A New Measure of Congruence: The Earth Mover's Distance

Noam Lupu[1], Lucía Selios[2] and Zach Warner[3]

[1] *Department of Political Science, Vanderbilt University, Nashville, TN 37203, USA. Email: noam.lupu@vanderbilt.edu*
[2] *Instituto de Ciencia Política, Universidad de la República, Montevideo 11200, Uruguay. Email: lucia.selios@cienciassociales.edu.uy*
[3] *Department of Political Science, University of Wisconsin–Madison, Madison, WI 53706, USA. Email: zwarner@wisc.edu*

Abstract

Scholars of representation are increasingly interested in mass–elite congruence—the degree to which the preferences of elected elites mirror those of voters. Yet existing measures of congruence can be misleading because they ignore information in the data, require arbitrary decisions about quantization, and limit researchers to comparing masses and elites on a single dimension. We introduce a new measure of congruence—borrowed from computer science—that addresses all of these problems: the Earth Mover's Distance (EMD). We demonstrate its conceptual advantages and apply it to two debates in research on mass–elite congruence: ideological congruence in majoritarian and proportional systems and the determinants of congruence across countries in Latin America. We find that improving measurement using the EMD has important implications for inferences regarding both empirical debates. Even beyond studies of congruence, the EMD is a useful and reliable way for political scientists to compare distributions.

1 Introduction

How well do elected officials represent the preferences of their constituents? Do some contexts or institutions yield a pool of elites whose preferences better mirror those of voters? Scholars of representation have long been interested in such questions of mass–elite congruence (Miller and Stokes 1963; Converse and Pierce 1986), but the topic has received renewed attention more recently (see Powell 2004; Canes-Wrone 2015). In some contexts, scholars suggest that incongruence deters citizens from engaging in politics and makes them distrustful of political institutions (Mainwaring, Bejarano, and Pizarro Leongómez 2006; Joignant, Fuentes, and Morales 2017). In others, political inequalities raise questions about how well elites represent voter preferences (Bartels 2008; Gilens 2012). Thus far, studies have found that proportional electoral systems tend to generate more mass–elite congruence than majoritarian systems (Powell 2009) and that congruence is greater when party systems are more institutionalized and economic outcomes are better (Luna and Zechmeister 2005; Kitschelt *et al.* 2010).

Measuring congruence, however, is challenging. In this paper, we focus on what Golder and Stramski (2010) call *many-to-many congruence*,[1] the similarity between the distribution of voter preferences and the distribution of elite positions. The goal is to come up with a summary measure that captures the similarity between the two distributions. Often, studies compare responses to mass surveys on the one hand with some measure of elite preferences, based either on party manifestos (McDonald and Budge 2005; Budge and McDonald 2007; Powell 2009; Golder and Lloyd 2014), voter perceptions about party platforms (Blais and Bodet 2006; Golder and Stramski 2010),

Authors' note: We are grateful to Manuel Alcántara, Matt Golder, and Juan Andrés Moraes for generously sharing their data, and to Andy Eggers, Alicia Fernández, Scott Gehlbach, Álvaro Gómez, Luis Schiumerini, Alex Tahk, Simon Urbanek, and seminar participants at Oxford and Wisconsin for helpful feedback. Previous versions of this paper were presented at the annual meetings of the European Political Science Association, Political Methodology Specialist Group of the Political Studies Association, and Society for Political Methodology. For replication materials, see Lupu, Selios, and Warner (2016). Supplementary materials are available in an online appendix on the *Political Analysis* website.

1 However, as we discuss below, the measure we introduce subsumes other definitions of congruence.

Political Analysis (2017)
vol. 25:95–113
DOI: 10.1017/pan.2017.2

Published
22 February 2017

Corresponding author
Noam Lupu

Edited by
Jonathan Katz

or elites' own responses to surveys (Lupu and Warner 2017; Siavelis 2009; Selios 2015; Buquet and Selios 2017). Regardless of the data one employs, the result hinges on using a measure that faithfully compares the distributions of preferences among voters and elites.

Political scientists have largely used three measures of this type of congruence. Most simply, some authors compute the difference between the means of mass and elite preference distributions (Powell 2000; McDonald, Mendes, and Budge 2004; McDonald and Budge 2005; Budge and McDonald 2007; Siavelis 2009). Greater distances are interpreted to imply more incongruence between the two populations. A more elaborate approach, proposed by Golder and Stramski (2010), compares the overlap between the cumulative distribution functions of mass preferences and elite positions. A final approach uses a similar calculation, but relies on overlap in the probability distribution functions, which is naturally constrained between zero and one, making interpretation more intuitive (Andeweg 2011; Lupu and Warner 2017; Buquet and Selios 2017). In both overlap measures, greater overlap between the mass and elite distributions indicates greater congruence.

All three approaches have substantial limitations. Most importantly, all of these measures throw out information in the data. With the difference-in-means approach, each distribution of responses gets collapsed into a single summary statistic before calculating congruence. We lose all information about differences between the samples' variances. With overlap measures, information loss is more subtle. The key point is that in order to compute overlap, a researcher must first decide what it actually means for data to overlap. In practice, this decision requires scholars to "bin" data into histograms, eliminating within-bin variation. Moreover, overlap measures ignore the data in bins that lack common support across histograms—that is, all the bins that do not overlap. As a result, scholars can construct bins that will return just about *any* amount of overlap for the same two distributions by arbitrarily ignoring portions of the data.

In addition, existing measures can only be computed in a single dimension. This is one reason scholars have limited their focus to the summary left–right ideological dimension. But the left–right scale can be problematic, particularly when used to measure mass opinion (Saiegh 2015; Zechmeister 2006).

In this paper, we propose a measure that overcomes these limitations. The Earth Mover's Distance (EMD), most commonly used in similarity-based image retrieval (Rubner, Tomasi, and Guibas 2000), originated as an eighteenth-century solution to a classic problem of resource allocation in transportation theory (Monge 1781). The EMD computes the minimum "work" required to transform two distributions so that they are identical. It evaluates all possible "flows" by which data can be "moved" so that the distributions match. Since its original application in transportation theory, the EMD has been generalized, adapted, and applied across a number of fields; today, it is associated with a broad class of similar measures (Deza and Deza 2006).[2] But to our knowledge, it has never been employed in political science.

We demonstrate, theoretically and using simulated data, that the EMD overcomes the limitations posed by the measures of congruence currently being employed by political scientists. For our purposes, the EMD has two key features. First, it works with variable-size *signatures*—generalized histograms—which eliminate the need for binning. In addition, the EMD calculates all pairwise distances across signatures, ensuring that both the amount and the location of all the data enters into the congruence calculation. These features also make it easier to calculate distance in multiple dimensions. The EMD can provide a summary measure of congruence across multiple questions, obviating scholars' reliance on problematic left–right

2 Recent studies have extended the EMD to approximations for big data applications or to "learn" an appropriate metric for a given data structure (Ling and Okada 2007; Wang and Guibas 2012). We focus on the vanilla EMD because these extensions are beyond the practical needs of most political science applications.

placements. Even better, it is already implemented in the R package `emdist` (Urbanek and Rubner 2015).[3]

We then apply the EMD to two current debates in political science: (1) whether different electoral systems are associated with different degrees of mass–elite congruence, focusing largely on advanced democracies; and (2) whether political characteristics like party system institutionalization and economic factors like growth correlate with mass–elite congruence in Latin America. In both cases, measuring congruence using the EMD yields results that depart substantively from existing findings. How we measure congruence has important implications for the inferences we draw about its determinants. Scholars interested in understanding what factors make elected elites better representatives of mass preferences should start employing the EMD.

2 Measuring Congruence

Scholars studying congruence are often interested in measuring the similarity between elite positions and mass preferences across samples. In some applications—like tests of the median voter theorem—doing so may be fairly simple, and existing measures may be adequate. But scholars often want to compare elite and voter distributions. Since these distributions are represented in Euclidean space, similarity is simply a measure of distance: more similar objects are physically closer to each other. But distance is defined contextually. Research in fields such as graph theory, pattern recognition, cryptography, and molecular biology have developed a staggering array of distance measures for particular applications (Deza and Deza 2006). Choosing among these measures should be of central concern to congruence scholars; as we will show, the choice of measure has important substantive implications.

To begin, let us define a *distance* $\mathbb{D}$ as a numerical description of how far apart objects are in a metric space, which is defined as an ordered pair $(\mathcal{M}, \mathcal{D})$. $\mathcal{M}$ is a set, while $\mathcal{D}$ is a *metric*, a function that defines the distance between objects in the set, i.e., $\mathcal{D} : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$.[4] A metric must satisfy nonnegativity, symmetry, Euclid's triangle inequality, and the identity of indiscernibles.[5] All distances between all elements of $\mathcal{M}$ must be identified.

