A measure of personalism in dictatorships

Barbara Geddes, Joseph Wright, and Erica Frantz*

October 9, 2017

Abstract

Geddes, Wright and Frantz (2017) employ a time-varying measure of personalism throughout their book. This document provides more details on this measure. It describes the variables and method used to construct the measure, offers a comparison of the measure with existing ones, and closes with a brief discussion of the measure's limitations and potential avenues for future research.

*This research is funded by the Minerva Research Initiative (ONR #N000141211004). The authors thank Anna Luhrmann, Mike Kenwick, and Bumba Mukherjee for helpful suggestions in writing this up. Correspondence on this paper should be directed to joseph.g.wright@gmail.com. Data collection for this project was funded by the National Science Foundation (BCS-0904478 and BCS-0904463).

Introduction

In their book, Geddes, Wright and Frantz (2017) (GWF, hereafter) describe personalism as the following:

> We refer to dictatorships in which the leader has concentrated power at the expense of his closest supporters as personalist. The defining feature of personalist dictatorship is that the dictator has personal discretion and control over the key levers of power in his political system. Key levers of power include the unfettered ability to appoint, promote, and dismiss high-level officers and officials, and thus to control the agencies, economic enterprises, and armed forces the appointees lead. In such regimes, the dictator's choices are relatively unconstrained by the institutions that can act as veto players in other dictatorships, especially the military high command and the ruling party executive committee. Personalist dictators juggle, manipulate, and divide-and-rule other powerful political actors. Like all dictators, they need some support, but they can choose from among competing factions which ones can join or remain in the ruling elite at any particular time. Personalist dictators are thus powerful relative to other members of the elite, but not necessarily relative to society or to international actors.

Throughout the book, GWF, employ a time-varying latent measure of personalism to quantify this concept.[1] This document provides more details on this measure of personalism.1 First, it describes the variables used to construct the measure, before discussing the method selected to do so. Next, it offers a comparison of this measure with existing ones. Lastly, it provides a brief assessment of the measure's limitations and potential avenues for future research.

There are a few important points to note before beginning this discussion. First, in line with the GWF conceptualization of personalism, the measure does *not* attempt to capture the extent to which the state has a monopoly on violence within the country's borders nor does it measure the state's coercive power relative to other states' military power. Second, this definition of personalism (along with the coding questions for the manifest indicators and the historical information gathered for these indicators) is conceptually distinct from the following:

- observed state-led repression
- government coercive capacity vis-a-vis non-state actors
- leader legitimacy vis-a-vis citizens or elites
- leader rhetoric
- leader use of propaganda
- leader ideology
- individual character traits of the leader, such as charisma or attractiveness
- electoral campaign strategies
- distribution of legislative seats

[1] We refer readers to Wright (2017) for an introduction to the larger data collection project.

- electoral rules that incentivize the cultivation of a 'personalist' vote
- leader time in power
- regime duration

Eight items measuring latent personalism

To construct a latent personalism measure, we use variables from the GWF data set described in the book. Each variable takes a value of 0 or 1 so that we can use a logistic item-response theory (IRT) model.[2] The variables include:

1. Does access to high office depend on personal loyalty to the regime leader? (`high office`)
2. Did the regime leader create a new support political party after seizing power? (`create new party`)
3. Does the regime leader control appointments to the party executive committee? (`party exec committee`)
4. Is the party executive committee absent or simply a rubber stamp for the regime leader's decisions (`rubber stamp party`)
5. Does the regime leader personally control the security apparatus? (`security apparatus`)
6. Does the regime leader promote officers loyal to himself or from his ethnic, tribal, regional, or partisan group, or are there widespread forced retirement of officers from other groups? (`military promotion`)
7. Does the regime leader create paramilitary forces, a president's guard, or new security force loyal to himself? (`paramilitary`)
8. Does the regime leader imprison/kill officers from groups other than his own without a reasonably fair trial? (`military purge`)

Table 1 provides summary statistics for the 8 variables. Appointment to high office and personal control over the security apparatus are the most common, occurring in over 60 percent of observations. Creating a new party is the least common, occurring in only 17 percent of observations. These patterns will be borne out in the IRT model when we examine the "difficulty" parameters: variables that are more commonly observed are less "difficult" for observations to obtain, while variables that are less commonly observed are more "difficult" for observations to obtain.

[2] Only one variable is ordinal: `military promotions`. We collapse the first two categories into one.

Item	*Mean*	*Std. Dev.*
`high office`	0.647	0.478
`create new party`	0.165	0.371
`party exec committee`	0.319	0.466
`rubber stamp party`	0.304	0.46
`military promotion`	0.424	0.494
`military purge`	0.365	0.482
`security apparatus`	0.603	0.489
`paramilitary`	0.355	0.479

Table 1: Summary statistics for items ($N = 4591$).

Exploratory factor analysis

An initial approach to measuring latent personalism is exploratory factor analysis (EFA), which attempts to reduce the dimensionality of data with minimal loss of information. EFA assumes there exists some common factor(s) in the data that are identified by examining how the measures attributes correspond to some underlying feature. It relies on a linear projection of the many-dimensional features into the lower-dimensional ones. This approach allows us to see which raw variables contribute the information to underlying dimensions, which enables a more precise interpretation of the estimated measure. Further, factor analysis allows us to look at other potential dimensions that may be extracted from the variation in the items (i.e. variables) we explore, which can help us assess whether to focus on one dimension or multiple dimensions given the data. This contrasts with the IRT model discussed below, which assumes that variation in the data is unidimensional.

The left plot of Figure 1 shows the eigenvalues for the eight dimensions constructed from the factor analysis of the eight items; higher eigenvalues indicate that more The first dimension has an eigen value greater than 3, while the second dimension has a value just above 1. This suggests that the first dimension extracted from the factor analysis captures a large share of the total variation among the eight items. Indeed the first dimension captures 38 percent of the variation, while the next three dimensions contain 16, 12, and 11 percent of the variation, respectively. The right plot shows the factor loadings on the first dimension. Importantly, all eight items load strongly on the same first dimension. Appointment to high office (`high office`) contributes the most information to the first dimension, while creating a new support party (`create new party`) and creating a paramilitary organization personally loyal to the leader (`personal paramilitary`) contribute the least.

