

# An Evaluation of Border Management Reforms in a Technical Agency

*Ana M. Fernandes*

*Russell Hillberry*

*Alejandra Mendoza Alcántara*



**WORLD BANK GROUP**

Development Research Group

Trade and International Integration Team

October 2017

## Abstract

Impact evaluations of trade facilitation reforms have almost exclusively focused on reforms by data-rich customs agencies. Other “technical” agencies also intervene in the logistics of international trade, and do so in ways that can cause significant interruptions in the flow of the imported products they oversee. This paper is the first to evaluate a reform by a technical agency, namely, the agency responsible for food safety and animal health in the former Yugoslav Republic of Macedonia. The data environment

is much more challenging than in customs, but enables the investigation of novel questions. The study finds that on-the-ground practices regarding sampling of import shipments departed substantially from those planned in the reform. It finds little evidence that the reform was successful in its attempt to improve the targeting of risky shipments. There is limited evidence that the reform increased trade flows, but circumstances make it difficult to establish a strong causal link to the specific reform studied.

---

This paper is a product of the Trade and International Integration Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at [afernandes@worldbank.org](mailto:afernandes@worldbank.org), [amendoza@worldbank.org](mailto:amendoza@worldbank.org), and [rhillber@purdue.edu](mailto:rhillber@purdue.edu).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. 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 for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# **An Evaluation of Border Management Reforms in a Technical Agency\***

Ana M. Fernandes <sup>a</sup>  
The World Bank

Russell Hillberry  
Purdue University

Alejandra Mendoza Alcántara  
The World Bank

**Keywords:** Impact evaluation, Trade Facilitation, Risk Management, Sanitary and Phytosanitary products.

**JEL Classification codes:** F13, F14, F15.

<sup>a</sup> Ana Margarida Fernandes (corresponding author). The World Bank. Development Research Group. 1818 H Street NW, Washington DC, 20433. Email: [afernandes@worldbank.org](mailto:afernandes@worldbank.org).

\* The authors would like to thank the Food and Veterinary Agency and the State Statistical Office of the FYR Macedonia for providing us with data, and Marija Manevska and Zorica Vasileva for valuable clarifications about the data and procedures. We thank Emine Elcin Koten for research assistance. Special thanks to Perica Vrboski, Todor Milchevski, and Violane Konar-Leacy for assisting with the data collection and with very helpful discussions and comments on the paper. We thank Jose-Daniel Reyes, Manon Schuppers and two anonymous reviewers for comments. We also thank Ma Soukha Ba, Vojdan Jordanov, and Simon Stoiljkovikj for help with data cleaning and Alejandro Forero-Rojas, Ganeshkumar Sathiyamoorthy, and Siddhesh Kaushik for help with WITS monthly data. Research for this paper has been supported in part by the World Bank's Multidonor Trust Fund for Trade and Development, the Strategic Research Program on Economic Development, the i2i Fund for Impact Evaluation, and the Investment Climate Impact Program. The findings expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank.

## Section I. Introduction

Each year hundreds of millions of dollars of development aid are spent supporting trade facilitation reforms.<sup>1</sup> Despite the enormous scale of the spending, relatively few of the specific reforms undertaken so far have been subject to scrutiny via formal impact evaluations. Those evaluations that have been undertaken have primarily concerned reforms involving national customs agencies.<sup>2</sup> As far as we know, this paper is the first impact evaluation of a border management reform undertaken by a “technical agency,” in this case the Food and Veterinary Agency (FVA) of the Former Yugoslav Republic of Macedonia (hereafter “FYR Macedonia”). The FVA is responsible for oversight of food safety and animal health laws in FYR Macedonia. In that capacity the FVA is responsible for oversight of imported products related to this mandate.

The reform we study is the introduction of a risk-based approach to sampling goods for laboratory testing that was undertaken by the FVA in 2014. The reform is similar in spirit to the introduction of risk-based inspection procedures undertaken by customs agencies, and it attempts to better target sampling and laboratory testing activities toward the import shipments with the highest risks while reducing the overall rate of sampling and testing activity. The reform of inspection regimes in customs agencies has been subject to some formal study.<sup>3</sup> But the operational capabilities and the mandates of technical agencies like the FVA are usually quite different from those of a customs agency. Typically, the technical agencies are not as well funded as customs. They also have important oversight responsibilities in the domestic sphere, which means that they are often not as focused on managing imports as are customs agencies. These differences are important, and mean that a study of reform in a technical agency provides new knowledge about the effectiveness of trade facilitation reforms in a very different setting.

We face a number of data challenges related to the organizational structure of the FVA’s control procedures and its data collection mechanisms. Because the mandate of the FVA differs from that of a standard customs agency we also have an opportunity to ask new questions. As in past studies of risk management in customs, we study changes in the frequency of oversight and (to the extent possible) the effects of the reform on trade flows. In this study, we also have access to planned rates of product sampling, so we are able to study the degree to which sampling activities in the field match those outlined in the *ex ante* sampling plan. We also have data on the outcomes of laboratory tests, which allows us to study the

---

<sup>1</sup> The Organization for Economic Cooperation and Development (OECD) maintains an Aid for Trade Facilitation database, which reports \$505.9 million disbursed in 2013 and \$480.8 million disbursed in 2014 (both figures in constant 2013 dollars). Data available at <http://www.oecd.org/dac/aft/Aid%20for%20Trade%20facilitation%20database%20%20.pdf>

<sup>2</sup> See Fernandes et al. (2015), Fernandes et al. (2016), Carballo et al. (2016b) and Carballo et al. (2016d). Carballo et al. (2016c) looks at the effects of postal trade reform using customs data, while Carballo et al. (2016a) investigate a multi-agency reform using customs data.

<sup>3</sup> Fernandes et al. (2015) conduct an evaluation of the introduction of risk management by the Albanian customs and study its effects on imports.

effects of the reform on the conditional probability that samples are not compliant with the FVA's technical standards.

Although the FVA has electronic communication and data storage capabilities, unlike most customs systems the FVA systems are not designed to isolate particular shipments for greater scrutiny. Electronic systems typically used by customs agencies exploit a range of data and statistical targeting methodologies to make shipment-by-shipment determinations about the need for physical inspection. FVA procedures rely more heavily on manual decision-making by staff stationed at border posts, and these staff have had substantial discretion over sampling activity. The premise of the FVA reform effort was that, *ex ante*, too many samples were taken, sampling activity was poorly targeted, and the costs associated with high rates of sampling and laboratory testing activity impeded imports.<sup>4</sup> The reform that we study intended to reduce the sampling rate and to provide administrative guidance to FVA border agents on appropriate levels of sampling activities for combinations of broad import product, country or set of countries of origin, and in some cases the specific hazard for which the sample should be tested (e.g., salmonella). One tangible outcome of the reform is that the FVA began to generate an annual sampling plan to guide the sampling activity in the forthcoming year. Our first empirical test estimates a statistical relationship between the actual samples taken in the field and the sampling activity projected in the sampling plan. Implementation delays and other important events not predicted by the sampling plan affected actual sampling rates, which partially explains our first key finding: that actual sampling activity adheres only weakly to the sampling plan.

The specific hazards that motivate sampling and testing of FVA products sometimes represent significant threats to public safety and to animal health. These public interests mean that the outcomes of laboratory tests are more readily available for evaluation than are outcomes of customs inspections.<sup>5</sup> The availability of data reporting the outcomes from the laboratory tests gives us an opportunity to estimate conditional correlations between non-compliant shipments (positive outcomes) and sampling activity. In principle, sampling activity should be positively correlated with the prevalence of positive outcomes (especially after the reform), although complicating factors such as cross-product variability in the severity of the potential harm are one reason that cross-commodity correlations might be weak.<sup>6</sup> We estimate a model with an indicator for positive outcomes from laboratory tests at the shipment level on the left-hand-

---

<sup>4</sup> The primary cost of sampling and laboratory activities are the delays associated with the wait for laboratory test outcomes. Goods in the shipment are not released onto the market until laboratory tests report a negative result (indicating compliance with FVA's technical standards).

<sup>5</sup> The identification of salmonella or aflatoxin in an import shipment would be communicated to the public, for example, while evidence of minor tariff avoidance activity identified in customs inspections would often be held confidentially by the customs agency.

<sup>6</sup> Because one operational response to positive outcomes is an increase in the sampling rate, there is the potential for endogeneity between the sampling rate and the prevalence of positive outcomes. The endogenous response should strengthen the conditional correlation rather than offset it.

side and sampling rates at the broad product-country-semester-year level by themselves and interacted with an indicator for the reform period, as well as product group, country group, and semester-year fixed effects on the right-hand-side. We find a negative and statistically significant coefficient on sampling rates after the reform, whether the reform period is defined as the six-month period immediately following implementation in July 2014 or as that period plus the entire year of 2015. Thus, our second key finding is that sampling rates were higher for products that were more likely to be compliant, suggesting poor targeting that did not improve after the reform. Two operational challenges complicate the simple empirical tests we offer. The 2014 sampling plan was implemented later than planned and 2015 was marked by a funding dispute between the FVA and the laboratories that test and evaluate the samples that resulted in a dramatic reduction in sampling. The effects of this funding dispute undermined the reform in 2015, as well as our ability to evaluate it.

It should be noted that our first and second key findings are obtained based only on the sets of FVA products that pose the largest potential harm: food of animal origin and animal feeds. The reform aimed to reduce sampling activity most among the sets of products with the least potential for harm (food of non-animal origin and food contact materials), but the data required to evaluate those effects were not suitable for evaluation. If the reform succeeded in sharply reducing sampling rates on these less risky sets of products, then it would have succeeded in ways that we are unable to measure due to data constraints.

Our ability to study the implementation of the reform makes this study novel among impact evaluations of trade facilitation reforms, but we also make an effort to ask a more standard question: what were the effects of the reform on import activity? Although trade responses to the reform might be expected to occur over a longer horizon, we estimate these responses over short horizons because the time available for evaluation is limited. Specifically, we estimate effects of a reform implemented in mid-2014 on trade outcomes in the second half of 2014 and in 2015. We also split the sample to acknowledge that 2015 results may be attributable to the funding dispute with the public laboratories as well as to the reform. We use two estimating strategies with different strengths and weaknesses and different interpretations.

The first strategy uses a difference-in-differences estimator to test for differential growth in import values among the products subject to FVA oversight, relative to other products. The effects estimated in these regressions may reflect aspects of the reform that went beyond the changes to the sampling and testing regime that are the focus of our study. The standard concerns about difference-in-differences estimators apply, but we employ a number of checks – namely the control for export supply and import demand shocks – that build some confidence in the approach. Our third key finding is that there is no statistically significant effect of the reform on import values in the second half of 2014, but modest positive effects of 3-5% higher import values for FVA products in 2015. Back-of-the-envelope calculations applied to these estimates suggest trade impacts that are roughly equivalent to those of a 1 percentage point reduction in tariffs applied

to imports of the products overseen by the FVA, which account for approximately 20 percent of FYR Macedonia's import value in a given year. Acknowledging possible threats to identification that are difficult to overcome, we offer these as our best estimates of short-run effects on imports of the cumulated shocks that include a) the FVA sampling and testing reforms, b) other reforms undertaken by the FVA at the same time, and c) the funding dispute with the public laboratories in 2015 that limited sampling activity.

The second strategy is to estimate the effects of the reform on imports relying on heterogeneity of the treatment. We link variation in the sampling rates to changes in import value at the broad product-country of origin-semester-year level, controlling for country-product and country-semester-year fixed effects. Specifications that exploit heterogeneity in sampling rates are more narrowly targeted than the differences-in-differences specifications and should provide a cleaner estimate of the effects of reduced sampling on import activity, though the resulting estimates could still conflate reduced sampling due to the reform with reduced sampling arising from the funding dispute with the public laboratories. Our last key finding is of very little evidence that reduced sampling activity alone led to increased imports during the period for which we have data. This may mean that sampling rates do not have sizable effects on imports, or it may mean that the time horizon used in the evaluation was too short to observe import responses.

Relative to earlier studies of customs reforms, we face a number of data challenges. Most customs agencies retain highly detailed electronic data on the universe of international trade shipments. The availability of these data means that studies of reforms undertaken by customs agencies can exploit the high-quality (and often confidential) data available from a single unified source. We collect and integrate data from four sources, and these data lack a common key that allows them to be easily merged. While we believe the outcomes we report are nonetheless credible, one key lesson of this study is that data available for study of technical agencies is more fragmented than those data commonly used to study reforms in customs agencies.<sup>7</sup> On the other hand, certain types of data are available for the evaluation of reforms in technical agencies that are difficult to obtain when studying customs agency reforms. One contribution of this study is to explore the data issues, many of which are likely to generalize beyond FYR Macedonia.

One operational lesson for the FVA is that the targeting procedures are not positively linked to the probability that a shipment fails to meet technical standards, as would be expected from the reform. This suggests room for improvement in the targeting strategy, especially as we also find a weak link between the samples planned and the actual number of samples taken. Even though it is to the advantage of the public in FYR Macedonia that high levels of compliance are observed in import shipments of food of animal origin and animal feeds, the risk-based methodology would aim to improve efficiency of sampling by

---

<sup>7</sup>The World Bank team supporting the reform indicates that data were also a problem in the implementation of the reform (Milchevski and Konar-Leacy (2014) describe the challenges collecting the data necessary for implementation). It is likely that several of these challenges will be common to studies of technical agencies for reasons we discuss further in the paper.

linking the probability of non-compliance with the frequency of laboratory testing. It is also likely that the weak link we observe is due to the fact that the sampling plans in 2014 and 2015 were not fully implemented due to delays and the funding dispute with the public laboratories.

The rest of the study is organized as follows: Section II describes the policy setting, beginning with a summary of SPS issues as they relate to FYR Macedonia's reforms, and then describing the specific reforms we study. Section III describes the data sources. Section IV describes the methodology. Section V presents results. Section VI provides policy recommendations. Section VII concludes.

## **Section II. Policy setting**

### ***II.1 Background***

Before turning to a description of the particular reform that we study, we first describe the international context. The FVA's adoption of sampling plans derived from an annual risk assessment exercise represents a small but significant reform within the complex and evolving systems for the oversight of food and animal health in FYR Macedonia. These systems are governed by international agreements and standards that motivate the reform. We review these briefly.

The interaction of FVA oversight systems with international trade rules falls under the World Trade Organization's 1994 "Agreement on Sanitary and Phytosanitary Measures" (the SPS Agreement).<sup>8</sup> The SPS Agreement in turn relies on standards, guidelines and codes of practice defined by three international institutions: the *Codex Alimentarius* Commission (which oversees food safety), the World Organization for Animal Health, and the International Plant Protection Convention. These organizations set international standards that apply to oversight agencies like the FVA.<sup>9</sup> The standards set by the international bodies are benchmarks, but more stringent standards can be applied if there are scientific justifications for them.

Another body of international law that is relevant for the reforms we study in FYR Macedonia is the *Acquis Communautaire* (the Acquis), which is the body of law governing the European Union (EU). FYR Macedonia aspires to EU membership, and has signed treaties with the EU that govern these efforts. Under the terms of the agreements, FYR Macedonia is reforming its laws and regulatory systems in order to comply with the Acquis. The food and animal health systems that we study relate to several chapters in the Acquis, especially Chapter 1 (Free movement of goods) and Chapter 12 (Food Safety, Veterinary and Phytosanitary policy).<sup>10</sup>

---

<sup>8</sup> The Agreement on Technical Barriers to Trade is also relevant, but we will focus here on the SPS agreement, which deals specifically with the oversight of food and animal health issues within the international trading system.

<sup>9</sup> The FVA's primary responsibilities relate to products that fall under the first two of these agreements. The Ministry of Agriculture oversees plant health issues in FYR Macedonia.

<sup>10</sup> FYR Macedonia's Stabilisation and Association Agreement with the EU entered into force in 2004. FYR Macedonia was granted candidate country status in 2005, and the EU adopted the Accession Partnership in 2008. The European Commission provides an annual report on FYR Macedonia's progress toward accession, the latest being European Commission (2016).



We refer to these larger bodies of law for two reasons. First, compliance with international standards is an important motivation for the particular reform that we study. Achieving compliance with the Acquis has been a particularly important spur to reforms of FVA oversight procedures.<sup>11</sup> Since accession to the WTO in 2003, FYR Macedonia has amended its *Food Safety, Veterinary Health, and Feed Safety* laws in an effort to comply with these larger international treaties and bodies of practice. Among other things, recent amendments to these laws provide the necessary legal backing for the reforms that we study. Second, the complex body of international agreements and law regarding international trade in SPS products highlights the existence of internationally agreed-upon best practices designed to balance trade facilitation objectives and the vital national interests related to food safety and animal health. Risk analysis and risk management approaches to oversight are central to the best practices sanctioned in the international agreements. Moreover, the agreements stipulate that oversight mechanisms should be the least trade restrictive methods that are capable of meeting national standards for animal and public health.

In light of these issues our study should be understood as more than just an attempt to establish the causal effects of a particular reform on the outcomes of interest. Our statistical investigation of the operation of sampling and testing regimes can also inform broader questions. For example, the challenges we face in acquiring data sets and making them consistent indicate weaknesses in data collection and management.<sup>12</sup> Evidence that laboratory tests rarely identify non-compliant shipments may indicate that the oversight system is more intrusive on international trade than is necessary to achieve the oversight objectives, even after the reform.<sup>13</sup> The descriptive statistical evidence that we provide about the operation of the current system can inform these questions, independent of our efforts to identify a causal impact of the reforms.

The role of border-based sampling and testing regimes within food safety systems also deserves further comment. While best-practice systems do intervene at the border, they also rely on several other mechanisms.<sup>14</sup> Mechanisms that generate private sector compliance throughout the food supply chain are

---

<sup>11</sup> See section 3.43 of World Trade Organization (2014), an official WTO review of trade policy in FYR Macedonia. This document offers a useful (but brief) summary of the legislative changes, and notes that that the FYR Macedonian authorities recognize the need for further training on the implementation of SPS measures within the relevant agencies.

<sup>12</sup> In a discussion of SPS measures and border management, van der Meer and Ignacio (2011) note that a lack of capacity in collecting and managing data is common in developing countries, and that the lack of capacities with respect to data can undermine both the public safety and trade facilitation objectives.

<sup>13</sup> The latter question is beyond our specific competencies, but we hope that our evidence is useful in informing others more qualified to make those judgements. We note that in the products studied here, agencies like the FVA play an important “monitoring” role to assure the public that food and animal health products are safe. High rates of compliance should be observed in these products because private operators have strong private incentives to ensure the safety of their products (i.e., brand value and legal liability). Acknowledging these issues as they relate specifically to food safety, one might still expect the frequency of laboratory testing to be linked to probabilities of non-compliance, especially when controlling for the type of product, which should capture most of the variation across shipments in the expected harm associated with the release of a non-compliant shipment onto the market.

