5-2015

The Good Judgment Project: Identifying Power Constraints to Improve Accuracy in Geopolitical Forecasting

Morgan Motzel
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/wharton_research_scholars

Part of the Business Commons

https://repository.upenn.edu/wharton_research_scholars/128

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/wharton_research_scholars/128
For more information, please contact repository@pobox.upenn.edu.
Abstract
This paper sets out to explore the relationships between institutional constraints and predictability in geopolitical forecasting. Despite the increasing complexity of our world today, researchers have found that institutional rules and norms still function to influence human behavior, and, therefore, the presence of well-functioning institutions may lead to greater stability and certainty in predicting world events.

Using forecasting data from the Good Judgment Project's recent prediction tournaments, we test the change in predictability—including accuracy, confidence, and difficulty—against the experimental constraints of diplomatic ties, rule of law, effective democracy, trade dependence, and freedom of the press. Our hypothesis is that each of these institutions, together and in conjunction with one another, are effectively able to constraint power player political behavior and reduce uncertainty in the geopolitical realm.

We find that all of the constraints, except for diplomatic ties, actually have a negative correlation with prediction accuracy. Democracy is the strongest negative correlation between the level of constraint and prediction accuracy. We propose that one possible explanation for this result is due to a potential quadratic relationship between democracy and predictability, such that countries who are transitioning from autocracy actually become less predictable than those that are autocratic. Further research in a larger sample set is needed to test this new hypothesis.

Disciplines
Business
The Good Judgment Project

Identifying Power Constraints to Improve Accuracy in Geopolitical Forecasting

Morgan Motzel
Wharton Research Scholar
University of Pennsylvania
Class of 2015
# Table of Contents

Table of Contents ............................................................................................................. 2

Abstract .............................................................................................................................. 4

Introduction to the Topic ................................................................................................. 5

  The Good Judgment Project ......................................................................................... 6
  Motivation ....................................................................................................................... 8

Literature Overview ......................................................................................................... 10

Research Methodology ................................................................................................. 14

  Sample of Observation ................................................................................................. 14
  Prediction Data ............................................................................................................... 15
  Dependent Variables .................................................................................................... 16
  Independent Variables ................................................................................................ 19

Identifying Constraints ................................................................................................. 21

  Five Institutional Constraints ..................................................................................... 22
  Applying Constraint Weights ...................................................................................... 25

Limitations of Investigation .......................................................................................... 28

Statistical Results ......................................................................................................... 30

  Focus on Accuracy ....................................................................................................... 32
  The Democracy Question ............................................................................................. 34
  Summary of Findings ................................................................................................... 36
Abstract

This paper sets out to explore the relationships between institutional constraints and predictability in geopolitical forecasting. Despite the increasing complexity of our world today, researchers have found that institutional rules and norms still function to influence human behavior, and, therefore, the presence of well-functioning institutions may lead to greater stability and certainty in predicting world events.

Using forecasting data from the Good Judgment Project’s recent prediction tournaments, we test the change in predictability—including accuracy, confidence, and difficulty—against the experimental constraints of diplomatic ties, rule of law, effective democracy, trade dependence, and freedom of the press. Our hypothesis is that each of these institutions, together and in conjunction with one another, are effectively able to constraint power player political behavior and reduce uncertainty in the geopolitical realm.

We find that all of the constraints, except for diplomatic ties, actually have a negative correlation with prediction accuracy. Democracy is the strongest negative correlation between the level of constraint and prediction accuracy. We propose that one possible explanation for this result is due to a potential quadratic relationship between democracy and predictability, such that countries who are transitioning from autocracy actually become less predictable than those that are autocratic. Further research in a larger sample set is needed to test this new hypothesis.
Introduction to the Topic

We are living in the Information Age, characterized by a universal reliance on research studies and immediate news coverage. Technology is providing humans with the capability to overcome barriers of time, distance, location, and physical human capacities in order to process information and make decisions. An increasing percentage of activity in the developed world has shifted into the “knowledge economy” whereby production and service offerings rely more heavily on intellectual capabilities than physical inputs or natural resources. The ability to exploit what management gurus call “big data” has permeated the way business, economic, and political strategists endeavor to accurately model the future and pursue courses of action which are statistically most likely to provide a desirable outcome.

Yet, across every academic discipline and economic sector, we continue to observe significant error in predicting future states of the world. This trend even holds true for so-called experts who are hired on the basis of their ability to calculate potential opportunities and threats. While some prediction errors are trivial in the grand scheme, many ensue considerable financial losses to the hard-earned dollars of organizations, investors, and taxpayers. Even graver are the human consequences that can result from forecasting failures, such as consequences from unexpected political coups, ethnic genocides, market failures, and natural disasters. More accurate prediction of geopolitical events such as these has the potential to save not just money but also human livelihoods. Places which are the most vulnerable to unpredictable circumstances seem to incur them the most often.
Researchers have begun to explore what actually constitutes good judgment in forecasting future events and why experts are often wrong in their forecasts. Four years ago, with some of these questions still left unanswered, an experiment called the Good Judgment Project set out with a mission to “harness the wisdom of the crowd to forecast world events.”¹ What could everyday people like you and me know about nuclear bombs and economic downturns? It came as a shock to many that the average forecasts of amateur individuals have been surprisingly accurate.

**The Good Judgment Project**

The Good Judgment Project (GJP) is a four-year research study begun in 2011 by psychology and management professors Phillip Tetlock, author of the award-winning Expert Political Judgment; Barb Mellers, an expert on judgment and decision-making; and Don Moore, an expert on overconfidence. The GJP was supported by research teams at the University of Pennsylvania and the University of California Berkeley including experts in psychology, economics, and statistics.

The GJP was organized as part of a government forecasting tournament sponsored by the Intelligence Advanced Research Projects Activity (IARPA) division of the US Office of the Director of National Intelligence. The specific tournament is IARPA’s Aggregative Contingent Estimation Program (ACE), which aims “to dramatically enhance the accuracy, precision, and timeliness of forecasts for a broad range of event types, through the development of

---

advanced techniques that elicit, weight, and combine the judgments of many intelligence analysts.”2 In essence, the goal of this initiative was to source better ways of collecting and manipulating raw forecast data to in order to ultimately improve the government’s performance in geopolitical forecasting.