While the second dimension captures substantially less information than the first, we still want to examine it to understand what it may be measuring. Figure 2 shows the factor loading for each item on the first two dimensions. Factor loadings assess the strength of the relationship between each variable and the underlying factor or dimension. The horizontal axis depicts the first dimension (which is the one we use throughout) and the vertical axis depicts the second dimension. Note that both axes have the same scale from -0.6 to 0.8. Each red dot indicates the location of an item in the two-dimensional space.

All eight items contribute in the same direction on the first dimension. Further, when we

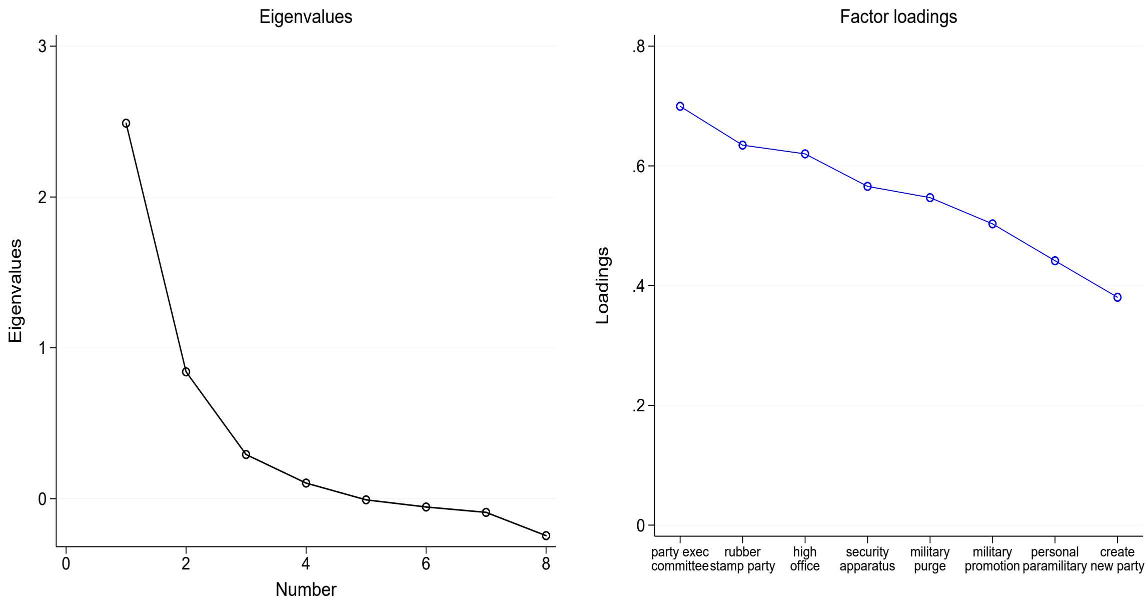


Figure 1: ***Exploratory factor analysis results.***

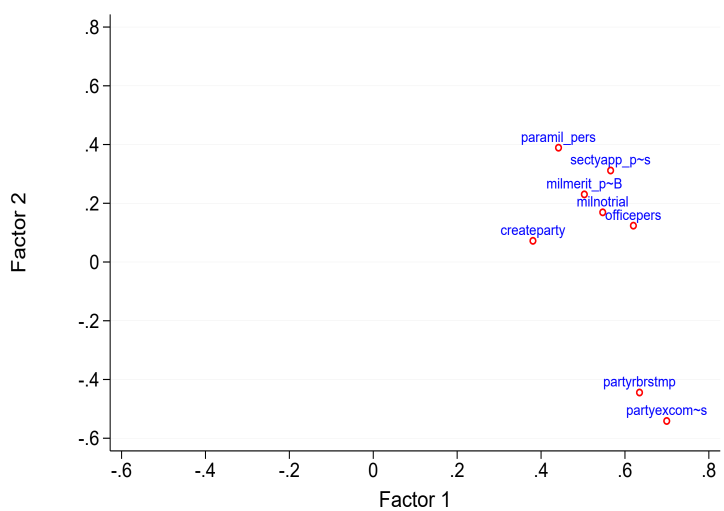


Figure 2: ***Factor loading for first two dimensions.***

visualize their locations in the first dimension on the same scale as the loadings on the second dimension (i.e. on a scale from -0.6 to 0.8), it becomes clear that each of the eight items contributes a roughly similar "amount" of information to the first dimension. The eight items do not load in the same direction on the second dimension (vertical axis). Rather the four variables relating to the military and security apparatus load in one direction while the two variables relating to leader control over the support party (`rubber stamp party` and `party exec committee`) load in the opposite direction. This suggests that the second dimension may be measuring a concept related to the distinction between regimes with support parties and those without them (including many military juntas).

Item	*Item-test correlation*
`high office`	0.6893
`create new party`	0.4849
`party exec committee`	0.6319
`rubber stamp party`	0.6178
`military promotion`	0.6163
`military purge`	0.6502
`security apparatus`	0.6522
`paramilitary`	0.5586

Table 2: Item-test correlations from Cronbach's α.

Table 2 shows similar results when calculating Cronbach's α, which is a measure of scale reliability or inter-correlation among the items. The test scale is 0.76, indicating that the eight items share some of the same dimensionality and are therefore highly inter-correlated. The individual items most highly correlated with the estimated scale are: `high office`, `security apparatus`, and `military purge`. Items with the lowest correlation with the scale are: `create new party` and `paramilitary`, which is similar to what the ordering shows in the factor loadings in the right plot of Figure 1. That said, these correlations are all relatively high, as shown when looking at the first dimension from PCA in Figure 2.

Together, these results indicate that the eight items contribute information to a latent trait, as captured in the first dimension of the PCA. If there is a second dimension that can be extracted from these eight items, it is likely related to the existence of a support party and contains substantially less variation in the items than the first dimension. We believe the first dimension is a plausible measure of personalism (as we define this concept) based on observed instances of power accumulating to the leader.

