ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Chalendard, Cyril et al.

Working Paper

Does Better Information Curb Customs Fraud?

CESifo Working Paper, No. 8371

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Chalendard, Cyril et al. (2020) : Does Better Information Curb Customs Fraud?, CESifo Working Paper, No. 8371, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/223443

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

CESifo WORKING PAPERS

8371
2020

June 2020

Does Better Information Curb Customs Fraud?

Cyril Chalendard, Alice Duhaut, Ana M. Fernandes, Aaditya Mattoo, Gael Raballand, Bob Rijkers

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8371*

Does Better Information Curb Customs Fraud?

Abstract

This paper examines how providing better information to customs inspectors and monitoring their actions affects tax revenue and fraud detection in Madagascar. First, an instrumental variables strategy is used to show that transaction-specific, third-party valuation advice on a subset of high-risk import declarations increases fraud findings by 21.7 percentage points and tax collection by 5.2 percentage points. Second, a randomized control trial is conducted in which a subset of high-risk declarations is selected to receive detailed risk comments and another subset is explicitly tagged for ex-post monitoring. For declarations not subject to third-party valuation advice, detailed comments increase reporting of fraud by 3.1 percentage points and improve tax yield by 1 percentage point. However, valuation advice and detailed comments have a significantly smaller impact on revenue when potential tax losses and opportunities for graft are large. Monitoring induces inspectors to scan more shipments but does not result in the detection of more fraud or the collection of additional revenue. Better information thus helps curb customs fraud, but its effectiveness appears compromised by corruption.

JEL-Codes: D730, F140, H260, K420.

Keywords: tariff evasion, tax enforcement, third-party information, performance monitoring, risk management, information provision, randomized control trial.

Cyril Chalendard
International Trade Centre
cchalendard@worldbank.org

Alice Duhaut
World Bank Group
aduhaut@worldbank.org

Ana M. Fernandes
World Bank Group
afernandes@worldbank.org

Aaditya Mattoo
World Bank Group
amattoo@worldbank.org

Gael Raballand
World Bank Group
graballand@worldbank.org

Bob Rijkers
World Bank Group
brijkers@worldbank.org

May 2020

We thank Madagascar Customs and GasyNet for sharing their data, knowledge, and their invaluable help implementing and designing the randomized control trial. We are especially grateful to Eric Rabenja, Dannick Fiononana, Antsa Rakotoarisoa, Cédric Catheline, Stephane Manouvrier, Prisca Michea, Tolotra Hasinantenaina Ramarosandratana, Tahina Rabesalama, and Jose Hermand, whose support was critical to the success of this project. We also thank, Henrik Kleven, Beata Javorcik, Joana Naritomi, Sandra Sequeira, Joel Slemrod, Shang-Jin Wei, Simeon Djankov, and seminar participants at the World Bank for useful comments and discussions. Robert Marty provided excellent research assistance on text mining analysis. Hibret Maemir and George Schaur helped with the differentiated and time-sensitive product classifications. This paper has been partly supported by the World Bank's Multidonor Trust Fund for Trade and Development and the Strategic Research Partnership on Economic Development. We also acknowledge the generous financial support from the World Bank research support budget and the Knowledge for Change Program (KCP), a trust funded partnership in support of research and data collection on poverty reduction and sustainable development housed in the office of the Chief Economist of the World Bank. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank of Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the countries they represent. All errors are our responsibility.

1. Introduction

Developing countries collect less tax revenue as a share of GDP than their developed counterparts and rely more on taxes collected at the border (Baunsgaard and Keen, 2010), while suffering higher levels of tariff evasion (Jean et al., 2018). Although there are many reasons for these differences in revenue mobilization along the development spectrum (Kleven et al., 2016), an important one may be the performance of civil servants entrusted with collecting taxes (Khan et al., 2015; Khan et al., 2019; Finan et al., 2015; Dincecco and Ravanilla, 2017). Given the reliance on trade taxes, customs inspectors play a central role in determining tax liabilities and enforcement in low-income countries. The inspectors may, however, not be able to determine taxes accurately because they are not fully informed and third-party reporting is inadequate. It is also possible that the incentives to enforce are weak compared to the gains from colluding with economic operators to reduce tax receipts – because inspectors are insufficiently monitored and the rewards for good performance are relatively poor.

This paper examines how the provision of information to, and the monitoring of, inspectors impacts customs performance and tax revenue mobilization in Madagascar. First, we estimate the returns to third-party provision of information on import values – henceforth designated as 'valuation advice.' The information service provider is GasyNet, a public-private partnership that assists customs with information technology (IT), equipment and risk analysis. The valuation advice comprises a detailed, high-quality report on what the price of each item included in the declaration should be. The report is issued only for a subset of high-risk declarations for which the risk of tariff evasion is especially high. Second, we document the results of a nationwide risk management randomized control trial (RCT) in which the customs' Risk Management Unit (RMU) was asked to provide better information to inspectors. This information was in the form of detailed comments related to valuation on a subset of randomly selected high-risk declarations. At the same time, a different subset of randomly selected high-risk declarations were flagged as being subject to *ex-post* monitoring.

The main methodological challenge in identifying the impact of information and monitoring on customs inspectors' performance is that neither is typically randomly assigned. Import declarations for which the risk of tax evasion is elevated are more likely to be the subject of valuation advice

and to be commented on by the RMU. Such declarations are also more likely to be monitored, given their potential for tax revenue recovery, and because risk information facilitates auditing of inspectors' behavior.

We address this challenge in several ways. To start with, our specifications include an unusually rich set of control variables that go a long way towards addressing selection bias in our regressions. Specifically, we are able to control for each import declaration's risk score, the recommended inspection channel, a rich set of product characteristics, and importer, inspector, product, source country, and port-time fixed effects. In robustness tests, we also include transaction-specific proxies for undervaluation as well as inspector-month and inspector-broker fixed effects.[1]

Second, to address the potential endogeneity of valuation advice, we use an instrumental variables strategy exploiting variation in the supply of valuation advice. This strategy relies on the fact that the propensity for declarations to be selected for valuation advice varies with their risk score and their port of arrival. It assumes that there are no direct effects of the share of declarations with a given risk score in a particular port being subject to valuation advice on inspector behavior, after conditioning on port-month fixed effects, risk score dummies, and a rich set of other controls. This assumption could be violated if that share of declarations was correlated with inspectors' workload and/or changes in importer behavior. We show that our results are robust to controlling for inspectors' workload and proxies for undervaluation that should capture potential changes in importers' behavior. In addition, we demonstrate that our instrument does not predict how inspectors handle declarations not receiving valuation advice.

Third, we implement what is (to our knowledge) the first risk management RCT in customs, thanks to a partnership with the customs administration and the service provider GasyNet who agreed to randomly select in real time import declarations for additional information provision and monitoring treatments.[2] The RCT enables us to identify separately the impact of monitoring from the impact of the provision of information, and is unique given that risk analysis and management

[1] The customs broker is an intermediary that acts on behalf of the importer to handle clearance of goods and related activities.

[2] The ability to implement this RCT was the culmination of four years of intensive World Bank engagement with the customs administration in Madagascar during which a relationship of trust was established.

are and must be highly secretive to deter tax evasion. Monitoring was conducted by the performance management unit that directly reports to the Director General of Customs.

Our main findings can be summarized as follows. First, according to our preferred instrumental variable estimates over the period January 2016-October 15, 2018, valuation advice increased the probability of goods being scanned by 10.3 percentage points and the probability of a declaration being deemed fraudulent by 21.7 percentage points. It also increased tax revenue by 5.2 percentage points on average and significantly prolonged clearance times. The gain in tax revenue is sizeable but small relative to the likely overall undervaluation. For perspective, importers declare values that are on average 37 percent lower than the valuation advice reference value. Furthermore, valuation advice is least effective in improving customs outcomes for declarations for which opportunities for graft are largest. We categorize import declarations as 'high-stakes' if their potential tax revenue yield (calculated over the reference value provided in the valuation advice) is higher than a cutoff of 10,000 USD. Undervalued high-stakes declarations are *less* likely to yield reported fraud, yield *less* tax adjustment, and are cleared more quickly than lower-stakes declarations that are similarly undervalued. Inspectors thus react significantly, but not optimally,[3] to valuation advice. If they were acting with integrity, we would expect a much stronger increase in revenue and fraud detected for high-stakes declarations.

Second, our RCT providing information and monitoring for high-risk import declarations over a period of 14 weeks — from October 15, 2018 to February 6, 2019 — resulted in a dramatic change in the quality of the comments provided by the customs' RMU. The length of the comments, the inclusion of price information, and the references to prior transactions increased significantly in response to the RCT, especially for declarations subject to high taxes for which the risk of tariff evasion is largest.

The more detailed comments significantly improved fraud detection and tax yield. Specifically, for declarations randomly selected to receive detailed comments (but not subject already to valuation advice) the likelihood of fraud being reported by inspectors increased by 3.1 percentage

[3] Given the high quality of the valuation advice from a revenue maximization perspective, the theoretically optimal response would be to apply the reference value and eliminate tax evasion entirely. Of course, valuation is prone to some measurement error but this is unlikely to be large enough to account for the discrepancy between actual and theoretically optimal adjustment. We are also not considering the possibility that more aggressive tax enforcement may deter trade (and could hence reduce revenue).

points and taxes collected increased by a full percentage point.[4] But the tax adjustments are again significantly smaller for high-stakes declarations for which opportunities for graft are largest, suggesting opportunistic behavior by inspectors. Detailed comments also prolonged clearance times by encouraging inspectors to upgrade more declarations to the red channel and to do more scanning of their contents.

Within the set of declarations that received detailed comments, only the subset that contain specific quantitative metrics, including information on prices and/or referring to previous transactions, saw significant improvements in customs outcomes. The extent to which information improves customs outcomes thus seems to be correlated with its quality and precision. That finding would also explain why valuation advice by the third-party service provider (which is more precise and detailed) seems to have a much larger impact than the detailed comments provided in the RCT.

Explicit monitoring tags significantly increase the prevalence of scanning but none of the other customs outcomes for any declarations. This lack of effectiveness could be related to the timing of the RCT, which took place after a period of strikes and during the 2018/2019 presidential elections. Customs inspectors may have anticipated a change of senior customs leadership, which could have diluted their fear of sanctions for improper behavior. However, the ineffectiveness is also consistent with an enduring culture of impunity and absence of strong performance incentives. Only 6 percent of customs inspectors believe that non-ethical behavior was appropriately sanctioned and only 13 percent believe promotions were merit-based, according to a survey we conducted prior to the start of the RCT.

To interpret these results, some limitations of the study must be kept in mind. Most importantly, we are unable to measure whether information provision increased the prevalence of extortion by customs inspectors. We did conduct structured interviews with representatives from the private sector before and after the RCT. Although there were some (seemingly isolated) complaints about a select few inspectors using the RCT as an opportunity to demand additional bribes and delay clearance, the RCT does not appear to have induced a widespread change in the prevalence of extortive practices. Second and related, we use fraud records and tax adjustments made by the

[4] While detailed comments appear to help inspectors detect fraud among declarations for which valuation advice was not issued, they do not change significantly most inspector actions and customs outcomes for declarations for which valuation advice is also issued. This may in part be due to imprecise estimates given a very limited number of declarations selected for valuation advice.

inspector as proxies for fraud detection. While these proxies may be prone to some reporting error, they are likely to be good indicators of actual fraud rather than reflecting inspectors' extortion of compliant economic operators, given that tariff evasion is widespread in Madagascar (Chalendard et al., 2019). Third, the information provided by means of the detailed comments treatment in the RCT and the valuation advice by the service provider is not private but can best be thought of as establishing common knowledge. Inspectors may respond to it not (only) because it helps them identify fraud, but (also) because they recognize that such information makes it easier to audit their behavior. This inherent inseparability of information and ease of auditing in this context is why we designed the RCT to include a (zero information) pure monitoring treatment. Finally, it is possible that the provision of information for and monitoring of a subset of declarations impacts how inspectors handle declarations for which such information is not made available and are not monitored, which could lead to biased estimates of the impact of information and monitoring. While such spillovers are hard to measure, we will present suggestive evidence that they are very limited at worst, and that results are unlikely to be driven by a reallocation of inspector effort from declarations for which information is limited to declarations for which information is made available.[5]

Notwithstanding these limitations, our findings suggest that better information improves fraud detection but that its effectiveness is undermined by limited incentives to optimally make use of this information. Bribery and weak sanctions for opportunistic behavior can explain why the provision of information to inspectors does not yield even stronger impacts.

This paper contributes to and builds on different strands of literature. To start, we add to the nascent literature on the performance of bureaucrats as a determinant of state effectiveness and tax collection (Olken and Pande, 2012; Dincecco and Ravanilla, 2017; Pepinsky et al., 2017; Xuo 2018) by focusing on customs inspectors and highlighting the pivotal role they play in revenue mobilization in low-income countries. The typically small number of inspectors account for a

[5] A related concern are spillovers across ports and mode of transportation but in practice such substitution is typically prohibitively costly, according to private sector representatives whom we interviewed before and after the implementation of the RCT.

sizeable share of government revenue. Each inspector in our study oversees the collection of 0.6% of total tax revenue on average.

Second and related, we contribute to the literature on tariff evasion (Bhagwati, 1964, 1967; Fisman and Wei, 2004; Yang, 2008; Dutt and Traca, 2010; Sequeira and Djankov, 2014; Sequeira, 2016; Wier, 2020), which has mostly focused on the determinants of tariff evasion. Our study is the first to focus on the effectiveness of remedial measures undertaken by customs agencies to reduce tariff evasion.[6]

Third, in doing so, we contribute to the tax enforcement literature (Slemrod, 2019) and more specifically to the growing body of evidence showing that third-party information can be leveraged to improve tax compliance (e.g., Kumler et al., 2015; Kleven et al., 2011; Pomeranz, 2015; Naritomi, 2019). We provide evidence of sizeable returns to the provision of declaration-specific valuation advice by an external organization (GasyNet). This is not standard third-party reporting as in the rest of the public finance literature, but it is close in spirit in that it is information on taxable revenue obtained by a third-party other than the customs department and the taxpayer.[7]

These findings suggest low-income countries can improve tax collection by hiring a third-party service provider to assist with risk analysis, since Gasynet valuation advice services substantially augmented tax revenues. However, the returns to internal restructuring of the customs RMU are arguably even higher since our RCT improved revenue yield at very low marginal costs given that no additional staff were recruited, and it only relied on changing the modus operandi of the RMU. This argument is consistent with the fact that most customs administrations in advanced economies have accumulated internal capacity for risk analysis and do not rely on extensive provision of risk diagnostics by external entities.

Finally, we add to the small body of literature that has used quasi-experimental methods to assess how risk management impacts customs clearance times and trade flows (Martincus et al., 2015;

[6] Other forms used to reduce customs duty avoidance are pre-shipment inspections in the exporting country studied in Anson et al. (2006) and Yang (2008).

[7] The discrepancies in mirror statistics also referred to as "evasion gaps" are shown to increase with tariff rates and vary with product characteristics (Javorcik and Narciso, 2008; Fisman and Wei, 2009), enforcement (Mishra et al., 2008), customs organization and country characteristics such as the level of corruption (Javorcik and Narciso, 2017; Jean et al., 2018) and political connections (Rijkers et al., 2016).

Fernandes et al., 2019; Laajaj et al., 2019). To our knowledge, we are the first to report the results of an RCT and to focus on fraud detection and inspector performance.

The remainder of the paper is organized as follows: Section 2 describes the Madagascar customs context, the clearance process, and elaborates on our hypotheses; Section 3 describes the data; Section 4 presents the results on the impact of valuation advice, while Section 5 describes the RCT and reports its results; Section 6 concludes.

