Supplementary Materials for Including Measurement Uncertainty in Time-Series, Cross-Sectional Analyses: The Case of Mood and Democracy

Contents

List of Figures

List of Tables

1. Modeling Approach

1.1. The dynamic latent variable model

Claassen (2019) develops six versions of a dynamic latent variable model for estimating cross-national public opinion, recommending the fifth and sixth of these. While both perform well, the fifth is simpler and is used to measure support for democracy in Claassen (2020*a*) and Claassen (2020*b*). It employs a beta-binomial link function between observed, nationally-aggregated survey responses y_{ikt} for each country i, year t, and survey questions k, and the probabilities of offering a supportive response π_{ikt}. These probabilities are modeled as a function of the latent country-year opinion estimates of interest θ_{it}, item bias parameters λ_k and item-country bias parameters δ_{ik}. The latent estimates are furthermore modeled as evolving over time via a random walk process:

$$y_{ikt} \sim \text{Beta-Binomial}(s_{ikt},\ \pi_{ikt},\ \phi) \tag{1}$$

$$\pi_{ikt} = \text{logit}^{-1}(\delta_{ik} + \lambda_k + \theta_{it}) \tag{2}$$

$$\theta_{it} \sim \text{N}(\theta_{i,t-1},\ \sigma^2_\theta) \tag{3}$$

Claassen's (2019) more complex sixth model is used in the present paper. It adds item slopes or discrimination parameters γ_k to the above model. Specifically, equation (2) above is expanded as follows:

$$\pi_{ikt} = \text{logit}^{-1}(\delta_{ik} + \lambda_k + \gamma_k\theta_{it}) \tag{4}$$

These item slopes allow analysts to test if survey questions fit the single dimension of opinion that is assumed to underlie the observed survey responses. It is therefore recommended by Claassen (2019) for general use and is incorporated into the unified model developed below.

1.2. Structural models of mood and democracy

Claassen (2020*a*) models democracy d in any given country i and year t as a function of its previous two lags, lagged mood m, and several covariates X. The parameter δ_1 captures the effect of mood on subsequent democracy, i.e., the Lipset hypothesis.

$$d_{it} = \mu^{(d)} + \zeta^{(d)}_1 d_{it-1} + \zeta^{(d)}_2 d_{it-2} + \delta_1 m_{it-1} + X^{(m)}_{it-1}\text{B}^{(d)} + \epsilon^{(d)}_{it} \tag{5}$$

Claassen (2020*b*) then models mood m in any given country-year as a function of its previous two lags, the first lag and first difference of democracy, and other covariates. The parameter δ_2 captures the effect of democracy on subsequent mood, i.e., the socialization hypothesis, while δ_3 captures the immediate effect of change in democracy on mood, i.e., the thermostatic hypothesis.

$$m_{it} = \mu^{(m)} + \zeta^{(m)}_1 m_{it-1} + \zeta^{(m)}_2 m_{it-2} + \delta_2 d_{it-1} + \delta_3 \Delta d_{it} + X^{(m)}_{it-1}\text{B}^{(m)} + \epsilon^{(m)}_{it} \tag{6}$$

1.3. Incorporating measurement uncertainty via the method of composition

Tai, Hu, and Solt (2022) adopt a two-step approach to including measurement uncertainty, described initially by Tanner (1996) as the "method of composition" and introduced to political science by Treier and Jackman (2008). In a situation with a country and year-varying latent variable θ_{it}, the analyst takes a number of draws p from the posterior distribution of their latent estimates

$\tilde{\theta}_{pit}$. These draws, rather than than a single set of point estimates, are then employed in subsequent analyses. We illustrate with the following model of mood, which incorporates draws from the posterior distributions of democracy $\tilde{\theta}^{(d)}$ and mood $\tilde{\theta}^{(m)}$:

$$\tilde{\theta}^{(m)}_{pit} = \mu^{(m)} + \zeta^{(m)}_1 \tilde{\theta}^{(m)}_{pit-1} + \zeta^{(m)}_2 \tilde{\theta}^{(m)}_{pit-2} + \delta_2 \tilde{\theta}^{(d)}_{it-1} + \delta_3 \Delta\tilde{\theta}^{(d)}_{it} + X_{it-1}\mathrm{B}^{(m)} + \epsilon^{(m)}_{it} \tag{7}$$

Such a model is fit p times, once for each posterior draw of mood and democracy. This captures measurement uncertainty. Parameter uncertainty is then included by simulating a draw from the j parameter by p draw matrix of coefficient estimates D and the p-length array of j by j by variance-covariance matrices Σ (which may be robust or conventional variance-covariance matrices). Point estimates, standard errors, and confidence intervals can be extracted from the resulting vector of draws of estimates for each parameter: the mean of the vector can be used as the point estimate, with the standard deviation providing an estimate of the standard error.

1.4. The unified model

An alternative approach to handling measurement error, proposed in the present paper, is to incorporate both the measurement and structural models in a single likelihood function. The analyst jointly estimates latent variables and the structural links between latent and observed variables. Such a joint, or unified model has appeared in a number of guises (e.g., Kellstedt, McAvoy, and Stimson 1993/94; Skrondal and Rabe-Hesketh 2004).

We describe our unified model below. The model begins in the same fashion as the dynamic latent variable model presented above. However the third line now integrates the structural model of mood (equation 6 above) and the dynamic model of latent opinion (equation 3 above):

$$y_{ikt} \sim \text{Beta-Binomial}(s_{ikt},\ \pi_{ikt},\ \phi) \tag{8}$$

$$\pi_{ikt} = \text{logit}^{-1}(\delta_{ik} + \lambda_k + \gamma_k \theta^{(m)}_{it}) \tag{9}$$

$$\theta^{(m)}_{it} = \mu^{(m)} + \zeta^{(m)}_1 \theta^{(m)}_{it-1} + \zeta^{(m)}_2 \theta^{(m)}_{it-2} + \delta_2 \theta^{(d)}_{it} + \delta_3 \Delta\theta^{(d)}_{it} + X^{(m)}_{it-1}\mathrm{B}^{(m)} + \epsilon^{(m)}_{it} \tag{10}$$

