A Balls and Bins Model of Trade: Comment

Bernardo S. Blum

Sebastian Claro

Ignatius J. Horstmann[1]

May 27, 2015

Blum: Rotman School of Management, University of Toronto, 105 St George Street, Toronto, ON M5S3E6, Canada (email: bblum@rotman.utoronto.ca); Claro: Central Bank of Chile, Agustinas 1180, Santiago, Chile (email: sclaro@bcentral.cl); and Pontificia Universidad Catolica de Chile, Av. Vicuna Mackenna 4860, Macul, Santiago, Chile; Horstmann: Rotman School of Management, University of Toronto, 105 St George Street, Toronto, ON M5S3E6, Canada (email: Ihorstmann@rotman.utoronto.ca). We thank an editor, two anonymous referees, and Roc Armenter and Miklos Koren for helpful feedback. Bernardo S. Blum and Ignatius J. Horstmann acknowledge the financial support of the Social Sciences and Humanities Council (SSHRC Standard Research Grant 410-2011-0475). The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

Introduction

In a recent paper, Armenter and Koren (2014) (AK hereafter) propose what they call "a parsimonious statistical benchmark" model of trade – "A Balls-and-Bins Model of Trade" – which they use "to discern which data moments are informative about the correct model of the extensive margin of trade" (pg. 2128). In this note, we argue that the AK model is not a statistical benchmark model but embodies a key economic assumption. We identify what this assumption is and show that i) the assumption is strongly rejected by the data and ii) the assumption is key to the AK firm-level trade results. We demonstrate the latter by showing that a modified balls-and-bins model that incorporates a data-consistent assumption fails to match the zeroes in the firm-product-country trade data. Based on this analysis, we conclude that statistics such as firm-product-country level zeroes may be informative and so allow us to distinguish among competing models of trade.[2]

In the AK benchmark (balls-and-bins) model, bins represent product or product-country trade outcomes while balls represent trade shipments. Shipments (balls) are matched to trade outcomes (bins) in a purely random fashion (balls are randomly allocated to bins). AK interpret the model as a statistical model in the sense that bin sizes are determined by aggregate trade flow data, with larger bins representing products and/or countries with a larger share of total US exports.

AK show that this benchmark model "quantitatively reproduces the pattern of zero product- and firm-level trade flows … and the frequency of multi-product, multi-destination exporters" (abstract, pg. 2127). They posit that the reason that the balls-and-bins model is successful at re-producing the pattern of trade zeroes in the data is the combination of sparse trade data and the highly skewed nature of trade distributions. In particular, because the number of export shipments from the US is small *relative to* the total number of possible product / product-country classifications, there must, of necessity, be many trade zeroes, even if shipments are assigned to classifications in a purely random fashion. The fact that the distribution of exporter sizes is skewed means that, in general, most exporters are small and carry few (one) shipments, and thus even a random allocation of shipments will match well with trade zeroes

[2] To quote AK, "If, on the other hand, the balls-and-bins model does not match a statistic, then a trade model has a chance to distinguish itself from the competition by positing the correct structure on the joint distribution of trade across countries, products and firms." (p. 2129)

in the data. AK argue that the ability of the balls-and-bins model to re-produce the pattern of trade zeroes in the data means that such statistics “are not helpful in differentiating models of the extensive margin of trade”. (p. 2129)

The AK product level results provide an important insight: Because shipments are few relative to the number of potential product exports, one should expect to find, even with a random allocation mechanism, that many products are not exported to a given country. As a result, statistics based on zeroes in the product-country trade flow data are unlikely to be informative about models of the extensive margin of trade.

Where we disagree with AK is in their analysis of firm-product-country trade flows. Specifically, we show in what follows that the AK firm-product-country results hinge critically on an economic modeling assumption; namely, that exporters, regardless of size, have identical shipment sizes.[3] In fact, the data reveal that smaller exporters have significantly smaller shipment sizes than do larger exporters, highlighting the fact that export shipment numbers and sizes are the outcomes of firms’ economic choices. In essence, balls in the AK data generating process are not a simple abstraction representing a draw from a skewed shipment distribution that is matched to a bin from a skewed product-country distribution. As such, the analogy to the Ellison and Glaeser (1997) benchmark of firm locations being determined by firms throwing darts at a map of the United States is inappropriate.

