Sirianne Dahlum and Carl Henrik Knutsen

What Counts as Evidence? Panel Data and the Empirical Evaluation of Revised Modernization Theory

Replying to our article (D&K) – which shows that the proposed evidence for a clear causal effect of self-expression values on democracy is highly questionable – Welzel, Inglehart and Kruse (WIK) criticize our empirical evaluation of "Revized Modernization Theory".[1] They claim that it is "irrelevant" (p.9) and "poses no challenge" (p.12) to the theory, asserting that "the evidence supports the emancipatory theory of democracy as it did in I&W's original analyses" (p.2).[2]

In particular, WIK question our use of time-series cross-sectional (TSCS) data and models, due for example to the "tectonic" nature of regime change, thereby also suggesting that their proposed theory of values and democratization is not weakened by the extant data (as the theory receives support from purely cross-sectional regressions, at least when using their favored measure of democracy). However, these assertions do not hold up to scrutiny. Unless one is willing to make various very strong assumptions, which we argue are implausible, it is incorrect to conclude that our empirical criticism is irrelevant and that their theory is strongly supported by extant evidence.

More specifically, WIK's simulation exercise does not deem our critique irrelevant. We show that their simulated world corresponds poorly with the real world, stacking the odds against TSCS models. Moreover, the types of models we use in D&K actually do pick up effects, even in this simulated world. Further, WIK's real-world replication tests based on more observations in fact corroborate our null-results, despite WIK's suggestions to the contrary.

When WIK present corroborating evidence, they draw on a particular specification (resting on very stark assumptions, such as the absence of unobserved country-specific effects on democracy). Moreover, their response when faced with specifications not corroborating their theory is problematic, given conventional norms of inference: All measures other than their own (problematic) Effective Democracy Index (EDI) are claimed less valid; standard panel data methods are deemed irrelevant; and, even widely acknowledged threats to inference (e.g., omitted variable bias, p.3) are downplayed. If WIK really had *known* the "true model", this would have been legitimate. However, without *knowing* the data-generating process one should be careful not to rely overly on one specification. If results do not hold up across various (plausible) tests, the appropriate response would be to doubt the hypothesis rather than discredit all models except the one producing results in line with the theory. One does not have to hold a strict Popperian view of scientific testing for questioning

---

[1] Dahlum and Knutsen 2016; Welzel, Inglehart and Kruse 2016.
[2] I&W: Ingehart and Welzel 2005.

emancipative/revised modernization theory. It is currently supported by far too brittle evidence to conclude that it is true, and accepting it amounts to what Leamer would term a "fragile inference".[3]

WIK's simulation and critique of TSCS models

WIK argue that standard TSCS/panel data models are inappropriate for testing their theory, which assumes a ""tectonic" model of incrementally accruing tensions, causing rare eruptive shifts to release them" (p.2). We agree that *some* panel data models can have difficulties picking up effects for slow-moving and "tectonic" processes; this is well-known *and explicitly discussed in D&K* (see, e.g., the discussions on Fixed Effects vs Random Effects/System GMM models, and D&K's inclusion of dynamic probit models on regime transitions). Yet, we find it highly unlikely that most TSCS models would be unable to pick up any effect of value on democracy if it is as strong as theorized by I&W, and we are not persuaded to the contrary by WIK's simulation exercise. There are several reasons for this.