Throughout the paper, we consider the distance between two random variables X and Y. The observed data are events drawn from X and Y, $P_x = \{p_x(x_1), \ldots, p_x(x_m)\}$ and $P_y = \{p_y(y_1), \ldots, p_y(y_n)\}$, respectively. The x_i and y_j form sequences over the metric space $(\mathcal{M}, \mathcal{D})$, and the $p_x(x_i)$ and $p_y(y_j)$ form the sequences' associated weights in P_x and P_y. Define the total weights (or simply "size") of each sequence such that $W_x = \sum_i p_x(x_i)$ and $W_y = \sum_j p_y(y_j)$. Sets such as P_x and P_y are known as *signatures* in the image-retrieval literature, and can be viewed as generalized histograms. Note that since X and Y are random variables, P_x and P_y are their distributions in $\mathbb{R}^q$, respectively (Levina and Bickel 2001).

Consider the examples in Figure 1. Each panel plots the empirical histogram and PDF of 1,000 draws from a random variable X that follows a standard-normal distribution in one dimension ($q = 1$). The top, middle, and bottom panels also plot these statistics for 1,000 draws each from random variables Y_1, Y_2, and Y_3, respectively, where $Y_1 \sim \mathcal{N}(3, 1)$, $Y_2 \sim \mathcal{N}(0, 64)$, and $Y_3 \sim \text{Mix}(\lambda\mathcal{N}(2.5, 1), (1 - \lambda)\mathcal{N}(13.1, 2))$. Here, Mix represents a mixture distribution, with mixing weight $\lambda = .65$. We highlight the overlap between histograms in dark gray.

What attributes are desirable when evaluating a measure of distance? One minimal requirement is ordinality: if X and Y_1 are closer together than are X and Y_2, then our distance

3 We recommend compiling from source; see our replication materials for details. Computation time using this implementation is not particularly onerous. We generated a hard test using two samples of simulated data, $m = 1,746$ and $n = 1,140$, and calculated distance in 15 dimensions, five of which are for noisy variables, with random cross-dimensional correlations, outliers, multimodality, and clustering. Computation time on a standard laptop for these data was approximately 8 minutes.

4 While some fields use "distance" and "metric" interchangeably, we follow convention in the image-retrieval literature, distinguishing these terms to help clarify the exposition below.

5 That is, for any x, y, $\mathcal{D}(x, y) \geq 0$; $\mathcal{D}(x, y) = \mathcal{D}(y, x)$; $\mathcal{D}(x, z) \leq \mathcal{D}(x, y) + \mathcal{D}(y, z)$; and $\mathcal{D}(x, y) = 0 \Leftrightarrow x = y$, respectively.

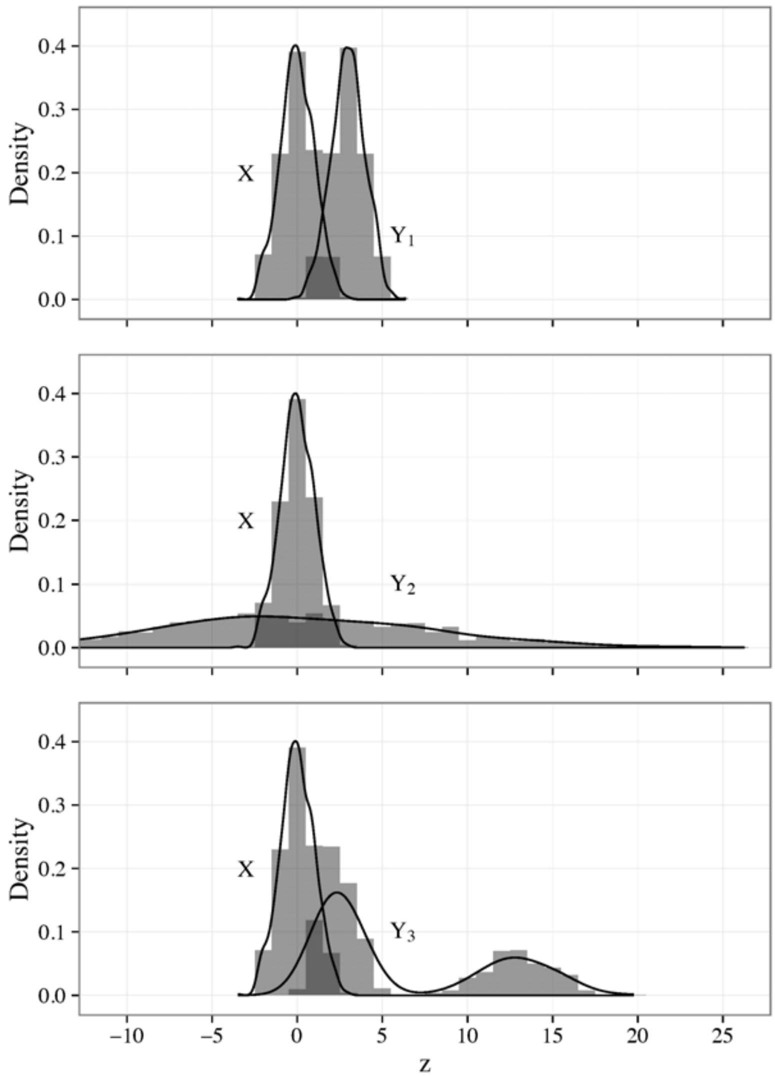

Figure 1. Three comparisons using simulated data. Three hypothetical distributional comparisons. A good measure of distance should at least indicate that the distributions in the top panel are the most similar.

function should always return $\mathbb{D}(X, Y_1) < \mathbb{D}(X, Y_2)$. In other words, a good measure of distance ought to at least recover the rank ordering of distributions' similarity. We also prefer a measure that preserves cardinality, though we consider this a second-order concern. If we know that Y_2 is twice as far from X as is Y_1, we prefer a measure that recovers $\mathbb{D}(X, Y_1) = 2\mathbb{D}(X, Y_2)$.[6] Finally, a good measure of distance should use all of the data across the relevant distributions.

The examples in Figure 1 highlight that these properties should hold whether distance is increasing because of differences across sample means or across sample variances. For instance, it is clear from visual inspection that $\mathbb{E}[Y_3] > \mathbb{E}[Y_1]$, so we would like a measure that does not report the samples in the bottom panel as more similar than those of the top panel. It is also clear that $\mathbb{V}[Y_2] > \mathbb{V}[Y_1]$, so we prefer a measure that reflects this. A good measure of distance should indicate that the distributions in the top panel of Figure 1 are the most similar.

6 This also implies linearity, which is reasonable since political scientists will likely want to include a measure of congruence in linear regression models.

3 Existing Measures

Three measures of congruence are especially popular in political science applications. The first is the simple difference in means, where the distance between X and Y is defined as

$$\mathbb{D}_{\mathrm{M}} = |\mathbb{E}[X] - \mathbb{E}[Y]|. \tag{1}$$

Here a larger distance indicates less similarity and thus less congruence.

The second is a measure that instead attempts to measure overlap between distributions. Golder and Stramski (2010) suggest computing the nonoverlap of empirical cumulative distribution functions (CDFs) for X and Y,

$$\mathbb{D}_{\mathrm{CDF}} = \sum_{z \in Z} |[\mathcal{F}_x(z) - \mathcal{F}_y(z)]|, \tag{2}$$

where Z is a set of points, chosen by the researcher, at which to evaluate the functions (generally covering the theoretical bounds or values observed in P_x and P_y). Other scholars instead use empirical probability density function (PDF) overlap since it is constrained to the unit interval, making regression estimates equivalent to effect magnitudes in percentage points (Andeweg 2011; Lupu and Warner 2017). This distance, also known as the Bhattacharyya coefficient, is computed as

$$\mathbb{D}_{\mathrm{PDF}} = \sum_{z \in Z} \min\{f_x(z), f_y(z)\}. \tag{3}$$

Unlike $\mathbb{D}_{\mathrm{M}}$ and $\mathbb{D}_{\mathrm{CDF}}$, larger values for $\mathbb{D}_{\mathrm{PDF}}$ indicate distributions that are closer together, and therefore more similar.

All of these measures discard information in the data. As the name "difference-in-means" signals, $\mathbb{D}_{\mathrm{M}}$ is computed by taking expectations over X and Y *separately*, and then finding the absolute difference between them. This method ignores substantial variation within each distribution. While such reduced summaries may be appropriate for other applications, this approach is too coarse to serve as a default measure of congruence.

