DISCUSSION PAPERS

DP/98/2015

International Merchandise Trade Flows: Defining Samples and Identifying Discoveries

Peter Stoyanov

BULGARIAN NATIONAL BANK

BULGARIAN
NATIONAL
BANK

DISCUSSION PAPERS
DP/98/2015

International Merchandise Trade Flows: Defining Samples and Identifying Discoveries

Peter Stoyanov

June 2015

ISBN 978-954-8579-68-1 (печатно издание)
ISBN 978-954-8579-69-8 (pdf)

Publishing, printing and binding: Publication Division of the BNB.

Send your comments and opinions to:
Publications Division
Bulgarian National Bank
1, Knyaz Alexander I Square
1000 Sofia, Bulgaria
Tel.: (+359 2) 9145 1351, 9145 1978
Fax: (+359 2) 980 2425
e-mail: Dimova.L@bnbank.org
Website: www.bnb.bg

Contents

SUMMARY: This paper aims to contribute to the study of export and import product discoveries by examining the definitions of trading samples and discoveries. It first looks at the definition of what constitutes a 'trading sample' by contrasting the results of applying the traditional dollar threshold (at $0, $10,000 and $100,000), four relative criteria (based on value and quantity) and two composite relative criteria (using both value and quantity). The nine examined definitions differ substantially both in the number of flows they tag as samples and in the individual flows they tag as samples. Second, a definition of 'product discovery' — similar to that used in Klinger & Lederman (2011) and Cadot et al. (2011) but allowing for a flexible middle window (the sending samples phase) — is presented and some of its potential benefits are discussed. The three examined definitions of discovery differ substantially in the flows they tag as discoveries. Despite the embedded flexibility in the proposed definition, in the majority of discovery episodes, for both exports and imports, products jump straight to established product status within less than a year. This result is robust across the nine examined definitions of trading samples, and could be linked to models such as Rauch & Watson (2003) and the empirical findings in Besedeš & Prusa (2006). The analysis is conducted at the bilateral level, using UN Comtrade data covering 1996–2012, at the 6-digit HS'1996 product level.

JEL: F14, F19, F10, O31

Keywords: export growth, export discoveries, trading sample, intensive and extensive margins

Acknowledgment: I would like to thank Andrey Vassilev from the Economic Research and Forecasting Directorate of the Bulgarian National Bank, for helpful discussions and detailed comments on earlier drafts of this paper.

Peter Stoyanov
Sofia University "St. Kliment Ohridski"
Faculty of Economics and Business Administration
peter.stoyanov@feb.uni-sofia.bg

Introduction

A strand of recent literature analyzes international trade via its extensive and intensive margins, *i.e.* whether growth happens within the same products or through the discovery (and the subsequent survival or disappearance) of new products. This paper is linked to the empirical work aimed at establishing stylized facts that are, at a later stage, examined and explained by a formal theoretical model. The aim is to contribute by (i) underlining the importance of how we define small trade flows, or 'trading samples', and (ii) by offering a definition of discovery that allows for more flexibility than the ones currently employed in the literature and thus helping enhance the understanding the dynamics of innovation implied by the emergence of new products in a country's export basket. Both issues are directly relevant to policymakers contemplating diversification opportunities. Considering the substantially different results produced by applying the approaches currently available in the literature — such as Klinger & Lederman (2011) and Cadot et al. (2011) — and the more general method described in this paper, there are potentially important implications in distilling the vast amounts of trade data into stylized facts and the subsequent analysis and policy formulation.

The analysis of product discoveries typically employs a base-line period (window 1), a second window during which some magic happens and a discovery is made, and a third window, during which it is verified whether the product has emerged as an established product or has failed to. While the literature often does not explicitly discuss the technical details on the construction of windows, it seems that the common practice is to employ fixed windows along at least one of the three main dimensions (starting point, ending point, and duration). This is unnecessarily restrictive, as it implies that discoveries in all possible commodities, and among all possible trading country pairs, follow the same temporal paths — something that is difficult to justify. This paper offers a more flexible approach, which (a) allows the middle window to be of variable length, within user-defined limits, and (b) allows for an interleaved discovery processes, instead of opting for waves of a fixed window structure whenever data is available over a longer period of time. The flexibility is applied at the lowest level of data availability (reporter-partner-commodity), making it possible for a single commodity to be 'discovered' several times within the same trading country pair, and to have different length of window 2 for each of the discovery episodes.

Some of the definitions of export discovery in the literature allow for the product to be exported as 'samples' before emerging as an established product but to the best of the author's knowledge there has been no systematic effort in defining what constitutes a trading sample. The current practice in the literature is to set a threshold dollar value, usually without formal justification, and to label all trade flows below this threshold value 'samples'. Similarly to the fixed windows structure, this is overly restrictive – the thresholds may well be different for different products within the same analysis. For example, consider the model developed in Rauch & Watson (2003). They model the decision of a developed-country buyer considering buying from unknown less-developed-country suppliers – the buyer may choose to place the full order (and face a potential loss if the supplier does not perform to expectations), or start with placing small orders until sufficient knowledge of the partner is accumulated. Since the size of the small orders is linked to the cost of training the supplier and the size of the full order, it is to be expected that what is considered a small order (or a sample in the terminology of the export discoveries literature) is dependent on both the trading partner and the product.

The discussion in the paper is not linked to any of the many theoretical models but most of the arguments are applicable in cases where the model allows for small vs large flows and existing vs new products. These include both models where the dynamics are generated on the production side (heterogeneous firms, multi-product firms) or on the demand side (searching for a dependable partner). In both cases the proper formulation of 'small flows', or samples, at the stylized facts phase of the analysis is important.

The proposed definition of discovery could be applied in establishing and classifying a set of 'patterns' in the discovery episodes – whether certain products for example require more years of sending samples, whether the length of the samples phase is dependent on the source and the destination of the flow (*i.e.* do north-north, north-south, south-south, and south-north flows require systematically different samples phases), whether the length of the samples phase is linked to the subsequent 'stability' of the flow, *etc.* One potential use of such set of patterns is illustrated in section 3.6, where the results generated by the approach employed in this paper relate to the well-known results in Besedeš & Prusa (2006) on the relatively short median duration of trade flows. Another possible use would be to combine the set of frequently-observed patterns with the 'product space' approach offered by Hidalgo et al. (2007). The product space reflects the typical connections between products, allowing the construction of a set of products that are close to the country's capabilities. This information can be combined with the set of typical discovery patterns (since moving to a new product will constitute a discovery), *i.e.* will allow to estimate the expected speed of the new product becoming an economically important flow.

The paper is structured as follows. Section 1 describes the dataset used in the analysis and the necessary adjustments that must be made prior to the analysis. Section 2 discusses the issue of defining what would constitute a trading sample, and provides quantitative comparison of the results applied to both exports and imports flows. Section 3 describes the need for a more flexible definition of what constitutes a new product discovery, presents the proposed definition with a flexible middle window, and provides a comparison to two of the approaches used in the literature. Section 4 motivates the need for further discussion on the topic and concludes.

1. Data

This paper uses UN Comtrade data on annual export and import flows over 1996–2012. The data is at the 6-digit commodity level, using the 1996 version of the HS nomenclature. Commodity aggregates and "UN special" codes have been removed, leaving a total of 5,111 commodities. Economy aggregates ('n.e.s.' entries and similar) have also been removed, as they do not represent economic entities and cannot be matched to country statistics. The data contains dollar value and quantity information, both in kilograms and in supplementary quantity units. A single quantity indicator was constructed, giving preference to the supplementary quantity units. Where supplementary units were not available in the original data (not reported or reported as zero), net weight in kg was used instead. About 1 percent of the total number of flows have no quantity information reported. In about 1.6 percent of the flows the value for quantity is reported as zero, which is considered to be unrealistic and has been treated as 'not available'.

An important aspect of UN Comtrade data is that different countries use different rules regarding the minimum size of flows that are reported. In 1996 and 2006 about 37 percent of national customs organizations used a threshold value below which the flow was not reported.[1] This issue is very important because it leads to bias in the discovery counts both across countries and across time, regardless of the definition of trading sample or product discovery used. Given the available information, there seems to

[1] See question 32 (1996 survey) and questions 9.01 and 9.02 (2006 survey) in the *International Merchandise Trade Statistics National Compilation and Reporting Practices Survey Results 2006 and 1996*.

be no objective way to distinguish between a flow which is genuinely new (*i.e.* did not exist before) from a flow which had previously existed but was not reported, or even a consistent source listing which countries used what thresholds in which years.

The issue is probably most visible in the case of Poland. Prior to 2004, there are no flows below $50,000 in the dataset. Starting in 2004, the threshold was apparently dropped, and there are flows with reported value as low as $1. Between 2003 and 2004 Poland's exports in the dataset grew by 39.8 percent in terms of dollar value and by 273 percent in terms of number of flows. The sum total of export flows which were below $50,000 in 2004 was just 0.78 percent of total export value but accounted for 67.8 percent of the number of flows. For Poland 2004 was not a 'normal' year — it joined the European Union on 1 May 2004 — and a boost to exports and export discovery activity is to be expected but likely not in such extreme amounts in terms of number of new flows.

In an attempt to mitigate the spurious emergence of false discoveries, flows with dollar value of $2500 and below have been removed from the dataset. This threshold is the one apparently applied by the USA, and is the second-highest after Poland.[2] This decision is costly in terms of number of data points but not so much in terms of total value of flows — between 10 and 25 percent of the number of export flows are eliminated each year but they account for less than 0.2 percent of the respective year's world trade in terms of total dollar value. Ultimately we're left with a dataset of slightly over 68 million exports and 69 million imports flows, covering 5,111 products, 187 reporter and 234 partner economies, over 1996–2012. Quantity data is missing in about 1.7 million of both the exports and imports flows, incl. cases where the zero quantity entries were replaced by missing data.

2. What Constitutes a 'Trading Sample'

For the purposes of research similar to this paper, 'trading samples' are implicitly taken to mean any flows that test a new market, *i.e.* flows before a product can be treated as established.

The traditional approach in the empirical literature seems to be to choose an arbitrary (dollar) value that, in the respective author's opinion, corresponds to the research question they pose, and to treat all flows with a value below that threshold as samples. Influential papers like Helpman

[2] Poland still remains an issue in 2004 — this threshold removes only about half of the sub-$50,000 flows in 2004, meaning there remain about 34 thousand flows (of value between $2500 and $50,000) which are reported in 2004 but could not have been reported in 2003.

et al. (2008) and Hummels & Klenow (2005) do not use thresholds at all. Evenett & Venables (2002) use five threshold levels between 0 and 500,000 US dollars, with the intention to "reduce the likelihood of misclassified imports or economically unimportant levels of imports distorting the analysis",[3] and most often report for a threshold of 50,000 US dollars. Klinger & Lederman (2011) use a threshold of 10,000 US dollars per year but provide no details on how the value was chosen. Cadot et al. (2011) use no threshold at all and explicitly discuss that using a threshold depends on whether the researcher is interested in the overall searching process of companies attempting to enter new markets (no threshold, so that all attempts, including subsequent failures, are captured), or just the successful discovery attempts (flows that eventually matured, *i.e.* ended up above the threshold). Agosin & Bravo-Ortega (2009) examine a more focused question (success stories in Chile), and as such impose a high threshold — one million dollars in 2000 constant prices.

While such definitions may seem intuitive in the concrete cases, as long as data is available,[4] there is little reason to support (a) choosing dollar value as the relevant metric, as opposed to physical quantity, unit value or some other characteristic, and (b) fixing the same value for all products and trading economies. There is no clear alternative, and there are many approaches that could prove useful in certain scenarios. For example, one could argue that, with respect to samples, a quantity-based measure would be better than dollar value, as different companies (or customs administrations) may have different accounting policies to determine the reported dollar value of the samples, but the physical quantity units should be more uniform.[5] Another possibility is to look, at the individual flow level, for step-like dynamics in dollar value, quantity or unit value, with jumping to a plateau signifying the transition from samples to established product. Or one may look for clustering of data points. In short, the possibilities are numerous.

[3] Evenett & Venables (2002), p. 7.

[4] It should be recognized that detailed trade data availability is a fairly recent phenomenon in economics, and a substantial portion of the literature originated before that.

5 Of course, this assumes the available quantity information is of good quality, which is often not the case. An example of present but unrealistic quantity information is the exports of product 870210 "Motor vehicles for the transport of ten or more persons, including the driver; With compression-ignition internal combustion piston engine (diesel or semi-diesel)" by Jordan to Iraq. In 2001 the dollar value was $17m, but quantity was reported as 284kg — a clearly unrealistic weight for even a single passenger motor vehicle carrying ten or more persons.

2.1. Alternative Definitions of 'Trading Sample'

This section of the paper tests several definitions of what constitutes a trading sample.[6] First is the standard approach, with three dollar values for the fixed threshold — no explicit threshold (labeled **s0**), which is equivalent to a threshold of $2500, since flows below $2500 were removed, as discussed in section 1, and thresholds of $10,000 and $100,000, labeled **s10** and **s100** respectively.[7,8] Second, a set of four relative criteria are used, separately based on value and on quantity. Definitions labeled **q1** and **v1** consider a flow as a sample if it falls, in terms of quantity or dollar value respectively, in the lowest percentile of all flows of the same commodity,[9] across all years and all trading pairs. Definitions **q5** and **v5** are the same but consider the bottom five percentiles as flows of samples. Lastly, two composite criteria, **c1** and **c5**, are constructed which require that both **q1** and **v1** (for **c1**) and **q5** and **v5** (for **c5**) tag a flow as a sample to consider it as such.

Table A.1 and Table A.2 in Appendix A provide information how the results of the application of the different definitions of 'sample' stack against each other, for exports and imports respectively. The alternative definitions of 'sample' were applied to all flows in the database, and each flow was classified as a trading sample or as an established product. Each 2x2 block in the table compares the result of two criteria — in how many cases they agree on the classification (both produce a 'sample' or both produce an 'established product'), and in how many they produce opposite results (the

---

[6] Outside of those intended to mimic the approaches in the literature, the examined definitions and the respective threshold values were largely selected only considering what type of information is available in the dataset (value and physical quantity), and should be seen as proof-of-concept examples rather than something with firm theoretic background.

[7] Again, because all flows below $2500 have been removed, a flow is defined as a sample if its value is between $2500 and $10,000, respectively $100,000. This consideration applies to all value-based definitions.

[8] Note that the s10 threshold is not equivalent to the threshold used by Klinger & Lederman (2011) — the dollar value is the same but is being applied at the bilateral level here, as opposed to aggregate exports in their paper.

[9] Comtrade provides quantity measures in net weight (kg) for all products, as well as a supplementary quantity unit which differs among products, and often even within products. As described in the data section, supplementary quantities were used where available, as they are more 'natural'. Percentiles were calculated on a (commodity, supplementary unit) pair basis. Each pair comprises all trade flows for a given product, across all years and all partners, whenever the same supplementary quantity was used in the reporting. A commodity reported in different supplementary units by different countries is taken to be different products for the sake of tagging samples. Thus, there are 5,111 individual commodities in the HS'1996 classification, but in the calculation of the relative thresholds there are 13,531 (commodity, supplementary quantity unit) pairs. This may introduce some bias as the decisions of different customs administrations to use different supplementary units are unlikely to be random in nature, but the alternative — using only net weight in kg — seems even more biased.

flow is classified as a sample by one of the criteria, and as an established product by the other).