<sup>14</sup> For example, the U.S. Food and Drug administration has a multi-dimensional strategy for oversight of imported seafood, available at <https://www.fda.gov/Food/GuidanceRegulation/ImportsExports/Importing/ucm248706.htm>. This strategy is notable for its use

especially important. In its technical document *Principles and Guidelines for National Food Control Systems*, Food and Agricultural Organization (2013) indicates that “food business operators have *the primary role and responsibility* (emphasis added) for managing the food safety of their products...”. In developed country markets, the responsibilities of food business operators are enforced via product recall and liability laws, among others.<sup>15</sup>

Well-functioning systems also leverage oversight by competent authorities within production facilities in the exporting country.<sup>16</sup> Even in the context of direct oversight of imported shipments, there are often operational benefits to conducting activities like sampling at inland control points (e.g., official control points inland or even at private warehouses), rather than at the physical border itself.<sup>17</sup> This is not to say that border-based sampling for the purposes of laboratory testing is illegitimate - clearly it is an important tool in the regulatory oversight kit. But a fundamental principle of international best practices is that any SPS measure should be evaluated against alternative mechanisms that would achieve the same objective while imposing a smaller burden on international trade activity. For example, in its *Guidelines for Food Import Control Systems*, Food and Agricultural Organization (2003) indicates that “(o)perational procedures should be developed and implemented to minimize undue delays at the point or points of entry without jeopardizing effectiveness of controls to meet requirements.”

Useful background reading on SPS issues as they relate to border management can be found in van der Meer and Ignacio (2011) and Cadot et al. (2011).<sup>18</sup> Van der Meer and Ignacio (2011) highlight a number of weaknesses that are common to developing country SPS oversight of imported products. These include (but are not limited to):

- a) a lack of systematic data gathering and an absence of risk profiles,
- b) a bias in interventions toward revenue generation from fees and informal payments,
- c) duplication of tasks and data gathering at the border by customs, quarantine agencies and border police,
- d) poor coordination of border processes and time consuming sequential processes,
- e) higher inspection rates than is necessary because of poor risk management,
- f) unnecessary testing, inspection and disinfestation treatment costs.

---

of multiple control points, and the use of information from those control points to improve risk predictions in the electronic risk management system used in border management, PREDICT.

<sup>15</sup> Laws that induce voluntary compliance in the private sector are valuable for food safety objectives because food safety risks may arise after border officials have tested the goods and cleared them for release onto the market.

<sup>16</sup> Usually this is accomplished via mutual recognition of inspection activities in foreign facilities. In some cases, authorities in large markets have negotiated the capacity to themselves conduct inspections of foreign facilities.

<sup>17</sup> The World Bank’s technical expert that accompanied us on our mission to Macedonia noted that it was likely that much of the oversight activity undertaken at the border by the FVA could be moved inland without compromising food safety objectives. The primary exceptions would be inspections for pests and animal diseases, which should occur at the border itself.

<sup>18</sup> Widdowson and Holloway (2011) discuss the role of risk-management at international borders in a setting more general than SPS issues.

Van der Meer and Ignacio make these statements about developing countries in general. It is quite likely that FYR Macedonia outperforms most developing countries along these and other dimensions.<sup>19</sup> That said, a case could be made that each of the above weaknesses is present to some degree in FVA oversight. On point (b), it is notable that the FVA's annual reports on its import control activities typically provide documentation on funds received through its import control activities before discussing incidents of successful identification and control of non-compliant imports. Our evaluation will offer some suggestive evidence on point (e), though we study sampling activities rather than inspection.

## II.2 Specifics of the reform

The reforms we study were undertaken as part of the Western Balkans Trade Logistics project under the auspices of the International Finance Corporation (IFC) of the World Bank Group. The IFC worked with two technical agencies with responsibilities over international trade, the Food and Veterinary Authority (FVA) and the State Sanitary and Health Inspectorate (SSHI). Both agencies implemented the reform, but only the FVA had data that are suitable for use in our evaluation. The FVA describes its role thusly:

“(The FVA) is an independent body of the Government responsible for activities in the field of food safety and animal feed, implementation, control, supervision and monitoring of veterinary activities in the field of animal health welfare, veterinary public health and control of national reference and authorized laboratories that provide adequate support for the Agency.”<sup>20</sup>

This mandate includes oversight over many different aspects of the food safety and animal health systems in FYR Macedonia. One of these responsibilities is that the FVA has oversight of imported products that fall under its mandate. Our study focuses on oversight of imports.

In the course of its duties, the FVA issues three import licenses that trading firms must acquire in order to import or transit certain products.<sup>21</sup> These types of products covered by each license are: 1) Food of animal origin (FVA license N853), 2) Food of non-animal origin and food contact materials (FVA license I007), and 3) Animal feeds (FVA license I047). FVA activities in the license-issuing process include the stationing of FVA staff at international border posts. The duties of these staff include the taking of samples

---

<sup>19</sup> For example, European Commission (2016), the EU Commission's progress report on FYR Macedonia's candidacy indicates “*some level of preparation* in the areas of food safety, veterinary and phytosanitary policy.” It appears that the primary EU concerns in these areas relate to the safety of FYR Macedonia's production processes, which may affect the safety of food and other exports to the EU. Nonetheless, the EU does sometimes comment on FYR Macedonia's management of SPS issues with respect to its own imports.

<sup>20</sup> [http://www.fva.gov.mk/index.php?option=com\\_content&view=article&id=364&Itemid=421&lang=mk](http://www.fva.gov.mk/index.php?option=com_content&view=article&id=364&Itemid=421&lang=mk), downloaded March 23, 2017. The Google Chrome browser will translate the page into English.

<sup>21</sup> We use the word license throughout our study, as this is the direct translation from the Macedonian word associated with the relevant documents. It should be noted that this is not an import license/permit as conventionally understood or applied, but rather an official electronic notification that goods have passed appropriate testing and may be released onto the market. Final release onto the market may depend on other criteria, such as payment of customs duties. We do not observe decisions made outside of the FVA.

from import shipments, and sending the samples to a laboratory for testing.<sup>22</sup> The testing process is contracted out to public laboratories, which test for the presence of bacterial and other pathogens as well as biological and chemical toxins. This joint process, known as sampling and testing, is the focus of our evaluation.

The IFC project had identified lengthy and intrusive sampling and testing regimes in the technical agencies as a significant problem for trade in FYR Macedonia. Very rarely did the laboratory test activities identify non-compliant shipments, suggesting poor targeting and/or excessive sampling.<sup>23</sup> In the data we have on sampling and testing activities (described below), we find that in the pre-reform year 2013, samples were taken from nearly 1 in 10 shipments, and that less than 0.2 percent of the samples taken were found to be non-compliant. Since sampling occurs mid-shipment (at the border) and is subject to queuing and the availability of FVA staff, the sampling regime imposes costs in terms of time delays during the movement of goods.<sup>24</sup> The longer time delays associated with the wait for laboratory approval prior to release of the goods onto the market are likely to be the primary source of trade costs for import activities.

The goal of the reform was to reduce costs of importing through the adoption of a risk-based approach by the FVA to sampling and testing imported products. Beginning in 2013 FVA representatives underwent training in various components of risk-based approaches to the oversight of imports. Under the guidance of IFC staff, FVA representatives went through a process of assessing the level of risk associated with each product imported into FYR Macedonia (e.g., what would be the scope of harm associated with non-compliance) and a review of past rates of non-compliance. The objective of this process was to focus their sampling and testing efforts on the products that were most dangerous if non-compliant and on product-country pairs that were most often non-compliant. Focusing the activities on risky shipments would allow the agency to retain or improve enforcement while reducing the overall rate of sampling and testing across all shipments.

On the basis of the review of past rates of non-compliance, the Rapid Alert System for Food and Feed (RASFF) system alerts, and EU requirements, the FVA set out a sampling plan (sometimes called a monitoring plan), which reports the number of shipments that were planned to be sampled for laboratory

---

<sup>22</sup> We observed this process on a field visit to a major border crossing in FYR Macedonia's north (Tabanovce). A truck filled with boxes of frozen vegetables was opened, and a single box selected for testing in the laboratory. The truck was then sealed and allowed to proceed inland for customs control. The products could not be released onto the market until the laboratory tests had been completed.

<sup>23</sup> Note that there are less intrusive methods of evaluation than laboratory testing (e.g., physical inspections) that can be used to screen goods before requesting laboratory analysis. There are also other subsequent points of control as well as recall and liability laws that serve to back-up these systems.

<sup>24</sup> In many developed countries, the food safety risks that are the object of this control regime are handled primarily via recalls. Developed country systems are less dependent on testing overall, and have more sophisticated approaches to targeting than those applied in FYR Macedonia.

testing at each border crossing in a year.<sup>25</sup> Expected sampling rates were allowed to vary across products and across origin countries or regions. While the plan specified the expected numbers of samples to be taken, FVA border agents were to continue having discretion in choosing the shipments that were sampled in terms of meeting the plan. Border agents also retained the authority to deviate from the plan and sample from any shipment if the circumstances justify it. Hence, the sampling plan was not entirely prescriptive, though border agents were expected to follow the plan to the extent possible.

The specific aspects of the reform that we study relate to these changes. Concretely, the direct outcomes that were meant to be achieved by the reform were:

- 1) Focusing sampling activity on the shipment types that were most harmful, and
- 2) Reducing the number of shipments that were sampled for laboratory testing.

It was expected that these reforms, if successfully undertaken, would reduce the costs for firms engaging in international trade. Economic theory suggests that if the costs of undertaking imports were reduced, the reform would lead to increases in international trade volumes. We are not in a position to study the effects of the reform on direct measures of trade costs, which are hard to capture in a single measure and would vary substantially across many dimensions of the traded products, the trading firms and modes of transportation. The availability of monthly trade data reporting import value with country and product detail allows us to study short run impacts of the broader reform, and of reduced sampling activity more specifically, on imports. The behavioral responses to the reform could occur through at least two channels: a) pre-announced reductions in sampling rates along with rational expectations by firms about the probability that they would be sampled generate more imports when sampling is less likely, or b) learning about the new lower sampling rates within the year and reasonably rapid updating of prior beliefs.<sup>26</sup> Delayed release onto the market that is associated with lengthy laboratory testing procedures is the primary source of trade costs that are avoided when sampling and testing activity is reduced.

### *Caveats*

As we detail in Section III and in Appendix A, the data on laboratory tests is weakest for imports of products covered by the license governing imports of food of non-animal origin and food contact materials. These products are the least likely of the FVA products to harm consumers. The reform was designed to have the most impact on sampling and testing activities in these products. Unfortunately, this means that most of our conclusions are limited analysis to the set of products with the fewest (planned) changes in sampling activity. This limitation should be kept in mind in interpreting our results.

---

<sup>25</sup> The EU's RASSF notifies governments about recent threats in the area of food safety and animal health. The 2015 annual report of RASSF is available here: [https://ec.europa.eu/food/sites/food/files/safety/docs/rasff\\_annual\\_report\\_2015.pdf](https://ec.europa.eu/food/sites/food/files/safety/docs/rasff_annual_report_2015.pdf).

<sup>26</sup> These channels are very difficult to pull apart, even with detailed shipment level data from a customs agency, which we lack here.

The effort to train the FVA agency representatives on the principles of risk management and produce the sampling plan began in the Spring of 2013. The first sampling plans were produced for use on January 2014, but implementation was delayed until July 2014. An updated sampling plan was produced for 2015. Our investigation of the effects of the FVA reforms includes data through the end of 2015.

A subsequent crucial event that is relevant for our evaluation is a funding dispute between the FVA and the laboratories that conduct the testing. Due to postponed public procurement for a certain sub-period within 2015, the FVA did not have a contract with the Veterinary Institute where the analysis for food of animal origin and animal feeds are conducted. This dispute limited FVA's use of laboratory testing facilities in 2015. We observe in our data a large drop in the number of samples sent to the laboratory for testing in 2015. Reduced sampling activity was a key objective of the reform, but a large share of the dramatic drop in sampling activity in 2015 is due to the funding dispute, not the reform. European Commission (2015) indicates that payment arrears jeopardized the FVA's ability to implement animal health programs. We lack further specific information on this issue, but we note that reduced sampling and testing generated by this dispute helps achieve some goals of the reform (lower sampling rates), while undermining others (targeting risky shipments). In our evaluation, we will sometimes include 2015 data in the estimation process, and other times exclude it. The funding dispute with the public laboratories makes it very difficult to identify separate effects of the reform (except the immediate impacts visible in the second half of 2014). Our estimates that include 2015 data should be interpreted as due to the joint effects of reform and the funding dispute with the public laboratories. We will discuss the implications of the funding dispute as we implement each portion of the evaluation.

The slow implementation of the reform, the funding dispute with the public laboratories, and the budget cycle funding for our evaluation mean that we are only able to investigate the short-run effects of the reform. As will be described in Section III, we have daily data on the timing of shipments and monthly data on import values so we are able to construct 6-month semesters and estimate impacts over the second semester of 2014 that follow the implementation of the sampling plans. We are also able to investigate sampling rates, measured compliance and import value for the entirety of 2015, but these estimates are conflated with the effects of the funding dispute with the public laboratories. The reforms were difficult to implement and so one might consider the latter half of 2014 to be a transition period. Our inability to cover a longer period in the evaluation of the reform is unfortunate. Trade facilitation reforms in the developing world frequently occur in environments associated with economic shocks and implementation difficulties, which is one reason why so few studies exist.

### **III. Data**

In this section, we describe briefly four sources of FYR Macedonia data used in the analysis (full details are provided in Appendix A). The data sets each offer detailed information along multiple dimensions (e.g., precisely defined import products, country of origin, date of arrival, etc.). Precision along multiple dimensions allows credible matching across data sets, even though the data lack unique identifiers that would tie the data sets together at the level of an individual import shipment.

The first data set provided by the FVA includes information on the annual sampling plans for 2014 and for 2015, which are the key outcome of the FVA reform and offer strong guidance to FVA border agents regarding the number of certain types of shipments to be sampled during the upcoming year. The shipment types are defined in terms of broad product groups, countries or country groups (e.g., CEFTA countries), the type of import license and the hazard for which the sample is to be tested.<sup>27</sup> For example, the 2014 plan calls for the taking of 12 samples from the product category *bovine meat, fresh or frozen* originating from EU or European Free Trade Agreement (EFTA) countries at the border post of Tabanovce for the purposes of testing for the hazards of *Salmonella*, *E.coli*, and antimicrobial substances. Some of our other data are not available at the level of the border post nor by the hazard, thus for each year we aggregate the sampling plan data by summing over border post and hazard to produce the planned number of shipments to be sampled at the level of a broad product group and a country or country group.

The second data set provided by the FVA, which we designate *the FVA lab sample data set*, includes information on the shipments from which FVA border agents took samples for the purpose of laboratory testing, along with the outcome of the tested samples for the period 2013-2015. Each row of this data set represents an import shipment from which samples were taken to the laboratory and for which information is recorded on product, name of the border post where it entered, date when the sample was taken for laboratory testing, country of origin, and health or other hazards tested in the laboratory. Since the raw data contain verbal descriptions of products that are not always consistent over time, these had to be aggregated into broad product groups, as will be described below. We use the FVA lab sample data set to calculate the number of shipments with samples taken to laboratory at the broad product group-country of origin-month-year or more aggregated levels. Combining these numbers with the numbers of import shipments from the EXIM data set described next allows calculation of sampling rates (i.e., the share of import shipments from which a sample is taken/sent to a laboratory for testing). We also use the FVA lab sample data set to generate an indicator variable at the shipment level identifying import shipments with samples taken to laboratory that had positive results/irregularities (i.e., that are non-compliant with FVA's technical standards). We calculate the number of samples taken to laboratory that are non-compliant at the

---

<sup>27</sup> The breadth of product and country groups is operationally useful because it allows some discretion by border agents over products that (presumably) pose similar risks. While more disaggregation would be useful for our analysis, it is not built into the sampling plan so we must conduct much of our analysis at this higher level of aggregation.

broad product group-country of origin-month-year, or at more aggregated levels. This allows to calculate the rate of non-compliant import shipments at those various levels of disaggregation.

The FVA lab sample data set suffers from two major weaknesses. The first weakness is that for two groups of products - food of non-animal origin and food contact materials - the data set records information *only* on shipments from which laboratory test results were positive for some hazard. This means that the data lack information on the total number of samples taken for those products, which precludes the calculation of sampling rates and rates of non-compliance. Therefore, our analysis of sampling focuses exclusively on the remaining sets of FVA products: food of animal origin and animal feeds. The second weakness is that the products we could study were sampled much less frequently in 2015 than in previous years. The raw total number of FVA samples taken from imports of food of animal origin and animal feeds was 1,487 and 1,235, respectively in 2013 and 2014, but just 63 in 2015. The reason for this decline was the funding dispute with the public laboratories described in Section II.2

The third data set provided by the FVA is the import shipment data from the single window for import, export and transit of goods and tariff quota (EXIM) for the products covered by licenses I007, I047, and N853 over the period 2013-2015. The EXIM data set records each license application, which has a firm identifier (which was not made available to us due to confidentiality concerns), a tariff line variable, product descriptions, volume imported (and unit of measurement), date of application, and country of origin.<sup>28</sup> We use the EXIM data set to calculate the total number of FYR Macedonian import shipments by broad product group-country of origin-month-year, or at more aggregated levels. The numbers of import shipments are combined with the number of samples taken to the laboratory (discussed above) to calculate sampling rates of import shipments. After a few data problems described in Appendix A are addressed the cleaned EXIM data set imports of food of animal origin and animal feeds (requiring licenses I047 and N853) used in our analysis covers 40,212 shipments over the period 2013-2015.

The fourth data set containing comprehensive data on the universe of imports by FYR Macedonia (not only those imports that require a license from the FVA) was obtained from the State Statistical office (Makstat). The Makstat data set available for the period 2009-2015 includes import values (measured in US dollars) and volumes (measured in kilograms) by month, HS 10-digit code, and border post of entry.<sup>29</sup> Makstat provides highly credible measures of import flows, as Makstat simply processes the data provided by the Macedonian customs agency.<sup>30</sup> Hence, the Makstat data set is used to measure our key outcome

---

<sup>28</sup> Note that applications for licenses are lodged at the border post when the goods have already been loaded and shipped to FYR Macedonia. We therefore operate under the assumption that license applications from EXIM represent real shipments.

<sup>29</sup> The ratio of FYR Macedonia's total imports in the Makstat data set to FYR Macedonia's total imports in the main international source of trade data, WITS, is very close to 100% in every year from 2009 to 2015.

<sup>30</sup> We requested more detailed data from Macedonian customs, but they were not willing to share it. Among other shortcomings, this means that our trade data lack firm-level information as well as information on the inspections undertaken by Macedonian customs. Separate inspections by the customs or other technical agencies would represent duplicative procedures of the kind that van der Meer and Ignacio (2011) describe as weaknesses common to developing country customs agencies. From an econometric



variable: import value. For the bulk of our analysis we will aggregate the raw Makstat data to measure that outcome variable at the HS 4-digit-country of origin-semester-year level. The HS 4-digit level of aggregation is chosen to be able to merge the Makstat data set with the list of products under the control of the FVA, which is available at the HS 4-digit level. We will designate products that underwent the FVA reform as treated products for all difference-in-differences and pre-treatment common trends testing regressions. The number of observations per year and some additional basic descriptive statistics based on the Makstat data set are provided in Appendix Table A1. The Makstat data set provides outcomes for 4.5 years of data prior to the reform and 1.5 years after the reform. The Makstat data are aggregated up to the HS4-digit level and concorded to the list of HS4 products overseen by the FVA.

A key challenge to conducting a more detailed analysis of the FVA reforms in terms of sampling and non-compliance is the different product groupings and verbal descriptions that are used to categorize products in the three other data sources: sampling plans, FVA lab sample data set, and EXIM data set. The only way to obtain a data set encompassing the information from the three sources is to define a set of broad product groups that serves as a relatively aggregated common denominator for the three data sets. The list of broad product groups that we define in order to encompass those various product categories and verbal descriptions is shown in Appendix Table A2. We construct several concordances to link across the various data sets, as detailed in Appendix A. Using those concordances, we build a data set which merges across the FVA lab sample data set, the EXIM data set, and the Makstat data set at the broad product group-country of origin-semester-year level, which we use to measure sampling rates and non-compliance rates for the analysis of sampling and compliance in Sections V.1 and V.2. As noted before this data set covers only information corresponding to food of animal origin and animal feeds. The numbers of observations in that merged data set are shown in Appendix Table A3.