To demonstrate the importance of the assumption on shipment size, we re-derive the AK firm-product-country results using the actual export shipment data to assign balls to exporters. Doing so amounts to explicitly embedding an alternative model of exporters’ choice of shipment sizes into the balls-and-bins structure. We show that the data strongly support the notion that firms export fewer products to significantly fewer destinations than a random allocation mechanism would predict. This is the case for exporters that sell one product and sell to one country, and also for exporters that sell multiple products to many countries. The evidence is overwhelming that the embedded economic model of shipment size determination is crucial to the AK results.

[3] AK make this assumption because they lack firm-level data on shipments.

The conclusion we draw from our analysis is that, at the firm-product-level, the balls-and-bins model proposed by AK is, in fact, an economic model of the extensive margin of trade and not a parsimonious statistical benchmark (as suggested by AK). Because it is an economic model and not a statistical benchmark, we conclude that it cannot be used "to discern which data moments are informative about the correct model of the extensive margin of trade". (p. 2128) Like any other economic model, the balls-and-bins model can have its relative performance evaluated by data, but cannot evaluate whether the data are "helpful in differentiating models" (p. 2129) As an economic model of the extensive margin of trade, not surprisingly, balls-and-bins does not perform well.

Data

The data we use are all Chilean export transactions in 2006, available from the Chile's Customs Office.[4] For each transaction in the data, we have information on the identity of the exporter, the HS 8-digit product classification code of the exported product, the destination country, and characteristics of the transaction, such as date, value, etc.

Table 1 reports summary statistics of the data. In 2006, almost 7,000 Chilean exporters sold USD 55 billion to 181 countries. These exports were classified under 5,218 different HS 8-digit, and 3,551 different HS 6-digit categories respectively. Following AK, we consider a shipment as "all merchandise sent from one firm to one foreign consignee, to a single foreign country of ultimate destination, on a single carrier, on the same day." Whenever a transaction comprises more than one product, we follow AK and define a shipment as a product-specific transaction. In our data, this is given by a HS 8-digit code within a customs form (DUS - *Documento Unico de Salida*). Based upon this definition, in 2006 there were 575,830 export transactions, with an average transaction value of USD 95.7 thousand.

[Table 1 around here]

One salient and well-known feature of Chilean export data is the importance of copper products. The last column in Table 1 shows summary statistics on Chilean non-copper exports.[5] This column reveals that copper products account for 50% of all Chilean exports in

[4] www.aduana.cl

[5] This sample excludes the two main HS 6-digit copper categories: 740311 o 260300

2006, although copper products are exported by few exporters who make few but large transactions.

Excluding or not copper exports, the number of Chilean export shipments is small relative to the possible number of product-destination pairs. In 2006 there were just over half a million export shipments from Chile to 944,458 possible (HS 8) product-destination pairs. As a comparison, AK report that, in 2005, the number of US export shipments were ten times larger than the number of US product-destination pairs. Therefore, by this metric, Chilean export data are substantially sparser than the US export data.

Another key feature of US export data that is also present in our data is the skewness of export shipments across destination countries and exported products. The U.S. is the most frequent destination of Chilean exports, receiving 23.3% of all Chilean export shipments in 2006, followed by Peru (9.1%), Bolivia (6.0%) and Argentina (5.1%). The most frequently shipped among the more than 5 thousand exported products account for between 2% and 4% of all export shipments. It is worth noting that, excluding copper transactions does not affect these results, since copper products account for very few transactions.

In the rest of the paper we will report results using the non-copper export sample. We do so to make sure that our findings are not driven by this idiosyncrasy of Chilean exports. However, we should note that including copper products into the analysis only strengthen our findings. Moreover, we will show that running AK's random trade mechanism in our data produces the same results as they found in US data.

The content of the firm-level extensive margin of trade

AK note that the international trade data are sparse in the sense that the number of export shipments is small *relative to* the total number of possible product / product-country classifications. At the firm level this implies, for instance, that most exporters do not export most products and do not export to most destinations. Given that inference under sparse data is difficult or even impossible, AK investigate whether or not the level of sparseness of trade data makes it impossible to test models of firm-level export behavior. To do so, they propose what they refer to as "a parsimonious statistical benchmark" – the balls-and-bins model – to "discern which data moments are informative" (p. 2128). Using US data, they examine which firm-product-destination export patterns observed in the data can be re-produced by the balls-and-bins model. Those that can be re-produced by balls-and-bins, they argue, are not informative about "the correct model of the extensive margin" (p. 2128).