There are two important results. First, the two sets of relative criteria (**q1** and **q5**, and **v1** and **v5**) produce substantially fewer samples than using a fixed threshold, at the chosen thresholds. For example, the definition **s10** tags 14,499,631 flows as samples, while **v5** only tags 3,407,133. Most of the flows tagged as a sample by **v5** are also tagged by **s10** though, indicating that flows that are in the bottom 5 percentiles tend not to exceed $10,000 in value. The issue, however, is in the remaining circa 11m flows that are tagged as samples using the $10,000 fixed dollar threshold but are not in the bottom 5 percentiles — these could be low-value flows in industries where low-value flows are the norm rather than an expression of search activity. Under the **s10** definition, such flows would never mature into established-level products. Whether a bilateral-level flow of less than $10,000 per year should be seen as a mature product of sufficiently high economic importance depends on the concrete case and the research question.[10] Among other considerations, the scope of the individual leaf elements in the HS nomenclature varies substantially, with some product definitions being more detailed (*i.e.* restrictive) than others.

Reconciling this issue would require either lowering the dollar amount of the fixed thresholds or increasing the relative cutoff point. To get a number of flows tagged as samples by the value-based relative definition broadly similar to those by **s10** would require defining the bottom 20–25 percentiles of the flows by value as samples. For exports, setting the cutoff at the 20th percentile yields 13,643,609 flows tagged as samples, vs. the 14,499,631 tagged by **s10**. The two definitions are much closer — they agree in 12,264,384 cases and disagree in 3,614,472 — but it is not clear whether a full one-fifth of the value distribution of the flows should be considered to be samples. Conversely, lowering the fixed dollar threshold would negatively impact the economic importance of those flows.

The second important result is that the disagreement between the different definitions is usually substantial, meaning that even if the alternative definitions tag a similar *number* of flows as samples, these are usually *different* flows. This is most visible when comparing the quantity-based to the value-based definitions — for exports, **q5** tags 3,006,594 flows as samples and **v5** tags 3,407,133. The overlap is just 995,451 flows though, and the defini-

[10] While the amount may seem small from the point of view of a developed-country business, the situation in a developing country may well be dramatically different. For example, analyzing Rwanda's export diversification, Chandra et al. (2007) (p. 161) conclude that "Each US$1000 increment of a non-traditional export is a precious achievement."

tions disagree on over 4.3 million flows. One possible explanation is that, if the analysis is carried out on the basis of dollar value alone, one cannot distinguish between flows of differing qualities within the same commodity – a flow may be an established product of low quality, hence low unit value, and even though it is being exported in sufficiently high quantities, it may remain under the value-based definition of 'established product' and never be tagged as one (a false negative). Similarly, it would be possible that sending out just a few samples of an expensive variety causes the flow to erroneously reach maturity status (a false positive).

Definition **s100** is an interesting case – it classifies more flows as samples (about 41-42m, for both exports and imports) than as established products (about 27m). Considering that a large number of small (sub-$2500) flows have already been removed from the dataset as discussed in section 1, this should probably be interpreted that setting a fixed threshold of $100,000 pushes a very high share of international trade in the realm of search activity rather than established product trade – a result that might be too strong for most general research efforts. This serves to underscore a tentative conclusion based on the considerations above – the choice of a definition of what constitutes a sample is important and should be tailored to the concrete research question and the theoretical basis.

2.2. Transition between States (no trade, sample and established product)

Another way to compare the definitions is to look at the probabilities for a given flow to transition from one state to another state in the subsequent period (the transition matrices) for the different definitions. The different sample definitions imply three possible states[11] in which a flow can be – non-existent state (**o**), sample state (**s**), or established product state (**X**).

As highlighted by Helpman et al. (2008), there is no trade between a significant share of all possible pairs of economies, esp. when looking at a very disaggregated product level, but those 'zeroes' carry important information. Generating the full set of relationships covering all possible country pairs for all commodities, however, is technically difficult to handle – 234 economies trading with 233 potential partners in 5,111 products over 17 years yields over 4.7 billion data points. To simplify the problem, this paper assumes that relationships (at the reporter-partner-commodity level) that never see realization in all years in the dataset have no economic meaning when estimating transition probabilities, since their state never

[11] Definition **s0** only has the non-existent flow (state **o**) and established product (state X) states.

changes. Including them in the analysis would just inflate the probability for a non-existent flow to remain non-existent.

Therefore estimates of the transition probabilities was carried out with only those bilateral trade relationships, at the 6-digit product level, where there have been exports, resp. imports, in at least one year during 1996–2012, *i.e.* where there has been at least one transition from one state to another. It is done in a naïve way — by creating a list of the states in all adjacent years and then counting the frequencies with which each possible pair of states occurs. No attempt has been made to include the prior history in the relationship, or to estimate what happens at the truncation points (the start and end of the period covered in the dataset).

The estimates of the transition probabilities for both the exports and imports flows at the level of the whole world[12] and for select countries are presented in Appendix B. Except for **s100**, which is again an outlier, several interesting observations can be made for the world-level flows.

First, both the non-existing-flow state (when a product is not traded) and the established-product state are quite stable, both in terms of high probability to remain in the same state and across the different definitions of what is a sample. The probability for a non-existing flow to remain non-existing in the subsequent period is about 86 percent in all cases, and as expected is constant across the definitions. The probability of an established product to remain so varies in a very narrow range, from 74 to 79 percent.

Second, the trading-sample state **s**, whatever the definition except **s100**, is not stable — a flow that is classified as a sample in the current period has a (usually substantially) higher probability to become extinct (go to state **o**) or to become an established product (go to state **X**) in the subsequent period than to remain a sample. This is to be expected — samples by definition are an expression of search activity and over time should either disappear or mature into an established product — but the speed with which this happens may be a little surprising (more than half of sample flows are expected to disappear each year).

Third, samples are typically about two times more likely to disappear than to become established products. The probability of not being exported in the subsequent period is typically in the range of 0.50–0.65, and the probability to mature to an established product is never higher than 0.37. This result can be linked to the findings in Besedeš & Prusa (2006) — since it is trade flows in samples that disappear quickly, the observed low median

[12] Here 'world' is defined as a list of all countries in the dataset, *i.e.* the transition probabilities have been estimated on all available bilateral flows, not on some sort of world aggregate.

duration and high mortality rates in the early years of trade flows is logical. Once a relationship matures into an established product, it is likely to survive substantially longer. The result is also in line with the theoretical model of Rauch & Watson (2003).

Fourth, when using the relative and the composite definitions of what constitutes a trading sample, flows that have just emerged are typically more likely to jump straight to established-product status than to be traded as a sample. The probability to transition from state **o** to state **X** is usually at least an order of magnitude higher than the probability to go from **o** to **s**. This is not the case for the fixed-dollar-value thresholds.

Definition **s100** again is an outlier. Under it non-existent flows and established products are stable, as in the other definitions. The transition probabilities of samples, however, are markedly different – there is a 0.5313 probability that a sample export flow will remain a sample, a 0.3654 probability that the flow will disappear, and a probability of just 0.1033 that an export sample will emerge as an established product. The probabilities in the imports flows are similar. These estimates are at odds with basic economic logic. Samples are supposed to reflect search activity – to either succeed or fail – not to remain stable. This contradiction is an indication that a fixed threshold of $100,000 is too high when working at the bilateral level.[13]

The observations above are for flows at the world level, where there is little expected difference between exports and imports flows – differences would come from differing reporting times and/or reporting requirements (f.o.b. or c.i.f.). Consequently the estimated probabilities for exports and imports are essentially the same. This does not apply at the economy level, as illustrated in the other tables in Appendix B.

Overall, the differences in the results of applying the alternative definitions of trading sample to the same dataset underlines the need to use a definition appropriate to the research question and its theoretical foundation. Additionally, sending samples to a potential partner is a micro phenomenon – it is firms that trade, not economies. Trade flows observed at the economy level are aggregates and may well depend on the size of the reporter and/or the partner economy or have other aggregation specifics when going from micro to the macro level.

[13] If similar results were obtained in the other definitions of trading sample, another explanation would be that samples take longer than one period to either disappear or mature into established products.

2.3. Defining Samples: an Look at Bulgaria

The implications of using different definitions of what constitutes a trading sample are well illustrated when looking at country-level data. Let's assume a hypothetical policymaker is interested in stimulating export diversification.[14] The economic literature offers a range of options, from export promotion (marketing and facilitating the search for partners), to mitigation of transportation costs and domestic barriers. In a setting of scarce public resources, some indication would be needed where to focus the attention, and the estimated transition probabilities offer a starting point.

One could start with the definition that does not allow for samples at all, **s0**, which says that there is a probability of 0.1172 for a new flow to emerge within the next period (Table B.2). The hypothetical policymaker would then be interested in distinguishing established-product flows from search activity. The distinction is important both from the point of view of value of trade and volatility of trade flows. This is where the differences in the definitions of what constitutes a trading sample become apparent. If we take the **s10** definition, we would conclude that the share of new exporting relationships that are able to find partners quickly (*i.e.* jump straight to established-product exports in the subsequent period) is very close to the share of those that need sending out samples — the probability for a new flow to emerge as an established product is 0.0616 vs. 0.0556 to emerge as a sample. Were we to adopt one of the relative definitions, however, this is no longer the case — for example, under **v1**, and with the rest of the relative definitions providing similar conclusions, the probability of a new flow emerging as an established product is more than an order of magnitude higher (0.1138 vs. 0.0034). This latter set of probabilities indicates that likely local businesses face little constraints in their ability to locate partners and quickly establish a relationship. Thus, as a first approximation, the focus would go to measures lowering domestic trading costs rather than trade promotion abroad.

Similarly, looking at the stability of existing established products — the probability for an established-product flow to remain so is about two thirds (between 0.6447 and 0.6611) for all definitions except **s100**. What happens to the other one third — the ones that drop back to samples or disappear altogether — is quite a different story, though, depending on the definition. According to definition **s10**, 26.08 percent disappear, and

[14] Export diversification may happen at the intensive margin (equalizing shares of existing exports) or the extensive margin (new products). This is further discussed in section 3.1. Also, see for example Dennis & Shepherd (2011), Persson (2010) and Feenstra & Ma (2014) on the link between trade facilitation and diversification.

7.81 percent drop from established-product status to trading-sample status. This dynamic is difficult to explain. At the micro level, it is plausible — an exporter's existing flow may be disrupted for whatever reason, and drop to zero. The exporter would naturally start looking for new partners, and will send out samples, so the observed exports flow would drop from established-product level to samples level. At the aggregate level, however, for a flow to drop from established-product level to samples level would require either that there is a single exporter who has lost a previous partner and is now seeking new partners by sending out samples, or that the cutting off of existing established-product relationships happened to all exporters simultaneously (and there are just a few exporters). Even in the simplest case with only two exporters which initially trade at established-product levels, if one were to suffer a disruption of the trade flow, the other exporter's flow would be sufficient to keep the aggregate export level at established-product levels.

The relative definitions of trading sample offer a more plausible story — the probabilities for an established product flow to drop back to samples level rather than to zero is very small, peaking at 2.2 percent in the case of **v5**.

3. Identifying Export Discoveries

This section outlines the approach to identifying product discoveries used in the literature, and describes a proposed extension. Most of the literature focuses on export products only or treats imports as a source of intermediate inputs but the same approach can be applied to import flows and to intra-industry trade (two-way flows) as well. For ease of presentation, the exposition here talks about export discoveries. Application to imports and two-way flows is straightforward – the former is essentially identical to exports and the latter would just require more possible states, *i.e.* would for example have to include states like 'exports at an established-product level, imports at samples level'.

3.1. Extensive and Intensive Margins in International Trade

There is no single, universally accepted definition of extensive and intensive margins in the literature. Besedeš & Prusa (2011) discuss some of the approaches in the context of getting to a definition of what constitutes a 'trade relationship'. Some authors define it at the product level alone (*e.g.* the emergence of a new product in an economy's exports basket, irrespective of partners). Others define it at the country level, *e.g.* the emergence of a new partner country. The third approach is to go lower, at the country-product level, *i.e.* discoveries to be defined as any of (a) exporting a 'new' product to a 'new' partner, (b) exporting an 'old' product to a 'new' partner or (c) exporting a 'new' product to an 'old' partner. This reporter-partner-commodity level definition is used in this paper. More formally, using the labels from Comtrade, a trade relationship is a triplet (reporter, partner, commodity), for any given year. This definition assumes that there are at least two groups of factors influencing the emergence of a discovery (and trade in general) — domestic factors, which affect the production side (the ability of the reporting economy to manufacture the product and bring it to market) and market factors at the partner economy, which affect the demand for the product.

For a discovery to emerge, we must have both the ability of the reporter to manufacture the product and the existence of demand at the partner economy for this product. The discovery event may be related to any of the steps in the process, from originally developing the product and/or mastering the production process to discovering or creating foreign demand for it (in each individual partner economy) and actually shipping the product to the destination. Thus exporting the same product (which may have been produced for domestic consumption for some time before being exported) for the first time or to a new partner-economy involves some element of discovery. In other words, as Klinger & Lederman (2006) put it, a 'discovery' may happen within the production possibility frontier and does not necessarily imply pushing the production possibility frontier outwards.

3.2. Identification of Export Discoveries: Overview

The identification of an export product discovery usually involves establishing an identification procedure, possibly with a filter for 'suspicious' cases. Klinger & Lederman (2004) and Klinger & Lederman (2011) look at the introduction of new export products as an aspect of economic development and diversification, using data covering a ten-year period (1994–2003). Their identification of an export product discovery (defined as the successful export of a product by an economy that had not exported it before) requires three periods:[15]

1) A baseline period, during which there are no exports of the product. The length of this window is 3 years (1994–1996).

[15] This description is based on Klinger & Lederman (2011). Their earlier working paper uses a more restrictive but fundamentally similar approach — a product is considered to be a discovery if exports were less than $10,000 at the beginning of the period but above $1,000,000 at the end.

2) A second period, during which something happens and the product is 'discovered'. For the purpose of the identification it is not necessary to know what and how caused the discovery to emerge. The length of this window is 5 years (1997–2001). A product can emerge as a discovery in any year within this period.

3) A third window (2002–2003), in which the product is confirmed to be established, *i.e.* it is being exported in values above a predefined threshold (export value of $10,000). Exports below the threshold are considered trading samples. The length of this window is 2 years (2002–2003).

Let's label this approach KL.[16]

Besedeš & Prusa (2006) look at the survival rates of trade flows (in a number of datasets). Even though they do not formally define and analyze discoveries, in the context of their analysis a discovery is simply a product which has not been traded in the previous period — substantially more relaxed that the approach in KL.