In 2011, the inaugural year of the ACE tournament, the GJP asked hundreds of people to join its team and help predict the likelihood of a broad set of nearly 200 global events that were of interest to the U.S. intelligence community. The GJP’s international group of forecasters signed up voluntarily and represented a diverse range of careers, backgrounds, and levels of expertise. The GJP beat out four other university-based groups in that year’s tournament, taking the raw individual forecast data and using sophisticated aggregation algorithms to combine them into the most accurate crowd-sourced forecasts possible. By the end of the tournament’s second year, the GJP team had performed so well that it became the only group to receive IARPA funding for the remainder of the ACE program.3

Each year, the GJP strives to improve its aggregate forecasts, implementing new strategies including cognitive and psychological training and introducing knowledge-sharing teams in addition to continually improving its algorithmic model. The successful strategies of the GJP can and have been used by the US Intelligence Community to improve the government’s ability to forecast global events. The massive quantities of prediction data

---


collected by the GJP provide an unparalleled opportunity for deeper analysis on how we can go about improving our country's ability to forecast.

**Motivation**

The art and science of prediction are deeply nested in the fields of international development and diplomatic relations. As a future professional within these disciplines, my ability to rely on proven rational and statistical models to precisely anticipate future global events will be critical to the success of any initiative I intend to pursue. The extent to which global leaders are able to design effective solutions for political peace in Syria, adequate health and sanitation in Bangladesh, and agricultural sustainability in Angola will depend heavily on the predictions of geopolitical forecasters to inform the allocation of resources and risks to where they might be most productive.

I believe that the findings from this study have great potential to inform the methods and techniques that intelligence analysts will be taught to employ as they engage in the formation of probabilities and conjectures regarding significant global events. More accurate predictions from these experts will ultimately enable decision-makers from across all sectors of society to make more efficient use of limited resources and become more effective in safeguarding the lives and livelihoods of human persons.

We hypothesize that individuals who forecast on a regular basis as part of their profession can be taught to make better predictions by improving their dissecting of forecasting prompts—real or experimental. Historical analysis of past predictions enables us to study where we have gone wrong (or right) and adjust our future strategy accordingly.
The identification of relevant institutional protocols and normal which constrain geopolitical outcomes is an important first step to understanding predictability. Essentially, if we can uncover specific question tags including countries, regions, or question types or combinations thereof which have seen statistically significant unpredictability, we should be able to improve our strategy going forward.

By analyzing both the correlation and potential causation between a question properties and the prediction quality of forecasters within the Good Judgment Project, we might uncover previously unknown insights into the most effectual cognitive strategies for achieving optimal results. If we succeed in our goal, we will be one step closer to the optimal training frontier in preparing accurate forecasters of global events.
Making accurate predictions in a world of increasingly complex geopolitics is a skill that brilliant minds from every academic discipline have sought to pinpoint. In particular, political scientists and international relations theorists have endeavored for centuries to explain the randomness of global events with a plethora of models, theories, and frameworks. While they have greatly improved the human ability to understand and articulate the complex nature of geopolitics, there has been little concrete progress in the way of anticipating future world events.

A common answer to this unsolvable problem is that of complexity theory which understands that processes having a large number of independent agents will interact and order themselves spontaneously into a coherent and stable system of behavior. Complexity theory which suggests that “organizations” (such as the sphere of international politics in the case of this paper) cannot be conceptualized in a linear or additive frame of mind. Instead, numerous experts agree that complex structures must be understood as both dynamic and adaptive. The theory also implies that such complex behavior is for the most part unpredictable and uncontrollable. The rapid pace of communication and evolving technologies today only serves to further entangle the interconnected web of actors within a constantly-changing global system.

Political scientist Stanley Feder suggests that creating prediction models, however nonlinear and intricate, still cannot replace what is to be learned by real-world experience in political science: “Analytical methods alone will not guarantee that policy makers and academics will not be surprised by political events. Preventing surprise depends on asking

Literature Overview
the right questions.” Feder argues that the basic value of a geopolitical forecasting lies in preventing surprises as opposed to attempting to pinpoint certainty, which is both futile and counter-productive. One can and should prevent surprises by becoming familiar with each of the moving parts in a complex system and understanding the norms of how a change in one influences a change in another, rather than attempting to predict the aggregate outcome of so many individual actions and motivations.

In geopolitics, these moving parts are often understood to be a wide range of relevant actors, from heads of state, to the military, to businessmen, to leaders of diplomatic entities. Our standard definition of complexity theory would seem to suggest that rather than having all of the power in the hands of a single actor, increasing the number of stakeholders interacting in a given situation will disperse the base of power, ultimately resulting in greater political stability. According to the international politics researcher Neil E. Harrison, decentralized decision-making further increases complexity. In his book *Complexity in World Politics*, Harrison writes that “complexity views politics as emerging from interactions among interdependent but individual agents within evolving institutional frameworks.”

The idea of institutions providing some guiding structure to the chaotic world of complexity theory is important to social science theory. Douglas North, a Nobel Prize-winning economist from the New Institutional Economics (NIE) school, defines institutions as the “the rules of the game” of a given society which are humanly-devised and govern

---


human interaction. Institutions can include both formal constraints such as statue law, common law, regulation as well as informal constraints including conventions, norms of behavior, and self-imposed codes of conduct. Arguing for the theory of institutional realism, political scientist Robert Graftstein proposes that individuals actively choose to participate in pre-existing institutions which reduce uncertainty and transaction costs.

While the various theories regarding institutions (historicism, realism, etc.) have been widely discussed and debated in political science literature, current scholarship has not yet closed the gap to approaching a clear and comprehensive understanding of how the rules and norms provided by institutions can play a role in increasing predictability. These institutional “constraints” as North defined then seen to follow Feder’s notion of forecasting not as an exercise to pre-determine the outcome, but rather understanding what norms do exist amidst the complexity that are likely to influence its behavior in a certain way. Thus strong institutions with programmed patterns of behavior may be able to serve as instruments of prediction in an otherwise complex system.

In the case of geopolitical forecasting, it is common to think of domineering political power players who make bold actions facing little to no constraint on their authority. Complexity theory suggests that the more actors who are involved in sharing authority with these power players will tend the collective outcome toward a “coherent and stable” system of behavior. Strong macro-level institutions can perhaps be those actors who not only disperse power away from a concentrated source but who also incorporate constraints in

the form of laws and norms on a power player’s otherwise rogue behavior and bring an entire community toward a more predictable outcome.

A wealth of research describes the nature of specific political and social institutions and how each operates through a combination of formal and informal norms to influence human behavior and interaction. A detailed analysis of the each of these types of institutions and their relative effectiveness is beyond the scope of this paper. We identified five well-researched global institutions which we intuitively believe may be capable of constraining power players from igniting geopolitical stress, either actively (invading a neighbor’s border) or passively (not controlling the growth of insurgency). If the institutions we have chosen to test fulfill their duty as effective constraints, geopolitical forecasting will become much more predictable in their presence.

