

Measuring party system institutionalization in democracies

Wooseok Kim 
University of Glasgow, UK

Party Politics
2023, Vol. 0(0) 1–12
© The Author(s) 2023



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13540688231211241
journals.sagepub.com/home/ppq



Abstract

Party system institutionalization (PSI) is regarded as a critical underpinning of democracy. However, the systematic study of PSI in democracies is constrained by weaknesses in existing measures, which are limited in coverage or comprehensiveness, and do not account for the latent nature of the concept, measurement error, and non-random missing data. This article presents a novel measure of PSI that uses a Bayesian latent variable measurement strategy to overcome extant measurement issues. The subsequent measure not only offers unmatched coverage and has demonstrated validity, but also exhibits more robust empirical associations with a range of outcomes related to the performance of democracy than existing measures. The measure should facilitate more integrated research on the causes and consequences of PSI in democracies around the world.

Keywords

democracy, institutionalization, party systems, political parties, latent variable

Introduction

Party system institutionalization (PSI) refers to the extent to which “a stable set of parties interacts regularly in stable ways” (Mainwaring, 2018:4) and come to “complete, colligate, and collaborate” in a predictable manner (Casal Bértoa and Enyedi, 2021:17). Stable and predictable party systems are regarded as a critical underpinning of democracy as they generate information about who the parties are, what they stand for, and how they might behave. Under-institutionalized party systems that exhibit frequent reshuffling of parties and erratic interparty interactions are thought to undermine accountability (Jensenius and Suryanarayan, 2022; Ridge, 2022; Robbins and Hunter, 2011; Schleiter and Voznaya, 2018), impede interparty coordination (Bernhard et al., 2020; Hicken, 2016), and diminish the capacity of governments to implement consistent policies that promote social and economic welfare (Hicken, 2016; Mauro, 2022; Robbins, 2010; Tommasi, 2006). Such deficiencies can erode the legitimacy of democratic institutions and create openings that anti-democratic actors can exploit to trigger the breakdown of democracy (Mainwaring, 2018; Mainwaring and Scully, 1995).

Despite this widely entrenched belief that democracy is unworkable—and perhaps even unsustainable—without institutionalized party systems, the supporting evidence is

not as conclusive as one might expect (Casal Bértoa, 2017). This is in part due to weaknesses in existing measurement approaches. Given the significant challenges associated with gathering comparative party system data, empirical studies often employ dissimilar indicators or focus on disjoint regions, which has at times produced conflicting results (Enyedi and Casal Bértoa, 2020). Other studies reduce data demands by mono-operationalizing PSI with Pedersen’s (1979) index of electoral volatility or the age of the main parties, but these measures only partially or indirectly capture PSI. While recent works have advanced the measurement of the concept, the coverage of the subsequent measures remains limited (e.g., Casal Bértoa and Enyedi, 2021; Chiaramonte and Emanuele, 2022; Rodriguez and Rosenblatt, 2020). Moreover, no measure directly accounts for the latent nature of PSI, the accompanying measurement error, and non-random missing data, but overlooking such issues can generate misleading inferences. Together, these

Paper submitted 17 January 2023; accepted for publication 16 October 2023

Corresponding author:

Wooseok Kim, University of Glasgow, Bute Gardens, Glasgow G12 8RT, UK.

Email: wooseok.kim@glasgow.ac.uk

unresolved measurement challenges have inhibited the development of systematic understandings of how PSI is related to party building and collapse, on one hand, and democratic consolidation and backsliding, on the other, across regions, contexts, and over time (Casal Bértoa, 2018).

This article fills this gap by presenting a novel measure of PSI that addresses these measurement issues. Since PSI is a latent concept that cannot be directly observed or measured (Casal Bértoa and Enyedi, 2021; Mainwaring, 2018), I employ a Bayesian latent variable measurement strategy, which leverages information from manifest (observable) indicators of the concept to estimate PSI. I overcome data limitations by gathering extensive data on the partisan composition of legislatures and governments, and use this data in conjunction with the Varieties of Democracy (V-Dem) dataset (Coppedge et al., 2020) to construct five manifest indicators of PSI that cover 96 democracies from 1945 to 2018. The subsequent Party System Institutionalization Scores (*PSI Scores*) measure not only offers unmatched coverage and is more comprehensive than single indicators, but also aligns with the latent nature of the concept, quantifies measurement uncertainty, and accounts for non-random missing data. Nonetheless, the utility of a measure depends more critically on its validity. As such, I conduct numerous validity tests while comparing the measure's performance to existing measures, and the results strongly favor the *PSI Scores*. Notably, this measure is a better predictor of a range of outcomes related to the performance of democratic institutions, and should be a useful empirical tool that can advance more integrated research on the causes and consequences of PSI across democracies.

Measuring PSI: Existing approaches

Following Mainwaring and Scully's (1995) seminal volume on party systems in Latin America, a wave of empirical studies highlighted the potential utility of the concept of PSI for explaining variations in the performance of democracies (e.g., Coppedge, 1998; Croissant and Völkel, 2012; Jones, 2010; Kuenzi and Lambright, 2001; Lindberg, 2007; Mainwaring and Torcal, 2006; Meleshevich, 2007; Stockton, 2001; Weghorst and Bernhard, 2014). However, given data constraints, these studies tend to employ dissimilar indicators or focus on disjoint regions, which makes it difficult to systematically compare results, explain discrepancies, and build unified theories. Furthermore, the usual strategy to aggregate multiple indicators is to average them into an index, but there is usually no justification given as to whether this strategy is appropriate for measuring PSI (Luna, 2014).

A popular workaround to reduce data demands has been to mono-operationalize PSI with electoral volatility or the age of the main parties, which can permit more

encompassing examinations of PSI (e.g., Mauro, 2022; Ridge, 2022; Robbins, 2010; Robbins and Hunter, 2011; Schleiter and Voznaya, 2018).¹ However, using a single indicator to measure a multi-faceted concept such as PSI can introduce bias in cross-space/cross-time analysis (Munck and Verkuilen 2002), which undermines the principal advantage of these measures. Furthermore, electoral volatility is a noisy measure (Casal Bértoa et al., 2017) that only accounts for the electoral/legislative arenas and, as discussed later, even discards pertinent information about stability and predictability by focusing on election-to-election changes. On the other hand, the age of the main parties is more closely aligned with *party* institutionalization since the measure simply aggregates characteristics (i.e., the age) of individual parties. Older parties may be more likely to be institutionalized, but this does not necessarily mean that interparty interactions will be stable and predictable. Hence, this measure may be less applicable for testing theories that specifically concern PSI.