Cadot et al. (2011) provide a different classification method, hereafter labeled as CS. A product is identified as a discovery (in the current year) if it has not been exported for the two preceding years, and has been exported in any value, however small, in each of the two subsequent years.[17] Since a discovery can be flagged using 5 years of data, they employ a "moving 5-year sub-sample" to cover the full period of the available data. Importantly, since their research question is different than that of KL, the CS approach does not distinguish between established products and samples. Due to the lack of a minimum threshold and the short length of the period, the definition captures the small-scale activity of entrepreneurs trying to discover new markets. For example, two years of no exports followed by three years of exports at samples level will be counted as a discovery.

[16] It should also be noted that they apply the identification process on data which is disaggregated only along the commodity dimension, *i.e.* they look at the aggregate exports at the HS 6-digit product level. In addition, they employ a filter to identify 'suspicious' cases. Both of these are outside of the scope of the current discussion.

[17] They define discoveries as "lines that were inactive in the country's export trade in the preceding two years but were exported in the following two years (two-years cutoff)" (p.596-597). This formulation is not explicit about whether there must be exports in the *current* year. The analysis in the present paper assumes that the current year's exports must be positive for a flow to be tagged as a discovery. The alternative — tagging a non-existent flow as a discovery — does not seem desirable. This paper uses only the two-year-cutoff version, which according to the authors "strikes a balance between the very conservative definition used by [Klinger & Lederman (2006)] and the very liberal one used by [Besedeš & Prusa (2006)]".

An important issue that both the CS and KL methods do not address is the possibility that the discovery process may be of different lengths for the different products. And, since both use data aggregated across partners, they implicitly assume away any differences with respect to partners.

The lengths of the first and third periods are more or less fixed along products, as it would be difficult to argue that product X will need a different number of years to be verified as an established product than product Y, or that one product will need fewer years for the baseline period test than another. The lengths of these periods are of course debatable – *e.g.*, does it take two years to verify that a product is not being traded, or should we require five or more years? – but once set, the lengths should apply equally to all products. This leaves the middle period, and it would seem sensible to allow that product X may require a longer period of being exported as a sample than product Y, before it becomes established (or fails to). In other words, the identification procedure should allow for a flexible duration of the middle period while the lengths of the first and third periods remain the same. This is even more valid when the analysis is done at the most disaggregated level. Attempting to export widgets to a developed market economy like the USA or Germany seems likely to take much less time than attempting to export the same product to a less-developed economy with poorly functioning markets and more difficult access to information regarding potential partner companies. This argument can be made for many of the traditional determinants of trade – the easier the trade between two partners, for whatever reasons, the shorter the expected duration of the discovery process should be, and *vice versa*. Finally, over time conditions in the reporting and/or partner economy may change sufficiently to impact the speed of discovery for one and the same product. That is, if a product is first 'discovered' and then 'forgotten', rediscovering it may require a different number of years of sending samples.

Let's illustrate the problem using the KL approach. Consider a product which has not been exported for three years (1994–1996 in their case), which then appears as an established product in 1997 and survives for five years (*i.e.* is exported in above-sample-threshold values for each year in 1997–2001). The product then disappears in 2002. In the KL approach, this product will not be flagged as a discovery, as it is not being exported in sufficiently high value in their control window which has fixed starting and ending points (2002 and 2003 respectively). Should this product have been identified as a discovery? Considering the results in Besedeš & Prusa (2006) – that the duration of most trade flows is relatively short (median duration of exporting a product to the US on the order of two to four years) — a product which has been exported for five years in large values should indeed have been deemed a discovery. It would have been flagged as a discovery using the CS method but it still uses a similar rigid 2+1+2 structure.

The positioning of the three periods (baseline, emergence, established product) within the available data is a second issue which has not seen sufficient discussion in the literature. The KL paper uses ten years of data, and constructs a single set of windows. An obvious question is how to proceed when data for a longer time period is available. There seem to be several possibilities:

- Use several sub-samples (a.k.a. waves, or moving windows) of a length less than the total number of years for which data is available. For example, the dataset used in this paper covers 17 years (1996–2012), and we could construct eight consecutive waves of ten years each (wave 1 covering 1996–2005, wave 2 covering 1997–2006, through wave 8 covering 2003–2012). This approach is used in CS.
- Use longer lengths of the three periods, so that the whole period is covered. It would seem difficult to defend this approach since the length of the windows should motivated by factors independent of the length of the available data. Recall that the window lengths and start/end points are fixed and are the same for all products.
- Use a more flexible definition of the lengths of at least some of the windows, *i.e.* allow for different *lengths* and different *starting points* of the three periods to be applied for different product-partner flows. As argued above, it would be difficult to motivate different *lengths* of the first and third windows but this does not apply to their *starting points*. It would seem that with respect to the middle window, when the magic happens, window length and window starting point should be as flexible as possible to reflect the different specifics of the partner-commodity relationships (*e.g.* a reporter might succeed in establishing a relationship for product $\mathbf{c}$ with partner $\mathbf{p}_1$ after only a single year of sending samples, but may need three or four years to establish a relationship for the same product with another partner, $\mathbf{p}_2$).

This paper uses the third option as it is the most flexible. The details are presented in the next subsection.

3.3. Identification of Product Discoveries with a Flexible Middle Window

This subsection outlines the procedure for the identification of product discoveries used in this paper. In general terms, the procedure uses the same three windows (baseline, emergence, established product test) but allows for the length of the middle period to differ among reporter-partner-product triplets, and allows it to start anywhere within the available data.[18]

Before proceeding, let us define a simple rule to convert the quantitative exports[19] data into qualitative data. For commodity **c** exported by reporter **r** to partner **p** during period **t** define:

$$F^c_{rpt} = \begin{cases} \mathbf{o} & \text{if no trade is reported} \\ \mathbf{s} & \text{if the flow is determined to be a sample} \\ \mathbf{X} & \text{if the flow is determined to be an established product} \end{cases}$$

We can then represent the available data on every reporter-partner-commodity relationship as a vector $F^c_{rpt} = (F^c_{rp1}, F^c_{rp2}, ..., F^c_{rpT})$, where **T** is the total number of years in the sample. From a software implementation point of view, the vector is simply a string made up of **T** characters, each of which corresponds to one period in the data. We can then identify patterns in the trade flows by identifying patterns in the text strings.

Using these text strings, we can construct arbitrary flexible definitions of windows patterns. Let's illustrate the approach by formalizing the following discovery definition (hereafter labeled PS), which is broadly similar to those in the papers cited above but allows for a flexible middle window. Define a discovery episode as a case where a product:

1. (*baseline test*) Has not been exported for at least three consecutive years, then
2. (*emergence*) Has been exported in above-zero levels[20] for zero to five consecutive years, inclusive, then

[18] A recognized deficiency of this approach, common with all explored models, is the lack of accounting for censored and/or truncated data. If a discovery episode has begun before the period of the available data, or ends after it, it will not be identified. A potentially fruitful line of further research would be to borrow from the survival analysis literature to augment the identification procedure, esp. when running explanatory models.

[19] This definition deals with one-way trade flows but can easily be extended to incorporate two-way trade flows as well by introducing more codes. Uni-directional flows (exports or imports) require three codes, as listed in the definition above. Bi-directional flows require nine codes, e.g. a code corresponding to 'exports of samples and imports of an established product', *etc*.

[20] 'Above-zero levels' means any combination of **s** and **X** that does not contain two consecutive **X**'s. The emergence period cannot contain two consecutive years with established-product level exports **(XX)**, as this would trigger the established-product criterion from Window 3.

3. (*established product test*) Has been exported as an established product for at least two consecutive years immediately following period 2.

That is, we would be looking for substrings **oooXX** (zero-length of window 2), **ooosXX** (window 2 of length one, with samples-level exports), through **ooosssssXX** or **oooXsXssXX** anywhere in the F^{c}_{rp} strings. Since our sample covers 17 years, it is possible to have several discovery episodes for a single **(r, p, c)** relationship – we can (and actually do) observe the shortest discovery episode (the five-character string **oooXX**) up to three times.

This definition of 'product discovery' is flexible in the following sense. First, only windows 1 and 3 have fixed lengths. Window 2 does not, and may vary between zero and five years in length, even within the same **(r, p, c)** triplet. Let's illustrate with an example. In our data, we have the string **oooXXoooooXsXXosX** that represents the trade for a single relationship **(r, p, c)** over the 17 years.[21] There are two discovery episodes: (1) over 1996–2000 inclusive (**oooXX**), with a zero-year length of window 2 and (2) over 2003–2009 (**oooXsXX**), with a two-year window 2, **Xs**, where the established-product exports in 2006 are counted as part of the samples window because they are not followed by another year of above-samples exports.

Defined like this, the combined length of all three windows varies between five and ten years, inclusive. While similar to the KL approach, the new element here is that the established-product test (window 3, two consecutive years of established-level exports) may start at any time, rather than being constrained to two fixed years at the end of the sample (or sub-sample if using waves). Second, the definition is flexible in the sense that it does not have a fixed starting point, *i.e.* the first year of window 1 may be any year within our sample, up to **T – 5**.[22] This is important if we think that discoveries will happen with different speeds (different lengths of the samples phase) for different products and/or trading partners, or may have begun at different points in time. It also takes into account that discovered products may disappear before the end of the sample.

Finally, there is some built-in flexibility in the definition of the length of window 3. It is the minimum length of the window that is fixed, and it will

[21] Exports from Brazil to India of HS'1996 product code 860719, "Bogies, bissel-bogies, axles and wheels, and parts thereof :– Other, including parts". Trading sample was defined as a flow with a value of less than 100,000 dollars per year (**s100**).

[22] Since we are looking for a pattern representing at least five years, it cannot be observed if we have fewer than five years of observations left. Similarly, since we require a three-year no exports in the baseline test, there can be no discoveries in the first three years in the data. Truncated and/or censored data is explicitly not accounted for.

expand as long as exports remain at established-product levels. This is not directly useful to the identification procedure here but may be used for example to analyze the survival rates of newly-established products.

3.4. Defining the Year When the Discovery is Made

Whatever the definition of 'discovery', it is a multi-year process, and once a discovery episode is flagged, it must be assigned to a particular year (or years). There seem to be two possibilities, putting the emphasis on different aspects of the discovery process. One is to count a product as a discovery in the year when it first appears as a positive flow (first year of window 2), *i.e.* to focus on when the first attempt is made irrespective of how long it took to reach maturity. The second is to count a flow as a discovery in the first year when it becomes mature, *i.e.* when it started being exported at above-sample levels (the first year of window 3). The longer the 'samples' phase allowed by the identification methodology, the larger the possible discrepancy between the two points.

KL and CS opt for the first approach. In the case of CS the choice is mostly irrelevant – as there is only a single year in window 2, the difference between the two approaches of assigning a discovery to a year is always exactly one year. In the KL approach, there is some uncertainty – the first positive flow may emerge in any of the five years in the fixed window 2.

Using the KL and CS definitions of a discovery episode but assigning it to the first year when the product becomes mature (first year of window3) is of little use. In the CS case, as already discussed, the difference is always exactly one year, and it is not clear what the benefits of shifting the counting from **t** to **t+1** would be. In the KL approach counting discoveries in the first year of window 3 is counterproductive – window 3 always starts at the same fixed year and all discoveries will be assigned to just a single year.

In the PS definition of discovery proposed in this paper, there is always a variable difference between the first year of window 2 and window 3 – from zero (*i.e.* when there is no window 2 at all, the product jumps straight to established product status), to a maximum of five years. In addition, since the start points of the windows are flexible, they would correspond to different years. To remain closer to the KL and CS approach, here a discovery is assigned to the first year of window 2, or if it is of zero length, to the first year of window 3.

3.5. Comparing the Results of the Three Definitions

This subsection discusses the results of applying the three definitions of 'discovery' (a.k.a. patterns[23] to be found in the text strings representing trade) to the Comtrade 1996–2012 dataset.

Two of the three patterns explored here — KL and CS — are intended to mimic the approach in the respective papers. The similarity is only in the way the identification windows are constructed. The present analysis is done at the reporter-partner-commodity-year level, an important difference compared to the original sources, which are done at the level of aggregate exports (*i.e.* reporter-commodity-year). In the KL case, this paper also uses several waves, since the time period covered in the dataset is longer.

Similarly to the discussion of what constitutes a trading sample, the different approaches to defining a discovery yield markedly different results. This is illustrated using three simple comparisons, out of the many that can be constructed on the basis of the individual trade flows.

The first comparison — Table C.1, Table C.2 and Table C.3 in Appendix C — is a simple count of the number of flows tagged as a discovery by each of the three discovery definitions, using each of the nine definitions of trading sample. Since the CS definition does not use samples, the numbers of discoveries found is the same for all definitions of trading sample. It tags as discoveries about 4 percent of the exports flows and a slightly higher number of the imports flows. As expected, the more stringent definition KL tags a substantially lower number of flows as discoveries than CS does, and the more flexible PS definition tags relatively more flows (except for **s10** and **s100**).

The second comparison is a cross-tab of the discovery counts for the same definition of trading sample and the different definitions of discovery, at the individual flow level (Table C.4, Table C.5 and Table C.6). Even if the total number of flows tagged is similar, *different* flows are being tagged by the different definitions. Since this comparison may be too restrictive — it requires that two definitions of discovery agree on the classification of a flow in the same year — Table C.7 and Table C.8 present a third comparison, a cross-tabulation of how many times a reporter-partner-commodity relationship (*i.e.* stripping away the time dimension) has been tagged as a discovery by the different definitions, even if not in the same year.

[23] In terms of implementation, the three verbal descriptions of what constitutes a discovery were translated into regular expressions, which in turn were applied on the text vectors describing trade flows. The regular expressions are as follows: **o{2}[sX]{3,}** for CS, **o{3}.{5}X{2,}** for KL and **(o{3})([sX]{0,5}?)(X{2,})** for PS.

3.6. The Length of Window 2

Since the main feature of the proposed definition of product discovery is the flexibility of the middle window, it only makes sense to exploit this flexibility. A look at the observed length of window 2 (Table D.1 and Table D.2) reveals a surprising result – most episodes of discovery have no samples phase, they jump from non-existence straight to established-product status. The results vary slightly depending on the definition of trading sample chosen, being stronger in the relative definitions and weaker in the fixed-dollar-threshold ones. Of course, if the thresholds for the sample definitions are raised, the no-samples episodes decline in number.

Considering the data is on an annual basis, a samples phase of length zero does not necessarily mean that there is no samples phase at all. Rather, the flow goes through the samples phase and into maturity within the same calendar year. The result is even stronger considering the second-frequent length of the samples phase is one year – something that may easily happen if the end of the calendar year comes before the flow reaches maturity, even if the total length of the samples phase is less than twelve months. An obvious line of future research is to examine the phenomenon using a higher-frequency dataset, e.g. using monthly data.

4. Discussion and Conclusion

This paper draws attention to and examines two technical issues in the analysis of new product discoveries that would benefit from further discussion as the current approach is lacking in terms of theoretic foundations and overly restrictive in some important aspects.