2. Context and hypotheses

With a tax-to-GDP ratio below 10 percent, Madagascar's revenue mobilization is among the weakest in the world.[8] Taxes and duties collected by customs accounted for 48.9 percent of overall tax revenue in 2019, in spite of substantial tariff evasion which has been estimated to result in revenue losses equivalent to at least 30 percent of non-oil revenues (Chalendard et al., 2019).[9]

The performance of individual inspectors has important fiscal implications. The number of frontline inspectors overseeing the collection of tariff revenue is limited and trade and tariff revenues are highly concentrated in a small number of ports. Toamasina, the main seaport of Madagascar, which accounts for 66.5 percent of non-oil imports and 75.8 percent of non-oil tax revenue, employed approximately 16 inspectors per year on average for clearing non-oil declarations over our sample period. Another 25 inspectors were employed in the seven other seaports considered in this study (Antanimena, Antsiranana, Mahajanga, Nosy Be, Toliary, Tolagnaro, Vohemar). Each inspector in our sample on average oversees the collection of approximately 10 million USD in terms of import revenues each year, or approximately 0.6 percent of the budget of the government of Madagascar. These numbers mask substantial heterogeneity across inspectors, with inspectors in Toamasina each overseeing the collection of more than 30 million USD.

Customs inspectors' positions are among the most lucrative and most sought after jobs in Madagascar. Customs inspectors typically earn a salary of approximately 11,000 USD per annum

[8] Madagascar ranks 112th out of 115 economies in terms of tax effort versus total tax potential (Godin et al., 2015).
[9] Taxes and duties collected by customs include tariffs but also consumption taxes such as value-added taxes.

(roughly 21 times annual GDP per capita, which is 527 USD), and receive between 5 and 20 percent of the fines they issue as a bonus.[10] Inspectors in Toamasina are employed under a performance monitoring scheme in which they can earn a bonus of 1,000 USD if they are the best performing inspector in terms of fraud detection, revenue mobilization, and speed of clearance in a given quarter. Their performance is evaluated relative to performance targets, which are revised regularly. Particular attention is paid to how they handle declarations allocated to the physical inspection (red) channel as well as declarations for which valuation advice was issued, since such declarations are associated with the largest potential tax losses. The shares of these types of declarations that are deemed fraudulent are among the metrics used to assess inspectors' performance.

In spite of these performance incentives, corruption appears rampant,[11] perhaps because official compensation is low relative to opportunities for graft or because of coercion by economic operators.[12] The customs department is characterized by a culture of impunity and absence of meritocracy. Inspectors are hardly ever fired, even in cases of obvious corruption.[13] According to a nationwide survey of inspectors that we conducted prior to the RCT, only 23 percent believe that their colleagues act with integrity, only 6 percent claim non-ethical behavior is sanctioned and only 12 percent believe promotions are merit-based.[14] Undervaluation of imports, which is the focus of our study, was widely agreed to be the most important type of customs fraud. This is consistent with formal fraud records which classify 67.2 percent of all fraud as underreporting of value, 27.4 percent as underreporting of quantities, and the remainder as misclassification (4.9 percent) and misreporting country of origin (0.5 percent).[15] However, only 38 percent of inspectors said they knew which firms were likely to commit customs fraud, suggesting that asymmetric

[10] In 2017, 15,000 candidates applied for a 300-position recruitment program launched by Madagascar customs. The best-ranked graduates from the national school of administration usually select customs as their preferred administration post.
[11] Madagascar ranked 158th out of 180 countries in Transparency International's 2019 Corruption Perception Index.
[12] According to a very senior civil servant, informal payments to some inspectors are north of 200,000 USD each year.
[13] Cantens et al. (2010) show similar environments with a culture of impunity across customs agencies in African countries.
[14] Due to a close collaboration with the World Bank, the Director General of Customs made the response to our survey be mandatory for all inspectors.
[15] The importance of undervaluation as the main form of tariff evasion in contrast to product and especially country of origin misclassification is also consistent with the very simple import tariff structure imposed by Madagascar with only four ad-valorem tariff rates (0, 5, 10, and 20) and no specific tariffs except on petroleum and gas (WTO, 2015).

information constrains fraud detection. However, it is also possible that inspectors are colluding with economic operators or are not sufficiently empowered to withstand the threats from operators that 31 percent of inspectors claim being subjected to on a regular basis. Reform proposals to enhance transparency and expand the use of performance management resulted in collective strikes at the customs department, suggesting inspectors did not welcome increased performance monitoring.

To understand how improved information and monitoring may ameliorate customs performance, it is helpful to consider the various steps of the customs clearance process, a stylized version of which is presented in Figure 1.

1. **Registration.** The first step is the electronic submission of an import declaration to the customs administration by the importer (or by a broker acting on behalf of the importer).
2. **(Pre-) Risk analysis by service provider GasyNet and potential issuance of valuation advice.** The second step is risk analysis of the declaration by GasyNet,[16] which issues for each import declaration (i) a risk score (on a scale of 1 to 9)[17] based on its proprietary risk model as well as (ii) a clearance channel recommendation along with qualitative comments justifying the recommendation. In addition (iii) for a subset of high-risk declarations for which the accuracy of the declared import value is questionable, specialized staff located abroad issue a valuation advice. Such valuation advice comprises a very detailed report (*Rapport d'Opinion sur la Valeur délivrée à destination*) on what the value of the specific declaration is likely to be, along with rigorous documentation and extensive justification for the proposed valuation. The advised value is calculated using specific valuation studies, official international market prices of goods, the value of similar transactions, and a network of offices/representatives in exporting countries, all of which is proprietary information

[16] Established in 2007, GasyNet has access to the customs electronic system where import declarations are submitted and it provides other services to Madagascar customs, including operating the Trade Single Window, inspecting second-hand imported vehicles in a separate terminal, and running cargo scanners (WTO, 2015). In exchange for these services, a fee – called Prestations GasyNet (PGN) – set up at 0.5 percent of the cost-insurance-freight value of the declaration has to be paid by the trader. If the free-on-board value does not exceed 25,000 euros, then a flat rate applies.

[17] The risk score ranges from 1 to 10 but we exclude declarations with risk score 10 since these are typically reserved for the importation of used vehicles, which follows a special procedure, as explained in Section 3.

to GasyNet, which is a subsidiary of the Swiss multinational group Société Générale de Surveillance (SGS) and leverages its extensive global expertise.

3. **Risk analysis by the customs Risk Management Unit (RMU).** The third step is risk analysis by the RMU (*Service du Renseignement et de l'Analyse des Risques)*. The RMU makes use of the information received from the service provider and its own risk model and recommends a clearance channel along with qualitative instructions meant to help frontline customs officers better detect fraud. Discussions with counterparts in customs indicated that prior to the start of the RCT, the RMU was widely regarded to be a dysfunctional unit with little value added. Their comments were typically either lacking altogether, or simply copied from the service provider's advice.
4. **Assessment by the inspector.** The fourth step is the assessment of the declaration by a frontline customs inspector. Based on the documentation submitted by the importer (or his broker acting as his representative) and the risk diagnostics provided by the service provider and the RMU (i.e., the risk score, inspection channel recommendations and comments, and a potential valuation report). She first has to take a decision on which clearance channel to effectively subject the declaration to. She may decide to follow the RMU's recommendation or to upgrade (i.e., move the declaration from the yellow to the red channel) or in exceptional cases, downgrade (i.e., move the declaration from the red to the yellow channel). If the yellow channel is selected, the inspector examines only the documents. If the red channel is selected, the inspector examines all documents, orders (a select number of) containers to be scanned, and/or physically inspects the container(s), possibly opening boxes inside the container(s) and taking samples of the cargo. Based on all information gathered in the documentary control, the results of the potential scan and the physical inspection, the inspector produces a report in which she indicates whether adjustments to the import value. quantity and/or product classification are to be made. The inspector then assesses the duties and taxes based on the (potentially revised) final value of the import declaration and any penalties if fraud is detected.
5. **Clearance.** Finally, in the last step the importer makes the payment of taxes, duties and any penalties, and then the goods are released.

Inspectors are thus responsible for detecting fraud and potentially adjusting values and taxes upwards. Verifying whether the values declared by the importer are accurate is the most challenging task for the inspector, who usually only has limited recourse to third-party information for verification.[18] Whereas underreporting of quantities and misclassification are readily detected by physical inspection, undervaluation is not. WTO provisions stipulate a presumption in favor of accepting the value on the invoice submitted by the importer as the primary basis for customs valuation.

We hypothesize that transaction-specific information on whether this value is correct will facilitate detecting fraud and enhance tax collection, not only because it helps identify mis-declaration of value but also because it may make it easier to justify insisting on value adjustments. We anticipate stronger effects when information is of higher quality, i.e., more precise and better justified and we therefore expect valuation advice (which comes with extensive documentation) to have a stronger effect on inspectors' actions and customs performance than the detailed comments provided in the RCT. In addition, we expect comments that contain quantitative information on reference prices to be more impactful than comments that do not contain such information.

The effectiveness of information may also vary with the amount of tax revenue at stake; if inspectors are honest, we would expect them to be especially likely to make use of the additional information when potential revenue losses are high. However, opportunities for graft and coercion by importers are also likely to be correlated with tax burden, so if inspectors are susceptible to bribes or coercion, information may be least impactful when the stakes are high.

Risk information may also impact inspectors' actions and customs performance because it is associated with enhanced monitoring of inspectors. Better risk analysis makes it easier to audit declarations and identify opportunistic behavior. The customs administration monitors declarations subject to valuation advice aggressively, especially for the port of Toamasina, because those declarations are typically highly risky but also because the quality of the valuation advice makes it easy to detect deviant behavior. To separately identify the impact of information from the impact of additional scrutiny, the RCT contains a pure monitoring (e.g., zero information)

[18] Documents that have to be attached to the customs declaration (e.g., the invoice, the bill of lading) are generally easily falsifiable. Sellers located abroad want to maintain their commercial relationship with importers and often do not respond to inquiries from foreign customs authorities — especially those originating from low-income countries. Some are known to be willing to produce a "second invoice" to the buyer that is only intended to facilitate tax evasion.

treatment arm. We anticipate monitoring to lead to an increase in inspector effort and improved customs outcomes as it lowers the expected return to colluding with importers to lower the customs value.

3. Data

3.1. Data sources and representativeness

The analysis combines a number of unique data sets from different providers.

Administrative customs data

From Madagascar's customs administration, we obtained highly disaggregated administrative data tracking imports at the transaction level spanning the period January 2016-February 2019 (extracted through the customs management system (ASYCUDA)). For each import declaration, ASYCUDA data cover several standard variables: the port of entry, the HS 8-digit products included (designated as items) and for each item the import value (in Madagascar Ariary), taxes paid (tariff and value added tax as well as exemptions, all in Madagascar ariary) and weight (in kilograms), the country of origin (transit). The data also include more unique and novel variables such as the time stamps for date of registration, date of assessment, and other relevant dates in the customs clearance process (day/month/year/hour), the import value initially declared by the broker on the behalf of the importer, the inspection channel initially recommended by: (i) the risk management system; and (ii) by the RMU (documentary control/yellow channel, physical frontline inspection/red channel or no inspection/blue channel), and unique numeric identifiers for the broker, the importing firm and, importantly, the customs inspector assigned to handle the declaration, which allows us to study inspector performance.

Internal control and fraud data

We merge these ASYCUDA data with additional data sets from the customs administration's internal control systems. First, we combine these data with information on modifications made by the customs inspector during the customs clearance process to the inspection channel (upgrade in the channel the import declaration was initially allocated by the RMU, as discussed in Section 2), the import value and the import weight. Second, we obtained information on fraud records from the Legal Department (*Service des Affaires Juridiques et du Contentieux*), namely which

declarations were deemed fraudulent and the amount of taxes recovered (if any). These variables are unique and crucial for our definition of inspector actions that may potentially respond to the provision of information and the monitoring. Third, we obtained information from the RMU on the qualitative comments made to each import declaration. For the declarations lodged during the period of the RCT whereby improvements to the qualitative comments were the treatment, we use text analysis methods to extract several characteristics of those comments, including their length, the presence of prices, of weights, and the reference to prior transactions, as will be further described in Section 5 and in the Appendix.

Risk scores, valuation advices and scanning results

We obtained from the service provider GasyNet additional information that we merge to the ASYCUDA data on the risk score assigned to each declaration (related to the risk of non-compliance with customs regulations ranging from 1 to 9), the recommendation of an inspection channel (documentary control, physical frontline inspection, or no inspection), and whether a declaration was subject to a scanner exam. Most importantly, and as discussed in Section 2, the service provider also issues valuation advices. Since this service is costly given that it requires additional research, valuation advices are carried out for only a subset of declarations, about 6 percent of the total per year, with the selection made by the service provider and the customs administration based on risk criteria. To our knowledge, ours is the first study with access to this type of risk and third-party valuation information. Further discussion on these valuation advices is provided in Section 4.

Madagascar's raw customs data cover the universe of Madagascar's formal import transactions, that is, imports made under five regimes: final imports for consumption (imports for home use), re-imports, temporary admissions, inward processing, warehouse, and other. Our sample for the econometric analysis includes only import declarations that are subject to taxation and to a physical or documentary control by customs frontline inspectors. This implies focusing only on imports for home use (customs regime IM4) and excluding declarations from importers that are members of the "*Procédure Accélérée de Dédouanement*" (PAD), a trusted trader program that allows member firms to benefit from expedited clearance procedures with minimal controls at the border.[19] We

[19] Only a very small share of declarations registered by PAD members are selected for random documentary or physical controls.

also exclude from the sample declarations that list petroleum products (HS chapter 27) or vehicles, which are subject to different clearance procedures.[20] Finally, we exclude from the sample import declarations that arrive to airports and to the minor seaports in Madagascar and focus on those arriving to the main port Toamasina as well as the next seven largest ports. The final estimating sample used in Section 4 to study the impact of valuation advice accounts for 79.3 percent of collected taxes and 68.5 percent of the total import value for non-oil import declarations cleared by seaports, on average across the period ranging from January 1, 2016 until mid-October 2018. For the analysis of the RCT presented in Section 5, the sample includes only high-risk declarations, that is, those subject to valuation advice and with a risk score of 8 or 9, as will be further discussed in Section 5. The RCT sample accounts for 53.6 percent of collected taxes and 45.3 percent of the total import value of non-oil import declarations cleared by seaports, on average across the 14 weeks of the RCT from October 15, 2018 to February 2019.

3.2. Measuring customs performance

Our analysis uses the set of outcome and control variables described below.

Outcome variables

Our outcomes of interest are measures of customs performance and inspector actions: a dummy indicating whether or not the inspector "upgraded" the declaration by moving it to the red channel; a dummy indicating whether or not the containerized goods in the declaration were scanned; log clearance time measured as the number of hours from the time the declaration was registered to the time of assessment;[21] a dummy for whether or not fraud was recorded;[22] the log change in

[20] In Madagascar, used cars are subject to specific clearance procedures and are examined by the service provider GasyNet.

[21] While we use the expression "clearance time" to designate our outcome variable, note that the variable is actually capturing what could be called 'assessment time' which is the only time that is under the control of the inspector. The complete clearance time (time from registration to clearance) is not only attributable to customs but depends also on other agents such as importers and banks if payments need to be made for imported goods to be cleared and if time is taken to pick up the goods from the port. The use of a measure of time that is entirely under control of the customs agent is an improvement over the other studies in the literature, namely Volpe Martincus et al. (2015) and Fernandes et al. (2019).

[22] Since we do not have direct measures of fraud, we have to rely on fraud records which could reflect a combination of true fraud, extortion, and reporting error. Given the widespread prevalence of undervaluation, it seems a priori reasonable to assume reported fraud is a good proxy for actual fraud.

value (finally registered – initially declared); and the log change in taxes (finally registered – initially declared).