An item-response theory (IRT) approach

In this section, we outline an item response theory (IRT) approach for modeling latent personalism. We employ a 2-parameter logistic (IRT-2PL) model, where i indexes regimes, t indexes calendar years, and j indexes the eight items that we claim are observable indicators (with measurement error) of personalism.

$$Pr(y_{j,i,t} = 1|personalism_{i,t}) = logit(\delta_j + \beta_j personalism_{i,t}) \quad (1)$$

In this equation, δ_j is the difficulty parameter; β_j is the discrimination parameter for item j; and the logit function is a logistic transformation of the data. The purpose of the equation is to estimate $personalism_{i,t}$, which is the estimated degree of personalism for each regime-year. The difficulty parameter (δ_j) reflects the extent to which regimes, on average, are observed to have political events that correspond with one of the items, while the discrimination parameter (β_j) reflects the extent to which one item predicts another item.

Local independence

One assumption of the IRT model is local independence of the item responses, meaning response to one item does not influence response to another item conditional on the value of the latent trait. The data-collection process in this project may lead to violations of the local independence assumption when observed (historical) political events serve as manifestations for more than one item. For example, historical evidence of personal control over the security apparatus `security apparatus` may include (among other types of events that constitute evidence for this item): the leader creating new security force under his control or the leader appointing a family member as the head of a security force. Evidence of the former event (creating a new security force) may also be used as evidence for the `paramilitary` item. And appointing a family relative as head of a security force can also serve as an indication for the `high office` item. By construction, these items may be related because the same historical event is recorded as a manifest indicator for more than one item (variable).

Violations of local independence can lead to artificially inflated scores that may compromise reliability. One suggestive piece of evidence that indicates a violation of local independence is the presence of a second dimension that measures whether the regime has a supporting political party, as shown in Figure 2. If the items used to construct the latent personalism measure have a second dimension, the variation in this second dimension means the items – net of the latent personalism dimension – will be related. Our goal in this application was to construct a comprehensive measure of personalism rather than to construct the most reliable measure. For this reason we chose to incorporate information on two underlying aspects of personalism: personalism related to the security sector, including the military, and personalism related to the support party. Both aspects of personalism might be modeled separately in future research (see below); in doing, so future research will likely produce measures that more closely meet the local independence assumption.[3]

One source of item dependence is structural missingness in the data. For example, regimes that do not have a supporting political party will not have created a new political party (`create new party`), will not have a rubber stamp party (`rubber stamp party`), and the leader will not control appointments to the party executive committee (`party exec committee`). Thus responses on these items will be related (they are coded as `0` for each of these items) due to an unmodeled regime feature, in this case, lack of a supporting political party. This fact cannot be avoided because the

[3] See Song and Wright (2017) for an application of this data to the North Korean case. They trace the sequence of personalism to show that the first regime leader, Kim Il-Sung, consolidated power of the security sector prior to personalizing power over the Korean Workers Party.

real world of dictatorships creates this type of structural missingness.[4]

One way to address potential violations of local independence is to drop items that are related. We chose not to do this in an effort to maximize the comprehensiveness of the personalism measure. Another way to address local dependence is to combine related indicators into an ordered indicator and use a graded response IRT model. We show in the accompanying code that this approach, while reducing between-item dependence, yields almost identical information as the latent measure we use throughout.[5] We leave future research to build a personalism measure that directly models the structural missingness with the aim of addressing dependence across items.

Static model

The IRT model implemented here is static: the estimates of the latent trait are determined only by the information conveyed in the items for that country year, without accounting for possible temporal dependence in the data. However, the estimates of the latent trait for the same leader over multiple years in power are unlikely to be independent because the data were coded with instructions that imply *dependence* across some item responses (over time) for the same individual leader. For example, instructions for coding the `military promotion` item state that coders should "use the same codes for a leader's entire time in office unless observers specifically mention changes in the leader's promotion strategy." This coding rule likely creates dependence over time in item responses for the same regime leader. This coding procedure assumes that once a leader has grabbed sufficient 'power' from other elites for coders to observe a manifestation of this indicator, this 'power' remains as a latent capacity held by the leader, despite manifestations not being observed continously thereafter for individual leaders' duration as de facto regime leader. The data therefore mark observed manifestations of an increase in the leader's power, not decreases in this power.

One way to address dependence over time for individual leaders is via a dynamic IRT model that uses the prior estimate of the latent trait ($\theta_{i,t-1}$) for a particular leader i as the center of the distribution from which to draw the prior of the latent trait estimate ($\theta_{i,t}$) for i in subsequent period t; i.e. $\theta_{i,t} \sim N(\theta_{i,t-1}, k)$ where k is fixed.[6]

One rationale for estimating dynamic models is to enhance efficiency so the model with yield smaller standard errors around the estimate of the latent trait. In the applications in the book, we do not incorporate the uncertainty around the estimated latent trait so choosing a static model is unlikely to alter any of the relationships we uncover in the book. An alternative rational for using dynamic models is account for missing data. However, there is no missing data for any of the manifest items, so a dynamic model does not provide information from the lagged estimate to help estimate the latent trait for observations for which some items are missing.

When weighing the benefits of a dynamic IRT in this application, it is important to note that some of the cross-section units (i.e. distinct autocratic leaders) over which we would want to relax the independence assumption are relatively short-lived. Roughly a third of leaders last three years or less; half of the leaders in the data endure less than six years; and two-thirds of leaders do not last ten years. The large number of short time-series for the dynamic panels means that for relatively short-lived leaders, the dynamic and the static models would likely yield very similiar

[4] An additive index of personalism cannot address structural missingness either.

[5] The Spearman correlation is 0.97, while the Spearman correlation for the *within-leader* variation is 0.93.

[6] For example, Martin and Quinn (2002) set $k = 1$, while Gandhi and Sumner (2016) set $k = 0.25$. Kenwick (2017) draws k from a uniform distribution over (0,1).

estimates since there is little "time" for the priors drawn from a distribution pegged at the lagged value of the dependent variable to depart from the initial prior for each leader, which is drawn from a distribution pegged at `0`.[7] If one were interested estimating a dynamic model with this data, we would encouage mixing (short) panels treated as locally independent with (longer) panels modeled dynamically to test whether this mixed model yield statistically different estimates than a static IRT model, as implemented for the latent estimate of personalism used throughout the book.