First is the definition of what constitutes a trading sample. Nine alternative definitions are examined – three fixed-dollar-value thresholds (the usual approach in the literature), four relative definitions (defining as samples the flows comprising the bottom 1 and 5 percent of the distribution, by value and by physical quantity), and two composite criteria (a flow is considered a sample if it falls in the bottom 1, resp. 5, percent of the distribution by both value and physical quantity). The alternative definitions produce substantially different results thereby making the choice of definition of trading sample an important step in any analysis. Further research is needed in defining what constitutes a 'trading sample', in particular linking the definition to concrete models of search activity and/or production structure. In this respect, the definitions examined in this paper should be seen as illustration of the importance of the issue and a starting point for further discussion rather than an exhaustive examination of possibilities.

Second, the paper discusses a deficiency in some of the empirical definitions of 'new product discovery' used in the literature, namely the fixed structure of the three windows of the identification procedure, and proposes a procedure which uses a middle window (the sending samples phase) of variable length to account for potential differences in the specifics of trade relationships at the reporter-partner-commodity level. The proposed procedure can be tailored to specific research needs. For example, by observing the length of the third window, one can obtain a rough approximation to a more rigorous duration (survival) analysis of discoveries. Or, the length of the middle window can be used to assess the importance of samples. The dynamics of the lengths of these windows can be especially interesting in cases where we have repeated discoveries of the same product in the same reporter-partner pair; or to contrast the differing experiences of a single reporter exporting the same product to several partners. Of course, as has been done in the literature, this definition of discovery can be inverted to tag episodes where a product has been 'forgotten' or dropped, rather than discovered.

An interesting preliminary result, coming from both lines of analysis in this paper, is that the observed length of the samples phase in the vast majority of discovery episodes is zero or one years, indicating that most product discoveries go through the samples phase and emerge as established products within twelve months. The result is observed for both exports and imports flows, and is qualitatively similar across all nine definitions of trading sample used in the paper. It seems also linked to well-known results in the literature (Besedeš & Prusa (2006)). Further discussion and research are needed, however, to determine whether the observed pattern is due to underlying economic logic or just an interesting-but-random pattern that has appeared in a moderately large dataset, or an improper choice of the thresholds for the different definitions.

The proposed definition of product discovery facilitates the process of better understanding the dynamics of innovation, and can serve as a starting point for analyzing the differences across products, trading economy pairs and time. At the same time, it remains just a proof-of-concept without firm theoretic background.

References

Agosin, M.R. & Bravo-Ortega, C., 2009. *The Emergence of New Successful Export Activities in Latin America: The Case of Chile*, Inter-American Development Bank Research Network Working Papers R-552.

Besedeš, T. & Prusa, T.J., 2006. Ins, outs, and the duration of trade. *Canadian Journal of Economics*, 39(1), pp.266–295.

Besedeš, T. & Prusa, T.J., 2011. The role of extensive and intensive margins and export growth. *Journal of Development Economics*, 96(2), pp.371–379.

Cadot, O., Carrère, C. & Strauss-Kahn, V., 2011. Export Diversification: What's behind the Hump? *Review of Economics and Statistics*, 93(2), pp.590–605.

Chandra, V., Li, Y. & Rodarte, I.O., 2007. Commodity Export Diversification in Rwanda - Many Export Discoveries with Little Scaling-Up. In *Rwanda: Toward Sustained Growth and Competitiveness Volume II, Report No. 37860-RW*.

Dennis, A. & Shepherd, B., 2011. Trade facilitation and export diversification. *World Economy*, 34(1), pp.101–122.

Evenett, S.J. & Venables, A.J., 2002. *Export Growth By Developing Countries: Market Entry and Bilateral Trade*, Mimeo.

Feenstra, R.C. & Ma, H., 2014. Trade Facilitation and the Extensive Margin of Exports. *Japanese Economic Review*, 65(2), pp.158–177.

Helpman, E., Melitz, M.J. & Rubinstein, Y., 2008. Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics*, 123(2), pp.441–487.

Hidalgo, C.A. et al., 2007. The product space conditions the development of nations. *Science*, 317(5837), pp.482–7.

Hummels, D. & Klenow, P.J., 2005. The Variety and Quality of a Nation's Exports. *American Economic Review*, 95(3), pp.704–723.

Klinger, B. & Lederman, D., 2004. *Discovery and Development: An Empirical Exploration of New Products*, Washington D.C. World Bank Policy Research Working Paper 3450.

Klinger, B. & Lederman, D., 2006. *Diversification, Innovation, and Imitation inside the Global Technological Frontier*, World Bank Policy Research Working Paper 3983.

Klinger, B. & Lederman, D., 2011. Export discoveries, diversification and barriers to entry. *Economic Systems*, 35(1), pp.64–83.

Persson, M., 2010. *Trade Facilitation and the Extensive Margin*, Stockholm, Sweden., Research Institute of Industrial Economics Working Paper 828.

Rauch, J.E. & Watson, J., 2003. Starting small in an unfamiliar environment. *International Journal of Industrial Organization*, 21(7), pp.1021–1042.

Appendix A. ***Comparing the Definitions of 'Sample'***

Comparison of the results of applying the different definitions of 'sample' to export flows. Each 2x2 block provides the number of cases where two definitions of 'sample' agree or disagree. For example, the bottom-left block in Table A.1 compares **c5** to **s10** and shows that 52,409,956 of the export flows are classified as established products by both the composite criterion **c5** and the **s10** criterion; 995,067 flows are classified by both criteria as samples; in 13,093,147 cases **s10** classified the flow as a sample while **c5** classified it as an established product, and in 384 cases the opposite happened.

Blocks that compare a definition to itself just provide information how many flows the given definition classifies as established products and how many as samples.

Table A.1. **Comparing the Definitions of 'Sample': Exports**

		q1		q5		v1		v5	
		established	**sample**	**established**	**sample**	**established**	**sample**	**established**	**sample**
q1	**established**	65,948,588	0	63,491,960	2,456,628	65,338,014	610,574	62,843,642	3,104,946
	sample	0	549,966	0	549,966	495,013	54,953	338,336	211,630
q5	**established**	63,491,960	0	63,491,960	0	63,062,214	429,746	61,170,835	2,321,125
	sample	2,456,628	549,966	0	3,006,594	2,770,813	235,781	2,011,143	995,451
v1	**established**	65,338,014	495,013	63,062,214	2,770,813	67,546,846	0	64,822,589	2,724,257
	sample	610,574	54,953	429,746	235,781	0	682,876	0	682,876
v5	**established**	62,843,642	338,336	61,170,835	2,011,143	64,822,589	0	64,822,589	0
	sample	3,104,946	211,630	2,321,125	995,451	2,724,257	682,876	0	3,407,133
s10	**established**	52,323,454	86,886	51,859,193	551,147	53,730,006	85	53,724,956	5,135
	sample	13,625,134	463,080	11,632,767	2,455,447	13,816,840	682,791	11,097,633	3,401,998
s100	**established**	26,766,113	6,160	26,743,969	28,304	27,405,620	4	27,405,566	58
	sample	39,182,475	543,806	36,747,991	2,978,290	40,141,226	682,872	37,417,023	3,407,075
c1	**established**	65,948,588	495,013	63,491,960	2,951,641	65,833,027	610,574	63,181,978	3,261,623
	sample	0	54,953	0	54,953	0	54,953	0	54,953
c5	**established**	65,164,767	338,336	63,491,960	2,011,143	65,073,357	429,746	63,181,978	2,321,125
	sample	783,821	211,630	0	995,451	759,670	235,781	0	995,451

		s10		s100		c1		c5	
		established	**sample**	**established**	**sample**	**established**	**sample**	**established**	**sample**
q1	**established**	52,323,454	13,625,134	26,766,113	39,182,475	65,948,588	0	65,164,767	783,821
	sample	86,886	463,080	6,160	543,806	495,013	54,953	338,336	211,630
q5	**established**	51,859,193	11,632,767	26,743,969	36,747,991	63,491,960	0	63,491,960	0
	sample	551,147	2,455,447	28,304	2,978,290	2,951,641	54,953	2,011,143	995,451
v1	**established**	53,730,006	13,816,840	27,405,620	40,141,226	65,833,027	0	65,073,357	759,670
	sample	85	682,791	4	682,872	610,574	54,953	429,746	235,781
v5	**established**	53,724,956	11,097,633	27,405,566	37,417,023	63,181,978	0	63,181,978	0
	sample	5,135	3,401,998	58	3,407,075	3,261,623	54,953	2,321,125	995,451
s10	**established**	53,730,091	0	27,405,624	26,324,467	52,410,299	41	52,409,956	384
	sample	0	14,499,631	0	14,499,631	14,033,302	54,912	13,093,147	995,067
s100	**established**	27,405,624	0	27,405,624	0	26,772,270	3	26,772,270	3
	sample	26,324,467	14,499,631	0	40,824,098	39,671,331	54,950	38,730,833	995,448
c1	**established**	52,410,299	14,033,302	26,772,270	39,671,331	66,443,601	0	65,503,103	940,498
	sample	41	54,912	3	54,950	0	54,953	0	54,953
c5	**established**	52,409,956	13,093,147	26,772,270	38,730,833	65,503,103	0	65,503,103	0
	sample	384	995,067	3	995,448	940,498	54,953	0	995,451

Source: Author's calculations based on UN Comtrade.

Notes: There are a total of 68,229,722 exports flows. Blocks containing quantity-based criteria sum to 66,498,554 as there are 1,731,168 flows with missing quantity data. The **s0** definition discussed in the text is not included in the table as it assumes there are no samples, *i.e.* all flows are classified as established products

Appendix B. *Transition Matrices by Definition of 'Sample': Exports and Imports*

Each table shows the transition matrix for a given definition of what constitutes a trading sample. Data is presented row-wise, *i.e.* for exports and definition **v1** in Table B.1 below, if a product is not exported in the current year (*i.e.* is in state **o**), for the subsequent period the probability to remain not exported is 0.8552, the probability to become a sample (to transition to state **s**) is 0.0035, and the probability to become an established product (**X**) is 0.1413. Please refer to section 2.2 in the main text for details on the calculation of the probabilities.

The probability for a non-existing flow to remain non-existing (*i.e.* to go from state **o** to state **o**) as expected is constant across the definitions. Non-existing flows will never be tagged as samples or mature products in the definitions used here, and indeed no 'sensible' definition of a sample would do so. The small differences in the estimated transition probabilities are due to missing quantity data in the dataset. All value-based definitions (**v1**, **v5**, **s0**, **s10** and **s100**) have the same estimated probability for the transition from state **o** to state **o**, and so do the definitions which use quantity information (**q1**, **q5**, and **c1** and **c5**).

Table B.1. **Transition Matrices for Different Definitions of 'Trading Sample'**

World

Exports

v1	o	s	X
o	0.8552	0.0035	0.1413
s	0.6199	0.0132	0.3668
X	0.2467	0.0034	0.7499
v5	o	s	X
o	0.8552	0.0168	0.1280
s	0.5980	0.0598	0.3421
X	0.2321	0.0166	0.7513
q1	o	s	X
o	0.8582	0.0028	0.1391
s	0.6283	0.0782	0.2935
X	0.2520	0.0024	0.7456
q5	o	s	X
o	0.8582	0.0136	0.1282
s	0.5623	0.1491	0.2886
X	0.2409	0.0131	0.7460
s0	o	s	X
o	0.8552	n.a.	0.1448
s	n.a.	n.a.	n.a.
X	0.2505	n.a.	0.7495
s10	o	s	X
o	0.8552	0.0623	0.0825
s	0.5166	0.2288	0.2545
X	0.1783	0.0624	0.7594
s100	o	s	X
o	0.8552	0.1255	0.0193
s	0.3654	0.5313	0.1033
X	0.0776	0.1334	0.7890
c1	o	s	X
o	0.8582	0.0003	0.1415
s	0.6656	0.0085	0.3259
X	0.2547	0.0003	0.7451
c5	o	s	X
o	0.8582	0.0049	0.1369
s	0.6153	0.0472	0.3374
X	0.2496	0.0048	0.7455

Imports

v1	o	s	X
o	0.8572	0.0034	0.1394
s	0.6306	0.0126	0.3567
X	0.2520	0.0034	0.7446
v5	o	s	X
o	0.8572	0.0164	0.1264
s	0.6059	0.0570	0.3371
X	0.2374	0.0163	0.7463
q1	o	s	X
o	0.8598	0.0028	0.1374
s	0.6460	0.0736	0.2803
X	0.2577	0.0023	0.7400
q5	o	s	X
o	0.8598	0.0137	0.1265
s	0.5726	0.1444	0.2829
X	0.2462	0.0132	0.7406
s0	o	s	X
o	0.8572	n.a.	0.1428
s	n.a.	n.a.	n.a.
X	0.2558	n.a.	0.7442
s10	o	s	X
o	0.8572	0.0640	0.0787
s	0.5144	0.2424	0.2432
X	0.1791	0.0652	0.7557
s100	o	s	X
o	0.8572	0.1237	0.0190
s	0.3645	0.5402	0.0954
X	0.0827	0.1307	0.7866
c1	o	s	X
o	0.8598	0.0003	0.1399
s	0.6842	0.0081	0.3078
X	0.2605	0.0002	0.7393
c5	o	s	X
o	0.8598	0.0048	0.1354
s	0.6261	0.0444	0.3295
X	0.2555	0.0046	0.7399

Source: Author's calculations based on UN Comtrade.