Control variables

Our specifications include a rich set of risk characteristics of the import declaration as control variables: a dummy for the declaration being initially assigned to the red channel, dummies for the risk score assigned to the declaration by the service provider GasyNet, the tax rate (tariffs and other taxes, in particular value-added taxes), a dummy for being a mixed shipment (that is, one that includes different HS 8-digit products), the shares of differentiated or time-sensitive products, and the initially declared weight.[23] In robustness tests, we also control for the workload of the inspector, defined as the number of import declarations he/she handled that month. Finally, in some specifications, we also control for measures of initial undervaluation. To construct such measures, we use two different ways to arrive at a hypothetical valuation of an import declaration. On the one hand, we calculate for each HS 8-digit product the median unit price declared (ratio of value to weight declared) across all importers.[24] Subsequently, we multiply these median unit prices by the weight of each HS 8-digit product imported to arrive at a hypothetical valuation of a particular declaration. The difference between this hypothetical valuation (in logarithms) and the initially declared import value (in logarithms) of a declaration is one of our measures of initial undervaluation. This measure only identifies undervaluation relative to other importers. On the other hand, for those declarations for which valuation advice is issued, we also compute undervaluation by simply subtracting the initially declared value (in logarithms) from the reference value issued in the valuation advice (in logarithms).

For outcome variables that are continuous and for undervaluation control variables, we trim the top and bottom 1 percent of observations to eliminate potential outliers. The exact definition of each of the variables used is provided in Appendix Table A1. Summary statistics for the outcome and control variables are provided for different sub-periods in Sections 4 and 5.[25] Given the

[23] In regressions whose outcome variable is upgrading of the declaration to the red channel, the control variable for the red channel is dropped as upgrading is not defined for such declarations. Regarding risk scores, the analysis of the returns to third-party valuation advice (Section 4) includes dummies for all risk scores from 2 to 9 (omitted is 1) while the analysis of the RCT (Section 5) includes only one dummy for risk score 9 as the estimating sample covers only declarations with risk score 8 or 9.

[24] The median is computed based on importers of the HS 8-digit product from a given country of origin.

[25] For the sub-period of the RCT and the subset of declarations subject to valuation advice, we are able to assess how aligned are the reported prices in the declaration to the various reference prices available: the price in the valuation

uniqueness of our measure of undervaluation relative to valuation advice, it is of interest to examine its distribution across import declarations in Madagascar. Figure 2 shows that undervaluation is rife. This evidence reinforces the relevance of our assessment of remedial measures to address undervaluation and tax evasion in this context.

4. The returns to valuation advice

4.1. Econometric strategy

To examine the impact of the provision of third-party valuation advice information on inspector actions and customs outcomes, the following regression is estimated:

$$\boldsymbol{Y_d = \beta_V Valuation\ Advice_d + \beta_X X_d + \mu_i + v_f + \pi_o + \varphi_{h2} + \tau_{pt} + \varepsilon_d} \quad \textbf{(1)}$$

where $\boldsymbol{Y_d}$ is one of the outcome variables for declaration *d* from importing firm *f* handled by inspector *i* in port *p* in month *t*. As discussed in Section 3, these outcome variables include upgrading, scanning, clearance time, detection of fraud, adjustments in import value, and adjustments in taxes. $\boldsymbol{Valuation\ Advice_d}$ is an indicator variable for declarations receiving valuation advice, $\boldsymbol{X_d}$ is the vector of characteristics of the declaration used as control variables (listed in Section 3) , $\boldsymbol{\mu_i}$ is a vector of inspector fixed effects, $\boldsymbol{v_f}$ is a vector of importer fixed effects, $\boldsymbol{\pi_o}$ is a vector of country of origin dummies, $\boldsymbol{\varphi_{h2}}$ is a vector of HS 2-digit product dummies, $\boldsymbol{\tau_{pt}}$ is a vector of port-month fixed effects, and $\boldsymbol{\varepsilon_d}$ is an error term. Standard errors are clustered by inspector. Q-values are calculated following Benjamini et al. (2006).

Valuation advice is not provided randomly by the service provider, but rather is typically assigned to declarations for which potential tax revenue losses are largest as is shown in Table 1, which presents descriptive statistics for declarations for which valuation advice was issued, as well as declarations for which it was not. Declarations for which valuation advice is issued have significantly higher risk scores, are subject to significantly higher taxes, are significantly more

advice, the reported prices by other importers, the price included in the extensive comments, and a price obtained based on mirror export data from COMTRADE. The sample of declarations for which all these prices are comparable is small as it requires them to include a single HS 6-digit product, receive valuation advice and extensive comments in the RCT and have mirror export price for that HS 6-digit product available. The correlations among all these prices is shown in Appendix Table 2 and are all positive and significant at the 95 percent confidence level. But most of the correlations are less than 0.5.

likely to be routed through the red channel, to be mixed shipments, and are more likely to be undervalued. For those declarations, average undervaluation relative to the median unit prices registered by other Madagascan importers is 18.2 percent, whereas for declarations not subject to valuation advice it is 7 percent.[26] However, this proxy for undervaluation is biased downwards because it only identifies undervaluation relative to other importers (which may also be underreporting). This may help explain why for declarations subject to valuation advice undervaluation relative to the reference value issued by the service provider is substantially higher, 37.5 percentage points on average.

To account for the differences shown in Table 1, Eq. (1) includes a very rich vector of declaration characteristics capturing inasmuch as possible all differences across declarations that could both affect their riskiness as well as their likelihood of receiving valuation advice. The fixed effects included in Eq. (1), in particular inspector and importer fixed effects, are very stringent, and account for inspectors' average ability and propensity to detect fraud, as well the average propensity of importers to commit it.

Still, OLS estimates of the coefficient $\boldsymbol{\beta_V}$ in Eq. (1) could still be biased. To address this endogeneity concern, we exploit the fact that the share of declarations for which valuation advice is provided varies over time and across ports, and, crucially, by their risk score. We instrument the provision of valuation advice with the share of declarations with the same risk score handled in the same port in the same month (of a given year) that were selected for valuation advice provision. This can be thought of as a measure of a particular declaration's propensity to be subjected to valuation advice.

The validity of our IV strategy relies on the assumption that the share of declarations receiving valuation advice in a port-month for a given risk category does not affect directly inspector actions and customs outcomes of a given declaration through any other channel than the probability of the specific declaration in question being subjected to valuation advice. This assumption could be violated, for instance, if the share of declarations subject to valuation advice were correlated with inspectors' workload; inspectors could simply be too busy to pay adequate attention to the

[26] Note that because we measure undervaluation relative to the median unit price computed across all importers (whether the declarations receive valuation advice or not), our proxy for undervaluation can be positive on average for both declarations subject to valuation advice as well as for declarations that are not.

information presented in the valuation advice. Alternatively, an increase in the share of declarations subject to valuation advice could be correlated with a change in importer behavior. These concerns are to a large extent mitigated by the fact that our instrument is specific to declarations with a particular risk score. Nonetheless, we will show in robustness tests that our results are maintained when we control explicitly for the inspector's workload. In addition, we will control for changes in importer behavior by conditioning on initial undervaluation.

4.2. Results

The provision of valuation advice strongly and significantly increases inspectors' scrutiny, the likelihood of fraud being detected, and it improves tax yield, but at the expense of protracted clearance times, as is shown in Table 2 which presents the main results. OLS estimates are presented in panel A and IV estimates in panel B.[27] According to our preferred IV estimates, declarations subject to valuation advice are 12.0 percentage points more likely to be upgraded but the effect is not statistically significant. They are also 13.2 percentage points more likely to be scanned and take more than twice as long to clear customs. To interpret these numbers, it is useful to recall that the unconditional probability of a declaration for which no valuation advice is issued being upgraded is 10 percent and being scanned is 12.7 percent. IV estimates are much lower than OLS estimates, which seem upward biased, presumably because valuation advice is issued for declarations which are on average riskier. Inspectors seem to spend more effort trying to uncover fraud when valuation advice is issued.[28]

The estimates demonstrate that valuation advice increases the likelihood of fraud being recorded by 21.7 percentage points according to our preferred IV estimates. This is a very large effect, especially when one considers that on average only 3.2 percent of all declarations not subject to valuation advice are deemed fraudulent. These large effects attest to the quality of the valuation advice but could also be due to the customs performance management unit's monitoring of how inspectors use valuation advice, and their use of the share of declarations deemed fraudulent as a

[27] First-stage estimates corresponding to the IV results in panel B show that the partial correlation between the selection for valuation advice and the exogenous regressor is positive and significant (see Appendix Table 3). In addition, the high F-statistics suggest our instruments are strong. The reason for having several first-stage regressions is the different number of observations included in the estimating samples of the different outcome variables in panel B of Table 2.

[28] Interestingly in unreported results, we show that longer clearance times are not solely driven by the fact that more fraud is being recorded. Clearance times also increase significantly for non-fraudulent declarations that receive valuation advice.

metric for assessing inspector performance. Valuation advice is also associated with a significant 4.9 percentage points increase in import value and a 5.2 percentage points increase in tax collected. On the one hand, these are large effects given that on average inspectors only increase the value of declarations that are not subject to valuation advice by 0.8 percentage point and taxes by 1 percentage point. On the other hand, inspectors are a long way from applying the recommended reference value, which, as discussed above, is on average 37 percent higher than the initially declared value for declarations subject to valuation advice. Note that OLS estimates again yield higher estimates of the impact of valuation advice on fraud and changes in value and taxes, presumably reflecting selection bias; recall that valuation advice is only issued for a very small share (approximately 5 percent) of observations for which suspected revenue losses are largest.

Several checks are conducted to assess the robustness of the impacts of valuation advice and are presented in Appendix Table 4. To start with, we re-estimate our preferred IV regressions adding a control for inspector workload, measured as the number of declarations she/he cleared that month (panel A). The coefficient on the workload measure is typically very close to zero and never statistically significant except when clearance time is the outcome; when inspector workload goes up, so do, intuitively, average clearance times. Importantly, however, the estimated impacts of valuation advice are hardly impacted. This robustness check helps to bolster the case for the adequacy of our instrument. We also replicate our main regressions controlling for, respectively, inspector-month fixed effects or inspector-broker fixed effects instead of plain inspector fixed effects (panels B and C). Again, this does not seem to impact our coefficient estimates. Last but not least, one may be concerned that a change in the number of declarations subjected to valuation advice may impact importer behavior. To guard against this possibility, we add a proxy for undervaluation to our IV regressions (panel D). The estimated impact of valuation advice is hardly impacted. Put differently, our results are very robust.

Finally, we also examine whether the provision of valuation advice has spillovers on how inspectors treat other declarations. In particular, one might worry that the large returns to valuation advice found in Table 2 come at the expense of inferior performance on declarations not subject to such advice. To address this concern, we focus exclusively on declarations not subject to valuation advice and regress each of the outcomes of interest on the share of declarations that a given inspector handled that were subjected to valuation advice, including all controls as in Eq.

(1). If there are spillovers within inspectors, the outcome variables should respond to the inspectors' own shares of declarations selected for valuation advice. However, the results, reported in Appendix Table 5, show that the share of declarations selected for valuation advice has no predictive power for any inspector action or customs performance measure. This suggests spillovers are limited and assuages potential concerns about results being driven by a reallocation of inspector effort across different types of declarations.

What are the returns to valuation advice? The average declaration in our sample (including declarations that are not taxed) yields 8,479 USD in taxes (before inspector assessment). Our estimates suggest that taxes increase by 5.2 percent in response to valuation advice, which corresponds to an increase in the amount of taxes paid of 441 USD per declaration. The costs of valuation advice are not fixed per declaration because they vary with how much research is required to formulate a recommendation (which in turn depends on which items are included in the declaration). The approximate cost per declaration is 250 USD. For our sample, we thus estimate a positive return to valuation advice of 191 USD per declaration, or approximately 76 percent. This number does not consider heterogeneity in the impact of valuation advice, which is likely to dampen aggregate returns, as we will show in the next section.

4.3. Heterogeneity

The results so far assess the average effects of valuation advice, but we show below that there is important heterogeneity in the effects, driven by inspectors' strategic behavior. We estimate an alternative set of specifications focusing on the sample of declarations with valuation advice, in which we condition on the extent of initial undervaluation relative to the valuation advice, $\boldsymbol{Undervaluation_d}$ and we examine whether the extent of undervaluation has differential effects depending on the declaration's potential tax revenue losses. Specifically, we define a high-stakes indicator $\boldsymbol{High\ Stakes_d}$ taking the value of 1 if hypothetical taxes exceed 10,000 USD (with hypothetical taxes being defined as the product between the declaration's tax rate and the value of the declaration as stipulated by the valuation advice). The specification to estimate is of the form:

$$\boldsymbol{Y_d = \beta_U Undervaluation_d + \beta_{UH} Undervaluation_d * High\ Stakes_d}$$
$$\boldsymbol{+\beta_H High\ Stakes_d + \beta_X X_d + \mu_i + v_f + \pi_o + \varphi_{h2} + \tau_{pt} + \varepsilon_d} \qquad (\mathbf{2})$$

where all variables are defined as above. The coefficient $\boldsymbol{\beta_{UH}}$ can be interpreted as measuring the differential effectiveness of valuation advice for high-stakes declarations.

Before turning to regression estimates, Figure 3 plots the extent of import value adjustment made by inspectors against initial undervaluation separately for both low-stakes and high-stakes declarations with valuation advice. Three facts stand out. First, the import value adjustment is significantly lower than the theoretically optimal adjustment which is depicted using a solid 45-degree line. If inspectors were following the valuation advice, upward adjustments to value should increase one-to-one with initial undervaluation but that is not the case at all. For example, for low-stakes declarations for which half of the initial value is declared, the observed upward valuation by inspectors is approximately 22 percent, i.e., only half of what would be expected on the basis of the valuation advice (the dotted line). Second, this pattern is particularly acute for high-stakes declarations (the solid line), for which the corresponding observed upward valuation is much lower than for low-stakes declarations for all observed levels of initial undervaluation. For instance, declarations for which the initially declared value is only half of the reference value are adjusted upwards by 10 percent on average. This is perturbing because, a priori, we would expect honest inspectors to be especially vigilant when potential tax yield is high. Third, and even more worryingly, the discrepancy between low- and high-stakes declarations is increasing with undervaluation. If anything, for highly undervalued declarations (at the right of the graph) we find that the extent of upward value adjustment *decreases* with the extent of initial undervaluation.

The OLS estimates of Eq. (2) presented in Table 3 confirm that highly undervalued high-stakes declarations are cleared *faster*, *less* likely to be deemed fraudulent and yield *less* tax adjustment, than lower-stakes declarations that are similarly undervalued. This finding attests to opportunistic behavior by inspectors, that act on the information they receive in the valuation advice but less so for declarations for which the returns to colluding with economic operators are higher.

In summary, we find sizeable returns to the provision of transaction-specific valuation advice, but we also show that inspectors are a long way from applying the recommended reference value. The fact that they are recording a lot of fraud but that the tax adjustments they are making are small relative to the theoretically optimal adjustment suggests they are engaging in box-ticking behavior. More worryingly, the impacts of valuation advice are muted for declarations for which potential

revenue losses (and opportunities for graft) are largest. Valuation advice could presumably be even more impactful if inspectors had stronger incentives not to act opportunistically and resist bribes.

5. Experimental evidence on the impact of information and monitoring

5.1. Design

To separately identify the role of information provision and that of monitoring on inspector actions and customs outcomes, an RCT was implemented in Madagascar's customs department with the full support of a project implementation unit established in the department. The RCT also sheds light on the relative returns to improving customs agencies' in-house risk analysis capacity versus those to hiring an external service provider. Whereas valuation advice is costly, the RCT mostly entailed internal restructuring and required only limited additional human and material resources. As such, the RCT provides evidence on the ability of customs administrations in low-income countries to conduct in-house risk analysis.

The RCT focused on high-risk declarations defined as those subject to valuation advice or those with risk scores of 8 or 9 arriving to one of Madagascar's eight main sea ports listed in Section 2. It contained two treatments which were implemented in staggered but mostly overlapping fashion.

