

Notes and Comments

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

SIRIANNE DAHLUM AND CARL HENRIK KNUTSEN*

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 ‘Revised Modernization Theory’.¹ They claim that it is ‘irrelevant’² and ‘poses no challenge’³ to the theory, asserting that ‘the evidence supports the emancipatory theory of democracy as it did in [Inglehart and Welzel’s] original analyses’.⁴

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 which rests on very stark assumptions, such as the absence of unobserved country-specific effects on democracy. Moreover, their response when faced with specifications that do not corroborate 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 (for example, omitted variable bias)⁵ 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 too heavily 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 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’.⁶

* Department of Political Science, University of Oslo (emails: s.a.dahlum@stv.uio.no, c.h.knutzen@stv.uio.no).

¹ Dahlum and Knutsen 2016; Welzel, Inglehart, and Kruse 2016.

² Welzel, Inglehart, and Kruse 2016, 5.

³ Welzel, Inglehart, and Kruse 2016, 10.

⁴ Inglehart and Welzel 2005, 2.

⁵ Welzel, Inglehart and Kruse 2016.

⁶ Leamer 1985, 308.

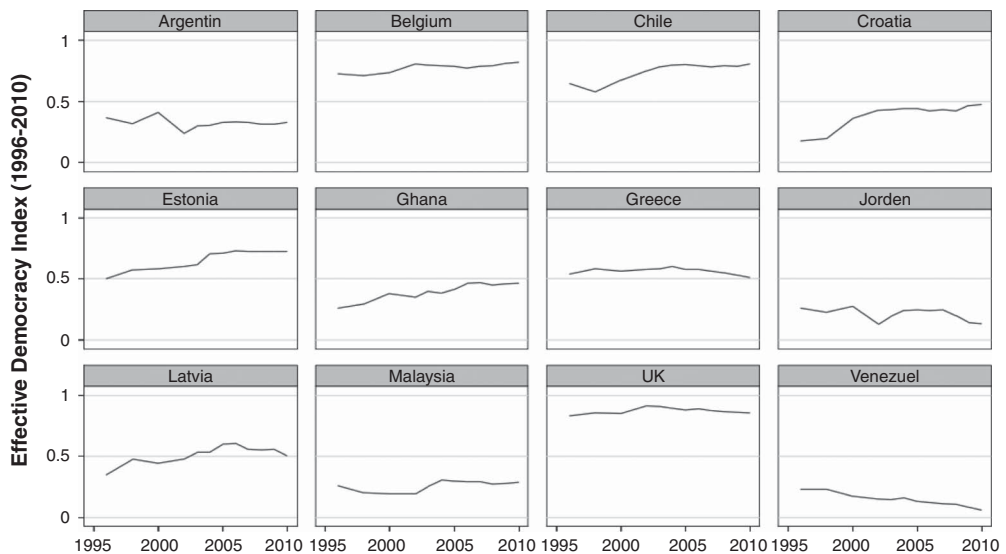


Fig. 1. EDI 1996–2010

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.⁷ 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, for example, 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 self-expression values on democracy if it is as strong as is theorized by I&W, and we are not persuaded to the contrary by WIK’s simulation exercise. There are several reasons for this.

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

Despite providing a ‘low-powered’ test (thirty-six hypothetical countries, twenty years), WIK’s inability to find effects in the hypothetical/simulated world where their theory applies would have been a concern 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, for example, assuming only one disruptive/monotonic change in EDI, which makes it harder to identify effects by construction. This, and other issues discussed below, implies that WIK’s 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, for example, on systematic and unsystematic measurement errors still stand).⁸ In fact, standard measures such

⁷ Welzel, Inglehart, and Kruse 2016, 2–3.

⁸ See Hadenius and Teorell 2005; Knutsen 2010; Teorell and Hadenius 2006.

