Simulating the Evolution of Global Democracy, 2001–2025*

Joachim Carlsen

Centre for the Study of Civil War; joachim@prio.no

Håvard Hegre

Centre for the Study of Civil War; hhegre@prio.no

Paper prepared for presentation at

the 6th Pan-European Conference on International Relations,

Turin, 12–15 September 2007.

September 7, 2007

Abstract

This paper presents the predicted global distribution of regime types for each year up to 2025. The predictions are made by a newly developed simulation tool that simulate on the basis of a multivariate multinomial logit model of changes to countries' level of democracy. The model takes into account the stability of different institutional setups, average income, diffusion through neighborhoods, and the instability of new institutions. In addition to the ability to predict outcomes, the simulation model allows us to interpret the aggregate effects of a set of model parameters that are mutually dependent on each other. Our statistical model replicates previous findings that economic development and neighborhood effects affect the distribution of regime types. The simulation, howvere, indicates that the substantive importance of these variables are somewhat limited. Increasing the global average growth rate with one percentage point over the 25-year simulation period, for example, only increases the final share of democracy with 0.7%.

*The research was funded by the Norwegian Research Council Grant No. 163115/V10. Thanks to Han Dorussen, Halvor Mehlum, Karl Ove Moene, Ole Magnus Theisen, Michael Ward, and participants at the National Political Science Conference 2007 for helpful comments and suggestions.

.

1 Introduction

This paper presents a simulated forecast up to 2025 of patterns of regime change in the world. We classify the regimes of the world into four types – autocracy, democracy, and two intermediate categories – and estimate how the distribution over these four categories change starting from 2000 (the latest year for which we have data). The simulation is based on a statistical model of changes in regime type estimated on a data set including most countries in the world for the 1950–2000 period. The statistical model accounts for a set of factors shown to be affect the probability of transitions from one regime type to another: economic development, economic growth, the instituional setup of neighboring countries, and the regime history of the country (see Epstein et al., 2006; Gates et al., 2006, 2007; Gleditsch & Ward, 2006; Przeworski et al., 2000). We predict that the proportion of consistent democracies in the world increases from 46.8% in 2000 to 52.2% in 2025.

These 'structural' variables are far from sufficient to predict perfectly transitions between regime types. Out-of-sample prediction validation of our model shows that only between 48 and 63% of the country years are correctly predicted. Still, the simulation allows us to infer a series of insights from statistical analysis of regime change:

1. Simulation-based prediction gives us a precise assessment of the substantive impact of variables in our models. Take economic growth, for example. Growth may directly affect the probability of transition into or sustaining democracy. However, growth also affects other variables in the model that again alter the expected trajectory of regime changes. First, growth increases average income, further increasing the odds for democratic institutions. But if economic growth leads to regime change in a country, this affects the probability of regime changes in the neighboring countries, and the regime change itself may lead to short- and long-term volatility in the institutional setup of the country. To assess the cascading effects of economic growth in a country, we need to simulate. To explore this, we simulate the impact on global and regional distribution of regime types of increasing annual growth in Africa south of Sahara by 1%.

2. Simulation allows us to capture the logic of democratic waves. Gates et al. (2007) show that democratic waves are partly an effect of system-wide shocks such as the two World wars or the fall of the Soviet Union, and partly to the transmission 'through space' (i.e., neighborhood effects). To adequately model transmissions in a country-year setup, we need a complex model as the one discussed above. To fully account for the wave pattern

reflected in these estimates, we have to simulate to be able to interpret the parameters jointly.

3. Simulation and prediction allows us to critically evaluate the quality of the statistical models we use to support theoretical claims. Variables may be clearly statistically significant in studies with thousands of more or less independent observations and still have very little impact on the dependent variables. The extent to which variables improve a model's out-of-sample predictive ability may tell us more about the relative importance of different variables. Prediction also discloses without mercy how partial a description of the world a statistical model is. The model reported below has a pseudo-R^2 of 0.819. Still, the out-of-sample prediction only gets between 48 and 63% of the country-years exactly right. It is clear that the model is fully able to account for the patterns of democratization, and the predictions emerging from the model are likely to be too 'flat'. Investigating the cases where prediction fails is a powerful tool to improve the model underlying the simulation.

4. Simulation helps avoiding drawing erroneous inferences because of collinearity problems. When collinearity is present, our parameter estimates may be very sensitive to sample variations. When studying a data set consisting of all countries for a recent period, we only have one sample to estimate from. To counter this, it is helpful to use the variance/covariance matrix to draw multiple possible realizations of parameters that are consistent with data and interpret based on the aggregates emerging from these multiple simulations.

In this paper, we specify and estimate a 'Markov regression' model – a multinomial logit model where the set of four regime types or categories we define is the dependent variable and their values for the preceding year are included as independent variables. This setup allows estimating the probabilities of transition between each of the four categories. We include as core variables income, growth, information on regime types in the neighborhood, and the recent history of regime changes within the country. We make use of out-of-sample validation to choose the optimal specification of the linear term of the model.

We then simulate the long-term behavior of a system described by the estimated transition probabilities. To account for the uncertainty in parameter estimate the simulations are based on a number of independent draws of realized parameters as specified by the estimate vector and the variance-covariance matrix. To account for the variability in system-level outcomes due to the stochastic nature of the process, we simulate the evolution of the system a number of times for each of these draws. For each year, we draw the regime type for each country as specified by the estimated transition probability matrix (which is a function of previous year's regime type and the independent variables). We then update the independent varaibles for each country and recalculate the transition probabilities.