The first experimental constraint, diplomatic ties, reflects research suggesting that international organizations can serve either a coercive or a support role for individual countries, as long as they comply with rigorous standards. The second, rule of law, consider studies which contrast corrupt and unjust legal institutions with those that are strictly bound to treat all men and women equal under the law. The third, effective democracy, implies the (historically-proven) potential for protest, impeachment, or even revolution in the even that a leader is acting in the explicit interests of the people. Fourth, trade dependence, echoes an ever-widening body of research suggesting that countries who rely almost exclusively on other nations for their revenues may be economically forced to enter into certain relationships. Finally, freedom of the press alludes to the widespread reputational damage which can quickly occur if a leader acts in a way that others, even small social groups, deem to be unfavorable.
Research Methodology

In brief, this paper will utilize linear regression methods to analyze numerical and descriptive data from the Good Judgment Project forecasting records, testing for relationships between observed levels of predictability and details of question content (both direct and indirect). Specifically, this paper endeavors to answer the following question:

*Are there certain institutions which are able to consistently constrain the agency of a power player into a more predictable pattern?*

A more detailed description of research methodology follows in the subsections below.

Sample of Observation

The majority of the data to be studied in this analysis come directly from the Good Judgment Project records itself. The Good Judgment Project question database and forecasting results is a place where we can begin to generate hypotheses about predictability. Given the wealth of data we have available, it should be very worthwhile to compare the questions forecasters predict well with those they predict poorly. Specifically, this paper will observe forecasting questions which “closed” during Year 3 of the GJP, conducted from August 2013 to May 2014. (All questions run for a pre-specified time frame, after which the final predictions of each forecaster will be scored.) Extreme outlier questions which were removed from analyses conducted by the GJP research team were also taken out of this data set. Furthermore, due to the research method chosen for this paper, questions
which did not identify one or two primary countries as subjects in the prompt were also eliminated. These initial data-cleaning measures left a sample of approximately 130 questions to test.

Each of these questions were identified by direct characteristics (country actor and question family) as well as indirect characteristics (relevant institutions). These descriptive measures formed the independent variables for these regressions. The dependent variable was determined to be observed predictability, based on forecaster performance in our sample set of GJP questions. We created three distinct measures of predictability (difficulty, confidence, and accuracy) to be tested in these regressions. The analysis will be primarily driven by a simple one-by-one linear regression of each of the dependent variables—difficulty, confidence, and accuracy—on each of the independent variables—country actor, question family, and institutional constraint and interpreting the results.

**Prediction Data**

Before discussing the formulation of variables, it is important to understand how predictions are actually submitted to the tournament. Almost all forecasting questions are created to be binary, such that one answer is correct and one answer is incorrect and there is no room for ambiguity in between. The questions are constructed around a specific time frame in the real world; whether the hypothetical event does or does not happen by a specifically identified date, will determine the outcome of the question, and thus whether an individual forecast is correct or incorrect.

While the questions are binary, the responses are more flexible. Forecasters are invited to submit a probability percentage ranging from 0% to 100% for each event which
they are predicting. For example, if I am certain that an event will occur, I will answer 100%; if I am less certain, but still fairly confident, I might answer 80%. If I am completely clueless, I would consider answering 50% as it sits perfectly in the middle of a yes or no response.

When scores are calculated to assess how correct the predictions were, there is a function that considers not only if a forecaster was on the correct side of 50% but also how far from 100% or 0% his or her prediction was.

It is also important to take into consideration that, in this tournament, predictions on a single question are made over a long time horizon, with a possibility to change your prediction during the duration of the question. Unlike the expert difficulty measurement which must be taken ex ante, measurements which are derived from an individual’s forecast can be seen as flow variables rather than stock variables. Typically in the GJP, variables such as these are reported as longitudinal weighted averages. (For example, if I score a 1 for 50% of the duration of the question, 0.5 for 25% and 0.25 for the remaining 25%, my overall reported score on this variable would be 1*0.5 + 0.5*0.25 + 0.25*0.25 = 68.75). Moreover, there are over 1000 forecasters in Year 3 of the tournament, a large sample which—in the spirit of the “wisdom of the crowds” theory—affords the opportunity for us to average the individual forecasting variables that correspond with each question for a more robust indicator of overall trends occurring.

**Dependent Variables**

Our research intends to test and compare the level of predictability of different post-mortem geopolitical forecasting questions in order to develop new insights regarding the anticipatory prediction of global events in the real world. The notion of predictability is
generally understood as the degree to which a correct prediction or forecast of a system’s state can be made either qualitatively or quantitatively; however, it is a complex idea that is actually quite hard to conceptualize and thus to concisely articulate. Precisely for this reason, we have chosen to identify and test three distinct notions of predictability—difficulty, confidence, and accuracy—to assess the phenomenon of predictability more carefully.

The variable of question **difficulty** aims to pinpoint how easy or hard a certain question is to predict, assuming a forecaster has all of the relevant information. These ratings of question difficulty are collected from subject-matter experts who place each question on a seven-point scale ranging from most to least difficult to accurately predict. The measure we will formulate for expert difficulty will be a mean of all expert difficulty ratings assigned to a certain question. Here high values (7) will denote that the question is easy and low values (1) that it is difficult. The rating task is done ex ante to avoid hindsight bias; however, this makes the rating task particularly challenging because, in effect, we are asking experts to predict unpredictability.

Following this logic, the variable of forecaster **confidence** intends to understand with how much certainty each prediction was made. Like predictability, the qualitative notion of confidence can take on a number of nuanced quantitative meanings. For the purposes of this analysis, we chose to create a scaled confidence index which would incorporate the statistical measures of extremity, entropy, and variance. For each of these variables (as stated in the paragraph above), we will take the weighted mean across all days the question is open and standardize it such that the variables have a mean of 0 and a standard deviation of 1. From these values, we create the confidence composite variable which is calculated as follows: scaled confidence index = mean extremeness – mean entropy – mean variance.
Extremeness is an evaluation of the distance of one’s prediction from 50%. If a forecaster predicts a 0% or 100% outcome, it can be assumed that the individual is quite confident in the outcome he or she has chosen, whereas a 50% prediction implies maximum uncertainty. High variance among the entire pool of forecasters denotes a lack of collective confidence, while low variance suggests that the majority of forecasters are in relative agreement. Finally, statistical entropy measures the dispersion of forecasters along the range of 0% to 100% outcomes. Entropy is highly correlated with variance, but rather than considering the range between forecasters (as variance does), entropy considers the location of dense pockets of similar predictions or very sparse ranges of very few predictions as a distinct way of assessing uncertainty. Because extremeness is an indication of confidence, and variance and entropy suggest a lack of confidence, the index measure has been formulated to subtract standardized variance and entropy values from the standardized extremeness value—a high index value represents confidence.