Democracy has a simpler measurement model because, with dozens of indicators available in every country-year, the degree of measurement error is lower than is the case for democratic mood.[1] In addition, annual point estimates and standard deviations are available from V-Dem for every country. Observed democracy scores d^{obs} are treated as a function of an unobserved, "true" democracy score $\theta^{(d)}$, with the degree of error measured using the observed standard deviations in annual democracy scores d^{sd}. These latent estimates, rather than the observed V-Dem scores, are then used in the structural model of democracy:

$$d^{obs}_{it} \sim N(\theta^{(d)}_{it}, d^{sd}_{it}) \tag{11}$$

$$\theta^{(d)}_{it} = \mu^{(d)} + \zeta^{(d)}_1 \theta^{(d)}_{it-1} + \zeta^{(d)}_2 \theta^{(d)}_{it-2} + \delta_1 \theta^{(m)}_{it-1} + X^{(d)}_{it-1}\mathrm{B}^{(d)} + \epsilon^{(d)}_{it} \tag{12}$$

All of these steps, equations 8 through 12, are estimated simultaneously in the unified model.

[1]Two percent of the total variance in liberal democracy (point estimates plus measurement error) is due to measurement error, compared with 17% of the total variance in democratic mood.

2. Monte Carlo tests of the three approaches

I test the accuracy of the three approaches using a Monte Carlo study. A data-generating process is set up that captures the main features of the case of interest, i.e., two variables varying across cross-sectional and temporal units, with one of these being directly observed and the other observed only via a fragmented and partial set of indicators.

2.1. Method

The accuracy of our three approaches is examined in the context of a latent variable that varies across 30 temporal and 50 cross-sectional units ($T = 30$ and $N = 50$). We consider three scenarios, running a Monte Carlo study for each:

1. the latent variable is exogenous, exerting a direct effect on a second, observed variable, but with no reciprocal effect;
2. the latent variable is endogenous to the observed variable, with the latter being exogenous;
3. both latent and observed variables are endogenous, with each exerting effects on the other

These situations are designed to approximate the democracy-support nexus, with the most complex third scenario allowing a full set of reciprocal effects assuming to hold in the real world. The other two scenarios provide simpler DGPs which allow us to identify under what scenarios the three approaches diverge and/or converge on the truth.

In all scenarios, the latent variable is treated as unobserved. What is observed is three indicators of the true latent variable. Each of these takes a TSCS form, but each is sparsely distributed across temporal and spatial units. We assume that only 20% of all temporal and spatial units are have observed values for each indicator, with the remaining 80% being missing. This approximates the survey measures which are available for support for democracy, which are scattered across time and space in the form of different survey items.

In scenario 1, the latent variable, y^{true} is assumed to follow an AR(1) process and to exert a causal effect on subsequent realizations of observed variable x:

$$x_{it} \sim N(\mu_1 + \zeta_1 x_{it-1} + \delta_1 y^{true}_{it-1}, \sigma_x) \tag{13}$$

$$y^{true}_{it} \sim N(\zeta_2 y^{true}_{it-1}, \sigma_{y^{true}}) \tag{14}$$

y^{true} is unobserved; instead we observe $k = 3$ indicators of y, y^{obs}. The DGP for these is kept rather simple: an item location parameter λ_k plus the true value of y at each temporal and cross-sectional unit plus some random noise:

$$y^{obs}_{ikt} \sim N(\lambda_k + y^{true}_{it}, \sigma_{y^{obs}}) \tag{15}$$

In scenario 2, y^{true} is modeled as endogenous to x, with x being exogenous. We allow x to exert two distinct effects on y^{true}, a lagged effect which may accumulate over the long run (δ_2), and an immediate effect which is exerted contemporaneously and fades quickly (δ_3). This parallels the specification of two similar effects of democracy on democratic mood in Claassen (2020*b*). The measurement model linking y^{true} and y^{obs} remains the same as previously specified.

$$x_{it} \sim N(\mu_1 + \zeta_1 x_{it-1}, \sigma_x) \tag{16}$$
$$y^{true}_{it} \sim N(\mu_2 + \zeta_2 y^{true}_{it-1} + \delta_2 x_{it-1} + \delta_3 \Delta x_{it}, \sigma_{y^{true}}) \tag{17}$$

Finally, **in scenario 3**, both y^{true} and x are modeled as endogenous, combining equations 13 and 17 above (as well as measurement model 15). Therefore, while in scenario one we estimate structural parameter δ_1, and in scenario two we estimate δ_2 and δ_3, in scenario three, we estimate all three of these structural parameters.

In all scenarios, I set $\zeta_1 = \zeta_2 = 0.9$ to reflect the strong serial correlation observed in both democracy and democratic mood. The residual standard deviation for both structural models, $\sigma_{y^{true}}$ and σ_x is set to 0.1, which is approximately the value which holds in dynamic regressions of democracy and democratic mood when these are standardized. The residual standard deviation for the measurement model, $\sigma_{y^{obs}}$ is set substantially higher, at 1. This provides for a substantial degree of measurement error, which is when we might expect to see differences between the three methods of including measurement error. The effect of the lagged latent variable on the observed variable, δ_1, is set to a small positive value, 0.1. The effect of the (lagged) observed variable on the latent variable, δ_2, is also fixed at 0.1, while we set the effect of the first difference of the observed variable on the latent variable, δ_3, to a larger but negative value, -0.8. These assumptions loosely parallel the findings of Claassen (2020*a*; 2020*b*) regarding the effects of mood on democracy and vice versa. Other parameters are set as follows: $\mu_1 = \mu_2 = 0$ and $\lambda = \{-1, 0, 1\}$.