Despite the possibility that the independence may be violated for some indicators, we emphasize that the applications using the IRT estimate in the book explicitly model non-independence within leaders or regimes using clustered errors.[8] We also directly model panel heterogeneity among distinct leaders (or regimes) to isolate variation over time within them.[9] These applied modeling choices are attempts to deal with non-independent data in particular applications that use personalism as a dependent or independent variable.

IRT estimates

Figure 3 plots the item information functions (IIF) for the eight items that contribute to the latent estimate of *personalism*, or θ .[10] The vertical axis measures the item discrimination parameter: higher values indicate more information in the latent estimate over a smaller range of θ values. The horizontal axis corresponds to the difficulty parameter: larger values indicate items for which observations have a higher estimate of θ. If the IRT model is accurately estimating latent *personalism*, more difficult items are those for which an observation must be highly personalist to observe a 1 for this item. This parameter captures how well an item splits high and low *personalism* cases at a particular point in the latent space.

The left plot of Figure 3 shows the IIF for four items that pertain to the support party and personnel. The item with the most information is `officepers`, while that with the least is `createparty`. Personal appointment to office (`officepers`) is the least discriminating, which is to be expected since it is the most commonly observed personalist item. Party creation by the leader (`createparty`) is the most discriminating item. The right plot shows the IIF for item relating to the military and security apparatus. The item with the most information (`sectyapp_pers`) is also the least discriminating, while the item with the least information (`paramil_pers`) is the most discriminating.

The plots in Figure 3 indicate that the information contained in the items splits the observations all along the latent space, as can be observed by noting that the peaks of the IIFs are spread across different values of θ. This means the items have different difficulty parameter estimates across most of the latent space.

There are additional items in the raw data set that are conceptually related to personalism:

[7] Unsuprisingly, leaders that last longer (i.e. more than five years) have higher personalism scores (at least as measured assuming local independence) than do short-lived leaders. The 244 leaders that last less than five years have an average personalism score of 0.19 while the 261 leaders that last longer have an average personalism score of 0.40. This difference is statistically significant at the 0.001 level (two-tailed test with 505 leader-observations).

[8] The IRT model can incorporate dependence by clustering on leader in the estimates of the latent trait.

[9] One drawback to using a dynamic model in this application is that this approach, by design, smooths the estimated latent variable over time within units. This means, for example, that an increase in personalism in year t based on the historical evidence will increase personalism in year *t-1*. The smoothed estimates will therefore be less precise in picking up sharp changes in personalism in the leader time-series.

[10] While not shown here, the test characteristics curve indicates that 95 percent of the observations fall between 0 and 7, meaning very few observations are positive for all eight items that contribute to the latent estimate.

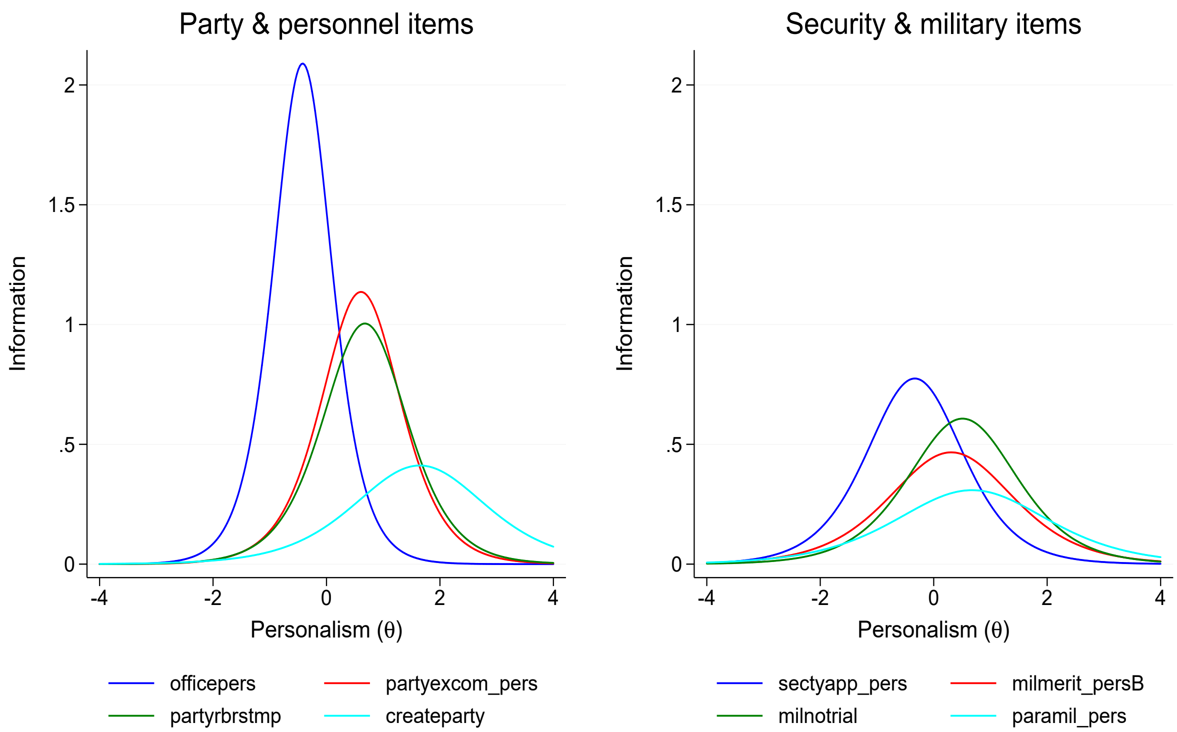


Figure 3: ***Item information functions***.