Recent works on PSI address some of these weaknesses, and Casal Bértoa and Enyedi (2021) provide one path-breaking contribution in this regard. The authors build on Mair (1997), who argues that the stability and predictability of patterns of government formation is a core characteristic of PSI since it captures key interactions between the most relevant parties. When measuring PSI, the authors accordingly deviate from existing approaches that typically focus on the electoral/legislative arenas. Instead, they construct the Party System Closure Index, which measures the stability and predictability of government formation based on the extent to which parties adhere to existing alliances and form familiar cabinets. Importantly, the authors find that their index offers additional insights about PSI that may not be readily apparent from looking at electoral patterns. Their work suggests that patterns of government formation encompass relevant and distinct information about PSI, and could be incorporated into more comprehensive measures of the concept.

However, one drawback of this index is its neglect of the electoral/legislative domains, which precede government formation and thus remain important arenas of interparty competition (Chiaromonte and Emanuele, 2022; Mainwaring, 2018). To build a more complete measure, Chiaromonte and Emanuele (2022) standardize and average time-weighted volatility measures across the electoral, legislative, and government arenas to form their own index. This index offers clear advantages as it tracks interparty competition across multiple arenas, and accounts for long-term stability by time-weighting observations from the past three elections/legislatures. However, the inclusion of both electoral and legislative volatility in the index may be problematic since the two measures are very highly correlated and convey almost identical information about PSI, which mechanically underweights the importance of

government volatility. In addition, the index does not account for measurement uncertainty even though volatility calculations can be quite noisy (Casal Bértoa et al., 2017).

In contrast to the preceding two works, Rodriguez and Rosenblatt (2020) recast PSI as having a necessary and sufficient conceptual structure that requires both (1) stability and predictability, and (2) the capacity of the party system to incorporate new societal demands. When constructing their accompanying measure, the authors correctly note that averaging indicators would misrepresent their concept since this assumes an additive conceptual structure. Instead, the authors use a combination of interactions and geometric/arithmetic means to reflect the necessary and sufficient conditions within their concept. Although this resulting measure more faithfully adheres to their underlying concept, the authors' conceptualization runs counter to recent trends that treat PSI as strictly revolving around stability and predictability (e.g., Casal Bértoa and Enyedi, 2021; Mainwaring, 2018),² and their second dimension might be more rightly viewed as one cause of PSI than as one of its internal components (Chiaramonte and Emanuele, 2022).

While these novel measures provide valuable insight into how PSI could be more effectively measured, their coverage remains limited, though this is understandable as they carry more burdensome data requirements.³ On the other hand, the Party Institutionalization (PI) Index in the V-Dem dataset offers almost universal coverage and has been utilized in more expansive studies of PSI (e.g., Mauro, 2022; Ridge, 2022).⁴ The index averages various indicators of party characteristics but—as its name suggests—it specifically measures the institutionalization of parties rather than the party system. Although PI and PSI are intimately intertwined, they may not always be complementary (Randall and Svåsand, 2002). Thus, conflating these concepts and associated measures can impede our understanding of how the two processes might be related, and whether it is the institutionalization of individual parties or the party system that drives outcomes of interest (Casal Bértoa, 2017).

Additional measurement issues have also been frequently overlooked. PSI is a latent concept since stability and predictability are a function of perceptions and expectations, which means that it cannot be directly observed or measured (Casal Bértoa and Enyedi, 2021; Mainwaring, 2018), but there has been no attempt to explicitly measure PSI as a latent variable. In turn, this has meant that estimates do not account for measurement uncertainty even though manifest indicators of latent concepts likely contain varying degrees of noise. Furthermore, the difficulties of acquiring comparative party system data—particularly for less institutionalized party systems—often lead to missing observations that are unlikely to be missing at random. Although such issues do not necessarily pose insurmountable hurdles, leaving

them unaddressed can lead to imprecise estimates and misleading inferences.

Measuring PSI as a latent concept

To deal with the aforementioned issues, I use a Bayesian latent variable measurement approach, which estimates latent levels of PSI by drawing on the common variance between its manifest indicators. This strategy offers numerous advantages for measuring the concept at hand. First, it incorporates information across multiple indicators, which should produce a measure that is more useful for cross-space/cross-time analysis relative to single indicators such as electoral volatility or the age of the main parties (Munck and Verkuilen, 2002). Second, it follows Rodriguez and Rosenblatt's (2020) example by aligning measurement with the underlying concept. Third, it provides a way to quantify measurement uncertainty, which is an important feature given the latent nature of the concept. Fourth, the measurement model can flexibly handle non-random missing data to mitigate bias.

Manifest indicators

The manifest indicators are selected based on two broad criteria. First, manifest indicators should capture observable characteristics of party systems that are a function of underlying levels of PSI.⁵ Some indicators may covary with PSI but represent features of related but distinct concepts. Including such indicators in the measurement model could generate imprecise estimates and produce a conflated measure that is less useful for theory-testing. As such, indicators of concepts such as party institutionalization (e.g., party age) and the quality of democracy (e.g., legitimacy of elections), or those that mainly focus on actors external to the party system such as the electorate (e.g., partisan identification) are not included in the measurement model.⁶

The second criteria used to select the manifest indicators are data coverage and generalizability since the goal is to develop a robust measure of PSI that also encompasses a global sample of democracies. Some indicators may convey pertinent information about PSI, such as the consistency in the ideological configuration of parties or patterns of legislative voting. However, such indicators are not included in the measurement model since requisite data are not consistently available for many countries. For similar reasons, indicators that are only applicable to a subset of democracies—notably those related to presidential elections—are also not included.

The manifest indicators used in the measurement model—*Aggregate Legislative Volatility*, *Aggregate Government Volatility*, *Minor Party Performance*, *Party Distinctiveness*, and *Party Switching*—satisfy these criteria.⁷ Moreover, the analyses presented later in this article

demonstrate that these indicators contain sufficient information to produce overall estimates of PSI that perform quite well in validity tests.

The first three indicators focus on the composition of the party system. In institutionalized party systems, interparty competition revolves around well-established parties that engage in consistent patterns of interactions across the electoral/legislative and government arenas (Casal Bértoa and Enyedi, 2021; Chiaramonte and Emanuele, 2022; Mainwaring, 2018). In such cases, the composition of parties that win seats and gain access to government offices should be generally stable over time. The *Aggregate Legislative Volatility* and *Aggregate Government Volatility* indicators modify conventional volatility calculations to capture such long-term stability.

