Working Paper DTC-2021-2

# How Misleading is Revealed Comparative Advantage?

Ben Shepherd, Principal, Developing Trade Consultants.[1]

315 W 99th St. #7C, New York, NY 10025, USA. Ben@Developing-Trade.com.

July 7th, 2021.

**Abstract**: Revealed Comparative Advantage (RCA) is a widely used indicator in applied work. Building on the analysis of French (2017), I show that not only is RCA not theoretically grounded, it in fact has very little to do with Ricardian productivity differences that are the basis of comparative advantage: productivity only explains around one percent of the observed variation in RCA, after accounting for the other elements of the index. Differences between RCA and a theory-consistent measure of comparative advantage are so major that applied work using RCA is likely to lead to incorrect policy conclusions.

**Keywords**: Gravity model; Comparative advantage; Trade policy.

**JEL Codes**: F10; F14.

[1] I am grateful to Anasuya Raj for helpful discussions.

# 1 INTRODUCTION

The Balassa (1965) Revealed Comparative Advantage (RCA) index is commonly used by applied researchers and policy practitioners. UNCTAD and the World Bank feature it prominently on their trade statistics websites,[2] while Hidalgo et al. (2007) use it as the basis of their product space literature. So it is important to understand the extent to which it in fact provides an accurate picture of comparative advantage in practice.

French (2017) investigates RCA from a theoretical perspective, and shows that it needs to be used with care and substantial modification. Costinot et al. (2012) use Eaton and Kortum (2002) Ricardian model to develop a theory-consistent measure of comparative advantage, but applied researchers and practitioners have proved reluctant to take it up (Leromain and Orefice, 2014, is an exception). Understanding the difference between RCA and the theory-consistent approach in an empirical setting is an important open question: Costinot et al. (2012) do not show how their measure differs from the traditional one in the data, while French (2017) provides limited empirical evidence for particular theory-consistent measures of comparative advantage, but does not compare them with traditional RCA.

The next section shows that the Balassa (1965) index incorporates Ricardian productivity as Costinot et al. (2012) understand it, but also a range of other factors related to country and sector size, as well as trade costs. I also extend the existing literature by highlighting the issue of domestic shipments in calculation of both indices. Section 3 implements the two measures in practice, and shows that RCA is largely uninformative as to Ricardian comparative advantage, as well as being very sensitive to the inclusion or exclusion of domestic shipments. The final section discusses the implications for applied researchers.

# 2 REVEALED VERSUS THEORETICAL COMPARATIVE ADVANTAGE

The Balassa (1965) RCA index claims that if a country specializes in a good relative to the world share, then it has a revealed comparative advantage in it. It therefore expresses the share of a good in a country's export bundle relative to its share in the world export bundle:

$$(1)\ RCA_{i_0}^{k_0} = \frac{\sum_{j=1}^{N} X_{i_0 j}^{k_0}}{\sum_{j=1}^{N}\sum_{k=1}^{G} X_{i_0 j}^{k}} \Bigg/ \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} X_{ij}^{k_0}}{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{G} X_{ij}^{k}}$$

Where $X_{i_0 j}^{k_0}$ is exports of product k0 by country i0 to country j.

Costinot et al. (2012) have shown that the Eaton and Kortum (2002) Ricardian gravity model takes the following form in a multi-sector context:

$$(2)\ X_{i0j}^{k} = t_{i0j} F_j^k {z_{i0}^{k}}^{\theta} e_{i0j}^{k}$$

[2] https://unctadstat.unctad.org/EN/RcaRadar.html; https://tcdata360.worldbank.org/indicators/h62a3e8cc?country=BRA&indicator=40085&viz=line_chart&years=1988,2016.

Where: $t_{i0j}$ is a country pair fixed effect capturing structural features of the model, such as trade costs; $\theta$ is a parameter from theory capturing intra-industry heterogeneity in productivity; $z_{i0}^{k}$ is the fundamental productivity of country i0 in sector k, taking account of factors like climate, infrastructure, and institutions; F is a fixed effect capturing importer-specific factors such as market size and relative prices; and $e_{i0j}^{k}$ is a standard error term.

