Democracy, Public Support, and Measurement Uncertainty Memo to Reviewers

Again, this is things to do only. Remember to go back to the reviews and mine them for compliments when writing the memo. Temporarily(?) listed by reviewer; reordering later to match the sequence of the article will likely be more effective, given the overlap and conversation among reviewers.

Uncertainty in DV and Control Variables

Reviewer #1

Issue 16 First, I believe that the measurement error of variables included in the model should be factored in. Many of these variables (democracy, corruption, economic output) cannot be directly observed and are therefore prone to measurement error. If one believes that (a) estimating the effect of one variable on another requires incorporating measurement error; and (b) control variable Z needs to be included because it likely has an effect on both X and Y, then to accurately adjust for the effects of Z requires that X, Y and Z all have included measurement error. ## Reviewer #2 Issue 16 Despite the MS's assertion of the "absolute necessity" of accounting for measurement error, the MS does not do so for any of the other variables in the analysis. Given that Treier and Jackman's analysis focused on uncertainty in measures of democracy, the most obvious omission is the V-DEM democracy measure itself (which, according to the V-DEM documentation, does appear to have associated uncertainty estimates).

Response

A major concern raised by Reviewers 1 & 2 was regard to the omission of measurement errors of DV and control variables in our analysis. It is a good point. Many variables in the analysis cannot be observed directly and inherently contaminated by measurement errors. To address this concern, since Claassen used V-Dem Version 8 in his analyses, we incorporated measurement error for V-Dem democracy index, electoral index, polyarchy index, and liberal index, by using the posterior distribution of Version 8 provided by V-Dem. Regarding the data used to replicate Claassen' analyses, we simulated 900 data set for democracy index from V-Dem Version 10 which did not

provide available posterior distribution yet but standard error. To account for measurement error in Corruption Perceptions Index, we simulated 900 draws from a distribution with the point estimate and standard error which provided by CPI for the year from 2012 to 2018. For the year before 2012, we estimated standard error by using the multiplication between country's maximum relative error (standard error/cpi index) and CPI index. We rerun the models following the same steps in our previous analysis but incorporating uncertainty in democracy index, democratic support, and corruption perception index. The relationship between public support and democracy development disappears when uncertainty is accounted for in models either using Claassen's original data or expanded data with 25% more country-year data. The updated results are presented in Figure XXX at Page xx.

Alternative Method for Incorporating uncertainty

Reviewer #1

Issue 17 My primary concern is that the method by which measurement error is included / added is not sufficiently explicated and demonstrated to be valid (e.g., using Monte Carlo experiments), especially in the time-series, cross-sectional context which is of interest. The authors cite several applied articles - none of which appears to be methodological in nature - as well as a classic book by Rubin on multiple imputation in surveys - which is a distinct topic. This is unfortunate because I have several concerns with the method as used and the results which are produced, as described below.

Issue 23 Alternatively, the authors might use a different method for factoring in the error due to measurement: structural equation models, which, in essence, combine a measurement model and the subsequent regression model in one step. In contrast to the method proposed in the present paper, SEMs benefits from a substantial methodological literature (e.g., Bollen & Noble 2011; Skrondal & Rabe-Hesketh 2004). Indeed, one of the articles cited by the current paper (Juhl 2019) adopts exactly this approach. Bollen & Noble. 2011. Structural Equation Models and The Quantification of Behavior. Proceedings of the National Academy of Sciences 108: 15639-15646; Skrondal & Rabe-Hesketh. 2004. Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. Taylor & Francis.

Reviewer #2

Issue 17 The MS deals with measurement error by treating mismeasured variables as if they were fully imputed, and following Rubin's standard multiple imputation rules. If I understand correctly, this is what Blackwell, Honaker, and King (2017) call "multiple overimputation" (MO)—a fact that should probably be noted in the MS, given that Blackwell et al. provide the most comprehensive exposition of this method that I am aware of. ... More can be done, however, to show that this critique itself is robust. First, the MS should provide more detail on the theoretical justification for MO and the conditions under which it performs well. How does it perform when measurement error is strongly correlated across variables, as is likely to be the case in the models with lagged DVs? Issue 23 It would be useful to show that the critique still holds with an alternative measurement-error approach, such as Treier and Jackman's (2008) method of composition. ## Reviewer #4 Issue 17 We need more information on (1) how uncertainty is integrated into the models of Claassen, and (2) the DCPO models.