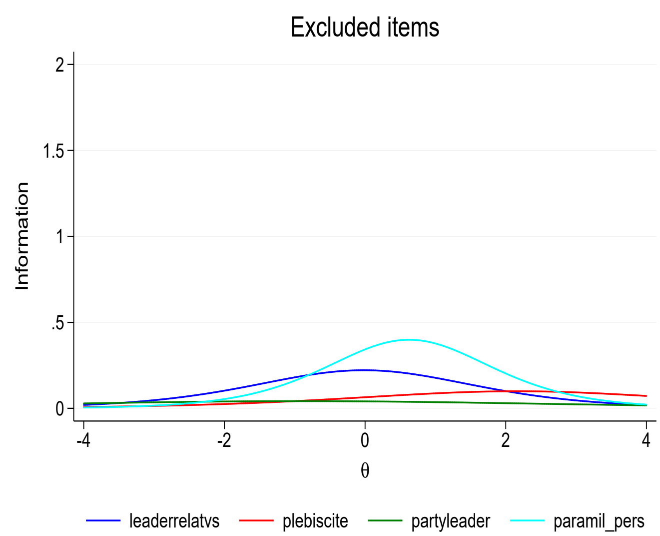


Figure 4: ***Item information functions***.

`leaderrelatvs`, `plebiscite`, and `partyleader`. These measure whether the leader appoints family members to high offices, whether the leader rules by plebiscite, and whether the leader is also the leader of the supporting political party (if there is one). To see if adding these items to the IRT model contributes information we look at the IFF's for these items after fitting the IRT (now with 11 items instead of the original 8).[11] Figure 4 shows the IIF for three additional items as well as the IIF for `paramilitary`, which has lowest information among the 8 items used for the initial IRT model. The plot shows that these other three items contribute very little additional information to the latent personalism estimate. For this reason, we excluded them from the model featured in the book. Thus, our selection of variables for the construction of the personalism measure is based on criteria that balance parsimony (just 8 variables are used!) with marginal gains in distinct information.

Comparing measures of personalism

In this section, we assess convergent validity by comparing our latent estimate of personalism to existing efforts to measure the same concept: the categorical indicator of personalist regime introduced by Geddes (1999, 2003) and updated in Geddes, Wright and Frantz (2014) and Weeks (2012). The latter builds on the former by using the original coding details that Geddes (1999) employed to measure categorical personalism and updating information for some specific leaders.

Figure 5 displays the pairwise correlations for the different measures.[12] The first column shows the correlations between the categorical GWF variable for personalist regime. It is correlated with the Weeks measure at 0.56 and the time-varying GWF data at roughly 0.40. The second column shows the correlations between the latent GWF measure used in the book and the other measures. It is highly correlated with the dimension constructed from factor analysis and the aforementioned latent measure that incorporates information from three additional variables (`leaderrelatvs`, `plebiscite`, and `partyleader`, meaning 11 items are included in the IRT model). Both of these correlation coefficients are greater than 0.99.

The correlations therefore suggest that various methods of combining the items into a latent index (IRT with 8 items, IRT with 11 item, factor analysis) yield very similar estimates of the level of personalism. Further, the latent GWF measure is not particularly highly correlated with extant measures, such as the categorical GWF variable and the Weeks data.

Next, in Figure 6, we compare the "within" variance for three measures of personalism: Latent (GWF), Categorical (GWF), and Weeks. The vertical axis measure the ratio of "within" to "total" variance[13] for different cross-section units (country, regime, and leader), while the horizontal axis measures the total variance. Higher values on the vertical axis indicate that more of the total variation in the variable is contained within the cross-section unit (e.g. over time within a country/regime/leader) rather than between cross-section units. This comparison is important because

[11] We used the original military promotion item, which has three ordered values, instead of the collapsed item which has only two values. To model the ordered item to the model we mixed a logistic link function with an ordered link function using a graded response model for the latter.

[12] Correlations between various GWF variables are for 4,591 observations; correlations with Weeks are based on 2,553 observations.

[13] Letting N be the total number of observations, k be the number of groups, n be the number of observations within a group, $\bar{x}$ be the group mean, and $\bar{\bar{x}}$ be the grand mean, then the "within" variance is: $\frac{\sum_{j=1}^{k}\sum_{i=1}^{n_i}(x_{ij}-\bar{x_j})^2}{N-k}$. Total variance, using this notation, is $\frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{\bar{x}})^2}{N-1}$.

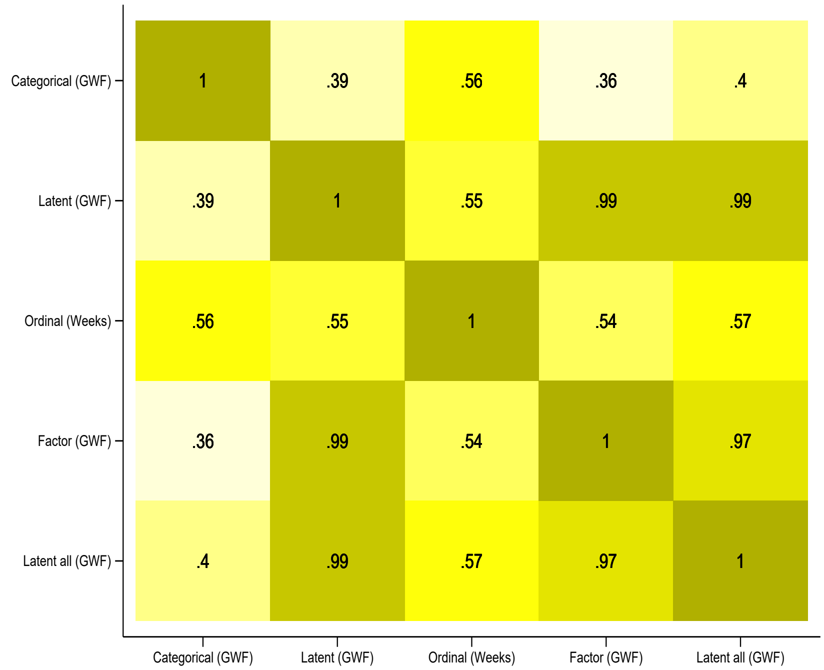


Figure 5: ***Correlation among personalism measures.***

in many of the applications in the book, we use the personalism measure in a statistical model that isolates variation over time within a cross-section unit (i.e. fixed effects models).