The final, and perhaps most intuitive, approach to assessing predictability is considering the accuracy of geopolitical forecasters in answering each question correctly. As mentioned previously, the forecaster’s score for each question awarded on the basis of “ending up on the right side of 50” as well as the certainty of each prediction. For example, if an event occurs, and my prediction was 100%, I will get the maximum score for that individual question. The technique used to make this method of assessment work is called a Brier Score, a measure which includes both an assessment for the correct outcome (i.e. +/- 0.5 for a binary question) and certainty or that outcome (i.e. distance from 0.5) and then taking a weighted average of a single individual's predictions across the duration of the question in case the predictor chose to update along the way, as most do.
The Brier Score traditionally reports a 0 as the best score and 1s and 2s to be very poor showings, so for the sake of consistency with other dependent variables, we will use the reverse of the Brier Score for the purposes of this analysis. The reason that accuracy cannot be our only measure of predictability is because the use of Brier Score as the dependent variable limits our analysis to only post-mortem evaluations, given that one cannot calculate the Brier Score until after the time frame for the question closes. Predictability, however, is a relevant factor before the outcome of a question is known; this understanding is captured by the previous measure of confidence. Moreover, Brier Score is not a perfect predictive mechanism for understanding real world geopolitical conclusions because the Brier scoring treats errors of under- and over-prediction as equally bad.

Independent Variables

The direct independent variable which is easiest to identify is “country actor.” All of the forecasting questions which were left in our data set after its initial cleaning had one or two countries identified in the context of the prompt. In the case of two country actors (e.g. *Will Iran and Russia officially sign an agreement regarding the exchange of oil for goods and services before 10 May 2014?*), we originally set out to repeat the question twice in our data set—once with each country as the primary actor. Doubling the presence of certain questions, however, created misleading results in the data, so instead we chose to identify the “least constrained” country as the primary actor (more explanation to follow on constraints below). After identifying the country actor in each question, we also assigned each question to one of the following seven regions: North America, Latin America, Europe, Africa, the Middle East, South Asia, and Southeast & East Asia.
The second direct variable we created, “question family,” was more subjective to identify. The GJP questions in our records had already been (subjectively) tagged with one of 18 question groupings, but this would have proven too granular for our small group of 130 observations. After carefully studying all of the questions in our data set, the following families were identified as mutually exclusive and collective exhaustive: conflict, domestic, economic, negotiation, and leader change. A number of the old question groupings were channeled directly into one larger question family (e.g. all questions previously identified as “Elections” or “Leader Entry/Exit” both became identified as “Leader Change” in the new family scheme.)

Conflict questions included the subjects of war, border aggression, deployment of troops, short-term ceasefire agreements, and nuclear threats. Domestic questions referenced local policy and legislative concerns, internal country factions and insurgency, national emergencies, and human rights issues including disease and refugees. Economic questions explored trade agreements, monetary policy, sovereign debt, interest rates, and commodity markets. Negotiation questions referenced international agreements, treaties, sanctions, and long-term peace talks. Finally, leader change questions included the topics of election fairness and results, political coups, and the removing or vacating of office for any reason.

The final independent variable is an indirect measure called an “institutional constraint” and is the primary measure that will be used to test our research question. The following section will describe the intuition and construction behind the creation of this unique variable.
Identifying Constraints

Following our review of the literature, we hypothesize that each of the following institutions may have the influence to limit the power of a given political actor and thus reduce the scope of acceptable choices and corresponding outcomes for his/her action: diplomatic ties, rule of law, effective democracy, trade dependence, & press freedom. By limiting the range of viable outcomes in a given geopolitical scenario, these institutions act as constraints on the behavior of a single powerful actor (in our analysis here, we identify the country itself as this actor). The presence of these classified institutional constraints (both independently and in conjunction with one another) we hypothesize will ultimately decrease the level of unpredictability for a given geopolitical outcome.

In order to scientifically evaluate how these proposed constraints can impact geopolitical predictability, it is critical that we measure these constraints as thoughtfully, systematically, and objectively as possible. For this reason, all constraints will be fit to a 0-100 scale in order to standardize values for the purpose of statistical analysis. We have also chosen to assign the relevant level of institutional constraint to the current political dynamics of each country (that is, to our primary actor in the geopolitical scenario). It is critical that each relevant constraint be linked to the country in question rather than the question itself because our hypothesis assumes a change in the power held by a country’s predominant political actor as civic institutions grow stronger and begin to disperse power away from the government. Our choice to organize data in this way also leaves our analysis open to the (likely) possibility that the central political figure or national regime of a
particular country changes; by assigning the institutional constraint index to a country, new leaders will not be misrepresented by the qualities of their predecessor.

**Five Institutional Constraints**

Diplomatic ties or **diplomacy** act as a constraint on political power when international pressures place limits on local or national sovereignty. We measure this constraint using data from the Center for Systemic Peace and its Integrated Network for Societal Conflict Research (INSCR). The index is constructed based on the following data:

Memberships in Conventional Intergovernmental Organizations (CIO), country data coded every fifth year, 1952-1997, denotes individual country membership in a) federations of inter-government organizations (1); b) universal membership organizations (39); c) inter-continental membership organizations (52); and d) regionally-defined membership organizations (288); characterized by “autonomous international governmental organizations of a non-profit nature”

This index has not been updated since 1997, however, it still should represent a useful tool for our analysis. There has been relatively little fluctuation in relevant IGO membership over the course of the past 20 years given that the major organizations were created much earlier in the 20th century.

The **rule of law** acts as a constraint on political power because it constitutes a civil order, strong legal enforcement, and just prosecution. We measure this constraint using the

---

Rule of Law index created by the World Justice Project. The index is constructed based on the following data:

The WJP Rule of Law Index offers a detailed, multidimensional view of the extent to which countries adhere to the rule of law in practice. The Index measures the rule of law using 47 indicators organized around 8 themes: constraints on government powers, absence of corruption, open government, fundamental rights, order and security, regulatory enforcement, civil justice, and criminal justice.\(^8\)

Unfortunately, this index is an incomplete indicator because it only includes data analysis compiled for 99 countries. A number of nations that are frequently identified as subjects in geopolitical forecasting questions (including Syria, North Korea, and the Democratic Republic of the Congo) are not evaluated or assigned a Rule of Law ranking by this index. Other potential rule of law indices we considered utilizing (but that were not as robust as the WJP) had identified all three of these countries near the very bottom of their lists, so we have taken the liberty of assigning these three nations each with an approximate rule of law score of 30, which is near the bottom of the range for this index.