Understanding the limitations of overlap measures requires closer attention to quantization, the process of restricting a continuous quantity to discrete values. While we would like to study the similarity between X and Y directly, our only information about these random variables is encoded in the sequences P_x and P_y. The problem for overlap measures is knowing which elements of these sequences to compare: which events should enter into the distance calculation, and how? In practice, overlap measures solve this problem by binning each set of observations into histograms. This transformation creates a shared index which identifies data that are considered to be "in the same place." Formally, a histogram of a sequence P_x, $\mathbf{H}_x = \{h_x^i\}$, is a mapping from a set of q-dimensional integer vectors $\boldsymbol{\omega}_x$ to $\mathbb{R}^+$. A histogram partitions the underlying space—here, the support of the probability distribution for P_x—into a fixed number of hyper-rectangular blocks known as "bins," centered at the points in $\boldsymbol{\omega}_x$, with h_x^{ω} the number of elements that have a value in the interval indexed by $\boldsymbol{\omega}_x$ (Rubner *et al.* 2000). Where $q = 1$, $\boldsymbol{\omega}_x$ is a vector, the number of bins is equal to the length of $\boldsymbol{\omega}_x$, and the bins are simply two-dimensional intervals. Allow $\mathbf{H}_y = \{h_y^j\}$ to be defined similarly for the set P_y. In practice, $\mathbb{D}_{\mathrm{CDF}}$ and $\mathbb{D}_{\mathrm{PDF}}$ sum overlap among each bin in $\mathbf{H}_x$ and $\mathbf{H}_y$, rather than the observed data P_x and P_y (or the random variables X and Y from which they are drawn).[7]

7 In principle, it may be possible to compute CDF overlap without first quantizing the data. Still, we recommend the EMD because CDF overlap functions built into statistical software automatically quantize the data and because the EMD performs as well or better in simulations.

This quantization requires scholars to make decisions about which variation in the data to ignore. Specifically, since overlap measures directly compare bins to their individual counterparts across histograms, $\mathbf{H}_x$ and $\mathbf{H}_y$ must be of a fixed size, and they must share an index. This constraint forces scholars to choose a single quantization for both distributions P_x and P_y (and any other distributions they wish to compare), with an identical range and bin width. Data that are naturally quantized—for example, variables which only take integer values—may be amenable to such techniques. However, for continuous variables, there is no rubric for making an objective decision about appropriate quantization.[8] Overlap measures are extremely sensitive to this choice. If the data are quantized too finely, many bins will be empty, driving overlap downward. If they are quantized too coarsely, then important information about the shape of the distributions will be lost as disparate data are grouped together, increasing overlap. In the extreme, for any X and Y, it is possible to construct bin widths such that PDF overlap takes *any value* on the unit interval (CDF overlap is similarly manipulable).

Even where the data are naturally quantized and share a common index, overlap measures throw out information. That is because overlap measures only compare bins that have the same index across histograms. Put formally, they compute common mass between h_x^i and h_y^i, but not h_x^i and h_y^j $\forall i \neq j$. This means that data from $\mathbf{H}_x$ that lie in bins beyond the index in $\mathbf{H}_y$ (and vice versa) are ignored. For example, if X and Y are one-dimensional and all observed data are integers, the elements of P_x will form a natural index such as $\boldsymbol{\omega}_x = \{-3, -2, \ldots, 3\}$. In that case, all data in P_y that fall outside the interval $[-3, 3]$ will be effectively discarded in computing overlap.[9] In some sense, this is precisely what overlap measures set out to accomplish: any bins lacking overlap do not contribute to the measure of congruence. Yet, in this example, congruence will be identical whether Y has a local maximum at 12 or at 10^{12}: since both locations fall outside the range of data observed in P_x, overlap will always be zero. An empty bin in $\mathbf{H}_x$ generates zero overlap, irrespective of how many observations lie in the corresponding bin in $\mathbf{H}_y$. By ignoring the amount and location of all data that do not overlap, these measures force scholars to overlook potentially important information about X and Y.

Overlap measures suffer from one final shortcoming. While it is possible to extend them to multidimensional space, the problems associated with quantization become even worse as more variables are studied simultaneously. Density concentrates near the surface of high-dimensional distributions, increasing the likelihood of empty bins—the curse of dimensionality (Bellman 2010). Moreover, cross-dimensional correlations and multimodality, which are also likely to increase when scholars use multiple survey responses to capture the same latent dimension, exacerbate this problem. Thus, extending overlap measures to multiple dimensions only frustrates the problems associated with them, making congruence even harder to measure reliably.

4 The Earth Mover's Distance

To address the limitations associated with existing measures of congruence, we propose the EMD. The EMD between X and Y is defined by the solution to a linear optimization problem, the optimal "flow" f_{ij}^* for moving the distance between P_x and P_y. Our goal is to minimize $\sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}$ subject to

$$f_{ij} \geq 0 \quad \text{for } 1 \leq i \leq m, 1 \leq j \leq n, \tag{4}$$

8 Algorithms to automatically select bin widths are inappropriate because they are built for visualization, not goodness of fit to the data. Most statistical packages, including R's `density` function, default to $q = 512$.

9 An intuitive solution to this problem might be to compute overlap over the interval $[\min\{\boldsymbol{\omega}_x, \boldsymbol{\omega}_y\}, \max\{\boldsymbol{\omega}_x, \boldsymbol{\omega}_y\}]$. In fact, this is what the R package `ggplot2`, used to create Figure 1, implements automatically. But building a common index by tacking empty bins onto either histogram will not change the distance computed, since overlap will always be zero for these extra bins.

$$\sum_{j=1}^{n} f_{ij} \leq p_x(x_i) \quad \text{for } 1 \leq i \leq m, \tag{5}$$

$$\sum_{i=1}^{m} f_{ij} \leq p_y(y_j) \quad \text{for } 1 \leq j \leq n, \text{ and} \tag{6}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min\left(\sum_{i=1}^{m} p_x(x_i), \sum_{j=1}^{n} p_y(y_j)\right), \tag{7}$$

where d_{ij} is some description of similarity between the ith and jth elements of the signatures P_x and P_y, respectively. Together, these pairwise descriptions form the ground distance matrix $\{d_{ij}\}$, defined by some metric $\mathcal{D}(x_i, y_j)$. Once the optimal flow is found, the EMD is defined as

$$\mathbb{D}_{\text{EMD}} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}^{*} d_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}^{*}}, \tag{8}$$

where the denominator is simply a normalization for cases where signatures differ in magnitude.[10] Larger values of $\mathbb{D}_{\text{EMD}}$ indicate greater distance, less similarity, and less congruence.

This definition is easier to understand with a simple example. Suppose we want to compute the EMD between two samples of respondents' self-placement on a 0–10 scale, $P_x = \{(2, .5), (5, .3), (8, .2)\}$ and $P_y = \{(4, .3), (5, .4), (6, .3)\}$, where each element in parentheses refers to a location and a density. We want to find the flow that minimizes the amount of "work" (or the "cost") of moving data such that the signatures are identical. Each panel in Figure 2 presents a different way of moving P_x, plotted in various grays, to the location of P_y, in dashed lines. One option, the "optimal flow" in the top panel, is to move from left to right, transferring data in P_x to the nearest data in P_y.[11] For instance, we start by moving data in P_x at $z = 2$ to $z = 4$ until the total mass there is 0.3, as in P_y —generating work of 0.6. Continuing this process yields total work equal to

$$(|4-2|)(0.3) + (|5-5|)(0.3) + (|5-2|)(0.1) + (|6-2|)(0.1) + (|6-8|)(0.2) = 1.7.$$

The constraint in Equation 4 ensures that all data movement counts toward the total work (flow cannot be negative); conditions 5 and 6 require that we cannot move more data than exist at any location (flow cannot exceed the density); and condition 7 ensures that all of the data are moved (the densities sum to 1) (Rubner *et al.* 2000).

Other flows achieve the same result, but may generate more work. The bottom panel displays an "inefficient flow" whereby we move data in P_x to a more distant location in P_y, yielding total work of 2.5. With more data, and as the set of feasible values $\{z\}$ increases, the number of possible flows grows dramatically. The EMD simply uses the one that minimizes total work. In this example, since the first flow is optimal, the EMD is $\frac{(1)(1.7)}{1} = 1.7$.

10 Note that this normalization accounts for differences in total weight, not sample size. In political science applications, signatures rarely differ in magnitude; we typically compare densities, whose total weight is equal to 1. Scholars ought to carefully consider whether a particular sample is representative of the population of interest—particularly when working with cross-national datasets. But the EMD will accurately compute the difference between two sample densities regardless of sample size.

