notified taxpayers' amendments. For each wave of notified taxpayers we track the cumulative changes in outcomes implied by their amendments (e.g. the cumulative increase in reported taxable sales). We then divide it by the number of notified taxpayers to create measures of the average response per notified taxpayer to allow us to compare across waves of different sizes. Panels A–C report cumulative increases in sales, purchases, and tax liabilities, respectively. Panels D–F study changes to the notified taxpayers' net position with respect to the tax authority. Panel D documents large increases in payments from notified taxpayers to the treasury, while panel E shows decreases in credits the taxpayers have from the tax authority. Panel F combines the increased payments and the reduced credits to create a measure of the net value of the notification. Specifically, the net value is the change in payments - $0.7 \times$ the change in credits.

FIGURE VIII: STACKED DATA SPECIFICATION USED TO PERFORM RANDOMIZATION INFERENCE



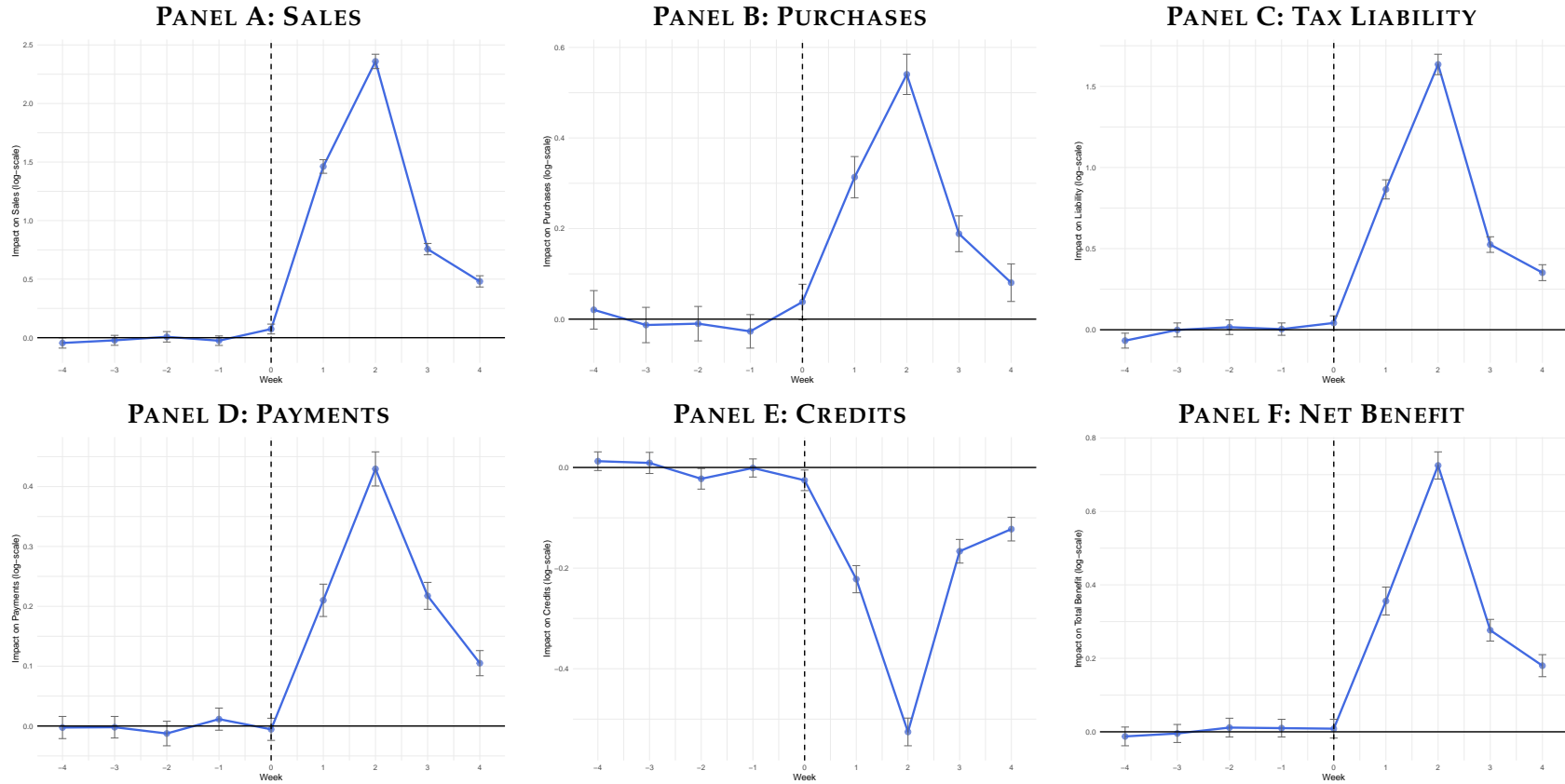
Notes: The figure illustrates our *stacked data specification*. The figure shows that for a positive-value week (e.g., 1, 2, . . .), our randomization tests compare the outcomes of eligible taxpayers observed in that particular week to the outcomes of the same taxpayers within 30 days leading to the treatment event (marked as “controls” in the figure). For a nonpositive-valued week (0, −1, −2, . . .), the randomization test compares the outcomes of eligible taxpayers in that week to their outcomes in the 30 days prior to that week.