Detailed risk-analysis comments (for brevity referred to as detailed comments in what follows): Starting on October 15, 2018, agents in the RMU were instructed to provide very detailed risk-analysis information for at least 18 of 24 randomly selected high-risk declarations tagged for detailed comments each day. Some latitude in the number of declarations to be commented on was offered to accommodate variability in the supply of high-risk declarations over time within a given day, and to allow for the possibility that the risk scores were inappropriate, i.e., some declarations that were flagged as high-risk might in fact be very compliant. The agents in the RMU were divided into three teams – each headed by a different manager – and each team was allocated 8 high-risk declarations per day from Monday to Friday – with the 8 declarations being sent in 4 different batches of two high-risk declarations each, at 9:30.a.m, 11:30 a.m., 2:30 p.m., and 5:00 p.m. each day. Declarations registered during the weekend were saved in a batch to be used for random allocation the following Monday. The random selection of the 24 declarations each day that would be receiving detailed comments was implemented by service provider GasyNet in real-time.

Agents in the RMU were instructed to provide detailed comments that helped with risk analysis and that ideally would include: (i) explicit references to similar prior transactions (when available); (ii) advice on what the relevant unit price should be for the goods in the declaration (based on those prior transactions and the RMU's own analysis); and (iii) explicit instructions for inspectors as to whether they should perform specific actions such as scanning, checking documents etc. They were told that at the end of the RCT, the performance of the RMU itself would be evaluated based on how useful their detailed comments had been.

Monitoring: Customs inspectors were informed that, in addition to the detailed comments they would be receiving on a selected number of declarations, they would be monitored more intensively for a period of approximately 12 weeks starting on November 5, 2018. They were informed that auditors would examine how well they did their job, and were reminded that they should follow procedures by using the appropriate inspection tools at their disposal, i.e., make sure that goods assigned to the red channel are scanned and if not, that the port manager should sign off on their formal justification for not having completed a scan; that photos are to be taken during physical inspections; that customs forms are filled in properly so as to enable ex-post audits, etc. This general announcement of intensified monitoring was communicated by means of a written communique from the director of customs, as well as a site visit by staff in the project implementation unit in charge of implementing the RCT.

Moreover, each week approximately 20-25 high-risk declarations were randomly selected for ex-post monitoring by the project implementation unit. Inspectors were informed that the declarations would be evaluated ex-post, by means of a message in the customs system (see Figure 4) and we designate those as declarations explicitly tagged for monitoring. To limit predictability of this treatment, inspectors were divided into two groups of roughly equal size and each week, a different group was considered eligible to receive declarations explicitly tagged for monitoring. The likelihood of receiving such a monitoring tag message varied widely across ports, because of differences in import volumes. To ensure that in smaller ports, inspectors would receive at least a few declarations explicitly tagged for monitoring, selection criteria favored selecting declarations from smaller ports. In the main port of Toamasina, each inspector would receive roughly 2 declarations explicitly tagged for monitoring per week in the weeks in which he/she received such messages, whereas in other ports, roughly 3 declarations were explicitly tagged for monitoring

each week per inspector. Inspectors were not informed that they were assigned to different groups (receiving explicitly tagged for monitoring declarations in different weeks), nor that only high-risk declarations were selected for ex-post monitoring. An important feature to note about this monitoring treatment is that it was not accompanied by the explicit announcement of sanctions or penalties for poor or inadequate inspector actions.

5.2. Implementation

Table 4 shows the composition of the sample resulting from the RCT. For both the detailed comments and the monitoring treatments, the treatment unit is the import declaration. Because the impact of detailed comments and monitoring is likely to be very different for declarations for which inspectors have not already benefitted from information on adequate valuation, we separately analyze: (i) the sub-sample of declarations that are not subject to the valuation advice studied in Section 4; and (ii) the sub-sample of declarations that are subject to valuation advice. In terms of numbers, we have 992 declarations in sub-sample (i) receiving detailed comments and 196 declarations explicitly tagged for monitoring over the RCT period and a total of 4,733 control declarations. And in sub-sample (ii), we have 123 declarations receiving detailed comments and 19 declarations explicitly tagged for monitoring over the RCT period and a total of 644 control declarations. As the sample sizes suggest, a caveat to note is that we have limited power for estimating the impacts of the treatments in the sub-sample of declarations subject to valuation advice.

Table 4 also shows the characteristics of the high-risk declarations in the control group and how those differ for declarations in the treatment groups, of detailed comments or monitoring, separately for declarations not subject and subject to valuation advice.[29] In the sub-sample not subject to valuation advice examined in columns (1)-(5), declarations selected for detailed comments do not have on average higher risk scores than declarations in the control group. By contrast, declarations selected for monitoring have a 7-percentage point higher probability of having the highest risk score than declarations in the control group, 20.6 percent of whom have the highest risk score, and the difference is statistically significant as indicated by the p-value in column (5). Put differently, randomization in terms of risk was fully successful for the detailed

[29] Appendix Table 6 provides balance tests for the pooled sample of all declarations.

comments treatment but unsuccessful for the monitoring treatment, perhaps in part because so far fewer declarations were selected.

We do find significant differences across treatment and control groups for a number of other characteristics that were not targeted in our randomization, but they are not systematically related to the risk of tariff evasion. Indeed, some of these differences, such as declarations subject to detailed comments being more undervalued and subject to higher tax rates, suggest higher evasion risk although the magnitude of the differences is small. But others, such as declarations subject to detailed comments containing significantly less time-sensitive goods and being less likely to be routed through the red channel, point in the opposite direction. Similarly, the fact that declarations selected for monitoring are significantly more likely to be routed through the red channel and are significantly more undervalued relative to control declarations suggest that they are more likely to be fraudulent. Yet, they contain significantly fewer differentiated and time-sensitive products, which are widely used proxies for evasion risk.

For the sub-sample subject to valuation advice examined in columns (6)-(10), the characteristics of declarations receiving detailed comments and control declarations are not significantly different on average, with the exception of the probability of being allocated to the red channel which is 9.1-percentage points lower for treated declarations. Declarations with valuation advice selected for monitoring differ along a few characteristics from control declarations; they are 22.9-percentage points more likely to be allocated to the red channel, which could indicate their higher risk of evasion. But they also contain a lower share of differentiated products which would point in the opposite direction.

All in all, the imbalances across treated and control samples that we identify in Table 4 do not conclusively point to systematically higher or lower evasion risk among declarations selected for detailed comments or monitoring. Yet they do attest to the crucial importance of including a comprehensive set of declaration and risk characteristics as controls in our regressions.

The analysis proceeds in two steps. First, we assess whether the RCT improved the quality of the comments provided by the RMU. Second, we assess to what extent the detailed comments and the announcement of monitoring impacted inspector actions and customs outcomes.

5.3. Did the comments become more informative?

To assess the impact of the RCT on the quality of the comments provided we estimate the following specification:

$$\boldsymbol{Y_d = \gamma_C C_d + \beta_X X_d + \pi_o + \varphi_{h2} + v_f + \tau_{pt} + e_d} \quad \textbf{(3)}$$

where the four outcome variables $\boldsymbol{Y_d}$ for each declaration $\boldsymbol{d}$ are dummies (i) for the length of the comment message measured as number of characters, (ii) for whether a price was explicitly referred to (captured by a currency detected next to numerical characters) or (iii) for whether a prior declaration was explicitly referred to, and (iv) for whether a weight was mentioned (captured by a unit of weight detected next to numerical characters) and C_d is a dummy variable that indicates whether the declaration was randomly selected for detailed comments.[30] The specification controls for the same rich vector of risk characteristics $\boldsymbol{X_d}$ as in Section 4 as well as for country of origin dummies ($\boldsymbol{\pi_o}$), HS 2-digit product dummies ($\boldsymbol{\varphi_{h2}}$), importer fixed effects ($\boldsymbol{v_f}$) and port-month fixed effects ($\boldsymbol{\tau_{pt}}$).

With a goal of assessing whether the comments are pertinent (e.g., is a price more likely to be mentioned when there is undervaluation etc.), we also estimate a variant of the specification above in which we interact the dummy for being selected for detailed comments with a proxy for evasion risk, notably the tax rate, as shown below:

$$\boldsymbol{Y_d = \gamma_C C_d + \gamma_{Ct} C_d T_d + \beta_X X_d + \pi_o + \varphi_{h2} + v_f + \tau_{pt} + e_d} \quad \textbf{(4)}$$

where $\boldsymbol{T_d}$ is the tax rate and all other variables are defined as above.

To set the scene for the econometric estimation that follows, Figure 5 plots the evolution of the length of comments and the number of times a price, weight, or prior transaction was explicitly referenced. The beginning of the RCT on October 15, 2018, is associated with a dramatic increase in the length of comments for declarations selected for detailed comments; with the length of comments more than doubling relative to the baseline. In addition, there is a clear increase in the share of declarations that mention a price and, to a lesser extent, a weight. Interestingly, the graphs also suggest that the RCT had some positive spillovers on comments for non-treated declarations.

[30] Appendix A describes in further detail how the outcome variables based on the comments are constructed.

This does not necessarily represent a problem; it could simply make it more difficult to estimate a significant impact of detailed comments on inspector actions and customs outcomes in Section 5.4.

The estimates presented in panel A of Table 5 confirm that declarations that were selected for detailed comments received on average comments that were 213 characters longer than those in the control group, for which the average length of comments is 169 characters. The treated declarations were, *ceteris paribus*, 44.3 percentage points more likely to include price information, which is a sizeable increase given that only 24.8 percent of comments on declarations in the control group mention a price of any kind. In addition, detailed comments are 10.8 percentage points more likely to refer to a prior transaction, and 6.3 percentage points more likely to refer to a weight measure. Moreover, the improvement in the quality of the comments was especially pronounced for declarations subject to higher taxes – as is shown in panel B of Table 5. This suggests that the comments became more relevant, as the RMU did a good job in providing them for the riskier declarations.

Discussions with staff from both customs and GasyNet indicated that the improvement in the RMU's performance was the most dramatic institutional change they had witnessed. Recall that, prior to the start of the experiment, the RMU was considered to be dysfunctional. At the launch of the experiment, the head of the RMU decided to divide the unit's work into three teams: one inspector and three analysts per team to increase internal accountability and optimize workflow. A standardized reporting framework was adopted for all teams. In less than two weeks, the RMU's work dramatically changed. The new organizational structure incentivized the inspectors who were appointed head of their respective teams to share their knowledge with newly recruited analysts on potential fraud cases and give them advice on what to look for, e.g., historical cases of fraud for some economic operators, abnormal price valuation for some imports. Moreover, they created a database of historical fraud cases and improved documentation of their risk analysis to facilitate internal knowledge transfer. Weekly meetings were organized with all senior managers in customs and head of unit/inspectors from the RMU to discuss contentious cases in terms of goods valuation and decisions were collectively taken on how to handle such cases in order to reduce discrepancies in customs procedures across ports and inspectors. At the end of the RCT, the work organization of the RMU instituted at the start of the RCT was maintained. The transformation of the RMU

staff's behavior was due in part to them fearing they might lose their jobs if they did not perform well. In addition, staff from the RMU claimed that they felt empowered by the RCT because it made the RMU's recommendations more salient to both inspectors and other arms of the customs administration.

5.4. Did information and monitoring impact inspector actions and customs outcomes?

To assess the impact of detailed comments as well as monitoring and the threat of ex-post audits on inspector actions and customs outcomes, we estimate (variants of) the following specification:

$$\boldsymbol{Y_d = \pi_C C_d + \pi_M M_d + \pi_X X_d + \mu_i + v_f + \pi_o + \varphi_{h2} + \tau_{pt} + u_d} \qquad \textbf{(5)}$$

where $\boldsymbol{Y_d}$ is again one of the outcome variables including upgrading, scanning, clearance time, detection of fraud, adjustments in import value, and adjustments in taxes, $\boldsymbol{C_d}$ is a dummy that indicates whether a particular declaration was selected for detailed comments, and $\boldsymbol{M_d}$ is a dummy that indicates a declaration was subject to ex-post monitoring. The specification controls for the vector of characteristics of the declaration used as control variables listed in Section 3 ($\boldsymbol{X_d}$) and for a very stringent set of fixed effects, namely for inspector ($\boldsymbol{\mu_i}$), importer ($\boldsymbol{v_f}$) as well as country of origin, HS 2-digit product, and port-month ($\boldsymbol{o_d}, \boldsymbol{p_d}, \boldsymbol{\tau_{pt}}$) and $\boldsymbol{\varepsilon_d}$ is an error term. Standard errors are clustered by inspector. Q-values are again calculated using the method proposed by Benjamini et al. (2006). We present results from estimating Eq. (5) separately for the sub-sample of declarations for which valuation advice was not issued and then for the smaller sub-sample of declarations for which valuation advice was issued. At the risk of belaboring the point, the reason for considering estimates across these two separate sub-samples is to enable us to uncover a role of detailed comments for declarations that have not already received information on adequate valuation.

Our main findings, presented in Table 6, are that detailed comments lead to significant changes in inspectors' actions and customs outcomes for declarations with no valuation advice (panel A). Specifically, detailed comments increase significantly the probabilities of both upgrading declarations to the more intrusive red channel - by 11.7 percentage points relative to the control group - as well as scanning the containerized goods - by 10.6 percentage points relative to the control group. Not surprisingly, detailed comments prolong clearance times by 27.4 percent on average. Inspectors are significantly more likely to report fraud in declarations receiving detailed

comments.[31] The magnitude of this increase is very large: the estimated 3.1 percentage point increase in fraud detection implies a doubling of fraud detection relative to average for control declarations of 3.2 percent. Detailed comments also lead to upward adjustments in import value by 0.8 percent and in taxes by 1 percent, *ceteris paribus*. While in absolute terms these improvements may appear small, they are meaningful when compared to the average changes in import value and taxes for control declarations of 0.5 percent and 0.7 percent, respectively. Overall, the impact of detailed comments is significant in reducing tax evasion for declarations receiving no other valuation advice.

Panel A of Table 6 also shows that explicit monitoring tags do not significantly impact most inspector actions and customs outcomes for declarations not subject to valuation advice. An exception to this pattern is found in column (2) where monitoring is estimated to increase the propensity to scan the containerized goods in the declaration, relative to the control group. Hence, just monitoring inspectors ex-post but not providing them with additional information has weak to no effects leading inspectors to only change easy-to-verify actions, again evidence of box-ticking behavior as in Section 4. This is consistent with the (perception of a) lack of punishment for improper conduct documented in Section 2: the absence of explicitly stated sanctions may have limited the deterrence effect of monitoring. But it may also be partly due to the timing of the RCT, which was implemented during Madagascar's general elections, which led to the subsequent removal of the Director General of customs and significant turnover among senior customs management, and on the heels of protracted strikes, all of which undermined the credibility of potential punitive actions.

While detailed comments appear to help inspectors detect fraud and adjust import value and taxes upward among declarations for which valuation advice was not issued, they do not significantly improve customs outcomes for declarations for which valuation advice is issued, as is shown in panel B of Table 6. This lack of impact may reflect the high quality of the information already contained in (the typically extensively documented valuation) advice. Even though the sign of the coefficients suggests they facilitate fraud detection, neither detailed comments nor monitoring has a significant impact on customs outcomes for declarations with valuation advice. The only

[31] In unreported results, we show that longer clearance times are not solely driven by the fact that more fraud is being recorded. Clearance times also increase significantly for non-fraudulent declarations that receive detailed comments (for the subsample not receiving valuation advice).

significant impact is seen in a higher likelihood of scanning for declarations subject to detailed comments, but the impact is significant only at the 90 percent confidence level (column (8)). This finding may be due to the treatment impacts being imprecisely estimated given the limited number of declarations selected for valuation advice.[32]

Robustness tests of the impacts of the detailed comments and the monitoring treatments are presented in Appendix Tables 7-9. Appendix Table 7 addresses the possibility that inspector behavior and customs outcomes may be affected by the degree of initial undervaluation of the declaration rather than by the RCT treatments by adding a proxy for undervaluation to Eq. (5). Appendix Tables 8 and 9 re-estimate Eq. (5) controlling, respectively, for inspector-month or for inspector-broker fixed effects instead of plain inspector fixed effects. All tables show mostly unchanged results, relative to those in Table 6.