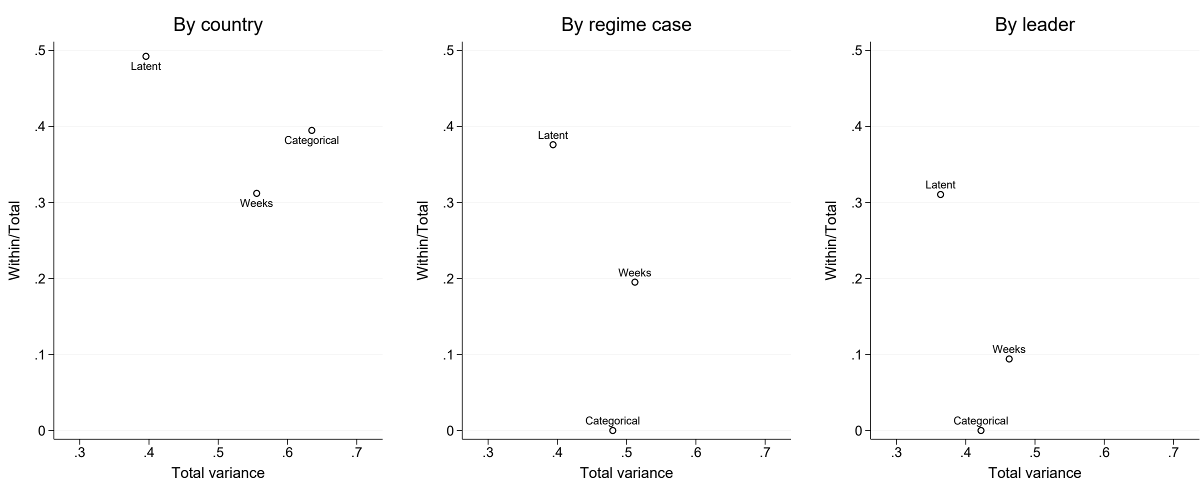


Figure 6: ***Variance decomposition.***

The left plot shows that roughly half of the variation in the Latent measure is within countries (and half is between countries). Roughly 40 percent of the variation in the Categorical measure is within countries, while just over 30 percent of the variation in the Weeks measure is within countries. The same pattern holds when we examine regimes and leaders as the cross-section units. Note that by construction the Categorical measure is constant within regimes and hence leaders. That is, it does not measure *any* within variation. In contrast, the Latent measure still has substantial within regime and within leader variation. Indeed, there is over three times as much within-leader variation in the Latent measure as in the Weeks measure. The variance decomposition plots therefore show that this measure of personalism may be useful in applications that require modeling unit heterogeneity with fixed or random effects.

Next we show how personalism estimates vary over the duration of a regime for different categories of regime (Party, Military, Personalist, and Monarchy), as outlined in GWF (2014). The top half of Figure 7 contains plots for each regime category. The horizontal axes plot regime duration in years, while the left vertical axes capture the number of observations stratified across regime during. The shaded gray bars show the distribution of regime duration.

First note that the frequency of regimes in each category varies substantially. Regimes in the Party category are numerous (roughly 70) and many endure as long as 25 years (roughly 40). Military regimes are also numerous but far fewer are long-lasting; only a handful live beyond 20 years. Personalist regimes are the most common, but many do not survive as long as Party regimes. Finally, Monarchies are rare but long lived. The steepness of the decline in the distribution of duration time (shaded grey) is a rough proxy for the hazard rate: steep declines mean the hazard rate is high and regimes are relatively short lived.

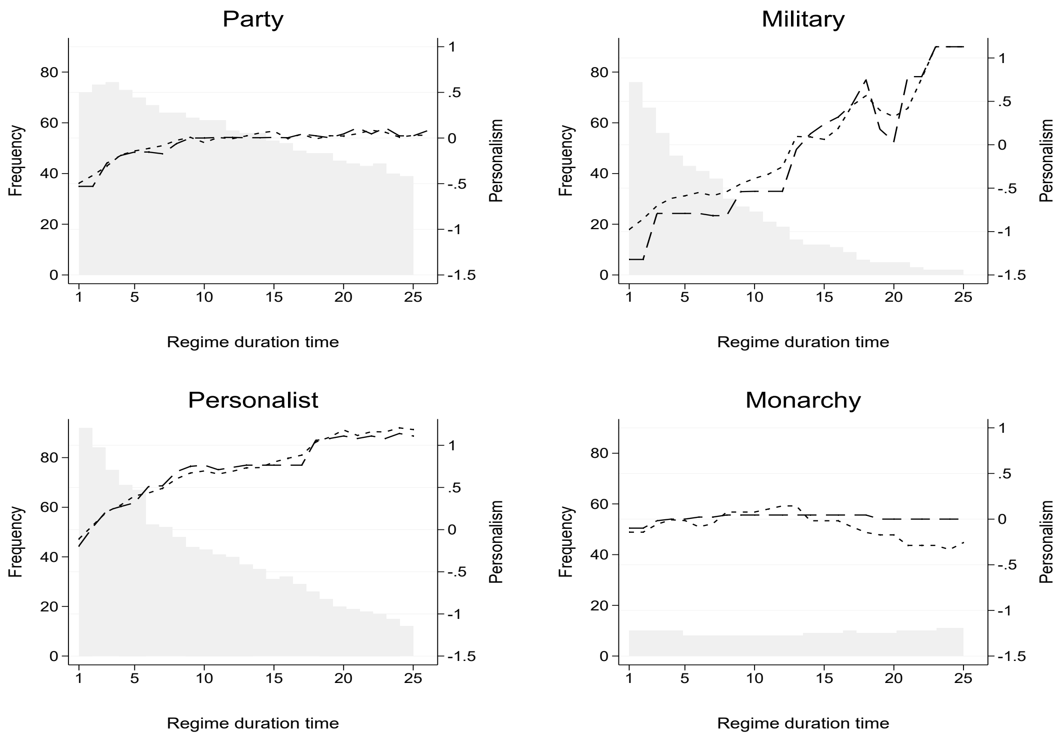
The solid (dash) lines in Figure 7 show the mean (median) levels of Latent personalism, with corresponding levels depicted on the right vertical axes, for each duration-year in each category of regime. Note that the vertical axes are scaled identically to facilitate visual comparison. In Party regimes, personalism is roughly -0.5 – on average in – the first years of the regime, rising to about 0 after 10 years and remaining constant afterwards. In Military regimes, personalism scores are very low – on average – in the first years of the regime but increase substantially over time. Though only a few regimes in this category last beyond 15 years, those that do tend to have high personalism scores. Personalist regimes have middling levels of personalism early in the regime but this increases substantially over time: long-lived Personalist regimes tend to have very high personalism scores. Finally, monarchies have middling levels of personalism that remain relatively constant over time.