Independently from the rule of law, effective democracy can also act as a constraint on political power because it evokes unrestrained participation and agency expressed by the citizenry. We will measure this constraint using the Democracy index created by The Economist. The index is constructed based on the following data:

The Democracy Index is based on five categories: electoral process and pluralism; civil liberties; the functioning of government; political participation; and political culture. Based on their scores on a range of

indicators within these categories, each country is then categorized as one of
four types of regime: “full democracies”; “flawed democracies”; “hybrid
regimes”; and “authoritarian regimes.”

This data set is complete, robust, and up to date.

Meanwhile in the economic sphere, trade dependence serves as a constraint on
political power to the extent of the international goods and commodities trade across
countries. We will measure this constraint using the Merchandise Trade as a percentage (\%) of GDP index source from the World Bank. The index is constructed based on the following:

Merchandise trade as a share of GDP is the sum of merchandise exports and imports divided by the value of GDP, all in current U.S. dollars. Merchandise trade only includes trade in goods, not services nor capital transfers and foreign investments.

This data set is complete, robust, and up to date. In some instances 2014 data was not available, so we inserted 2013 numbers as a close approximation.

Finally, we include freedom of the press as a constraint on political power which can be assessed by the extent, accuracy, and honesty of local media and reporting coverage. We will measure this constraint using the World Press Freedom index created by the Reporters Without Borders. The index is constructed based on the following data:

The Reporters Without Borders World Press Freedom Index ranks the performance of 180 countries according to a range of criteria that include media pluralism and independence, respect for the safety and freedom of

---


journalists, and the legislative, institutional and infrastructural environment in which the media operate.\textsuperscript{11} This data set is also complete, robust, and up to date.

We can imagine other “institutions” which might act as constraining factors on the range of acceptable political activity as well. For example, reliance on foreign aid, religious homogeneity, and progressive social values can undoubtedly limit the agency of individual political leaders to act completely freely. For the purposes of this analysis, however, we chose to focus on constraints that are characterized by institutions which are likely to share the same realm of political power with the country’s Head of State, or another lead national actor. Each of these institutions identified above would be considered legitimate civic actors which can have a broad impact reaching nearly all citizens, if the institution is strong. Practice examples of these for each identified constraint might be the United Nations, the judiciary system, the electoral system, capitalist market forces, the popular press.

\textbf{Applying Constraint Weights}

Yet, while it logically follows to assign these constraint indices to each individual country, we need to still take into account the content of the question itself in evaluating a certain event’s predictability. Different institutional constraints will not necessarily apply equally to different questions. For example, for a question regarding a possible upcoming election, the level of effective democracy in a country is going to be a much more relevant potential constraint than the level of trade dependence.

Table 1: Summary of Institutional Constraints and Accompanying Indices

<table>
<thead>
<tr>
<th>INSTITUTION</th>
<th>COUNTRY INDEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diplomatic Ties</td>
<td>Conventional IGO Membership (Center for Systemic Peace)</td>
</tr>
<tr>
<td>Rule of Law</td>
<td>Rule of Law Index (World Justice Project)</td>
</tr>
<tr>
<td>Effective Democracy</td>
<td>Democracy Index (The Economist)</td>
</tr>
<tr>
<td>Trade Dependence</td>
<td>Merchandise Trade as % of GDP (World Bank)</td>
</tr>
<tr>
<td>Press Freedom</td>
<td>World Press Freedom Index (Reporters without Borders)</td>
</tr>
</tbody>
</table>

To this end, we chose to create an additional variable which attempts to quantify the qualitative traits of the content in each particular forecasting question for the purposes of our analysis. Essentially, this variable will act as a multiplier to ensure that the constraints relevant to each question are given enough weight in determining predictability, while those that are not particularly relevant are not overemphasized.

We call this variable question relevance, and it will be decidedly orthogonal to the country-linked institutional constraint variables described in the previous section. In practice, this means the relevance of each identified constraint is assessed independently of the country in question (i.e. as if the country name was not given). The most systematic way to accomplish this goal will be to automatically assign quantitative question relevance values based on the previously referenced qualitative question family variable, thus avoiding a more haphazard approach to assigning relevance based on each individual question.
Table 2: Institutional Constraint Weightings Assigned by Question Family

<table>
<thead>
<tr>
<th>QUESTION FAMILIES</th>
<th>DIPLM</th>
<th>RULAW</th>
<th>DEMOC</th>
<th>TRADE</th>
<th>PRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conflict</td>
<td>2.0</td>
<td>1.5</td>
<td>2.0</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Domestic</td>
<td>1.5</td>
<td>2.0</td>
<td>2.0</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>Economic</td>
<td>1.5</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Negotiation</td>
<td>2.0</td>
<td>1.5</td>
<td>1.0</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Leader Change</td>
<td>1.0</td>
<td>2.0</td>
<td>2.0</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

The weighting will be assigned on a three-point scale. A “1.0” will indicate that a forecasting question’s outcome will not likely be influenced by this constraint. A “1.5” will indicate that a forecasting question’s outcome is likely to be influenced by this constraint. A “2.0” will indicate that a forecasting question’s outcome will be significantly influenced by this constraint. Essentially, we are aiming to determine to what extent this constraint actually affects this type of question. The following example question illustrates why this weighting is necessary for our analysis: *Will inflation in Japan reach 2 percent at any point before 1 April 2014?* For a prompt like this which falls in the Economics question family, it is easy to see that Press Freedom is unlikely to be an influential constraint for this question. On the flip side, Trade Dependence will be a significant constraint in questions that determine the future of Japan’s economy. By assigning these weights to each constraint, we are strengthening the robustness of our analysis to test for correlation between institutional constraints and predictability.
We will summarize our full variable methodology in the following example forecasting question, illustrated by the inset to the right. After isolating each question, we identify the primary actor country (and associated region) and then sort each question into one of five question families. From there, we refer to our external indices such as the Economist Democracy Index to determine the country actor’s score on each of the five institutional constraints we are testing (all scores are scaled from 0-100). Finally, we multiply each located constraint value with the appropriate weighting based on its question family. This final weighted value, which you see on the far right of the table, will be our primary independent variable in this analysis.

**Limitations of Investigation**

This investigation faces one important potential limitation in that the amount of data observations to be used in conducting this analysis is limited by the quantity of forecasts conducted by the Good Judgment Project. With a smaller sample size, correlations must be particularly high in order to be statistically significant. This also means that the potential relationships identified as a part of this analysis may be essentially false positives when considered outside the scope of this data set. With this limitation in mind, the GJP and I recommend that the results to be discussed in the following section be understood as
correlations recognized within the GJP-specific context that provide a starting point for forecasting hypotheses which can later be tested in the broader field of political science with a larger scale and scope of data collection.
Statistical Results

The initial analysis tested for predictability—difficulty, confidence, and accuracy—against the weighted institutional constraint variables designed for this analysis. Our institutional constraint variables turn out to be poor predictors of prediction difficulty (measured by the expert difficulty rating) and prediction confidence (measured by a composite variable of forecaster extremeness, variance, and entropy). Out of our three dependent variables, prediction accuracy (measured by the Brier Score indicator) turned out to be much more highly correlated to weighted institutional constraints on average.