An additional source of data we use to construct control variables are the export and import flows from WITS at the annual and monthly levels. We use total exports to the world (except to FYR Macedonia) at the exporting country-HS 4-digit-year level or at the exporting country-HS 4-digit-semester-year level as our foreign supply shock variables and total imports from the world (except from FYR Macedonia) in for each of the four neighbors of FYR Macedonia - Albania, Bulgaria, Greece, and Serbia - at the HS 4-digit-year level or at the HS 4-digit-semester-year level as our import demand shock variables at the HS 4-digit-year and the HS 4-digit-month-year levels, as will be discussed in Section IV.3.<sup>31</sup>

---

point of view, our lack of information on inspections by other agencies will raise the possibility of omitted variables bias. We are unable to address this because we lack the specific data necessary to control for those actions.

<sup>31</sup> WITS does not report data for Kosovo, another neighbor of FYR Macedonia, thus we focus only on the other four neighbors. As described in Appendix A, WITS data at the country-HS 4-digit product-month-year level is available only from 2010 onwards and for a subset of reporting countries, relative to the data at the country-HS 4-digit product-year level. Thus, our estimating sample will be smaller in specifications that include the shock variables at the monthly (aggregated to the semester) level.

## **Section IV. Methodology**

In this section, we describe our empirical strategy. Neither the FVA reform nor the data lend themselves directly to clean tests of causal mechanisms, but we are able to use a rich set of fixed effects and a rich set of variables to control for many (if not all) sources of plausible confounding variation. Given the complete absence of studies in this area, we view this as an important step forward for the literature. We are also able to study various phenomena that have not been studied before, including the conformity of sampling activities on the ground to the sampling plan and the empirical study of the rates of non-compliance.

### ***IV.1 Adherence to sampling plan***

Our first empirical exercise studies the implementation of the reforms. In the new system, the FVA conducts a risk analysis using available evidence on the frequency of non-compliant shipments and the severity of the harm that would occur if a non-compliant shipment would be allowed onto the market. The main outcome of this process is the sampling plan indicating the number of shipments to be sampled at each border post for groups of products originating from certain source countries or regions in a given year. In principle, the plan represents the informed view of the FVA administration about levels of risk of non-compliance and the degree of harm, combined with resource constraints as well as the obligations of FYR Macedonia under the SPS agreement and its other trade agreements. If the FVA administration accurately forecasts numbers of import shipments and levels of non-compliance in each group, one should expect variation across products and regions in samples actually taken to match variation in the sampling plan. Border agents have discretion over which samples they take, but the sampling plan should offer strong guidance as to the number of samples taken within each category.<sup>32</sup>

There are perfectly legitimate reasons not to follow the sampling plan in certain circumstances. FVA border agents have discretion, and they are empowered to deviate from the plan when individual shipments are deemed worthy of further scrutiny. During the year, there may also be news, such as a finding of aflatoxin or salmonella in particular imported products from a given source region, which would warrant additional sampling activity for those products if imported from that source region.<sup>33</sup> It may also be that the number of import shipments differs from what is expected or that new products or shipments from new source countries are being imported and therefore the border agent is obliged to take samples from that shipment.<sup>34</sup>

---

<sup>32</sup> One would also hope that FVA border agents would select riskier shipments within each category in the sampling plan. The empirical implications of this behavior for our statistical tests are not immediately obvious. We find very low rates of non-compliant shipments overall, which probably indicates that shipments are generally compliant.

<sup>33</sup> The EU's RASSF database reports the identification of hazardous samples anywhere in the EU. Such reports are likely to trigger increased scrutiny of shipments with the same characteristics as the shipments identified in the database.

<sup>34</sup> Taking samples from new shipments is a recommendation set forth in the FVA sampling plan.

There may also be illegitimate reasons for border agents to deviate from the sampling plan. Any management, coordination or recordkeeping problems might lead to deviations from the plan. Border agents tasked with sampling a certain number of shipments may over- or under-sample in early parts of the year and miss the target. In principle, petty corruption schemes could also cause deviations from the plan.

One way or the other the degree to which actual sampling activity adheres to the plan is an interesting empirical question. The sampling plan is the output of a careful process designed to balance resources and risks, and represents the optimal plan, *ex ante*, as determined by experts in the FVA central administration. Its connection to sampling activity on the ground is critical for understanding the degree to which the reforms were implemented as had been planned.<sup>35</sup> In order to evaluate this we estimate a series of regressions with sampling activity as dependent variable and planned samples as explanatory variables. Let  $SP_{kit}$  represent the number of import shipments to be sampled according to the sampling plan for product group  $k$ , region/group of countries  $i$ , and year  $t$ , while  $S_{kit}$  represents the number of shipments actually sampled for the same product group - region/group of countries - year triplet. Our strictest test of the plan is an ordinary least squares (OLS) regression that takes the form:

$$1) S_{kit} = \alpha_0 + \alpha_1 SP_{kit} + \varepsilon_{kit}$$

where  $\alpha_0$  is the regression constant,  $\alpha_1$  is the coefficient on the planned number of samples, and  $\varepsilon_{kit}$  is a normally distributed error term. If border agents adhere strictly to the sampling plan the coefficient estimates from this regression would be  $\alpha_0 = 0$  and  $\alpha_1 = 1$ , and the regression as a whole would produce an R-squared of 1.<sup>36</sup> Deviations from this coefficient pattern are useful for understanding the ways in which border agents are deviating from the sampling plan. For example,  $\alpha_0 > 0$  would indicate that the border agents are over-sampling in a non-discriminating way,  $\alpha_1 > 1$  could mean that border agents are over-sampling shipments deemed risky by the FVA administration, while  $\alpha_1 < 1$  could suggest an under-sampling of risky shipments. Of further interest is the manner in which deviations from the sampling plan occur. For example, we can replace the constant in Equation (1) with product and/or region/group of countries-specific or year fixed effects. These exercises can inform tests for a common constant across the members of each group. The exercises could also isolate an across-the-board deviation from the sampling plan in a given product group.

## ***IV.2 Testing compliance***

---

<sup>35</sup> Treating the plan itself as the ‘reform’ is somewhat reductive, but it is the aspect of the reform that lends itself most cleanly to evaluation. The reorientation of the FVA towards methods that incorporate assessment and communication of risks involved substantial training, legislative changes, and more. Here we are simply evaluating the link between the FVA administration’s plan and the sampling activity, one of the latest stages of the reform.

<sup>36</sup> This pattern of statistics would also require a perfect forecast by the FVA of the annual number of import shipments of each product group from each country group.

In our second exercise we consider an outcome that has not yet been investigated in the literature on impact evaluation of trade facilitation reforms: rates of compliance with import regulations. Specifically, we investigate the outcomes of laboratory tests ordered by the FVA. One of the objectives of the risk-based sampling reform was to more closely align testing activity with the likelihood of non-compliance. We design econometric tests of this hypothesis.

Let  $I_{kits}^{NC}$  be an indicator variable that takes the value of one if a sample taken from shipment  $s$  of product group  $k$  from country  $i$  at time  $t$  is found to violate FVA standards and zero if it is compliant.<sup>37</sup> For example, *salmonella* might be found in a shipment of imported chicken. We shall call  $I_{kits}^{NC}$  a non-compliance indicator. Next, we calculate sampling rates  $SR_{kit}$  as the ratio of the number of shipments that were sampled to the number of shipments arriving for import clearance at the  $kit$  level. Under the principles of risk management, sampling rates should be higher when probabilities of non-compliance are higher.<sup>38</sup> Our initial model takes the form:

$$2) I_{kits}^{NC} = \alpha_0 + \alpha_1 SR_{kit} + \varepsilon_{kits},$$

to which we add product group, country of origin, and time-period fixed effects in succession. Our initial test simply asks whether the probability of non-compliance is higher when the sampling rates are higher, i.e.,  $\alpha_1 > 0$ . Sampling rates can vary across product groups for reasons unrelated to the probability of non-compliance. Most notably the level of harm can vary substantially across products, and we use product group fixed effects to capture this variation. It could also be that products imported from distant countries or countries with low levels of per capita income are more likely to be non-compliant, and the country fixed effects are intended as controls for such variation. Time-period fixed effects may soak up time-based variation in threats, though the source of such variation is not as obvious as in the case of product group and country effects. We estimate the model on pre-reform data, i.e., up to June 30, 2014.

In order to understand whether the FVA reform affected relationships between non-compliance and sampling rates we estimate a model with interaction terms. We augment Equation (2) in the following way:

$$3) I_{kits}^{NC} = \alpha_0 + \alpha_1 SR_{kit} + \alpha_2 D_t + \alpha_3 SR_{kit} D_t + \varepsilon_{kits}$$

where  $D_t$  is an indicator variable that takes the value of 1 in the period from July 2014 onwards when the reform is in place and 0 otherwise. An effective reform should have strengthened the relationship between

---

<sup>37</sup> In principle, more than one sample can be taken from a single shipment, but our data do not specify systematically the number of samples taken from each shipment. Thus, we use only the information on whether or not the shipment was subject to sampling and laboratory testing.

<sup>38</sup> A literature on racial profiling in policing considers a similar exercise when testing for racial discrimination in police search activity. Knowles et al. (2001) build a model in which compliance rates are responsive to policing, and propose an “outcome test” which posits police bias if the searches find drugs in the same proportions across races. The focus is on outcomes rather than search activity. The assumption of fully responsive compliance rates is probably not suitable for our exercise (i.e., food spoilage may be mechanically higher for shipments that travel larger distances). We focus instead on the question of whether or not inspectors are searching more among groups of shipments that are more likely to be non-compliant. This is one of the goals of the reform effort, and it may or not be true of inspection activity prior to reform.

non-compliance rates and sampling rates, which we test by evaluating the hypothesis that  $\alpha_3 > 0$ . We conduct this test controlling for product group and country of origin fixed effects.

The need to control for various sets of fixed effects combined with the very small number of shipments found to be non-compliant (as will be shown in Section V) leads us to estimate Equations (2) and (3) by OLS as linear probability models, acknowledging the shortcoming that the estimated predicted probabilities may not be meaningful since they can lie outside of the  $[0,1]$  interval. But this is not a limitation for the qualitative predictions from the regressions which are our focus.<sup>39</sup>

One important challenge for Equation (2) and (3) tests is the endogeneity of the sampling rate. A responsive border agency will react to evidence of non-compliance by raising the sampling rates, even within the current time period. This is a feedback effect that would strengthen the coefficient on  $\alpha_1$  in cases where the non-compliance rate is unexpectedly high.<sup>40</sup> Another possibility for an increase in sampling activity in a given product group-country pair could be due to a notification of non-compliance somewhere else in the world, for example via the EU's RASSF. We will estimate Equations (2) and (3) excluding some products where such feedback effects might be stronger and using lagged sampling rates to address this endogeneity concern.

### *IV.3 Effect on imports*

One of the central questions in the literature on impact evaluation of trade facilitation reforms concerns the effects of reforms on trade. Even in the enviable situations where complete and integrated data sets are available from a customs agency, estimating this effect is challenging. In cases like ours, where an integrated data set is not available, the challenge is even greater. The policy reform that we consider lacks the specific aspects that are useful for evaluation (such as randomized selection of shipments for inspection). While the identification of trade effects of this reform is considerably more difficult, we undertake it anyway because reforms of this kind have not been studied in the literature and our evidence can begin to fill in this important gap.

Our baseline model is a simple difference-in-differences (D-i-D) model that relies on variation across products in their exposure to the reform. Using the Makstat data, we ask whether the products for which the FVA has oversight responsibilities saw faster import growth than other products in the period after the FVA reform, relative to the period before the reform. Consider as a baseline, the following regression model:

$$4) \ln(M_{kit}) = \alpha FVA_k D_t + \gamma_{ki} + \gamma_{it} + \varepsilon_{kit},$$

---

<sup>39</sup> We were unable to obtain results for Equations (2) and (3) defined as logit regressions due to convergence problems.

<sup>40</sup> Note that a within-year feedback effect would be the correct management response by the border agency. Insofar as we are evaluating the link between non-compliance rates and sampling rates as a correlation, the feedback effect is not a problem so long as we do not interpret the coefficient as a causal effect of sampling rates on non-compliance rates.

where  $M_{kit}$  is a measure of the value of imports of HS 4-digit product  $k$  from country  $i$  at time  $t$ ,  $FVA_k$  is an indicator that product  $k$  is overseen by the FVA,  $D_t$  is an indicator variable that takes the value of 1 in the period from July 2014 onwards when the reform is in place and 0 otherwise, and  $\gamma_{ki}$  and  $\gamma_{it}$  are fixed effects at the product-country and country-year levels, respectively.<sup>41</sup> This equation serves as a D-i-D specification because the  $\gamma_{ki}$  term sweeps out the  $FVA_k$  fixed effects, and  $\gamma_{it}$  sweeps out the  $D_t$  fixed effects.<sup>42</sup> The coefficient of interest is  $\alpha$ , which is positive if the effect of the FVA reforms on imports of FVA products is positive.

A first key challenge in estimating a specification like Equation (4) is the question of whether or not the non-FVA products are suitable as a control group for the FVA products. Ideally, we would have a set of products with similar characteristics that are not subject to control by the FVA. Such products could serve as the control group against which we could evaluate developments in the FVA products. But because the FVA has comprehensive authority over plant and animal-based food products as well as animals and feed (and none over most non-food manufactured goods), there are very few products that are near enough in product characteristic space to serve as credible controls.<sup>43</sup> One exception concerns the set of products known as food contact materials, which are subject to FVA import license I007. Food contact materials (cutlery, food packaging materials, etc.) have near neighbors in product characteristic space that are not subject to oversight by the FVA because they do not regularly come into contact with food. We estimate Equation (4) for the entire set of imported products, but also for a narrower set of food contact materials and the control group of neighbor products (defined as HS 4-digit products belonging to the same HS 2-digit chapters as those of FVA's food contact materials).

A second key challenge with Equation (4) is that its estimate of  $\alpha$  will be biased if imports at the  $kit$  level are subject to time-varying shocks that are correlated with treatment status (i.e., with  $FVA_k D_t$ ). The most plausible sources of these shocks would be export supply shocks in foreign countries. In order to address this challenge, we exploit global data on international trade flows from WITS to measure these shocks directly. We construct for each origin country's its world export supply of product  $k$  at time  $t$ ,  $WES_{kit}$  (where world export supply does not include exports to FYR Macedonia), following Hummels et al. (2014). Because FYR Macedonia is a small country, developments in FYR Macedonia are very unlikely

---

<sup>41</sup> One might expect to see product-country of origin time-varying tariff rates enter into Equation (4) as a determinant of imports. Tariffs applied to imports of FVA products are remarkably stable over our period of study. Less than 1.5 percent of HS-10 digit tariff lines saw changes in MFN tariffs during the period of our sample. Given the inclusion of product-country of origin fixed effects in Equation (4), it is not useful to add these virtually stable tariffs as controls.

<sup>42</sup> Equation (4) is a more stringent specification than a more standard D-i-D equation which would include in levels the two variables that enter the interaction term,  $FVA_k$  and  $D_t$ .

<sup>43</sup> In other settings, the structure of international trade statistics can offer a quite good set of controls. For example, Hillberry and McCalman (2016) study antidumping orders in the United States using untreated products within still rather narrow product groups (HS4) as controls for products within the same HS4 that become subject to an antidumping order.

to affect export supply activity in the origin countries of its imports.<sup>44</sup> Hummels et al. (2014) employ WES as an instrumental variable, but in our case we simply include it as a control in our D-i-D regressions. Its role is to soak up variation in FYR Macedonia's imports across products and origin countries that arise through developments abroad.

Another potential threat to identification are import demand shocks in FYR Macedonia that are correlated with the FVA reform episode. A key source of import demand shocks would be shocks to national income, but is not clear if these would differentially affect FVA products.<sup>45</sup> Expenditure data is not available at anywhere near the level of disaggregation of the Makstat data set, so we are not able to address shocks to expenditures. One quite plausible source of import demand shocks that we can make some attempt to control for are domestic supply shocks. A positive supply shock in a narrowly defined product would likely reduce import demand for that product, and vice versa. Any available data on production would also be at much more aggregated level than the Makstat data set, so we attempt to use international trade statistics from WITS to construct control variables, as we did for the export supply shocks. FYR Macedonia's export statistics might well be informative of domestic supply considerations, but these data are possibly affected by the FVA reform and thus not useful to address this threat to identification.<sup>46</sup> Hence, we address the possibility of domestic supply shocks driving import demand shocks taking a plausible, if unorthodox, strategy. We note that most of the products made in FYR Macedonia that can plausibly replace imports over a short time horizon would be commodities and food and agricultural products. The most plausible source of shocks for these products would be weather-related shocks. Since FYR Macedonia is a very small landlocked country, it shares its major weather patterns with its neighboring countries. We therefore use fluctuations in imports of each of the countries that share a land border with FYR Macedonia as control variables to soak up variation in FYR Macedonian imports that would occur through this channel.<sup>47</sup> As with the  $WES_{kit}$  variable, we exclude each neighbor country's imports from

---

<sup>44</sup> There is a slight possibility that developments in FYR Macedonia could affect the WES variables of its neighbors. In robustness checks we show that our results are maintained when we exclude from the estimating sample the countries that share a land border with Macedonia.

<sup>45</sup> In other settings, one might worry that international supply chains are organized so that very specific inputs would be imported into FYR Macedonia for further processing, and that shocks to the processing firm in FYR Macedonia would produce shocks to import demand for very specific inputs. Supply chain trade of this kind is not a significant part of FYR Macedonian imports, so we are not overly concerned this possibility. As we describe next, the inclusion of import demand shocks of FYR Macedonia's immediate neighbors should do a reasonable job of controlling any remaining variation so long FYR Macedonia's role in particular supply chains is not too different from one or more of its neighboring countries.

<sup>46</sup> If the FVA reform makes it easier to import milk, then Macedonian cheese exports might surge, for example.

<sup>47</sup> The countries are Albania, Bulgaria, Greece, and Serbia. Kosovo is another neighbor country but cannot be covered as it does not report data to WITS. Note that the variables we include might also be successful in soaking up other sources of transitory import demand shocks that we have not listed. For example, if there were a sudden increase in the demand for a fashion product from an idiosyncratic location, it is plausible that this demand shock would be felt in some of the neighboring countries at the same time. One-quarter of the population in FYR Macedonia is ethnically Albanian, for example, which may mean that fashion trends in Albania are correlated with fashion trends in FYR Macedonia.

FYR Macedonia when calculating its import demands. We label these variables  $WID_{kjt}$ , where  $j$  designates the subset of neighbor countries  $i$  that share a land border with FYR Macedonia.

With these control variables, we rewrite Equation (4) to generate our primary D-i-D estimating equation for observing trade effects of the FVA reform as:

$$5) \ln(M_{kit}) = \alpha FVA_k D_t + \gamma_{ki} + \gamma_{it} + \beta_{ES} WES_{kit} + \sum_{j \in J} \beta_{IDj} WID_{kjt} + \varepsilon_{kit}.$$

Once again, our coefficient of interest is  $\alpha$ , which measures the semi-elasticity of trade to the FVA reform. The coefficient  $\beta_{ES}$  measures the impact of variation in foreign export supply (that is exogenous to FYR Macedonia) and the collection of coefficients  $\beta_{IDj}$  jointly capture variation in import demand within the region neighboring FYR Macedonia (that proxies for supply shocks in FYR Macedonia).