While the volatility indicators reflect the stability of partisan composition, the *Minor Party Performance* indicator instead focuses on the distribution of the seats among parties in the legislature, and more specifically on the cumulative seat shares won by minor parties (and independents). As party systems become institutionalized, political competition comes to be channeled through and largely restricted around a set of well-established parties. This raises the barriers of entry for other contenders and increases the costs of challenging these main parties (Casal Bértoa and Enyedi, 2021; Mainwaring, 2018; Meleshevich, 2007), which diminishes the likelihood that many minor parties will both compete and successfully win seats in elections. As such, PSI should be associated with higher levels of major party dominance and, correspondingly, a reduced presence of minor parties in the party system.⁸

PSI also alters the expectations of actors within the party system (Mainwaring, 1999), and thus their subsequent behaviors can also inform about latent levels of PSI. As interparty interactions become more stable and predictable, actors come to expect such patterns to persist and develop longer time horizons (Hicken, 2016; Mainwaring, 2018). This magnifies the costs of short-term opportunistic behaviors that defy such expectations and generate uncertainty (Tommasi, 2006), which weakens the appeal of ephemeral parties or those that frequently change their policy platforms (Lupu and Riedl, 2013). Consequently, PSI fosters *Party Distinctiveness* by generating stronger incentives for parties to carve out clear and consistent identities within the party system (Greene and Sánchez-Talanquer, 2018; Jones, 2010). PSI should also have comparable implications for the behaviors of individual politicians. In particular, the reputational costs that legislators incur when switching parties increase as parties become more distinct and actors develop longer time horizons. Hence, PSI should also be associated with reduced occurrences of *Party Switching* within the party system (Desposato, 2006; Kreuzer and Pettai, 2009).⁹

These indicators offer related but distinct information about PSI, and requisite data are also generally available across most democracies in the post-WWII period. The following section discusses the data sources and the construction of the indicators.

Data

To construct the aggregate volatility indicators, I gather extensive data on the partisan composition of around 1050 lower-house legislatures and 1300 cabinets across 96 post-WWII democracies.¹⁰ These indicators modify Pedersen's (1979) index of electoral volatility:

$$EV = \frac{\sum_{i=1}^N |p_{i,t} - p_{i,t-1}|}{2}$$

where $p_{i,t}$ is the seat (or vote) share obtained by party i in election t . Pedersen's index captures the extent to which parties win similar shares of seats across consecutive elections, and ranges from 0 to 1 with lower values associated with greater PSI. Although it has never been discussed as such, the index is in fact one specific application of dissimilarity measures that are widely employed in the ecology literature. Dissimilarity measures are used to assess the resemblance of different communities by leveraging information about the number and diversity of the species in those communities. Pedersen's index is essentially a dissimilarity measure with communities replaced by party systems, species by parties, and similarity in species (across space) cast as stability in partisan composition (over time). However, measures of dissimilarity in ecology are often calculated across all pairs of communities in the sample and—as I demonstrate in Appendix C—this set of dissimilarity scores contains meaningful information about the similarity of *all* communities. In other words, Pedersen's index essentially discards relevant information about long-term stability by only considering two consecutive elections.

Therefore, I calculate the *Aggregate Legislative Volatility* indicator by adapting existing applications of dissimilarity measures. At each year t in country i , I use seat share data¹¹ from the most recent election to calculate a linear weighted moving average of the volatility scores¹² between year t and each of the twenty years (or $t-1$ if $t < 20$) preceding t .¹³ Substantively, the indicator summarizes how similar the partisan composition of the legislature at year t is to those from the past twenty years, with recent legislatures and those that lasted for a longer period receiving greater weight.¹⁴ I also calculate the *Aggregate Government Volatility* indicator following a similar procedure, weighting parties in government by their legislative seat shares.^{15,16} Both indicators range from 0 to 1, with lower scores indicating greater long-term stability. The *Minor Party*

Performance indicator is calculated by summing the seat shares won by minor parties (i.e., those that secure less than two percent of the seats) and independents, which are excluded from the legislative volatility calculations due to data limitations.¹⁷

Data for the *Party Distinctiveness* and *Party Switching* indicators are sourced from the V-Dem dataset. The former indicator is constructed using a Bayesian Item Response Theory measurement model that aggregates expert assessments of whether parties “have publicly available platforms that are publicized and relatively distinct from one another.”¹⁸ The indicator is rescaled to a unit interval so that its scale is equivalent to the other indicators. The latter indicator represents expert assessments of the percentage of members of the national legislature who change or abandon their parties in between elections.¹⁹

Measurement model

In the model, $x_{k,i,t}$, which refers to manifest indicator x_k for country i in year t , is specified to be a linear function of latent levels of PSI ($\xi_{i,t}$) and random noise, i.e., $x_{k,i,t} = c_k + \lambda_k \xi_{i,t} + \epsilon_k$. Coefficients are assigned Normal(0,2) priors, and the error terms are specified to have mean zero independent normal priors with InverseGamma(1,1) as the prior for the variance. Since latent variable models require additional constraints for identification (Bollen, 1989), I reverse the scale of all indicators except the *Party Distinctiveness* indicator so that higher values are associated with more institutionalized party systems, constrain λ_k to be positive, and set the priors for the latent factors ($\xi_{i,t}$) as Normal(0,1).

Despite extensive efforts and consultation across numerous sources, it was not possible to compile consistent cabinet composition data for some democracies. *Aggregate Government Volatility* cannot be calculated in such cases since the indicator requires complete data across all twenty-year periods. These observations, which account for around fourteen percent of the total observations, could be excluded or their missingness ignored—as is usually the case when averaging indicators into an index—but this could bias estimates since missingness is likely to be non-random, i.e., more likely when party systems are under-institutionalized and consistent cabinet data are harder to come by given the frequent entry and exit of parties. As such, I adopt the approach presented by Lee (2007) and Arel-Bundock and Mebane (2011), which leverages missingness to obtain more information about the latent factor. More specifically, m_{it} is specified to be a missingness dummy variable, and the probability of missingness is linked to latent levels of PSI, i.e., $Pr(m_{it} = 1) = \Lambda(b_0 + b_1 \xi_{i,t})$.

Since stability and predictability cannot be evaluated given a short time span, observations for democracies²⁰ begin after the first two consecutive democratic elections or five years have passed since the inauguration of democracy, depending on whichever occurs later. For similar reasons, democratic regimes that survive for less than ten years are excluded. I estimate the model using Stan in R. Each of the four independent chains discards the first 2000 iterations as burn-in, and the next 5000 iterations are treated as draws from the joint posterior density. Standard MCMC diagnostics indicate that all chains have sufficiently converged. Substantive and statistical summaries of the manifest indicators and their posterior distributions are presented in Appendix E.^{21,22}

PSI Scores

The measurement model uses 15,988 data points across the five manifest indicators to generate 3,313 posterior distributions of the latent factors, which provide country-year estimates of PSI across 96 post-WWII democracies. In my discussion, I refer to the medians of these posterior distributions as *PSI Scores*, which serve as reasonable point estimates, and the full posterior distributions as *PSI Scores (D)*, which incorporate measurement uncertainty.²³ The coverage of the *PSI Scores* far outstrips those of existing measures, which should facilitate tests of PSI that encompass a global sample of democracies. Nonetheless, the utility of a measure is determined principally by its validity, and thus I examine the face, convergent, and construct validity of the *PSI Scores*.²⁴

In the latter two tests of validity, I compare the *PSI Scores* measure with four alternative measures of PSI. The first two are electoral volatility (*EV*) and the age of the main parties (*Party Age*), which provide a useful baseline given their frequent deployment in the literature. *EV* is calculated using the same data as the *Aggregate Legislative Volatility* indicator, and its scale is reversed so that higher values are associated with more institutionalized party systems. The *Party Age* variable calculates the average age of the two largest governing parties and the largest opposition party (Cruz et al., 2018) and is logged given its skew. The third measure is Casal Bértoa and Enyedi’s (2021) Party System Closure Index (*Closure*), which measures PSI in the government arena across Europe. Comparisons with the index should reveal how the *PSI Scores* perform relative to a detailed within-region measure. The fourth measure is simply the average of the five manifest indicators used in the measurement model (*MI Avg*). It is included to assess whether using the more complex Bayesian latent variable measurement model to aggregate indicators provides any noticeable advantages. To facilitate comparisons, I limit the observations to those that cover the 96 post-WWII democracies in my sample.