In addition, we examine whether the RCT incited the customs inspectors to reduce the effort paid to their non-treated declarations, that is, whether there is evidence of within-inspector spillovers. First, Appendix Table 10 addresses the impact of the intensity of the valuation advice, detailed comments, and monitoring on the attention inspectors pay to and on the customs outcomes of the control declarations during the RCT. Specifically, we re-estimate a variant of Eq. (5) where the main regressor is the share for each inspector of its declarations that receive valuation advice, detailed comments, or are tagged for monitoring and the sample includes only declarations with none of those. The estimates in Appendix Table 10 show that inspector actions and customs outcomes for control declarations do not respond to the share of treated or valuation advice declarations among the workload of an inspector. As in Section 4, this evidence suggests that our findings are not driven by a reallocation of inspector effort towards these highly scrutinized declarations. Second, we look into the possibility that the onset of the RCT itself changed inspector behavior. Appendix Table 11 addresses the possibility of a break in the outcome variables for the declarations potentially eligible for treatment around Monday, October 15, 2018, the announced starting date of the RCT. It shows the results of a local linear polynomial (panel A) and a local quadratic polynomial (panel B). The bandwidth was selected following Cattaneo et al. (2015). The optimal bandwidth suggested includes about two weeks of controls. No significant breaks appear

[32] The attendant lack of power may also explain why we cannot reject the null hypothesis that the impact of either detailed comments or monitoring across the sub-sample of declarations with no valuation advice versus the sub-sample of declarations with valuation advice (see the bottom of Table 6).

except for clearance time for clearance in panel B (but only at a 90 percent confidence level), suggesting that the inspectors did not react strongly to the beginning of the RCT.[33]

Our evidence in Section 4 showed an important role of high-quality information, in the form of the service provider's valuation advice, leading inspectors to catch more fraud and reduce tax evasion (albeit with some opportunism) while our evidence in Section 5.3 shows that the RCT significantly improved the quality of the comments provided. Yet, there is heterogeneity in terms of the quality of the detailed comments provided through the RCT. We assess whether inspector actions and customs outcomes respond differently to information of higher quality and more precision using the specification below:

$$\boldsymbol{Y_d = \pi_{CH} C_H_d + \pi_{CO} C_O_d + \pi_M M_d + \pi_X X_d + \mu_i + v_f + o_d + p_d + \tau_{pt} + u_d} \quad \textbf{(6)}$$

where $\boldsymbol{C_H_d}$ is a dummy that indicates whether a particular declaration was selected for detailed comments with exact references, that is, those including a price, a weight, or explicitly referring to a prior transaction and $\boldsymbol{C_O_d}$ is a dummy for other declarations receiving less precise comments. All other variables in Eq. (6) are defined as above.

The OLS estimates for Eq. (6) are reported in Table 7. For the sub-sample of declarations not receiving valuation advice, only detailed comments with exact references induce inspectors to significantly increase fraud detection and make upward adjustments to import value and taxes. Those customs outcomes are not improved for declarations receiving less precise comments, relative to the control group. The estimated significant increases in the likelihood of upgrading or scanning declarations and in clearance times are of same magnitude regardless of the type of comments provided, as indicated by the p-values at the bottom of the table.[34] For the sub-sample of declarations receiving valuation advice, detailed comments with exact references play a significant role in increasing inspectors' probability of upgrading the clearance channel or scanning, relative to the control group. Put simply, better information seems to yield higher returns.

[33] This is confirmed by Appendix Figure 1 that depicts the binned averages of the outcome variables over the same period of time.

[34] There is a potential endogeneity problem in that comments with exact references may precisely be those more likely to be fraudulent and more undervalued. In unreported results we used an IV estimator where an instrument for comments with exact references was constructed using information on which team in the RMU provided the comments. The instrument was the share of comments with exact references provided by each of the three teams in each month which differs importantly across the teams. The IV estimates are qualitatively similar to the OLS estimates shown in Table 7.

We end by investigating the prevalence of opportunistic behavior by inspectors during the RCT period focusing on declarations receiving detailed comments but no valuation advice. Analogous to the analysis presented in Section 4.3, we examine whether inspector actions and customs outcomes vary, depending on the amount of tax revenue at stake, controlling for initial undervaluation.[35] The initial undervaluation is measured relative to the internal reference price and a new high-stakes indicator is equal to one if the hypothetical taxes exceed 10,000 USD (with hypothetical taxes being defined as the product between the declaration's tax rate and the hypothetical value of the declaration calculated using the internal reference unit price). We find evidence of differential treatment by inspectors of high-stakes highly undervalued declarations in that they detect less fraud and adjust significantly less their value and taxes, but the impact on fraud is not statistically significant. Not surprisingly, these results are generally weak due to the small sample size and the effects are not robust to adjustments for multiple hypothesis testing. But overall, this evidence is consistent with detection of fraud combined with a lower than theoretically optimal adjustment, especially for declarations for which opportunities for graft are largest.

In summary, we find a significant response by inspectors to the provision of detailed comments. The fact that inspectors increase substantially their recording of fraud but that the tax adjustments are small and that they are smaller for high-stakes declarations suggests that inspectors are engaging in box-ticking behavior.

6. Conclusion

The performance of customs inspectors is an important determinant of tax collection in developing countries but may be constrained by asymmetric information and inadequate monitoring of their behavior in conjunction with limited performance incentives. This paper has used uniquely rich administrative data and a randomized control trial to examine the impact of the provision of information to inspectors and of monitoring them on fraud detection, tax revenue, and clearance times.

Using an instrumental variables strategy exploiting variation in the supply by an external third party of declaration-specific valuation advice, we have shown that such advice increases scrutiny.

[35] This analysis is not conducted for the sub-sample of declarations with valuation advice during the RCT due to the very small number of observations (fewer than 100) with valuation advice and detailed comments.

It makes inspectors more likely to scan declarations, protracts clearance times, and improves fraud detection. Issuing valuation advice increases tax collection by 5.2 percent on average. This is a sizeable amount but much less than the theoretically optimal adjustment from a tax revenue mobilization perspective, given that importers declare values that are, on average, 37 percent lower than the valuation advice reference value. Moreover, the impacts of valuation advice are muted for declarations for which potential tax revenue losses and kickbacks are highest, consistent with opportunistic behavior of inspectors.

We also report the results of a nationwide RCT in the customs administration in which a subset of high-risk declarations were randomly selected to receive detailed risk-analysis comments from the RMU, and another subset received explicit tags informing inspectors that such declarations would be monitored. The experiment induced a dramatic change in the quality of the comments provided by the RMU. The length of the comments, the inclusion of price information and the references to prior transactions increased significantly in response to the RCT, and especially so for high-risk declarations.

The detailed comments induced inspectors to upgrade the declarations' inspection channel to red, to scan their contents, and spend a longer time assessing them. Among declarations not subject to valuation advice[36] but randomly selected to receive detailed risk-analysis comments, the probability of fraud being recorded increased by 3.1 percentage points, and tax yield increased by a percentage point on average. The reaction of inspectors to detailed comments is even stronger when these comments include a reference value and/or explicitly refer to prior transactions. More precise information thus seems to yield higher returns. But the adjustments in value and taxes are smaller for high-stakes declarations which present greater opportunities for bribery, again suggesting opportunistic behavior of inspectors.

Informing inspectors that the declaration they were handling was being monitored and would be evaluated *ex-post* did not impact fraud detection and revenue collection. The monitoring treatment did make inspectors significantly more likely to subject the declaration in question to a scan. The lack of revenue responsiveness to increased monitoring is surprising but consistent with a lack of

[36] Similar, but statistically insignificant impacts are documented for declarations subject to valuation advice. The lack of significance is likely in part driven by a lack of power.

strong performance rewards, virtual absence of punishment for improper conduct and large monetary gains associated with corruption.

Taken together, our results suggest that better risk-analysis information can significantly enhance customs performance by facilitating the determination of tax liabilities and the detection of fraudulent behavior. Yet, its effectiveness appears compromised by corruption and inadequate incentives, which help explain why inspectors do not optimally exploit the information provided to them, and why information provision is least successful in remedying fraud for declarations presenting the greatest opportunities for graft.

7. References

Anson, Jose, Olivier Cadot, and Marcelo Olarreaga. 2006. Tariff evasion and customs corruption: Does pre-shipment inspection help?. *Contribution to Economic Analysis and Policy* 5, (1): 1600–1623.

Baunsgaard, Thomas, and Michael Keen. 2010. Tax revenue and (or?) trade liberalization. *Journal of Public Economics* 94, (9-10): 563–577.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. 2006. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93, (3): 491–507.

Bhagwati, Jagdish. 1964. On the Underinvoicing of Imports. *Oxford Bulletin of Economics and Statistics* 26, (4): 389–397.

Bhagwati, Jagdish. 1967. Fiscal Policies, The Faking of Foreign Trade Declarations, and The Balance of Payments. *Bulletin of the Oxford University Institute of Economics and Statistics* 29, (1): 61–77.

Cantens, Thomas, Gaël Raballand, and Samson Bilangna. Reforming Customs by measuring performance: a Cameroon case study. *World Customs Journal* 4, (2): 55-74.

Cattaneo, Matias D., Brigham R. Frandsen, and Rocío Titiunik. 2015. Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate. *Journal of Causal Inference* 3, (1): 1–24.

Chalendard, Cyril, Gaël Raballand, and Antsa Rakotoarisoa. 2019. The use of detailed statistical data in customs reforms: The case of Madagascar. *Development Policy Review* 37, (4): 546–563.

Dincecco, Mark, and Nico Ravanilla. 2017. *The Importance of Bureaucrats in a Weak State: Evidence from the Philippines*. SSRN 2773884.

Dutt, Pushan, and Daniel A. Traca. 2010. Corruption and Bilateral Trade Flows: Extortion or Evasion? *Review of Economics and Statistics* 92, (4): 843–860.

Fernandes, Ana Margarida, Russell Hillberry, and Alejandra Mendoza Alcántara. 2019. Trade Effects of Customs Reform: Evidence from Albania. *The World Bank Economic Review*.

Finan, Frederico, Benjamin A. Olken, and Rohini Pande. 2015. *The Personnel Economics of the State* (No. w21825). National Bureau of Economic Research.

Fisman, Raymond, and Shang-Jin Wei. 2004. Tax Rates and Tax Evasion: Evidence from 'Missing Imports' in China. *Journal of Political Economy* 112, (2): 471–496.

Fisman, Raymond, and Shang-Jin Wei. 2009. The smuggling of art, and the art of smuggling: Uncovering the illicit trade in cultural property and antiques. *American Economic Journal: Applied Economics*, *1*(3), 82-96.

Godin, Mattéo, and Jean Hindriks. 2015. A Review of Critical Issues on Tax Design and Tax Administration in a Global Economy and Developing Countries. *Core Discussion Paper 2015/28*. Louvain: UCL.

Hummels, David L., and Georg Schaur. 2013. Time as a Trade Barrier. *American Economic Review* 103, (7): 2935–2959.

Javorcik, Beata S., and Gaia Narciso. 2008. Differentiated products and evasion of import tariffs. *Journal of International Economics* 76, (2): 208–222.

Javorcik, Beata S., and Gaia Narciso. 2017. WTO accession and tariff evasion. *Journal of Development Economics* 125, 59-69.

Jean, Sébastien, Cristina Mitaritonna, and Antoine Vatan. 2018. *Institutions and Customs Duty Evasion*. Working Paper 2018-24. CEPII.

Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken. 2016. Tax farming redux: Experimental evidence on performance pay for tax collectors. *The Quarterly Journal of Economics* 131, (1): 219–271.

Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken. 2019. Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings. *American Economic Review* 109, (1): 237–270.

Kleven, Henrik J., Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica* 79, (3): 651–692.

Kleven, Henrik J., Claus Thustrup Kreiner, and Emmanuel Saez. 2016. Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries. *Economica* 83, 219–246.

Kumler, Todd, Eric Verhoogen, and Judith A. Frías (2013). *Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico* (No. w19385). National Bureau of Economic Research.

Laajaj, Rachid, Marcela Eslava, and Tidiane Kinda. 2019. *The Costs of Bureaucracy and Corruption at Customs: Evidence from the Computerization of Imports in Colombia*. Documentos CEDE 017173. Universidad de los Andes - CEDE.

Mishra, Prachi, Arvind Subramanian, and Petia Topalova. 2008. Tariffs, enforcement, and customs evasion: Evidence from India. *Journal of Public Economics* 92, (10): 1907–1925.

Naritomi, Joanna. 2019. Consumers as tax auditors. *American Economic Review 109*, (9): 3031–3072.

Olken, Benjamin, and Rohini Pande. 2012. Corruption in Developing Countries. *Annual Review of Economics* 4, (1): 479–509.

Pepinsky, Thomas B., Jan H. Pierskalla, and Audrey Sacks. 2017. Bureaucracy and Service Delivery. *Annual Review of Political Science* 20, 249–268.

Pomeranz, Dina. 2015. No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review* 105, (8): 2539–2569.

Rauch, James E. 1999. Networks versus markets in international trade. *Journal of International Economics* 48, (1): 7–35.

Rijkers, Bob, Leila Baghdadi, and Gael Raballand. 2017. Political Connections and Tariff Evasion: Evidence from Tunisia. *The World Bank Economic Review* 31, (2): 459–482.

Sequeira, Sandra, and Simeon Djankov. 2014. Corruption and firm behavior: Evidence from African ports. *Journal of International Economics* 94, (2): 277–294.

Sequeira, Sandra. 2016. Corruption, trade costs, and gains from tariff liberalization: Evidence from Southern Africa. *American Economic Review* 106, (10): 3029–3063.

Slemrod, Joel. 2019. Tax compliance and enforcement. *Journal of Economic Literature*, *57*(4), 904-54.

Volpe Martincus, Christian, Jerónimo Carballo, and Alejandro Graziano. 2015. Customs. *Journal of International Economics* 96, (1): 119–137.

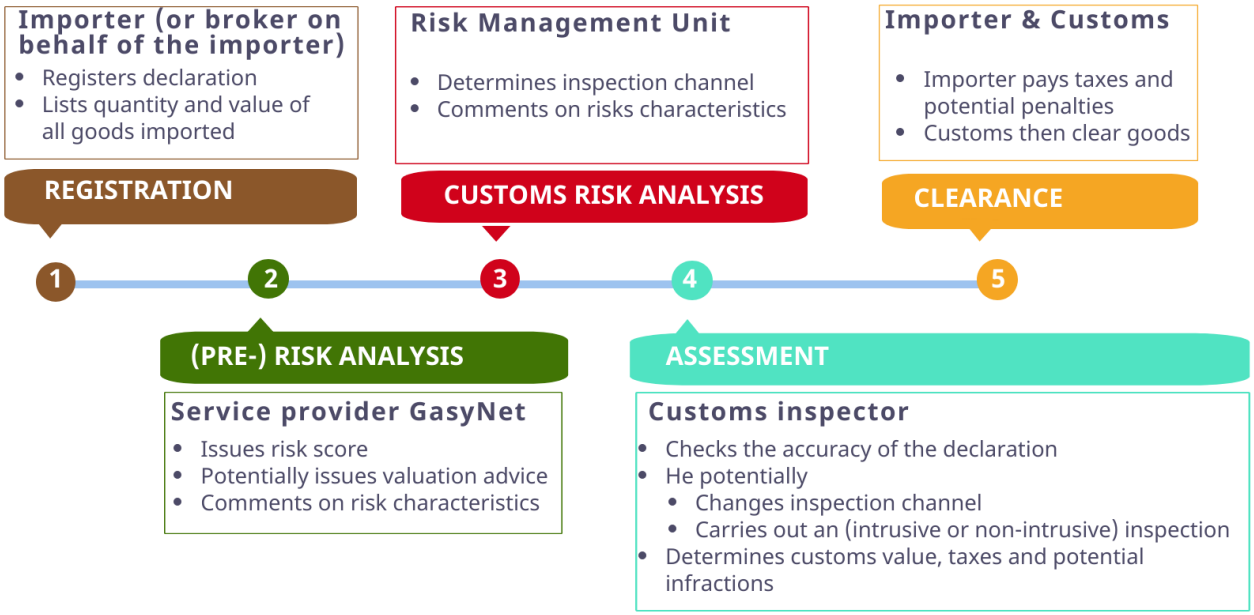
Wier, Ludvig. 2020. Tax-motivated transfer mispricing in South Africa: Direct evidence using transaction data. *Journal of Public Economics* 184, 104153.