The Monte Carlo simulation for each scenario proceeds as follows:

1. Generate values for x and y^{true} in time period $t = 0$ by sampling $N = 50$ observations from U $(-1, 1)$ distributions.
2. Generate values for x and y^{true} in time periods $t \in \{1, \dots, 30\}$ using the equations described above for each scenario.
3. Produce $K = 3$ observed indicators of y^{true}, y^{obs}_k, using equation 15. A random 20% of $N \times T$ observations for each y^{obs}_k vector are retained; the rest are discarded and treated as missing.
4. Produce estimates of y^{true}, y^{est}, using a measurement model which closely follows model 16 (MCMC simulation is used):
$$y^{obs}_{ikt} \sim N(\lambda_k + y^{est}_{it}, \sigma_{y^{obs}}) \tag{18}$$
$$y^{est}_{it} \sim N(\zeta_2 y^{est}_{it-1}, \sigma_{y^{est}}) \tag{19}$$
5. For the MEX test, fit the following model(s):
$$x_{it} \sim N(\mu_1 + \zeta_1 x_{it-1} + \delta_1 y^{est}_{it-1}, \sigma_x) \text{ – in scenarios 1 and 3} \tag{20}$$
$$y^{est}_{it} \sim N(\mu_2 + \zeta_2 y^{est}_{it-1} + \delta_2 x_{it-1} + \delta_3 \Delta x_{it}, \sigma_{y^{est}}) \text{ – in scenarios 2 and 3} \tag{21}$$

   Simulate 500 draws from the parameters of interest using D, the estimated coefficient vector, and Σ, the estimated parameter variance covariance matrix.
6. For the MOC test: fit models 20 and/or 21 using 500 draws of y^{est} from step 4. For each of these 500 draws, fit model(s) 20 and/or 21 once, capturing structural modeling uncertainty by simulating a draw from the parameters of interest using D and Σ.

7. For the UM test: fit the unified model corresponding to the scenario (various combinations of equations 18 to 21). Save 500 draws from the posterior density of the parameters of interest.

2.2. Specification and estimation of MCMC models

In our Monte Carlos analyses, four different MCMC models are fit across the three scenarios: (1) a simple univariate latent variable model used to measure the latent variable without covariates (specified in equations 18 and 19 above); (2) a unified model that estimates an exogenous latent variable (equations 18 and 19) and exerts effects on a second observed variable (equation 20); (3) a second unified model that estimates a latent variable which is endogenous to an observed variable (equations 18 and 21); (4) a third unified model that estimates an endogenous latent variable and includes an endogenous observed variable (equations 18, 20, and 21).

These MCMC models are generally specified in line with the principles described in section 3.5 below (i.e., redundant parameterizations are used whenever possible). Certain other features are worth highlighting. The univariate measurement model includes an autoregressive parameter, ζ_2. This parameter is estimated, allowing the measurement model to (in principle) model the data-generating process accurately, where this parameter was fixed at 0.9. This contrasts with the measurement model employed by Claassen (2020*a*;*b*) and below, where this parameter was fixed at 1. To identify this parameter, an informative prior of N(0.9, 0.01) is used. Weakly informative priors are otherwise used (e.g., $N^+(0, 1)$ for $\sigma_{y^{est}}$).

The three unified model, used in scenario 1, employ weakly informative priors – N(0, 1) – for the key parameters of interest, δ_1, δ_2, and δ_3. The informative prior for ζ_2 – N(0.9, 0.01) – is retained, being necessary to identify these models in certain situations.

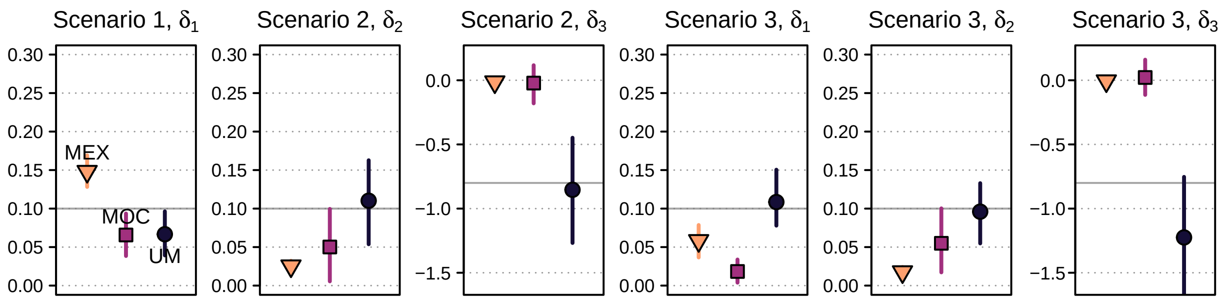
These Bayesian models are estimated using Bayesian Markov-Chain Monte Carlo (MCMC) methods and Stan software (Carpenter et al. 2017). Three serial chains, with randomly selected starting values drawn from a Uniform $(-1, 1)$ distribution, are run using 500 warmup and 1000 post-warmup samples each. The maximum R-hat statistic for each model in each iteration of the Monte Carlo test is saved and examined further. Any iterations of the Monte Carlo tests in which models showed a maximum R-hat of 1.1 or greater were dropped. There were six such instances, leaving 494 Monte Carlo iterations for analysis.

2.3. Additional Monte Carlo results

We focus on the accuracy of our three methods in estimating the true values of the three structural parameters, δ_1, δ_2, and δ_3. In the main paper we discuss the bias of each method and its uncertainty (confidence or credible) interval coverage. Here we present additional findings, which directly compare the estimates obtained for each method/scenario/parameter combination with the true values that were used to generate the underlying datasets. With 496 Monte Carlo iterations and 500 draws for each, we have 248,000 estimates for each combination. This incorporates sampling, structural modeling, and measurement uncertainty. We calculate the mean parameter estimate for each parameter and modeling approach as well as the central 95% quantiles. These results are presented in Figure S1.

In essence, UM estimates are generally close to the true parameter values, with the latter always falling within the 95% uncertainty intervals shown in the figure. MOC varies in accuracy, being comparable to UM in accuracy when the latent variable y is exogenous (scenario 1). How-

Figure S1. Results of the Monte Carlo Studies: Average Parameter Estimate by Scenario and Methos



Estimates for the three structural parameters (δ_1, δ_2, and δ_3) by the three methods (MEX, MOC, and UM). These results combine measurement and estimation error from both posterior simulation draws (using MCMC or MOC) with sampling error from multiple simulated Monte Carlo datasets.

ever, MOC is less accurate when the latent variable is endogenous to x, and especially when it comes to the first difference of x on y, i.e., δ_3, with estimates close to 0 rather than the true value of -0.8. While the MOC estimates of δ_2 are less accurate than those obtained by UM, the true value of 0.1 falls within the 95% confidence intervals, which is more than can be said for excluding measurement error (MEX). Nevertheless, when estimating δ_1 in scenario 3, MOC is more inaccurate than MEX.