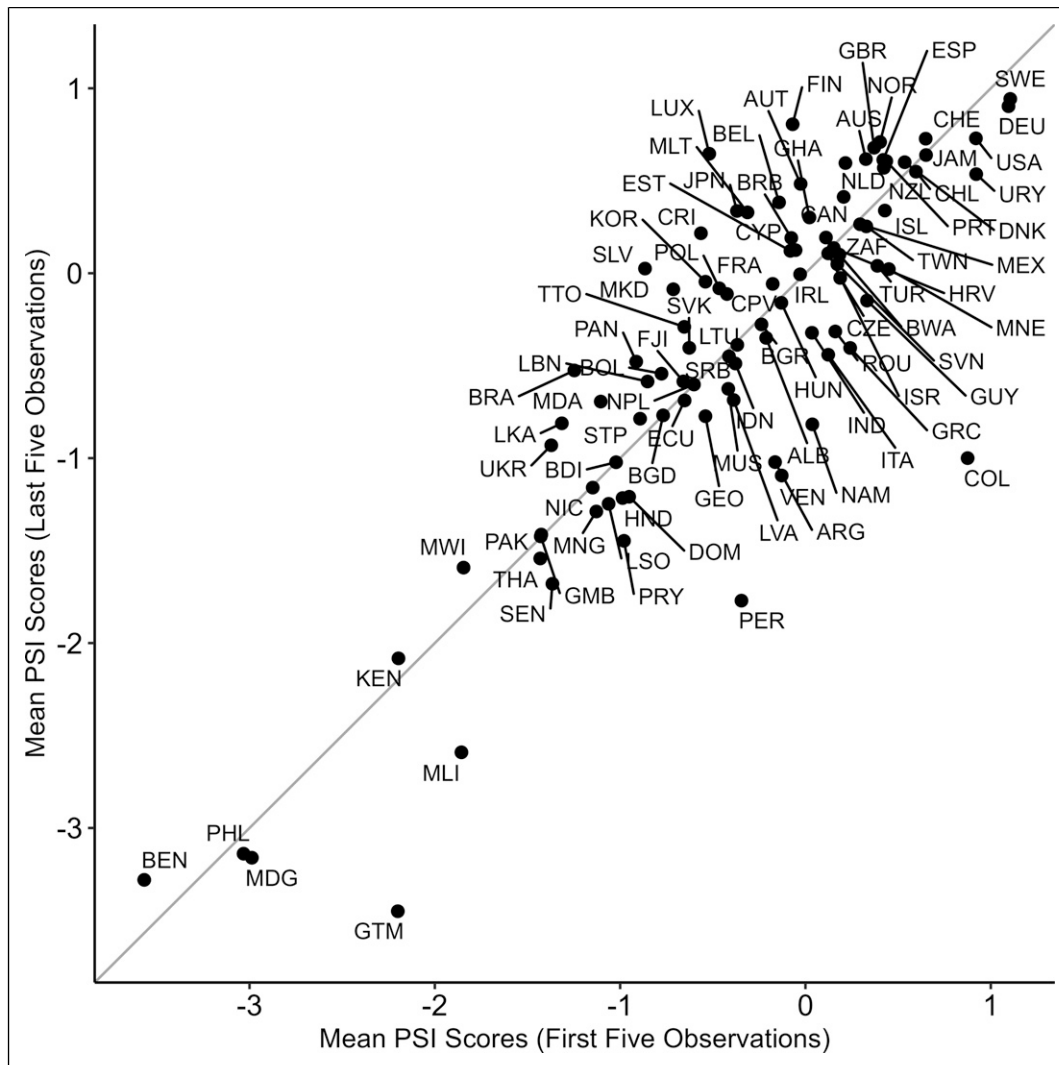


Figure 1. Variations in the PSI Scores across space and time.

Face validity: PSI across space and time

A measure with face validity should conform to existing expectations (Adcock and Collier, 2001). To give a sense of the spatial and temporal variation in the *PSI Scores*, I plot the mean *PSI Scores* for the first and last five observations for each country in Figure 1.

The spatial distribution generally aligns with expectations. Western democracies dominate the upper right spectrum, whereas countries that have been noted as having consistently under-institutionalized party systems such as Benin (Kuenzi and Lambright, 2001), the Philippines (Hicken and Kuhonta, 2015), and Guatemala (Mainwaring, 2018) occupy the lower left spectrum.²⁵

In addition, countries that lie above (below) the 45-degree line are those that are estimated to have become more (less) institutionalized over time. The points are evenly divided by the grey line and do not exhibit a clear

trend, which affirms existing arguments that party systems do not necessarily become and stay institutionalized over time, and that even institutionalized party systems can also undergo de-institutionalization (Casal Bértoa and Enyedi, 2021; Chiamonte and Emanuele, 2022; Lindberg, 2007). On a pessimistic note, this suggests that there has been little convergence in the institutionalization of party systems over time.

Convergent validity: Correlations with existing measures

Convergent validation examines whether a measure is correlated with other measures of the same concept (Adcock and Collier, 2001). Figure 2 plots the *PSI Scores* and the four alternative measures of PSI. Reassuringly, the correlations are relatively strong, but there remain noticeable