The productivity term can be consistently estimated with an exporter fixed effect, so (2) takes the familiar structural gravity form, where the F terms are fixed effects:

$$(3)\ X_{i0j}^{k} = F_{i0}^{k} F_{j}^{k} t_{i0j} e_{i0j}^{k}$$

A theory based measure of comparative advantage (Theoretical Comparative Advantage, TCA) is given by the productivity term indexed to a base country and sector (i1, k0):

$$(4)\ TCA_{i_0}^{k_0} = \frac{F_{i_0}^{k_0\frac{1}{\theta}}}{F_{i_1}^{k_0\frac{1}{\theta}}}$$

To compare the two measures, I substitute (3) into (1):

$$(5)\ RCA_{i_0}^{k_0} = \frac{F_{i_0}^{k_0} \sum_{j=1}^{N} F_j^{k_0} t_{i_0j}^{k_0}}{\sum_{k=1}^{G} F_{i_0}^{k} \sum_{j=1}^{N} F_j^{k} t_{i_0j}^{k}} \Bigg/ \frac{\sum_{i=1}^{N} F_i^{k_0} \sum_{j=1}^{N} F_j^{k_0} t_{ij}^{k_0}}{\sum_{k=1}^{G} \sum_{i=1}^{N} F_i^{k} \sum_{j=1}^{N} F_j^{k} t_{ij}^{k}}$$

Equation (5) makes clear that RCA depends not just on exporter-specific Ricardian productivity, but also importer-specific factors like market size, as well as trade costs. Even with frictionless trade, RCA mixes exporter and importer effects.

The sums in equation (5) are typically calculated for the case of $i \neq j$, so domestic shipments are excluded. But there is no particular reason why that should be the case. For example, just because a country does not export a good due to prohibitively high trade costs does not mean that it does not have comparative advantage in its production relative to some baseline. Indeed, recent practice for gravity models like (3) is to include domestic shipments, for reasons such as consistency of estimates (Yotov, 2012), identification of policy effects (Heid et al., 2021), and achieving equality of estimated fixed effects and theoretical measures (Failly, 2015). Based on this literature, it is straightforward to define RCA' and TCA' through equations (1) and (5) above but including domestic shipments.

I take the logarithm of (5) and rearrange:

$$
\begin{aligned}
(6)\ \log RCA_{i_0}^{k_0} &= \log F_{i_0}^{k_0} + \log\left(\sum_{j=1}^{N} F_j^{k_0} t_{i_0 j}^{k_0}\right) \\
&- \log\left(\sum_{k=1}^{G} F_{i_0}^{k} \sum_{j=1}^{N} F_j^{k} t_{i_0 j}^{k}\right) - \log\left(\sum_{i=1}^{N} F_i^{k_0} \sum_{j=1}^{N} F_j^{k_0} t_{ij}^{k_0}\right) \\
&+ \log\left(\sum_{k=1}^{G}\sum_{i=1}^{N} F_i^{k} \sum_{j=1}^{N} F_j^{k} t_{ij}^{k}\right) \\
&\equiv \log F_{i_0}^{k_0} + \log\left(\frac{\sum_{j=1}^{N} X_{i_0 j}^{k_0}}{F_{i_0}^{k_0}}\right) - \log\left(\sum_{j=1}^{N}\sum_{k=1}^{G} X_{i_0 j}^{k}\right) - \log\left(\sum_{i=1}^{N}\sum_{j=1}^{N} X_{ij}^{k_0}\right) \\
&+ \log \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{G} X_{ij}^{k} \equiv \log F_{i_0}^{k_0} + \log\left(\frac{\sum_{j=1}^{N} X_{i_0 j}^{k_0}}{F_{i_0}^{k_0}}\right) + D_i + D^k + c
\end{aligned}
$$