3. Application of the methods to the case of democracy and mood

3.1. Democratic mood

Democratic mood captures the extent to which a national public offers explicit support for a democratic system and rejects any autocratic alternatives. It is principled or diffuse support for democracy itself, rather than instrumental support for the outputs of government or the incumbent office-holders. Consequently, mood is measured using existing survey questions which ask respondents to evaluate the appropriateness or desirability of democracy; compare democracy to some undemocratic alternative; or evaluate one of these undemocratic forms of government. Such items are widely used to measure democratic support (e.g., Bratton, Mattes, and Gyimah-Boadi 2005; Magalhães 2014; Norris 2011). Questions focusing on related concepts such as satisfaction with the performance of democracy and trust in national political institutions were not included because neither is a valid measure of principled support for democracy (e.g., Canache, Mondak, and Seligson 2001; Linde and Ekman 2003). A full list of the survey items which were used are available in the replication materials folder, here (url to be added).

Two datasets are used to estimate democratic mood in the applied section of the paper. The first is the original dataset collected by Claassen (2020*a*; 2020*b*), which runs from 1988 (in some cases) to 2017. Following Claassen's initial coding rules, countries with less than two years' of survey measures were dropped, as were countries lacking V-Dem democracy data (either non-independent territories or micro-states). Data from items that were fielded in only one wave of surveys were also dropped to ensure that item-country parameters were identifiable. There are 3,531 nationally aggregated survey responses remaining, drawn from 13 survey projects and 135 countries.

The models were also run on an expanded dataset. Although Tai, Hu, and Solt (2022) collect and publish their own expanded dataset, it appears to include a number of errors and inconsistencies (see next section), necessitating that this data be recollected. This is accomplished by extending Claassen's original dataset to 2020 by adding new survey measures from the survey projects that were originally included. Data from two smaller cross-national survey projects – the Arab Transformation Project and the Consolidation of Democracy in Central and Eastern Europe – were also added. In addition, two errors in Claassen's original dataset, identified by Hu, Tai, and Solt (2022) are rectified here; these are:

- In the 2010 Canadian LAPOP sample, the "strong leader" question is asked of only half the sample meaning most missing values are not non-responses.
- The "democracy suitability" question in the second wave (2005-8) of the Asianbarometer was originally dichotomized differently than the same question in the first and third waves of the Asianbarometer.

In the expanded dataset, countries with less than two years' data are once again dropped. Survey items fielded only once are now retained however: Claassen's initial concern – that item-country parameters would not otherwise be identified – seems to have been too cautious. The resulting dataset includes 4,445 nationally aggregated survey responses gathered by 16 survey projects in 141 countries.

3.2. Irregularities in Tai et al "expanded" dataset of democratic mood

A brief consideration of Tai et al's expanded dataset (`exp_claassen_input.rda`, available here: `https://doi.org/10.7910/DVN/XAUF3H`) reveals a number of errors, as well as inconsistencies with Claassen's initial coding rules.

1. Many of the sample sizes appear to be incorrect. The measurement model treats observed survey data as a binomial count, which requires the total sample size and the number of respondents who offered a supportive view of democracy. Many of the sample sizes reported in Tai et al's expanded dataset appear to be incorrect, however, which influences the measurement error attached to each data point in the latent variable model. Some of these errors are obvious, e.g., the implausibly small sample sizes reported in many of the Gallup Voices surveys. For example, a sample of 26 respondents is reported for Iceland in 2004 although the codebook states that the sample size is 502 (see `https://www.icpsr.umich.edu/web/ICPSR/studies/24681/datadocumentation`).

2. Non-representative samples are included. Tai et al's expanded dataset includes samples such as Senegal's 2002 Pew Global Attitudes survey, which Pew themselves describe as not representative of the adult population, but rather, as a "disproportionately urban" sample (see `https://www.pewresearch.org/wp-content/uploads/sites/4/legacy-pdf/165.pdf`). Given that Senegal's population was roughly 60% rural-dwelling in 2002, these data may be considerably biased and should not be included. Many of the other samples fielded this wave of the Pew Global Attitudes surveys were also unrepresentative (e.g., Pakistan, Cote d'Ivoire, India). Other unrepresentative samples, e.g., Morocco and Pakistan, were fielded in the next (2005) wave (see `https:`

//www.pewresearch.org/wp-content/uploads/sites/2/pdf/251.pdf).

3. Respondents who do not provide a response are (apparently) excluded. Respondents who answered "don't know" or refused to respond appear to have been removed from the Tai et al. expanded dataset before the percentage supporting democracy was calculated. Whether or not this should be done is perhaps a matter of debate. Dropping such respondents is not, however, consistent with the original Claassen coding scheme. As described in the supplementary materials to Claassen (2020*b*), "all other possible responses (i.e., the difference between the sample size and the number of supportive respondents) were treated, similarly, as not supportive of democracy. These non-supportive respondents may have actively opposed democracy, (e.g., 'an authoritarian government can be preferable to a democratic one'), chosen an intermediate response (e.g., 'for someone like me, it does not matter what kind of government we have'), responded with 'don't know,' or refused to provide any response." (p. 6).

Consider the case of Brazil, the question on whether democracy is important, and the most recent wave of the World Values Survey (2018). The number of observations in the dataset (`WVS_Cross-National_Wave_7_spss_v20200720.sav`) is 1,762. The Tai et al expanded dataset has this sample size as 3,214 however. The raw number of respondents who offered support for democracy in response to this question (defined as offering an opinion above the median on the 1-10 scale; i.e., 6 or above) is 1,271, or 72.1%. Weighting each respondent by the including survey weight, `W_WEIGHT`, produces a weighted response percentage of 72.5%. However the Tai et al dataset reports that 2,557 respondents supported democracy in this question (out of 3,214), for a response percentage of 79.6%. Similar results to Tai et al can be obtained by dropping respondents who declined to provide a response to this particular question: excluding these 161 respondents results in an unweighted percentage agreement of 79.4% and a weighted percentage agreement of 79.8%.