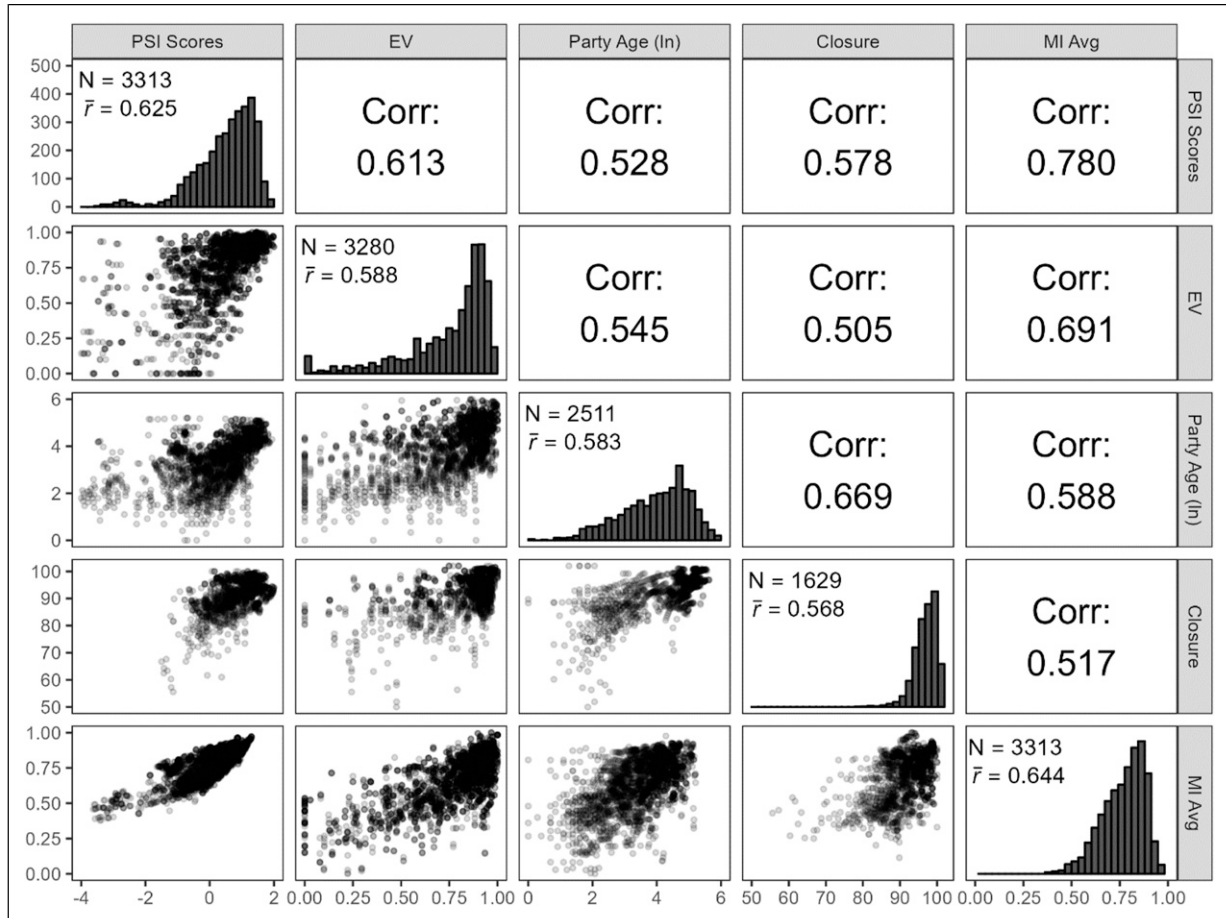


Figure 2. Scatterplots, histograms, and correlations of the PSI measures.

differences between the measures, which could impact inference. Interestingly the correlation between the *PSI Scores* and *MI Avg* measures is only 0.78 even though they are constructed from the same indicators, which suggests that the aggregation method could make a meaningful difference. Furthermore, the correlation between the *PSI Scores* and V-Dem's Electoral Democracy Index is 0.56, which signals that the measure is capturing a concept that is related to but distinct from democracy.

Construct validity: Democracy and PSI

Construct validation assesses whether a measure corroborates well-established hypotheses, and builds on the premise that a valid measure should be correlated with measures of distinct but theoretically related concepts (Adcock and Collier, 2001).²⁶ PSI is widely regarded as being beneficial for the performance of democratic institutions as it generates information about parties and lengthens their time horizons, which enhances the ability of parties to hold each other accountable, cooperate and sustain intertemporal agreements, and implement policies

more consistently and efficiently (Hicken, 2016; Mainwaring, 2018; Schleiter and Voznaya, 2018; Tommasi, 2006).

I identify six dependent variables to test this relationship, which are summarized in Table 1. To evaluate the broad applicability of the PSI measures, the dependent variables are each selected from different sources that employ distinct data and methodologies. To further increase the challenging nature of the test, I use variables from the V-Dem dataset to control for the strength of other relevant political actors—the bureaucracy, civil society, judiciary, media, and the state—that can impact democratic performance.²⁷ All variables are averaged over five-year intervals to mitigate the impact of random shocks and political cycles. Following recent cross-national studies of PSI (e.g., Chiamonte and Emanuele, 2019; Mainwaring et al., 2017; Mainwaring, 2018), I estimate the models using Generalized Estimating Equations with an AR1 autoregressive error-correlation structure and Huber-White standard errors to account for autocorrelation and panel-heteroskedasticity. This method estimates population-averaged effects and is a relatively conservative estimation strategy (Mainwaring, 2018).

Table 1. Summary of the dependent variables.

Dependent Variable	Description (Source)
Performance of Democratic Institutions (<i>PDI</i>)	Democratic institutions are effective and efficient, and political decisions are prepared, made, implemented, and reviewed in legitimate procedures by the appropriate authorities (Bertelsmann Transformation Index).
State Fragility Index (<i>SFI</i>)	The capacity of the state to manage conflict; make and implement public policy; deliver essential services and maintain system coherence, cohesion, and quality of life; respond effectively to, and sustain progressive development (Center of Systemic Peace).
Factionalized Elites (<i>FE</i>)	The extent to which local and national leaders engage in deadlock and brinkmanship for political gain (Fund for Peace).
Corruption Perception Index (<i>CPI</i>)	Perceived levels of public sector corruption (Transparency International).
Transparent and Predictable Laws (<i>TPL</i>)	Laws of the land are clear, well-publicized, coherent/consistent, relatively stable year to year, and enforced in a predictable manner (V-Dem).
Regulatory Quality (<i>RQ</i>)	The ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development (Worldwide Governance Indicators).

Note: the dependent variables are rescaled so that higher values represent normatively superior outcomes. See [Appendix H](#) for more details.

Table 2. Summary of the coefficient estimates.

	PDI	SFI	FE	CPI	TPL	RQ
PSI Scores	0.35** (0.16)	1.28*** (0.34)	0.67*** (0.17)	4.26*** (1.18)	0.18*** (0.06)	0.23*** (0.04)
PSI Scores (D)	0.23* [0.02, 0.43]	0.47** [0.13, 0.79]	0.3** [0.13, 0.5]	1.92* [0.06, 3.9]	0.05 [-0.01, 0.11]	0.08*** [0.03, 0.13]
EV	-0.05 (0.34)	0.45 (0.50)	0.63* (0.35)	-0.03 (2.76)	0.07 (0.07)	0.13** (0.07)
Party Age (ln)	-0.13 (0.12)	0.56*** (0.19)	0.02 (0.13)	2.75*** (0.87)	0.02 (0.03)	0.05** (0.03)
Closure	0.01 (0.02)	0.10** (0.04)	0.07** (0.03)	0.42 (0.31)	0.01 (0.01)	0.02*** (0.01)
MI Avg	1.04 (0.89)	2.35* (1.42)	2.29*** (0.87)	-6.32 (6.39)	0.22 (0.24)	0.34* (0.20)

Note: parentheses show the standard errors and brackets show the 90% credible interval (see footnote 23). ***= $p < 0.01$; **= $p < 0.05$; *= $p < 0.10$.