Where the D terms are fixed effects in the relevant dimensions, and c is a constant. Equation (6) shows that RCA is related to the productivity parameter underlying TCA, i.e. TCA without indexing by a base country and sector, but with important adjustments. The second term indicates that countries with larger potential markets (exports without exporter-side factors) receive a higher RCA score relative to TCA. The constant is just an index effect, since the decomposition is in logarithms. But the exporter fixed effects show that countries with larger total exports are penalized relative to TCA, and the same goes for sectors with larger total exports. So the two measures are connected, but RCA includes extraneous information, in the sense of being uninformative of an exporter's level of productivity in a given sector. The damage done to RCA as an indicator is not minor: it is plausible that RCA may not say anything of interest about productivity, given the influence of factors like market size and trade costs. The extent to which that proposition is true, however, depends on the data.

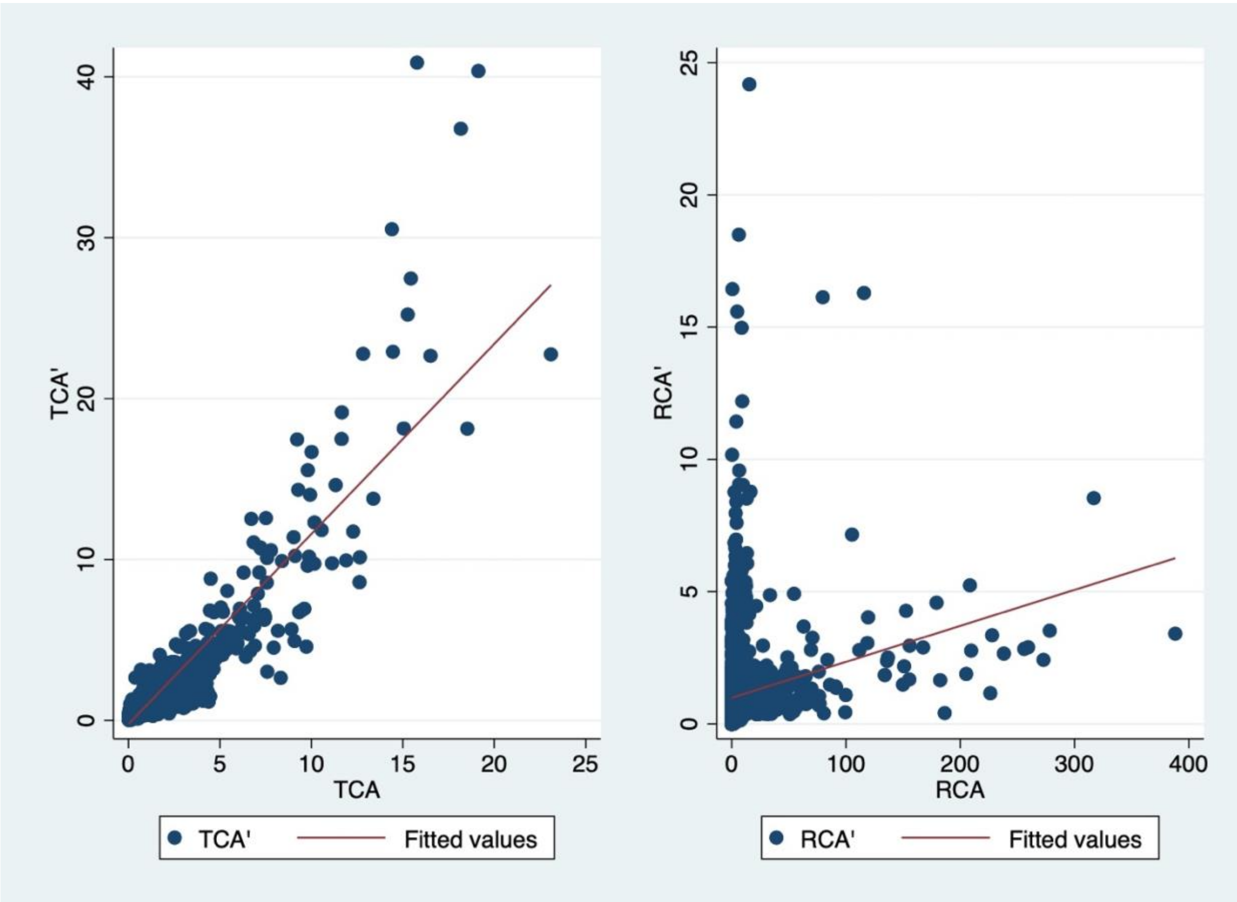
# 3 How much do the Differences Matter Empirically?