3.3. Other variables used in the applied analysis

Democracy is measured using the liberal democracy index from the Varieties of Democracy (V-Dem). We use both the point estimates and standard deviations. Analyses of the original dataset use version 8 of the V-Dem dataset; analyses of the extended dataset use version 11 (there is some variation in national liberal democracy scores in various versions of the V-Dem dataset). Several other covariates are included in our structural models:

1. Regional democracy is measured using V-Dem's liberal democracy index aggregated to United Nations sub-regions.
2. The logarithm of GDP per capita, included in lagged levels and first differences (the latter being a measure of economic growth). To remove missing values, we use GDP measures from several sources: IMF, World Bank World Development Indicators, Penn World Tables, and Maddison. These data are combined using hierarchical linear models with country-varying intercepts and slopes.
3. Natural resource dependence: an indicator taking a value of one if natural resource products (natural gas, oil, and coal) were valued at more than USD 1,000 per person in any given country-year observation (see, e.g., Haber and Menaldo 2011). Data are primarily drawn from the World Bank World Development Indicators. Missing values are imputed using data

from Haber and Menaldo (2011); remaining missing values are given country modes (i.e., 0 or 1).

4. Proportion of the population identifying as Muslim in 1990, from Pew Research.

3.4. Differences with the original specifications

Claassen (2020*a*) reports the results of the Lipset hypothesis (mood affects change in democracy) and Claassen (2020*b*) reports the tests of the thermostatic hypothesis (change in democracy affects change in mood). These two papers employ slightly different specifications. The unified model used in the present paper requires that these be aligned however. To accomplish this, I make some changes to the original specifications.

1. I shift from the error correction models of democratic mood used by Claassen (2020*b*) to the autoregressive distributed lag model, as used by Claassen (2020*a*). In practice, this only affects the first lag parameter.
2. I rescale democracy and mood to be unit-normal standardized. This is consistent with Claassen (2020*b*), but departs from Claassen (2020*a*), where democracy was scaled using the original V-Dem 0-100 scale.
3. While Claassen (2020*a*) included lagged growth in GDP per capita as a measure of economic conditions, I follow Claassen (2020*b*) in using the immediate first difference of log GDP per capita as a measure of growth.
4. I re-estimate democratic mood using the most comprehensive measurement model proposed by Claassen (2019), which includes item discrimination parameters as well as item location parameters. This facilitates a comparison with the unified model, where such features are incorporated. Note that the new estimates of mood (using on the original 1988-2017 dataset) correlate at 0.99 with the estimates provided by Claassen (2020*a*;*b*).

3.5. Specification and estimation of MCMC models

When applying the methods to the mood and democracy datasets, I fit three models using Bayesian MCMC methods: (1) the univariate latent variable model described in section 1.1 above and originally proposed and used by Claassen (2019; 2020*a*;*b*); (2) the unified model that estimates an overall effect of mood on democracy; (3) the unified model which allows this effect of mood on democracy to vary by regime. Each of these models is employed twice – once using Claassen's original data and again using an expanded dataset. The estimation of these models is described here.

3.5.1 Specifying the univariate latent variable models

The dynamic latent variable described in section 1 is used to estimate mood for later use in the MEX and MOC analyses. Several include several computational refinements, compared the model developed by Claassen (2019), are included. First, I allow for ragged country-by-year arrays to accommodate the varying length of national latent opinion time-series (due to the varying years in which survey measurement commenced). I also make use of non-centered parameterizations for all variance terms, e.g., $\sigma_{\theta^{(m)}}$, $\sigma_{\gamma^{(m)}}$, and $\sigma_{\delta^{(m)}}$. Non-centered parameterizations include standard-normally distributed redundant parameters, e.g., $\nu_{ik}^{\delta^{(m)}}$ which shift variance and covariance terms away from zero, making MCMC sampling more efficient:

$$\delta_{ik}^{(m)} = \sigma_{\delta^{(m)}}^2 \times \nu_{ik}^{\delta^{(m)}} \tag{22}$$

The item-country variances are given weakly-informative half-Normal priors, e.g., $\sigma_\delta^{(m)} \sim N^+(0,\ 1)$. The variance-covariance matrix for the item intercepts λ and slopes γ is split into two variances and correlation term, with the former receiving a half-Normal $(0, 1)$ prior and the latter an LKJ (2) prior. Item intercepts and slopes are identified by setting their expectations: the former at the log of the mean proportion expressing support for democracy, and the latter at 0.5. The beta-binomial dispersion parameter ϕ receives a gamma$(3,\ 0.04)$ prior. Since latent opinion is modeled as a function of its value in the previous year, I estimate initial values for each country in the year preceding the first estimates based on data. These initial values receive a $N(0,\ 1)$ prior.

3.5.2 Specifying the unified models

Non-centered parameterizations are again used for all variance terms. Variance parameters all receive half-normal $N^+(0,\ 1)$ priors. Parameters capturing the effects of lagged outcome variables are restricted to ensure stationarity of each time series. The second lag is restricted to lie between -1 and +1 and is given a Uniform $(-1, 1)$ prior. The sum of first and second lags is restricted to lie between 0 and 1; it receives a weakly informative Beta $(3, 1)$ prior. The first lag is then defined as the difference between the lag sum and the second lag.

Since the models of mood and democracy require two lags of each, I estimate two years' worth of initial values for mood and democracy. These initial values are given $N(0,\ 1)$ priors. Regression / structural parameters are given $N(0,\ 1)$ priors. Structural models only take the initial values of democracy or mood as inputs into lagged outcomes (and necessarily only in years 1 and 2). The values used as outcome variables are based only on estimates obtained for years in which mood survey data are available. Structural model residuals are also given a non-centered parameterization.