WTO. 2015. Trade Policy Review: Madagascar. Geneva, Switzerland: World Trade Organization (WTO).

Xu, Guo. 2018. The Costs of Patronage: Evidence from the British Empire. *American Economic Review* 108, (11): 3170–3198.

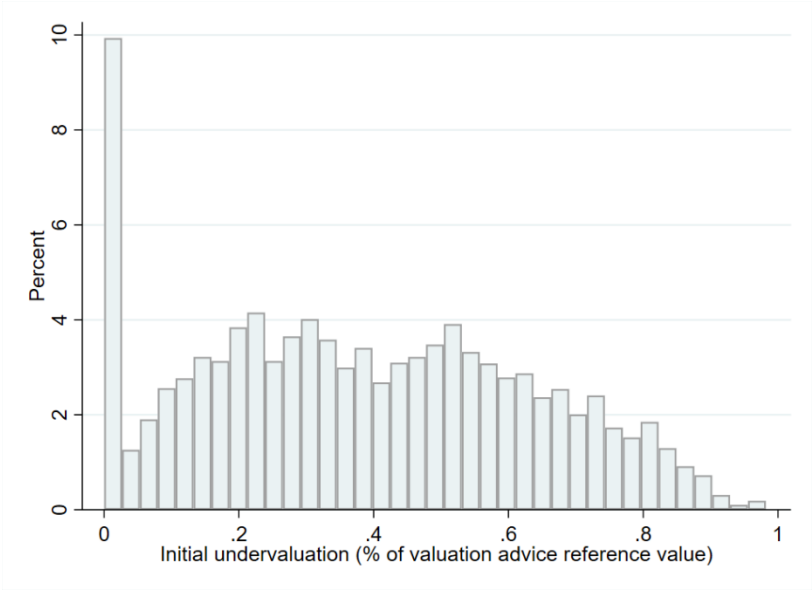
Yang, Dean. 2008. Can enforcement backfire? Crime displacement in the context of customs reform in the Philippines. *The Review of Economics and Statistics* 90, (1): 1–14.

Figure 1: The customs clearance process in Madagascar



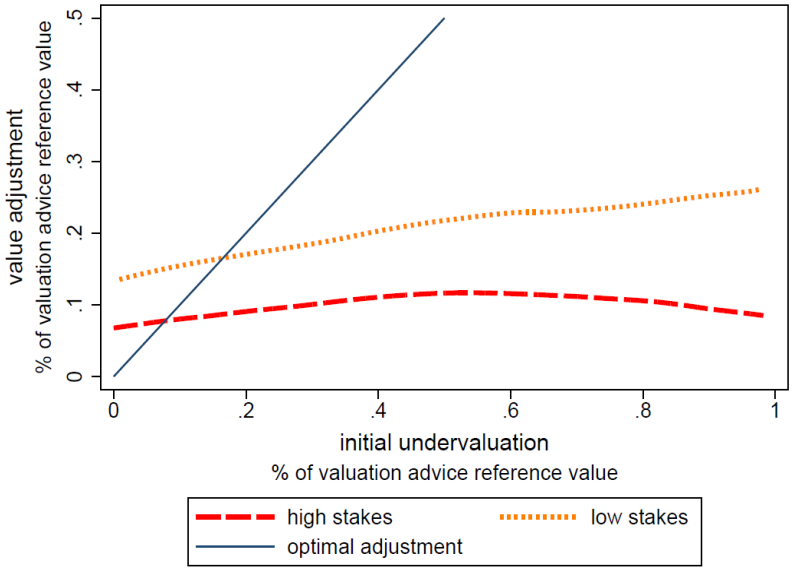
Notes: The figure is a stylized representation. For instance, as part of the clearance process, a pre-arrival notification has to be submitted prior to the cargo's arrival (that is prior to step 1 in the figure). GasyNet uses information from this pre-arrival notification in its risk analysis.

Figure 2: Distribution of initial undervaluation relative to valuation advice across declarations



Note: Sample covers import declarations subject to a valuation advice and registered between January 1, 2016 and October 14, 2018. The valuation advice reference value corresponds to the declaration's value advised by the service provider.

Figure 3: Value adjustment versus initial undervaluation for declarations with valuation advice



Notes: High-stakes declarations are those for which the product between the tax rate and the advised value by the service provider exceeds 10,000 USD and low-stakes declarations are all other declarations also subject to valuation advice. The 45-degree blue line represents the theoretically optimal adjustment from a revenue mobilization perspective (assuming the reference valuation is correct).

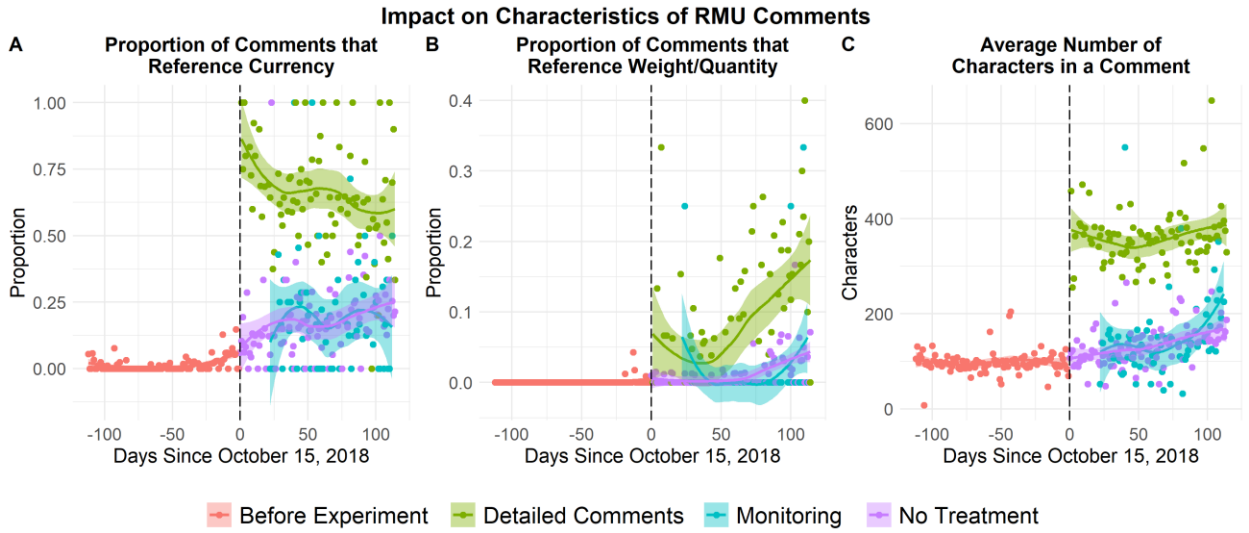
Figure 4: Monitoring announcement for specific declarations during RCT

Instructions sur base DAU et BSC

Date	Service	Login	Commentaire	Action
12/11/2018 15:45:44	DGD	System	DAU sélectionné par le système pour contrôle	Contrôle accru

Notes: The message from the Directorate General of Customs: (i) indicates that the declaration has been selected and will be subject to monitoring; and (ii) is displayed in the internal customs application The application is used by frontline customs inspectors to find, for each customs declaration, all relevant background information (attached documentation, valuation advice, comments, scanner images, etc.).

Figure 5: Evolution of comments



Notes: Each dot represent a daily average. The lines reflect the locally estimated scatterplot smoothing (LOESS) fits and the shaded areas the associated 95% confidence intervals. Sample covers import declarations subject to RMU comments with completed registration between October 15, 2018 and February 6, 2019. The start of the RCT (October 15) is marked as day 0.

Table 1: Descriptive statistics for the third-party valuation advice analysis

	Non Valuation Advice (N)			Valuation Advice (VA)			(N)=(VA)
	Mean	**SD**	**N. obs**	**Mean**	**SD**	**N. obs**	**P-value**
Declaration characteristics							
Tax rate	0.280	0.147	104,926	0.371	0.080	5,898	0.000
Red channel	0.277	0.447	104,926	0.267	0.442	5,898	0.000
Risk score	5.349	3.283	92,067	7.473	2.076	5,801	0.000
Differentiated	0.753	0.418	104,926	0.753	0.410	5,898	0.000
Mixed	0.375	0.484	104,926	0.491	0.500	5,898	0.000
Time-sensitive	0.039	0.182	104,926	0.050	0.208	5,898	0.000
Weight (log)	8.912	2.397	104,609	10.136	1.001	5,804	0.000
Undervaluation (rel to internal reference price)	0.070	0.614	102,409	0.182	0.481	5,748	0.000
Undervaluation (rel. to valuation advice)				0.376	0.324	5,803	0.000
Inspector actions & customs outcomes							
Upgrade	0.105	0.306	75,879	0.491	0.500	4,323	0.000
Scanned	0.133	0.339	104,926	0.516	0.500	5,898	0.000
Time (log hours)	3.634	1.285	99,246	4.695	1.076	5,550	0.000
Fraud	0.033	0.180	104,926	0.448	0.497	5,898	0.000
Δlog value	0.008	0.066	104,496	0.114	0.183	5,787	0.000
Δlog tax	0.010	0.073	90,744	0.112	0.181	5,850	0.000
Miscellaneous							
Workload (log)	5.055	0.815	104,926	5.051	0.858	5,898	0.000

Note: Sample covers import declarations registered between January 1, 2016 and October 14, 2018.

Table 2: The impact of valuation advice on inspector actions and customs outcomes

Panel A: OLS						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Valuation advice (VA)	0.260***	0.213***	0.938***	0.375***	0.094***	0.092***
	(0.024)	(0.017)	(0.036)	(0.020)	(0.005)	(0.005)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Risk score dummies	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Average declarations without VA	0.102	0.127	3.627	0.032	0.008	0.010
Observations	75,916	103,635	98,478	103,635	103,521	89,842
R-squared	0.412	0.484	0.309	0.348	0.273	0.270
Panel B: IV						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Valuation advice (VA)	0.120	0.132**	0.745***	0.217***	0.049***	0.052***
	(0.077)	(0.066)	(0.269)	(0.077)	(0.013)	(0.014)
	[0.299]	[0.149]	[0.029]	[0.029]	[0.005]	[0.005]
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Risk score dummies	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Average declarations without VA	0.102	0.127	3.627	0.032	0.008	0.010
Kleibergen Paap F-statistic	176.732	165.391	143.667	165.391	164.980	255.736
Observations	75,916	103,635	98,478	103,635	103,521	89,842
R-squared	0.031	0.084	0.059	0.116	0.060	0.057

Notes: Robust standard errors clustered by inspector in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al., 2006 and presented in brackets. Declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed (i.e., contained multiple products), and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations. In Panel B, valuation advice is instrumented using the share of declarations with a similar risk score registered in the same port in the same month that received valuation advice.

Table 3: Heterogeneity in responses of declarations subject to valuation advice

Dependent variable:	Upgrade	Scan	Time (log hours)	Fraud	Δlog value	Δlog tax
	(1)	(2)	(3)	(4)	(5)	(6)
High stakes (VA)	0.031*	0.035	0.180***	0.102***	0.019***	0.018**
	(0.017)	(0.028)	(0.050)	(0.020)	(0.007)	(0.008)
	[0.219]	[1.000]	[0.022]	[0.337]	[1.000]	[1.000]
High stakes (VA)*Initial undervaluation (VA)	-0.012	-0.050	-0.477***	-0.343***	-0.141***	-0.134***
	(0.063)	(0.063)	(0.136)	(0.080)	(0.029)	(0.027)
	[1.000]	[1.000]	[0.219]	[0.140]	[0.050]	[0.204]
Initial undervaluation (VA)	0.246***	0.269***	1.327***	0.628***	0.353***	0.336***
	(0.058)	(0.052)	(0.150)	(0.077)	(0.039)	(0.034)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Risk score dummies	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Average non-high stakes	0.337	0.463	4.619	0.396	0.103	0.102
Observations	6,601	6,601	6,320	6,601	6,582	6,542
R-squared	0.574	0.512	0.430	0.501	0.474	0.446

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al. (2006) and presented in brackets. Declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed (i.e., contained multiple products), and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations. OLS estimation is used. High-stakes declarations are those for which the product between the tax rate and the advised value by the service provider exceeds 10,000 USD.

Table 4: Balance tests

Panel A: Non Valuation Advice					
	Controls	**Comments (C)**	**(C)=Controls**	**Monitoring (M)**	**(M)=Controls**
	Mean	**Mean**	**P-value**	**Mean**	**P-value**
Targeted characteristics					
Risk Score 9	0.206	0.217	0.467	0.276	0.020
Declaration characteristics					
Tax rate	0.313	0.325	0.018	0.316	0.770
Red channel	0.292	0.264	0.092	0.459	0.000
Mixed	0.440	0.464	0.168	0.449	0.800
Differentiated	0.753	0.756	0.844	0.674	0.009
Time-sensititive	0.044	0.028	0.016	0.014	0.035
Weight (log)	9.372	9.421	0.473	9.762	0.014
Undervaluation (rel to. internal reference price)	-0.031	0.005	0.038	0.089	0.003
Joint significance of differences					
F-statistic			2.667		6.390
P-value			0.006		0.000
Observations	4,733	992		196	

Panel B: Valuation Advice					
	Controls	**Comments (C)**	**(C)=Controls**	**Monitoring (M)**	**(M)=Controls**
	Mean	**Mean**	**P-value**	**Mean**	**P-value**
Targeted characteristics					
Risk Score 9	0.264	0.268	0.953	0.368	0.320
Other characteristics					
Tax rate	0.371	0.370	0.916	0.348	0.199
Red channel	0.245	0.154	0.034	0.474	0.021
Mixed	0.651	0.602	0.313	0.526	0.270
Differentiated	0.801	0.853	0.163	0.560	0.006
Time-sensititive	0.031	0.048	0.278	0.060	0.426
Weight (log)	10.064	9.967	0.240	10.548	0.021
Undervaluation (rel to. internal reference price)	0.118	0.050	0.116	0.078	0.690
Joint significance of differences					
F-statistic			1.553		2.856
P-value			0.136		0.004
Observations	644	123		19	

Note: Sample covers import declarations with a risk score of 8 or 9 and registered between October 15, 2018 and February 6, 2019.

Table 5: Quality of comments

Panel A: Variable Comments only				
Dependent variable:	**Comment's length**	**Price reference**	**Past transaction reference**	**Weight reference**
	(1)	(2)	(3)	(4)
Comments	213.081***	0.443***	0.108***	0.063***
	(4.292)	(0.014)	(0.009)	(0.006)
	[0.000]	[0.000]	[0.000]	[0.000]
Declaration characteristics	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes
Observations	5,720	5,720	5,720	5,720
R-squared	0.604	0.476	0.246	0.273
Panel B: Interaction with tax rate variable				
Dependent variable:	**Comment's length**	**Price reference**	**Past transaction reference**	**Weight reference**
	(1)	(2)	(3)	(4)
Comments	173.401***	0.295***	0.079***	0.081***
	(11.084)	(0.036)	(0.023)	(0.014)
	[0.000]	[0.000]	[0.004]	[0.000]
Tax rate	82.197***	0.324***	0.107*	0.039
	(27.033)	(0.089)	(0.057)	(0.035)
	[0.012]	[0.002]	[0.269]	[1.000]
Comments*Tax rate	122.795***	0.457***	0.088	-0.054
	(31.635)	(0.104)	(0.066)	(0.041)
	[0.001]	[0.000]	[0.742]	[0.742]
Declaration characteristics	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes
Observations	5,720	5,720	5,720	5,720
R-squared	0.605	0.478	0.246	0.273

Notes: Robust standard errors clustered by inspector in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al. (2006) and presented in brackets. In Panel A, declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed (i.e., contained multiple products), and the log of the initial weight of the declaration. Declarations characteristics. In Panel B, declaration characteristics include the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed, and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations.