In contrast with our hypothesis, however, the overall effect was found to be negative rather than positive. In other words, while our initial assumption was that the presence of these institutions would constrain political actors and thus make forecasting questions more predictable, our data suggest instead that the more institutionally constrained a question is, the less likely it is to be predictable. The sum of all weighted constraint values (“total”) showed a small correlation with accuracy ($r =-0.15$, $t(110)=-1.80$, $p=0.07$). In fact, with varying strengths of correlation, all of the constraint variables we tested were negatively associated with accuracy, except for diplomacy which was observed to have a slight positive correlation with accuracy. The strongest of these negative correlations was for democracy ($r =-0.19$, $t(110)=-2.00$, $p=0.05$), suggesting that as the level of democracy in a country increases, forecasters have an increasingly hard time make accurate predictions.
Table 3: Statistical Output of Predictability Measures against Institutional Constraints

**Difficulty by Weighted Constraint**

<table>
<thead>
<tr>
<th>Constraint</th>
<th>R</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diplomacy</td>
<td>6%</td>
<td>0.58</td>
<td>97</td>
<td>0.57</td>
</tr>
<tr>
<td>Rule of Law</td>
<td>6%</td>
<td>0.59</td>
<td>84</td>
<td>0.56</td>
</tr>
<tr>
<td>Democracy*</td>
<td>-6%</td>
<td>-0.60</td>
<td>97</td>
<td>0.55</td>
</tr>
<tr>
<td>Trade Dep</td>
<td>-16%</td>
<td>-1.44</td>
<td>81</td>
<td>0.15</td>
</tr>
<tr>
<td>Free Press</td>
<td>0%</td>
<td>0.02</td>
<td>97</td>
<td>0.98</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>-4%</td>
<td>-0.42</td>
<td>97</td>
<td>0.67</td>
</tr>
</tbody>
</table>

**Confidence by Weighted Constraint**

<table>
<thead>
<tr>
<th>Constraint</th>
<th>R</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diplomacy</td>
<td>-3%</td>
<td>-0.31</td>
<td>110</td>
<td>0.75</td>
</tr>
<tr>
<td>Rule of Law</td>
<td>-6%</td>
<td>-0.57</td>
<td>95</td>
<td>0.57</td>
</tr>
<tr>
<td>Democracy</td>
<td>-8%</td>
<td>-0.79</td>
<td>110</td>
<td>0.43</td>
</tr>
<tr>
<td>Trade Dep</td>
<td>3%</td>
<td>0.27</td>
<td>93</td>
<td>0.79</td>
</tr>
<tr>
<td>Free Press</td>
<td>-9%</td>
<td>-0.92</td>
<td>110</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>-8%</td>
<td>-0.81</td>
<td>110</td>
<td>0.42</td>
</tr>
</tbody>
</table>

**Accuracy by Weighted Constraint**

<table>
<thead>
<tr>
<th>Constraint</th>
<th>R</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diplomacy</td>
<td>10%</td>
<td>1.04</td>
<td>110</td>
<td>0.30</td>
</tr>
<tr>
<td>Rule of Law</td>
<td>-15%</td>
<td>-1.47</td>
<td>95</td>
<td>0.15</td>
</tr>
<tr>
<td>Democracy*</td>
<td>-19%</td>
<td>-2.00</td>
<td>110</td>
<td>0.05</td>
</tr>
<tr>
<td>Trade Dep</td>
<td>-7%</td>
<td>-0.64</td>
<td>93</td>
<td>0.52</td>
</tr>
<tr>
<td>Free Press</td>
<td>-15%</td>
<td>-1.56</td>
<td>110</td>
<td>0.12</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>-17%</td>
<td>-1.80</td>
<td>110</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Focus on Accuracy

These findings that contradicted our initial hypothesis led us to look deeper into the relationship between prediction accuracy and weighted institutional constraints. Our additional independent variables of geographic region and question family offered an opportunity to subset the data to observe the correlations between accuracy and institutional constraint more closely.

The correlations become larger and more significant when looking at smaller subsets of data. When looking across geographies, we see that within the subset region of Southeast and East Asia—which includes numerous questions on China, Japan, the Koreas, Thailand, and Myanmar—there is a fairly strong negative correlation between prediction accuracy and the “total” existence of institutional constraints ($r = -0.34$, $t(32) = -2.08$, $p = 0.05$), observing a particularly significant relationship with democracy ($r = -0.36$, $t(32) = -2.19$, $p = 0.04$).

Table 3: Statistical Output of Accuracy against Institutional Constraints by Region Subset

<table>
<thead>
<tr>
<th>GEOGRAPHIC REGIONS: Accuracy by Weighted Constraint</th>
<th>R</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe Diplomacy</td>
<td>36%</td>
<td>1.67</td>
<td>19</td>
<td>0.11</td>
</tr>
<tr>
<td>Middle East Rule of Law</td>
<td>34%</td>
<td>1.67</td>
<td>21</td>
<td>0.11</td>
</tr>
<tr>
<td>SE &amp; East Asia Rule of Law</td>
<td>-33%</td>
<td>-1.97</td>
<td>32</td>
<td>0.06</td>
</tr>
<tr>
<td>SE &amp; East Asia Democracy*</td>
<td>-36%</td>
<td>-2.19</td>
<td>32</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>SE &amp; East Asia Total</strong>*</td>
<td>-34%</td>
<td>-2.08</td>
<td>32</td>
<td>0.05</td>
</tr>
</tbody>
</table>

We can also analyze the correlations subdivided by question family. The most significant results from this analysis are located within the domestic question family, which
again include the topics of local policy and legislative concerns, internal country factions and insurgency, national emergencies, and human rights issues including disease and refugees. Within this question family, there is a statistically significant negative correlation between prediction accuracy and the “total” existence of institutional constraints (r = -0.38, t(32) = -2.30, p=0.03) as well as with the rule of law constraint in particular (r = -0.41, t(27) = -2.36, p=0.03). This breakdown suggests that domestic question accuracy tends to be negatively correlated with the presence of institutional constraints, whereas economic question accuracy is positively correlated with the presence of institutional constraints.

