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 reanalysis. It is a good and reasonable suggestion.

Many variables, dependent variables, key independent variables, and control variables in the original analysis, are latent variables, which cannot be observed directly and are inherently contaminated by measurement errors. To address this concern and justify the necessity of incorporating

measurement uncertainty in latent variables, we take uncertainty into account for five variables, public support for democracy, liberal democracy index, electoral democracy index, liberal component index, and corruption perception index, in the revised version.

In respect of the key independent variable, public support for democracy, we used Claassen's Model 5 and recollected original source data to replicate the measure of public support and keep its posterior distribution. For expanded data, still, we used Model 5 and expanded source data to measure public support and preserve its posterior distribution. In terms of three indices of democratic development, we incorporated measurement error using their posterior distribution from V-Dem Version 8 which covers 201 countries from 1789 to 2017 and is the version Claassen used in his analyses. The posterior distributions are provided by V-Dem Project Team. Regarding the expanded data used to replicate Claassen's analyses, we use the same three democratic indices from V-Dem Version 10 which covers 202 countries from 1789 to 2019. However, since V-Dem project has not released posterior distributions for indices in version 10, we simulated 900 data set for democracy indices using their 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 provided by CPI for the year from 2012 to 2018. However, since CPI does not provide standard error for the years before 2012, we first estimated the standard error for these years by using the product of countries' maximum relative error (standard error/cpi index) and their CPI index and then used the estimated standard error to simulate 900 draws.

We used the posterior distributions of five variables in method of composition to address measurement errors. By incorporating uncertainty in democracy indices, democratic support, and corruption perception index, the relationships between public support and democratic development and between the change of public support and the change of democracy disappear either using Claassen's original data or expanded data with one-third more source data. The results from models including measurement error for outcome, key independent, and control variables again justify the importance of incorporating uncertainty in analyses.

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.

Response

Reviewers 1 & 2 & d concern the validity of the approach used to incorporate measurement uncertainty. We appreciate Reviewers' suggestion on method of composition (**?**) (MOC) to take into account measurement error. After reviewing the logic of MOC and its application in articles with latent variables (**???**), we believe that MOC is the approach that has been justified theoretically and methodologically to address measurement error in latent variables. We follow reviewers' suggestion and employ MOC in our revised version.

The idea of MOC is to account for uncertainty in both latent variables and effects of latent variables and other variables on outcome variable through drawing samples from the joint posterior distribution of the measurement and analysis model. Although MOC is known to methodologists, to approach a broader audiance and make MOC more accessible, we present each step in employing MOC in our reanalysis and provide R script for faciliating reproducibility and replication. We followed MOC precedure and the results from MOC demonstrated that when uncertainty is allowed, the coefficients on public support and the change of democratic level are no longer distinguishable from zero on Figure 1 page 8 and Figure 2 page 9. The null results hold in robust check models and models with more control variables in Figure xx Page xx in Online Supplementary material.

We appreciate reviewers' suggestions on MOC, and the results from MOC not only justify but also provide us more confidence in our conclusion that the importance of accounting for uncertainty.

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.'

Response

Thanks to reviewers' feedback on Figure 2, we first need to clarify that the DV in Figure 2 is the *change* of public support for democracy rather than current public support. Claassen examined the thermostatic effect of the change of democratic level on the *change* of public support in his original APSR analysis. In our replication of the APSR article, the DV is also the *change* of public support for democracy. Therefore, the question is that is there any theoretical reasons to suppor a significant relationship between the current *change* of pubic support and the lagged public support.

Claassen's original results indicate the *change* of public support in current year is positively changed by the public support in the previous year and negatively shifted by the previous two years. For example, Claassen's results indicate that a high public support in 2020 leads to a bigger *change* (either upturn or downturn) in mood in 2021 but a high support in 2019 leads to a smaller *change* (either upturn or downturn) in mood in 2021. However, we do not have grounded theories

to expect that public support can produce its positive demand in one year but backsliding demands in two years. In contrast, our null results of lagged support on the *change* of support mood indicate that the *change* of moods might not be dependent on its previous situation.

Theoretically, we should expect the change of public support is more likely to be shaped by exogenous factors, especially a shock in democratic or economic environment. However, in terms of the effects of social and economic environment on the change in mood, even in Claassen's original results (the uppermost results in Figure @ref(fig:plot-mainAPSR)), point estimates of the change in democracy, economic growth, corruption index as well as their lagged level, are close to zero in magnitude. Except for the change in democracy, all other estimated effects are not distinguishable from zero. After acknowledging uncertainty, the estimated effect of the change in democracy is indistinguishable from zero, and the magnitude of the coefficient moves further to zero.

Methodologically, given the measurement errors in public support, democratic indices, and corruption index, it should not be surprising to see the diminution of the estimated effects. Although it is rare that accounting of uncertainty disappears the significant effect of democracy, **?**, p213 emphasized, after allowing uncertainty, the significant effect of democracy holds mainly in applications where democracy strongly affects the quantity of interest and models are simple enough. Our results demonstrate that there at least is no a strong relationship between democracy and public support. Theoretically, as Reviewer3 pointed out, democratic value is formed and shaped by institution, history, and culture, and the nature of democratic value determines that it cannot be impacted swiftly by the change of democratic level. Moreover, the understanding of democracy which shapes democratic value has cultural bounds (**???**) but not yet incorporated in most current measure of democratic support.