A standard concern about D-i-D models is the strong identifying assumption they require in order to be valid. In our setting, this assumption would be that, in the absence of the FVA reform, import growth for treated and control products would have experienced similar time trends. It is important to test this assumption even though we do not observe the counterfactual scenario where the FVA reform did not take place. We approximate this by estimating a regression based on data for the pre-reform period (from January 2009 until June 2014) of import values at the  $kit$  level on a common time trend  $T_t$  and the interaction of the time trend with an indicator for the FVA products (that will be treated in the future by the FVA reform), as well as on the export supply shock and import demand shock variables included in the D-i-D regressions. The estimable pre-treatment common trend test regression is:

$$6) \ln(M_{kit}) = \theta_1 T_t + \theta_2 T_t * FVA_k + \gamma_{ki} + \gamma_{it} + \theta_{ES} WES_{kit} + \sum_{j \in J} \theta_{IDj} WID_{kjt} + u_{kit}.$$

The significance of the coefficient  $\theta_2$  on the interaction term serves as our test statistic for the validity of the common trends assumption and thus the adequacy of D-i-D models in our setting.

#### ***IV.4 Exploiting heterogeneous treatment effects***

While our D-i-D exercises control for the most plausible sources of shocks to FYR Macedonia's imports, the distinctiveness of the products overseen by the FVA makes it difficult to treat any D-i-D exercise as dispositive. In principle, any such exercise can be biased if the outcome of interest is affected by unobservable shocks that coincide with the FVA reform and for which there are no specific controls. In order to supplement the evidence from the D-i-D exercises we conduct a different exercise, this time exploiting heterogeneity in the treatment, which in this case we measure as the sampling rate. In this exercise variation across product groups and countries in the level of sampling undertaken is used to generate an estimate of the causal impacts of reduced sampling on imports. The estimating equation for this exercise is as follows:

$$7) \ln(M_{ikt}) = \alpha_1 SR_{ikt} + \gamma_{ik} + \gamma_{it} + \beta_{ES} WES_{ikt} + \sum_{j \in J \subset I} \beta_{IDj} WID_{jkt} + \varepsilon_{ikt}.$$



where  $SR_{ikt}$  is the sampling rate for product group  $k$  from country  $i$  at time  $t$ . The coefficient of interest is  $\alpha_1$ , which should be negative if reduced sampling increases imports flows. As we did above, we include a rich set of fixed effects to sweep out idiosyncratic shocks. We also include trading partners' world export supply shock and the import demand shocks of FYR Macedonia's immediate neighbors, for the same reasons outlined above.<sup>48</sup> The specification exploits variation in the sampling rate within product group-country pairs to identify the link between the sampling rate and the level of imports. Importantly, the sample used to estimate this specification includes only products in the categories of food of animal origin and animal feeds, so it does not include any control products outside of the FVA's mandate nor food of non-animal origin and food contact materials (which are part of the estimating sample for the exercise described in Section IV.3).

Note that while  $\alpha_1$  measures the effect of reduced sampling activity on import values, it cannot be interpreted as the impact of the FVA reform itself. As noted in Sections II and III, the funding dispute with the public laboratories led to a sharp reduction in the number of samples taken in 2015. This variation in the sampling rate is not due to the FVA reform, but it has a large impact on the sampling rate in 2015. Arguably, the unexpected and sharp change in the sampling rate aids the identification of the effect we wish to estimate here. Since the sampling rate drops almost to zero, the change in the sampling rate is virtually determined by the sampling rate prior to the reform. If one thought that political economy reasons led the authorities to conduct more sampling on products whose imports were sensitive to sampling rates, a sudden elimination of sampling activity can help to identify the parameter of interest.<sup>49</sup>

## **Section V. Results**

### ***V.1 Adherence to sampling plan results***

The sampling plan for 2014 establishes that a total of at least 598 import shipments should be sampled for food of animal origin and animal feed products while the number grows by 270% in the sampling plan for 2015.<sup>50</sup> The last row in Panels A or B in Table 1 shows that in total, across all product groups or groups of countries, the FVA sampled substantially more shipments than planned in 2014 - at a ratio of three more shipments sampled than required by the plan - while it sampled substantially fewer shipments than planned in 2015, 23 times fewer shipments than planned. The over-sampling in 2014 may be partly due to the fact

---

<sup>48</sup> These shocks are defined here at the broad product group-country-semester-year level of disaggregation.

<sup>49</sup> Similarly, Topalova and Khandelwal (2011) exploit cross-product variation in the changes in India's tariff rates that arises because tariffs were quite high and variable initially, and were reduced to low and more uniform levels by an IMF structural adjustment program.

<sup>50</sup> Some of the rules in the sampling plan require FVA border agents to sample all import shipments of a particular product group, therefore the number of planned import shipments to be sampled depends on the total number of import shipments. The trebling in the number of shipments to be sampled in the sampling plan for 2015 is due to a change in the rule requiring to sample all import shipments of pet food from any country of origin in 2015.

that while the sampling plan indicated numbers of import shipments to be sampled over the course of the entire year, the FVA reform (the effective use of the sampling plan) only became operational on July 1 2014, hence sampling activity in the first half of 2014 was conducted in an environment where the risk management process was not active yet. The dramatic under-sampling observed in 2015 can be explained by the funding dispute with the public laboratories.

**Table 1. Number of import shipments sampled: planned and actual in 2014 and 2015 and adherence to plan**  
**Panel A. By product group**

Product Group	Planned number of import shipments sampled	Actual number of import shipments sampled	Adherence to plan (actual/planned)	Planned number of import shipments sampled	Actual number of import shipments sampled	Adherence to plan (actual/planned)
	2014	2014	2014	2015	2015	2015
Milk and milk products	48	1072	22.33	31	6	0.19
Fish, fresh, frozen, and canned	55	356	6.47	42	20	0.48
Poultry meat, fresh or frozen	130	189	1.45	83	38	0.46
Meat products (not canned)	15	64	4.27	14	5	0.36
Pet food	9	46	5.11	1072	0	0.00
Compound feed	18	34	1.89	20	0	0.00
Pork meat, fresh or frozen	10	29	2.90	10	0	0.00
Bovine meat, fresh, chilled or frozen	45	23	0.51	28	0	0.00
Honey	23	21	0.91	38	0	0.00
Molluscs and crustaceans	220	11	0.05	224	0	0.00
Offal	5	7	1.40	18	0	0.00
Products from fish and crustaceans	20	1	0.05	9	0	0.00
<i>Total</i>	<i>598</i>	<i>1853</i>	<i>3.10</i>	<i>1589</i>	<i>69</i>	<i>0.04</i>

**Panel B. By country or group of countries**

Sampling Rule	Planned number of import shipments sampled	Actual number of import shipments sampled	Adherence to plan (actual/planned)	Planned number of import shipments sampled	Actual number of import shipments sampled	Adherence to plan (actual/planned)
	2014	2014	2014	2015	2015	2015
All countries	389	1021	2.62	1450	37	0.03
EU & EFTA	36	503	13.96	23	0	0.00
Asia & Africa	25	131	5.24	4	9	2.25
CEFTA	40	119	2.96	12	2	0.17
North America	55	41	0.75	14	3	0.21
Brazil	20	20	1.00	4	0	0.00
All excluding EU & CEFTA	25	19	0.76	.	.	.
Switzerland	0	0		.	.	.
Baltic countries	5	0	0.00	3	0	0.00
All excluding EU & CEFTA & EFTA	3	0	0.00	.	.	.
All excluding Brazil				24	0	0.00
All excluding North America & South America				40	16	0.40
<i>Total</i>	<i>598</i>	<i>1853</i>	<i>3.10</i>	<i>15</i>	<i>2</i>	<i>0.13</i>

Notes: in both panels the table shows information based on sampling plans (1<sup>st</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, and 6<sup>th</sup> columns), FVA lab sample data set (2<sup>nd</sup>, 3<sup>rd</sup>, 5<sup>th</sup>, and 6<sup>th</sup> columns), and EXIM data set (1<sup>st</sup> and 3<sup>rd</sup> columns when all shipments are to be sampled). The data sets are combined based on the broad product group. The sample covers food of animal origin and animal feeds imported under licenses I047 and N853. The actual number of shipments sampled is overestimated in this table as the sampling plan includes for a specific product group regions or groups of countries that overlap. Therefore, samples taken from a specific region may be counted twice. This is mainly because for a given product group, different hazards are tested for different groups of countries.

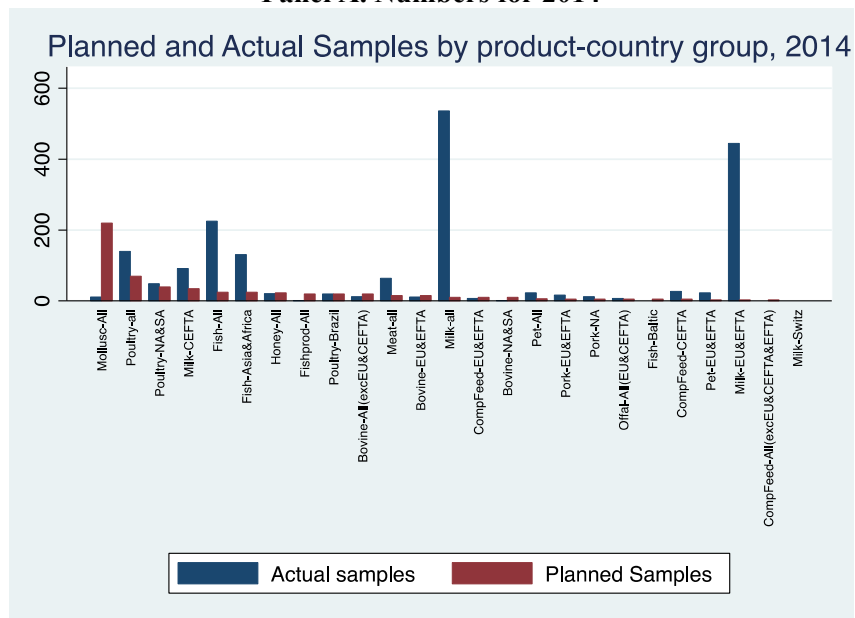
As described in Section III, the sampling plan establishes the number of samples to be taken based on the country or group of countries, the product and the hazard. Panel B of Table 1 shows that according to the sampling plan North America is the group of countries from which more import shipments would be

sampled. However, in practice for certain products, the EU, CEFTA, EFTA and Asia and Africa are the groups of countries from which import shipments are most sampled. The high sampling in EU, CEFTA, and EFTA and the dramatic over-sampling of the product group milk and milk products in 2014 seen in Panel A was due to a widespread contamination of milk products in some European countries in 2013.<sup>51</sup> This contamination forced FVA border agents to sample a large proportion of shipments of those products originating from the EU, CEFTA, and EFTA. Specifically, in 2014 FVA border agents sampled 22 times more import shipments of milk and milk products than indicated in the sampling plan and in that year sampled shipments of milk and milk products accounted for a fifth of the total number of sampled shipments. This constitutes an example of a legitimate case where the sampling plan was not followed, as news warranted substantial additional sampling activity.

It is also useful to assess the adherence to the sampling plan visually by plotting side-by-side planned and actual import shipments sampled for each product group and country or group of countries. Figure 1 shows these plots in Panel A for 2014 and in Panel B for 2015.

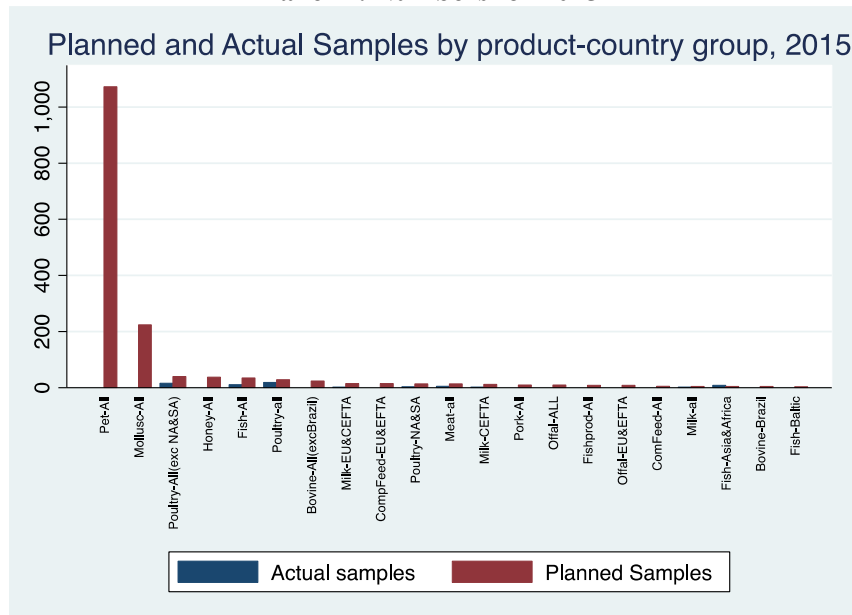
**Figure 1. Planned and actual number of import shipments sampled**

**Panel A. Numbers for 2014**



<sup>51</sup> [https://en.wikipedia.org/wiki/2013\\_European aflatoxin contamination](https://en.wikipedia.org/wiki/2013_European aflatoxin contamination) and <https://www.bloomberg.com/news/articles/2013-03-05/balkan-states-to-step-up-food-control-on-aflatoxin-scare>.

**Panel B. Numbers for 2015**



Notes: In both panels the table shows information based on sampling plans, FVA lab sample data set, and EXIM data set. The data sets are combined based on the broad product group.

To formally test the degree to which actual sampling activity adheres to the sampling plan we estimate Equation (2), which links sampling activity with planned samples by OLS. For this analysis we rely on the sampling plan data at the product group-country or group of countries-year level and the actual number of shipments sampled from the FVA lab sample data set at the same level. Table 2 presents the estimates and inference is based on Huber-White standard errors robust to heteroskedasticity. Columns (1)-(4) make use of the full sample of food of animal origin and animal feeds' product groups while columns (5)-(8) exclude from the sample milk and milk products which deviated largely from the sampling plan in 2014 due to the toxic scare in Europe. Pet products in 2015 were also excluded as the number of planned shipments to be sampled changed dramatically due to a change in the group of countries from which shipments had to be sampled and the number of shipments to be sampled. The estimated coefficient on the planned number of import shipments to be sampled is actually negative and in some cases statistically significant when the full sample is used. However, when the two product groups are excluded, there is a positive and significant coefficient on the planned number of import shipments to be sampled in column (6) when broad product group fixed effects are included as controls. Product group fixed effects are especially important in the context of FVA samples, because they act as partial controls for cross-product variation in the expected harm from a non-compliant shipment.

Albeit weak, our regression evidence is indicative of rough adherence to the sampling plan by FVA border agents in their sampling activity. Moreover, the fact that the coefficient on the planned number of import shipments to be sampled is larger than 1 suggests that FVA border agents are deviating from the

sampling plan in over-sampling shipments deemed risky *ex ante* (i.e. those with higher numbers of planned samples). But it is important to note again the two major challenges that affect our assessment of the adherence to the sampling plan. On one hand, the sampling plan for 2014 was meant to be implemented at the start of the year. However, implementation started only on July 1, 2014. Therefore, the actual number of import shipments sampled counts sampling of import shipments conducted before the implementation of the sampling plan. On the other hand, the funding dispute with the public laboratories experienced by the FVA in 2015 affected the number of import shipments sampled, so we are unable to fully separate the effect of the funding dispute with the public laboratories and the adherence to the sampling plan.

**Table 2. Adherence to the sampling plan regressions**

	Dependent variable: actual number of shipments sampled							
	Sample period: 2014-2015							
	Sample excludes milk products (in 2014-2015) and pet food (in 2015)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Planned number of import shipments to be sampled	-0.047* (0.026)	-0.017* (0.010)	-0.011 (0.040)	0.060 (0.047)	0.053 (0.109)	2.240** (0.872)	1.929 (1.105)	1.101 (1.244)
Constant	44.478** [16.9]	43.04*** (15.49)	79.05*** (27.58)	70.08* (37.80)	20.76*** (7.23)	-37.32* (19.48)	-15.08 (38.35)	6.323 (37.91)
Broad product group fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Country or group of countries fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Year fixed effects	No	No	No	Yes	No	No	No	Yes
Observations	46	46	46	46	39	39	39	39
R-squared	0.005	0.294	0.502	0.658	0.003	0.457	0.553	0.680

Notes: Robust standard errors in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% confidence levels, respectively.

## V.2 Results linking sampling rates and observed compliance

This section presents the results from testing the hypothesis that the risk-based sampling reform led to a closer alignment between testing activity and rates of non-compliance with import regulations. As mentioned before, non-compliance is a novel outcome that has not been previously studied in the literature on impact evaluation of trade facilitation reforms. To calculate sampling and non-compliance rates, we use the total number of import shipments, the number of shipments sampled, and the number of shipments sampled that are non-compliant at the product group-country of origin-semester-year level from the merged FVA lab sample and EXIM data sets. Summary statistics on the various numbers of shipments, sampling rates and non-compliance rates are provided in Table 3, Panel A covers all import shipments, Panel B considers only import shipments of food of animal origin, and Panel C considers only import shipments of animal feeds.

The total number of import shipments sampled exhibits a decreasing trend with the dramatic fall in 2015 being due to the funding dispute with the public laboratories rather than the FVA reform itself. A clear pattern that emerges from Table 3 is that the number of import shipments sampled that exhibit non-

compliance with FVA’s technical standards is extremely small, less than 30 over the period 2013-2015. Also, between 2013 and 2014 the non-compliance rate increased while the sampling rate decreased. This is rough evidence suggesting that the FVA was implementing a better targeting strategy, although firmer conclusions require conditional statements. One issue is that the aflatoxin outbreak in 2013 could have raised sampling rates and measured rates of non-compliance at the same time.

**Table 3. Summary statistics on numbers of shipments, sampling and non-compliance rates**  
**Panel A. All import shipments**

	Number of import shipments sampled	Total number of import shipments	Number of sampled import shipments non-compliant	Sampling rate	Non-compliance rate
2013-1	835	8167	12	10.30%	1.40%
2013-2	446	8479	1	5.30%	0.20%
<i>Total 2013</i>	<i>1281</i>	<i>16646</i>	<i>13</i>	<i>8.00%</i>	<i>1.00%</i>
2014-1	451	7932	2	5.70%	0.40%
2014-2	780	8643	21	9.10%	2.70%
<i>Total 2014</i>	<i>1231</i>	<i>16575</i>	<i>23</i>	<i>7.00%</i>	<i>2.00%</i>
2015-1	4	8844	0	0.00%	0.00%
2015-2	59	9615	0	0.60%	0.00%
<i>Total 2015</i>	<i>63</i>	<i>18459</i>	<i>0</i>	<i>0.00%</i>	<i>0.00%</i>

**Panel B. Import shipments of food of animal origin (requiring license N853)**

	Number of import shipments sampled	Total number of import shipments	Number of sampled import shipments non-compliant	Sampling rate	Non-compliance rate
2013-1	753	7678	2	9.80%	0.30%
2013-2	418	8011	0	5.20%	0.00%
2014-1	422	7510	2	5.60%	0.50%
2014-2	744	8136	20	9.10%	2.70%
2015-1	4	8320	0	0.00%	0.00%
2015-2	56	9050	0	0.60%	0.00%
<i>Total</i>	<i>2397</i>	<i>48705</i>	<i>24</i>	<i>4.90%</i>	<i>1.00%</i>

**Panel C. Import shipments of animal feeds (requiring license I047)**

	Number of import shipments sampled	Total number of import shipments	Number of sampled import shipments non-compliant	Sampling rate	Non-compliance rate
2013-1	82	489	10	16.80%	12.20%
2013-2	28	468	1	6.00%	3.60%
2014-1	29	422	0	6.90%	0.00%
2014-2	36	507	1	7.10%	2.80%
2015-1	0	524	0	0.00%	0.00%
2015-2	3	565	0	0.50%	0.00%
<i>Total</i>	<i>178</i>	<i>2975</i>	<i>12</i>	<i>6.00%</i>	<i>6.70%</i>

Notes: in all panels the table shows information calculated based on the FVA lab sample data set (1<sup>st</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, and 5<sup>th</sup> columns) and the EXIM data set (2<sup>nd</sup> and 4<sup>th</sup> columns). The data sets are combined based on the broad product group.