In AK's balls-and-bins model, each export shipment is a ball that is randomly allocated across different bins. Each bin represents a cell in a matrix in which each row represents a different product exported by any US firm, and each column represents a different export destination. In other words, each bin represents a different product-country pairing. In the 2005 US export data there are 8,867 10-digit HS products and 229 export destinations, resulting in a total of 2 million plus bins. The size of each bin is computed as the share of total U.S. export shipments for that product times the share of total U.S. export shipments for that destination country. This construction guarantees that, with enough shipments, the random allocation mechanism reproduces the aggregate US export pattern in terms of product and country shipment share, including those associated with the gravity equation.

Each exporting firm is assigned a number of balls, meant to capture the number of export shipments it makes. Because AK do not have data on the number of shipments by firm, they cannot assign balls to exporters based on actual shipment data. Instead they impute a number of balls by assuming that all shipments by all exporters are of a constant dollar value, USD 36,000. This imputation means that any firm exporting USD 36,000 or less is assigned one ball, while firms with exports between USD 36,000 and USD 72,000 are assigned two balls, and so on. Each firm is assumed to allocate the balls randomly across all possible bins.

The outcomes under a balls-and-bins mechanism can be computed as follows. Any shipment (ball) represents an export by exporting firm $f = 1, 2, \dots, F$ of product $i = 1, 2, \dots, I$ to country $k = 1, 2, \dots, K$. Let s_i be product $i's$ share of total export shipments. Then, the probability that a firm, *f*, with n_f balls exports product i only is $s_i^{n_f}$ and the probability that a firm f with n_f balls exports a single product only is $\Pr(k = 1 / n_f) = \sum_{i=1}^{I} s_i^{n_f}$. Finally, the fraction of firms exporting a single product becomes $\Pr(k = 1/n_1, n_2, \dots, n_F) = \frac{1}{F} \cdot \sum_{f=1}^{F} \sum_{i=1}^{I} s_i^{n_F}$.

The expected fraction of single-destination exporters is computed similarly. Let s_k be the share of all shipments to country k. The share of firms that export to a single country is equal to $\frac{1}{F} \cdot \sum_{f=1}^{F} \sum_{k=1}^{K} s_k^{n_f}$. The share of single product-single destination exporters is computed as $\frac{1}{F} \cdot \sum_{f=1}^{F} \sum_{k=1}^{K} \sum_{i=1}^{I} s_{ik}^{n_f}$, where $s_{ik} = s_i \cdot s_k$.

The first two columns in Table 2 summarize empirical regularities 3, 4 and 5 from AK and the analogous outcomes from the AK balls-and-bins model. So, for instance, in the data 42% of firms export only one product while the balls-and-bins model predicts 43%; the share of total exports accounted for by these firms is 0.3%, slightly lower than the 0.4% in the data. These outcomes lead AK to argue that the evidence that many exporters sell one product, sell to one country, and sell one product to one country is not informative in the sense that these export outcomes are consistent with a purely random process. To illustrate, even if firms did not face country-specific fixed costs of exporting, many of them should still be expected to sell to one destination country. In general, these results cast doubts on the empirical content of the evidence on the extensive margin of trade.

[Table 2 around here]

The feature of the data that yields these results is straightforward and is recognized by AK. Specifically, it is the skewness of the exporter distribution – there are many small exporters sending one shipment under the constant shipment size assumption – that enables the balls-and-bins model to match the data. We show below that this assumption of constant shipment size significantly overstates the actual number of single shipment exporters in the Chilean data and so biases the results in favor of the random allocation model.[6]

Figure 1 shows, for each Chilean exporter in 2006, the relationship between how much this exporter sells and the average size of its export transactions. The pattern is clear: large exporters make larger transactions. The systematic relationship between the size of an exporter and shipment size depicted in Figure 1 rejects the constant shipment size assumption made by AK. More importantly, it provides evidence consistent with the notion that shipment size (and frequency) is the outcome of economizing decisions on the part of individual

[6] What proves to be important here is not the imputed dollar value but the fact that large and small exporters are assumed to have a constant shipment size.

exporters.[7] In this case, the balls-and-bins model cannot be considered a purely statistical model since it embodies an economic assumption about individual exporter behavior.