The above analysis has shown that RCA suffers from two flaws relative to its theoretical counterpart TCA: it excludes domestic shipments, and it is influenced by demand-side factors and trade costs, in addition to supply-side productivity. But how important are these problems in practice? To investigate, I compare TCA and RCA for a single dataset. To include domestic shipments, I take the Eora multi-region input-output (MRIO) and use it to construct standard measures of bilateral exports. I retain a single year of data (2015, the latest available) and all 26 sectors, which covers services as well as goods. This process yields a dataset of 870,714 observations: 183 exporters and importers, and 26 sectors. Estimates of the trade elasticity $\theta$ come from Egger et al. (2018), concorded manually to Eora sectors. The gravity models used to obtain estimates of $F_{i_0}^{k_0}$ are estimated by Poisson Pseudo Maximum Likelihood (Santos Silva and Tenreyro, 2006), specifically the high dimensional fixed effects implementation due to Correia et al. (2020).[3]

[3] Costinot et al. (2012) log-linearize their model and estimate by OLS. Santos Silva and Tenreyro (2006) show that this approach yields inconsistent parameter estimates unless the error term is homoskedastic, which is a condition that does not appear to hold in empirical settings.

Figure 1 shows the extent of the first issue, namely exclusion of domestic shipments. RCA is much more sensitive to inclusion or exclusion of intranational trade than TCA, as evidenced by a much weaker correlation between RCA and RCA' (0.219) than between TCA and TCA' (0.889). There is no example in the published literature of an RCA analysis using domestic shipments, so the implication is that the way in which the index is typically used by practitioners, including the World Bank and UNCTAD websites cited above, could result in substantial errors of interpretation. Even at the impressionistic level of RCA > 1 (traditionally interpreted as indicating comparative advantage), only 28.631% of the sample has the same inequality holding for both measures.

*Figure 1: TCA and RCA with and without domestic shipments, 2015.*




The second question relates to the extent to which RCA misstates productivity-based comparative advantage by including extraneous factors. Since the two indices are on different scales, I use ranks to compare the baseline measures and the measures including domestic shipments. Irrespective of whether or not domestic shipments are included, there are major differences of interpretation between the two indicators. The Spearman rank correlation coefficient for RCA and TCA is 0.430, while for RCA' and TCA' it is 0.093. So including domestic shipments—as theory suggests is important—results in an even greater gap between the two measures than do the baseline indices typically used in the literature. But in any case, the two comparisons show that use of RCA instead of TCA could lead to radically different policy conclusions, as the two indicators identify completely different sectors as having comparative advantage.

The log decomposition in equation (6) can be taken one step further by using it as the motivation for calculating squared semi-partial correlations for each of the terms identified in it.[4] The squared semi-partial correlation between RCA and TCA gives the proportion of the observed variance in the former that is accounted for by variation in the latter only, i.e. excluding the effects of the other independent variables in equation (6). It is therefore equivalent to the increase in R2 given by adding TCA to a regression with the other variables.

Table 1 shows that productivity explains only a very small proportion of observed variation in RCA. Its contribution is dwarfed by those of the other factors. Based on these results, it is not going too far to say that RCA is essentially a measure of size and market potential rather than productivity-based comparative advantage.