Table 3 also shows that in the second semester of 2014 the sampling rate increased and the non-compliance rate increased, reaching its maximum value in the sample period. Imports of food of animal origin represent the largest share, covering approximately 93% of the total number of shipments and of the total number of shipments sampled of the combination of food of animal origin and animal feeds. Imports

of food of animal origin exhibit lower sampling rates as well as non-compliance rates than import shipments of animal feeds during the sample period.

Table 4 shows the product groups whose import shipments are most sampled and the product groups whose sampled import shipments are most non-compliant. The two lists of product groups differ. Also, a key feature that Table 4 shows is that there are product group-country of origin-semester-year cells whose sampling rate is larger than 1, which should be theoretically impossible to observe. However, this anomaly is due to our need to use different data sets to compute the number of sampled import shipments (FVA lab samples) and the total number of shipments (EXIM) given that there is no centralized system collecting all the information in FYR Macedonia. Our conjecture is that there is an imperfect concordance in the timing of the recording of the import shipments being sampled in the FVA lab samples data and the shipments being imported in the EXIM data.

**Table 4. Most sampled and most non-compliant product groups**

Product group	Sampling rate	Non-compliance rate
<i>Most sampled</i>		
1 Compound feed	382.00%	10.00%
2 Honey	43.00%	0.00%
3 Eggs and egg products	37.00%	0.00%
4 Meat of other animals, fresh or frozen	17.00%	0.00%
5 Fish, fresh, frozen, and canned	15.00%	0.00%
<i>Most non-compliant</i>		
1 Prepared meat	0.00%	22.00%
2 Offal	0.00%	10.00%
3 Compound feed	382.00%	10.00%
4 Molluscs and crustaceans	3.00%	6.00%
5 Poultry meat, fresh or frozen	3.00%	5.00%

Note: the table shows information calculated based on the FVA lab sample data set (both columns) and the EXIM data set (2<sup>nd</sup> column). The data sets are combined based on the broad product group.

Table 5 documents the specific hazards identified in non-compliant import shipments of food of animal origin in Panel A and of animal feeds in Panel B in years 2013 and 2014. The information is aggregated across FVA laboratory samples with common characteristics. Shipments of milk and milk products had the most irregular samples, most of them linked to the aflatoxin outbreak in 2013.

**Table 5. Identified irregularities and their characteristics**

Panel A. Identified irregularities in license N853 - food of animal origin				
Product group	Origin countries	Irregularity, if reported	Border post	Number of irregular shipments sampled
2013 Milk and Milk products	Bosnia	Aflatoxin	Tabanovce	8
2013 Milk and Milk products	Croatia	Aflatoxin	Skopje	2
2013 Milk and Milk products	Morocco	-	Skopje	1
2013 Milk and Milk products	Serbia	-	Kavardarci	1
2013 Milk and Milk products	Serbia	Aflatoxin	Skopje	6
2013 Pork meat, fresh or frozen	Canada	-	Bogorodica	2
2013 Pork meat, fresh or frozen	Italy	-	Veles	1
2014 Meat products (not canned)	Bosnia	Listeria	Skopje	1
2014 Molluscs and Crustaceans	Greece	Listeria	Medzitlija	1
2014 Offal	Italy	Salmonella	Bogorodica	1
2014 Pork meat, fresh or frozen	Not reported	E.Coli	Tabanovce	3
2014 Pork meat, fresh or frozen	Spain	Salmonella	Tabanovce	1
2014 Poultry meat, fresh or frozen	Brazil, Canada, Italy	Salmonella	Bogorodica	3
2014 Poultry meat, fresh or frozen	Croatia	Salmonella	Skopje	1
2014 Poultry meat, fresh or frozen	Germany	Salmonella	Tabanovce	1
2014 Poultry meat, fresh or frozen	Germany	E. Coli	Tabanovce	1
2014 Poultry meat, fresh or frozen	Greece	E. Coli	Bogorodica	2
2014 Poultry meat, fresh or frozen	Italy	Listeria	Bogorodica	1
2014 Poultry meat, fresh or frozen	Not reported	Salmonella	Tabanovce	3
2014 Prepared meat	Not reported	Salmonella	Tabanovce	1
2014 Prepared meat	Not reported	Listeria	Tabanovce	1

Panel B. Identified irregularities in license I047 - animal feeds				
Product group	Origin countries	Irregularity, if reported	Border post	Number of irregular shipments sampled
2013 Concentrated feeds	Serbia, Slovenia	-	K. Palanka	11
2014 Concentrated feeds	Serbia	Clostridium	K. Palanka	1

Notes: the table shows information calculated based on the FVA lab sample data set. “-“ indicates that the identified hazard was not reported in the FVA lab sample data set.

Table 6 presents the results from estimating by OLS Equation (2) in Panel A based on a sample covering pre-reform data only (ending on June 30, 2014) and Equation (3) in Panel B including data until December 2015 and in Panel C including data only until end of 2014, relying on data at the sampled shipment level where each shipment belongs to a given product group-country of origin-semester-year. Inference is based on Huber-White standard errors robust to heteroscedasticity, which we cluster by broad product group-country of origin because the specifications explain shipment level non-compliance with sampling rates at that more aggregate level (Moulton, 1990). In each of the panels the first half of the columns use as estimating sample all shipments sampled while the second half of the columns exclude from the estimating sample shipments whose corresponding sampling rate at the product group-country of origin-semester-year is larger than 1.

The specifications in Panel A allow us to check, in principle, whether the likelihood of non-compliance is higher when the sampling rates are higher - before the reform. But given the extremely small number of sampled shipments that are non-compliant the regressions are unable to identify any significant link, even when we do not control for any type of fixed effects.

The specifications in Panels B and C show that once the post-reform semester-year data are included in the estimating sample there is generally a negative relationship between non-compliance rates and sampling rates per se, though with unstable significance across specifications. The coefficient on the post-reform indicator in columns (4) and (9) in both panels suggests that the FVA reform increased



significantly the probability of non-compliance, conditional on sampling rates. This might hint at some improvement in targeting of sampling after the reform. But that is not the case, as verified by the worrying coefficient on the interaction term in column (5) of Panels B and C, which suggests that the relationship between sampling and non-compliance became significantly more negative after the reform was implemented. This finding is true even based on the sample where shipments with corresponding sampling rates higher than 1 are dropped, in column (10) of Panels B and C. Such negative relationships are unexpected and counter-intuitive because we would expect the probability of non-compliance to be higher when the sampling rates are higher even prior to the reform, but especially after the reform. This follows from the risk management principle that sampling rates should be higher when probabilities of violations are higher. Thus, if the reform was effective it should have strengthened the positive relationship between non-compliance rates and sampling rates. That does not seem to be the case for FYR Macedonia, even if we consider only the second semester of 2014 as the reform period (Panel C) ignoring the funding dispute with the public laboratories that plagues our analysis. Overall our findings suggest that the targeting of sampling is not effective in FYR Macedonia.

**Table 6. Non-compliance testing regressions**  
**Panel A. Period pre-reform**

Dependent variable: indicator for import shipment sampled being non-compliant								
Sample period: 2013-1 <sup>st</sup> semester of 2014								
	Excluding sampling rates larger than 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sampling rate	-0.000 (0.002)	0.003 (0.003)	0.006 (0.004)	0.005 (0.004)	-0.000 (0.003)	0.004 (0.004)	0.006 (0.004)	0.005 (0.005)
Broad product group fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Country of origin fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Semester-Year fixed effects	No	No	No	Yes	No	No	No	Yes
Observations	1,594	1,594	1,594	1,594	1,579	1,579	1,579	1,579
R-squared	0.000	0.007	0.020	0.021	0.000	0.007	0.020	0.021

**Panel B. Reform period until 2015**

Dependent variable: indicator for import shipment sampled being non-compliant										
Sample period: 2013-2015										
	Excluding sampling rates larger than 1									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sampling rate	-0.015*** (0.005)	-0.006* (0.003)	-0.017* (0.010)	-0.011 (0.007)	0.008 (0.007)	-0.017*** (0.006)	-0.006* (0.003)	-0.018* (0.010)	-0.011 (0.008)	0.008 (0.008)
Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)				0.017** (0.008)	0.040** (0.015)				0.017** (0.008)	0.040** (0.016)
Sampling rate * Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)					-0.033** (0.013)					-0.033** (0.013)
Broad product group fixed effects	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Country fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Semester-Year fixed effects	No	No	Yes	No	No	No	No	Yes	No	No
Observations	2,296	2,296	2,296	2,296	2,296	2,281	2,281	2,281	2,281	2,281
R-squared	0.007	0.044	0.079	0.073	0.079	0.007	0.044	0.079	0.073	0.079

### Panel C. Reform period only 2<sup>nd</sup> semester of 2014

Dependent variable: indicator for import shipment sampled being non-compliant										
Sample period: 2013-2015										
						Excluding sampling rates larger than 1				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sampling rate	-0.016*** (0.005)	-0.007* (0.004)	-0.017* (0.010)	-0.015* (0.009)	0.005 (0.008)	-0.017*** (0.006)	-0.007* (0.004)	-0.018* (0.010)	-0.016* (0.009)	0.006 (0.008)
Post-reform indicator (2 <sup>nd</sup> semester of 2014)				0.020** (0.009)	0.047*** (0.018)				0.020** (0.009)	0.047*** (0.018)
Sampling rate * Post-reform indicator (2 <sup>nd</sup> semester of 2014)					-0.038*** (0.015)					-0.039*** (0.015)
Broad product group fixed effects	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Country fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Semester-Year fixed effects	No	No	Yes	No	No	No	No	Yes	No	No
Observations	2,259	2,259	2,259	2,259	2,259	2,244	2,244	2,244	2,244	2,244
R-squared	0.007	0.047	0.081	0.080	0.087	0.008	0.047	0.081	0.080	0.087

Notes: Robust standard errors clustered by broad product group-country of origin in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% confidence levels, respectively. The sample covers import shipments of food of animal origin (license N853) and animal feeds (license I047).

One potential endogeneity problem in these estimates is the feedback from observed non-compliance to higher rates of sampling and laboratory testing. Evidence of non-compliant shipments might invite more scrutiny. Alternatively, news of non-compliance elsewhere in Europe might trigger increased sampling activity even in the absence of direct evidence of non-compliance on imports into FYR Macedonia. These possibilities could bias the estimated coefficient on sampling rates in either direction. We are not able to address such problems fully, but for purposes of robustness it is useful to exclude from the estimating sample milk and milk products in 2014-2015 (because these were products in which feedback effects were most likely) as well as pet food in 2015 (as the number of planned samples increased substantially likely due to feedback effects). The estimates shown in Appendix Table B1 are similar to those in Table 6, despite a smaller number of observations (since most shipments that are sampled are of milk and milk products and pet food, as seen in Table 3). We also estimate Equations (2) and (3) using lagged rather than contemporaneous sampling rates, as an alternative to mitigate potential endogeneity. The estimates shown in Appendix Table B2 are also similar to those in Table 6.

### V.3 Estimates of trade effects

We estimate the impact of the FVA reform on trade via D-i-D regressions with corresponding pre-treatment common trend testing regressions relying on the Makstat data set's import values at the HS 4-digit product-country of origin-semester-year level.<sup>52</sup>

<sup>52</sup> While the data are available at a more disaggregated HS 6-digit product-country of origin-month-year level, we decided against using that level of aggregation because the many zero observations at that very low level of aggregation made it nonsensical to estimate pre-treatment trends.

We first report on the tests for the adequacy of using D-i-D models for identification of the impact of the FVA reform. As discussed in Section IV.3, we estimate Equation (6) based on data for the period ranging from January 2009 until June 2014 to check whether products subject to the FVA reform experienced differential growth patterns in their imports before the reform, relative to other products not subject to the FVA reform. Table 7 presents the results from estimating Equation (6) by OLS with standard errors robust to heteroscedasticity. Panel A provides results for the full sample of all HS 4-digit products under FVA control using all other HS 4-digit products imported by Macedonia as control group. Column (1) provides estimates of a simplified version of Equation (6) which controls for HS 4-digit product-country of origin (panel) fixed effects only, while Columns (2) and (3) provide estimates including the rich set of control variables and fixed effects shown in Equation (6).<sup>53</sup> All columns in Panel A of Table 7 report an insignificant coefficient on the interaction between the time trend and the indicator variable for FVA products. Hence, there is no evidence of differential trends in import growth for FVA products relative to other imported products before the FVA reform.

---

<sup>53</sup> Note that these supply and demand shocks all have highly significant coefficient estimates of the expected sign.

**Table 7. Pre-reform common trends tests regressions**  
**Panel A. Full Sample**

	Dependent variable: log import value		
	Sample period:	Sample period:	Sample period:
	2009-1 <sup>st</sup> semester	2010-1 <sup>st</sup> semester	2010-1 <sup>st</sup> semester
	of 2014	of 2014	of 2014
	Annual shocks	Semester shocks	
	(1)	(2)	(3)
Semester time trend	0.013*** (0.001)	0.109*** (0.007)	0.089*** (0.009)
Semester time trend * Indicator for FVA products	0.002 (0.002)	0.000 (0.002)	0.003 (0.004)
Export supply shock - log country of origin total exports to world		0.110*** (0.005)	0.102*** (0.010)
Import demand shock - log Albania imports from world		0.018*** (0.003)	0.042*** (0.003)
Import demand shock - log Bulgaria imports from world		0.042*** (0.006)	0.049*** (0.011)
Import demand shock - log Greece imports from world		0.049*** (0.006)	0.059*** (0.012)
Import demand shock - log Serbia imports from world		0.046*** (0.004)	0.083*** (0.009)
HS 4-digit product * country of origin fixed effects	Yes	Yes	Yes
Country of origin * year fixed effects	No	Yes	Yes
Observations	172,644	162,260	112,794
R-squared	0.785	0.790	0.798

**Panel B. Sample of food contact materials and products in same HS 2-digit**

	Dependent variable: log import value		
	Sample period:	Sample period:	Sample period:
	2009-1 <sup>st</sup> semester	2010-1 <sup>st</sup> semester	2010-1 <sup>st</sup> semester
	of 2014	of 2014	of 2014
	Annual shocks	Semester shocks	
	(1)	(2)	(3)
Semester time trend	0.013*** (0.002)	0.108*** (0.009)	0.090*** (0.012)
Semester time trend * Indicator for food contact materials	0.005 (0.003)	0.002 (0.003)	0.001 (0.005)
Export supply shock - log country of origin total exports to world		0.110*** (0.006)	0.102*** (0.015)
Import demand shock - log Albania imports from world		0.018*** (0.004)	0.044*** (0.004)
Import demand shock - log Bulgaria imports from world		0.015 (0.009)	0.019 (0.017)
Import demand shock - log Greece imports from world		0.040*** (0.009)	0.031* (0.017)
Import demand shock - log Serbia imports from world		0.081*** (0.007)	0.126*** (0.015)
HS 4-digit product * country of origin fixed effects	Yes	Yes	Yes
Country of origin * year fixed effects	No	Yes	Yes
Observations	99,121	93,686	65,899
R-squared	0.769	0.774	0.781

Notes: Robust standard errors in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% confidence levels, respectively. In Panel A FVA products includes food of animal origin (license N853), animal feeds (license I047), food of non-animal origin and food contact materials (license I007)). In column (3) of Panels A and B the import demand shock for Albania is annual (as in column (2)) due to data constraints.

One problem with the above results is that non-FVA products may not serve as a good control group for FVA products as a whole. Panel B provides results for a smaller sample that includes only FVA-controlled products in the category of food contact materials along with other HS 4-digit products included in the corresponding HS 2-digit chapters. As discussed in Section IV.3, the advantage of using this sample

is that it includes a better control group for the FVA-controlled HS 4-digit products, whereas the disadvantage is that it covers just a subset of the FVA-controlled products.<sup>54</sup> All columns in Panel B of Table 7 also show an insignificant coefficient on the interaction term. Thus, there is no evidence of differential trends in import growth for food contact materials FVA-controlled products relative to other imported products in the same HS 2-digit chapters before the FVA reform. Hence, we believe that the common trends assumption required for the appropriateness of D-i-D regressions to evaluate the impact of the FVA reform is valid.

Table 8 presents the results from estimating our D-i-D regressions - Equation (5) - by OLS with standard errors robust to heteroscedasticity. Again, Panel A provides results for the full sample while Panel B provides results for a smaller sample limited to food contact materials and the other HS 4-digit products included in the corresponding HS 2-digit chapters. Recall that the D-i-D estimator allows us to examine the question of whether the products for which the FVA has oversight responsibilities saw faster import growth than other products in the period after the FVA reform, relative to the period before the reform.

Focusing first on column (1) in Panels A and B of Table 8, the D-i-D estimates from a simpler version of Equation (5) including only the HS 4-digit product-country of origin (panel) and semester-year fixed effects suggest that the FVA reform led to a significant increase in import values of FVA products, relative to other products in the post-reform period ranging from July 2014 until the end of 2015. Our concern that the impact of the FVA reform could be potentially biased due to foreign supply shocks or import demand shocks in Macedonia correlated with the FVA reform episode led us to include explicit control variables for foreign supply shocks and import demand shocks based on WITS trade data from column (2) onwards. These variables absorb changes in import values that are unrelated to the FVA reform. The controls for foreign supply and import demand shocks result in a decrease in the magnitude and significance of the estimated trade impact of the FVA reform, but the impact in columns (2) and (3) is still significant at the 10 percent confidence level. In turn, note that variation in foreign supply and import demand matters for Macedonian imports: both foreign supply and import demand shocks have significant effects on Macedonian imports, with the expected positive sign.

---

<sup>54</sup> Specifically, this smaller sample does not include FVA-controlled products food of non-animal origin products, nor food of animal origin and animal feeds.