The plots in the top half of Figure 7 reveal that much of the "within" variation occurs in regimes categorized as Military or Personalist in Geddes (1999) and GWF (2014). This also means that categorical indicators of "regime type" used in prior research[14] mask considerable variation over time in the level of personalism within regimes.

The bottom half of Figure 7 shows a similar set of trends but measures average personalism over leader tenure (or duration) instead of regime duration. These patterns reveal that the level of personalism increases, on average, as leaders last longer in power – but not in Monarchies. Further, the (positive) slopes of the lines differ in the other regime categories: the slope is steepest in Military regimes and more gradual in Party regimes. This indicates that using leader duration – as Wahman, Teorell and Hadenius (2013) propose – is a poor proxy for the level of personalism, at least as we have measured it here. For example, during their first years in power, Military regime leaders have relatively low personalism scores (on average) compared with Personalist regime leaders. Similarly, after 15 years in power leaders in Party regimes and Monarchies have, on average,

[14]See, for example, Peceny, Beer and Sanchez-Terry (2002), Weeks (2008), Wright (2009), Gurses and Mason (2010), Aksoy, Carter and Wright (2012), Conrad, Conrad and Young (2014), Wilson and Piazza (2013), and Frantz and Kendall-Taylor (2017).

(a) Regimes



(b) Leaders

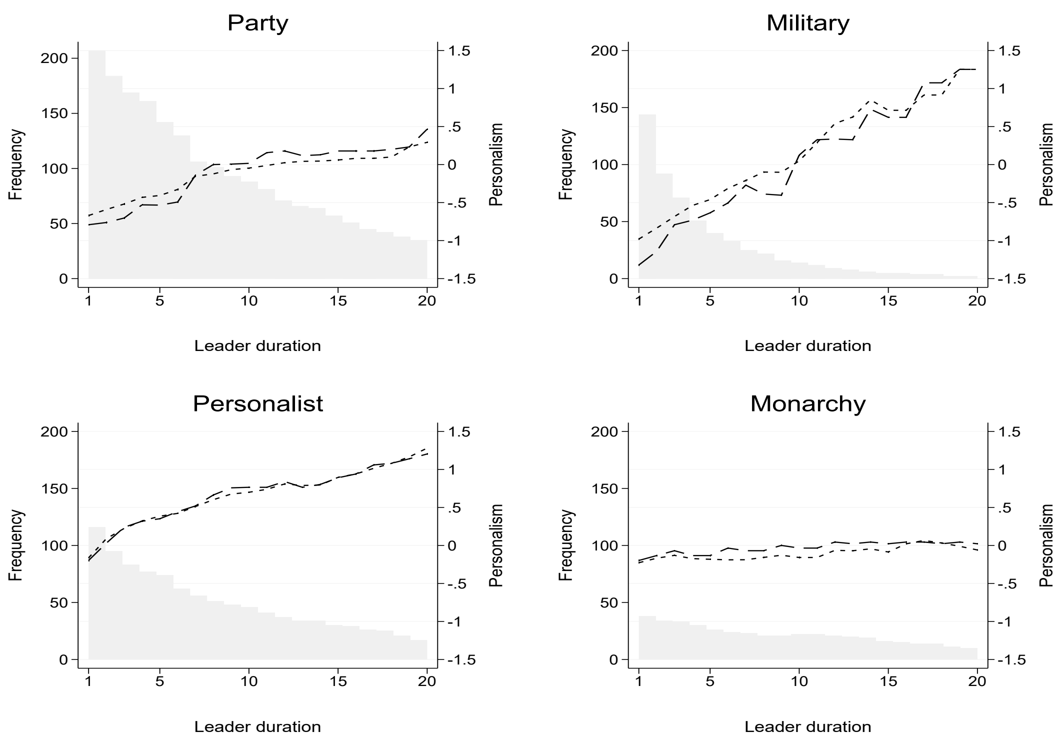


Figure 7: ***Latent personalism over regime duration, by categorical type.***

middling levels of personalism (roughly `0`) while leaders in Military and Personalist regimes have much higher levels (approaching `1`).

Paths forward

The latent measure of personalism introduced in the book combines information from real-world political events.[15] The events themselves are objective; the process of determining whether the event qualifies as an indicator of an attribute of personalism is less so.

The GWF data collection entailed reading a substantial amount of material to organize and systematize information that exists in case studies, biographies, and news reports. In this respect, it differs from recent research that aggregates either existing off-the-shelf data sources (Gandhi and Sumner, 2016) or the subjective assessment of a concept from multiple human experts (Coppedge et al., 2016). The human hours required to compile the GWF data set were many but remain uncounted; the data collection process lasted over eight years. This measure of personalism is therefore extremely costly to create, but we hope provides researchers with new variation to explore in empirical analysis. This is perhaps particularly important given evidence suggesting a rise in personalism around the globe (Kendall-Taylor, Frantz and Wright, 2016, 2017).

We see three potential paths forward in this research agenda. First, future research might explore using more sophisticated methods for aggregating the existing data. It could probe the local independence assumption using dynamic latent models, or consider other techniques such as nonlinear dimensionality reduction. Directly modeling structural missingness may also be a fruitful way forward.