[Figure 1 around here]

This observation, of course, becomes important only if the constant shipment size assumption is practically relevant in balls-and-bins re-producing the data patterns in Table 2. If it were the case that, under any assumed relationship between exporter size and shipment size, a random allocation mechanism could generate the data patterns on the extensive margin of trade, the observation that the balls-and-bins model is not a purely statistical model would be largely irrelevant: its economic part is not determinant to its ability to re-produce the patterns on the extensive margin of trade.

The evidence in Figure 1 suggests that this will not be the case. The figure clearly shows that AK are grossly underestimating the number of transactions made by small exporters when they assume a constant value of export transactions across small and large export firms. In the actual shipment data, many small exporters carry multiple export transactions that often involve the same product and destination. Indeed, over half of the exporters that would get one ball if shipments were allocated based on a constant transaction size equal to USD 48,400—the average in the data—actually make multiple shipments. Of these, 25 percent are single product exporters, 63 percent export to a single destination and almost 20 percent are single product and single destination country exporters. In a random setting, almost surely these multiple shipment firms will not be single product and single destination exporters, which will generate a mismatch between the data and the balls-and-bins model.

To confirm that the assumption about the relationship between exporter size and shipment size has a material impact on the predictions of a balls-and-bins model, we re-simulate the model, assigning shipments to firms based on the actual shipment data.[8] This re-simulation is performed at the exact same level of aggregation as the one performed by AK, i.e., the shipment level. The only difference is that AK assume that shipment sizes are constant across

[7] See for instance the seminal work of Arrow, Harris, and Marschak (1951) and the literature that followed. Recently, Hornok and Koren (2015) and Kropf and Sauré (2014) provide evidence that per-shipment costs affect exporters' shipment size and frequency.
[8] In this simulation, firms are assigned their actual number of shipments as well as their actual shipment sizes.

exporters while we use the actual shipments sizes, which vary across exporters in a systematic way.

The last three columns in Table 2 show, respectively, the key empirical regularities in the Chilean data, and simulations of the balls-and-bins model for the Chilean data using constant- and actual-size shipments. The third column in table 2 shows the actual Chilean data in 2006, where 52% of exporters sell to a single destination, 40% sell a single product, and 33% sell a single product to a single destination country. The next column shows the balls-and-bins model when we assume, as AK does, that all export shipments are of the same dollar value (the average shipment size). In this case, the balls-and-bins model closely matches the share of single-destination exporters, and it even over predicts the number of single-product and single-product-single-destination exporters. As in AK, the high incidence of single-destination, single-product, and single-product-single-destination exporters observed in trade data can be generated by the random setting when shipment sizes are assumed to be constant.

The last column of Table 2 reports the results when we use the actual shipment data for each exporter. The modified balls-and-bins model consistently misses the data. It predicts that exporters with multiple balls export many products to many destinations. However, in the data, a large share of small exporters with multiple shipments are single product and/or single destination firms. These patterns cannot be replicated by the model. Likewise, the number of single-destination single-product firms is markedly smaller than in the data (and also than in the constant shipment size balls-and-bins model, which predicts a very large number of firms in this category). [9]

These simulations demonstrate that assumptions about the way that exporters determine shipment size – economic modeling assumptions – are not innocuous to the ability of the balls-and-bins model to re-produce key data patterns on the extensive margin of trade. This fact lead us to conclude that the balls-and-bins model of AK is not just statistical benchmark model but rather a simple economic model, with its assumption on the economic behavior of

[9] More generally, we can ask whether exporting firms specialize in destination countries and products in a way that is inconsistent with the random allocation mechanism. The data strongly show that both small and large Chilean exporters are much more specialized than a random allocation model of trade would suggest. These results are not reported to conserve space but are available upon request.

exporters in determining shipment size being fundamental to its ability to re-produce key aspects of the data.

Conclusion

In this note we show that the assumption made by Armenter and Koren in their model of country-product-firm trade, namely that export shipment size is constant across all exporters, is both not supported by the data and crucial to their results. In the Chilean data, shipment size is increasing in exporter size. Further, when actual shipment data are used to determine the number of balls assigned to each firm, a ball-and-bins model significantly underestimates the fraction of single product, single country exporters.

Perhaps more importantly, any assumption on shipment size, whether constant or increasing in exporter size, is an assumption about the economizing behavior of export firms. As a consequence, whether one utilizes the actual shipment data or some other assumption on shipment size, the resulting balls-and-bins model is not a purely statistical benchmark for evaluating trade zeroes in the data. One way or another, the model incorporates the economic decision of firms on the frequency of their shipments and these decisions matter to the balls-and-bins outcome. As such, the balls-and-bins model is correctly described as an economic model. Like any economic model, balls-and-bins can be evaluated by how well it matches the data but cannot be used to evaluate whether the data are "helpful in differentiating models" (p. 2129).