The Change of Democratic Support and Lagged Support

Reviewer #1

Issue 24 Second, the results in Figure 2, based on models using democratic support as a dependent variable, are puzzling. They appear to show that democratic support is entirely disconnected from such prominent features of the political environment as the level of democracy and economic development. Not only are these effects insignificant, but their point estimates are almost exactly zero in magnitude. The effects of lagged democratic support are also zero, which is even more implausible because it suggests that democratic support in one year is entirely disconnected from its levels in previous years. In other words, Denmark is as likely to have a low value next year as China is likely to have a high value. This stands in contrast with much of existing research on democratic support and political culture.

Reviewer #2

Issue 24 Also, I could not help but be struck by the extreme attenuation of the coefficient estimates in Figure 2, which almost suggest that the variables were generated from independent distributions.

In particular, can it really be the case that lagged support has no conditional association with current support? If so, then I think some explanation is required, or else readers will think that some mistake has been made.

Reviewer #4

Issue 15 Second, I was a bit confused by Figures 1 and 2. My understanding is that accounting for uncertainty should only affect the standard errors, but in the figures it also affects the point estimates. This is the case even in the models 'Claassen W/Uncertainty.'

Test Full Models

Reviewer #1

Issue 22 I have two additional concerns with the paper. First, the authors have not replicated all of Claassen's main results. Claassen 2020a argues that within-country analyses are desirable and uses generalized method of moments (GMM) estimators to obtain these. Claassen (2020b) uses first difference models; he also focuses on the separate effects of the electoral and liberal components of democracy. I suggest that the authors replicate all the main results from Claassen (2020a; 2020b), rather than an ad hoc subset of these.

Justification of DCPO

Reviewer #1

Issue 18 Second, in addition to the results obtained from the latent variable model developed and used by Claassen, the authors present the results obtained from an entirely different latent variable model. Although intriguing, no evidence at all is presented to support the claim (repeated throughout) that this method is "superior" to Claassen's (reference is made to an unpublished working paper). Indeed, it would appear from the short summary of this model that it adds additional parameters to Claassen's model. Yet more complicated models may overfit the data and are not necessarily more accurate. I propose that evidence supporting the superiority of this method be supplied, or alternatively, the authors just focus on replicating Claassen's models. ## Reviewer #2 Issue 20 Finally, I am concerned about the over-time smoothing induced by the random-walk

4

prior in measurement model. In county-years without survey data, this prior will impute the latent variable by interpolating between adjacent years, and in years with data it may still make over-time changes smoother than the raw data suggest. From a descriptive perspective this is largely a virtue, but in models that rely for causal identification on sharp over-time shifts, such as the thermostatic model, it can be a big problem. There is no magic fix for this problem, but readers should still be given more information on how much "work" the prior is doing—what proportion of country-years are missing data, how informative is the prior relative to the likelihood, and so on. ## Reviewer #4 Issue 18 Third, the author should provide more information on the differences between the models employed by Claassen and those that she/he employs (perhaps in an appendix).

Thanks to editor's suggesions, we do agree that we should present and assess the effect of intermediate modification in an incremental way. To do this, in the revised manuscript, we replicated Claassen's measure for public democratic support by applying his measurement Model5 to his original data file and used the posterior distribution of the public support estimates to incorporated uncertainty. Then, we added more survey data to improve the quality of source data and applied Model5 to this expanded data set, still keeping the full posterior distribution of estimates. In this way, we isolate the effects of measurement uncertainty and more data. In the APSR analyses, it is uncertainty that accounts for the disappearing significance of public support.

Substantive Discussion

Reviewer #4

Issue 19 First, I found the 'Discussion' section somewhat underwhelming. The discussion is mainly about data issues that could explain the lack of a relationship, not the substantive question. If that is the case, the letter makes a negligible contribution because it only finds that the findings of two specific studies are not robust (not that there is no relationship between public support for democracy and regime survival). I would like the author to engage a bit more in substantive debates.