Second, future research might explore other approaches to collecting similar data. For example, scholars could aggregate multiple subjective (human) perceptions of the concept of personalism to produce a consistent scaled measure. Another approach is to employ machine learning that uses a corpus of texts to quickly find (real-world) textual information that is then combined to approximate the measure of personalism we have constructed (Bieler, Ulfelder and Wright, 2017; Minhas, Ulfelder and Ward, 2015). This approach requires an existing measure with which to train the learning algorithm but can relatively cheaply and quickly process new information to extend coding into more recent periods (our coding ends in 2010). Finally, researchers with substantial financial resources might explore using paid country experts – including academics, journalists, and independent researchers – to collect objective information pertaining to a concept from the case study literature. Using country experts would greatly speed up the data collection process without jeopardizing data quality (Wright, 2015; Morgenbesser, 2017).

Third, scholars might extend the conceptual exercise to more narrow and distinct dimensions of non-institutionalized politics. For example, (Song, 2017) uses measures of personalism related to the military and security sector from the GWF data to construct a latent estimate of this subdimension of personalism as conceptually distinct from the leader's power relative to the support party.[16] In a similar vein, new research could, for example, extend measures of personalism to measure the leader's family's penetration of key economic sectors. Finally, the data collection effort might be extended to more narrow units in time (Wright, 2015). This might, for example,

[15] For example, in North Korea in 1966 the regime leader, Kim Il Sung, used the October Party Congress to reorganize the Korean Worker's Party (KWP) leadership structure, resulting in his de facto control over the party executive committee. By 1968, Kim "faced no further challenges from within the KWP" (Buzo, 1999, 34); the party had become a rubber-stamp for Kim's decisions.

[16] See Song and Wright (2017) as well.

entail coding historical events that meet the coding criteria for manifest indicators of personalism on a monthly basis rather than on a yearly basis. Doing this for periods in the ancient past (the 1950s!) might be too difficult but doing so for the contemporary period might be realistic.

Other researchers, having read this far, will undoubtedly have many new ideas of their own.

References

Aksoy, Deniz, David B Carter and Joseph Wright. 2012. "Terrorism in Dictatorships." *The Journal of Politics* 74(3):810–826.

Bieler, John, Jay Ulfelder and Joseph Wright. 2017. "Report on Automated Production of Annual Data on Political Regime Types.".

Buzo, Adrian. 1999. *The Guerilla Dynasty: Politics and Leadership in North Korea*. Boulder, CO: Westview Press.

Conrad, Justin, Courtenay Conrad and Joseph K. Young. 2014. "Tyrants and Terrorism: Why some Autocrats are Terrorized while Others are Not." *International Studies Quarterly* 58(3).

Coppedge, Michael, John Gerring, Staffan I Lindberg, Svend-Erik Skaaning, Jan Teorell, Frida Andersson, Kyle L Marquardt, Valeriya Mechkova, Farhad Miri, Daniel Pemstein et al. 2016. "V-Dem Methodology v6.".

Frantz, Erica and Andrea Kendall-Taylor. 2017. "Pathways to democratization in personalist dictatorships." *Democratization* 24(1):20–40.

Gandhi, Jennifer and Jane Lawrence Sumner. 2016. Measuring Personalism in Non-Democracies. Technical report Paper presented at Varieties of Democracy Latent Variable Week.

Geddes, Barbara. 1999. "What Do We Know About Democratization After Twenty Years?" *Annual Review of Political Science* 2:115–144.

Geddes, Barbara. 2003. *Paradigms and Sand Castles*. Ann Arbor: Univ. of Michigan Press.

Geddes, Barbara, Joseph Wright and Erica Frantz. 2014. "New Data Set: Autocratic Breakdown and Regime Transitions." *Perspectives on Politics* 12(2):313–331.

Geddes, Barbara, Joseph Wright and Erica Frantz. 2017. *How Dictatorships Work*.

Gurses, Mehmet and T. David Mason. 2010. "Weak states, regime types, and civil war." *Civil Wars* 12(1-2):140–155.

Kendall-Taylor, Andrea, Erica Frantz and Joseph Wright. 2016. "The New Dictators: Why Personalism Rules." *Foreign Affairs* .

Kendall-Taylor, Andrea, Erica Frantz and Joseph Wright. 2017. "The Global Rise of Personalized Politics: It's Not Just Dictators Anymore." *The Washington Quarterly* 40(1):7–19.

Kenwick, Michael. 2017. "Is Civilian Control Self-Reinforcing? A Measurement-Based Analysis of Civil-Military Relations.".

Martin, Andrew D and Kevin M Quinn. 2002. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." *Political Analysis* 10(2):134–153.

Minhas, Shahryar, Jay Ulfelder and Michael D Ward. 2015. "Mining texts to efficiently generate global data on political regime types." *Research & Politics* 2(3):2053168015589217.

Morgenbesser, Lee. 2017. "Misclassification on the Mekong: the origins of Hun Sen's personalist dictatorship." *Democratization* pp. 1–18.

Peceny, Mark, Caroline C. Beer and Shannon Sanchez-Terry. 2002. "Dictatorial Peace?" *American Political Science Review* 96(1):15–26.

Song, Wonjun. 2017. "Personalization of Autocratic Security Forces and Coups.".

Song, Wonjun and Joseph Wright. 2017. "The North Korean Autocracy in Comparative Perspective.".

Wahman, Michael, Jan Teorell and Axel Hadenius. 2013. "Authoritarian regime types revisited: updated data in comparative perspective." *Contemporary Politics* 19(1):19–34.

Weeks, Jessica L. 2008. "Autocratic Audience Costs: Regime Type and Signaling Resolve." *International Organization* 62(1):35–64.

Weeks, Jessica L. 2012. "Strongmen and Straw Men: Authoritarian Regimes and the Initiation of International Conflict." *American Political Science Review* 106(2):326–347.

Wilson, Matthew C. and James A. Piazza. 2013. "Autocracies and Terrorism: Conditioning Effects of Authoritarian Regime Type on Terrorist Attacks." *American Journal of Political Science* 57(4):941–955.

Wright, Joseph. 2009. "How foreign aid can foster democratization in authoritarian regimes." *American Journal of Political Science* 53(3):552–571.

Wright, Joseph. 2015. "Leadership-Security Ties Codebook.".

Wright, Joseph. 2017. "The Latent Characteristics That Structure Autocratic Rule.".