*Table 4: Statistical Output of Accuracy against Institutional Constraints by Family Subset*

<table>
<thead>
<tr>
<th>QUESTION FAMILIES: Accuracy by Weighted Constraint</th>
<th>R</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leader Change Diplomacy</td>
<td>43%</td>
<td>1.79</td>
<td>14</td>
<td>0.09</td>
</tr>
<tr>
<td>Domestic Rule of Law*</td>
<td>-41%</td>
<td>-2.36</td>
<td>27</td>
<td>0.03</td>
</tr>
<tr>
<td>Economic Rule of Law</td>
<td>42%</td>
<td>1.71</td>
<td>14</td>
<td>0.11</td>
</tr>
<tr>
<td>Domestic Democracy</td>
<td>-31%</td>
<td>-1.87</td>
<td>32</td>
<td>0.07</td>
</tr>
<tr>
<td>Economic Democracy</td>
<td>45%</td>
<td>1.97</td>
<td>15</td>
<td>0.07</td>
</tr>
<tr>
<td>Domestic Total*</td>
<td>-38%</td>
<td>-2.30</td>
<td>32</td>
<td>0.03</td>
</tr>
</tbody>
</table>

As an aside, the output in Table 4 above also calls attention to the potential for positive correlation between rule of law and democracy, which indeed does exist very strongly (r = 0.83, t(95) = 14.70, p<2.2e-16). While these two indices appear to be somewhat redundant, a deeper look into the data confirms the relevance of keeping them both as separate constraints. After each institutional constraint index is adjusted such that it scales from 0-100, we observe that some nations like Iran have a much higher rule of law than
democracy score (44.34 vs. 19.80 respectively) whereas countries like India have a much lower rule of law than democracy score (47.71 vs. 79.20 respectively). Once again, the democracy constraint seems to create a lot of questions within this analysis, thus we dedicate the following section to the exploration of democracy as an institutional constraint and a potential agent for geopolitical predictability.

**The Democracy Question**

The significant negative correlation between institutional constraints (particularly democracy) and prediction accuracy was a surprising finding which deserves more in-depth analysis. Rather than trying to deduce conclusions from statistical analysis, the method of inductive reasoning may provide a better window into helping us understand this somewhat counter-intuitive result. As the initial goal of this study was to identify the content qualities which made certain questions more predictable than others, we return to conduct an inductive analysis on the questions themselves.

We being looking at the five highest accuracy questions in the data set:

1. Before 1 May 2014, will Iran abolish the office of President of the Islamic Republic?
2. Before 1 May 2014, will General Abdel Fattah al-Sisi announce that he plans to stand as a candidate in Egypt's next presidential election?
3. Will Syria's *mustard agent and key binary chemical weapon components be destroyed on or before the 31 March 2014 deadline established by the Executive Council of the Organization for the Prohibition of Chemical Weapons (OPCW)?
4. Before 1 May 2014, will China *attempt to seize control of Zhongye Island?
5. Will the six-party talks with North Korea resume before 1 May 2014?
The prompts come from a wide variety of question families, but we observe that all five of them are housed in the geographic regions of the Middle East or Southeast / East Asia. Most of these countries (North Korea, Syria, Iran, etc.) are actually highly unstable which seems to follow from our preliminary finding that institutional constraints and prediction accuracy are negatively correlated.

Next, we select on the dependent variable for all questions which have a reverse Brier Score (accuracy measure) of >0.998 to obtain a slightly larger sample of the 17 most accurately predicted questions in our data set. This subset is likely too small and too biased to interpret any results as statistically significant; nevertheless it could be a useful tool for dissecting the data further. We run the initial thread of analysis exploring the relationship between prediction accuracy and “total” institutional constraints and find that, indeed, the negative correlation is even stronger between the two within this subset (r =-0.45, t(16)=-2.02, p=.06), further reinforcing our initial result.

We do the same with the five lowest accuracy questions in the data set:

1. Before 1 January 2014, will the Prime Minister of Japan visit the Yasukuni Shrine?
2. Will Viktor Yanukovich vacate the office of President of Ukraine before 10 May 2014?
3. Before 1 May 2014, will Chinese armed forces or maritime law enforcement forces attempt to interdict ... vessel or airplane that it claims is in its territorial waters or airspace?
4. Which of the following will occur first with regard to the state of emergency declared by the government of Thailand on 21 January 2014?
5. Which party will win the largest number of seats in the next elections for Colombia’s Chamber of Representatives?
This time a number of geographic regions are represented, but there is a potential pattern in the question families, with two prompts falling in the domestic category and two in the leader change category.

Next, we select for all questions which have a reverse Brier Score <0.75 to again obtain a sample of the 17 least accurately predicted questions in our data set. With an even stronger result in this subset relative to the previous one, the negative correlation between accuracy and “total” institutional constraints is observed ($r = -0.52$, $t(16) = -2.44$, $p = 0.03$). Moreover, among the subset of least accurately predicted questions, the constraints of rule of law and democracy present a particularly strong negative correlation—($r = -0.71$, $t(12) = -3.47$, $p = 0.005$) and ($r = -0.57$, $t(16) = -2.76$, $p = 0.01$) respectively. Acknowledging that this final step of analysis is both biased from selecting on the dependent variable and limited due to its very small sample size, it is safe to say that this investigation may contribute some additional strength to our original surprising result that the presence of institutional constraints, and democracy, in particular, make it harder for forecasters to accurately predict future geopolitical events.

**Summary of Findings**

This research project was focused on identifying patterns based on the content characteristics of geopolitical forecasting questions that may make them more predictable for geopolitical forecasters (both amateur and professional) in the future. The goal of this study was to generate hypotheses within the realm of forecasting data that could possibly form the basis for future political science theories.
The most noteworthy finding from this study is that the presence of institutional constraints is negatively associated with prediction accuracy. Diplomacy was the only constraint tested which seemed positively correlated with accuracy, suggesting that perhaps political actors in countries which are party to numerous international treaties (as this metric tests) do actually feel a restriction in their ability to pursue “unpredictable outcomes.” All other constraints tested were negatively correlated with prediction accuracy across a number of different statistical analyses run. The particularly strong negative relationship between democracy and predictability offers a hypothesis for further research in the future.

Lastly, prompts located in the region of Southeast and East Asia and the question family of “domestic” were seen as fairly unpredictable relative to other question in this data set, so future forecasters would do well to note this observation and consider exercising greater caution on prediction questions which include these two categorizations.
Possible Interpretation

Can democracy really make things less predictable? This is a question to be answered by a future paper. Nonetheless, our data from this study allow us to make a brief conjecture regarding this initially surprising finding. We begin by reviewing our chosen indicator for this particular institutional constraint—The Democracy Index, created by the Economist Intelligence Unit—which incorporates an assessment of a country’s electoral process and pluralism, civil liberties, functioning of government, political participation, and political culture.\textsuperscript{12} Why would a country that scores higher on these factors be subject to so much unpredictability?