11 This flow is an implementation of the greedy algorithm, which looks for a globally optimal solution by making a locally optimal choice at each stage. While the EMD is always optimal, the greedy algorithm produces the optimal flow only where a Monge sequence exists, as in this case. See Alon *et al.* (1989) for a definition and discussion.

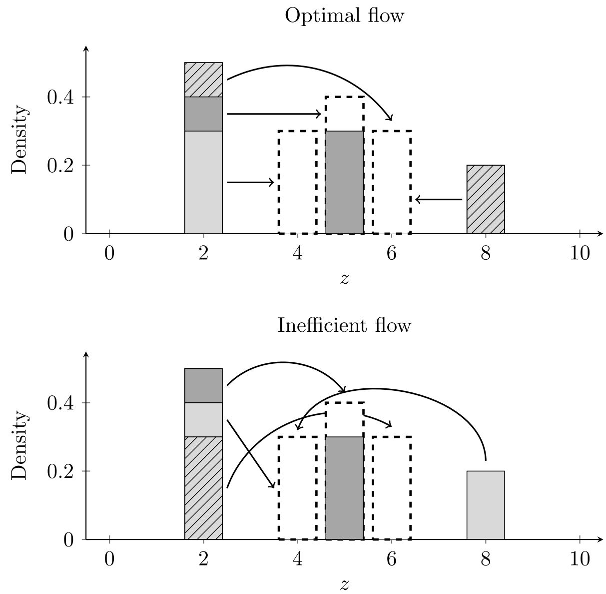

Figure 2. Two examples of "flow." P_x is in gray, while P_y is represented in dashed lines. Data plotted in light gray "ends up" at $z = 4$, dark gray at $z = 5$, and stripes at $z = 6$. In the top panel, data in P_x are moved to the nearest position available, yielding total work of 1.7. In the bottom panel, a less efficient flow yields more work (2.5). The EMD finds the optimal flow that minimizes work.

This example assumes that the ground distance matrix is defined by the L_1 norm. This metric, also known as the city-block or Manhattan distance, is given by

$$\mathcal{D}_{L_1}(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\| = \sum_{r \in q} |a_r - b_r|, \tag{9}$$

where **a** and **b** are points in q-dimensional space.[12] The EMD can take any metric, and although this choice is generally dependent on the application (Wang and Guibas 2012), we recommend the L_1 norm for ease of interpretability. In the foregoing example—as in all one-dimensional calculations—the L_1 yields an average distance *on the scale of the original response*. To see this, observe that the maximum we could calculate for a question on a 0–10 scale would be if one sample was entirely located at 0, and the other entirely at 10, yielding $(|10 - 0|)(1) = 10$. Unlike other metrics, this property also holds for multidimensional comparisons (i.e., $q > 1$) where the underlying scales are identical. For instance, if we compare three dimensions, all on the same 0–10 scale, then an EMD of 6 can be directly interpreted as the average sum of three-dimensional differences. Further, if the differences between distributions are relatively evenly spread across dimensions, then we could say that the distributions are, on average, 2 points away on each of the original 11-point scales. And if instead of three ordinal scales we have a series of binary questions, then the EMD-L_1 across all of them is simply a measure of the average number of questions on which respondents differ.

Although the EMD is suitable for measuring distance between any objects in metric space, it is especially useful for problems like the example above, which are most familiar to congruence scholars: computing the cost of transforming one statistical distribution into another. Recall that the signatures P_x and P_y are sequences of events ("observations"), or empirical distributions drawn from random variables X and Y. If we assume that each has finite pth moments, then

12 This metric is a special case of the more general L_p norm. Another special case, more familiar to political scientists, is the Euclidean or L_2 norm, given by $\mathcal{D}(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|_2 = \sqrt{(a_1 - b_1)^2 + \cdots + (a_q - b_q)^2}$. We use the L_1 metric throughout the paper, but all of our results are robust to using the L_2 norm (see online appendix).

the flow $\{f_{ij}\}$ is equivalent to the joint distribution $\mathcal{F}(X, Y)$. Moreover, the signatures' weights are simply $W_x = W_y = 1$, so the constraints in Equations (5)–(6) must hold with equality, while Equation (7) and the denominator in Equation (8) both equal 1 (Levina and Bickel 2001). With the L_1 norm, we then have the EMD for P_x and P_y as the minimum sum of simple pairwise distances:

$$\mathbb{D}_{\text{EMD}} = \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}^* \|x_i - y_j\|, \tag{10}$$

which we can equivalently express as $\mathbb{D}_{\text{EMD}} = \min\{(\mathbb{E}_{\mathcal{F}} \|X - Y\|)\}$.

Written this way, the EMD appears similar to the measures of congruence that political scientists regularly employ. However, while the mathematical distinctions are subtle, the EMD differs substantially from existing measures. First, the EMD can be used to calculate distances between histograms which need not cover the same range or share an index. More importantly, we need not even use histograms: the EMD can compute distances directly from the data. We simply define each signature as the observations in each sample, weighted by their respective sample sizes so that $W_x = W_y = 1$, and then calculate pairwise distances between all data points.

The EMD also outperforms overlap measures even when the data are naturally quantized into bins x_i and y_j. As is evident in Equation 8, the EMD incorporates *all* pairwise distances between these bins—it calculates distance between h_x^i and $h_y^j \; \forall i, j$. Unlike overlap measures, which treat all nonoverlap as equal, the EMD avoids ignoring local maxima and information in the tails of X and Y. The amount and location of data always matter for the EMD.

In one dimension, the EMD also subsumes other definitions of congruence. For instance, suppose P_x is a sample of citizens within a particular district, and P_y is the (single) representative for that district, so that $P_y = p_y(y)$ and $p_y = W_y = 1$. It is clear from Equation 10 that, in this case, the EMD reduces to the total absolute ideological distance between P_x and P_y. If we were to divide through by m (in this example, the size of the citizen sample), the resulting quantity would be identical to what Golder and Stramski (2010) call *absolute citizen congruence*. Similarly, if we also restrict P_x to a single respondent, x, then the EMD reduces to simple absolute distance, $\|x - y\|$; if x is the median voter in a district, then we get what Golder and Stramski (2010) term *absolute median citizen congruence*. Similarly, their *relative citizen congruence* can easily be recovered from EMD calculations. Thus, depending on how the signatures of interest are defined, the EMD can be employed to measure many conceptualizations of congruence.

The EMD has an important final advantage: it is easily extended to comparing distributions in high dimension.[13] This allows scholars to study congruence across a range of questions, preferences, or issue areas with a single summary statistic. The EMD generalizes directly to multiple dimensions, and since all d_{ij} contribute to total distance, increasing the number of empty bins does not affect the total distance computed. The cost of adding variables is only computation time.[14]

Scholars should ensure that the multidimensional EMD is appropriate for their application with minimal data processing. We recommend that the variables entering into a multidimensional congruence calculation have similar scales. Each variable in a high-dimensional congruence calculation is weighted equally by default. However, if this is inappropriate for a particular

13 This is another advantage of the L_1 norm. Scholars have uncovered some performance issues with the Euclidean norm in high dimension (Aggarwal, Hinneburg, and Keim 2001), demonstrating that lower-p L_p metrics perform better. These problems are beyond the scope of this paper, but we reiterate that the EMD can be calculated using whatever metric scholars believe most appropriate for their data (Rubner *et al.* 2000).

14 These features also make the EMD more attractive than a less common technique, procrustes analysis (Luna 2014). Like the EMD, this approach measures congruence in multidimensional space. However, procrustes analysis relies on data transformations that distort the underlying shape of the data, which is unnecessary with the EMD. It also only calculates pairwise distances between points that share the same index, like overlap measures. Finally, it produces a parametric estimate of distance, whereas the EMD produces a true distance.

Table 1. Comparing measures of distance.

	$\mathbb{D}_M$	$\mathbb{D}_{\text{CDF}}$	$\mathbb{D}_{\text{PDF}}$	$\mathbb{D}_{\text{EMD}}$
Top panel (Y_1)	2.98	155.49	0.15	**2.98**
Middle panel (Y_2)	**0.08**	**54.85**	0.24	5.65
Bottom panel (Y_3)	6.22	136.99	**0.26**	6.22

Cells report the values that each measure of congruence generates for each of the panels in Figure 1. For each measure, the most similar pair of distributions is bolded.

application, scholars can change this by simply including particular variables multiple times in the data. The EMD affords substantial flexibility in constructing a single measure of multidimensional congruence.