Table B.2. **Transition Matrices for Different Definitions of 'Trading Sample'**

Bulgaria

Exports				Imports			
v1	**o**	**s**	**X**	**v1**	**o**	**s**	**X**
o	0.8828	0.0034	0.1138	**o**	0.8530	0.0039	0.1431
s	0.6747	0.0131	0.3122	**s**	0.6034	0.0108	0.3858
X	0.3502	0.0046	0.6452	**X**	0.2512	0.0041	0.7448
v5	**o**	**s**	**X**	**v5**	**o**	**s**	**X**
o	0.8828	0.0162	0.1010	**o**	0.8530	0.0190	0.1279
s	0.6496	0.0634	0.2870	**s**	0.5674	0.0602	0.3724
X	0.3304	0.0222	0.6474	**X**	0.2354	0.0207	0.7440
q1	**o**	**s**	**X**	**q1**	**o**	**s**	**X**
o	0.8830	0.0016	0.1154	**o**	0.8536	0.0023	0.1441
s	0.7115	0.0573	0.2313	**s**	0.6550	0.0527	0.2922
X	0.3529	0.0015	0.6457	**X**	0.2526	0.0019	0.7455
q5	**o**	**s**	**X**	**q5**	**o**	**s**	**X**
o	0.8830	0.0098	0.1072	**o**	0.8536	0.0145	0.1319
s	0.6454	0.1151	0.2395	**s**	0.5628	0.1230	0.3142
X	0.3419	0.0103	0.6478	**X**	0.2402	0.0149	0.7449
s0	**o**	**s**	**X**	**s0**	**o**	**s**	**X**
o	0.8828	n.a.	0.1172	**o**	0.8530	n.a.	0.1470
s	n.a.	n.a.	n.a.	**s**	n.a.	n.a.	n.a.
X	0.3553	n.a.	0.6447	**X**	0.2553	n.a.	0.7447
s10	**o**	**s**	**X**	**s10**	**o**	**s**	**X**
o	0.8828	0.0556	0.0616	**o**	0.8530	0.0742	0.0728
s	0.5777	0.2138	0.2085	**s**	0.4736	0.2640	0.2624
X	0.2608	0.0781	0.6611	**X**	0.1719	0.0829	0.7452
s100	**o**	**s**	**X**	**s100**	**o**	**s**	**X**
o	0.8828	0.1044	0.0128	**o**	0.8530	0.1338	0.0132
s	0.4449	0.4772	0.0779	**s**	0.3286	0.5789	0.0926
X	0.1251	0.1652	0.7097	**X**	0.0758	0.1766	0.7476
c1	**o**	**s**	**X**	**c1**	**o**	**s**	**X**
o	0.8830	0.0002	0.1168	**o**	0.8536	0.0003	0.1461
s	0.7391	0.0067	0.2542	**s**	0.6477	0.0052	0.3472
X	0.3549	0.0003	0.6448	**X**	0.2549	0.0002	0.7449
c5	**o**	**s**	**X**	**c5**	**o**	**s**	**X**
o	0.8830	0.0042	0.1128	**o**	0.8536	0.0057	0.1406
s	0.6723	0.0536	0.2741	**s**	0.5931	0.0415	0.3654
X	0.3492	0.0050	0.6458	**X**	0.2492	0.0059	0.7449

Source: Author's calculations based on UN Comtrade.

Table B.3. **Transition Matrices for Different Definitions of 'Trading Sample'**

Romania

Exports

v1	o	s	X
o	0.8686	0.0021	0.1293
s	0.6371	0.0111	0.3519
X	0.3232	0.0029	0.6739
v5	**o**	**s**	**X**
o	0.8686	0.0139	0.1175
s	0.6333	0.0473	0.3194
X	0.3068	0.0171	0.6762
q1	**o**	**s**	**X**
o	0.8730	0.0016	0.1254
s	0.6912	0.0679	0.2409
X	0.3223	0.0015	0.6762
q5	**o**	**s**	**X**
o	0.8730	0.0090	0.1180
s	0.6214	0.1205	0.2581
X	0.3129	0.0098	0.6773
s0	**o**	**s**	**X**
o	0.8686	n.a.	0.1314
s	n.a.	n.a.	n.a.
X	0.3260	n.a.	0.6740
s10	**o**	**s**	**X**
o	0.8686	0.0501	0.0812
s	0.5686	0.1918	0.2397
X	0.2524	0.0614	0.6861
s100	**o**	**s**	**X**
o	0.8686	0.1079	0.0235
s	0.4405	0.4537	0.1058
X	0.1321	0.1418	0.7262
c1	**o**	**s**	**X**
o	0.8730	0.0002	0.1268
s	0.7376	0.0000	0.2624
X	0.3243	0.0002	0.6755
c5	**o**	**s**	**X**
o	0.8730	0.0035	0.1235
s	0.6463	0.0393	0.3144
X	0.3199	0.0042	0.6758

Imports

v1	o	s	X
o	0.8282	0.0026	0.1693
s	0.5703	0.0094	0.4203
X	0.2107	0.0026	0.7867
v5	**o**	**s**	**X**
o	0.8282	0.0199	0.1520
s	0.5421	0.0611	0.3968
X	0.1955	0.0175	0.7870
q1	**o**	**s**	**X**
o	0.8365	0.0023	0.1612
s	0.6264	0.0772	0.2964
X	0.2141	0.0017	0.7842
q5	**o**	**s**	**X**
o	0.8365	0.0123	0.1512
s	0.5313	0.1487	0.3200
X	0.2050	0.0118	0.7832
s0	**o**	**s**	**X**
o	0.8282	n.a.	0.1718
s	n.a.	n.a.	n.a.
X	0.2130	n.a.	0.7870
s10	**o**	**s**	**X**
o	0.8282	0.0767	0.0951
s	0.4556	0.2537	0.2908
X	0.1413	0.0700	0.7888
s100	**o**	**s**	**X**
o	0.8282	0.1491	0.0228
s	0.3035	0.5809	0.1156
X	0.0575	0.1494	0.7931
c1	**o**	**s**	**X**
o	0.8365	0.0003	0.1632
s	0.6489	0.0097	0.3414
X	0.2161	0.0002	0.7837
c5	**o**	**s**	**X**
o	0.8365	0.0045	0.1590
s	0.5752	0.0412	0.3836
X	0.2121	0.0043	0.7835

Source: Author's calculations based on UN Comtrade.

Table B.4. **Transition Matrices for Different Definitions of 'Trading Sample'**

Germany

Exports

v1	o	s	X
o	0.8311	0.0018	0.1670
s	0.5531	0.0128	0.4341
X	0.1388	0.0013	0.8599
v5	**o**	**s**	**X**
o	0.8311	0.0234	0.1455
s	0.5344	0.0757	0.3899
X	0.1250	0.0135	0.8615
q1	**o**	**s**	**X**
o	0.8323	0.0011	0.1665
s	0.5417	0.0805	0.3777
X	0.1403	0.0006	0.8591
q5	**o**	**s**	**X**
o	0.8323	0.0137	0.1540
s	0.4701	0.1490	0.3810
X	0.1334	0.0091	0.8574
s0	**o**	**s**	**X**
o	0.8311	n.a.	0.1689
s	n.a.	n.a.	n.a.
X	0.1402	n.a.	0.8598
s10	**o**	**s**	**X**
o	0.8311	0.0827	0.0862
s	0.4236	0.2957	0.2806
X	0.0849	0.0483	0.8668
s100	**o**	**s**	**X**
o	0.8311	0.1488	0.0201
s	0.2512	0.6405	0.1083
X	0.0316	0.0901	0.8783
c1	**o**	**s**	**X**
o	0.8323	0.0001	0.1675
s	0.5882	0.0156	0.3962
X	0.1407	0.0001	0.8592
c5	**o**	**s**	**X**
o	0.8323	0.0062	0.1614
s	0.5342	0.0609	0.4049
X	0.1371	0.0037	0.8592

Imports

v1	o	s	X
o	0.8440	0.0012	0.1549
s	0.6096	0.0096	0.3807
X	0.1688	0.0010	0.8303
v5	**o**	**s**	**X**
o	0.8440	0.0174	0.1386
s	0.5958	0.0532	0.3511
X	0.1555	0.0115	0.8330
q1	**o**	**s**	**X**
o	0.8450	0.0009	0.1541
s	0.5523	0.0806	0.3672
X	0.1696	0.0006	0.8298
q5	**o**	**s**	**X**
o	0.8450	0.0088	0.1463
s	0.5201	0.1180	0.3619
X	0.1642	0.0066	0.8292
s0	**o**	**s**	**X**
o	0.8440	n.a.	0.1560
s	n.a.	n.a.	n.a.
X	0.1699	n.a.	0.8301
s10	**o**	**s**	**X**
o	0.8440	0.0661	0.0899
s	0.5091	0.2129	0.2779
X	0.1128	0.0449	0.8423
s100	**o**	**s**	**X**
o	0.8440	0.1350	0.0210
s	0.3343	0.5373	0.1284
X	0.0387	0.0946	0.8667
c1	**o**	**s**	**X**
o	0.8450	0.0001	0.1549
s	0.5983	0.0042	0.3975
X	0.1700	0.0001	0.8299
c5	**o**	**s**	**X**
o	0.8450	0.0036	0.1514
s	0.5920	0.0425	0.3655
X	0.1673	0.0024	0.8303

Source: Author's calculations based on UN Comtrade.

Table B.5. **Transition Matrices for Different Definitions of 'Trading Sample'**

USA

Exports				Imports			
v1	**o**	**s**	**X**	**v1**	**o**	**s**	**X**
o	0.8143	0.0048	0.1809	**o**	0.8389	0.0038	0.1573
s	0.5467	0.0178	0.4355	**s**	0.6130	0.0132	0.3738
X	0.2101	0.0042	0.7857	**X**	0.1861	0.0026	0.8113
v5	**o**	**s**	**X**	**v5**	**o**	**s**	**X**
o	0.8143	0.0227	0.1630	**o**	0.8389	0.0176	0.1435
s	0.5455	0.0599	0.3946	**s**	0.6050	0.0458	0.3493
X	0.1969	0.0188	0.7843	**X**	0.1745	0.0117	0.8138
q1	**o**	**s**	**X**	**q1**	**o**	**s**	**X**
o	0.8271	0.0039	0.1690	**o**	0.8501	0.0057	0.1442
s	0.5516	0.0552	0.3932	**s**	0.5964	0.0895	0.3141
X	0.2339	0.0036	0.7626	**X**	0.2035	0.0040	0.7925
q5	**o**	**s**	**X**	**q5**	**o**	**s**	**X**
o	0.8271	0.0181	0.1548	**o**	0.8501	0.0202	0.1298
s	0.5208	0.1187	0.3605	**s**	0.5391	0.1606	0.3004
X	0.2233	0.0165	0.7602	**X**	0.1912	0.0154	0.7934
s0	**o**	**s**	**X**	**s0**	**o**	**s**	**X**
o	0.8143	n.a.	0.1857	**o**	0.8389	n.a.	0.1611
s	n.a.	n.a.	n.a.	**s**	n.a.	n.a.	n.a.
X	0.2135	n.a.	0.7865	**X**	0.1892	n.a.	0.8108
s10	**o**	**s**	**X**	**s10**	**o**	**s**	**X**
o	0.8143	0.0799	0.1058	**o**	0.8389	0.0691	0.0920
s	0.5058	0.1912	0.3030	**s**	0.5325	0.1939	0.2736
X	0.1490	0.0630	0.7879	**X**	0.1272	0.0466	0.8262
s100	**o**	**s**	**X**	**s100**	**o**	**s**	**X**
o	0.8143	0.1640	0.0217	**o**	0.8389	0.1388	0.0223
s	0.3481	0.5194	0.1325	**s**	0.3639	0.5082	0.1279
X	0.0551	0.1434	0.8015	**X**	0.0473	0.0948	0.8579
c1	**o**	**s**	**X**	**c1**	**o**	**s**	**X**
o	0.8271	0.0005	0.1724	**o**	0.8501	0.0006	0.1494
s	0.5868	0.0115	0.4017	**s**	0.6323	0.0108	0.3568
X	0.2364	0.0004	0.7632	**X**	0.2080	0.0004	0.7916
c5	**o**	**s**	**X**	**c5**	**o**	**s**	**X**
o	0.8271	0.0069	0.1661	**o**	0.8501	0.0063	0.1436
s	0.5590	0.0459	0.3951	**s**	0.6082	0.0385	0.3533
X	0.2316	0.0062	0.7622	**X**	0.2029	0.0047	0.7923

Source: Author's calculations based on UN Comtrade.

Table B.6. **Transition Matrices for Different Definitions of 'Trading Sample'**

China

Exports				Imports			
v1	o	s	X	**v1**	o	s	X
o	0.8301	0.0032	0.1667	o	0.8432	0.0034	0.1534
s	0.5340	0.0123	0.4537	s	0.5773	0.0109	0.4119
X	0.1642	0.0022	0.8337	X	0.1858	0.0027	0.8115
v5	o	s	X	**v5**	o	s	X
o	0.8301	0.0151	0.1548	o	0.8432	0.0158	0.1409
s	0.5146	0.0469	0.4385	s	0.5574	0.0475	0.3951
X	0.1554	0.0104	0.8342	X	0.1754	0.0124	0.8122
q1	o	s	X	**q1**	o	s	X
o	0.8300	0.0004	0.1695	o	0.8435	0.0037	0.1528
s	0.5734	0.0605	0.3661	s	0.5835	0.0846	0.3319
X	0.1682	0.0002	0.8316	X	0.1870	0.0024	0.8107
q5	o	s	X	**q5**	o	s	X
o	0.8300	0.0034	0.1666	o	0.8435	0.0179	0.1386
s	0.5454	0.0835	0.3711	s	0.5005	0.1653	0.3342
X	0.1660	0.0019	0.8320	X	0.1770	0.0129	0.8101
s0	o	s	X	**s0**	o	s	X
o	0.8301	n.a.	0.1699	o	0.8432	n.a.	0.1568
s	n.a.	n.a.	n.a.	s	n.a.	n.a.	n.a.
X	0.1665	n.a.	0.8335	X	0.1887	n.a.	0.8113
s10	o	s	X	**s10**	o	s	X
o	0.8301	0.0637	0.1062	o	0.8432	0.0637	0.0931
s	0.4434	0.2007	0.3559	s	0.4761	0.2126	0.3113
X	0.1185	0.0441	0.8374	X	0.1338	0.0501	0.8161
s100	o	s	X	**s100**	o	s	X
o	0.8301	0.1473	0.0226	o	0.8432	0.1311	0.0257
s	0.2889	0.5464	0.1647	s	0.3275	0.5255	0.1470
X	0.0415	0.1009	0.8576	X	0.0603	0.1073	0.8323
c1	o	s	X	**c1**	o	s	X
o	0.8300	0.0001	0.1699	o	0.8435	0.0004	0.1561
s	0.5855	0.0024	0.4120	s	0.6742	0.0042	0.3216
X	0.1685	0.0000	0.8315	X	0.1895	0.0002	0.8103
c5	o	s	X	**c5**	o	s	X
o	0.8300	0.0014	0.1686	o	0.8435	0.0053	0.1512
s	0.5719	0.0277	0.4004	s	0.5702	0.0358	0.3940
X	0.1674	0.0008	0.8317	X	0.1857	0.0038	0.8105

Source: Author's calculations based on UN Comtrade.

Appendix C. ***Comparing the Three Definitions of Discovery***

Table C.1. **Flows Tagged/Not Tagged as a Discovery, by Trading Sample Definition**

CS Definition of Discovery

Exports

sample	discovery	number	percent
v1	not discovery	65,543,498	96.06
	discovery	2,686,224	3.94
v5	not discovery	65,543,498	96.06
	discovery	2,686,224	3.94
q1	not discovery	65,443,304	95.92
	discovery	2,786,418	4.08
q5	not discovery	65,443,304	95.92
	discovery	2,786,418	4.08
s0	not discovery	65,543,498	96.06
	discovery	2,686,224	3.94
s10	not discovery	65,543,498	96.06
	discovery	2,686,224	3.94
s100	not discovery	65,543,498	96.06
	discovery	2,686,224	3.94
c1	not discovery	65,443,304	95.92
	discovery	2,786,418	4.08
c5	not discovery	65,443,304	95.92
	discovery	2,786,418	4.08

Imports

sample	discovery	number	percent
v1	not discovery	66,232,284	95.71
	discovery	2,971,501	4.29
v5	not discovery	66,232,284	95.71
	discovery	2,971,501	4.29
q1	not discovery	66,154,061	95.59
	discovery	3,049,724	4.41
q5	not discovery	66,154,061	95.59
	discovery	3,049,724	4.41
s0	not discovery	66,232,284	95.71
	discovery	2,971,501	4.29
s10	not discovery	66,232,284	95.71
	discovery	2,971,501	4.29
s100	not discovery	66,232,284	95.71
	discovery	2,971,501	4.29
c1	not discovery	66,154,061	95.59
	discovery	3,049,724	4.41
c5	not discovery	66,154,061	95.59
	discovery	3,049,724	4.41

Source: Author's calculations based on UN Comtrade.
Note: Since the CS definition does not use samples, the numbers of discoveries found is the same for all definitions of trading sample. The differences in the values in the table are due to flows with missing quantity data.