**Table 8. Difference-in-differences regressions for impact of FVA reform on trade**  
**Panel A. Full Sample**

	Dependent variable: log import value						
	Sample period: 2009-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2014	Sample period: 2010-2014
	(1)	Annual shocks (2)	Semester shocks (3)	Annual shocks (4)	Semester shocks (5)	Annual shocks (6)	Semester shocks (7)
Indicator for FVA products * Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)	0.034** (0.016)	0.029* (0.017)	0.032* (0.018)				
Indicator for FVA products * Post-reform indicator (2015)				0.044** (0.019)	0.041* (0.021)		
Indicator for FVA products * Post-reform indicator (2 <sup>nd</sup> semester of 2014)						0.003 (0.026)	0.009 (0.028)
Export supply shock - log country of origin total exports to world		0.117*** (0.004)	0.099*** (0.008)	0.117*** (0.004)	0.099*** (0.008)	0.114*** (0.004)	0.098*** (0.009)
Import demand shock - log Albania imports from world		0.019*** (0.002)	0.038*** (0.002)	0.019*** (0.002)	0.039*** (0.002)	0.019*** (0.002)	0.038*** (0.002)
Import demand shock - log Bulgaria imports from world		0.040*** (0.005)	0.059*** (0.009)	0.040*** (0.005)	0.059*** (0.009)	0.037*** (0.006)	0.046*** (0.010)
Import demand shock - log Greece imports from world		0.043*** (0.005)	0.059*** (0.010)	0.043*** (0.005)	0.059*** (0.010)	0.046*** (0.006)	0.054*** (0.011)
Import demand shock - log Serbia imports from world		0.050*** (0.004)	0.062*** (0.006)	0.050*** (0.004)	0.061*** (0.006)	0.048*** (0.004)	0.088*** (0.008)
HS 4-digit product * country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Semester-year fixed effects	Yes	No	No	No	No	No	No
Country of origin * semester-year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	224,053	210,446	153,302	210,446	153,302	178,309	126,378
R-squared	0.772	0.777	0.783	0.777	0.783	0.785	0.792

**Panel B. Sample of food contact materials and products in same HS 2-digit**

	Dependent variable: log import value						
	Sample period: 2009-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2014	Sample period: 2010-2014
	(1)	Annual shocks (2)	Semester shocks (3)	Annual shocks (4)	Semester shocks (5)	Annual shocks (6)	Semester shocks (7)
Indicator for food contact materials * Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)	0.073*** (0.023)	0.050** (0.023)	0.051** (0.025)				
Indicator for food contact materials * Post-reform indicator (2015)				0.068** (0.027)	0.072** (0.029)		
Indicator for food contact materials * Post-reform indicator (2 <sup>nd</sup> semester of 2014)						0.012 (0.036)	0.005 (0.039)
Export supply shock - log country of origin total exports to world		0.113*** (0.006)	0.100*** (0.012)	0.113*** (0.006)	0.100*** (0.012)	0.112*** (0.006)	0.101*** (0.014)
Import demand shock - log Albania imports from world		0.020*** (0.003)	0.039*** (0.003)	0.020*** (0.003)	0.039*** (0.003)	0.018*** (0.003)	0.038*** (0.004)
Import demand shock - log Bulgaria imports from world		0.022*** (0.008)	0.032** (0.014)	0.022*** (0.008)	0.032** (0.014)	0.017* (0.009)	0.021 (0.016)
Import demand shock - log Greece imports from world		0.029*** (0.007)	0.038*** (0.014)	0.029*** (0.007)	0.038*** (0.014)	0.033*** (0.008)	0.02 (0.016)
Import demand shock - log Serbia imports from world		0.075*** (0.006)	0.075*** (0.009)	0.075*** (0.006)	0.075*** (0.009)	0.079*** (0.007)	0.119*** (0.014)
HS 4-digit product * country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Semester-year fixed effects	Yes	No	No	No	No	No	No
Country of origin * semester-year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	128,148	121,037	89,169	121,037	90,799	102,764	75,125
R-squared	0.756	0.762	0.766	0.763	0.767	0.770	0.776

Notes: Robust standard errors in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% confidence levels, respectively. In Panel A FVA products includes food of animal origin (license N853), animal feeds (license I047), food of non-animal origin and food contact materials (license I007)). In columns (3), (5), and (7) of Panels A and B the import demand shock for Albania is annual (as in columns (2), (4), and (6)) due to data constraints.

Focusing on the estimate in column (3) of Panel A in Table 8, our estimated semi-elasticity of imports to the FVA reform suggests that import values increased by 3.2% more for FVA products than other products after the reform. As a way to illustrate the economic magnitude of this estimate, we can calculate what would be the equivalent reduction in an *ad valorem* tariff that would generate the same import growth. Following Hummels and Schaur (2014) and others in the gravity literature, we can decompose a given percentage change in the log of imports into the product of two structural parameters: a trade elasticity  $\sigma$  that summarizes the response of imports to changes in prices, and an *ad valorem* tariff-equivalent trade cost  $\tau$ . While our data do not allow us to estimate  $\sigma$  we borrow an estimate of  $\sigma$  from the literature: the intensive margin elasticity of -3.8 estimated by Bernard et al (2003), which is adequate for evaluating import growth along the intensive margin (at the HS 4-digit-country of origin-semester-year level). Assuming this trade elasticity, our estimate of a causal increase in imports of 3.2% is consistent with a 0.8 ( $=3.2\%/-3.8$ ) percentage point reduction in an *ad valorem* tariff. This is the implied effect of the FVA reform (together with the subsequent reduction in sampling rates caused by the funding dispute with the public laboratories). If we believe the best D-i-D estimates are those in Panel B where the treated group includes only food contact materials and the control group is made up of closely related products within the same HS 2-digit chapters, then the estimated semi-elasticity of imports to the FVA reform suggests that import values increased by 5.1% more for food contact materials. Assuming the same trade elasticity of -3.8, this estimate of a causal increase in imports of 5.1% is consistent with a 1.3 percentage point ( $=5.1\%/-3.8$ ) reduction in the tariff equivalent.

Panels A and B in Table 8 provide additional D-i-D estimates in columns (4)-(7), which consider a different definition of the post-FVA reform period. Columns (4) and (5) consider only year 2015 as the post-reform period. The rationale for isolating 2015 is that in that year not only were the monitoring plans from the FVA reform being implemented, but there was the funding dispute with the public laboratories that produced a much lower sampling rate than even recommended by the sampling plan for 2015. As such it is useful to examine what impacts are observed on import growth for FVA products, relative to other products. The estimates show a significant impact of the changes in sampling procedures (along with any other relevant aspects of the reform) on import growth of FVA products. These estimates are of higher magnitude and (in some cases) stronger significance than in columns (2)-(3).

Columns (6) and (7) consider only the second semester of 2014 as the post-reform period and exclude observations for year 2015 from the estimating sample. The rationales for this choice are a) to measure the very short-term impacts of the FVA reform, and b) to attempt to isolate the effects of reform from the effects of the funding dispute with the public laboratories in 2015. These estimates show an insignificant impact of the FVA reform on import growth of FVA products. We are not able to identify a

separate effect of the reform, but the D-i-D regressions do seem to indicate that the changes undertaken after July 1 of 2014 led to increased imports of products under FVA oversight.

### ***Robustness***

One concern about the D-i-D estimates in Panel A of Table 8 is that the control group for FVA products subject to the reform includes HS 4-digit products under HS 2-digit chapter 01 (live animals) whose imports are controlled by the FVA and are subject to veterinary checks at the border post of entry but are not subject to sampling plans. Those products enter under a different import license for live animals (N640) and require a quarantine procedure which was not subject to a reform of the same type as the FVA reforms we are studying, but whose import procedures were also subject to some improvements over our sample period (FVA, 2014). Thus, the inclusion of such products in the control group could be biasing (possibly downwards) the impact of the FVA reforms on imports of FVA products. We estimate all specifications in columns (2)-(7) in Panel A of Table 8 excluding HS 4-digit products under HS 2-digit chapter 01 from the sample and present the results in Panel A of Appendix Table B3. The estimated coefficients on the FVA reform dummy are almost unchanged relative to those in Table 8.

While we believe that the inclusion of the import demand shocks helps to isolate the impact of the FVA reform, relative to other supply and demand developments, there is a slight possibility that developments in FYR Macedonia could affect the import demand of its neighbors. If that were the case, a bias in the import demand shocks' impacts could bias the estimated impact of the FVA reform. To check this possibility, we estimate the specifications in columns (2)-(7) of Table 8 excluding the countries that share a land border with FYR Macedonia from the sample. The results presented in Appendix Table B4 imply an almost unchanged impact on the FVA reform, relative to Table 8.

As another robustness check to the findings in Table 8, we also estimate the specifications in columns (2)-(7) in Panel A of Table 8 excluding from the treated group of FVA all HS 4-digit products mapped onto the broad product group Milk and milk products in 2014-2015 and onto pet food in 2015. These products saw anomalies in the rates of sampling in 2014 and 2015. The estimates shown in Panel B of Appendix Table B3 show the trade impact of the FVA reform is not affected by the exclusion of these products from the sample.

In unreported results we estimated specifications similar to those in Table 8 but allowing the impact of the FVA reform to impact imports differently depending on whether the country of origin of the imports is in a free-trade agreement with FYR Macedonia. The results suggested no significant difference in the impact of the FVA reform for such imports. In unreported results, we also estimated specifications similar to those in Table 8 but focusing on the extensive margin of imports, defined as the number of HS 6-digit products imported in each of the HS 4-digit-country of origin-semester-year cells used in Table 9. Such D-



i-D estimates help us understand whether the variety of imported FVA products grew faster than the variety of non-FVA products imported after the FVA reform. The results showed an insignificant effect of the FVA reform on the extensive margin of imports.

#### ***V.4 Trade effects of heterogeneous treatment***

As a supplement to the evidence from the D-i-D regressions, in this section we discuss the results from specifications that exploit heterogeneity in the degree of treatment by the FVA reform, which we capture by variation across product groups and origin countries over time in the level of sampling undertaken, to identify the link between the sampling rate and the level of imports. There are two caveats to keep in mind regarding these specifications. First, they are estimated on a sample including only food of animal origin and animal feeds, which are the FVA products for which we have the data necessary to construct sampling rates (as discussed in Section IV.2). Second, the product classification applied in these regressions is more aggregated than in the earlier D-i-D regressions using HS 4-digit products, due to the difficulties concurring the FVA sampling data with the HS codes.<sup>55</sup> We estimate the impact of the sampling rate on log imports relying on the merged data set that combines FVA lab samples, EXIM and Makstat for import values and sampling rates measured at the level of the broad product group-country of origin-semester-year in the period 2013-2015.

Table 9 shows the results from estimating Equation (7) by OLS with standard errors robust to heteroscedasticity. All specifications control for the rich set of fixed effects specified in that equation to sweep out idiosyncratic shocks as well as trading partners' world export supply shock and import demand shocks of FYR Macedonia's immediate neighbors. The estimating samples used in columns (1)-(4) includes all observations while those used in columns (5)-(6) exclude observations with a sampling rate larger than 1, an anomaly discussed in Section VI.2. The evidence in Table 9 is generally weak and not consistent across specifications.

The coefficients on the sampling rate in columns (1), (4) and (5) capture not only the effect of reduced sampling activity due to the FVA reform but also of sharply reduced sampling activity due to the funding dispute with the public laboratories in 2015. We had hoped that in face of this unexpected and sharp decline the sampling rate we might be able to observe an effect of reduced sampling rates on imports.<sup>56</sup> However, the positive (though weak) estimates are of the opposite sign as would be expected if reduced

---

<sup>55</sup> The much smaller numbers of observations in these regressions relate to the dropping of non-FVA products as a control group. But they also reflect aggregation across products associated with difficulties concurring the trade classification with the verbal descriptions of product categories used in FVA sampling activity.

<sup>56</sup> Because the funding dispute causes the sampling rate to fall nearly to zero, the change in the sampling rate in 2015 is virtually determined by the sampling rate prior to the reform. If higher sampling activity was conducted on products whose imports were sensitive to sampling rates, the sudden elimination of sampling activity could help to identify the causal effect of reduced sampling activity.

sampling rates had the expected impacts of facilitating trade. Focusing on columns (3)-(4) and (7)-(8), where observations from year 2015 are excluded, the estimated impacts of sampling on imports are negative and in one case significant. This is quite weak evidence suggesting that reduced sampling increases imports in the period in which the effects of the reform were not affected by the finding dispute (the second semester of 2014). In general, we are left with little evidence to suggest an effect of reduced sampling activity on imports.

As a robustness check to the findings in Table 9, we also estimate the specifications excluding from the treated group of FVA all HS 4-digit products mapped onto the broad product group milk and milk products in 2014-2015 and onto pet food in 2015 as these saw for different reasons anomalies in the rates of sampling in 2014 and 2015. The estimates shown in Appendix Table B5 show again no evidence of reduced sampling on imports.

**Table 9. Heterogeneity of FVA reform treatment regressions**

	Dependent variable: log import value							
	Sample period: 2013-2015 Annual shocks	Sample period: 2013-2015 Semester shocks	Sample period: 2013-2014 Annual shocks	Sample period: 2013-2014 Semester shocks	Sample period: 2013-2015 Annual shocks	Sample period: 2013-2015 Semester shocks	Sample period: 2013-2014 Annual shocks	Sample period: 2013-2014 Semester shocks
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sampling rate	0.036 (0.140)	0.011 (0.152)	-0.186 (0.187)	-0.394* (0.239)	0.164 (0.212)	0.123 (0.253)	-0.214 (0.260)	-0.378 (0.343)
Export supply shock - log country of origin total exports to world	-0.346** (0.151)	-0.102 (0.095)	-0.683*** (0.199)	-0.395*** (0.138)	-0.348** (0.153)	-0.102 (0.096)	-0.685*** (0.200)	-0.389*** (0.139)
Import demand shock - log Albania imports from world	0.034 (0.026)	0.000 (0.032)	0.0669* (0.034)	0.003 (0.038)	0.031 (0.026)	-0.004 (0.032)	0.0679* (0.035)	0.002 (0.039)
Import demand shock - log Bulgaria imports from world	0.262 (0.204)	0.426*** (0.159)	0.010 (0.521)	0.421* (0.235)	0.256 (0.205)	0.419*** (0.156)	0.037 (0.531)	0.433* (0.239)
Import demand shock - log Greece imports from world	0.169 (0.264)	-0.332** (0.166)	0.549 (0.716)	-0.474* (0.241)	0.179 (0.265)	-0.278* (0.168)	0.567 (0.722)	-0.477* (0.244)
Import demand shock - log Serbia imports from world	0.130 (0.094)	0.080 (0.061)	0.103 (0.172)	0.029 (0.109)	0.136 (0.095)	0.080 (0.061)	0.098 (0.174)	0.027 (0.110)
Broad product group * country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin * semester-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,480	1,347	1,003	854	1,468	1,338	990	846
R-squared	0.907	0.905	0.899	0.912	0.906	0.905	0.899	0.911

Notes: Robust standard errors in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% confidence levels, respectively. In column (3) of Panels A and B the import demand shock for Albania is annual (as in column (2)) due to data constraints.

## Section VI. Data Management Recommendations

A key challenge to credible evaluation of the FVA sampling and testing reforms has been the quality of the available data. The FVA staff with whom we worked have been very responsive and helpful in their efforts to help us understand the data and to use it. That said, there are important shortcomings in the architecture of the data, and these shortcomings pose important challenges to credible evaluation. More importantly,

the shortcomings also limit the operational effectiveness the risk-based system itself. In particular, the inability to assemble credible data on all the lab samples can make it difficult for the FVA to improve its sampling plans in the future.

We therefore offer some suggestions that we think are necessary to improve the ability of the FVA to monitor its own progress. As we understand it, the FVA is one of the more advanced technical agencies in FYR Macedonia (in terms of its capabilities to oversee import shipments). It is therefore likely that these suggestions apply to other agencies in FYR Macedonia. It is also likely that these suggestions are useful to technical agencies in other countries in the Western Balkans and elsewhere. We limit ourselves to suggestions for improved data management, although it is likely that there are operational suggestions that would improve the current system.<sup>57</sup>

1. FVA should itself retain information on all test outcomes in an electronic format, rather than relying on the laboratories to retain this information and to provide it when needed. Negative test outcomes should be retained as well as positive outcomes. The information should be stored in a way that allows precise matching of both positive and negative test results from each sample to the identification number in EXIM of the shipment from which the sample was taken. Because more than one sample may be taken from a shipment, the system used for matching must allow matching on a many-to-one basis. This recommendation can be implemented by the FVA without the cooperation of any external agencies (although additional funding and technical support may be necessary). Improved data on test results is not only useful for evaluation, it is a resource for improved targeting of sampling in the future. An electronic format would allow updating the sampling plan in real time.
2. A more ambitious data management reform would seek to integrate information from laboratory test results with data recorded by other agencies, especially data from the customs agency. Risk analysis on a shipment-by-shipment basis is preferable to an annually updated sampling plan. It could be that the capabilities of the customs agency could be leveraged to improve the sampling and testing regime. Risk analysis that also incorporates shipment-level data drawn from the customs declaration would allow the improvement of targeting.
3. The FVA should update the sampling plan more frequently. New information - such as reports of outbreaks coming through RASSF or evidence of failed inspections in other agencies - should be incorporated in the sampling plan more frequently than on an annual basis. While emerging risks are currently communicated to border agents through other channels, it would be best to have the

---

<sup>57</sup> We leave these to experts in the implementation of trade facilitation reforms. It is our experience that trade facilitation reforms are context specific, with potentially many interlocking pieces (legislation, staff capabilities, IT capabilities, etc.) that make it difficult to paint with too broad a brush.

sampling plan updated as well. Frequent updating of the sampling plan would better tie actual sampling behavior to the plan. It would also help the agency to learn about the effectiveness of its approaches to updating the plan.

## **Section VII. Conclusion**

The recent negotiation and ratification of the Trade Facilitation Agreement by the members of the World Trade Organization indicates that trade facilitation is an important component of international trade policy. Trade facilitation is also an important tool of economic development policy, as the hundreds of millions of dollars of international aid money spent supporting these activities make clear. As yet there is still very little formal evidence documenting the impacts of specific trade facilitation reforms. The small existing literature tends to focus on reforms in customs agencies, and focuses almost exclusively on estimating the effects of reforms on international trade flows.

While developing country customs agencies have their own challenges, they tend to be much better resourced than other agencies that operate at international borders. The agencies that oversee Sanitary and Phytosanitary standards are particularly important for trade logistics, as they often intervene in a time-consuming manner and their interventions often affect products whose value degrades over time. Because these agencies' border operations are often under-resourced in developing countries, reform is challenging and so is evaluation. We evaluate a reform by the Food and Veterinary Agency in FYR Macedonia that attempts to better target sampling and laboratory testing activities towards the import shipments with the highest risks, and to reduce the overall rate of sampling activity at the same time.

Both the reform and the data present important challenges for our study, but the challenges were greatest in areas in which the existing literature provides the least prior knowledge. We are able to evaluate the agency's implementation of the reform program by comparing actual sampling activities to the sampling activities proposed in the annual sampling plans. A funding dispute between the agency and the public laboratories that conduct the tests meant that the sampling rates in 2015 fell well below the expected rates of activity. We were also able to evaluate the match between sampling activities and levels of non-compliance with FVA's standards, which was low before the reform and did not improve after the reform. Our evidence suggests that there is still substantial room to improve the targeting of sampling and testing activities of import shipments.

The standard exercise in impact evaluations of customs reforms has been an attempt to isolate causal effects of the reforms on international trade. We conduct two exercises that are intended to inform this question for the reform that we study. Because the funding dispute with the public laboratories affected sampling rates, the effects of the dispute are conflated with those of the reform, but the reduced sampling

rates produced by the dispute were consistent with one objective of the reform, which was to achieve lower rates of overall sampling.

Our first set of estimates are indicative of the causal effects of all policy changes related to the products overseen by FVA that occurred in FYR Macedonia in late 2014 and 2015. These changes include the portions of the reform not related to sampling as well as the changes in sampling that occurred but were not planned in the reform. The difference-in-differences specifications suggest a range of estimates, depending on the specification. Our preferred estimates lie in the 3-5% increase in imports, which implies tariff equivalent impacts of approximately 1 percentage point arising from all the policy changes. Our cleanest estimate of the reform alone employs only data from late 2014 data and thus estimates a very short-term effect. This estimate is not statistically different from zero.