5 Simulation Evidence

Simulated data illustrate the superior performance of the EMD relative to existing measures. Table 1 reports the measure of congruence for each panel in Figure 1, comparing existing methods with the EMD. As the Table makes clear, the EMD produces better results than do existing measures. In this simulation, the EMD determines that Y_1, in the top panel of Figure 1, is closest to X. In contrast, difference-in-means calculations suggest that the most similar distributions are X and Y_2, in the middle panel. In fact, they are considered nearly identical, since their empirical means differ only by 0.08, compared to approximately 3 and 6 in the top and bottom panels, respectively. Yet it is clear from visual inspection that this does not capture significant variation in Y_2, giving a false impression of similarity. This result follows directly from the definition: recall that $\mathbb{D}_{\text{M}}$ is computed by taking expectations over the marginal distributions of X and Y and then finding the absolute difference between them. As expected, ignoring within-sample variance overstates congruence in this case.

The EMD also outperforms overlap measures. $\mathbb{D}_{\text{PDF}}$ determines that the X-Y_3 pairing, in the bottom panel, is the most similar. Yet, again, this is a misleading conclusion: while roughly two-thirds of the data in Y_3 appear to be very close to that of X, another third is substantially distant. Meanwhile, the similar distributions in the top panel fare the worst. Perhaps most disconcerting, the overlap measures do not agree on the ordering of each distribution's similarity to X, with $\mathbb{D}_{\text{CDF}}$ favoring the X-Y_2 pairing in the middle panel. Since both $\mathbb{D}_{\text{PDF}}$ and $\mathbb{D}_{\text{CDF}}$ measure overlap and the PDF is just the first difference of the CDF, we would expect that they produce similar results. Here, not only do they disagree on which distributions are most similar, they also disagree on the relative magnitudes of the distances between distributions. According to CDF overlap, Y_2 is decisively the closest distribution to X, with the next closest (Y_3) being 2.5 times farther away. In contrast, according to PDF overlap, the standout is the top panel, with Y_1's overlap less than 65 percent of Y_2's.

Finally, it is worth noting that the EMD returns a distance identical to difference-in-means calculations for the top and bottom panels. This is an intuitive result given the construction of these distributions: Y_1 is equivalent to a secular shift in X of magnitude three, while Y_3 is essentially an average of two shifts in X, adjusted for the mixing weight λ. Encouragingly, when two samples are drawn from distributions that are essentially identical in shape, the EMD recovers the true distance in their locations—even when those shifts are embedded as components of more complex mixture distributions.

These encouraging results hold more generally. We conduct a Monte Carlo experiment to investigate each measure's performance in recovering the ordinality and cardinality of distance comparisons. First, we generate a target sample (T) with 100 draws from $\mathcal{N}(0, 1)$. Then we randomly generate 1,000 means ($y_1, y_2, \ldots, y_{1000}$), and draw 100 samples from each $\mathcal{N}(y_i, 1)$.

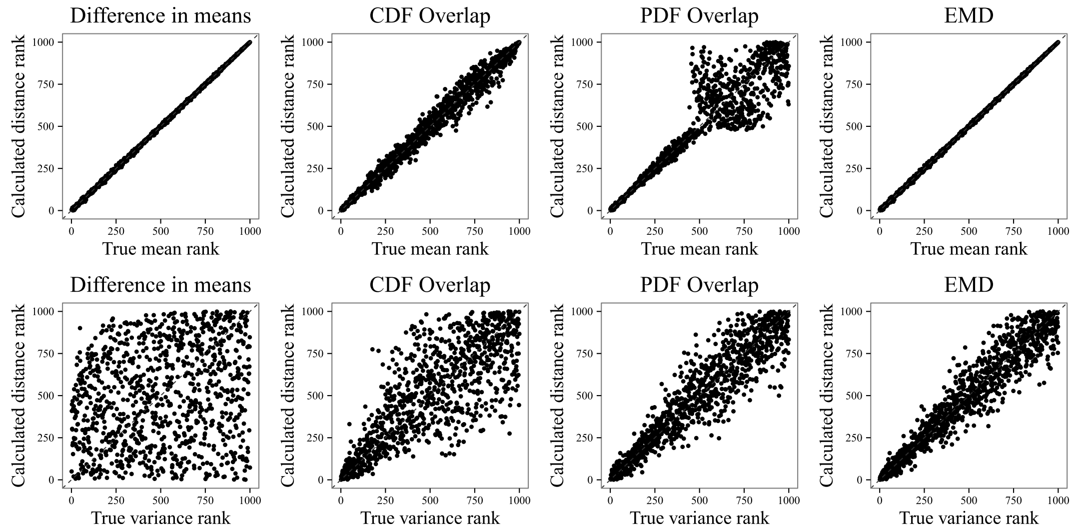

Figure 3. Monte Carlo simulations of measures of congruence: ordinality. We compare the rank order of computed distances between a "target" sample and 1,000 samples with randomly generated means (top row) and 1,000 samples with randomly generated variances (bottom row). We expect a good measure of congruence to preserve ordinality: it should return a one-to-one correspondence between the ranked differences in means (variances) and the ranked computed distance. See the online appendix for the (very similar) cardinality results.

Table 2. RMSE for Monte Carlo simulations.

	Ordinality		Cardinality	
Distance	**Means**	**Variances**	**Means**	**Variances**
Difference in Means	5.89	363.09	0.00	0.41
CDF Overlap	42.68	184.48	0.44	0.20
PDF Overlap	115.96	106.35	0.17	0.26
EMD	5.90	87.57	0.00	0.11

Root mean squared error for each measure of congruence across 2,000 simulations. Note that the scale is 1 to 1,000 for ordinality but 0 to 1 for cardinality.

Since each of these samples differs from T only in location, the distance between them should be very close to their difference in means: $|y_i - 0|$, or just $|y_i|$. We then generate 1,000 variances $(z_1, z_2, \ldots, z_{1000})$ and draw 100 samples from each $\mathcal{N}(0, z_i)$. Here, distance should be increasing in the true difference in variances, $|z_i - 1|$, since each sample has the same mean as T.[15] Finally, we calculate each of these 2,000 distances using the three existing measures and the EMD.[16]

Preserving ordinality is our primary concern. We expect that comparing the orderings of the $|y_i|$ ($|z_i - 1|$) and the distance calculations should yield a one-to-one correspondence: samples with larger differences in means (differences in variances) should be considered farther away from the target sample. For that reason, Figure 3 plots the ranked means (top row) and variances (bottom row) against the ranked calculated distances. Root mean squared error (RMSE) calculations—the mean difference between a sample's actual rank and its distance rank—are presented in Columns 2 and 3 of Table 2.

As with the examples in Figure 1, the EMD outperforms existing measures. Both the EMD and difference in means have a neat correspondence between the ranked distances and the real $|y_i|$

15 Specifically, the means and variances are drawn from $\mathcal{N}(10, 100)$ and Unif(1, 10), respectively. See the online appendix for the empirical densities of these samples.
16 We invert $\mathbb{D}_{\text{PDF}}$ before scaling it so that we could directly compare calculated PDF overlap to true distances.

ordering, with an average error of about 0.6%. In contrast, while CDF overlap does not appear to suffer from any systematic bias, it struggles to recover ordinality, with an RMSE eight times larger than the EMD. PDF overlap's difficulty recovering a rank ordering in the top row highlights the problem of very little overlap. For the larger half of $|y_i|$, the results become essentially random, as PDF overlap is zero, and $\mathbb{D}_{\text{PDF}}$ is unable to tell how far apart the samples are. Moving to the bottom row, the EMD continues to recover ordinality, although with more difficulty. RMSE in the rank-ordering variances rises to about 9%, but this is a substantial improvement over any of the other measures. Thus, the most important rubric by which to judge a measure of distance—its ability to preserve ordinality—provides clear evidence that we should favor the EMD.

To examine cardinality, we scale the means, variances, and distances computed to the unit interval and compare their (linear) correspondence. The results look similar. Columns 4 and 5 in Table 2 indicate that the EMD is increasing perfectly in the true difference in means between samples, and errs by approximately 11% in the true difference in variances. Both of these substantially outperform the other measures. These results provide clues as to why overlap measures produce such different conclusions about congruence using the same data: they are largely unable to recover accurate descriptions of distributions' similarity. Across all of these 2,000 samples, the EMD best preserves both ordinality and cardinality of distance comparisons.[17]

6 Empirical Applications

There are good theoretical reasons to think that the EMD is a better measure of congruence than those currently employed in political science. We have also demonstrated its advantages using simulated data. How might adopting the EMD affect empirical results of substantive interest to political scientists?