Table 6: Impact of comments and monitoring on inspector actions and customs outcomes

Panel A: Non Valuation Advice (Non VA)						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments	0.117***	0.106***	0.274***	0.031***	0.008***	0.010***
	(-0.028)	(-0.024)	(-0.059)	(-0.010)	(-0.003)	(-0.003)
	[0.008]	[0.005]	[0.005]	[0.053]	[0.051]	[0.051]
Monitoring	0.079	0.095**	0.089	0.002	0.001	0.003
	(-0.067)	(-0.042)	(-0.121)	(-0.014)	(-0.003)	(-0.003)
	[1.000]	[0.423]	[1.000]	[1.000]	[1.000]	[1.000]
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Average controls	0.252	0.209	3.953	0.032	0.005	0.007
Observations	3,503	4,898	4,687	4,898	4,893	4,338
R-squared	0.594	0.602	0.440	0.411	0.349	0.398
Comparison with VA, p-value	0.942	0.709	0.399	0.835	0.883	0.417
Panel B: Valuation Advice (VA)						
Dependent variable:	**Upgrade**	**Scan**	**Time**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments	0.122	0.127*	0.111	0.024	0.007	0.003
	(0.076)	(0.069)	(0.226)	(0.067)	(0.020)	(0.021)
	[1.000]	[0.960]	[1.000]	[1.000]	[1.000]	[1.000]
Monitoring	0.066	0.036	0.224	0.127	-0.014	0.003
	(0.223)	(0.183)	(0.360)	(0.182)	(0.054)	(0.047)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Average controls	0.641	0.585	4.884	0.508	0.105	0.107
Observations	467	623	608	623	622	623
R-squared	0.801	0.803	0.677	0.766	0.749	0.773

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al. (2006) and presented in brackets. Declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed, and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations.

Table 7: Heterogeneity in comments precision

Panel A: Non Valuation Advice						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments with exact references (CERs)	0.116***	0.109***	0.304***	0.047***	0.013***	0.013***
	(0.028)	(0.023)	(0.082)	(0.014)	(0.004)	(0.004)
	[0.019]	[0.009]	[0.033]	[0.047]	[0.052]	[0.033]
Other comments (OCs)	0.119***	0.100***	0.215***	-0.001	-0.001	0.001
	(0.039)	(0.033)	(0.059)	(0.008)	(0.002)	(0.002)
	[0.088]	[0.088]	[0.033]	[1.000]	[1.000]	[1.000]
Monitoring	0.079	0.095**	0.089	0.001	0.001	0.003
	(0.067)	(0.042)	(0.121)	(0.015)	(0.003)	(0.003)
	[1.000]	[0.471]	[1.000]	[1.000]	[1.000]	[1.000]
Observations	3,503	4,898	4,687	4,898	4,893	4,338
R-squared	0.594	0.602	0.440	0.413	0.352	0.400
Comparison CERs = Ocs						
P-value	0.923	0.752	0.381	0.007	0.005	0.001
Q-value	1.000	1.000	1.000	0.036	0.036	0.011
Panel B: Valuation Advice						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments with exact references (CERs)	0.137*	0.129**	0.125	0.047	0.012	0.008
	(0.072)	(0.061)	(0.225)	(0.070)	(0.017)	(0.018)
	[0.850]	[0.653]	[1.000]	[1.000]	[1.000]	[1.000]
Other comments (OCs)	-0.079	0.088	-0.101	-0.294*	-0.059	-0.078
	(0.267)	(0.265)	(0.302)	(0.147)	(0.100)	(0.094)
	[1.000]	[1.000]	[1.000]	[0.740]	[1.000]	[1.000]
Monitoring	0.069	0.038	0.234	0.140	-0.011	0.007
	(0.225)	(0.184)	(0.371)	(0.186)	(0.055)	(0.048)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Observations	467	623	608	623	622	623
R-squared	0.802	0.803	0.677	0.768	0.750	0.774
Comparison CERs = Ocs						
P-value	0.392	0.860	0.390	0.034	0.465	0.343
Q-value	1.000	1.000	1.000	0.499	1.000	1.000

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al. (2006) and presented in brackets. Regressions include declaration characteristics, inspector fixed effects, importer fixed effects, HS2 product fixed effects, country of origin fixed effects and port-month fixed effects. Declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed, and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations. Comments with exact references are comments that either reference a price, a past transaction, or a weight.

Table 8: Heterogeneity in responses of declarations not subject to valuation advice – declarations receiving extensive comments only

Dependent variable:	Upgrade	Scan	Time (log hours)	Fraud	Δlog value	Δlog tax
	(1)	(2)	(3)	(4)	(5)	(6)
High stakes (C)	0.028	0.051	0.028	-0.002	-0.002	-0.006
	-0.119	-0.063	-0.241	-0.036	-0.010	-0.010
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
High stakes (C)*Initial undervaluation (internal reference price)	-0.174	-0.173	0.115	-0.161	-0.037*	-0.042*
	(0.215)	(0.214)	(0.435)	(0.129)	(0.019)	(0.021)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Initial undervaluation (internal reference price)	-0.002	0.038	-0.106	0.082*	0.021	0.024
	(0.139)	(0.067)	(0.269)	(0.046)	(0.014)	(0.017)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Average non high stakes	0.401	0.320	4.268	0.072	0.018	0.022
Observations	403	534	519	534	534	477
R-squared	0.831	0.790	0.709	0.762	0.666	0.771

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al. (2006) and presented in brackets. The sample used is non valuation advice declarations receiving comments. High stakes are defined as declaration with more than 10,000 USD in potential tax yield based on the hypothetical valuation calculated using internal reference prices. The hypothetical valuation for the declaration is calculated multiplying the initial weights of the items in the declaration by the corresponding internal reference prices. The initial undervaluation is defined as the difference between the log of the hypothetical valuation and the log of the initially submitted value by the importer or his representative Regressions include declaration characteristics, inspector fixed effects, importer fixed effects, HS2 product fixed effects, country of origin fixed effects and port-month fixed effects. Declaration characteristics include the VAT rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed (i.e. contained multiple products) and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations.

Appendix for online publication

Appendix A. Data

Use of text analysis to analyze comments in the RCT

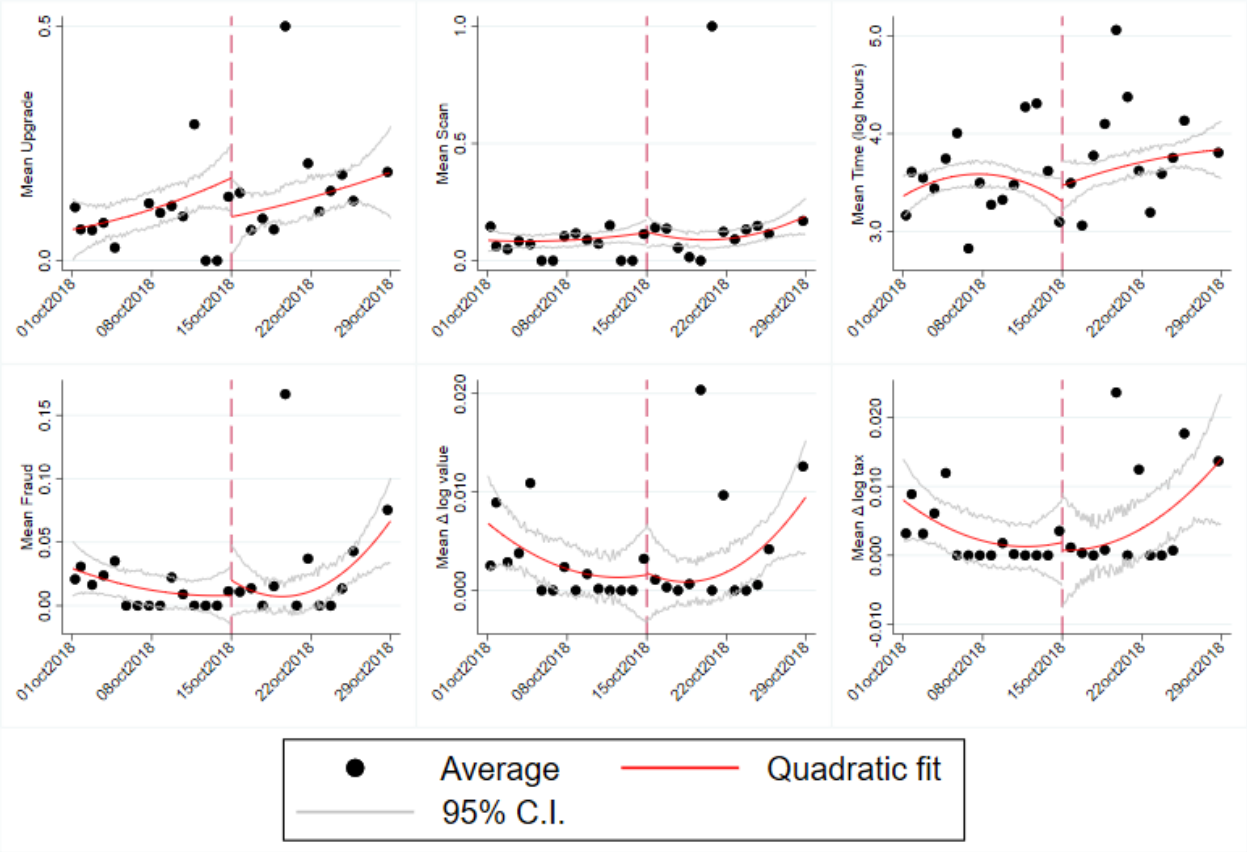
We used text analysis to construct outcome variables measuring the quality of the comments made by the risk management unit to each declaration. Comments to all declarations were considered, whether or not they were extensive comments in the treatment group or regular comments in the control group.

The first step in the text analysis is to clean the comments for example by adding spaces between all letters and numbers (whenever those were missing) or by removing extra spaces and removing punctuation. Then an algorithm is applied to the comments to look for specific types of information.

1. Length of message: is the number of characters in a message.
2. Whether price was explicitly referred: the algorithm searches for whether a currency is mentioned next to a number. Specifically, the algorithm looks for whether “usd” or “eur” is mentioned next to a number. Finding “usd 10,000” in the text counts as a price but finding just “usd” does not count as a price.
3. Whether a prior transaction was explicitly referred to: the algorithm searches for “bsc” (the declaration number) being mentioned next to a number.
4. Whether a weight was mentioned: the algorithm searches for a weight term ("carton", "kg","metre", "piece", "sac", "tonne") being mentioned next to a number.

For each of the types of information 1 through 4, an indicator variable is constructed equal to 1 if that type of information is found in the comments of a declaration, and 0 otherwise.

Appendix Figure 1: Inspectors actions and customs outcomes around start of RCT



Notes: Sample covers import declarations with a risk score of 8 or 9 and registered between September 15, 2018 and February 6, 2019. Averages are the binned average of the data. The fitted values show the result of a quadratic fit on the data.

Appendix Table 1: Definitions of variables

Variable name	Variable definition and data source
Instrumental variables	
Share of valuation advice	Share of declarations with a similar risk score registered in the same port in the same month that received the valuation advice. Data sources: Madagascar customs and GasyNet.
Outcome variables: inspector actions / customs outcomes	
Upgrade of clearance channel	Dummy variable equal to 1 if the customs inspector upgrades the clearance channel from yellow to red and 0 otherwise. This variable is defined only if the initial clearance channel is red. Data source: Madagascar customs.
Scanning	Dummy variable equal to 1 if the customs inspector asked the service provider to scan the containerized goods in the declaration. Data source: GasyNet.
Clearance time	Log of the difference in time (hours) between the date of assessment of the declaration by the inspector and the date of registration of the declaration in the customs electronic system. Data source: Madagascar customs.
Fraud recorded	Dummy variable equal to 1 if the customs inspector identifies fraud in the import declaration and 0 otherwise. Data source: Madagascar customs.
Value adjustment	Difference between the log of the declaration value retained by customs and the log of the initially submitted value by the importer or his representative. Variable is measured in percent. Data source: Madagascar customs.
Value adjustment (VA)	Ratio of the difference between the declaration value retained by customs and the initially submitted value by the importer or his representative (numerator) to the valuation advice reference value (denominator). Missing value for declarations not subject to a valuation advice. Sources: Madagascar customs and GasyNet.
Tax adjustment	Difference between the log of the taxes paid (including tariff, VAT) on the declaration and the log of taxes that should have been paid in the absence of customs controls (which equal taxes paid minus tax adjustment by the customs inspector). Variable is measured in percent. Data source: Madagascar customs.
Control variables: ex-ante risk characteristics of import declarations	
Tax rate	Sum of taxes (including tariff, VAT) paid divided by the import value retained by customs. Data source: Madagascar customs.
Red channel	Dummy variable equal to 1 if the customs risk management system routed the declaration to the frontline inspection channel (red channel) and 0 otherwise. Data source: Madagascar customs.
Mixed shipment	Dummy variable equal to 1 if the import declaration includes more than 1 HS 6-digit product and 0 otherwise. Source: Madagascar customs.

Differentiated	Share of HS 6-digit products in the import declaration that are defined as differentiated according to the classification by Rauch (1999). Data source: Rauch (1999) and a concordance between HS 6-digit revision 2012 classification and SITC revision 2 classification from COMTRADE.
Time-sensitive	Share of HS 6-digit products in the import declaration that are defined as time-sensitive according to the classification by Hummels and Schaur (2013). Data source: Hummels and Schaur (2013).
Weight	Log of the initially submitted total weight of the declaration by the importer or his representative. Data source: Madagascar customs.
Initial undervaluation	Difference between the log of the hypothetical valuation and the log of the initially submitted value by the importer or his representative. The hypothetical valuation for the declaration is calculated multiplying the initial weights of the items in the declaration by the corresponding internal reference prices. Source: Madagascar customs.
Initial undervaluation (VA)	One minus the ratio of the initially submitted value by the importer or his representative (numerator) to the valuation advice reference value (denominator). Missing value for declarations not subject to a valuation advice. Sources: Madagascar customs and GasyNet.
Inspector workload	Log of the number of import declarations the inspector handled in a month. Source: Madagascar customs.
VAT rate	Value Added Tax (VAT) paid divided by the import value retained by customs. Data source: Madagascar customs.
Comments quality: characteristics of extensive comments	
Comment's length	Length of the comment expressed as number of characters. Data source: GasyNet.
Comment with exact reference	Dummy equal to one if the comment references a price, a past transaction, or a weight. Data source: GasyNet.
Other comments	Dummy equal to one if the declaration was subject to comments that did not contain exact references. Data source: GasyNet
Price reference	Dummy equal to one if the comment references a price as measured by a currency detected next to numerical characters. Data source: GasyNet.
Past transaction reference	Dummy equal to one if the comment references a past transaction as measured by a currency detected next to numerical characters. Data source: GasyNet.
Weight reference	Dummy equal to one if the comment reference a weight as measured by a unit of weight detected next to numerical characters. Units include abbreviation of metric units ton and kilogram. Data source: GasyNet.

Appendix Table 2: Correlation across prices

		Price provided in the comments (treatment group)	Price based on COMTRADE mirror export data	Price based on median across declarations	Price provided in the valuation advice
Price based on COMTRADE mirror export data	ρ	0.113***			
	N	569			
Price based on median across declarations	ρ	0.375***	0.182***		
	N	710	569		
Price provided in the valuation advice	ρ	0.342***	0.300***	0.411***	
	N	114	100	114	
Price in declaration (import value / import weight)	ρ	0.360***	0.154***	0.840***	0.489***
	N	710	569	710	114

Notes: *** indicates significance at the 1% level. Sample covers import declarations subject to valuation advice as well to RMU comments with exact references, registered between October 15, 2018 and February 6, 2019, including only one HS 6-digit product, and with available COMTRADE mirror export data.

