Benchmarking across Borders: An Update and Response

Mark A. Kayser[*]
Michael Peress[†]

August 2018

Arel-Bundock, Blais and Dassonneville (2018, ABD hereafter) offer an unusual critique of our article, *Benchmarking across Borders.* They find no methodological flaws, produce identical empirical results, and concede that their proposed specification (Model 5) is mathematically identical to that used in Kayser and Peress (2012, KP hereafter). ABD make two claims: (1) that their preferred specification is an innovation that improves interpretation and (2) that the empirical evidence presented in KP does not support benchmarking. The first is unpersuasive and the second depends on a selective reading of the evidence. We address these issues below and update the individual-level dataset from KP to increase statistical power, finding additional evidence for benchmarking.

1 Value added?

Let us examine the improvement in clarity from ABD's innovation. KP define local growth as the deviation of national economic growth from an international benchmark yielding coefficients on local growth and on the international growth benchmark. This allows for the testing of three relevant null hypotheses (the null of no-benchmarking, the null of benchmarking, and the null of no economic voting) as well as the possibility of partial benchmarking. We concisely and correctly discuss how to interpret the three relevant null hypotheses and how to interpret the relative size of the local and international (a.k.a., global) coefficients as the degree of benchmarking in the last paragraph of page 668 of KP. ABD spend 4 pages describing how to test only one of the relevant nulls and then present a model (Model 5) with coefficients on un-benchmarked growth (G_y) and international growth (G_i). Readers confronted with coefficients such as in ABD's Table 1 are supposed to know that the coefficient on international growth does not capture the effect of international growth on the vote but, rather, the effect of cross-national benchmarking. This is a dubious improvement to clarity.

Nor is ABD's specification superior because it does not require a Wald test of the equality of two coefficients: it requires a Wald test for the null of benchmarking and the null of no

[*]Hertie School of Governance, Berlin, `kayser@hertie-school.org`
[†]SUNY-Stony Brook, `michael.peress@stonybrook.edu`

economic voting. Moreover, unlike ABD, the KP specification allows easy calculation of a point estimate of the degree of partial benchmarking (by dividing the global coefficient by the local coefficient). ABD's preferred specification is mathematically equivalent, as they acknowledge (e.g., footnote 40), less intuitive for the interpretation that matters and adds no value.

2   Mechanism

ABD additionally argue that "when correctly interpreted, the results [in KP] do not support the contention that voters make rational comparative evaluations." They arrive at this broad-stroke claim, however, by impugning an admittedly improbable mechanism that we do not espouse, neglecting the evidence for the mechanism that we do espouse, ignoring additional evidence of the effect that we report, and then testing only one possible null. Consider the mechanism. ABD consider implausible the assumption that voters are sufficiently informed and rational to assess foreign growth rates, compare them to those in their country, and hold politicians accountable. We also do not hold this to be the most plausible mechanism, which is why we develop and test a much different mechanism in KP that ABD barely engage.

We believe that voters (both low and high information) learn about economic growth primarily through media coverage, and that the most likely way that voters incorporate expectations is through consuming media that already contains this information. The media pre-benchmark economic news by reporting more positively (negatively) on the economy when it outperforms (underperforms) the international economy. We explicitly develop and refer to this pre-benchmarking mechanism throughout our article and test it empirically with newspaper data (KP pp. 679-80, Table 7). KP also present evidence that low and high information voters engage in similar amounts of benchmarking, which is not consistent with a sophisticated voter mechanism for benchmarking (KP appendix, pp. 3-9).[1] ABD, in turn, critique a straw man mechanism of hyper-informed and sophisticated voters that we never propose. Addressing an established mechanism rather than an inferred caricature would be a fairer way to assess an argument.

3   The Evidence for Benchmarking

ABD are no less selective in choosing how to interpret results and which empirical evidence to engage. Focusing solely on statistical tests of the equality of local and international growth effects, ABD ignore more than just our evidence for the media mechanism (Table 7 in KP) and against the voter sophistication mechanism for benchmarking. They also neglect KP's

---

[1]In the one brief passage in which ABD do acknowledge our mechanism, they mischaracterize it. They improperly dismiss our mechanism by referring to our finding that low and high information voters engage in an equal amount of benchmarking. While we interpret this as evidence against a voter-sophistication origin for benchmarking, ABD argue that information level is a proxy for media consumption and that high information voters should engage in more benchmarking if our media mechanism is correct. ABD seem to assume that voters learn about the international economy through media coverage and about the local economy elsewhere, which is strange position that we do not hold.

finding that un-benchmarked, relative to benchmarked, growth has a less stable effect across time and political institutions (Figure 2 and Table 4 in KP). The point estimates in almost all of our models for the degree of (partial) benchmarking suggest substantial benchmarking. The coefficients on global growth are usually near zero, the coefficients on local growth are large, positive, and statistically significant, but the two coefficients are statistically distinguishable only in the aggregate data (KP Table 1) and in the test of the media mechanism (KP Table 7). So, should we be troubled by the failure to reject the no-benchmarking null elsewhere?

Given that the individual-level sample in KP includes only 31 independent observations of the economy—one for each election study—the most likely explanation for the failure to reject the null of no-benchmarking is a lack of statistical power. We added the analysis of aggregate-level data in KP to see whether the no-benchmarking null could be rejected in a sample with more independent observations of the economy—it was (KP Table 1)—and whether the overall pattern in the coefficients would hold up—it did. That the coefficients on local and global growth in KP are not statistically differentiable at standard levels outside of Tables 1 and 7 could arise from voter behavior or simply from low statistical power.