*Figure 2: Squared semi-partial correlations based on equation (6).*

| Interpretation | Term | Sq. Semi-Partial Correlation |
|---|---|---|
| Productivity | $\log F_{i_0}^{k_0}$ | 0.012 |
| Foreign market potential | $\log\left(\frac{\sum_{j=1}^{N} X_{i_0 j}^{k_0}}{F_{i_0}^{k_0}}\right)$ | 0.514 |
| Total exports (country) | $\log\left(\sum_{j=1}^{N}\sum_{k=1}^{G} X_{i_0 j}^{k}\right)$ | 0.819 |
| Total exports (sector) | $\log\left(\sum_{i=1}^{N}\sum_{j=1}^{N} X_{ij}^{k_0}\right)$ | 0.407 |

# 4 CONCLUSION

RCA has a long history among applied researchers and policy practitioners. But I have shown that only around one percent of the observed variation in RCA is due to differences in productivity, as estimated rigorously, after accounting for other components of the index. It has much more to do with market size and trade costs than productivity, and so is a poor guide to policy advice based on notions of comparative advantage.

Thankfully, the problem is a simple one to remedy. The theory-consistent measure of Costinot et al. (2012) can easily be implemented using the high dimensional fixed effects estimator of Correia et al. (2020). Their approach was able to estimate each of the models used here—870,531 observations with over 43,000 fixed effects—in less than one minute on a standard desktop. So computational feasibility is no longer a reason for persisting with RCA, given that theory and data both provide compelling reasons for preferring TCA.

[4] An OLS regression in the form of equation (6) has an R2 of exactly one, with exact estimates of the coefficients implied by the decomposition. This result may appear surprising, since $F_{i_0}^{k_0}$ is estimated using a gravity model, not directly observed in the data. The reason is that as Arvis and Shepherd (2013) and Failly (2015) show, actual and predicted values for group sums in a PPML regression are exactly identical.

## REFERENCES


Arvis, J.-F., and B. Shepherd. 2013. "The Poisson Quasi-Maximum Likelihood Estimator: A Solution to the Adding Up Problem in Gravity Models." *Applied Economics Letters*, 20(6): 515-519.

Balassa, B. 1965. "Trade Liberalization and Revealed Comparative Advantage." *The Manchester School*, 33(2): 99-123.

Correia, S., P. Guimaraes, and T. Zylkin. 2020. "Fast Poisson Estimation with High Dimensional Fixed Effects." *Stata Journal*, 20(1): 95-115.

Costinot, A., D. Donaldson, and I. Komunjer. 2012. "What Goods do Countries Trade? A Quantitative Exploration of Ricardo's Ideas." *Review of Economic Studies*, 79: 581-608.

Eaton, J., and S. Kortum. 2002. "Technology, Geography, and Trade." *Econometrica*, 70(5); 1741-1779.

Fally, T. 2015. "Structural Gravity and Fixed Effects." *Journal of International Economics*, 97(1): 76-85.

French, S. 2017. "Revealed Comparative Advantage: What is it Good For?" *Journal of International Economics*, 106(C): 83-103.

Heid, B., M. Larch, and Y. Yotov. 2021. "Estimating the Effects of Non-Discriminatory Trade Policies Within Structural Gravity Models." *Canadian Journal of Economics*, 54(1): 376-409.

Hidalgo, C., B. Klinger, A.-L. Barabasi, and R. Hausmann. 2007. "The Product Space Conditions the Development of Nations." *Science*, 317(5837): 482-487.

Leromain, E., and G. Orefice. 2014. "New Revealed Comparative Advantage Index: Dataset and Empirical Distribution." *International Economics*, 139: 48-70.

Santos Silva, J., and S. Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4): 641-658.

Yotov, Y. 2012. "A Simple Solution to the Distance Puzzle in International Trade." *Economics Letters*, 117(3), 794-798.