Appendix Table 3: First-stage regressions – Valuation Advice

Dependent variable :	**Valuation Advice (VA)**					
Corresponding second-stage outcome:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Share of valuation advice	0.833***	0.775***	0.776***	0.775***	0.774***	0.892***
	(0.063)	(0.060)	(0.065)	(0.060)	(0.060)	(0.056)
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Risk score dummies	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	75,916	103,635	98,478	103,635	103,521	89,842
F-statistic	176.732	165.391	143.667	165.391	165.391	255.736
P-value	0.000	0.000	0.000	0.000	0.000	0.000

Notes: OLS estimator. Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Declarations characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed, and the log of the initial weight of the declaration. Share of valuation advice is defined as the share of declarations with a similar risk score registered in the same port in the same month that received valuation advice.

Appendix Table 4: Robustness checks to impacts of Valuation Advice

Panel A: Controlling for inspectors' workload						
Dependent variable:	**Upgrade**	**Scan**	**Time**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Valuation advice (VA)	0.121	0.132**	0.741***	0.217***	0.049***	0.052***
	(0.077)	(0.066)	(0.269)	(0.077)	(0.013)	(0.014)
	[0.754]	[0.377]	[0.076]	[0.076]	[0.012]	[0.012]
Workload (log)	-0.001	-0.003	0.042	-0.001	0.001	0.000
	(0.003)	(0.004)	(0.032)	(0.002)	(0.000)	(0.001)
	[1.000]	[1.000]	[1.000]	[1.000]	[0.934]	[1.000]
Average declarations without VA	0.102	0.127	3.627	0.032	0.008	0.010
Kleibergen Paap F-statistic	176.293	165.420	143.699	165.420	165.002	255.889
Observations	75,916	103,635	98,478	103,635	103,521	89,842
R-squared	0.031	0.084	0.059	0.116	0.060	0.057
Panel B: Controlling for inspector-month fixed effects						
	(7)	(8)	(9)	(10)	(11)	(12)
Valuation advice (VA)	0.106	0.123*	0.803***	0.222***	0.049***	0.052***
	(0.079)	(0.069)	(0.278)	(0.079)	(0.013)	(0.015)
	[0.453]	[0.239]	[0.026]	[0.026]	[0.007]	[0.007]
Average declarations without VA	0.102	0.127	3.627	0.032	0.008	0.010
Kleibergen Paap F-statistic	167.540	153.439	134.173	153.439	153.558	235.375
Observations	75,877	103,612	98,462	103,612	103,498	89,815
R-squared	0.029	0.083	0.060	0.117	0.060	0.057
Panel C: Controlling for inspector-broker fixed effects						
	(13)	(14)	(15)	(16)	(17)	(18)
Valuation advice (VA)	0.153*	0.126**	0.854***	0.234***	0.052***	0.059***
	(0.076)	(0.058)	(0.228)	(0.078)	(0.013)	(0.015)
	[0.125]	[0.105]	[0.002]	[0.015]	[0.002]	[0.002]
Average declarations without VA	0.102	0.127	3.627	0.032	0.008	0.010
Kleibergen Paap F-statistic	209.386	183.644	158.897	183.644	182.723	251.735
Observations	75,621	103,360	98,207	103,360	103,245	89,558
R-squared	0.036	0.085	0.065	0.122	0.063	0.060
Panel D: Controlling for initial undervaluation						
	(19)	(20)	(21)	(22)	(23)	(24)
Valuation advice (VA)	0.109	0.127*	0.751***	0.218***	0.048***	0.051***
	(0.076)	(0.068)	(0.274)	(0.077)	(0.013)	(0.014)
	[0.480]	[0.223]	[0.032]	[0.027]	[0.006]	[0.007]
Initial undervaluation (rel to.	0.007***	0.005***	0.026***	0.012***	0.008***	0.008***
internal reference price)	(0.002)	(0.002)	(0.008)	(0.002)	(0.001)	(0.001)
	[0.015]	[0.019]	[0.019]	[0.027]	[0.006]	[0.000]
Average declarations without VA	0.102	0.127	3.627	0.032	0.008	0.010
Kleibergen Paap F-statistic	170.934	166.066	144.420	166.066	165.950	260.731
Observations	74,535	101,700	96,682	101,700	101,619	88,227
R-squared	0.029	0.084	0.059	0.117	0.063	0.059

Notes: Robust standard errors clustered by inspector in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al. (2006) and presented in brackets. Regressions include declaration characteristics, risk score dummies, inspector fixed effects, importer fixed effects, HS2 product fixed effects, country of origin fixed effects and port-month fixed effects. Panel B includes inspector-month fixed effects instead of inspector fixed effects. Panel C includes inspector-broker fixed effects instead of inspector fixed effects. Declarations characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed (i.e., contained multiple products), and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations. Valuation advice is instrumented using the share of declarations with a similar risk score registered in the same port in the same month that received valuation advice.

Appendix Table 5: Inspector-level spillovers - Impact of share of Valuation Advice on declarations not subject to Valuation Advice

Panel A: Issuance of Valuation Advice

Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Valuation Advice	-0.002	0.015	0.477	-0.029	-0.001	-0.007
	(0.038)	(0.033)	(0.377)	(0.025)	(0.005)	(0.006)
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Risk score dummies	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	71,617	97,918	93,005	97,918	97,824	84,167
R-squared	0.347	0.433	0.256	0.165	0.155	0.156

Panel B: Issuance of Valuation Advice and workload

Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Valuation Advice	-0.362*	-0.056	-0.810	0.051	-0.006	-0.002
	(0.193)	(0.212)	(0.895)	(0.058)	(0.023)	(0.024)
Workload (log)	-0.006	-0.003	-0.001	0.002	0.001	0.001
	(0.004)	(0.004)	(0.034)	(0.002)	(0.000)	(0.001)
Share of Valuation Advice x Workload (log)	0.072*	0.015	0.264	-0.017	0.001	-0.001
	(0.041)	(0.043)	(0.237)	(0.011)	(0.005)	(0.005)
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Risk score dummies	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	71,617	97,918	93,005	97,918	97,824	84,167
R-squared	0.347	0.433	0.256	0.165	0.155	0.156

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed, and the log of the initial weight of the declaration. Sample includes only declarations not receiving valuation advice. Observations refers to the number of non-singleton observations. Share of valuation advice is defined as the share of declarations of an inspector receiving valuation advice.

Appendix Table 6: Balance - Pooled sample

	Controls	Comments (C)	(C)=Controls	Monitoring (M)	(M)=Controls
	Mean	**Mean**	**P-value**	**Mean**	**P-value**
Targeted characteristics					
Valuation advice (VA)	0.135	0.118	0.122	0.101	0.143
Risk Score 9	0.210	0.220	0.423	0.280	0.013
Declaration characteristics					
Tax rate	0.321	0.330	0.038	0.320	0.880
Red channel	0.283	0.252	0.032	0.463	0.000
Mixed	0.468	0.478	0.514	0.459	0.798
Differentiated	0.760	0.766	0.680	0.664	0.001
Time-sensititive	0.044	0.031	0.043	0.023	0.118
Weight (log)	9.460	9.486	0.702	9.833	0.010
Undervaluation (relative to internal reference price)	-0.015	0.009	0.176	0.092	0.005
Joint significance of differences					
F-statistic			2.311		7.094
P-value			0.014		0.000
Observations	5,407	1,125		218	

Note: Sample covers import declarations with a risk score of 8 or 9 and registered between October 15, 2018 and February 6, 2019.

Appendix Table 7: Robustness check- Controlling for initial price undervaluation

Panel A: Non Valuation Advice						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments	0.120***	0.110***	0.285***	0.031***	0.007***	0.008***
	(0.030)	(0.025)	(0.069)	(0.009)	(0.003)	(0.003)
	[0.013]	[0.012]	[0.013]	[0.037]	[0.137]	[0.137]
Monitoring	0.091	0.100**	0.044	-0.002	-0.003	-0.002
	(0.069)	(0.046)	(0.124)	(0.016)	(0.003)	(0.003)
	[1.000]	[0.423]	[1.000]	[1.000]	[1.000]	[1.000]
Average controls	0.252	0.209	3.953	0.032	0.005	0.007
Observations	3,124	4,297	4,112	4,297	4,297	3,768
R-squared	0.594	0.601	0.455	0.387	0.334	0.374
Panel B: Valuation Advice						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments	0.204**	0.143*	0.080	0.084	0.028*	0.024
	(0.086)	(0.073)	(0.240)	(0.079)	(0.015)	(0.019)
	[0.371]	[0.641]	[1.000]	[1.000]	[0.680]	[1.000]
Monitoring	-0.082	-0.079	0.289	0.152	-0.031	-0.009
	(0.252)	(0.228)	(0.441)	(0.284)	(0.075)	(0.077)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Average controls	0.641	0.585	4.884	0.508	0.105	0.107
Observations	373	492	482	492	492	492
R-squared	0.820	0.808	0.713	0.791	0.808	0.836

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al. (2006) and presented in brackets. Regressions include declaration characteristics, inspector fixed effects, importer fixed effects, HS2 product fixed effects, country of origin fixed effects and port-month fixed effects. Declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed (i.e. contained multiple products) and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations.

Appendix Table 8: Robustness check- Controlling for inspector-month fixed effects

Panel A: Non Valuation Advice						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments	0.118***	0.107***	0.294***	0.033***	0.008***	0.010***
	(0.024)	(0.022)	(0.062)	(0.010)	(0.002)	(0.003)
	[0.002]	[0.002]	[0.002]	[0.035]	[0.035]	[0.035]
Monitoring	0.093	0.096**	0.108	-0.009	-0.000	0.001
	(0.067)	(0.042)	(0.120)	(0.015)	(0.003)	(0.003)
	[1.000]	[0.379]	[1.000]	[1.000]	[1.000]	[1.000]
Average controls	0.252	0.209	3.953	0.032	0.005	0.007
Observations	3,481	4,870	4,662	4,870	4,865	4,310
R-squared	0.637	0.639	0.485	0.450	0.379	0.434
Panel B: Valuation Advice						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments	0.051	0.049	-0.033	-0.042	-0.006	-0.010
	(0.058)	(0.062)	(0.334)	(0.069)	(0.022)	(0.023)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Monitoring	0.194	0.179	0.023	0.335	0.041	0.023
	(0.283)	(0.158)	(0.312)	(0.218)	(0.066)	(0.055)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Average controls	0.641	0.585	4.884	0.508	0.105	0.107
Observations	361	481	469	481	479	481
R-squared	0.930	0.901	0.817	0.894	0.863	0.882

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al. (2006) and presented in brackets. Regressions include declaration characteristics, inspector-month fixed effects, importer fixed effects, HS2 product fixed effects, country of origin fixed effects and port-month fixed effects. Declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed (i.e. contained multiple products) and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations.

Appendix Table 9: Robustness check - Controlling for inspector-broker fixed effects

Panel A: Non Valuation Advice						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments	0.120***	0.101***	0.265***	0.025**	0.004**	0.007**
	(0.029)	(0.024)	(0.079)	(0.010)	(0.002)	(0.003)
	[0.014]	[0.014]	[0.076]	[0.280]	[0.280]	[0.284]
Monitoring	0.010	0.067*	0.019	-0.022**	-0.003	-0.002
	(0.055)	(0.034)	(0.124)	(0.009)	(0.002)	(0.002)
	[1.000]	[0.646]	[1.000]	[0.280]	[1.000]	[1.000]
Average controls	0.252	0.209	3.953	0.032	0.005	0.007
Observations	3,271	4,652	4,452	4,652	4,645	4,114
R-squared	0.682	0.672	0.546	0.527	0.529	0.539
Panel B: Valuation Advice						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Comments	0.147	0.109	0.102	-0.021	0.012	0.000
	(0.186)	(0.138)	(0.303)	(0.134)	(0.025)	(0.034)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Monitoring	0.098	0.187	0.197	0.124	0.001	0.003
	(0.173)	(0.238)	(0.257)	(0.195)	(0.041)	(0.038)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Average controls	0.641	0.585	4.884	0.508	0.105	0.107
Observations	264	378	371	378	377	378
R-squared	0.971	0.933	0.914	0.934	0.918	0.937

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Q-values are calculated following Benjamini et al. (2006) and presented in brackets. Regressions include declaration characteristics, inspector-broker fixed effects, importer fixed effects, HS2 product fixed effects, country of origin fixed effects and port-month fixed effects. Declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed (i.e. contained multiple products) and the log of the initial weight of the declaration. Observations refers to the number of non-singleton observations.

Appendix Table 10: Inspector-level spillovers - Impact of intensity of information and monitoring received on control declarations

Panel A: Intensity of valuation advice, extensive comments or monitoring

Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Share of valuation advice, comments or monitoring	0.039	0.062	0.285	0.006	-0.016	0.001
	(0.235)	(0.149)	(0.842)	(0.106)	(0.015)	(0.026)
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,737	3,819	3,651	3,819	3,815	3,350
R-squared	0.596	0.598	0.448	0.424	0.422	0.441

Panel B: Impact of intensity of valuation advice, extensive comments, or monitoring and workload

Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Share of valuation advice, comments or monitoring	1.960	0.823	4.193	-0.344	-0.058	-0.073
	(1.541)	(0.560)	(3.728)	(0.291)	(0.046)	(0.056)
Share of valuation advice, comments or monitoring x Workload (log)	-0.371	-0.169	-0.601	0.070	0.009	0.016
	(0.290)	(0.123)	(0.692)	(0.064)	(0.009)	(0.012)
Workload (log)	0.033	-0.008	0.360**	-0.012	-0.000	-0.001
	(0.063)	(0.028)	(0.161)	(0.013)	(0.002)	(0.003)
Declaration characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Inspector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Importer fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
HS2 product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Port - Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,737	3,819	3,651	3,819	3,815	3,350
R-squared	0.597	0.599	0.450	0.424	0.423	0.442

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Declaration characteristics include the tax rate, the share of value accounted for by differentiated products, the share of value accounted for by time-sensitive goods, a dummy indicating whether the declaration was mixed (i.e., contained multiple products), and the log of the initial weight of the declaration. Sample includes only control declarations. Share of valuation advice, comments or monitoring is defined as the share of the high risks declarations of an inspector receiving valuation advice, extensive comments or monitoring. Observations refers to the number of non-singleton observations.

Appendix Table 11: Evidence of inspectors' reaction to the onset of the RCT

Panel A: Local linear polynomial						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Estimated difference	-0.081	-0.040	-0.154	-0.013	-0.003	-0.004
	(0.074)	(0.066)	(0.146)	(0.017)	(0.005)	(0.005)
Observations (Full sample)	16,355	22,045	20,767	22,045	21,699	19,999
Observations used left of cutoff	1,426	1,951	1,624	1,672	1,310	1,195
Observations used right of cutoff	1,274	1,785	1,413	1,503	1,368	1,172
Panel B: Local quadratic polynomial						
Dependent variable:	**Upgrade**	**Scan**	**Time (log hours)**	**Fraud**	**Δlog value**	**Δlog tax**
	(1)	(2)	(3)	(4)	(5)	(6)
Estimated difference	-0.127	-0.007	-0.404*	0.009	-0.001	-0.002
	(0.097)	(0.091)	(0.220)	(0.026)	(0.006)	(0.007)
Observations (Full sample)	16,355	22,045	20,767	22,045	21,699	19,999
Observations used left of cutoff	1,565	1,934	1,761	1,951	2,332	2,068
Observations used right of cutoff	1,390	1,589	1,413	1,785	1,948	1,761

Notes: Standard errors are clustered by inspector and presented in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Bandwidth selection and point estimates were obtained using the method described in Cattaneo et al. (2015).