Fortunately, in one of our two samples, new data and new methods allow us to markedly increase statistical power. Since KP, the Comparative Study of Electoral Systems (CSES) project has released modules 3 and 4, which increases the number of elections and independent economic observations in our individual-level sample from 31 to 58. In addition, we previously employed a conditional logit model with clustered standard errors, which is conservative because it allows for an arbitrary correlation structure within cluster. Given the nature of the CSES sample, the data are exchangeable within cluster so can also be analyzed using a conditional logit model with election-choice random effects, further improving power.

Table 1 below reports the results from running conditional logit models with clustered standard errors (as in KP) and election-choice random effects on the updated CSES election sample.[2] The conditional logit model with clustered standard errors yields similar coefficients to KP with reduced standard errors for global growth. The Wald test reports that the local and global growth coefficients are now statistically distinguishable using the median measure. When applying the random effects estimator, the coefficients are statistically significantly different for both the median and principal components measures of international growth. Neither model using the trade measure of international growth can reject the no-benchmarking null but the model results with the trade measure were also the weakest in KP suggesting that a country's five largest trade partners may not constitute the most relevant benchmark.[3] These results show that ABD's solitary critique of KP — the failure to reject the null of no-benchmarking outside of Tables 1 and 7 of KP — is nothing more than a power issue.

---

[2]The expanded economic data are taken from Kayser and Peress (2018).

[3]China, which is distant and politically and economically distinct from many sample countries, frequently ranks in the top five.

3

	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	Conditional Logit			Conditional Logit w/ R.E.		
Decomposition Method	Median	PC	Trade	Median	PC	Trade
Party Utility Parameters						
Distance	-0.486***	-0.487***	-0.487***	-0.490***	-0.490***	-0.492***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.018)	(0.016)
Outside Option	-2.078***	-2.081***	-2.071***	-2.434***	-2.354***	-2.514***
	(0.170)	(0.170)	(0.171)	(0.084)	(0.099)	(0.104)
Leader Party	1.420**	0.764**	1.069*	1.157***	0.901***	1.023***
	(0.503)	(0.234)	(0.460)	(0.213)	(0.093)	(0.159)
Local Growth * Leader Party	0.141*	0.077+	0.070	0.264***	0.095***	0.094***
	(0.058)	(0.044)	(0.051)	(0.021)	(0.017)	(0.018)
Global Growth * Leader Party	0.005	0.028	0.042	-0.016	0.034*	0.033
	(0.039)	(0.041)	(0.048)	(0.010)	(0.015)	(0.027)
Local Unem. * Leader Party	0.007	-0.016	-0.013	-0.036***	-0.051***	-0.095***
	(0.033)	(0.034)	(0.034)	(0.009)	(0.013)	(0.012)
Global Unem. * Leader Party	-0.103	-0.001	-0.051	-0.089**	-0.074***	-0.085***
	(0.086)	(0.032)	(0.065)	(0.034)	(0.012)	(0.019)
p-Value from Wald Test						
Local Growth = Global Growth	0.027*	0.210	0.659	0.000***	0.004**	0.101
Local Unem. = Global Unem.	0.232	0.508	0.417	0.106	0.066+	0.534
Number of Voters	71196	71196	70523	71196	71196	70523
Number of Elections	58	58	57	58	58	57

Table 1: Expanded Individual-level Results for Benchmarking in the Economic Vote — Columns (1)-(3) report the results for a conditional logit model with standard errors clustered by election study. Columns (4)-(6) report the results for a conditional logit model with election-choice random effects. $^{+}p < .10,^{*}p < .05,^{**}p < .01,^{***}p < .001$.

4 Conclusion

ABD paint a dire picture of the state of benchmarking research, however they critique a straw man mechanism, neglect an explicitly developed and tested mechanism, ignore evidence in support of benchmarking, and selectively interpret any ambiguity as evidence for their preferred theory.[4] Their positive contribution is a mathematically identical specification that requires readers to interpret the coefficient on its key variable, international growth, as something other than the effect of international growth. A correct reading of the evidence in KP suggests that benchmarking is supported and no-benchmarking is not. Analysis of an extended individual-level sample provides additional evidence in support of benchmarking.

References

Arel-Bundock, Vincent, Andre Blais and Ruth Dassonneville. 2018. "Do Voters Benchmark Economic Performance?" *British Journal of Political Science* .

Aytac, Selim Erdem. 2018. "A Response to Arel-Bundock, Blais and Dassonneville." Unpublished manuscript. http://home.ku.edu.tr/~saytac/Aytac_response_ABD.pdf.

---

[4]See Aytac (2018) for a response to ABD's critique of his benchmarking findings.

Kayser, Mark A. and Michael Peress. 2012. "Benchmarking across Borders: Electoral Accountability and the Necessity of Comparison." *American Political Science Review* 106(3):661–684.

Kayser, Mark A. and Michael Peress. 2018. "Accuracy and Bias in Media Coverage of the Economy: Evidence from Sixteen Developed Democracies." Paper presented at the 2018 annual meetings of the European Political Science Association, Vienna, Austria.