FIGURE IX: NOTIFICATIONS INCREASE TAXPAYERS' AMENDMENT RATE



Notes: The figure shows the randomization-based 95% confidence intervals (CIs) for the direct treatment effect — assuming no interference— for the amendment rate in each week $w \in \{-4, \dots, 0, \dots, 4\}$. In each week we test the hypothesis $H_0^{te}(\tau)$, defined in Equation (8): $H_0^{te}(\tau) : f_i(1, w) = f_i(0, w) + \tau$, for all $i \in [N], w \in \mathbb{Z}$. We then invert the test by keeping those values of τ that correspond to a randomization p -value that is no larger than 0.05, producing the randomization-based confidence interval.

FIGURE X: TREATMENT EFFECTS ON AMENDMENTS' CONTENTS



Notes: The figure shows the randomization-based 95% confidence intervals (CIs) for the direct treatment effect—assuming no interference—for the amendments' contents in each week $w \in \{-4, \dots, 0, \dots, 4\}$. In each week we test the hypothesis $H_0^{te}(\tau)$, defined in Equation (8): $H_0^{te}(\tau) : f_i(1, w) = f_i(0, w) + \tau$, for all $i \in [N], w \in \mathbb{Z}$. We then invert the test by keeping those values of τ that correspond to a randomization p -value that is no larger than 0.05, producing the randomization-based confidence interval. Panels A–C report cumulative increases in sales, purchases, and tax liabilities, respectively. Panels D–F study changes to the notified taxpayers' net position with respect to the tax authority. Panel D documents large increases in payments from notified taxpayers to the treasury, while panel E shows decreases in credits the taxpayers have from the tax authority. Panel F combines the increased payments and the reduced credits to create a measure of the net value of the notification. Specifically, the net value is the change in payments - $0.7 \times$ the change in credits.

TABLE I: TREATMENT BALANCE

Covariate		Cluster Design		Split Graph Design	
		Wave 1	Wave 2	Wave 3	Wave 4
1	Legal Form	0.43	0.22	0.97	0.75
2	Size	0.05	0.93	0.49	0.18
3	Region	0.94	0.54	0.13	0.10
4	Economic Activity	0.57	0.99	0.27	0.12
5	Age Group	0.35	0.60	0.80	0.58

Notes: The table shows the results of balance checks on the randomization. As our randomization protocols varied across waves, we present balance results separately for Waves 1–2 and Waves 3–4. Waves 1–2 followed the saturated cluster randomized design described in Section II.v. In clusters assigned to 0 or 100% saturation we cannot compare treatment and control taxpayers. Instead, we summarize the covariates of taxpayers in the cluster into *cluster-level covariates*, and then test balance at the cluster level. Specifically, we perform the following steps to diagnose balance: 1: For covariate Z with levels $k = 1, \dots, K$, calculate the cluster covariate for each cluster J : $C_J = \text{entropy}(\{c_k\}_{k=1}^K)$, $c_k = \sum_{i \in J} 1\{Z_i = k\}/|J|$. Here, $\{c_k\}_{k=1}^K$ is the empirical distribution of X within cluster J , while C_J is the entropy of this distribution, which serves as the cluster-level covariate; 2: Regress C_J on p_J , where p_J is the saturation rate for cluster J . Report the p -value of the t -test from the regression. The results from this calculation are shown in columns 1 and 2.

In Waves 3–4, the randomization changed to a stratified network experiment as described in Section II.v. Randomization consisted of multiple simple randomized experiments within sets of taxpayers with the same type x and anti-clique membership status. As a result, we can test for balance at the taxpayer level as follows: 1: Within each wave, calculate each taxpayer i 's type x_i (degree, orientation); treatment $D_i \in \{0, 1\}$; and anti-clique status $A_i \in \{0, 1\}$; 2: For each covariate Z , fit two multinomial regression models: $M_0 : Z_i \sim x_i + A_i$, and $M_1 : Z_i \sim D_i + x_i + A_i$, and report the p -value from the ANOVA comparison between M_0 and M_1 . If treatment is predictive of the covariate, then such p -value should be small. The results from this calculation are shown in columns 3 and 4.