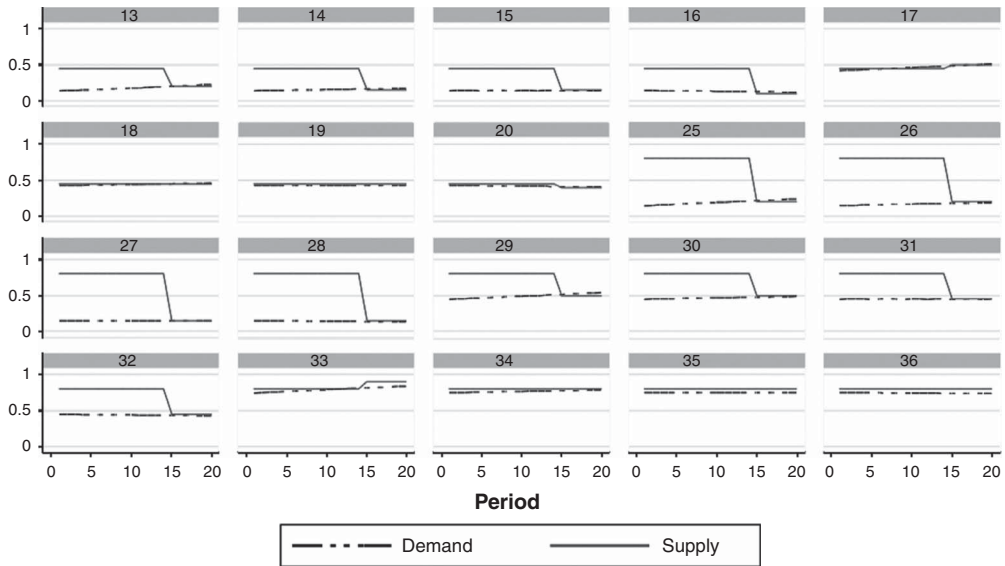


Fig. 2. Twenty (of thirty-six) simulated cases with democratic 'over-supply'

as the Democracy-Dictatorship measure (DD),⁹ Polity and Freedom House display real-world patterns that resemble 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, rather than temporal, comparisons *and* use the problematic EDI yield support.

Secondly, 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 (with panel-corrected standard errors (PCSE)), random effects (RE) and fixed effects (FE) models without lagged dependent variables (LDVs). Operating with LDVs can be problematic when the dependent variable exhibits as little temporal variation as it does 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 null-result in Model A1, Table 1. But when throwing out the problematic LDV (A2), thus following D&K's baseline models, the PCSE model does, in fact, pick up a positive significant (1 per cent) coefficient on demand.

Still, A2 draws heavily on cross-country variation, and we therefore tested panel data models in D&K. Indeed, an RE model (A3) also 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.

Thirdly, 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 – which matches poorly with real-world patterns *and* narratives in I&W and Welzel – is driving this result:¹⁰ twenty of WIK's thirty-six 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 twenty, sixteen have an *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 feature does not reflect that some countries are 'over-democratic' due to gradually falling 'emancipative values'.

It is unclear which real-world/historical patterns could have generated so many 'artificially high' democracy scores; importantly, they cannot come from emancipative values previously being very high if

⁹ See Przeworski et al. (2000).

¹⁰ Welzel 2013.

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 (0.92)	6.188*** (2.16)	7.450*** (0.49)	-21.154* (11.00)	6.116*** (0.33)	10.586** (4.83)
Lagged dependent variables (supply)	0.932** (0.09)					
Lagged development	0.033 (0.88)	-5.522* (2.31)	-6.734** (0.49)	20.849+ (10.42)	-5.166** (0.32)	-4.448 (5.23)
N	684	684	684	684	304	304

Note: 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. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the world has evolved as Welzel describes.¹¹ In any case, this, by construction, makes it unlikely that FE models will uncover the true relationship; almost half of the sample turns increasingly 'emancipative' and simultaneously experiences de-democratization. When throwing out the twenty over-democratic cases, *even an FE model* (A6) finds a positive significant (5 per cent) relationship, despite only sixteen 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 that are relevant for the credibility of the results and conclusions in D&K.¹²

First, directly testing WIK's fine-tuned theory concerning demand being higher/lower than supply is far more problematic than WIK realize, hinging, for example, 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 (the probability of increases in) supply, which is exactly what D&K's models test.

Secondly, 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 reiterate that D&K also tested dynamic probit models explicitly designed to capture transitions (and Generalized Method of Moments (GMM) models designed to capture slow-moving processes). These models yielded no evidence that values affect 'tectonic' regime transitions.

Thirdly, WIK note that their imputation model is superior to D&K's because it includes some additional values survey data and excludes variables that predict democracy. The latter is a problematic argument, breaking with conventional advice on the construction of multiple imputation models (including more

¹¹ Welzel 2013.

¹² There are several additional issues. Some are mere details (e.g., we did not argue that emancipative values are 'never' learned under autocracy (Welzel, Inglehart, and Kruse 2016, 8)), 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 turn out to have substantial within-country variation empirically).