[Appendix H](#) offers more detailed information about the dependent and control variables.

I iteratively regress the six dependent variables on each of the six measures of PSI and the control variables. To expedite comparisons across the subsequent thirty-six regression models, [Table 2](#) presents just the coefficient estimates for the PSI measures, and the full results are reported in [Appendix H](#). As seen in the table, the *PSI Scores* measure clearly outperforms the alternative measures as a predictor of the dependent variables. Strikingly, this pattern holds even when using the *PSI Scores (D)*, which incorporate measurement uncertainty and subsequently suffer from attenuation bias (i.e., a bias of the coefficient estimates towards 0). In other words, the *PSI Scores* measure exhibits more robust empirical associations with the dependent variables even when it is the only variable handicapped by attenuation bias, which provides persuasive support for the improved validity of the measure.

The substantive implications are also meaningful. A one standard deviation increase in the *PSI Scores*, which is about the mean difference between the United States and Turkey, is associated with a 0.22 standard deviation increase in the dependent variables, on average. Moreover, such estimated magnitudes are larger or comparable to those associated with the bureaucracy, civil society, judiciary, media, and the state,²⁸ and the fact that the *PSI Scores* measure remains a strong predictor even when controlling for the strength of these actors suggests that the measure does seem to measure a distinct concept.²⁹

Interestingly, the *EV* measure performs relatively poorly in this analysis, which warns against the tendency to default to electoral volatility when carrying out cross-national tests of PSI. Moreover, the *MI Avg* measure is a weaker predictor than the *PSI Scores*. This affirms the benefits of measuring PSI using a latent variable

measurement strategy over the more straightforward but less appropriate method of averaging indicators.

Conclusion

The comparative study of PSI currently lacks a comprehensive measure of the concept, which has limited the formulation of unified understandings of PSI's role in democracies. This article fills this gap by constructing a novel measure of the concept that covers 96 post-WWII democracies, addresses extant measurement issues, and has demonstrated validity. Importantly, the measure exhibits robust empirical associations with numerous outcomes that are linked to the performance of democratic institutions, and should contribute to more systematic and encompassing studies of the relationship between PSI and democracy.

Although there is still much to be learned about this relationship, there is another strand of research on PSI that generally remains unexplored. Most of the PSI literature focuses on democracies, but the rise of authoritarian regimes that incorporate interparty competition means that PSI could have meaningful implications for regime performance even in non-democratic contexts (Kim et al., 2022). Moreover, the patterns of interparty competition developed during past authoritarian regimes could cast long shadows that continue to shape the party system after democratization (Hicken and Kuhonta, 2015; Riedl, 2014). However, given disjointed efforts to measure PSI in democracies and the increased difficulty of gathering comparative party system data in autocracies, it is unsurprising that there is no comprehensive measure of the concept that extends to non-democratic regimes. Nonetheless, the measurement strategy presented in this article could be expanded to cover non-democratic regimes, and the subsequent measure could facilitate the development of holistic theories about the long-term causes and consequences of PSI across regime types and regime transitions.

Acknowledgements

I would like to thank Fernando Casal Bértoa, Christopher Fariss, Robert J. Franzese, Laura Gamboa, Allen Hicken, Woojin Kim, Walter Mebane, Anja Neundorf, Yuki Shiraito, George Tsebelis, Fabricio Vassellai, and participants at Michigan's Comparative Politics workshop and MPSA 2021 conference for their helpful comments and suggestions. I am also indebted to Octavio Amorim Neto and Don Lee for sharing their data. Lastly, I am grateful to the editor and three anonymous reviewers for their constructive feedback.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Wooseok Kim  <https://orcid.org/0000-0003-3624-2864>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. The former is available for at least 67 post-WWII democracies (Mainwaring et al., 2017), while the latter covers most countries from 1975 onwards (Cruz et al., 2018).
2. For example, Mainwaring (2018) identifies stability and predictability as the principal dimension of PSI and construes other dimensions of Mainwaring and Scully's (1995) original framework (i.e., strong party roots, legitimacy of elections/parties, and solid organizations) as underpinnings rather than features of PSI.
3. Casal Bértoa and Enyedi's index is limited to European countries, but does offer extensive temporal coverage from 1848 to 2019. Chiaramonte and Emanuele's index covers 20 post-WWII European countries, while Rodriguez and Rosenblatt's measure covers Latin American countries at three points in time.
4. See Bizzaro et al. (2017) for a more detailed overview of the index.
5. It may be possible to regard some of the manifest indicators as drivers of PSI since—as with any system—some endogeneity is likely present. Nonetheless, it is the characteristics of party systems, including levels of institutionalization, that set the overarching parameters that condition all observable behaviors and interactions within the party system.
6. Appendix A provides an overview of other commonly used indicators of PSI and the reasons for their exclusion from the measurement model.
7. Although the non-volatility indicators have been theoretically linked to PSI in the literature, they have typically not been used to measure the concept. This is understandable, as existing works do not treat PSI as a latent concept and instead seek to directly measure its dimensions, or requisite data may not have been previously available. In the context of the latent variable measurement approach, the indicators can be characteristics of party systems that are a function of underlying levels of PSI.
8. While some minor parties can win and keep their seats in institutionalized party systems as one reviewer points out, the rationale here is that PSI makes it less likely that minor parties can collectively amass substantial presence comparable to that of the main parties. In support of this reasoning, Gerring (2005) finds that major party hegemony is

associated with low electoral volatility (i.e., more institutionalized party systems). Moreover, as detailed in footnote 17, this indicator provides one strategy for handling data availability issues without losing relevant information about PSI.