We answer these questions with regard to two debates in the study of mass–elite congruence. The first centers on whether electoral rules affect the degree of congruence between voters and their elected officials. Scholars within this debate ask whether proportional electoral systems offer higher levels of mass–elite congruence than do majoritarian systems. A second debate regards whether party system institutionalization and socioeconomic factors positively correlate with congruence in Latin America. Since we lack an empirical benchmark for measuring true congruence, we do not intend for these applications to further demonstrate the advantages of the EMD. Rather, we use them to show that adopting the EMD significantly changes how we answer these important substantive questions.

6.1 Electoral rules and ideological congruence

Scholars of representation have become interested in the degree of congruence between citizens' preferences and policymakers' ideological positions (Powell 2006, 2000). At the most basic level, these studies have shown that mass–elite congruence varies across space and time—that is, that some governments more closely reflect the preferences of the citizenry than others (Dalton 1985; Miller and Stokes 1963). One possible reason for this cross-national variation is that some political institutions produce more congruent governments than others. In particular, an "ideological congruence controversy" (Powell 2009) has emerged regarding the role of electoral systems in promoting mass–elite congruence on left–right ideology. The scholarly debate is between those who find that electoral systems of proportional representation generate more mass–elite ideological congruence than majoritarian electoral systems (Huber and Powell 1994; Powell and Vanberg 2000; McDonald and Budge 2005; Powell 2006, 2009; Ezrow 2007; Budge and McDonald 2007) and those who find little or no difference across electoral systems (Blais and

17 Like all measures of many-to-many congruence based on samples, the EMD likely overstates the distance between the populations. This upward bias is the result of measurement error in sampling. We recommend that scholars concerned about such bias use resampling methods like bootstrapping to calculate the EMD with uncertainty bounds.

Table 3. Electoral rules and ideological congruence.

	G & S replication		G & S sample, EMD		Full sample, EMD	
	(1)	(2)	(3)	(4)	(5)	(6)
Majoritarianism dummy	9.51		0.00		0.04	
	(12.79)		(0.12)		(0.08)	
Disproportionality		2.41*		0.01		0.01
		(0.99)		(0.01)		(0.01)
Observations	34	37	34	37	81	79
Countries	20	22	20	22	33	32
R^2	0.02	0.14	0.00	0.02	0.00	0.01

Sources: Bormann and Golder (2013); CSES; Gandrud (2015); Golder and Stramski (2010). *Notes*: OLS estimates. Positive values mean less congruence. The dependent variable in (1) and (2) is CDF overlap, as calculated by Golder and Stramski (2010, "G & S"). $^*p < 0.05$.

Bodet 2006; Golder and Lloyd 2014; Ferland 2016). Yet, as Powell (2009) demonstrates, scholars engaging in these empirical debates use different sets of data covering different time periods. Most importantly for our purposes, they also employ different measures of congruence.

One prominent example is an analysis by Golder and Stramski (2010) of data from the parliamentary systems included in the Comparative Study of Electoral Systems (CSES). At the time they conducted their analysis, there were 37 election studies in the CSES from parliamentary systems with the necessary variables. They covered 22 countries during the period 1996–2005. Today, the CSES includes 81 surveys from 33 parliamentary systems that have the necessary variables, covering the period 1996–2013.[18]

The CSES surveys ask respondents to place themselves on a 0–10 left–right ideological scale. They also ask them to place up to nine political parties on the same scale. Golder and Stramski (2010) use the average placement of each party by the 40 percent most educated respondents as a measure of the party's ideological position.[19] Using the share of seats obtained by each party in the lower legislative chamber, they construct an ideological distribution of legislators. They then use the CDF overlap measure to compare the ideological distribution of survey respondents and the ideological distribution of legislators. Most importantly, they compare CDF overlap among majoritarian and PR electoral systems, finding that legislatures in PR systems are significantly more congruent with mass ideology than are majoritarian systems.[20]

Columns 1 and 2 of Table 3 replicate the analysis of electoral rules and ideological congruence that Golder and Stramski (2010) report.[21] Like these authors, we find a positive, but statistically insignificant relationship between electoral rules and CDF overlap when we use a dummy for majoritarianism. However, when we regress CDF overlap on a continuous measure of disproportionality, the positive association reaches statistical significance. For Golder and Stramski (2010, 104), this result offers, "strong evidence that countries with PR electoral rules are more likely to have legislatures that are congruent with the ideological preferences of the citizenry than countries with majoritarian ones."

In columns 3 and 4 of Table 3, we use the same set of CSES cases, the same data transformations, and the same quantization, but now measure congruence using the EMD. In

18 See the online appendix for details about the sample.
19 They focus on these respondents because less educated respondents tend to be less informed about the party system and, as a result, place parties at the midpoint of the scale. One might reasonably object that this measures mass *perceptions* of party positions, rather than their true positions (Merrill, Grofman, and Adams 2001), but we follow the authors' approach for the sake of replication.
20 Elsewhere in the paper, they find no difference between electoral systems in terms of the congruence between citizens and their government.
21 Our results are identical to those provided in their replication archive.

other words, the only difference with the analysis in columns 1 and 2 is that we apply the EMD optimization equation rather than calculating CDF overlap. Regardless of whether we measure majoritarianism dichotomously or continuously, our results now reveal no statistically significant relationship between congruence and electoral rules, and our estimates suggest a substantively small effect at best.

Given that the CSES has expanded dramatically since Golder and Stramski (2010) conducted their analysis, we also updated their dataset to see if additional data points reveal any relationships. In columns 5 and 6 of Table 3, we recalculate the EMD on the full sample of parliamentary elections now available in the CSES. In keeping with Golder and Stramski's approach, we employed Gandrud's (2015) continuous measure of disproportionality and constructed a majoritarianism dummy variable based on how each country is coded in Bormann and Golder's (2013) database of electoral systems.[22] And we limited our sample of mass respondents to educated citizens by dropping those who did not complete high school. Finally, since our updated sample includes a number of observations within the same country over time, we cluster our standard errors by country (see Golder and Lloyd 2014).[23]

With the EMD and an expanded dataset, we still find no relationship between electoral rules and mass–elite congruence. Regardless of whether we use a majoritarianism dummy variable or a continuous measure of disproportionality, we never find evidence for the theory that PR systems better represent citizen preferences. At best, our estimates suggest a substantively minuscule effect of electoral rules that sometimes runs in the opposite direction. When we replace the problematic CDF overlap measure of congruence with the EMD, we find no convincing evidence that legislatures in PR systems better reflect the ideological preferences of citizens.

6.2 Mass–elite congruence in Latin America

Debates about congruence among parliamentary systems have revolved primarily around electoral rules. Scholars of the developing world instead point to other factors to explain why democracies in these contexts seem to poorly represent voter preferences. In Latin America, scholars regularly cite presidential policy switches (Stokes 2001; Lupu 2016), institutions that centralize power in the executive (O'Donnell 1994), weak and uninstitutionalized party systems (Mainwaring and Scully 1995), widespread clientelism (Calvo and Murillo 2013), low levels of development, and high economic inequality. Only recently have studies emerged that test these claims by comparing mass–elite congruence across countries and over time.

These studies use a wide variety of measures of congruence for different subsets of countries in the region, making their findings difficult to compare directly. But a substantial number find that mass–elite congruence correlates with more institutionalized party systems (Luna and Zechmeister 2005; Kitschelt *et al.* 2010; Bornschier 2013; Otero Felipe and Rodríguez Zepeda 2014). Several studies also find a positive relationship between economic prosperity and congruence (Luna and Zechmeister 2005; Selios 2015).

Part of the interest in studying mass–elite congruence in Latin America is the unique set of data available for the region. Since the mid-1990s, the University of Salamanca's Project on Parliamentary Elites in Latin America (PELA) has been conducting representative surveys of national legislators across the region's major democracies. Mass surveys are also available across the region, including the AmericasBarometer conducted by Vanderbilt University's Latin

22 We also use their classification to identify parliamentary systems. For the most recent observations, we updated the coding ourselves.

23 The online appendix reports four alternative versions of this analysis: (1) without clustered standard errors, (2) including all mass respondents, (3) dropping countries coded as semipresidential, and (4) using the (slightly different) electoral-rule classifications in the Database of Political Institutions (DPI; see Beck *et al.* 2001). No matter how we slice it, we never find a statistically significant or remotely substantial association between electoral rules and mass–elite congruence using the EMD.