We also conduct an exercise that exploits within-broad product group-country of origin variation in sampling rates as an attempt to estimate a causal effect of reduced sampling on imports. We find no statistically significant link between sampling activity and imports. We might have estimated a stronger relationship between reduced sampling and imports if complete sampling data on food of non-animal origin and food contact materials had been collected, because these products were less risky and larger reductions in sampling were planned for these products.

Overall our estimates of the effect of the reform on trade are inconclusive. Our exercises highlight the challenges for evaluation that arise when the reforms are conducted by agencies that lack detailed and comprehensive data in a format consistent with the Harmonized System used to classify international trade flows. The reform we study took place in a challenging period for FYR Macedonia and its bureaucracy. Many if not most reforms will lack a clean experiment amenable to evaluation. Our study's outcome may be typical in that regard.

## References

- Cadot, Olivier, Maryla Maliszewska and Sebastian Saez (2011). "Non-Tariff Measures: Impact, Regulation and Trade Facilitation," in McLinden, G., Fanta, E., Widdowson, D. and Doyle, T., (Eds.). *Border Management Modernization*. The World Bank.
- Carballo, Jerónimo, Alejandro Graziano, Georg Schaur, and Christian Volpe Martincus (2016a). "The Border Labyrinth: Information Technologies and Trade in the Presence of Multiple Agencies," No. IDB-WP-706. IDB Working Paper Series.
- Carballo, Jerónimo, Georg Schaur, Alejandro Graziano, and Christian Volpe Martincus (2016b). "Transit Trade," No. IDB-WP-704. IDB Working Paper Series.
- Carballo, Jerónimo, Georg Schaur, and Christian Volpe Martincus (2016c). "Posts as Trade Facilitators," No. IDB-WP-701. IDB Working Paper Series.
- Carballo, Jerónimo, Georg Schaur, and Christian Volpe Martincus (2016d). "Trust No One? Security and International Trade," No. IDB-WP-703. IDB Working Paper Series.
- European Commission (2016). "Commission Staff Working Document on the former Yugoslav Republic of Macedonia," accompanying the *Communication from the Commission on EU Enlargement Policy*, available at [https://ec.europa.eu/neighbourhood-enlargement/sites/near/files/pdf/key\\_documents/2016/20161109\\_report\\_the\\_former\\_yugoslav\\_republic\\_of\\_macedonia.pdf](https://ec.europa.eu/neighbourhood-enlargement/sites/near/files/pdf/key_documents/2016/20161109_report_the_former_yugoslav_republic_of_macedonia.pdf), accessed April 16, 2017.
- European Commission (2015). "Commission Staff Working Document on the former Yugoslav Republic of Macedonia," accompanying the *Communication from the Commission on EU Enlargement Strategy*, available at [https://ec.europa.eu/neighbourhood-enlargement/sites/near/files/pdf/key\\_documents/2015/20151110\\_report\\_the\\_former\\_yugoslav\\_republic\\_of\\_macedonia.pdf](https://ec.europa.eu/neighbourhood-enlargement/sites/near/files/pdf/key_documents/2015/20151110_report_the_former_yugoslav_republic_of_macedonia.pdf), accessed April 20, 2017.
- Fernandes, Ana M., Russell H. Hillberry, and Claudia N. Berg (2016). "Expediting Trade: Impact Evaluation of an In-House Clearance Program," Policy Research Working Paper 7708.
- Fernandes, Ana M., Russell H. Hillberry, and Alejandra Mendoza Alcantara (2015). "Trade Effects of Customs Reform: Evidence from Albania," Policy Research Working Paper 7210.
- Food and Agricultural Organization (2003) *Guidelines for Food Import Control Systems*, available at [http://www.fao.org/input/download/standards/10075/CXG\\_047e.pdf](http://www.fao.org/input/download/standards/10075/CXG_047e.pdf), accessed April 16, 2017.
- Food and Agricultural Organization (2013) *Principles and Guidelines for National Food Control Systems*, available at [http://www.fao.org/input/download/standards/13358/CXG\\_082e.pdf](http://www.fao.org/input/download/standards/13358/CXG_082e.pdf), accessed April 16, 2017.
- FVA (2014). *Annual Report on the Activity of the Border Veterinary Inspection Department for 2014*. Food and Veterinary Agency FYR Macedonia.
- Hillberry, Russell, and Phillip McCalman (2016). "Import Dynamics and Demands for Protection." *Canadian Journal of Economics* 49 (3): 1125-1152.
- Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang (2014) "The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data," *The American Economic Review* 104 (6): 1597-1629.

- Knowles, John, Nicola Persico, and Petra Todd (2001). "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109 (1): 203-229.
- Martincus, Christian Volpe, Jerónimo Carballo, and Alejandro Graziano (2015). "Customs," *Journal of International Economics* 96 (1): 119-137.
- Milchevki, Todor and Violane Konar-Leacy (2016). "Go the Extra Mile: Changing Long-standing Practices and Adopting New Ways of Thinking," *Smart Lessons*, International Finance Corporation, publisher: pp. 1-4.
- Moulton, B. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics*, 72(2): 334-38.
- Topalova, Petia and Amit Khandelwal (2011). "Trade Liberalization and Firm Productivity: The Case of India," *Review of Economics and Statistics* 93 (3): 995-1009.
- van der Meer, Kees and Laura Ignacio (2011). "Sanitary and Phytosanitary Measures and Border Management", in McLinden, G., Fanta, E., Widdowson, D. and Doyle, T. (Eds.). *Border Management Modernization*. The World Bank.
- Widdowson, David and Stephen Holloway (2011). "Core Border Management Disciplines: Risk-based Compliance Management," in McLinden, G., Fanta, E., Widdowson, D. and Doyle, T., (Eds.). *Border Management Modernization*. The World Bank.
- World Trade Organization (2013). Trade Policy Review of the former Yugoslav Republic of Macedonia, Revision 1. Available online at [https://www.wto.org/english/tratop\\_e/tpr\\_e/tp390\\_e.htm](https://www.wto.org/english/tratop_e/tpr_e/tp390_e.htm), accessed April 16, 2017.

## Appendix A: Data details

Below we describe in detail the four sources of FYR Macedonia data used in the analysis. The data sets each offer detailed information along multiple dimensions (e.g., precisely defined import products, country of origin, date of arrival, etc.). Precision along multiple dimensions allows credible matching across data sets, even though the data lack unique identifiers that would tie the data sets together at the level of an individual shipment. When linking laboratory testing data to data on import flows it is necessary to aggregate across products to improve the concordance. We describe each data set in turn and our efforts to concord the data series.

### *A.1 Sampling plans*

The sampling plan produced by the FVA headquarters offers strong guidance to FVA border agents regarding the number of shipments of certain types to be sampled during the upcoming year. The shipment types are defined in terms of broad product categories, broad country groups (e.g., CEFTA countries), the type of import license and the hazard for which the sample is to be tested. Our sampling plan data for 2014 are taken from the *Manual for implementation of the annual programme for monitoring of food of animal and non-animal origin, feedingstuffs and food contact materials from import for 2014*. Although the plan was constructed with the entire year 2014 in mind, implementation of the 2014 plan was delayed, so sampling according to the plan only began on July 1, 2014. The sampling plans for 2015 were provided by the FVA in electronic format. Our other data are not available at the level of the border post - and some data are not available by the hazard - so for each year we aggregate the sampling plan data by summing over border post and hazard to produce a planned number of shipments to be sampled at the level of a broad product group and countries or groups of countries.

### *A.2 Laboratory sample data from FVA (FVA lab samples)*

The second data source provides information on the number of shipments from which FVA border agents took samples for the purpose of laboratory testing, along with the outcome of the tested samples. These data sets were made available to us by each of the two FVA officials in charge of each subgroup of FVA products (one official handles food of animal origin and animal feeds and the other handles food of non-animal origin and food contact materials). The data sets are not available in any systematic electronic format, rather they are compiled by the FVA officials based on information sent to them by the laboratories. The data sets were provided separately by import license and year. Each row of the FVA lab sample data set represents an import shipment from which samples were taken to the laboratory. The information reported for that import shipment includes the product, the name of the border post where it entered, the date when the sample was taken for laboratory testing, the country of origin, and the health or other hazards tested in the laboratory. We use this data set to calculate the number of shipments with samples taken to laboratory at the broad product group-country of origin-month-year or more aggregated levels.

The FVA lab sample data set presents the following problems for the analysis. The first problem is that for food of non-animal origin and food contact materials (covered by license I007) the FVA lab sample data set records information *only* on shipments from which laboratory samples were taken and laboratory test results were positive for some hazard at levels that indicate non-conformity with regulations. Because only positive outcomes were reported for products under license I007, we lack information on the total number of samples taken under this import license. FVA officials indicated that the data on the universe of laboratory samples taken for all products under license I007 had to be collected from five different institutions, i.e., the laboratories that conducted the testing: the Center for Public Health in Skopje, the Center for Public Health in Veles, the Center for Public Health in Kumanovo, the Institute of Public Health of Republic of Macedonia, and the Veterinary Institute in Skopje. We requested the data from those five agencies but obtained refusals from three of them, only receiving data from the Institute of Public Health



and the Veterinary Institute in Skopje.<sup>58</sup> Therefore all our analysis on sampling focuses exclusively on the other products: food of animal origin and animal feeds (covered by licenses I047 and N853).

The second problem is that the products covered by these two licenses were sampled much less frequently in 2015 than in previous years. The FVA official in charge indicated that the reason for this was that due to a funding dispute with the public laboratories for a certain sub-period within 2015, the FVA did not have a contract with the Veterinary Institute where the analysis for food of animal origin and feeds are conducted.

The third problem relates to the product classification available in these data sets. While occasionally in some years and for some licenses there is an HS 4-digit or HS 6-digit code associated with the sampled shipments, that is not verified across all licenses and years and therefore we had to rely on verbal descriptions of the products, which were not always consistent over time, contained typos, and were sometimes broad and sometimes narrow, relative to an HS 4-digit or HS 6-digit code. A couple of examples of the diversity of the product categories listed in the FVA samples data sets are:

- bovine meat frozen which corresponds well to HS 4-digit code 0202;
- chicken breaded products which corresponds well to HS 6-digit code 160232.

The fourth problem relates to the date when the laboratory samples were taken. Although theoretically all shipments in the FVA lab samples data set should include information on the date when they were taken, in practice several shipments (corresponding to 7.5% of the initial number of shipments in the FVA samples data set for licenses I047 and N853 over the period 2013-2015) do not report that information and therefore only the year of the sample can be used, not the month-year.

The fifth problem relates to the country of origin of the shipment from which laboratory samples were taken. While theoretically all shipments in the FVA lab samples should report the country of origin, in practice several shipments (corresponding to 4.9% of the initial number of shipments in the FVA samples data set for licenses I047 and N853 over the period 2013-2015) do not report that information and therefore can be used only for calculations of numbers of samples by product and date.

Problems three through five are addressed via aggregation along various dimensions.

### ***A.3 Import shipment data from the single window for import, export and transit of goods and tariff quota (EXIM)***

The EXIM e-government portal collects the submission of applications for import licenses (as well as licenses allowing exports and transit shipments). EXIM reports trade volume and license data for a total of 16 agencies including the FVA.<sup>59</sup> The EXIM data set records each license application, which has a firm identifier (which was not made available to us due to confidentiality concerns), a tariff line variable, product descriptions, volume imported, unit in which volume is measured, date of application, and country of origin.<sup>60</sup> The EXIM data set was provided to us by the FVA for the products covered by licenses I007, I047, and N853 for the period 2013-2015. Despite being an electronic platform for data collection, the data are not available in a harmonized and systematized way. We use the EXIM data set to calculate the total number of Macedonian import shipments by product-country of origin-month-year, or at more aggregated levels. The EXIM data set had observations that were incomplete for a variety of reasons. While the number of incomplete observations is small we describe the problems and our treatment of these data in greater detail.

First, although theoretically all shipments in the EXIM data set should include information on the type of product imported at the HS 10-digit level, in practice some shipments report that information at the HS 4-digit level or at the HS 2-digit level. While information at the HS 4-digit level can be used, information

---

<sup>58</sup> Unfortunately, these data cannot be included in our analysis as they represent only a partial set of all samples taken, according to FVA officials. Since we lack data on the total number of import shipments under license I007 that were sampled, we do not even know the share of samples available in the samples data that we do have.

<sup>59</sup> Aside from the FVA, the other agencies are in charge of issuing licenses for the import of controlled products, for example munition, chemicals, drugs and medical products, forestry products, and vehicles.

<sup>60</sup> Note that applications for licenses are lodged at the border post when the goods have already been loaded and shipped to FYR Macedonia. We therefore operate under the assumption that license applications from EXIM represent real shipments.

at the HS 2-digit level cannot as it would cause difficulties in merging across data sets. We exclude from the final EXIM data set - based on which numbers of import shipments by product-country-time period are counted - any shipments whose product information is available only at the HS 2-digit level.<sup>61</sup> This eliminates 0.68% of the initial 287,426 shipments included in the EXIM data set for licenses I007, I047 and N853 over the period 2013-2015.

Second, although theoretically all shipments in the EXIM data set should report the date when they entered Macedonia, in practice some shipments do not report that information. In these cases, only the year of import can be used, not the month-year. Such shipments account for 0.67% of the initial number of shipments in the EXIM data set for licenses I007, I047 and N853 over the period 2013-2015.

Third, although theoretically all shipments in the EXIM data set should report the country of origin, in practice some shipments do not report that information and therefore can be used only for calculations of numbers of shipments by product and date (those shipments account for 0.14% of the initial number of shipments in the EXIM data set for licenses I007, I047 and N853 over the period 2013-2015). Additionally, some shipments report FYR Macedonia as the origin country for the imports (0.19% of the initial number of shipments for licenses I007, I047 and N853 over the period 2013-2015) and thus need to be dropped from the final EXIM data set.

After data problems are addressed, the cleaned EXIM data set includes 284,948 shipments for licenses I007, I047 and N853 over the period 2013-2015. But since the laboratory data for license I007 are not suitable for analysis, we are restricted to imports requiring licenses I047 and N853, for which there are 40,212 import shipments.

#### ***A.4 Trade data from the Macedonian Institute of Statistics (Makstat)***

The Makstat data set was obtained from the State Statistical Institute of FYR Macedonia, which compiled it from the raw import customs data. It contains comprehensive data for the period 2009-2015 on the universe of imports by Macedonia (not only those imports that require a license from the FVA). The Makstat data set includes import values (measured in US dollars) and import volumes (measured in kilograms) by month, HS 10-digit code (called tariff line), and border post of entry. The Makstat data are recorded using codes from different revisions of the Harmonized System (HS) classification (HS2007 and HS2012). To obtain a consistent product classification over time, which is crucial for our analysis, we concord the HS 6-digit codes to a set of ‘consolidated’ HS 6-digit codes that are consistent over time following Fernandes et al. (2016).<sup>62</sup> Import value from the Makstat data set is aggregated to the HS 4-digit-country of origin-semester-year level. Due to the use of products defined at the HS 4-digit level of aggregation in our analysis, we do not use information on import volumes as it would be erroneous to sum them across very different HS 6-digit products within those HS 4-digit categories. The number of observations per year as well as some additional basic descriptive statistics based on the Makstat data set are provided in Appendix Table A1.

---

<sup>61</sup> Several HS 2-digit codes cannot be unambiguously concurred to just one of the broad product categories that we will use in our analysis (defined in section III.5) and concordances that are not 1-to-1 are impossible to consider.

<sup>62</sup> The principles behind this consolidation are to combine all the codes existing under the different HS classifications into a list of unique HS 6-digit codes then the basic principle of ‘consolidation’ is to identify the HS codes related to each other (e.g., codes that were split or merged with the modifications introduced by HS2012) and to replace them with a single code for the entire period. Cebeci (2012) provides additional details.

**Appendix Table A1. Basic statistics on Makstat data set**

	Number of observations at HS 6-digit-country of origin-month-year level	Number of distinct HS 6-digit products imported	Number of distinct origin countries for imports	Total imports (in Billions of USD)
2009	200,261	3,765	180	2,620
2010	201,975	3,759	177	2,850
2011	207,341	3,745	179	3,630
2012	210,186	3,736	177	3,120
2013	221,669	3,757	175	3,080
2014	229,065	3,766	175	3,210
2015	231,750	3,746	177	2,870

***A.4 Usable product classification and a merged sampling and compliance data set***

For a more detailed analysis on the FVA reforms in terms of sampling and non-compliance, we have to define a set of broad product groups that can be concorded to the different product categories and the verbal descriptions used in the sampling plans, FVA lab samples data set, and EXIM data set. The list of broad product groups that we defined in order to encompass those various product categories and verbal descriptions is shown in Appendix Table A2. We constructed a variety of different concordances:

- a concordance between the product groups included in the sampling plans data and the broad product groups – which is used in order to aggregate the sampling plan data to the level of the broad product group-country of origin-year level;
- a concordance between the verbal product descriptions in the FVA samples data set and the broad product groups – which is used in order to aggregate the FVA lab sample data set to the level of the broad product group-country of origin-month-year level;
- concordances between HS 6-digit codes and HS 4-digit codes (in HS 2012 revision) included in the EXIM data set and the broad product groups – which is used in order to aggregate the EXIM data set to the level of the broad product group-country of origin-month-year level;
- a concordance between consolidated HS 6-digit codes used in the Makstat data set and the broad product groups – which is used in order to aggregate the Makstat data set to the level of the broad product group-country of origin-month-year level.

With all these concordances at hand, we construct a data set which merges across the FVA lab samples data set, the EXIM data set and the Makstat data set at the product group-country of origin-semester-year level, which we use to measure sampling rates and non-compliance rates at various levels of disaggregation. Note that this data set covers only information corresponding to food of animal origin and animal feeds (under licenses I047 and N853). In our analysis of sampling and compliance in Sections V.1 and V.2 we rely on sampling and compliance data sets at the product group-country of origin-semester-year level and at the product group-country of origin-year level. The numbers of observations in those data sets are also shown in Appendix Table A3.