TABLE II: AMONG NOTIFIED TAXPAYERS: RESPONSES BY TYPES OF TAXPAYERS, INTERACTIONS

	1[Amendment]	Δ Sales	Δ Purchases	Δ Tax Liability	Δ Payment	Δ Credit	Net Benefit
Unincorporated X Small Taxpayer	0.7554*** (0.0066)	177322411*** (6397185)	29696102*** (3275113)	12749234*** (536528)	1971559*** (158306)	-1782806*** (110613)	3219523*** (183296)
Incorporated X Small Taxpayer	0.6014*** (0.0202)	553409959*** (49299112)	81129622*** (21659882)	62550316*** (5193608)	7553385*** (984461)	-7226746*** (838371)	12612107*** (1177654)
Unincorporated X Medium Taxpayer	0.2778*** (0.0757)	249001324*** (36097221)	26623084** (9132347)	22250785*** (2965695)	9452988*** (1493819)	-6710725*** (1408340)	14150496*** (2453741)
Incorporated X Medium Taxpayer	0.5354*** (0.0504)	529787981*** (79043820)	23758595* (9582732)	208518286*** (52699566)	14450181*** (3258257)	-15841085*** (4785661)	25538941*** (5056076)
Incorporated X Large Taxpayer	0 (0)	97898871*** (27152261)	-584180055* (289308763)	300551875*** (71452081)	0 (0)	-2141232** (685743)	1498863** (480020)
Num.Obs.	4975	4975	4975	4975	4975	4975	4975

Notes: The table shows estimates of the heterogeneity of the responses across types of notified taxpayers, focusing on their size classification (small/medium/large) and their legal form (incorporated/unincorporated). In the sample of amendments by notified taxpayers, we regress the outcome of taxpayer i on indicators for the different groups $g(i) \in \{\text{Incorporated, Unincorporated}\} \times \{\text{Small, Medium, Large}\}$ that the taxpayers i belong to: $y_i = \beta_h \mathbf{1}[g(i) = h] + \varepsilon_i$, and we report the β_h coefficients.

TABLE III: REPORTED SALES, PURCHASES, AND TAX LIABILITY BY TAXPAYER TYPE**(A) MEDIAN VALUES**

Type	N	Reported Sales	Reported Purchases	Reported Tax Liability
Unincorporated X Small	4239	30934198	27536364	0
Incorporated X Small	587	181956567	132041760	0
Unincorporated X Medium	36	1088716775	571590510	12671677
Incorporated X Medium	99	613819719	211464037	4510789
Incorporated X Large	14	3658694296	2405293225	-45408485
Total	4975	4975	4975	4975

(B) MEAN VALUES

Type	N	Reported Sales	Reported Purchases	Reported Tax Liability
Unincorporated X Small	4239	254892186	189985567	148540
Incorporated X Small	587	1329616126	1048966522	-13537108
Unincorporated X Medium	36	1435296620	742567195	62006771
Incorporated X Medium	99	2412371339	1836405125	-3675766
Incorporated X Large	14	9361276024	5777092862	-12856772
Total	4975	4975	4975	4975

Notes: The tables show the heterogeneity of the reported sales, purchases, and tax liabilities across types of notified taxpayers, focusing on their size classification (small/medium/large) and their legal form (incorporated/unincorporated).

TABLE IV: CAUSAL IMPACTS OF NOTIFICATIONS

	Amendment	Sales	Purchases	Liability	Payments Due	Credit	Value
Notified	0.6103 (0.0181) [0.010]	0.5462 (0.0152) [0.006]	0.1192 (0.0101) [0.003]	0.5161 (0.0149) [0.002]	0.164 26 (0.008 22) [0.004]	0.1698 (0.0089) [0.005]	0.2452 (0.0106) [0.006]
Control mean	0.141	0.0794	0.0718	0.0783	0.014	0.0184	0.0268
Num.Obs.	14 918 211	14 918 211	14 918 211	14 918 211	14 918 211	14 918 211	14 918 211