Table C.2. **Flows Tagged/Not Tagged as a Discovery, by Trading Sample Definition**

KL Definition of Discovery

Exports				Imports			
sample	discovery	number	percent	sample	discovery	number	percent
v1	not discovery	67,752,924	99.30	v1	not discovery	68,687,736	99.25
	discovery	476,798	0.70		discovery	516,049	0.75
v5	not discovery	67,770,503	99.33	v5	not discovery	68,704,926	99.28
	discovery	459,219	0.67		discovery	498,859	0.72
q1	not discovery	67,623,669	99.11	q1	not discovery	68,571,053	99.09
	discovery	606,053	0.89		discovery	632,732	0.91
q5	not discovery	67,645,188	99.14	q5	not discovery	68,593,102	99.12
	discovery	584,534	0.86		discovery	610,683	0.88
s0	not discovery	67,748,713	99.30	s0	not discovery	68,683,765	99.25
	discovery	481,009	0.70		discovery	520,020	0.75
s10	not discovery	67,852,775	99.45	s10	not discovery	68,795,734	99.41
	discovery	376,947	0.55		discovery	408,051	0.59
s100	not discovery	68,067,194	99.76	s100	not discovery	69,025,303	99.74
	discovery	162,528	0.24		discovery	178,482	0.26
c1	not discovery	67,619,780	99.11	c1	not discovery	68,566,901	99.08
	discovery	609,942	0.89		discovery	636,884	0.92
c5	not discovery	67,627,039	99.12	c5	not discovery	68,574,044	99.09
	discovery	602,683	0.88		discovery	629,741	0.91

Source: Author's calculations based on UN Comtrade.

Table C.3. **Flows Tagged/Not Tagged as a Discovery, by Trading Sample Definition**

PS Definition of Discovery

Exports				Imports			
sample	discovery	number	percent	sample	discovery	number	percent
v1	not discovery	64,792,383	94.96	v1	not discovery	65,377,441	94.47
	discovery	3,437,339	5.04		discovery	3,826,344	5.53
v5	not discovery	65,060,531	95.36	v5	not discovery	65,662,415	94.88
	discovery	3,169,191	4.64		discovery	3,541,370	5.12
q1	not discovery	64,651,087	94.76	q1	not discovery	65,276,277	94.32
	discovery	3,578,635	5.24		discovery	3,927,508	5.68
q5	not discovery	64,880,362	95.09	q5	not discovery	65,526,228	94.69
	discovery	3,349,360	4.91		discovery	3,677,557	5.31
s0	not discovery	64,722,760	94.86	s0	not discovery	65,301,769	94.36
	discovery	3,506,962	5.14		discovery	3,902,016	5.64
s10	not discovery	65,968,824	96.69	s10	not discovery	66,720,920	96.41
	discovery	2,260,898	3.31		discovery	2,482,865	3.59
s100	not discovery	67,455,599	98.87	s100	not discovery	68,350,002	98.77
	discovery	774,123	1.13		discovery	853,783	1.23
c1	not discovery	64,601,891	94.68	c1	not discovery	65,221,910	94.25
	discovery	3,627,831	5.32		discovery	3,981,875	5.75
c5	not discovery	64,697,183	94.82	c5	not discovery	65,322,100	94.39
	discovery	3,532,539	5.18		discovery	3,881,685	5.61

Source: Author's calculations based on UN Comtrade.

The following three tables provide a cross-tabulation of flows according to whether they have or have not been tagged as a discovery by the respective pair of discovery definitions. For example, Table C.4 compares how the CS and KL definitions stack against each other. For sample definition v1, 65,268,668 flows have been tagged as 'not discoveries' by both the CS and KL definitions. There are 274,830 flows that have been tagged as a discovery by KL, and tagged as 'not discovery' by CS; 2,484,256 flows have been tagged as discoveries by CS and not by KL, and only 201,968 flows have been tagged as a discovery by both definitions.

Table C.4. **Crosstab of Discoveries by Different Discovery Definitions**

			CS vs. KL			
			Exports		Imports	
sample		kl =	not discovery	discovery	not discovery	discovery
v1	cs =	not discovery	65,268,668	274,830	65,957,661	274,623
	cs =	discovery	2,484,256	201,968	2,730,075	241,426
v5	cs =	not discovery	65,283,770	259,728	65,972,373	259,911
	cs =	discovery	2,486,733	199,491	2,732,553	238,948
q1	cs =	not discovery	65,052,954	390,350	65,760,223	393,838
	cs =	discovery	2,570,715	215,703	2,810,830	238,894
q5	cs =	not discovery	65,070,790	372,514	65,778,260	375,801
	cs =	discovery	2,574,398	212,020	2,814,842	234,882
s0	cs =	not discovery	65,264,986	278,512	65,954,302	277,982
	cs =	discovery	2,483,727	202,497	2,729,463	242,038
s10	cs =	not discovery	65,350,402	193,096	66,044,214	188,070
	cs =	discovery	2,502,373	183,851	2,751,520	219,981
s100	cs =	not discovery	65,482,117	61,381	66,174,878	57,406
	cs =	discovery	2,585,077	101,147	2,850,425	121,076
c1	cs =	not discovery	65,049,648	393,656	65,756,735	397,326
	cs =	discovery	2,570,132	216,286	2,810,166	239,558
c5	cs =	not discovery	65,055,742	387,562	65,762,789	391,272
	cs =	discovery	2,571,297	215,121	2,811,255	238,469

Source: Author's calculations based on UN Comtrade.

Table C.5. **Crosstab of Discoveries by Different Discovery Definitions**

			CS vs. PS			
			Exports		Imports	
sample		**ps =**	**not discovery**	**discovery**	**not discovery**	**discovery**
v1	cs =	not discovery	62,145,774	3,397,724	62,448,796	3,783,488
	cs =	discovery	2,646,609	39,615	2,928,645	42,856
v5	cs =	not discovery	62,551,380	2,992,118	62,884,134	3,348,150
	cs =	discovery	2,509,151	177,073	2,778,281	193,220
q1	cs =	not discovery	61,884,754	3,558,550	62,249,118	3,904,943
	cs =	discovery	2,766,333	20,085	3,027,159	22,565
q5	cs =	not discovery	62,196,851	3,246,453	62,591,849	3,562,212
	cs =	discovery	2,683,511	102,907	2,934,379	115,345
s0	cs =	not discovery	62,036,536	3,506,962	62,330,268	3,902,016
	cs =	discovery	2,686,224	0	2,971,501	0
s10	cs =	not discovery	63,748,959	1,794,539	64,274,013	1,958,271
	cs =	discovery	2,219,865	466,359	2,446,907	524,594
s100	cs =	not discovery	65,126,200	417,298	65,760,728	471,556
	cs =	discovery	2,329,399	356,825	2,589,274	382,227
c1	cs =	not discovery	61,818,016	3,625,288	62,174,902	3,979,159
	cs =	discovery	2,783,875	2,543	3,047,008	2,716
c5	cs =	not discovery	61,958,714	3,484,590	62,323,705	3,830,356
	cs =	discovery	2,738,469	47,949	2,998,395	51,329

Source: Author's calculations based on UN Comtrade.
Note: The table provides a cross-tabulation of flows according to whether they have or have not been tagged as a discovery by the respective pair of discovery definitions.

Table C.6. **Crosstab of Discoveries by Different Discovery Definitions**

			KL vs. PS			
			Exports		Imports	
sample		ps =	not discovery	discovery	not discovery	discovery
v1	kl =	not discovery	64,319,185	3,433,739	64,865,276	3,822,460
	kl =	discovery	473,198	3,600	512,165	3,884
v5	kl =	not discovery	64,618,478	3,152,025	65,181,910	3,523,016
	kl =	discovery	442,053	17,166	480,505	18,354
q1	kl =	not discovery	64,046,625	3,577,044	64,645,365	3,925,688
	kl =	discovery	604,462	1,591	630,912	1,820
q5	kl =	not discovery	64,305,521	3,339,667	64,926,293	3,666,809
	kl =	discovery	574,841	9,693	599,935	10,748
s0	kl =	not discovery	64,241,751	3,506,962	64,781,749	3,902,016
	kl =	discovery	481,009	0	520,020	0
s10	kl =	not discovery	65,648,250	2,204,525	66,376,117	2,419,617
	kl =	discovery	320,574	56,373	344,803	63,248
s100	kl =	not discovery	67,354,507	712,687	68,241,214	784,089
	kl =	discovery	101,092	61,436	108,788	69,694
c1	kl =	not discovery	63,992,107	3,627,673	64,585,211	3,981,690
	kl =	discovery	609,784	158	636,699	185
c5	kl =	not discovery	64,098,395	3,528,644	64,696,595	3,877,449
	kl =	discovery	598,788	3,895	625,505	4,236

Source: Author's calculations based on UN Comtrade.
Note: The table provides a cross-tabulation of flows according to whether they have or have not been tagged as a discovery by the respective pair of discovery definitions.

Table C.7. **Exports: Crosstab of Discoveries by Different Discovery Definitions, Ignoring Time**

CS vs. KL

v1	kl = 0	kl = 1
cs = 0	8,975,857	155,615
cs = 1	2,179,655	301,328
cs = 2	82,309	19,710
cs = 3	256	145
v5	**kl = 0**	**kl = 1**
cs = 0	8,986,274	145,198
cs = 1	2,185,636	295,347
cs = 2	83,477	18,542
cs = 3	269	132
q1	**kl = 0**	**kl = 1**
cs = 0	8,864,635	169,872
cs = 1	2,161,421	413,266
cs = 2	82,548	22,764
cs = 3	218	151
q5	**kl = 0**	**kl = 1**
cs = 0	8,875,857	158,650
cs = 1	2,170,514	404,173
cs = 2	83,740	21,572
cs = 3	230	139
s0	**kl = 0**	**kl = 1**
cs = 0	8,973,155	158,317
cs = 1	2,178,464	302,519
cs = 2	82,001	20,018
cs = 3	246	155
s10	**kl = 0**	**kl = 1**
cs = 0	9,025,544	105,928
cs = 1	2,223,506	257,477
cs = 2	88,558	13,461
cs = 3	320	81
s100	**kl = 0**	**kl = 1**
cs = 0	9,097,492	33,980
cs = 1	2,356,436	124,547
cs = 2	98,036	3,983
cs = 3	383	18
c1	**kl = 0**	**kl = 1**
cs = 0	8,862,388	172,119
cs = 1	2,159,989	414,698
cs = 2	82,341	22,971
cs = 3	215	154
c5	**kl = 0**	**kl = 1**
cs = 0	8,866,631	167,876
cs = 1	2,162,568	412,119
cs = 2	82,771	22,541
cs = 3	222	147

CS vs. PS

v1	ps = 0	ps = 1	ps = 2	ps = 3
cs = 0	7,859,618	1,231,091	40,507	256
cs = 1	536,108	1,872,326	72,240	309
cs = 2	21,451	54,731	25,789	48
cs = 3	161	200	40	0
v5	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	8,045,730	1,056,138	29,448	156
cs = 1	589,266	1,832,285	59,220	212
cs = 2	23,723	54,561	23,700	35
cs = 3	173	194	34	0
q1	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	7,742,281	1,249,714	42,215	297
cs = 1	520,347	1,978,199	75,842	299
cs = 2	20,199	57,720	27,345	48
cs = 3	144	184	41	0
q5	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	7,888,249	1,111,973	34,068	217
cs = 1	581,079	1,927,360	66,017	231
cs = 2	22,468	57,466	25,339	39
cs = 3	154	178	37	0
s0	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	7,807,720	1,279,540	43,918	294
cs = 1	526,218	1,878,725	75,710	330
cs = 2	21,026	54,755	26,185	53
cs = 3	158	201	42	0
s10	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	8,544,180	576,820	10,433	39
cs = 1	922,131	1,531,740	27,049	63
cs = 2	39,306	48,555	14,146	12
cs = 3	235	147	19	0
s100	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	9,024,658	105,837	974	3
cs = 1	1,842,043	635,412	3,522	6
cs = 2	80,579	19,072	2,367	1
cs = 3	356	44	1	0
c1	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	7,707,843	1,281,934	44,411	319
cs = 1	510,685	1,985,678	78,001	323
cs = 2	19,879	57,719	27,663	51
cs = 3	141	185	43	0
c5	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	7,776,361	1,217,803	40,062	281
cs = 1	527,152	1,973,775	73,474	286
cs = 2	20,490	57,769	27,008	45
cs = 3	143	184	42	0

PS vs. KL

v1	kl = 0	kl = 1
ps = 0	8,301,417	115,921
ps = 1	2,827,489	330,859
ps = 2	109,101	29,475
ps = 3	70	543
v5	**kl = 0**	**kl = 1**
ps = 0	8,537,439	121,453
ps = 1	2,628,573	314,605
ps = 2	89,592	22,810
ps = 3	52	351
q1	**kl = 0**	**kl = 1**
ps = 0	8,160,483	122,488
ps = 1	2,835,081	450,736
ps = 2	113,195	32,248
ps = 3	63	581
q5	**kl = 0**	**kl = 1**
ps = 0	8,369,056	122,894
ps = 1	2,663,219	433,758
ps = 2	98,021	27,440
ps = 3	45	442
s0	**kl = 0**	**kl = 1**
ps = 0	8,240,635	114,487
ps = 1	2,878,653	334,568
ps = 2	114,509	31,346
ps = 3	69	608
s10	**kl = 0**	**kl = 1**
ps = 0	9,384,692	121,160
ps = 1	1,910,702	246,560
ps = 2	42,522	9,125
ps = 3	12	102
s100	**kl = 0**	**kl = 1**
ps = 0	10,897,449	50,187
ps = 1	649,087	111,278
ps = 2	5,810	1,054
ps = 3	1	9
c1	**kl = 0**	**kl = 1**
ps = 0	8,116,239	122,309
ps = 1	2,871,819	453,697
ps = 2	116,807	33,311
ps = 3	68	625
c5	**kl = 0**	**kl = 1**
ps = 0	8,200,717	123,429
ps = 1	2,801,767	447,764
ps = 2	109,650	30,936
ps = 3	58	554

Source: Author's calculations based on UN Comtrade.