First, the EDI does *not* follow the "tectonic" process that WIK describe, with large shifts and long periods of constant scores, empirically. Figure 1 plots EDI trends for 12 countries, displaying more incremental year-to-year variation. Moreover, EDI (even over 15 years) does not always change monotonically (the same holds, empirically, for self-expression/emancipatory values, cf. WIK), and seldom tectonically. In contrast, WIK's simulated "Supply" changes only *once*, offering little variation for panel-data estimators (see Figure 2).
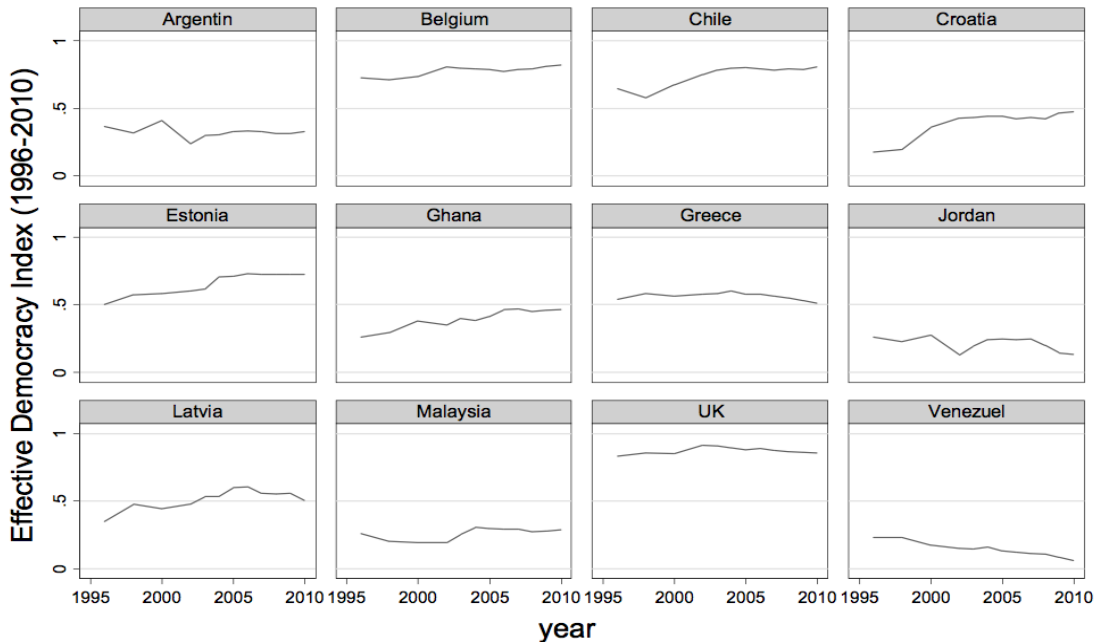

Figure 1: EDI 1996-2010

---

[3] Leamer 1985, 308

Despite providing a "low-powered" test (36 hypothetical countries, 20 years), WIK's inability to find effects in the hypothetical/simulated world where their theory applies would have been concerning for the appropriateness of TSCS models if the simulated world offers the same prospects for identifying effects as the real world. (Obviously, statistical testing would yield the same null-result in a world where the theory *does not* apply). However, their simulated universe is constructed in (unrealistic) ways, by e.g. assuming only one disruptive/monotonic change in EDI, which make it harder to identify effects by construction. This, and other issues discussed below, implies that WIKs simulation exercise does not invalidate the use of TSCS models.

Now, EDI is a highly problematic democracy measure, despite WIK's claim that it is the most appropriate and that previous criticisms have been debunked (we strongly disagree; many criticisms, e.g., on systematic and unsystematic measurement errors still stand).[4] Actually, standard measures such as ACLP/DD, Polity and Freedom House display real-world patterns resembling tectonic patterns somewhat better. But, these measures do not yield robust support for their theory even on cross-country variation. Only designs that draw heavily on cross-country- comparisons rather than temporal *and* use the problematic EDI yield support.

Table 1: Democracy/supply and values/demand in WIK's simulated world

	Full simulated sample (36 countries, 19 years)				Excl. 20 over-democratic countries	
	OLS PCSE	OLS PCSE	Random Effects	Fixed Effects	Random Effects	Fixed Effects
	(A1)	(A2)	(A3)	(A4)	(A5)	(A6)
Lagged demand	0.042	6.188***	7.450***	-21.154*	6.116***	10.586**
	(0.92)	(2.16)	(0.49)	(11.00)	(0.33)	(4.83)
Lagged dep. var.	0.932**					
(supply)	(0.09)					
Lagged development	0.033	-5.522*	-6.734**	20.849+	-5.166**	-4.448
	(0.88)	(2.31)	(0.49)	(10.42)	(0.32)	(5.23)
N	684	684	684	684	304	304

Notes: ***p<0.01, **p<0.05, *p<0.1. Errors (in parenthesis) adjusted for panel-specific AR(1) autocorrelation, contemporaneous correlation and heteroskedastic panels in PCSE, and clustered by country in RE/FE. Constant and country dummies omitted.

Second, even if we were to accept WIK's simulated world, the type of specifications that we *actually employed* in D&K outperform WIK's regressions in terms of picking up effects. We ran our OLS PCSE, Random Effects (RE) and Fixed Effects (FE) without lagged dependent variables (LDVs).