A comparison with the Ellison and Glaeser model is informative on this point. In their approach, the map of the United States has exogenously given topography and state boundaries and a dart is only a metaphor for a random matching process between firms and locations. In short, there is no economic content whatsoever in this random location choice process. For this reason, the outcome of the process can be utilized as a benchmark for a data set for the United States. The random benchmark can be built into an agglomeration index that shows, for any data set, the degree of agglomeration beyond the random benchmark. This is not the case for the balls-and-bins model because balls in the balls-and-bins model have real economic meaning and content. In particular, a firm's decision on shipment sizes and frequencies (the balls) are jointly determined with the products shipped and destinations served (in aggregate, the bins) by a set of underlying economic factors.

The fact that the number of balls is not independent from bin characteristics (and even bin labels) implies one of two possible approaches to a balls-and-bins analysis. In the first one, the economic factors that determine shipment sizes and frequencies (i.e.: the balls) are imposed during the allocation of balls to bins. In this case, the balls-and-bins outcome is not a purely random benchmark. Alternatively, the economic factors that determine shipment sizes and frequencies are not imposed during the allocation of balls to bins. In this case, the benchmark generated is random, but it is not useful as a benchmark, and the model is nothing more than a formal data diagnostic tool. To illustrate this latter point, suppose that the actual distribution of shipment sizes and frequencies consists of many single-shipment firms and that this is the result of the characteristics of the products being exported and the countries they are being exported to. If we simulate the balls-and-bins model using this actual distribution of shipment sizes and frequencies, because of the random assignment process, the outcome of the simulation is likely to be a different set of products and countries. However, were the same firms exporting those products to those countries, they likely would not have chosen the distribution of shipment sizes and frequencies (i.e.: balls) used in the simulation. In other words, the products exported and the countries served by firms in the simulation will be inconsistent with the observed shipment sizes and frequencies in the data. In this case, the random "benchmark" outcome generated by the balls-and-bins analysis cannot be claimed to be a useful benchmark for the choices observed in the actual data.

References

Armenter, Roc and Miklos Koren, 2014, "A Balls-and-Bins Model of Trade", *American Economic Review*, 104(7): 2127-2151.

Arrow, Kenneth J., Theodore Harris, and Jacob Marschak, 1951, "Optimal Inventory Policy", *Econometrica*, 19(3): 250-272.

Ellison, Glenn and Edward Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", *Journal of Political Economy* 105 (5): 889-927.

Hornok, Cecilia, and Miklos Koren, 2015, "Per-Shipment Costs and the Lumpiness of International Trade", *Review of Economics and Statistics* 97(2): 525-530.

Kropf, Andreas, and Philip Sauré, 2014. "Fixed Costs per Shipment", *Journal of International Economics*, 92(1): 166-184.

Table 1: Chilean Export Data - Summary Statistics 2006.

	All Exports	Non-Copper Exports
Number of Exporters	6,924	6,912
Exports (USD millions)	55,085	27,537
Number of Transactions/Shipments	575,830	568,906
Average Shipment Value (USD thousands)	95.7	48.4
Number of Destination Countries	181	181
Number of Exported Products:		
- HS8	5,218	5,216
- HS6	3,551	3,549

Source: Chile's Customs Office
Shipment Definition: NDUS-HS8-RUT

Table 2: Firm-level results

	US-Data 2005 (AK 2014)*		Chilean-Data 2006		
	Data	Balls and Bins Imputed Balls	Data	Balls and Bins Imputed Balls	Balls and Bins Real Balls
Single-destination Exporters					
Fraction of total exporters	64%	45%	52%	48%	25%
Share of total exports	3.3	0.3	2.9	0.6	0.2
Single-product Exporters					
Fraction of total exporters	42%	43%	40%	48%	24%
Share of total exports	0.4	0.3	7.9	0.6	0.2
Single-destination & Single-product Exporters					
Fraction of total exporters	40%	43%	33%	48%	24%
Share of total exports	0.2	0.3	0.9	0.6	0.2

Notes: *Armenter and Koren (2014) Table 1

Figure 1: Mean Transaction Size by Exporter – 2006