Note: The row and column names show how many times a relationship (reporter-partner-commodity) has been tagged as a discovery by the respective definition in any of the 17 years in the sample. For sample definition **v1**, 8,975,857 relationships have never been tagged as discoveries by either CS or KL; 155,615 relationships have been tagged once by KL and never by CS, *etc.*

Table C.8. **Imports: Crosstab of Discoveries by Different Discovery Definitions, Ignoring Time**

CS vs. KL

v1	kl = 0	kl = 1
cs = 0	9,350,148	158,733
cs = 1	2,414,930	335,951
cs = 2	88,492	21,218
cs = 3	253	147
v5	**kl = 0**	**kl = 1**
cs = 0	9,360,489	148,392
cs = 1	2,420,508	330,373
cs = 2	89,744	19,966
cs = 3	272	128
q1	**kl = 0**	**kl = 1**
cs = 0	9,260,769	170,179
cs = 1	2,388,947	439,556
cs = 2	87,186	22,856
cs = 3	238	141
q5	**kl = 0**	**kl = 1**
cs = 0	9,272,633	158,315
cs = 1	2,397,787	430,716
cs = 2	88,520	21,522
cs = 3	249	130
s0	**kl = 0**	**kl = 1**
cs = 0	9,347,650	161,231
cs = 1	2,413,784	337,097
cs = 2	88,170	21,540
cs = 3	248	152
s10	**kl = 0**	**kl = 1**
cs = 0	9,403,746	105,135
cs = 1	2,462,042	288,839
cs = 2	95,706	14,004
cs = 3	327	73
s100	**kl = 0**	**kl = 1**
cs = 0	9,476,056	32,825
cs = 1	2,609,334	141,547
cs = 2	105,616	4,094
cs = 3	384	16
c1	**kl = 0**	**kl = 1**
cs = 0	9,258,256	172,692
cs = 1	2,387,516	440,987
cs = 2	86,984	23,058
cs = 3	232	147
c5	**kl = 0**	**kl = 1**
cs = 0	9,262,580	168,368
cs = 1	2,389,897	438,606
cs = 2	87,417	22,625
cs = 3	237	142

CS vs. PS

v1	ps = 0	ps = 1	ps = 2	ps = 3
cs = 0	8,136,532	1,328,275	43,732	342
cs = 1	541,390	2,126,233	82,957	301
cs = 2	20,615	61,977	27,073	45
cs = 3	166	197	37	0
v5	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	8,336,229	1,140,330	32,095	227
cs = 1	596,916	2,084,004	69,745	216
cs = 2	22,969	61,741	24,966	34
cs = 3	180	188	32	0
q1	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	8,051,081	1,335,230	44,292	345
cs = 1	528,650	2,216,006	83,542	305
cs = 2	19,538	62,659	27,798	47
cs = 3	157	186	36	0
q5	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	8,212,391	1,182,904	35,399	254
cs = 1	593,294	2,161,897	73,081	231
cs = 2	21,815	62,392	25,791	44
cs = 3	176	171	32	0
s0	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	8,079,955	1,381,092	47,452	382
cs = 1	530,764	2,133,067	86,726	324
cs = 2	20,201	61,961	27,496	52
cs = 3	163	200	37	0
s10	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	8,907,804	590,193	10,827	57
cs = 1	996,117	1,721,714	32,994	56
cs = 2	41,377	53,866	14,456	11
cs = 3	248	138	14	0
s100	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	9,399,231	108,589	1,056	5
cs = 1	2,039,910	705,313	5,655	3
cs = 2	85,702	21,624	2,384	0
cs = 3	359	39	2	0
c1	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	8,012,299	1,371,540	46,735	374
cs = 1	518,485	2,223,883	85,815	320
cs = 2	19,272	62,578	28,143	49
cs = 3	156	187	36	0
c5	**ps = 0**	**ps = 1**	**ps = 2**	**ps = 3**
cs = 0	8,085,327	1,303,071	42,221	329
cs = 1	535,250	2,211,669	81,298	286
cs = 2	19,861	62,599	27,535	47
cs = 3	163	180	36	0

PS vs. KL

v1	kl = 0	kl = 1
ps = 0	8,582,108	116,595
ps = 1	3,148,367	368,315
ps = 2	123,265	30,534
ps = 3	83	605
v5	**kl = 0**	**kl = 1**
ps = 0	8,834,010	122,284
ps = 1	2,934,019	352,244
ps = 2	102,917	23,921
ps = 3	67	410
q1	**kl = 0**	**kl = 1**
ps = 0	8,477,475	121,951
ps = 1	3,137,434	476,647
ps = 2	122,150	33,518
ps = 3	81	616
q5	**kl = 0**	**kl = 1**
ps = 0	8,705,309	122,367
ps = 1	2,948,133	459,231
ps = 2	105,681	28,622
ps = 3	66	463
s0	**kl = 0**	**kl = 1**
ps = 0	8,516,051	115,032
ps = 1	3,204,620	371,700
ps = 2	129,088	32,623
ps = 3	93	665
s10	**kl = 0**	**kl = 1**
ps = 0	9,823,619	121,927
ps = 1	2,089,131	276,780
ps = 2	49,054	9,237
ps = 3	17	107
s100	**kl = 0**	**kl = 1**
ps = 0	11,477,102	48,100
ps = 1	706,351	129,214
ps = 2	7,937	1,160
ps = 3	0	8
c1	**kl = 0**	**kl = 1**
ps = 0	8,428,326	121,886
ps = 1	3,178,489	479,699
ps = 2	126,087	34,642
ps = 3	86	657
c5	**kl = 0**	**kl = 1**
ps = 0	8,517,758	122,843
ps = 1	3,103,588	473,931
ps = 2	118,705	32,385
ps = 3	80	582

Source: Author's calculations based on UN Comtrade.
Note: The row and column names show how many times a relationship (reporter-partner-commodity) has been tagged as a discovery by the respective definition in any of the 17 years in the sample.

Appendix D. ***Length of the Samples Phase in the PS Definition of Discovery***

Table D.1 and Table D.2 show the share of flows tagged as a discovery by the PS definition, in any year, that had a samples phase of the respective length. At the world level, for exports and the **v1** definition of trading sample, 0.98848 of all discoveries had a zero-length window 2, *i.e.* moved from samples to established-product trade in the subsequent year, 0.00856 had a samples phase of one year, *etc.* Sample definition **s0** does not distinguish samples from established-product flows, hence all discoveries have a zero-length sample phase.

Table D.1. **Distribution of the Length of the Samples Phase of Exports Discoveries, by Sample Definition**

	length	v1	v5	q1	q5	s0	s10	s100	c1	c5
World	0	0.98848	0.94413	0.99439	0.96928	1.00000	0.79373	0.53906	0.99930	0.98643
	1	0.00856	0.03912	0.00394	0.01998	0.00000	0.11782	0.19234	0.00054	0.00971
	2	0.00286	0.01416	0.00140	0.00805	0.00000	0.05348	0.11027	0.00016	0.00341
	3	0.00009	0.00198	0.00019	0.00177	0.00000	0.02093	0.07283	0.00000	0.00034
	4	0.00002	0.00052	0.00007	0.00069	0.00000	0.00973	0.05065	0.00000	0.00010
	5	0.00000	0.00010	0.00002	0.00025	0.00000	0.00431	0.03486	0.00000	0.00002
	length	v1	v5	q1	q5	s0	s10	s100	c1	c5
Bulgaria	0	0.98799	0.94113	0.99626	0.97260	1.00000	0.78588	0.50173	0.99940	0.98640
	1	0.00881	0.04299	0.00291	0.01961	0.00000	0.13371	0.23564	0.00051	0.01025
	2	0.00303	0.01366	0.00077	0.00572	0.00000	0.05000	0.11732	0.00009	0.00296
	3	0.00016	0.00186	0.00003	0.00163	0.00000	0.01879	0.07100	0.00000	0.00036
	4	0.00000	0.00028	0.00003	0.00041	0.00000	0.00852	0.04473	0.00000	0.00003
	5	0.00000	0.00007	0.00000	0.00003	0.00000	0.00309	0.02958	0.00000	0.00000
	length	v1	v5	q1	q5	s0	s10	s100	c1	c5
Romania	0	0.99160	0.94697	0.99627	0.97465	1.00000	0.81555	0.54967	0.99938	0.98748
	1	0.00622	0.03940	0.00257	0.01780	0.00000	0.11565	0.21838	0.00041	0.00956
	2	0.00209	0.01155	0.00107	0.00579	0.00000	0.04377	0.10632	0.00021	0.00251
	3	0.00009	0.00166	0.00009	0.00123	0.00000	0.01556	0.06135	0.00000	0.00036
	4	0.00000	0.00042	0.00000	0.00044	0.00000	0.00613	0.03673	0.00000	0.00009
	5	0.00000	0.00000	0.00000	0.00009	0.00000	0.00334	0.02755	0.00000	0.00000
	length	v1	v5	q1	q5	s0	s10	s100	c1	c5
Germany	0	0.99383	0.92114	0.99793	0.96343	1.00000	0.74658	0.60677	0.99973	0.98086
	1	0.00419	0.05343	0.00153	0.02294	0.00000	0.13220	0.15302	0.00021	0.01325
	2	0.00198	0.02099	0.00048	0.01058	0.00000	0.06881	0.09938	0.00007	0.00511
	3	0.00000	0.00331	0.00004	0.00209	0.00000	0.03000	0.05783	0.00000	0.00058
	4	0.00000	0.00098	0.00001	0.00078	0.00000	0.01519	0.04779	0.00000	0.00020
	5	0.00000	0.00014	0.00001	0.00018	0.00000	0.00722	0.03522	0.00000	0.00000
	length	v1	v5	q1	q5	s0	s10	s100	c1	c5
USA	0	0.98631	0.93818	0.99260	0.96500	1.00000	0.79041	0.54695	0.99904	0.98469
	1	0.00954	0.04138	0.00512	0.02228	0.00000	0.12075	0.18910	0.00074	0.01038
	2	0.00391	0.01704	0.00203	0.00966	0.00000	0.05613	0.10969	0.00022	0.00434
	3	0.00021	0.00267	0.00020	0.00196	0.00000	0.01986	0.07201	0.00001	0.00043
	4	0.00004	0.00066	0.00004	0.00080	0.00000	0.00927	0.04992	0.00000	0.00015
	5	0.00000	0.00007	0.00001	0.00031	0.00000	0.00357	0.03233	0.00000	0.00001
	length	v1	v5	q1	q5	s0	s10	s100	c1	c5
China	0	0.98737	0.93902	0.99878	0.99033	1.00000	0.74021	0.32223	0.99985	0.99591
	1	0.00927	0.04218	0.00088	0.00675	0.00000	0.14372	0.21503	0.00012	0.00310
	2	0.00324	0.01648	0.00031	0.00248	0.00000	0.07323	0.16421	0.00004	0.00094
	3	0.00010	0.00185	0.00002	0.00034	0.00000	0.02681	0.13035	0.00000	0.00004
	4	0.00002	0.00040	0.00001	0.00008	0.00000	0.01166	0.09798	0.00000	0.00002
	5	0.00000	0.00007	0.00000	0.00003	0.00000	0.00437	0.07019	0.00000	0.00000

Source: Author's calculations based on UN Comtrade.

Table D.2. **Distribution of the Length of the Samples Phase of Imports Discoveries, by Sample Definition**

	length	v1	v5	q1	q5	s0	s10	s100	c1	c5
World	**0**	0.98880	0.94544	0.99425	0.96864	1.00000	0.78871	0.55231	0.99932	0.98678
	1	0.00827	0.03805	0.00403	0.02045	0.00000	0.11702	0.17814	0.00052	0.00949
	2	0.00283	0.01406	0.00147	0.00825	0.00000	0.05570	0.10889	0.00016	0.00332
	3	0.00008	0.00187	0.00017	0.00176	0.00000	0.02263	0.07120	0.00000	0.00032
	4	0.00002	0.00048	0.00005	0.00066	0.00000	0.01086	0.05158	0.00000	0.00008
	5	0.00000	0.00010	0.00002	0.00025	0.00000	0.00507	0.03788	0.00000	0.00001
	length	**v1**	**v5**	**q1**	**q5**	**s0**	**s10**	**s100**	**c1**	**c5**
Bulgaria	**0**	0.98452	0.92268	0.99460	0.95756	1.00000	0.69116	0.40426	0.99912	0.97930
	1	0.01128	0.05434	0.00407	0.02773	0.00000	0.16313	0.21308	0.00067	0.01484
	2	0.00401	0.01954	0.00112	0.01132	0.00000	0.08243	0.13332	0.00021	0.00527
	3	0.00016	0.00275	0.00021	0.00244	0.00000	0.03778	0.10210	0.00000	0.00053
	4	0.00003	0.00060	0.00000	0.00084	0.00000	0.01716	0.08612	0.00000	0.00005
	5	0.00000	0.00009	0.00000	0.00011	0.00000	0.00835	0.06111	0.00000	0.00000
	length	**v1**	**v5**	**q1**	**q5**	**s0**	**s10**	**s100**	**c1**	**c5**
Romania	**0**	0.99049	0.92370	0.99490	0.96512	1.00000	0.69969	0.39554	0.99933	0.98511
	1	0.00704	0.05431	0.00359	0.02315	0.00000	0.16293	0.21806	0.00055	0.01094
	2	0.00242	0.01898	0.00125	0.00917	0.00000	0.07973	0.14206	0.00013	0.00364
	3	0.00005	0.00235	0.00019	0.00176	0.00000	0.03428	0.10644	0.00000	0.00028
	4	0.00000	0.00057	0.00006	0.00054	0.00000	0.01636	0.08067	0.00000	0.00002
	5	0.00000	0.00010	0.00000	0.00025	0.00000	0.00700	0.05723	0.00000	0.00000
	length	**v1**	**v5**	**q1**	**q5**	**s0**	**s10**	**s100**	**c1**	**c5**
Germany	**0**	0.99713	0.94730	0.99849	0.97860	1.00000	0.80528	0.54488	0.99984	0.99016
	1	0.00205	0.03685	0.00101	0.01471	0.00000	0.11844	0.21124	0.00014	0.00726
	2	0.00082	0.01342	0.00047	0.00493	0.00000	0.04775	0.10539	0.00002	0.00224
	3	0.00000	0.00184	0.00004	0.00123	0.00000	0.01830	0.06807	0.00000	0.00031
	4	0.00000	0.00051	0.00000	0.00030	0.00000	0.00708	0.04458	0.00000	0.00002
	5	0.00000	0.00008	0.00000	0.00023	0.00000	0.00315	0.02584	0.00000	0.00000
	length	**v1**	**v5**	**q1**	**q5**	**s0**	**s10**	**s100**	**c1**	**c5**
USA	**0**	0.98965	0.95361	0.99044	0.96631	1.00000	0.81591	0.55013	0.99897	0.98699
	1	0.00758	0.03348	0.00669	0.02212	0.00000	0.11264	0.21699	0.00072	0.00955
	2	0.00271	0.01111	0.00231	0.00826	0.00000	0.04535	0.10434	0.00031	0.00301
	3	0.00005	0.00145	0.00039	0.00214	0.00000	0.01680	0.06278	0.00000	0.00029
	4	0.00000	0.00033	0.00013	0.00090	0.00000	0.00662	0.04090	0.00000	0.00016
	5	0.00000	0.00002	0.00004	0.00028	0.00000	0.00268	0.02486	0.00000	0.00000
	length	**v1**	**v5**	**q1**	**q5**	**s0**	**s10**	**s100**	**c1**	**c5**
China	**0**	0.98847	0.94425	0.99112	0.95025	1.00000	0.76095	0.47989	0.99901	0.98230
	1	0.00831	0.03929	0.00610	0.03076	0.00000	0.13795	0.20178	0.00074	0.01282
	2	0.00316	0.01424	0.00226	0.01358	0.00000	0.06116	0.11683	0.00025	0.00438
	3	0.00006	0.00173	0.00036	0.00369	0.00000	0.02521	0.08827	0.00000	0.00042
	4	0.00000	0.00047	0.00015	0.00129	0.00000	0.01031	0.06453	0.00000	0.00008
	5	0.00000	0.00002	0.00002	0.00043	0.00000	0.00442	0.04870	0.00000	0.00000

Source: Author's calculations based on UN Comtrade.