Notes: This table presents the effects of the notifications on the likelihood of amending tax returns. Each observation is a firm-day. Each outcome equals 100 for a given firm-day if the firm filed a given type of amendment on that day, and 0 otherwise. Column 1 shows the impacts on the likelihood of filing any amendment; Col. 2: filing an amendment that increases the firm's reported sales; Col. 3: filing an amendment that increases the firm's reported purchases; Col. 4: filing an amendment that increases the firm's tax liability; Col. 5: filing an amendment that increases the taxpayer's payments due to the tax authority; Col. 6: filing an amendment that reduces the taxpayer's outstanding tax credits from the tax authority; and Col 7: filing an amendment that increases the net value of the taxpayer's declaration to the tax authority, where the net value is the change in tax payment - $0.7 \times$ the change in tax credit. The results come from estimation of equation (9): $y_{kit} = \beta D_{kit} + \omega p_{kit} + \mu_{ki} + \delta_{kt} + \epsilon_{kit}$ where y_{kit} is the outcome for taxpayer i at date t , within notification wave k ; D_{kit} is an indicator equal to 1 if taxpayer i was notified within wave k and if date t is after the notification date for that wave; p_{kit} takes value 0 before the notification date, and is otherwise equal to the probability that taxpayer i was randomly selected to receive a notice within wave k ; μ_{ki} are notification wave-by-taxpayer fixed effects; and δ_{kt} are notification wave-by-time fixed effects. Standard errors clustered at the taxpayer level are reported in parentheses, and randomization inference p-values in square brackets.

TABLE V: CAUSAL IMPACTS OF NOTIFICATIONS: EFFECTS BY TAXPAYER TYPE

	Amendment	Sales	Purchases	Liability	Payments Due	Credit	Value
Notified:Large,Incorporated	−2.41 (2.36) [0.186]	−0.663 (0.581) [0.576]	−0.940 (0.585) [0.733]	−1.52 (1.83) [0.011]	0.0659 (0.0547) [0.043]	−1.92 (1.72) [0.473]	−1.87 (1.71) [0.078]
Notified:Medium,Unincorporated	0.521 (0.203) [0.299]	0.615 (0.147) [0.312]	−0.124 (0.139) [0.122]	0.653 (0.159) [0.060]	0.1346 (0.0674) [0.887]	0.0419 (0.0349) [0.769]	0.142 (0.067) [0.582]
Notified:Medium,Incorporated	0.341 (0.149) [0.455]	0.307 (0.101) [0.639]	−0.0406 (0.0824) [0.967]	0.325 (0.127) [0.494]	0.0428 (0.0366) [0.509]	0.0191 (0.0557) [0.279]	0.0559 (0.0620) [0.685]
Notified:Small,Unincorporated	0.6432 (0.0193) [0.784]	0.5730 (0.0167) [0.875]	0.1369 (0.0109) [0.131]	0.5373 (0.0161) [0.534]	0.172 31 (0.009 06) [0.644]	0.180 08 (0.009 52) [0.708]	0.2606 (0.0114) [0.586]
Notified:Small,Incorporated	0.496 (0.056) [0.666]	0.4422 (0.0433) [0.910]	0.0562 (0.0306) [0.380]	0.4492 (0.0428) [0.184]	0.1449 (0.0241) [0.554]	0.1656 (0.0263) [0.444]	0.2158 (0.0313) [0.593]
Control mean	0.142	0.0804	0.0727	0.0792	0.0142	0.0187	0.0271
Num.Obs.	14 714 447	14 714 447	14 714 447	14 714 447	14 714 447	14 714 447	14 714 447

Notes: This table presents estimates of the heterogeneity of the effects of the notifications on the likelihood of amending tax returns by type of taxpayer. Each observation is a firm-day. Each outcome equals 100 for a given firm-day if the firm filed a given type of amendment on that day, and 0 otherwise. Column 1 shows the impacts on the likelihood of filing any amendment; Col. 2: filing an amendment that increases the firm's reported sales; Col. 3: filing an amendment that increases the firm's reported purchases; Col. 4: filing an amendment that increases the firm's tax liability; Col. 5: filing an amendment that increases the taxpayer's payments due to the tax authority; Col. 6: filing an amendment that reduces the taxpayer's outstanding tax credits from the tax authority; and Col 7: filing an amendment that increases the net value of the taxpayer's declaration to the tax authority, where the net value is the change in tax payment - $0.7 \times$ the change in tax credit. The results come from estimation of Equation (9) $y_{ijkt} = \sum_j \beta_j D_{kit} \times S_{ij} + Post_t \times S_{ij} + \omega p_{kit} + \mu_{ki} + \delta_{kt} + \varepsilon_{ijkt}$ where terms are defined as in the notes to table IV, S_{ij} indicates taxpayers of type j , where j indexes the subgroups; and $Post_t$ is an indicator for time periods after the notification. Standard errors clustered at the firm-level are reported in parentheses, and randomization inference p-values in square brackets.