---

[4] See Hadenius and Teorell 2005; Teorell and Hadenius 2006; Knutsen 2010.

Operating with LDVs is problematic when the dependent variable exhibits as little temporal variation as in the simulated world, and the key independent variable, unrealistically, changes at a completely constant rate. No wonder that WIK's PCSE models fail to identify any relationship, and we replicate this (A1, Table 1). But, when throwing out the problematic LDV (A2), thus following D&K's baseline models, the PCSE models does pick up a positive significant (1%) coefficient on demand.

Still, A2 draws heavily on cross-country variation, and we therefore tested panel data models in D&K. Indeed, also an RE model (A3) yields a highly significant effect – RE models in D&K found no such real-world effect, even for far more countries and longer time series.

Third, the more conservative FE specification (A4) does not uncover the relationship. As highlighted in D&K, such models might be overly inefficient. But, another feature of the simulated world – matching poorly with real-world patterns *and* narratives in I&W and Welzel – is driving this result:[5] 20 of WIK's 36 simulated countries are "over-democratic" (Figure 2), starting out with higher democratic supply than demand. This contrasts with WIK's end-of-Cold-War analogy (suggesting several "under-democratic" countries) motivating that regimes can only switch in t=15. Further, among these 20, 16 have *increasing* demand for democracy, but still display downward supply-shifts in t=15 since they started out *much more democratic than theoretically expected*. Thus, this does not reflect that some countries are "over-democratic" because of gradually falling "emancipative values" either.
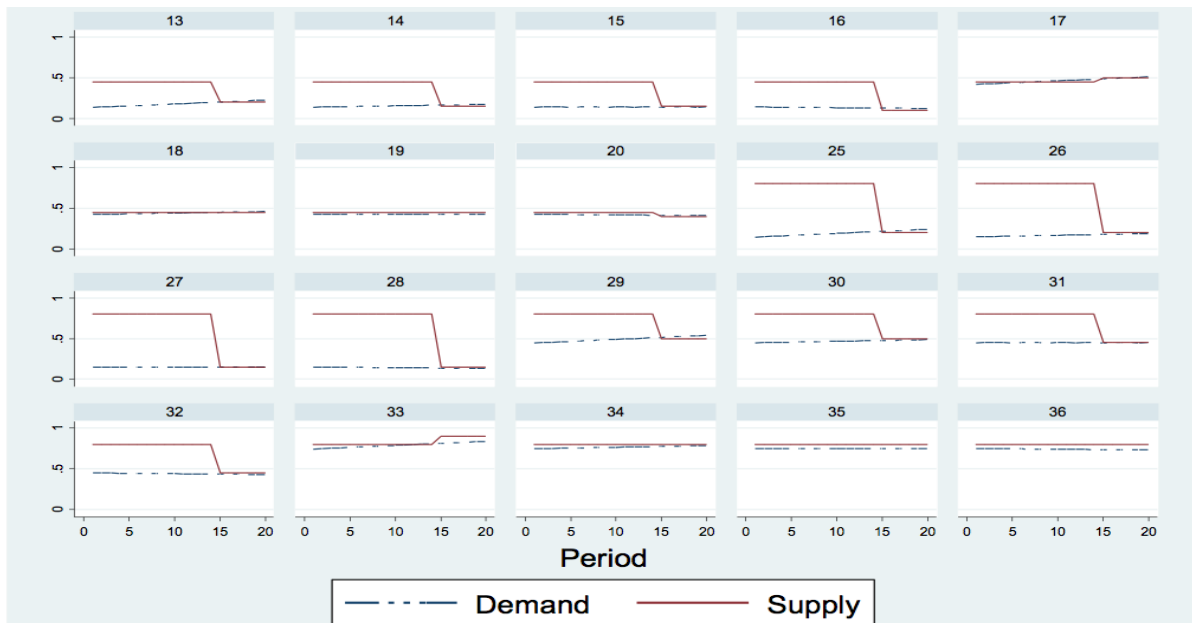

Figure 2: 20 (of 36) simulated cases with democratic "over-supply"

---

[5] Welzel 2013.