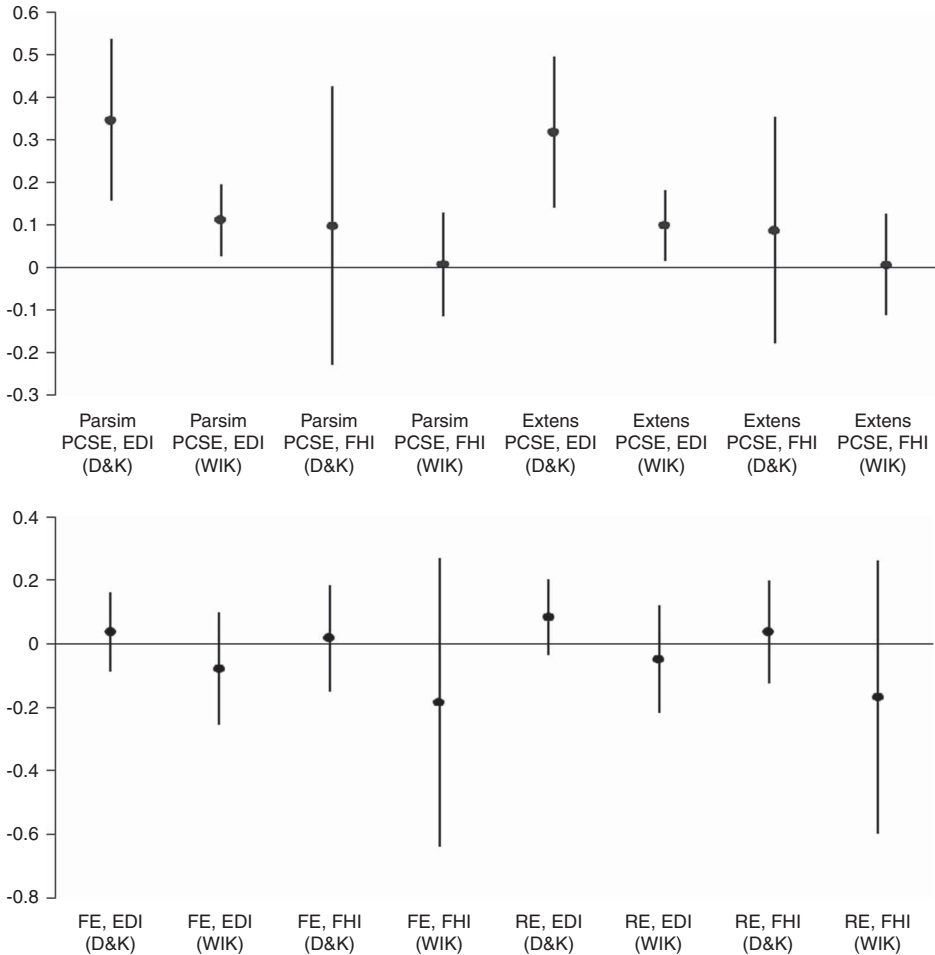


Fig. 3. Comparing results from equivalent models (D&K/WIK)

Note: Coefficients for values (self-expression/emancipative) on democracy measures, with 95 per cent 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.

predictors is better). We carefully evaluated the predictive power of our model, which performs very well, whereas WIK did not.

Fourthly, referring to Achen and Clark's work, WIK argue that our TSCS models' ability to add several controls 'does nothing to improve a model'.¹³ 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 that were highlighted as relevant in I&W.

Finally, WIK's empirical analysis expands on D&K by adding World Values Survey 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'.¹⁴ However, this statement is inaccurate; the replication results produced by WIK, reported in their Appendix, actually corroborate D&K's findings.

¹³ Welzel, Inglehart, and Kruse 2016, 2.

¹⁴ Welzel, Inglehart, and Kruse 2016, 10.

For transparency, Figure 3 displays estimates (with 95 per cent confidence intervals), from D&K's regressions alongside 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, whereas much knowledge on other questions – for example, 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 used in D&K nonetheless detect the values–democracy relationship. Further, WIK's empirical replication of D&K, if anything, casts 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'.¹⁵

REFERENCES

- Dahlum, Sirianne, and Carl Henrik Knutsen. 2016. Democracy by Demand? Reinvestigating the Effect of Self-Expression Values on Political Regime Type. *British Journal of Political Science* 46: 437–61.
- 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.
- Knutsen, Carl Henrik. 2010. Measuring Effective Democracy. *International Political Science Review* 31 (2):109–28.
- Leamer, Edward E. 1985. Sensitivity Analyses Would Help. *American Economic Review* 75 (3):308–13.
- Przeworski, Adam, Michael Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950–1990*. Cambridge: Cambridge University Press.
- 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, Ronald Inglehart, and Stefan Kruse. 2016. Pitfalls in the Study of Democratization. Testing the Emancipatory Theory of Democracy. *British Journal of Political Science* 46: 463–72.

¹⁵ Welzel, Inglehart, and Kruse 2016, 2.