DISCUSSION PAPERS

DP/1/1998 The First Year of the Currency Board in Bulgaria
Victor Yotzov, Nikolay Nenovsky, Kalin Hristov, Iva Petrova, Boris Petrov

DP/2/1998 Financial Repression and Credit Rationing under Currency Board Arrangement for Bulgaria
Nikolay Nenovsky, Kalin Hristov

DP/3/1999 Investment Incentives in Bulgaria: Assessment of the Net Tax Effect on the State Budget
Dobrislav Dobrev, Boyko Tzenov, Peter Dobrev, John Ayerst

DP/4/1999 Two Approaches to Fixed Exchange Rate Crises
Nikolay Nenovsky, Kalin Hristov, Boris Petrov

DP/5/1999 Monetary Sector Modeling in Bulgaria, 1913–1945
Nikolay Nenovsky, Boris Petrov

DP/6/1999 The Role of a Currency Board in Financial Crises: The Case of Bulgaria
Roumen Avramov

DP/7/1999 The Bulgarian Financial Crisis of 1996–1997
Zdravko Balyozov

DP/8/1999 The Economic Philosophy of Friedrich Hayek (The Centenary of his Birth)
Nikolay Nenovsky

DP/9/1999 The Currency Board in Bulgaria: Design, Peculiarities and Management of Foreign Exchange Cover
Dobrislav Dobrev

DP/10/1999 Monetary Regimes and the Real Economy (Empirical Tests before and after the Introduction of the Currency Board in Bulgaria)
Nikolay Nenovsky, Kalin Hristov

DP/11/1999 The Currency Board in Bulgaria: The First Two Years
Jeffrey B. Miller

DP/12/2000 Fundamentals in Bulgarian Brady Bonds: Price Dynamics
Nina Budina, Tzvetan Manchev

DP/13/2000 Currency Circulation after Currency Board Introduction in Bulgaria (Transactions Demand, Hoarding, Shadow Economy)
Nikolay Nenovsky, Kalin Hristov

DP/14/2000 Macroeconomic Models of the International Monetary Fund and the World Bank (Analysis of Theoretical Approaches and Evaluation of Their Effective Implementation in Bulgaria)
Victor Yotzov

DP/15/2000 Bank Reserve Dynamics under Currency Board Arrangement for Bulgaria
Boris Petrov

DP/16/2000 A Possible Approach to Simulate Macroeconomic Development of Bulgaria
Victor Yotzov

DP/17/2001 Banking Supervision on Consolidated Basis *(in Bulgarian only)*
Margarita Prandzheva

DP/18/2001 Real Wage Rigidity and the Monetary Regime Choice
Nikolay Nenovsky, Darina Koleva

DP/19/2001 The Financial System in the Bulgarian Economy
Jeffrey Miller, Stefan Petranov

DP/20/2002 Forecasting Inflation via Electronic Markets Results from a Prototype Experiment
Michael Berlemann

DP/21/2002 Corporate Image of Commercial Banks (1996–1997) *(in Bulgarian only)*
Miroslav Nedelchev

DP/22/2002 Fundamental Equilibrium Exchange Rates and Currency Boards: Evidence from Argentina and Estonia in the 90's
Kalin Hristov

DP/23/2002 Credit Activity of Commercial Banks and Rationing in the Credit Market in Bulgaria *(in Bulgarian only)*
Kalin Hristov, Mihail Mihailov

DP/24/2002 Balassa – Samuelson Effect in Bulgaria *(in Bulgarian only)*
Georgi Choukalev

DP/25/2002 Money and Monetary Obligations: Nature, Stipulation, Fulfilment
Stanislav Natzev, Nachko Staykov, Filko Rosov

DP/26/2002 Regarding the Unilateral Euroization of Bulgaria
Ivan Kostov, Jana Kostova

DP/27/2002 Shadowing the Euro: Bulgaria's Monetary Policy Five Years on
Martin Zaimov, Kalin Hristov

DP/28/2002 Improving Monetary Theory in Post-communist Countries – Looking Back to Cantillon
Nikolay Nenovsky

DP/29/2003 Dual Inflation under the Currency Board: The Challenges of Bulgarian EU Accession *(in Bulgarian only)*
Nikolay Nenovsky, Kalina Dimitrova

DP/30/2003 Exchange Rate Arrangements, Economic Policy and Inflation: Empirical Evidence for Latin America
Andreas Freytag

DP/31/2003 Inflation and the Bulgarian Currency Board
Stacie Beck, Jeffrey B. Miller, Mohsen Saad

DP/32/2003 Banks – Firms Nexus under the Currency Board: Empirical Evidence from Bulgaria
Nikolay Nenovsky, Evgeni Peev, Todor Yalamov

DP/33/2003 Modelling Inflation in Bulgaria: Markup Model *(in Bulgarian only)*
Kalin Hristov, Mihail Mihailov

DP/34/2003 Competitiveness of the Bulgarian Economy
Konstantin Pashev

DP/35/2003 Exploring the Currency Board Mechanics: a Basic Formal Model
Jean-Baptiste Desquilbet, Nikolay Nenovsky

DP/36/2003 A Composite Tendency Indicator for Bulgaria's Industry *(in Bulgarian only)*
Tsvetan Tsalinsky

DP/37/2003 The Demand for Euro Cash: A Theoretical Model and Monetary Policy Implications
Franz Seitz

DP/38/2004 Credibility Level of the Bulgarian Exchange Rate Regime, 1991–2003: First Attempt at Calibration *(in Bulgarian only)*
Georgi Ganev

DP/39/2004 Credibility and Adjustment: Gold Standards Versus Currency Boards
Jean-Baptiste Desquilbet, Nikolay Nenovsky

DP/40/2004 The Currency Board: 'The only game in town' *(in Bulgarian only)*
Kalin Hristov

DP/41/2004 The Relationship between Real Convergence and the Real Exchange Rate: the Case of Bulgaria
Mariella Nenova

DP/42/2004 Effective Taxation of Labor, Capital and Consumption in Bulgaria *(in Bulgarian only)*
Plamen Kaloyanchev

DP/43/2004 The 1911 Balance of Payments of the Kingdom of Bulgaria *(in Bulgarian only)*
Martin Ivanov

DP/44/2004 Beliefs about Exchange Rate Stability: Survey Evidence from the Currency Board in Bulgaria
Neven T. Valev, John A. Carlson

DP/45/2005 Opportunities of Designing and Using the Money Circulation Balance *(in Bulgarian only)*
Metodi Hristov

DP/46/2005 The Microeconomic Impact of Financial Crises: The Case of Bulgaria
Jonathon Adams-Kane, Jamus Jerome Lim

DP/47/2005 Interest Rate Spreads of Commercial Banks in Bulgaria *(in Bulgarian only)*
Mihail Mihailov

DP/48/2005 Total Factor Productivity Measurement: Accounting of Economic Growth in Bulgaria *(in Bulgarian only)*
Kaloyan Ganev

DP/49/2005 An Attempt at Measurement of Core Inflation in Bulgaria *(in Bulgarian only)*
Kalina Dimitrova

DP/50/2005 Economic and Monetary Union on the Horizon
Dr Tsvetan Manchev, Mincho Karavastev

DP/51/2005 The Brady Story of Bulgaria *(in Bulgarian only)*
Garabed Minassian

DP/52/2005 General Equilibrium View on the Trade Balance Dynamics in Bulgaria
Hristo Valev

DP/53/2006 The Balkan Railways, International Capital and Banking from the End of the 19th Century until the Outbreak of the First World War
Peter Hertner

DP/54/2006 Bulgarian National Income between 1892 and 1924
Martin Ivanov

DP/55/2006 The Role of Securities Investor Compensation Schemes for the Development of the Capital Market *(in Bulgarian only)*
Mileti Mladenov, Irina Kazandzhieva

DP/56/2006 The Optimal Monetary Policy under Conditions of Indefiniteness *(in Bulgarian only)*
Nedyalka Dimitrova

DP/57/2007 Two Approaches to Estimating the Potential Output of Bulgaria
Tsvetan Tsalinski

DP/58/2007 Informal Sources of Credit and the 'Soft' Information Market (Evidence from Sofia)
Luc Tardieu

DP/59/2007 Do Common Currencies Reduce Exchange Rate Pass-through? Implications for Bulgaria's Currency Board
Slavi T. Slavov

DP/60/2007 The Bulgarian Economy on Its Way to the EMU: Economic Policy Results from a Small-scale Dynamic Stochastic General Equilibrium Framework
Jochen Blessing

DP/61/2007 Exchange Rate Control in Bulgaria in the Interwar Period: History and Theoretical Reflections
Nikolay Nenovsky, Kalina Dimitrova

DP/62/2007 Different Methodologies for National Income Accounting in Central and Eastern European Countries, 1950–1990
Rossitsa Rangelova

DP/63/2008 A Small Open Economy Model with a Currency Board Feature: the Case of Bulgaria
Iordan Iordanov, Andrey Vassilev

DP/64/2008 Potential Output Estimation Using Penalized Splines: the Case of Bulgaria
Mohamad Khaled

DP/65/2008 Bank Lending and Asset Prices: Evidence from Bulgaria
Michael Frömmel, Kristina Karagyozova

DP/66/2008 Views from the Trenches: Interviewing Bank Officials in the Midst of a Credit Boom
Neven Valev

DP/67/2008 Monetary Policy Transmission: Old Evidence and Some New Facts from Bulgaria
Alexandru Minea, Christophe Rault

DP/68/2008 The Banking Sector and the Great Depression in Bulgaria, 1924–1938: Interlocking and Financial Sector Profitability
Kiril Danailov Kossev

DP/69/2008 The Labour Market and Output in the UK – Does Okun's Law Still Stand?
Boris Petkov

DP/70/2008 Empirical Analisys of Inflation Persistence and Price Dynamics in Bulgaria
Zornitsa Vladova, Svilen Pachedjiev

DP/71/2009 Testing the Weak-form Efficiency of the Bulgarian Stock Market
Nikolay Angelov

DP/72/2009 Financial Development and Economic Growth In Bulgaria (1991–2006). An Econometric Analysis Based on the Logic of the Production Function)
Statty Stattev

DP/73/2009 Autonomy vs. Stability: the Relationship between Internal and External Money in Bulgaria (1879–1912)
Luca Fantacci

DP/74/2009 The Size of the Shadow Economy in Bulgaria: A Measurement Using the Monetary Method
Hildegart Ahumada, Facundo Alvarado, Alfredo Canavese, Nicolàs Grosman

DP/75/2009 Efficiency of commercial banks in Bulgaria in the wake of EU accession
Kiril Tochkov, Nikolay Nenovsky

DP/76/2009 Structural Current Account Imbalances: Fixed Versus Flexible Exchange Rates?
Slavi T. Slavov

DP/78/2009 Explanations for the Real Exchange Rate Development in the New EU Member States in Transition
Galina Boeva

DP/79/2009 The Great Depression in the Eyes of Bulgaria's Inter-war Economists (How History of Economic Thought Could Matter for Today's Policy Advice)
Stefan Kolev

DP/80/2010 Modeling Interest Rates on Corporate Loans in Bulgaria *(in Bulgarian only)*
Mihail Mihailov

DP/81/2010 A Small Open Economy Model with Financial Accelerator for Bulgaria: The Role of Fiscal Policy and the Currency Board
Ivan Lozev

DP/82/2010 The Impact of the Global Economic Crisis on Bulgaria's Accession to the Euro Area *(in Bulgarian only)*
Tsvetelina Marinova

DP/83/2011 Are Long-term Inflation Expectations Well-anchored? Evidence from the Euro Area and the United States
Tsvetomira Tsenova

DP/84/2011 Relative Inflation Dynamics in the EU Accession Countries of Central and Eastern Europe
Hiranya K Nath, Kiril Tochkov

DP/85/2011 Trade, Convergence and Exchange Rate Regime: Evidence from Bulgaria and Romania
Emilia Penkova-Pearson

DP/86/2011 Short-Term Forecasting of Bulgarian GDP Using a Generalized Dynamic Factor Model
Petra Rogleva

DP/87/2011 Wage-setting Behaviour of Bulgarian Firms: Evidence from Survey Data
Ivan Lozev, Zornitsa Vladova, Desislava Paskaleva

DP/88/2012 The Predictive Power of Some Market Liquidity Risk Measures: An Empirical Approach
Assoc. Prof. Tsvetan Manchev, Ph. D. Daniel Simeonov, Hristo Ivanov, Christian Hausmann

DP/89/2012 Survey Evidence on Price-setting Behaviour of Firms in Bulgaria
Zornitsa Vladova

DP/90/2013 Fiscal Policy and Economic Growth in Bulgaria
Kristina Karagyozova-Markova, Georgi Deyanov, Viktor Iliev

DP/91/2013 Financial Contagion and Network Models of the Banking System *(in Bulgarian only)*
Tsvetelina Nenova

DP/92/2013 Agent-based Systems and Their Applications in Macroeconomic and Financial Modelling *(in Bulgarian only)*
Andrey Vassilev, Georgi Deyanov, Svilen Pachedjiev

DP/93/2014 Yield Curve Fitting with Data from Sovereign Bonds
Yavor Kovachev, Daniel Simeonov

DP/94/2014 Constant Market Shares Analysis beyond the Intensive Margin of External Trade
Marina Dyadkova, Georgi Momchilov

DP/95/2014 Factors of Credit Dynamics outside the Euro Area *(in Bulgarian only)*
Petar Peshev

DP/96/2014 Factors Affecting Bank Interest Rate Spread Dynamics in Bulgaria *(in Bulgarian only)*
Mihail Mihailov

DP/97/2014 Lending Interest Rates in Bulgaria: the Role of Euro Area Monetary Conditions and Economic Activity Effects *(in Bulgarian only)*
Mihail Mihailov