Compliement

Reviewer #3

Issue 19 I, for my part, however, have been left in disbelief by this mechanistic model of the link between public opinion and regime qualities. Although I am quite confident that policy preferences and policy change react thermostatically to each other in yearly rhythms, I find it hard to believe that such fundamental things as regime preferences and regime qualities underly such short-term cycles. Regime preferences and regime qualities are more enduring, inert and their changes are too glacial for a thermostatic model of short-term cycles to *plausibly capture the underlying dynamic in the co-evolution of public opinion and regime qualities. Since Claassen uses V-Dem data to measure democratic regime qualities, which are entirely expert judgements, his models demonstrate at best that expert and lay assessments of democracy react thermostatically to each other.*

Issue 19 Another reason why I doubt Claassen's findings is evidence showing that the same level of support for democracy hides over firmly encultured differences in how people in different countries understand democracy (Kirsch & Welzel 2018), which often leads people to mis-estimate their own countries' democraticness (Kruse, Ravlik & Welzel 2018). Hence, levels of support for democracy are strictly speaking incomparable across culturally diverse sets of societies. Also relevant in this context are findings showing that what matters for democratic regime stability and change is not how much people say that they support democracy but what values motivate them to do so. Specifically, mass support for democracy operates in favor of democracy only in conjunction with emancipative values but not in disconnection from these values (Brunkert, Kruse & Welzel 2018). Hence, to establish the regime-relevance of public opinion, looking merely at levels of democratic mass support is misleading. Instead, it is more promising to use value priorities to distinguish different types of democracy supporters and estimate their demographic distribution.

Kirsch, H. & C. Welzel (2018). "Democracy Misunderstood: Authoritarian Notions of Democracy around the Globe." Social Forces 91: 1-33 (DOI: 10.1093/sf/soy114).

Kruse, S., M. Ravlik & C. Welzel (2018). "Democracy Confused: When People Mistake the Absence of Democracy for Its Presence." Journal of Cross-Cultural Psychology 49: 1-21 (DOI: 10.1177/0022022118821437).

Brunkert, L., S. Kruse & C. Welzel (2018). "A Tale of Culture-bound Regime Evolution:

The Centennial Democratic Trend and Its Recent Reversal." Democratization 25: 1-23 (DOI: 10.1080/13510347.2018.1542430).

Editor

Both Reviewers 1 and 2 want more evidence, explanation, and/or justification of the alternative method used to incorporate measurement error. Reviewer 1 seems to specifically suggest a simulation study. If such a study already addresses the performance of the method (either analytically or with simulations) in a context with temporally and serially correlated errors, it would make sense to reference that work in the main text. Reviewer 2 also asks for clarification and/or justification of the imputation and smoothing processes used in the analysis. Reviewers 1 & 2 also ask why measurement error in other variables are not included in the analysis. Reviewers 1 & 2 also ask why lagged support is unrelated to current support.

Reviewers 1, 2, and 4 all ask why the results for Claassen's model but accounting for measurement error (second set of results) also yield different point estimates. The revised version should address this directly, including by better explaining the method(s) used. If the different in point estimates is due to multiple deviations from Claassen's approach (e.g., in both the smoothing *and* the use of the errors), I would suggest introducing and presenting each intermediate modification of the research design incrementally to avoid confusion and help readers assess how much of the change in results is due to the incorporation of measurement error vs. other modelling or measurement decisions. Analytical transparency is key.

Evidence to address these various methodological questions and issues issues may be included in the response to reviewers and/or the Supplemental Information (if it would help readers appreciate the methods used) as needed. For instance, I could imagine some readers benefitting from having the estimating equations included in the Supplemental Information in order to appreciate how measurement error is incorporated into the estimates as well as more information about the levels of measurement used (mentioned as an important difference by Reviewer 3). While Reviewer 1 asks for all results from the original two studies be replicated as well, these additional replications need only be included in the paper or SI to the extent that they are relevant to the relationship between support and democracy at the core of this Letter's focus.

As you expand the SI, please keep in mind our limit of 25 pages. Material solely for editors and

reviewers that would not be relevant for readers can remain in the response to reviewers.

Finally, Reviewers 2, 3, and 4 offer various suggestions for how to more effectively frame or discuss the substantive or theoretical (rather than measurement or methodological) implications of your contribution. We hope their feedback is useful as you revise your Letter.