It is unclear which real-world/historical patterns that could have generated so many "artificially high" democracy scores; importantly, they cannot come from emancipative values previously being very high if the world has evolved as according to Welzel.[6] In any case, this, by construction, makes it unlikely that FE models will uncover the true relationship; almost half the sample turns increasingly "emancipative" and simultaneously experiences de-democratization. When throwing out the 20 over-democratic cases, *even an FE model* (A6) finds a positive significant (5%) relationship, despite only 16 countries remaining (and the short simulated time series). Hence, if the real world had looked like the theorized world described in I&W or Welzel, many panel models in D&K would likely have identified a values–democracy relationship.

Interpreting WIK's empirical tests and other issues

WIK's reply contains numerous other problematic points, including the interpretation of their empirical replication of D&K on extended data material. Before discussing this, we briefly note four other issues, with relevance for the credibility of the results and conclusions in D&K.[7]

First, directly testing WIK's fine-tuned theory concerning demand being "higher"/"lower" than supply is far more problematic than WIK realize, hinging, e.g., on arbitrary scaling properties of (non-comparable) values and democracy measures. For plausible distributions of initial supply/demand levels (cf. the many "over-democratic" countries in WIK's simulation exercise), a more robust empirical implication is that increased demand enhances (probability of increases in) supply, which is exactly what D&K's models test.

Second, while some of WIK's "conceptual criticism" of TSCS models, and their appropriateness for picking up transitions, seems to confuse deterministic and stochastic processes, we remind that D&K also did test dynamic probit models explicitly designed to capture transitions (and GMM models designed to capture slow-moving processes). These models yielded no evidence that values affect "tectonic" regime transitions.

Third, WIK note that their imputation model is superior to D&K's because it includes additional values survey data and excludes variables predicting democracy. The latter is a problematic argument, breaking with conventional advice on the construction of multiple imputation models (more predictors is better). We carefully evaluated the predictive power of our model, which performs very well, whereas WIK did not.

Fourth, referring to Achen and Clark's work WIK argue that our TSCS models' ability to add several controls "does nothing to improve a model" (WIK, p.3). This, however, represents a misreading of this work (and of D&K): If correctly specified, models including all relevant controls constitute an improvement and reduce bias. D&K only included controls highlighted as relevant in I&W.

---

[6] Welzel 2013.

[7] There are several additional issues. Some are mere details (e.g., we did not claim that emancipative values are "never" learned under autocracy (p.11)), whereas others are relevant for the choice of research design and interpretation of results (e.g., D&K do not "double-treat" omitted variable bias by including both democratic history variables and country-fixed effects, since the former actually have substantial within-country variation).

Finally, WIKs empirical analysis expands on D&K by adding WVS Wave 6 data. *If* the theory is correct, it should be easier to observe patterns when including more data/longer time series. WIK conclude that "[r]eplicating D&K's TSCS models with a larger set of countries disconfirms their findings" (p.13). However, this statement is inaccurate; the replication results produced by WIK, reported in their Appendix, actually corroborate D&K's findings.
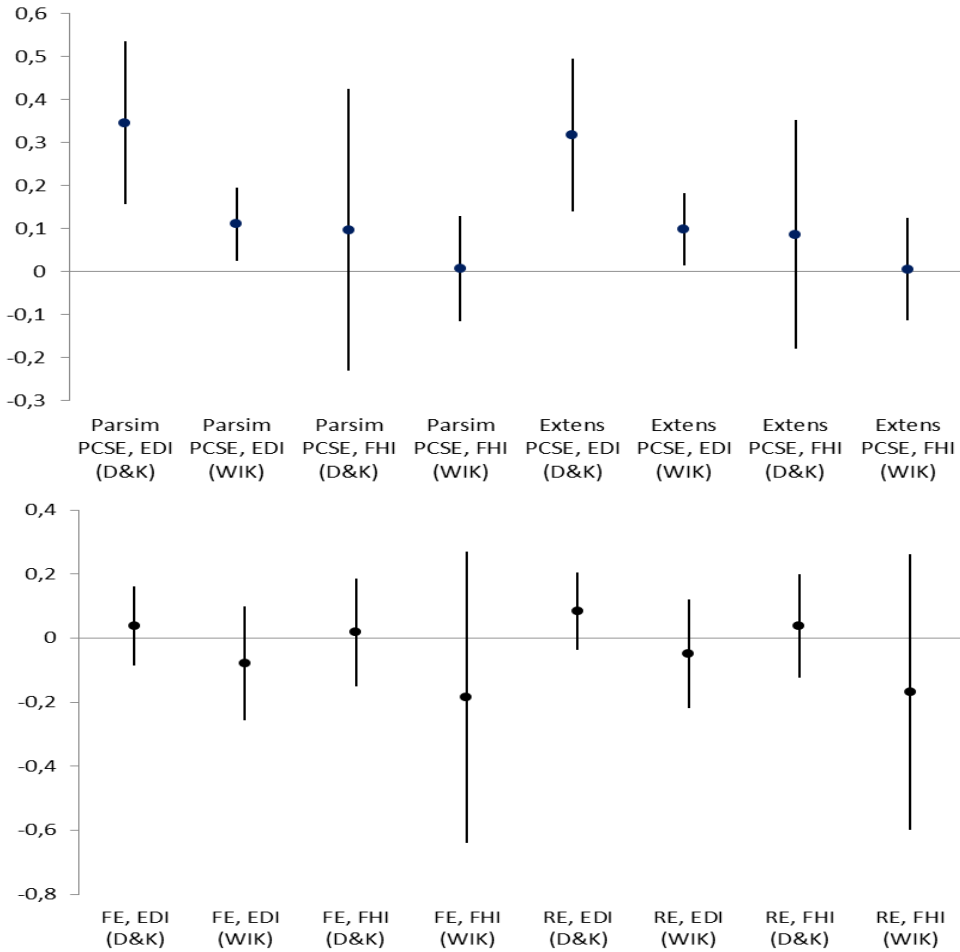