3.5.3 Estimation

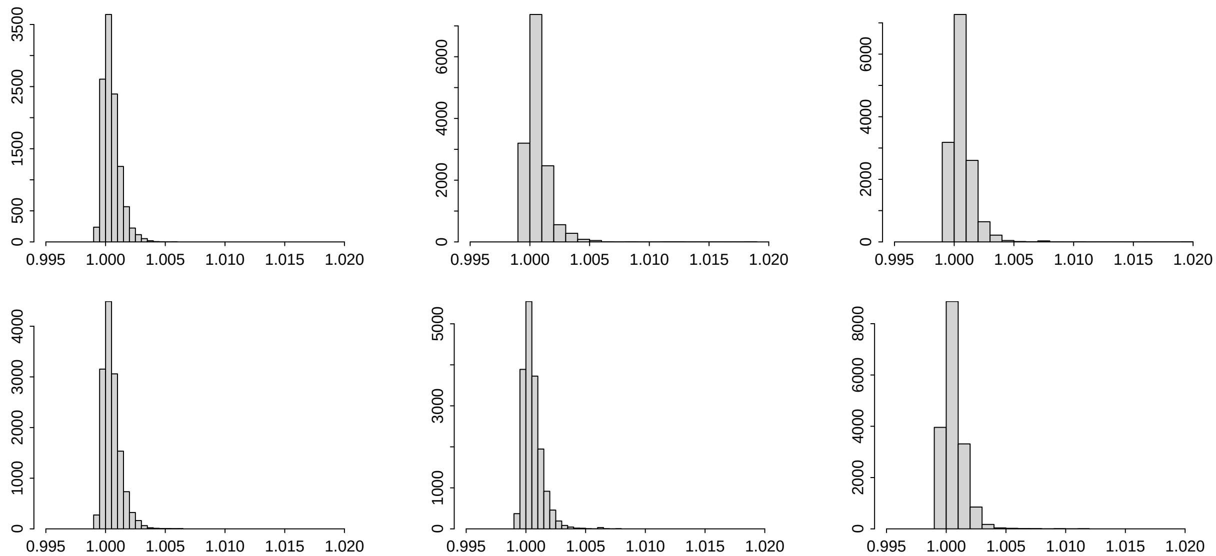
These Bayesian models are estimated with Bayesian Markov-Chain Monte Carlo (MCMC) methods. Stan software, which implements Hamiltonian Monte Carlo sampling (Carpenter et al. 2017), is employed. Four parallel chains, with randomly selected starting values drawn from a Uniform $(-1, 1)$ distribution, are run, with 500 warmup and 1,500 post-warmup samples each. The 4,000 post-warmup samples are saved and analyzed further.

3.6. Model Checking

Convergence of the MCMC models is assessed using a variety of diagnostics, including traceplots of multiple parameters and Gelman-Rubin R-hat statistics. The latter were close to one for all models (Figure S2), indicative of convergence.

The models can be further verified using posterior predictive checking: simulating data conditional on the estimated parameters and comparing the simulated data against the actual data

Figure S2. Model convergence as shown by the Gelman-Rubin R-hat diagnostic

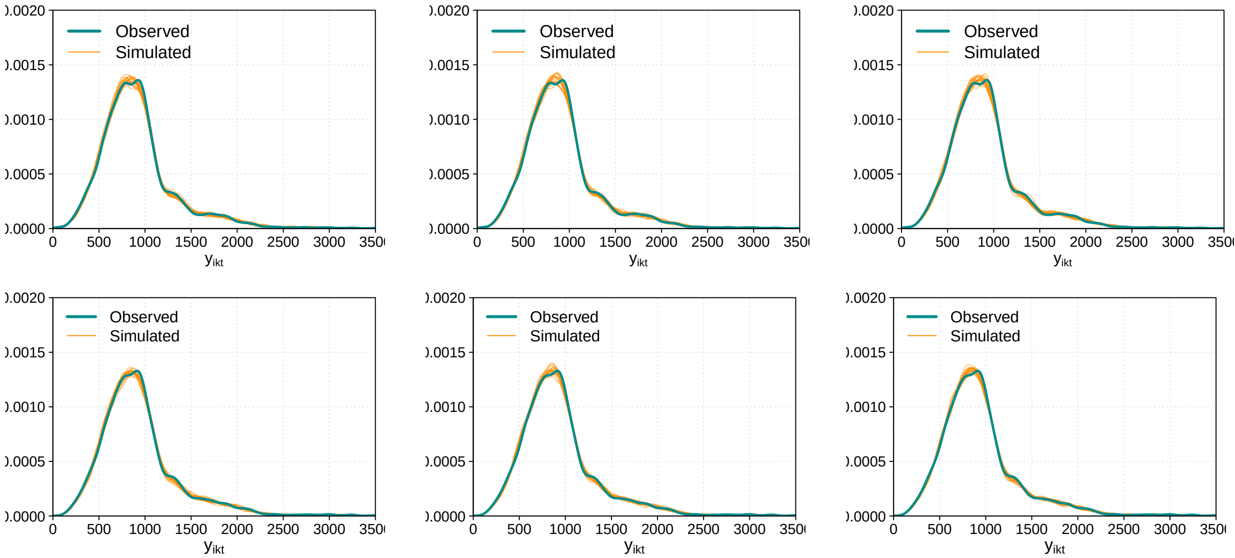


Distribution of the Gelman-Rubin R-hat statistic for all parameters, one plot for each model. First column: univariate measurement model; second column: unified model with overall mood effects; third column: unified model with regime-specific mood effects. Top row: original data; bottom row: extended data.