American Public Opinion Project (LAPOP). Starting in 2010, the PELA surveys adopted the AmericasBarometer wording of a number of ideological and issue questions, making it easy to directly compare mass and elite preferences.[24]

One challenge in using these data is matching observations from the two survey projects. Whereas the AmericasBarometer is fielded every two years, the PELA surveys are conducted in the first year of each legislative session. We matched each survey in the 2010 and 2012 rounds of the AmericasBarometer with a PELA survey fielded within a year (but not prior to 2010, before the questions were harmonized). We were able to match 22 surveys across 14 countries.

A second challenge is how to aggregate multiple survey items into a measure of overall congruence. Since the EMD can be computed in high dimension, we treat each survey item as a dimension and generate a single measure of congruence for each country-year, with each variable equally weighted.[25] To test the claims of prior research about the correlates of mass–elite congruence, we relate these congruence measures to a number of variables that measure features of the regime, the political elite, the party system, electoral institutions, and levels of development and inequality.[26] Given the small number of cases in the dataset and the correlations among these variables, we regress our measure of congruence on each independent variable separately.[27]

For each covariate, the top value in Figure 4 plots the predicted change in the multidimensional EMD as a result of shifting each independent variable across its interquartile range. For comparison, the bottom value plots the same predicted change, but using a unidimensional EMD based only on the left–right variable. Unlike previous studies, we find no relationship between congruence and political factors like the age of democracy or the institutionalization of the party system (proxied by average party age and electoral volatility). We also find no association between congruence and electoral institutions like compulsory voting or district magnitude. This is the case with both the unidimensional and multidimensional EMD. Whereas the number of years a president has been in office seems to be associated with only left–right congruence, the number of years the ruling party has been in office seems to significantly reduce overall congruence. In contrast to prior studies, we also find a statistically significant association between economic development and diminished mass–elite congruence, but only when we measure congruence across multiple dimensions.

By measuring mass–elite congruence in Latin America with the EMD, we take advantage of its ability to summarize distances in high dimension. Our results stand in marked contrast to the claims of prior studies. Unlike previous scholars, we find no relationship between the institutionalization of a country's party system and the level of congruence. Instead, we find that a ruling party's tenure in office and economic development are correlated with congruence. Improving the way mass–elite congruence is measured yields very different substantive results that suggest that it is the countries in the region that are most developed where mass–elite congruence is lowest.

7 A New Measure for Comparing Distributions

How well do elected representatives reflect the preferences of the people who elected them? How and why does this level of mass–elite congruence vary over space and time? These are important questions not only for scholars of democratic representation but also for policymakers and

24 In particular, the comparable items asked about left–right ideology, support for same-sex marriage, government versus private ownership of industries, government social provision, government responsibility for creating jobs, government redistribution to reduce inequality, and government provision of healthcare. The precise question wording is available in the online appendix.

25 The online appendix reports the unidimensional and multidimensional EMDs for each country-year.

26 Saiegh (2015) recently highlighted the importance of rescaling survey responses to account for measurement problems such as differential item functioning. Unfortunately, our data do not include the "anchor" questions necessary to generate a common scale.

27 Data definitions, sources, summary statistics, and regression estimates are provided in the online appendix.

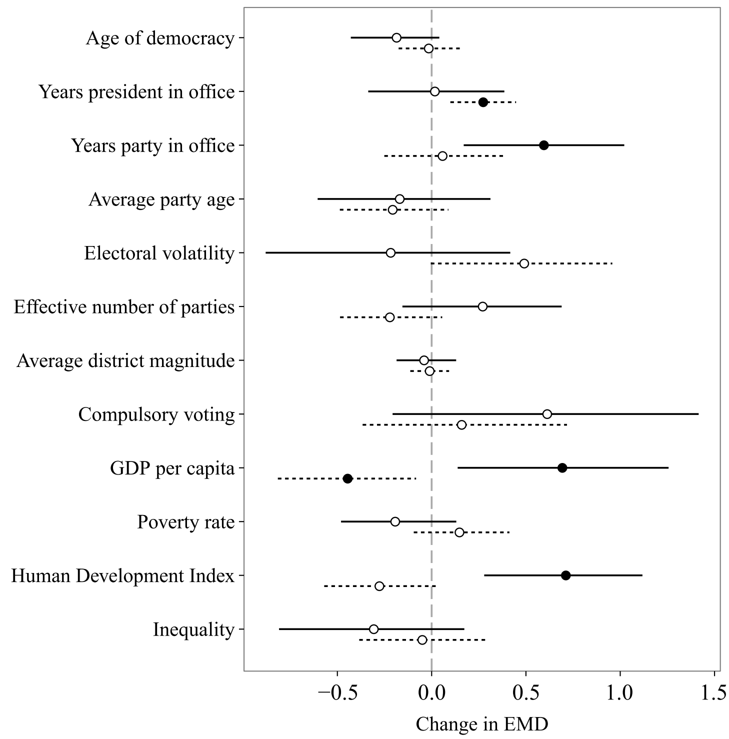

Figure 4. The correlates of mass–elite congruence in Latin America. Values represent the predicted effect on congruence of shifting each variable across its interquartile range, based on regression estimates reported in the online appendix. For each variable, the top value uses the multidimensional EMD measure and the bottom value measures the EMD with just left–right placement. Lines represent the bootstrapped 95% confidence intervals. *Sources*: See online appendix.

activists seeking democratic reforms. They are fundamental if we wish to understand democratic representation and the contexts that help improve it. But to understand congruence and how it varies, we need to be able to measure it accurately.

The measures that currently prevail in political science— difference in means and CDF or PDF overlap—are woefully inadequate. They throw out sometimes enormous amounts of information in the data, they require arbitrary quantization choices that affect the measure, and they can only compute congruence along a single dimension. In contrast, the EMD resolves these problems. Because it uses signatures, the EMD does not require quantization or any other transformation of the data. It uses all of the information in the data and it works in high dimension. There are good theoretical reasons to prefer the EMD, and our simulations offer good empirical reasons as well.

The choice of measure is consequential. Measuring mass–elite congruence with the EMD affects our substantive claims. Whereas prior measures suggested a link between electoral rules and congruence, our reanalysis with the EMD uncovers no such relationship. We find no reason to think that PR systems yield more congruent legislatures than majoritarian ones. Whereas prior measures found positive correlations in Latin American democracies between congruence on the one hand and party system institutionalization and economic prosperity on the other, we find no such relationships. Instead, we find that socioeconomic factors like development are negatively associated with congruence.

The EMD gives a more accurate measure of congruence than those currently being used in political science. Moreover, it is simple to implement and offers scholars the ability to measure congruence in multiple dimensions. It should become the standard way we compute the distance between mass and elite preference distributions. Other research on representation could also employ the EMD: instead of comparing distributions of preferences, many studies of responsiveness compare mass opinion with policy outcomes (Bartels 2008; Gilens 2012; Canes-Wrone 2015). Here too the EMD can improve inferences.

More broadly, the EMD could be adopted whenever researchers want to compare distributions. Breunig and Jones (2011) provide an extended discussion of how these kinds of comparisons can be used to test theories about budgetary processes. Scholars of U.S. politics have been debating whether American voters and political elites have become more polarized over time, and could employ the EMD to study long-term ideological shifts (e.g., Fiorina and Abrams 2008). Observers of Russian elections have compared the distribution of precincts' reported turnout to uncover electoral fraud, and scholars of election forensics could use the EMD to measure similar electoral deviations from a baseline distribution (e.g., Myagkov, Ordeshook, and Shakin 2005). Studies of between-group inequality could also use the EMD to quantify differences in income distributions across ethnic groups (Baldwin and Huber 2010). For these and many other problems, the EMD is a useful and reliable way for political scientists to compare distributions.

Supplementary material

For supplementary material accompanying this paper, please visit
https://doi.org/10.1017/pan.2017.2.

References

Aggarwal, C. C., A. Hinneburg, and D. A. Keim. 2001. On the surprising behavior of distance metrics in high dimensional space. In *Database theory—ICDT 2001*, ed. J. Van den Bussche and V. Vianu. Berlin: Springer, vol. 1973, pp. 420–434.

Alon, N., S. Cosares, D. S. Hochbaum, and R. Shamir. 1989. An algorithm for the detection and construction of monge sequences. *Linear Algebra and its Applications* 114-115:669–680.

Andeweg, R. B. 2011. Approaching perfect policy congruence: measurement, development, and relevance for political representation. In *How democracy works: political representation and policy congruence in modern societies*, ed. M. Rosema, B. Denters, and K. Aarts. Amsterdam: Amsterdam University Press, pp. 39–52.