Figure 3: Comparing results from equivalent models (D&K/WIK)
Notes: Coefficients for values (self-expression/emancipative) on democracy measures, with 95% CIs, for WIK's Appendix Table A3 (left in pair), and corresponding models in D&K (A7-A8, Table 1; B1-B6, Table 2). D&K's FHI coefficients are scaled/transformed for direct comparison.

For transparency, Figure 3 displays estimates, with 95% CIs, from D&K's regressions side-by-side with all replication estimates from WIK's Appendix Table 3. WIK's results, based on more extensive data, actually yield lower point estimates for *all* models. Hence, the replication results should strengthen faith in D&K's conclusions, contrary to WIK's assertion.

Conclusion
Arguing against our empirical criticism of "Revised Modernization Theory", WIK discard standard TSCS/panel data models as inappropriate. If WIK's arguments are right, their own theory is true,

6

whereas much knowledge on other questions, e.g., on other structural causes of democratization such as inequality or education that have been investigated using such models, remains unfounded. However, WIK's claims falter under closer scrutiny. For example, their simulated world has many peculiar characteristics, and models resembling the ones we actually used in D&K detect the values—democracy relationship nonetheless. Further, WIK's empirical replication of D&K, if anything, cast even stronger doubts on the theorized values—democracy relationship.

In sum, only scholars with very clear (and unconventional), preferences over research designs, models and measures should accept WIK's assertion that "the evidence supports the emancipatory theory of democracy" (p.2).

References

Dahlum, Sirianne and Knutsen, Carl Henrik (Forthcoming). "Democracy by Demand? Reinvestigating the Effect of Self-expression Values on Political Regime Type." British Journal of Political Science.

Hadenius, Axel, and Jan Teorell. 2005. "Cultural and Economic Prerequisites of Democracy: Reassessing Recent Evidence." *Studies in Comparative International Development* 39(4): 87-106.

Inglehart, Ronald and Christian Welzel. 2005. *Modernization, cultural change and democracy. The Human Development Sequence*. New York: Cambridge University Press.

Leamer, Edward E. 1985. "Sensitivity Analyses Would Help." American Economic Review 57(3): 308-13.

Knutsen, Carl Henrik. 2010. "Measuring Effective Democracy." *International Political Science Review* 31(2): 109-128.

Teorell, Jan and Axel Hadenius. 2006. "Democracy without Democratic Values: A Rejoinder to Welzel and Inglehart. *Studies in Comparative International Development* 41(3): 95-111.

Welzel, Christian. 2013. *Freedom Rising: Human Empowerment and the Quest for Emancipation*. Cambridge: Cambridge University Press

Welzel, Christian, Inglehart, Ronald and Kruse, Stefan (Forthcoming). "Pitfalls in the Study of Democratization.Testing the Emancipatory Theory of Democracy." British Journal of Political Science.