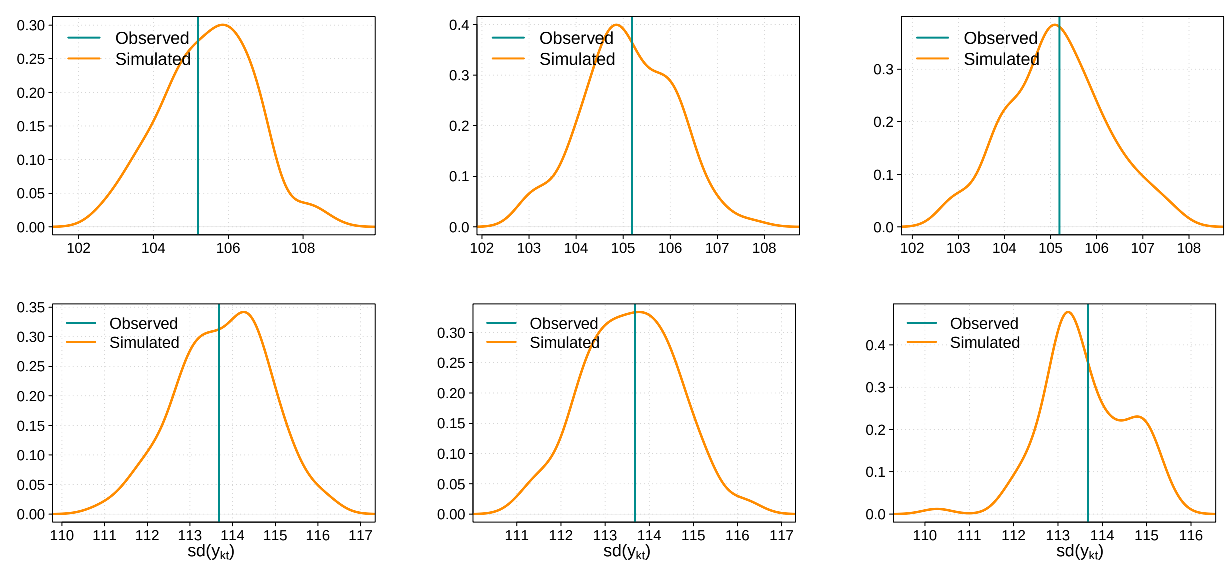
used to produce the estimated parameters (Gelman et al. 2014). As the plots in the top half of Figure S3 show, there is a close correspondence between the aggregated survey responses for each of the national mood items in our dataset $y_{ikt}^{(m)}$ and those we simulate $\tilde{y}_{ikt}^{(m)}$ in each of our six Bayesian models.[2] The plots in the bottom half of the figure then compare the within-country variability from 100 draws of the response count vector simulated from each model ($\sqrt{(}Var(\tilde{y}_{kt}^{(m)})))$ to that observed in the observed vector of response counts ($\sqrt{(}Var(y_{kt}^{(m)})))$. These posterior predictive checks suggest that each model fits the observed survey responses.

[2]These models are the measurement model to generate estimates for MEX and MOC analyses and two unified models, with either overall effects of mood, or regime-specific effects. Each of these three is run using the original and extended datasets.

Figure S3. Posterior predictive checks



Each plot above compares the distributions of response counts in the observed survey measures of mood (teal line) against 20 draws from the posterior distributions of the response counts y_{ikt} estimated by the particular model (orange lines). First column: univariate measurement model; second column: unified model with overall mood effects; third column: unified model with regime-specific mood effects. Top row: original data; bottom row: extended data.



Each plot above compares the within-country standard deviation in response counts to the distribution of within-country standard deviations from 100 simulated response counts $\tilde{y}_{ikt}$ generated by each model (orange lines). First column: univariate measurement model; second column: unified model with overall mood effects; third column: unified model with regime-specific mood effects. Top row: original data; bottom row: extended data.

3.7. Tables of results

Table S1. Replication of Claassen (2020*a*) AJPS results

	Dependent variable: Democracy					
	MEX		MOC		UM	
Intercept	−.008	−.003	−.096	−.096	.002	.005
	(.027)	(.027)	(.071)	(.068)	(.003)	(.003)
Democracy_{t-1}	1.142	1.144	.583	.584	1.445	1.439
	(.020)	(.020)	(.027)	(.026)	(.038)	(.038)
Democracy_{t-2}	−.164	−.165	.352	.351	−.458	−.454
	(.020)	(.020)	(.027)	(.027)	(.038)	(.037)
Mood_{t-1}	.009		.015		.010	
	(.003)		(.006)		(.002)	
Mood, democracies only_{t-1}		.011		.013		.005
		(.003)		(.007)		(.003)
Mood, autocracies only_{t-1}		.003		.019		.018
		(.005)		(.013)		(.004)
Log GDP per capita_{t-1}	.001	.000	.010	.010	−.001	.000
	(.003)	(.003)	(.008)	(.007)	(.002)	(.002)
Δ log GDP capita	.152	.151	.153	.146	.097	.093
	(.048)	(.048)	(.125)	(.135)	(.055)	(.053)
Regional democracy_{t-1}	.007	.006	.028	.027	.004	.005
	(.004)	(.004)	(.009)	(.010)	(.002)	(.002)
Resource dependence_{t-1}	−.011	−.012	−.036	−.036	−.006	−.006
	(.010)	(.010)	(.024)	(.024)	(.006)	(.006)
Proportion Muslim	−.006	−.008	−.022	−.020	−.007	−.005
	(.009)	(.009)	(.022)	(.022)	(.005)	(.006)
N	2435	2435	2435	2435	2300	2300

Original, 1988-2017 dataset used. MEX: measurement error excluded; cell entries are coefficient estimates and conventional standard errors (to facilitate comparison with UM). MOC: method of composition; cell entries are MOC parameter estimates and standard errors (see section 2 for details). UM: unified model; cell entries are posterior means and standard deviations for each parameter. Note that the latent variables variances in the UM differ slightly from presented in the main paper; there, reported parameter estimates for key structural parameters are standardized based on the empirical distributions of latent variables.

Table S2. Replication of Claassen (2020*b*) APSR results

	Dependent variable: Democratic mood		
	MEX	MOC	UM
Intercept	−.037	−.020	−.081
	(.025)	(.081)	(.025)
Mood_{t-1}	1.473	.961	.422
	(.018)	(.033)	(.057)
Mood_{t-2}	−.487	.003	.471
	(.018)	(.031)	(.059)
Democracy_{t-1}	.007	.019	.073
	(.003)	(.011)	(.020)
Δ democracy	−.063	−.004	−.950
	(.021)	(.030)	(.299)
Log GDP per capita_{t-1}	.003	.001	.010
	(.003)	(.009)	(.021)
Δ log GDP capita	.071	.114	1.151
	(.051)	(.186)	(.451)
N	2300	2300	2300

Original, 1988-2017 dataset used. MEX: measurement error excluded; cell entries are coefficient estimates and conventional standard errors (to facilitate comparison with UM). MOC: method of composition; cell entries are MOC parameter estimates and standard errors (see section 2 for details). UM: unified model; cell entries are posterior means and standard deviations for each parameter. Note that the latent variables variances in the UM differ slightly from presented in the main paper; there, reported parameter estimates for key structural parameters are standardized based on the empirical distributions of latent variables.