Baldwin, K., and J. D. Huber. 2010. Economic versus cultural differences: forms of ethnic diversity and public goods provision. *American Political Science Review* 104(4):644–662.

Bartels, L. M. 2008. *Unequal democracy*. Princeton, NJ: Princeton University Press.

Beck, T., G. Clarke, A. Groff, P. Keefer, and P. Walsh. 2001. New tools in comparative political economy: the database of political institutions. *World Bank Economic Review* 15(1):165–176.

Bellman, R. E. 2010. *Dynamic programming*. Princeton, NJ: Princeton University Press.

Blais, A., and M. A. Bodet. 2006. Does proportional representation foster closer congruence between citizens and policy makers? *Comparative Political Studies* 39(10):1243–1262.

Bormann, N.-C., and M. Golder. 2013. Democratic electoral systems around the world, 1946-2011. *Electoral Studies* 32(2):360–369.

Bornschier, S. 2013. Trayectorias históricas y responsiveness del sistema de partidos en siete países de América Latina. *América Latina Hoy* 65:45–77.

Breunig, C., and B. D. Jones. 2011. Stochastic process methods with an application to budgetary data. *Political Analysis* 19(1):103–117.

Budge, I., and M. D. McDonald. 2007. Election and party system effects on policy representation: bringing time into a comparative perspective. *Electoral Studies* 26(1):168–179.

Buquet, D., and L. Selios. 2017. Political congruence in Uruguay, 2014. ed. A. Joignant, M. Morales, and C. Fuentes. *Malaise in representation in Latin American Countries: Chile, Argentina, Uruguay*. New York: Palgrave Macmillan, pp. 187–210.

Calvo, E., and M. V. Murillo. 2013. When parties meet voters: assessing political linkages through partisan networks and distributive expectations in argentina and chile. *Comparative Political Studies* 46(7):851–882.

Canes-Wrone, B. 2015. From mass preferences to policy. *Annual Review of Political Science* 18(1):147–165.

Converse, P. E., and R. Pierce. 1986. *Political representation in France*. Cambridge, MA: Harvard University Press.

Dalton, R. J. 1985. Political parties and political representation: Party supporters and party elites in nine nations. *Comparative Political Studies* 18(3):267–299.

Deza, E., and M.-M. Deza. 2006. *Dictionary of distances*. Amsterdam: Elsevier.

Ezrow, L. 2007. The variance matters: How party systems represent the preferences of voters. *Journal of Politics* 69(1):182–192.
Ferland, B. 2016. Revisiting the ideological congruence controversy. *European Journal of Political Research* 55(2):358–373.
Fiorina, M. P., and S. J. Abrams. 2008. Political polarization in the american public. *Annual Review of Political Science* 11(1):563–588.
Gandrud, C. 2015. Gallagher Electoral Disproportionality Data. Available at http://bit.ly/Ss6zDO, last updated 20-MAR-2015. Accessed 21-MAR-2016.
Gilens, M. 2012. *Affluence and influence*. Princeton, NJ: Princeton University Press.
Golder, M., and G. Lloyd. 2014. Re-evaluating the relationship between electoral rules and ideological congruence. *European Journal of Political Research* 53(1):200–212.
Golder, M., and J. Stramski. 2010. Ideological congruence and electoral institutions: conceptualization and measurement. *American Journal of Political Science* 54(1):90–106.
Huber, J. D., and G. B. Powell. 1994. Congruence between citizens and policymakers in two visions of liberal democracy. *World Politics* 46(3):291–326.
Joignant, A., C. Fuentes, and M. Morales, eds. 2017. *Malaise in representation in Latin American Countries: Chile, Argentina, Uruguay*. New York: Palgrave Macmillan.
Kitschelt, H., K. A. Hawkins, J. P. Luna, G. Rosas, and E. J. Zechmeister. 2010. *Latin American party systems*. Cambridge: Cambridge University Press.
Levina, E., and P. Bickel. 2001. The Earth mover's distance is the mallows distance: some insights from statistics. *Proceedings of ICCV 2001*, pp. 251–256.
Ling, H., and K. Okada. 2007. An efficient Earth mover's distance algorithm for robust histogram comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(5):840–853.
Luna, J. P. 2014. *Segmented representation: political parties in unequal democracies*. Oxford: Oxford University Press.
Luna, J. P., and E. J. Zechmeister. 2005. Political representation in latin america: a study of elite-mass congruence in nine countries. *Comparative Political Studies* 38(4):388–416.
Lupu, N. 2016. *Party brands in crisis: partisanship, brand dilution, and the breakdown of political parties in Latin America*. Cambridge: Cambridge University Press.
Lupu, N., L. Selios, and Z. Warner. 2016. Replication Data for: A New measure of Congruence: The Earth Mover's Distance. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NO90AJ, Harvard Dataverse.
Lupu, N., and Z. Warner. 2017. Mass-elite congruence and representation in Argentina. ed. A. Joignant, M. Morales, and C. Fuentes. *Malaise in representation in Latin American Countries: Chile, Argentina, Uruguay*. New York: Palgrave Macmillan, pp. 281–302.
Mainwaring, S., A. M. Bejarano, and E. Pizarro Leongómez, eds. 2006. *The crisis of demcoratic representation in the Andes*. Palo Alto, CA: Stanford University Press.
Mainwaring, S., and T. R. Scully, eds. 1995. *Building democratic institutions: party systems in Latin America*. Palo Alto, CA: Stanford University Press.
McDonald, M. D., and I. Budge. 2005. *Elections, parties, democracy: conferring the median mandate*. Oxford: Oxford University Press.
McDonald, M. D., S. M. Mendes, and I. Budge. 2004. What are elections for? Conferring the median mandate. *British Journal of Political Science* 34(1):1–26.
Merrill, S. III, B. Grofman, and J. Adams. 2001. Assimilation and contrast effects in voter projections of party locations: evidence from Norway, France, and the USA. *European Journal of Political Research* 40(2):199–221.
Miller, W. E., and D. E. Stokes. 1963. Constituency influence in congress. *American Political Science Review* 57(1):165–177.
Monge, G. 1781. Mémoire sur la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pp. 666–704.
Myagkov, M., P. C. Ordeshook, and D. Shakin. 2005. Fraud or fairytales: Russia and ukraine's electoral experience. *Post-Soviet Affairs* 21(2):91–131.
O'Donnell, G. A. 1994. Delegative democracy. *Journal of Democracy* 5(1):55–69.
Otero Felipe, P., and J. A. Rodríguez Zepeda. 2014. Vínculos ideológicos y éxito electoral en América Latina. *Política y gobierno* 21(1):159–200.
Powell, G. B. Jr. 2000. *Elections as instruments of democracy: majoritarian and proportional visions*. New Haven: Yale University Press.
Powell, G. B. Jr. 2004. Political representation in comparative politics. *Annual Review of Political Science* 7(1):273–296.
Powell, G. B. Jr. 2006. Election laws and representative governments: beyond votes and seats. *British Journal of Political Science* 36(2):291–315.

Powell, G. B. Jr. 2009. The ideological congruence controversy: the impact of alternative measures, data, and time periods on the effects of election rules. *Comparative Political Studies* 42(12):1475–1497.

Powell, G. B. Jr, and G. S. Vanberg. 2000. Election laws, disproportionality and median correspondence: implications for two visions of democracy. *British Journal of Political Science* 30(3):383–411.

Rubner, Y., C. Tomasi, and L. J. Guibas. 2000. The Earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision* 40(2):99–121.

Saiegh, S. M. 2015. Using joint scaling methods to study ideology and representation: evidence from Latin America. *Political Analysis* 23(4):363–384.

Selios, L. 2015. Receptividad democrática en América Latina: un análisis de congruencia ideológica en perspectiva diacrónica. Paper presented at VIII Congreso Latinoamericano de Ciencia Política (ALACIP).

Siavelis, P. M. 2009. Elite-mass congruence, partidocracia and the quality of chilean democracy. *Journal of Politics in Latin America* 1(3):3–31.

Stokes, S. C. 2001. *Mandates and democracy: neoliberalism by surprise in Latin America*. Cambridge: Cambridge University Press.

Urbanek, S., and Y. Rubner. 2015. *Emdist: earth mover's distance*., R package version 0.3-2.

Wang, F., and L. J. Guibas. 2012. Supervised earth mover's distance learning and its computer vision applications. ed. A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid. *Computer vision—ECCV 2012*, vol. 7572, Berlin: Springer, pp. 442–455.

Zechmeister, E. 2006. What's left and who's right? A q-method study of individual and contextual influences on the meaning of ideological labels. *Political Behavior* 28(2):151–173.