**Appendix Table A2. List of broad product groups and corresponding HS 4-digit product codes**

Broad product groups	HS2012 revision 4-digit products mapped
Animal fats	1501; 1502; 1504; 1517; 1521; 1503; 1505; 1506; 1518; 1522
Bovine meat, fresh, chilled or frozen	0201; 0202
Coffee products	0901; 2202
Compound feed	2301; 2308; 1213; 1214
Eggs and egg products	0407; 0408
Fish fillets, fresh or chilled	0304
Fish, fresh, frozen, and canned	0302; 0303; 1604
Honey	0409
Live fish	0301
Meat of other animals, fresh or frozen	0204; 0205
Meat products (not canned)	0209; 0210; 1601
Milk and milk products	0401; 0402; 0403; 0404; 0405; 0406; 1702; 2105
Molluscs and crustaceans	0306; 0307
Offal	0206; 0504
Other edible animal products	0208; 0410
Other animal parts	0505; 0506; 0507; 0510; 0511; 3001; 3101; 4101; 4102; 4103; 4205; 4206; 4301; 5101; 5102; 5103
Other animal related products	9508; 9705
Pet food	2309
Pork meat, fresh or frozen	0203
Poultry meat, fresh or frozen	0207
Prepared meat	1602
Products from fish and crustaceans	0305; 0306; 1605
Protein substances and their derivatives	3502; 3503; 3504

**Appendix Table A3. Basic statistics on sampling and compliance data set**

	Number of observations at product group-country of origin-semester-year level (for licenses I047 & N853)	Number of observations at product group-country of origin-year level (for licenses I047 & N853)	Number of product groups imported	Number of distinct origin countries for imports
2013	1,008	606	24	114
2014	1,052	619	23	126
2015	1,075	644	24	133

### ***A.5 Complementary international trade data from WITS***

We rely on export and import flows from WITS at the annual and at the monthly levels. We use total exports to the world (except to FYR Macedonia) in thousands of US dollars at the exporting country-HS 4-digit-year level or at the exporting country-HS 4-digit-semester-year level as our foreign supply shock variable. In parallel we use total imports from the world (except from FYR Macedonia) in thousands of US dollars for each of the four neighbors of FYR Macedonia - Albania, Bulgaria, Greece, and Serbia - at the HS 4-digit-year level or at the HS 4-digit-semester-year level as our import demand shock variables at the HS 4-digit-year and the HS 4-digit-month-year levels. Note that WITS data at the country-HS 4-digit product-month-year level is available only from 2010 onwards and for a subset of reporting countries, relative to the data at the country-HS 4-digit product-year level. We use data for the subset of 74 reporting countries that report monthly data every year in 2010-2015. The monthly exports of these 74 countries sum to 60% of total exports on average in a given year in the 2010-2015 period. Note also that WITS does not report annual or monthly data for Kosovo, another neighbor of FYR Macedonia, thus our choice of focusing only on the other four neighbors.

## Appendix B: Robustness checks

**Appendix Table B1. Non-compliance testing regressions  
excluding milk and milk products (in 2014-2015) and pet food (in 2015)**

### Panel A. Period pre-reform

Dependent variable: indicator for import shipment sampled being non-compliant								
Sample period: 2013-1 <sup>st</sup> semester of 2014								
Excluding milk and milk products (in 2014-2015) and pet food (in 2015)								
Excluding sampling rates larger than 1								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sampling rate	0.000 (0.003)	0.006 (0.004)	0.008 (0.005)	0.008* (0.004)	0.001 (0.004)	0.007 (0.005)	0.009 (0.005)	0.010** (0.005)
Broad product group fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Country of origin fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Semester-Year fixed effects	No	No	No	Yes	No	No	No	Yes
Observations	1,343	1,343	1,343	1,343	1,328	1,328	1,328	1,328
R-squared	0.000	0.008	0.020	0.023	0.000	0.008	0.020	0.023

### Panel B. Reform period until 2015

Dependent variable: indicator for import shipment sampled being non-compliant										
Sample period: 2013-2015										
Excluding milk and milk products (in 2014-2015) and pet food (in 2015)										
Excluding sampling rates larger than 1										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sampling rate	-0.021*** (0.008)	-0.0134* (0.007)	-0.015 (0.011)	-0.013 (0.011)	-0.001 (0.009)	-0.025*** (0.009)	-0.014* (0.008)	-0.016 (0.013)	-0.015 (0.012)	-0.001 (0.010)
Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)				0.038** (0.016)	0.073*** (0.024)				0.039** (0.016)	0.073*** (0.024)
Sampling rate *Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)					-0.070** (0.030)					-0.070** (0.030)
Broad product group fixed effects	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Country fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Semester-Year fixed effects	No	No	Yes	No	No	No	No	Yes	No	No
Observations	1,726	1,726	1,726	1,726	1,726	1,711	1,711	1,711	1,711	1,711
R-squared	0.007	0.045	0.088	0.085	0.095	0.008	0.045	0.087	0.085	0.095

### Panel C. Reform period only 2<sup>nd</sup> semester of 2014

Dependent variable: indicator for import shipment sampled being non-compliant										
Sample period: 2013-2014										
Excluding milk and milk products (in 2014-2015) and pet food (in 2015)										
Excluding sampling rates larger than 1										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sampling rate	-0.020*** (0.007)	-0.0119* (0.006)	-0.0160 (0.011)	-0.00697 (0.009)	0.00510 (0.008)	-0.024*** (0.009)	-0.0127* (0.007)	-0.0171 (0.012)	-0.00784 (0.010)	0.00580 (0.009)
Post-reform indicator (2 <sup>nd</sup> semester of 2014)				0.0325** (0.014)	0.0581*** (0.020)				0.0328** (0.014)	0.0583*** (0.020)
Sampling rate *Post-reform indicator (2 <sup>nd</sup> semester of 2014)					-0.0534** (0.023)					-0.0537** (0.023)
Broad product group fixed effects	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Country fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Semester-Year fixed effects	No	No	Yes	No	No	No	No	Yes	No	No
Observations	1,761	1,761	1,761	1,761	1,761	1,746	1,746	1,746	1,746	1,746
R-squared	0.007	0.042	0.085	0.077	0.084	0.008	0.042	0.085	0.077	0.084

Notes: Robust standard errors clustered by broad product group-country of origin in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% confidence levels, respectively. The sample covers import shipments of food of animal origin (license N853) and animal feeds (license I047).

## Appendix Table B2. Non-compliance testing regressions using lagged sampling rates

### Panel A. Period pre-reform

Dependent variable: indicator for import shipment sampled being non-compliant								
Sample period: 2013-1 <sup>st</sup> semester of 2014								
	Excluding sampling rates larger than 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lagged sampling rate	-0.008 (0.005)	0.003 (0.003)	0.010 (0.007)	0.011 (0.011)	-0.019 (0.011)	0.013 (0.012)	0.023 (0.026)	0.096 (0.070)
Broad product group fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Country of origin fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Semester-Year fixed effects	No	No	No	Yes	No	No	No	Yes
Observations	779	779	779	779	699	699	699	699
R-squared	0.003	0.024	0.040	0.043	0.006	0.025	0.041	0.053

### Panel B. Reform period until 2015

Dependent variable: indicator for import shipment sampled being non-compliant										
Sample period: 2013-2015										
	Excluding sampling rates larger than 1									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged sampling rate	-0.026*** (0.009)	-0.003 (0.002)	-0.013 (0.010)	-0.017 (0.012)	0.005 (0.013)	-0.046*** (0.016)	-0.003 (0.004)	0.010 (0.038)	-0.001 (0.028)	0.044 (0.032)
Lagged sampling rate *Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)				0.017* (0.009)	0.044** (0.021)				0.031* (0.016)	0.060** (0.025)
Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)					-0.034** (0.017)					-0.077** (0.033)
Broad product group fixed effects	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Country fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Semester-Year fixed effects	No	No	Yes	Yes	No	No	No	Yes	Yes	No
Observations	1,367	1,367	1,367	1,367	1,364	1,058	1,058	1,058	1,058	1,058
R-squared	0.013	0.071	0.130	0.128	0.135	0.015	0.070	0.131	0.130	0.138

### Panel C. Reform period only 2<sup>nd</sup> semester of 2014

Dependent variable: indicator for import shipment sampled being non-compliant										
Sample period: 2013-2014										
	Excluding sampling rates larger than 1									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged sampling rate	-0.026*** (0.009)	-0.005 (0.003)	-0.006 (0.010)	-0.009 (0.009)	0.004 (0.009)	-0.044*** (0.015)	-0.010 (0.007)	0.018 (0.033)	0.007 (0.017)	0.019 (0.021)
Lagged sampling rate *Post-reform indicator (2 <sup>nd</sup> semester of 2014)				0.013 (0.008)	0.032* (0.018)				0.026* (0.014)	0.038* (0.019)
Post-reform indicator (2 <sup>nd</sup> semester of 2014)					-0.024* (0.013)					-0.036 (0.022)
Broad product group fixed effects	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Country fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Semester-Year fixed effects	No	No	Yes	Yes	No	No	No	Yes	Yes	No
Observations	1,399	1,399	1,399	1,399	1,396	1,088	1,088	1,088	1,088	1,088
R-squared	0.012	0.067	0.122	0.117	0.121	0.014	0.065	0.124	0.118	0.121

Notes: Robust standard errors clustered by broad product group-country of origin in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% confidence levels, respectively. The sample covers import shipments of food of animal origin (license N853) excluding milk and milk products.

**Appendix Table B3. Difference-in-differences regressions for impact of FVA reform on trade**  
**Panel A. Excluding HS 2-digit 01 from control group**

	Dependent variable: log import value					
	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015
	Annual shocks	Semester shocks	Annual shocks	Semester shocks	Annual shocks	Semester shocks
	Excluding HS 4-digit products in HS 2-digit chapter 01 from control group					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator for FVA products * Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)	0.029* (0.017)	0.032* (0.018)				
Indicator for FVA products * Post-reform indicator (2015)			0.044** (0.019)	0.041* (0.021)		
Indicator for FVA products * Post-reform indicator (2 <sup>nd</sup> semester of 2014)					0.003 (0.026)	0.009 (0.028)
Export supply shock - log country of origin total exports to world	0.117*** (0.004)	0.098*** (0.008)	0.117*** (0.004)	0.098*** (0.008)	0.114*** (0.004)	0.098*** (0.009)
Import demand shock - log Albania imports from world	0.019*** (0.002)	0.039*** (0.002)	0.019*** (0.002)	0.039*** (0.002)	0.019*** (0.002)	0.038*** (0.002)
Import demand shock - log Bulgaria imports from world	0.040*** (0.005)	0.060*** (0.009)	0.040*** (0.005)	0.060*** (0.009)	0.037*** (0.006)	0.047*** (0.010)
Import demand shock - log Greece imports from world	0.043*** (0.005)	0.059*** (0.010)	0.043*** (0.005)	0.059*** (0.010)	0.046*** (0.006)	0.055*** (0.011)
Import demand shock - log Serbia imports from world	0.049*** (0.004)	0.062*** (0.006)	0.049*** (0.004)	0.061*** (0.006)	0.048*** (0.004)	0.089*** (0.008)
HS 4-digit product * country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Semester-year fixed effects	No	No	No	No	No	No
Country of origin * semester-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	210,211	153,102	210,211	153,102	178,110	126,214
R-squared	0.777	0.783	0.777	0.783	0.785	0.792

**Panel B. Excluding milk and milk products, and pet food from treatment group**

	Dependent variable: log import value					
	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015	Sample period: 2010-2015
	Annual shocks	Semester shocks	Annual shocks	Semester shocks	Annual shocks	Semester shocks
	Excluding milk and milk products (in 2014-2015) and pet food (in 2015) from treatment group					
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator for FVA products * Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)	0.032* (0.017)	0.036* (0.019)				
Indicator for FVA products * Post-reform indicator (2015)			0.047** (0.020)	0.046** (0.022)		
Indicator for FVA products * Post-reform indicator (2 <sup>nd</sup> semester of 2014)					0.004 (0.026)	0.010 (0.028)
Export supply shock - log country of origin total exports to world	0.116*** (0.004)	0.098*** (0.008)	0.116*** (0.004)	0.098*** (0.008)	0.113*** (0.004)	0.097*** (0.009)
Import demand shock - log Albania imports from world	0.019*** (0.002)	0.038*** (0.002)	0.019*** (0.002)	0.038*** (0.002)	0.019*** (0.002)	0.038*** (0.002)
Import demand shock - log Bulgaria imports from world	0.040*** (0.005)	0.058*** (0.009)	0.040*** (0.005)	0.058*** (0.009)	0.037*** (0.006)	0.046*** (0.010)
Import demand shock - log Greece imports from world	0.043*** (0.005)	0.059*** (0.010)	0.043*** (0.005)	0.059*** (0.010)	0.046*** (0.006)	0.054*** (0.011)
Import demand shock - log Serbia imports from world	0.050*** (0.004)	0.062*** (0.006)	0.050*** (0.004)	0.061*** (0.006)	0.048*** (0.004)	0.088*** (0.008)
HS 4-digit product * country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Semester-year fixed effects	No	No	No	No	No	No
Country of origin * semester-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	209,099	152,137	209,099	152,137	177,224	125,464
R-squared	0.776	0.782	0.776	0.782	0.784	0.792

Notes: Robust standard errors in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5% and 10% confidence levels, respectively. In Panel A FVA products includes food of animal origin (license N853), animal feeds (license I047), food of non-animal origin and food contact materials (license I007). In Panel B FVA products includes food of animal origin (license N853) except milk and milk products in 2014-2015, animal feeds (license I047) except pet food in 2015, food of non-animal origin and food contact materials (license I007). In columns (2), (4), and (6) the import demand shock for Albania is annual (as in columns (1), (3), and (5)) due to data constraints.

**Appendix Table B4. Difference-in-differences regressions for impact of FVA reform on trade excluding neighbor countries from sample**

**Panel A. Full Sample**

	Dependent variable: log import value					
	Sample period:	Sample period:	Sample period:	Sample period:	Sample period:	Sample period:
	2010-2015	2010-2015	2010-2015	2010-2015	2010-2014	2010-2014
	Annual shocks	Semester shocks	Annual shocks	Semester shocks	Annual shocks	Semester shocks
Excluding imports from neighbor countries from estimating sample						
	(1)	(2)	(3)	(4)	(5)	(6)
Indicator for FVA products * Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)	0.029* (0.017)	0.032* (0.018)				
Indicator for FVA products * Post-reform indicator (2015)			0.044** (0.019)	0.041* (0.021)		
Indicator for FVA products * Post-reform indicator (2 <sup>nd</sup> semester of 2014)					0.003 (0.026)	0.009 (0.028)
Export supply shock - log country of origin total exports to world	0.117*** (0.004)	0.099*** (0.008)	0.117*** (0.004)	0.099*** (0.008)	0.114*** (0.004)	0.098*** (0.009)
Import demand shock - log Albania imports from world	0.019*** (0.002)	0.038*** (0.002)	0.019*** (0.002)	0.039*** (0.002)	0.019*** (0.002)	0.038*** (0.002)
Import demand shock - log Bulgaria imports from world	0.040*** (0.005)	0.059*** (0.009)	0.040*** (0.005)	0.059*** (0.009)	0.037*** (0.006)	0.046*** (0.010)
Import demand shock - log Greece imports from world	0.043*** (0.005)	0.059*** (0.010)	0.043*** (0.005)	0.059*** (0.010)	0.046*** (0.006)	0.054*** (0.011)
Import demand shock - log Serbia imports from world	0.050*** (0.004)	0.062*** (0.006)	0.050*** (0.004)	0.061*** (0.006)	0.048*** (0.004)	0.088*** (0.008)
HS 4-digit product * country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Semester-year fixed effects	No	No	No	No	No	No
Country of origin * semester-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	210,446	153,302	210,446	153,302	178,309	126,378
R-squared	0.777	0.783	0.777	0.783	0.785	0.792

**Panel B. Sample of food contact materials and products in same HS 2-digit**

	Dependent variable: log import value					
	Sample period:	Sample period:	Sample period:	Sample period:	Sample period:	Sample period:
	2010-2015	2010-2015	2010-2015	2010-2015	2010-2014	2010-2014
	Annual shocks	Semester shocks	Annual shocks	Semester shocks	Annual shocks	Semester shocks
Excluding imports from neighbor countries from estimating sample						
	(2)	(3)	(4)	(5)	(6)	(7)
Indicator for food contact materials * Post-reform indicator (2 <sup>nd</sup> semester of 2014 & 2015)	0.050** (0.023)	0.051** (0.025)				
Indicator for food contact materials * Post-reform indicator (2015)			0.068** (0.027)	0.072** (0.029)		
Indicator for food contact materials * Post-reform indicator (2 <sup>nd</sup> semester of 2014)					0.012 (0.036)	0.005 (0.039)
Export supply shock - log country of origin total exports to world	0.113*** (0.006)	0.100*** (0.012)	0.113*** (0.006)	0.100*** (0.012)	0.112*** (0.006)	0.101*** (0.014)
Import demand shock - log Albania imports from world	0.020*** (0.003)	0.039*** (0.003)	0.020*** (0.003)	0.039*** (0.003)	0.018*** (0.003)	0.038*** (0.004)
Import demand shock - log Bulgaria imports from world	0.022*** (0.008)	0.032** (0.014)	0.022*** (0.008)	0.032** (0.014)	0.017* (0.009)	0.021 (0.016)
Import demand shock - log Greece imports from world	0.029*** (0.007)	0.038*** (0.014)	0.029*** (0.007)	0.038*** (0.014)	0.033*** (0.008)	0.02 (0.016)
Import demand shock - log Serbia imports from world	0.075*** (0.006)	0.075*** (0.009)	0.075*** (0.006)	0.075*** (0.009)	0.079*** (0.007)	0.119*** (0.014)
HS 4-digit product * country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Semester-year fixed effects	No	No	No	No	No	No
Country of origin * semester-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	121,037	89,169	121,037	90,799	102,764	75,125
R-squared	0.762	0.766	0.763	0.767	0.770	0.776

Notes: Robust standard errors in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% confidence levels, respectively. In Panel A FVA products includes food of animal origin (license N853), animal feeds (license I047), food of non-animal origin and food contact materials (license I007). In columns (3), (5), and (7) of Panels A and B the import demand shock for Albania is annual (as in columns (2), (4), and (6)) due to data constraints.



**Appendix Table B5. Heterogeneity of FVA reform treatment regressions  
excluding milk and milk products, and pet food**

	Dependent variable: log import value							
	Sample period: 2013-2015	Sample period: 2013-2015	Sample period: 2013-2014	Sample period: 2013-2014	Sample period: 2013-2015	Sample period: 2013-2015	Sample period: 2013-2014	Sample period: 2013-2014
	Annual shocks	Semester shocks	Annual shocks	Semester shocks	Annual shocks	Semester shocks	Annual shocks	Semester shocks
	Excluding sampling rates larger than 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sampling rate	0.037 (0.149)	0.004 (0.164)	-0.177 (0.201)	-0.410 (0.266)	0.170 (0.229)	0.127 (0.281)	-0.199 (0.279)	-0.350 (0.384)
Export supply shock - log country of origin total exports to world	-0.294* (0.160)	-0.050 (0.100)	-0.686*** (0.210)	-0.405*** (0.144)	-0.295* (0.161)	-0.050 (0.101)	-0.689*** (0.212)	-0.400*** (0.145)
Import demand shock - log Albania imports from world	0.036 (0.026)	0.010 (0.032)	0.075** (0.035)	0.017 (0.039)	0.033 (0.027)	0.007 (0.033)	0.0760** (0.036)	0.016 (0.040)
Import demand shock - log Bulgaria imports from world	0.118 (0.214)	0.285* (0.162)	-0.386 (0.585)	0.294 (0.249)	0.111 (0.215)	0.300* (0.164)	-0.364 (0.595)	0.314 (0.253)
Import demand shock - log Greece imports from world	0.081 (0.271)	-0.332* (0.172)	0.244 (0.757)	-0.500** (0.252)	0.092 (0.273)	-0.331* (0.173)	0.261 (0.763)	-0.503** (0.254)
Import demand shock - log Serbia imports from world	0.141 (0.096)	0.080 (0.065)	0.169 (0.180)	0.049 (0.113)	0.149 (0.097)	0.079 (0.0656)	0.166 (0.182)	0.045 (0.114)
Broad product group * country of origin fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country of origin * semester-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,316	1,187	882	738	1,305	1,179	870	731
R-squared	0.914	0.914	0.904	0.919	0.914	0.913	0.904	0.919

Notes: Robust standard errors in parentheses. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% confidence levels, respectively. In column (3) of Panels A and B the import demand shock for Albania is annual (as in column (2)) due to data constraints.