Table S3. Extension of Claassen (2020*a*) AJPS results

	Dependent variable: Democracy					
	MEX		MOC		UM	
Intercept	−.037	−.030	−.174	−.185	.004	.004
	(.024)	(.024)	(.057)	(.057)	(.002)	(.002)
Democracy$_{t-1}$	1.156	1.157	.643	.642	1.467	1.471
	(.015)	(.015)	(.023)	(.023)	(.033)	(.034)
Democracy$_{t-2}$	−.182	−.182	.282	.282	−.480	−.483
	(.015)	(.015)	(.023)	(.023)	(.033)	(.034)
Mood$_{t-1}$	.008		.014		.006	
	(.003)		(.006)		(.002)	
Mood, democracies only$_{t-1}$		.010		.013		.007
		(.003)		(.007)		(.002)
Mood, autocracies only$_{t-1}$		.001		.016		.005
		(.005)		(.012)		(.003)
Log GDP per capita$_{t-1}$	.006	.005	.018	.020	.002	.002
	(.003)	(.003)	(.006)	(.006)	(.002)	(.002)
Δ log GDP capita	.042	.042	.008	.006	.050	.049
	(.023)	(.023)	(.060)	(.059)	(.021)	(.021)
Regional democracy$_{t-1}$	.006	.006	.028	.028	.004	.004
	(.004)	(.004)	(.008)	(.009)	(.002)	(.002)
Resource dependence$_{t-1}$	−.024	−.025	−.062	−.062	−.011	−.011
	(.008)	(.008)	(.021)	(.021)	(.004)	(.005)
Proportion Muslim	−.011	−.013	−.035	−.033	−.004	−.005
	(.008)	(.008)	(.021)	(.019)	(.004)	(.004)
N	2927	2927	2927	2927	2786	2786

Extended, 1988-2020 dataset used. MEX: measurement error excluded; cell entries are coefficient estimates and conventional standard errors (to facilitate comparison with UM). MOC: method of composition; cell entries are MOC parameter estimates and standard errors (see section 2 for details). UM: unified model; cell entries are posterior means and standard deviations for each parameter. Note that the latent variables variances in the UM differ slightly from presented in the main paper; there, reported parameter estimates for key structural parameters are standardized based on the empirical distributions of latent variables.

Table S4. Extension of Claassen (2020*b*) APSR results

	Dependent variable: Democratic mood		
	MEX	MOC	UM
Intercept	$-.033$	$-.079$	$-.104$
	(.019)	(.063)	(.018)
Mood_{t-1}	1.504	.968	.491
	(.016)	(.025)	(.057)
Mood_{t-2}	$-.514$	.000	.426
	(.016)	(.026)	(.059)
Democracy_{t-1}	.009	.020	.067
	(.003)	(.008)	(.016)
Δ democracy	$-.040$	.003	$-.609$
	(.017)	(.024)	(.250)
Log GDP per capita_{t-1}	.002	.007	.029
	(.002)	(.007)	(.016)
Δ log GDP capita	.038	.067	.503
	(.020)	(.069)	(.176)
N	2786	2786	2786

Extended, 1988-2020 dataset used. MEX: measurement error excluded; cell entries are coefficient estimates and conventional standard errors (to facilitate comparison with UM). MOC: method of composition; cell entries are MOC parameter estimates and standard errors (see section 2 for details). UM: unified model; cell entries are posterior means and standard deviations for each parameter. Note that the latent variables variances in the UM differ slightly from presented in the main paper; there, reported parameter estimates for key structural parameters are standardized based on the empirical distributions of latent variables.

References

Bratton, Michael, Robert Mattes, and E. Gyimah-Boadi. 2005. *Public Opinion, Democracy, and Market Reform in Africa*. Cambridge, UK: Cambridge University Press.

Canache, Damarys, Jeffery J. Mondak, and Mitchell A. Seligson. 2001. "Meaning and Measurement in Cross-National Research on Satisfaction with Democracy." *Public Opinion Quarterly* 65(4): 506–528.

Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76(1): 1–32.

Claassen, Christopher. 2019. "Estimating Smooth Country-Year Panels of Public Opinion." *Political Analysis* 27(1): 1–20.

Claassen, Christopher. 2020*a*. "Does Public Support Help Democracy Survive?" *American Journal of Political Science* 64(1): 118–134.

Claassen, Christopher. 2020*b*. "In the Mood for Democracy? Democratic Support as Thermostatic Opinion." *American Political Science Review* 114(1): 36–53.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2014. *Bayesian Data Analysis, 3rd Ed.* CRC Press.

Haber, Stephen, and Victor A. Menaldo. 2011. "Do Natural Resources Fuel Authoritarianism? A Reappraisal of the Resource Curse." *American Political Science Review* 105(1): 1–26.

Hu, Yue, Yuehong Cassandra Tai, and Frederick Solt. 2022. "On Data 'Janitor Work' in Political Science: The Case of Thermostatic Support for Democracy." Unpublished paper, `https://doi.org/10.31235/osf.io/kd7mu`.

Kellstedt, Paul, Gregory E. McAvoy, and James A. Stimson. 1993/94. "Dynamic Analysis with Latent Constructs." *Political Analysis* 5: 113–50.

Linde, Jonas, and Joakim Ekman. 2003. "Satisfaction With Democracy: A Note on a Frequently Used Indicator in Comparative Politics." *European Journal of Political Research* 42(3): 391–408.

Magalhães, Pedro C. 2014. "Government Effectiveness and Support for Democracy." *European Journal of Political Research* 53(1): 77–97.

Norris, Pippa. 2011. *Democratic Deficit: Critical Citizens Revisited*. Cambridge: Cambridge University Press.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: CRC Press.

Tai, Yuehong 'Cassandra', Yue Hu, and Frederick Solt. 2022. "Democracy, Public Support, and Measurement Uncertainty." *American Political Science Review* `https://doi.org/10.1017/S0003055422000429`.

Tanner, Martin A. 1996. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York: Springer.

Treier, Shawn, and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1): 201–217.