Thanks to reviewer2's comments on random walk, we do think we should discuss more on problems caused by random walk, which actually provides us another powerful explanation for our results. In order to isolate the effect of different priors on estimated measures, in revised version, we only employed Claassen's Model5 with original data, recollected data and expanded data, separately.

By using the same measurement model, we exclude the effects which might be brought by different priors and a bigger size of observed data should reduce the possible dominance of priors.

Since our expanded data added 278 country-year, the estimates from the expanded data should

be influenced less by priors than original data by inputing more data. However, even though we put much effort to collect as much data as possible. At aggregated level, country-year data is super sparse.

However, if we have observations for every year in each country surveyed, the number would be 4,640, and a complete set of country-year-items would encompass 283,040 observations. In fact, even collecting as many available national and cross-national data as possible, the current source data has 1,443 country-years and a total of 4,717 country-year-item observations. The spareness is reflected from the complete-data perspective but also from the available data in the country-years span. Although we have the 2,746 country-years, only 53% of it has available information.

The sparseness of available data makes it possible that the changes of public mood are "smooth" out by modeling the temporal evolution in a simple local-level dynamic linear, which is also referred to as the random walk plus noise model. With random noise, the smoothed time series data might fail to capture a sudden change which is caused by exogenous factors. Given the sparsed available data and effects from random walk prior, it is highly possible that we cannot detect any significant effect from institution changes and other economic and social factors. Futhermore, scholars should be careful of using estimated latent variable as a dependent variable when source data is sparsed. We are planning to carry out more work on figuring out at what rate the sparseness of source data will post severe problems in statistical inference, which will faciliate scholars' usage of latent variable measures.

In terms of the impact of measurement error on estimated effects, as we illustrated in Appendix, point estimates of coefficients are calculated from samples drawn from joint distribution of estimated latent variable and coefficients, which incorporate both uncertainty in latent variable measurements and the relationship between outcome variables and covariates. Given the big uncertainty in two parts, we should expect the estimated effects are different from point estimates from models without accounting uncertainty. Many studies have already shown that when taking measurement uncertainty into account, the magnitude of coefficients of latent variables can be either attenuated or enlarged substantively (e.g., **?**, p. 211; **?**, p. 254; **?**, p. 6).

7

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.

Response

As the editor suggested, we should focus on our main point in this letter, that is, reanalyzing the relationship between democracy and public support. We replicate additional models which examine this main relationship.

We replicate **?**'s results using generalized method of moments (GMM) estimators for both pooled data and split data (democracy only and authoritarian country only). The results hold and presented in Figure @ref(fig:plot-mocAJPS). We also replicate **?**'s results using First Different Models, including corruption index, and examing the separate effects of the electoral and liberal components of democracy. Still, the results hold in these robustness check models and specifications in Figure @ref(fig:plot-mocAPSR).

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 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).

Response

Thanks to editor's suggestion, we agree that we should present and assess the effect of intermediate modification in an incremental way, holding analytical transparency. To do this, in the revised manuscript, we focus on replication by only employing Claassen's measurement Model 5 to measure public support for democracy.

We apply Model 5 to Claassen's original data file, recollected data based on his provided data source and expanded data. In this way, the only two changed things are acccounting for uncertainty and adding more data. We incorporate uncertainty when using Model 5 estimated public support from recollected data to examine the effect of uncertainty. Then, we apply Model5 to the expanded data set which improves source data quality and should reduce the uncertainty.
In this way, we eliminate potential effects caused by different measurement models and make efforts to reduce the uncertainty caused by fragmented data. Our results show that it is the incorporation of uncertainty that accounts for the differences between Claassen's original analyses and our reanalyses.

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.

Response

Although we do acknowledge uncertainty in extant measurements and we can not unambiguously make a statement that there is no relationship between public support and regime survival, the relationship, if exists, is probably negligible, because the incorporation of measurement error swamping the significant effects of democracy in a strong relationship is rare (**?**).

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.

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.

Response

We did not do a simulation study. Instead, we employ MOC suggested by Review 1 and 2 in this revised version. We illustrate MOC process in detail in SI. Measurement errors in other variables, including three democracy indices, public support and corruption perception index, have been included in our reexaminations. As for the concerns in Fig2, the DV is not current public support but the change of public support. We elaborate on why our null results make more sense than Claassen's original results.

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.

Response

We appreciate editor's suggestion on presenting each intermediate modificantion. Following this suggestion, we only focus on uncertainty in this revised version. It is a misunderstanding that incorporating uncertainty only changes standard error. First, incorporating uncertainty might or might not cause the changes of standard error. For example, standard errors did not change after accounting for uncertainty in (**?**) became smaller in (**?**) and bigger in (**?**). Second, the propagation of uncertainty can either attenuate or magnify the magnitude of estimated effects (see **???**).

Editor

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.

Response

In SI, we discussed methodology used in this article to address measurement errors.

Editor

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.

Response