9. Likewise, [Heller and Mershon \(2009:38\)](#) note that the party system “sets the stage” for party switching by conditioning the decision-making of party leaders and legislators.
10. The data sources are listed in [Appendix B](#).
11. I use seat shares rather than vote shares since seat shares are (1) highly correlated with vote shares, (2) more relevant for legislative outcomes, and (3) missingness is more prevalent in vote share data.
12. A party is coded as being the same party across time if there is a continuation of the party label and/or its organizational composition. This approach lies between the strict and relaxed linkages described by [Casal Bértoa et al. \(2017\)](#).
13. This assigns the majority of the weight to the most recent two to three elections on average. Similarly, [Mainwaring and Torcal \(2006\)](#) and [Mainwaring \(2018\)](#) calculate electoral volatility scores between elections spaced roughly two decades apart to assess long-term trends, and [Chiaromonte and Emanuele \(2022\)](#) average volatility calculations across three elections. The analysis presented in this paper are robust to varying the timeframe.
14. [Appendix D](#) compares aggregate volatility and electoral volatility scores. In sum, aggregate volatility better distinguishes different types of party system change (e.g., collapse vs. re-alignment) by incorporating information from prior elections, which allows it to produce more intuitive volatility scores. Electoral volatility, on the other hand, is a forgetful measure that equates all party system change with de-institutionalization, which can lead to volatility scores that do not accurately reflect long-term systemic change.
15. Although using portfolio shares would be preferred, such data are not consistently available for many post-WWII democracies. Nonetheless, portfolio shares should generally be proportional to seat shares ([Gamson, 1961](#)). When exact seat share data are not available, which is usually the case for minor actors that hold one portfolio, the seat share is coded as one percent.
16. This indicator is updated whenever there is a change in the cabinet or after a general election.
17. Election reports frequently do not provide detailed information about minor parties and independents, and thus scholars frequently apply similar exclusion thresholds when calculating volatility scores ([Casal Bértoa et al., 2017](#)). While I also follow this strategy, I leverage the excluded data to construct the *Minor Party Performance* indicator since the data should contain relevant information about PSI. In addition, I accordingly adjust the denominator of the *Aggregate Legislative Volatility* by subtracting the excluded seat share from 2. Leaving the denominator unchanged despite having an exclusion threshold can introduce a bias in the volatility calculations (see [Birch \(2003\)](#) for a similar adjustment). The denominator for the government volatility indicator is also similarly adjusted to be the sum of the seat shares obtained by parties in the two governments under comparison.
18. While this variable is included in V-Dem’s PI Index, it is more closely related to PSI since the *distinctiveness* of party platforms can only be assessed at the *system* level.
19. Data for this indicator is available up to 2015. Since party switching is generally stable within countries, I impute values for the few years leading up to 2018 using ARIMA forecasting ([Hyndman and Khandakar, 2008](#)). These missing observations can be estimated by the measurement model, but this does not produce any meaningful differences while causing unintuitive shifts in the estimates when encountering missingness. In the case of Luxembourg and Malta, which are not covered in the V-Dem dataset, missing observations are treated as unknown parameters that are estimated by the measurement model.
20. Democracies are identified using a threshold of six in the Polity V dataset ([Marshall and Gurr, 2020](#)). One reviewer raises the possibility of sampling using V-Dem’s regime measure. While there is a large overlap in the classification of regimes across Polity and V-Dem, V-Dem classifies fewer country-years as being democratic. As such, using V-Dem’s classification would likely produce a highly correlated measure with fewer observations.
21. One reviewer notes that institutionalized party systems may not always exhibit all the characteristics captured by the manifest indicators. The manifest indicators are expected to covary with but not share a deterministic relationship with PSI, and this assumption is reflected in the measurement model, which includes both a systematic and non-systematic component (i.e., ϵ_k). Moreover, as shown in [Appendix E](#), the selected manifest indicators are robustly associated with the latent factor, i.e., all factor loadings (λ_k) are statistically significant.
22. The measurement model also confirms that missingness does seem to be associated with less institutionalized party systems, i.e., $b_1 < 0$.
23. [Appendix F](#) outlines the Monte Carlo procedure that uses the *PSI Scores (D)* to incorporate measurement uncertainty. It also provides Stata code to implement the procedure.
24. The indicators that comprise the *PSI Scores* are subject to measurement noise not only because they are imperfect reflections of a latent concept, but also due to potential inconsistencies in data, coding, or expert judgment. Such issues can have implications for reliability, which is another important attribute of a measure. Nonetheless, the *PSI Scores* measure also offers an improvement over other measures in this aspect since it provides information about the accompanying measurement uncertainty.
25. [Appendix G](#) presents the distribution of the *PSI Scores* for all countries and discusses how measurement uncertainty can impact inference.
26. As one reviewer correctly notes, such correlations neither establish causality nor preclude the possibility that other

measures could exhibit stronger empirical associations. Nonetheless, Adcock and Collier (2001) suggest that these correlations contribute positive (albeit not definitive) evidence about the validity of a measure by demonstrating that the measure behaves as expected. Stated otherwise, it would be more difficult to claim that a measure is valid if contradicts well-established hypotheses. Moreover, and as this section demonstrates, highly correlated measures of the same concept can perform quite differently in empirical evaluations, and thus construct validity tests offer unique leverage for comparing the validity of such measures (ibid.).

27. These variables already capture substantial variation in levels of economic development and their mean correlation with logged GDP per capita is 0.55. As such, I exclude GDP per capita from the main analysis to avoid exacerbating multicollinearity issues. In Appendix H, I repeat the main analysis including logged GDP per capita as an additional control variable. While statistical significance unsurprisingly diminishes across the board, the *PSI Scores* measure still outperforms the other measures of PSI.
28. See Table H5 in Appendix H.
29. Sample sizes vary depending on data availability as shown in Appendix H. When repeating the analysis after limiting the sample to observations that are common across all PSI measures, the *PSI Scores* measure remains the strongest predictor. Incorporating measurement uncertainty weakens the results, though this is to be expected given the drastically reduced sample size. Furthermore, hyper-institutionalized party systems have been hypothesized as being potentially detrimental to democracy (e.g., Mainwaring, 2018). However, additional models suggest that PSI does not seem to have such non-linear implications in this analysis. Summaries of these results are presented in Tables H17 and H18 in Appendix H.