\textit{Table 5: The Economic Intelligence Unit’s Democracy Index 2013}

<table>
<thead>
<tr>
<th>The Democracy Index --- Highest / Lowest States of Democracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norway</td>
</tr>
<tr>
<td>Sweden</td>
</tr>
<tr>
<td>Iceland</td>
</tr>
<tr>
<td>New Zealand</td>
</tr>
<tr>
<td>Denmark</td>
</tr>
<tr>
<td>Switzerland</td>
</tr>
<tr>
<td>Canada</td>
</tr>
<tr>
<td>Finland</td>
</tr>
<tr>
<td>Australia</td>
</tr>
<tr>
<td>The Netherlands</td>
</tr>
</tbody>
</table>

---

\textsuperscript{12} Democracy Index
A potential explanation lies in the nature of the forecasting tournament data itself. Prediction prompts, which are designed to be challenging, are much less likely to include countries like Norway and Canada (which are highly democratic) as they are to include North Korea and Syria (which are highly undemocratic). In fact, the average democracy score across all 167 countries surveyed was 55.5 (after our scaling mechanism from 0-100), but the average unweighted country democracy score assigned to all questions in our data set (n=129) was 44.3. This reference point suggests that the most democratic countries tend to be underrepresented in forecasting tournament data.

Looking back on our inductive analysis of the most and least predictable questions, we see evidence to suggest that questions referencing the least democratic countries (North Korea, Syria, Iran, etc.) are surprisingly well-predicted by forecasters. Questions that are most poorly predicted, on the other hand, seem to fall in the middle ranges of the Democracy Index—again, according to our previous inductive analysis, these would be nations like Colombia, Thailand, and the Ukraine. All of these data points begin to build a hypothesis that perhaps it is not the most democratic countries that are the least predictable, but rather the most democratic countries in our forecasting data set which actually represent more a middle-democracy group of countries than one that is truly high-democracy countries.
This hypothesis is reminiscent of an economics model known as the Kuznets curve. In his landmark 1955 paper, *Economic Growth and Income Inequality*, Nobel laureate Simon Kuznets introduced a theory that while poor societies start off relatively equal, as their economies grow, they are subject to increasing inequality as employment models shift to adjust to changing external circumstances such as industrialization. Eventually, however, as the economy develops further, this inequality is reduced again, as the market adjusts to its new normal standard of activity.  

![Image 1: An illustration of the Kuznet’s Curve](image)

The resemblance between this phenomenon and that which we experience in our own data set—namely, that both the most and least democratic states appear to be the most predictable—implies that there may perhaps be a similar quadratic equation that could help explain the relationship between relative levels of state democracy and the predictability of geopolitical events. We will test this postulate by fitting the Brier score (not reversed) prediction accuracy measure of against the level of (unweighted) national democracy as determined by the index. Each individual question will serve as a unique data point.

After plotting all of the data points and fitting a quadratic polynomial to the observations, we detect that there may indeed be a curvilinear relationship between

---

predictability and democracy as follows from the similar economic model planted by Kuznets in 1955. In our limited data set, however, the relationship is weak ($R^2 = 0.0254$, $y = -8E-05x^2 +0.0089x - 0.0568$), and there are likely to be many other factors which play a role. It is worth noting that forecasting data is intended to be non-linear, or even, in this case, non-quadratic, as the goal of these tournaments is to pose distinctly unpredictable questions to the participants. Thus, even a small observed relationship such as this one could warrant a very interesting and explanatory study when real world data is incorporated.

*Image 2: Geopolitical unpredictability as a function of national democracy levels*

In closing, I suggest that more research be done on the evolution of democracy in emerging political powers and how this progression could perhaps be a source of local instability and thus geopolitical unpredictability. The idea of a turbulent democracy in transition has been promoted by many scholars, including Edward Mansfield and Jack Snyder who assert that in such nations “mass politics mixes with authoritarian elite politics in a volatile way” that can lead to wars, aggression, and other negative political
consequences. Such a hypothesis would also fit our data results in this paper suggesting that questions falling within the “domestic” family are among the least predictable along with questions located on Asian nations like China and Thailand—both of which are frequently referenced as countries undergoing a (long) process of democratization. Among the BRICS and MINT countries, which are generally cited as nations undergoing serious economic transition, we observe from the graph below that each of them is situated somewhere along the process of democratization as well—a condition which may perhaps be a contributing factor in their frequent mention in geopolitical news.

Image 2: States in economic transition as a function of national democracy levels

---

Conclusion

Summary of Paper

This paper set out to explore the relationships between institutional constraints and predictability in geopolitical forecasting. Despite the increasing complexity of our world today, researchers have found that institutional rules and norms still function to influence human behavior, and, therefore, the presence of well-functioning institutions may lead to greater stability and certainty in predicting world events.

Using forecasting data from the Good Judgment Project’s recent prediction tournaments, we tested the change in predictability—including accuracy, confidence, and difficulty—against the experimental constraints of diplomatic ties, rule of law, effective democracy, trade dependence, and freedom of the press. Our hypothesis was that each of these institutions, together and in conjunction with one another, would be effectively able to constraint power player political behavior and reduce uncertainty in the geopolitical realm.

We found that all of the constraints, except for diplomatic ties, actually had a negative correlation with prediction accuracy. Democracy was the strongest negative correlation between the level of constraint and prediction accuracy. We proposed that one possible explanation for this result is due to a potential quadratic relationship between democracy and predictability, such that countries who are transitioning from autocracy actually become less predictable than those that are autocratic. Further research in a larger sample set is needed to test this new hypothesis.
Opportunities for Future Research

Beyond furthering the democracy hypothesis presented in the conclusion, there are a number of other research projects that could follow this paper. As was mentioned numerous times throughout this paper, the small and very particular sample of data used in this analysis is sufficient for hypothesis generation but not true hypothesis testing. Future research could be done to expand the validity of this study from the forecasting world into the real world through a choice of parallel data points that form a part of empirical political science research.

In addition, a number of additional variables could be incorporated into this analysis to better understand the role that institutions can play in encouraging or diminishing question predictability. Some suggestions on the individual forecaster level include education background and cognitive/political science training to gain a better understanding as to the knowledge base with which each forecaster is operating. Another interesting study would be to test the time spent researching, collecting new information, and updating each forecast as a measure of how well the forecaster understands the complex power dynamics taking place. These variables, though harder to measure, would add a significant component to the study of predictability based on specific question details.
Bibliography