References

- Adcock R and Collier D (2001) Measurement validity: a shared standard for qualitative and quantitative research. *American Political Science Review* 95(3): 529–546.
- Arel-Bundock V and Mebane W (2011) Measurement error, missing values and latent structure in governance indicators. In: 107th Annual Meeting of the American Political Science Association, Seattle, US, September, 2011.
- Bernhard M, Hicken A, Reenock C, et al. (2020) Parties, civil society, and the deterrence of democratic defection. *Studies in Comparative International Development* 55(1): 1–21.
- Birch S (2003) *Electoral Systems and Political Transformation in Post-Communist Europe*. New York: Palgrave Macmillan.
- Bizzaro F, Hicken A and Self D (2017) The V-Dem party institutionalization index: a new global indicator. V-Dem Working. Paper 48.
- Bollen KA (1989) *Structural Equations with Latent Variables*. New York: Wiley.
- Casal Bértoa F (2017) Political parties or party systems? Assessing the ‘myth’ of institutionalisation and democracy. *West European Politics* 40(2): 402–429.
- Casal Bértoa F (2018) The three waves of party system institutionalisation studies: a multi- or uni-dimensional concept? *Political Studies Review* 16(1): 60–72.
- Casal Bértoa F and Enyedi Z (2021) *Party System Closure: Party Alliances, Government Alternatives, and Democracy in Europe*. Oxford: Oxford University Press.
- Casal Bértoa F, Deegan-Krause K and Haughton T (2017) The volatility of volatility: measuring change in party vote shares. *Electoral Studies* 50: 142–156.
- Chiaromonte A and Emanuele V (2019) Toward turbulent times: measuring and explaining party system (de-)institutionalization in western Europe. *Italian Political Science Review* 49(1): 1–23.
- Chiaromonte A and Emanuele V (2022) *The Deinstitutionalization of Western European Party Systems*. London: Palgrave Macmillan.
- Coppedge M (1998) The dynamic diversity of Latin American party systems. *Party Politics* 4(4): 547–568.
- Coppedge M, Gerring J, Knutsen CH, et al. (2020) *V-Dem Codebook V10*. Varieties of Democracy (V-Dem) Project.
- Croissant A and Völkel P (2012) Party system types and party system institutionalization: comparing new democracies in east and southeast Asia. *Party Politics* 18(2): 235–265.
- Cruz C, Keefer P and Scartascini C (2018) *The Database of Political Institutions*. Washington: Inter-American Development Bank.
- Desposato SW (2006) Parties for rent? Ambition, ideology, and party switching in Brazil’s chamber of deputies. *American Journal of Political Science* 50(1): 62–80.
- Enyedi Z and Casal Bértoa F (2020) Party systems: types, dimensions, and explanations. In: Thompson WR (ed), *Oxford Research Encyclopedia of Politics*. Oxford: Oxford University Press.
- Gamson WA (1961) A theory of coalition formation. *American Sociological Review* 26(3): 373–382.
- Gerring J (2005) Minor parties in plurality electoral systems. *Party Politics* 11(1): 79–107.
- Greene KF and Sánchez-Talanquer M (2018) Authoritarian legacies and party system stability in Mexico. In: Mainwaring S (ed), *Party Systems in Latin America*. Cambridge: Cambridge University Press.
- Heller WB and Mershon C (eds), (2009) *Political Parties and Legislative Party Switching*. New York: Palgrave Macmillan.
- Hicken A (2016) Party systems and the Politics of development. In: Lancaster C and Van de Walle N (eds), *Oxford Handbook of Political Development*. Oxford: Oxford University Press.
- Hicken A and Kuhonta EM (eds), (2015) *Party System Institutionalization in Asia*. Cambridge: Cambridge University Press.

- Hyndman RJ and Khandakar Y (2008) Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* 27(3): 1–22.
- Jensenius FR and Suryanarayan P (2022) Party system institutionalization and economic voting: evidence from India. *The Journal of Politics* 84(2): 814–830.
- Jones MP (2010) Beyond the electoral connection: the effect of political parties on the policymaking process. In: Scartascini C, Stein E and Tommasi M (eds), *How Democracy Works: Political Institutions, Actors, and Arenas in Latin American Policymaking*. Washington: Inter-American Development Bank.
- Kim W, Bernhard M and Hicken A (2022) *Electoral volatility and regime survival in competitive authoritarian regimes*. V-Dem Working Paper 129.
- Kreuzer M and Pettai V (2009) Party switching, party systems, and political representation. In: Heller WB and Mershon C (eds), *Political Parties and Legislative Party Switching*. New York: Palgrave Macmillan.
- Kuenzi M and Lambright G (2001) Party system institutionalization in 30 African countries. *Party Politics* 7(4): 437–468.
- Lee S (2007) *Structural Equation Modeling: A Bayesian Approach*. Hoboken: Wiley.
- Lindberg SI (2007) Institutionalization of party systems? Stability and fluidity among legislative parties in Africa's democracies. *Government and Opposition* 42(2): 215–241.
- Luna JP (2014) Party system institutionalization: do we need a new concept? *Studies in Comparative International Development* 49(4): 403–425.
- Lupu N and Riedl RB (2013) Political parties and uncertainty in developing democracies. *Comparative Political Studies* 46(11): 1339–1365.
- Mainwaring S (1999) *Rethinking Party Systems in the Third Wave of Democratization*. Stanford: Stanford University Press.
- Mainwaring S (ed), (2018) *Party Systems in Latin America: Institutionalization, Decay, and Collapse*. Cambridge: Cambridge University Press.
- Mainwaring S and Scully TR (eds), (1995) *Building Democratic Institutions: Party Systems in Latin America*. Stanford: Stanford University Press.
- Mainwaring S and Torcal M (2006) Party system institutionalization and party system theory after the third wave of democratization. In: Katz RS and Crotty W (eds), *Handbook of Party Politics*. London: Sage Publications.
- Mainwaring S, Gervasoni C and España-Najera A (2017) Extra- and within-system volatility. *Party Politics* 23(6): 623–635.
- Mair P (1997) *Party System Change*. Oxford: Oxford University Press.
- Marshall MG and Gurr TR (2020) *Polity5: Dataset Users' Manual*. Center for Systemic Peace.
- Mauro V (2022) Party systems and redistribution in democratic Latin America. *Comparative Politics* 54(3): 429–452.
- Meleshevich AA (2007) *Party Systems in Post-Soviet Countries*. New York: Palgrave Macmillan.
- Munck GL and Verkuilen J (2002) Conceptualizing and measuring democracy. *Comparative Political Studies* 35(1): 5–34.
- Pedersen MN (1979) The dynamics of European party systems: changing patterns of electoral volatility. *European Journal of Political Research* 7(1): 1–26.
- Randall V and Svåsand L (2002) Party institutionalization in new democracies. *Party Politics* 8(1): 5–29.
- Ridge HM (2022) Party system institutionalization, partisan affect, and satisfaction with democracy. *Party Politics*. Advance online publication. DOI: [10.1177/13540688221130492](https://doi.org/10.1177/13540688221130492).
- Riedl RB (2014) *Authoritarian Origins of Democratic Party Systems in Africa*. Cambridge: Cambridge University Press.
- Robbins JW (2010) Party system institutionalization and government spending. *Comparative Politics* 42(2): 229–248.
- Robbins JW and Hunter LY (2011) Impact of electoral volatility and party replacement on voter turnout levels. *Party Politics* 18(6): 919–939.
- Rodriguez RP and Rosenblatt F (2020) Stability and incorporation: toward a new concept of party system institutionalization. *Party Politics* 26(2): 249–260.
- Schleiter P and Voznaya A (2018) Party system institutionalization, accountability, and governmental corruption. *British Journal of Political Science* 48(2): 315–342.
- Stockton H (2001) Political parties, party systems, and democracy in east Asia. *Comparative Political Studies* 34(1): 94–119.
- Tommasi M (2006) The institutional foundations of public policy. *Economía* 6(2): 1–36.
- Weghorst KR and Bernhard M (2014) From formlessness to structure? The institutionalization of competitive party systems in Africa. *Comparative Political Studies* 47(12): 1707–1737.

Author biography

Wooseok Kim is a Postdoctoral Fellow in Democratic Politics at the University of Glasgow. His research focuses on regime resilience and distributive policymaking. He received his PhD in Political Science at the University of Michigan.