We finally present the central tendency and the confidence bands for a set of aggregate statistics of future distributions of regime types. We break the predictions down on regions and income groups.

2 An Empirical Model of Regime Transitions

This section presents an empirical model of regime transitions to study the phenomenon of waves of democratization. The model is developed on the basis of Gates et al. (2004, 2006), Gleditsch & Ward, 2006, Epstein et al. (2006), Przeworski et al. (2000), and Fjelde & Hegre (2007). The model attempts to capture several mechanisms of "transmission": First, institutional changes often lead countries to unstable intermediate constellations of political institutions (Gates et al., 2006). Second, the institutional changes themselves are likely to destabilize the country, such that new changes are likely the first few years after a transition. Third, changes may affect the probability of changes in a neighboring country.

Underlying the model is a conception of two stable institutional equilibria: An autocratic equilibrium where all power is concentrated in the hands of a few individuals, and a democratic equilibrium where power is dispersed as widely as possible (Gates et al., 2006). The model also allows studying whether states tend to adopt the same political systems as the neighborhood.

2.1 Measuring Democracy and Democratization

Gates et al. (2006) develop a three-dimensional conception of democracy, placing each political system on axes representing the extent to which the chief executive is elected, the extent to which the executive is constrained by a parliaments or other insitutions, and the extent to which the system allows competitive participation. Data for the executive and constraints dimensions were taken from Polity IVd (Jaggers & Gurr, 1995; McLaughlin et al., 1998), whereas data on participation was taken from Vanhanen (2000). To model what determines the direction of change, however, Gates et al. (2004) condense democracy to one dimension.

To empirically identify the mechanisms, determining regime change, we develop a model

Institutional Type	**Ab-brev.**	**SIP range**	**No. of countries** 1950	1975	2000	**Age in** 2000	**Examples**
Consistent Autocracy	CA	0.00–0.15	29	76	27	27.3	Cuba 1961 – 2000 Eq. Guinea 1969 – 2000 Libya 1969 – 2000 Pakistan 1999 – 2000
Inconsistent Autocracy	IA	0.15–0.35	12	15	22	9.3	Haiti 2000 – 2000 Chad 1996 – 2000 Egypt 1976 – 2000 Bhutan 1953 – 2000
Inconsistent democracy	ID	0.35–0.80	14	10	39	6.7	Mexico 1988 – 2000 Nigeria 1999 – 2000 Russia 1990 – 2000 Malaysia 1969 – 2000
Consistent Democracy	CD	0.80–1.00	27	36	68	33.7	Argentina 1983 – 2000 Benin 1991 – 2000 Switzerland 1872 – 2000 Bangladesh 1991 – 2000

Table 1: The four institutional types used in analysis

that relates a country's probability of having a particular regime type in one year to the regime type it had the year before, and to other variables that influence regime stability and the nature of transition. We distinguish between four different regime types: Autocracy, inconsistent autocracy, inconsistent democracy, and consistent democracy. The categorization is based on the 'Scalar Index of Polities' (SIP) developed by Gates et al. (2004; 2006). This index combines information from Polity IV on the method for selecting executives, the extent to which the executive is constrained by other institutions in the political system (Jaggers & Gurr, 1995), and information on the extent and competitiveness of participation from Vanhanen's Polyarchy dataset (Vanhanen, 2000). Table 1 summarizes the four categories and provides some selected actual examples as of 2000, currently the last year in the dataset used here. The columns labeled 'No. of countries' report the number of countries coded for each type in the years 1950, 1975, and 2000. The column labeled 'Age' reports for each institutional type the average number of years countries had remained unchanged as of the year 2000.

2.2 The matrix of transition probabilities

We are interested in modeling under various conditions the average stability of these four institutional types, and what they tend to change into when they change. The analysis presented below builds on the concept of a transition matrix. The observed transition matrix for the 1950 – 2000 period is presented in Table 2. The rows in the matrix represent the

	Institutional Type at t				
	CA: Consistent Autocracy	IA: Inconsistent Autocracy	ID: Inconsistent Democracy	CD: Consistent Democracy	All Countries
Inst. Type at $t-1$					
Consistent Autocracy	2,426 96.35%	50 1.99%	28 1.11%	14 0.56%	2,518 100.00%
Inconsistent Autocracy	48 6.37%	675 89.64%	26 3.45%	4 0.53%	753 100.00%
Inconsistent democracy	26 3.13%	25 3.01%	743 89.52%	36 4.34%	830 100.00%
Consistent Democracy	16 0.75%	2 0.09%	23 1.08%	2,079 98.07%	2,120 100.00%
All Countries	2,156 40.44%	752 12.09%	820 13.18%	2,133 34.29%	6,221 100.00%

Table 2: Transition Matrix, four institutional types, 1950–2000

institutional type at $t-1$, the year before the year of observation. The columns represent the institutional types at the year of observation t. The right-most column and the bottom row sum over all institutional types.

The first row shows what happened to countries that were consistent autocracies (CA) at $t-1$. There were a total of 2,518 country-years of this type. Of these, 2,426 or 96.35% remained CA. Fifty (1.99%) changed into inconsistent autocracy (IA), 28 (1.11%) into inconsistent democracy (ID), and 14 (0.56%) into consistent democracy (CD). The second row shows the same transition probabilities for countries that were IA at $t-1$.

Political systems are relatively stable, as shown in Table 1. The transition frequencies and probabilities along the diagonal of Table 2 show the probability of remaining in the same state for the four institutional types. Consistent democracies are most stable – 98.07% of the CD country-years were followed by another CD country-year in the same country. Consistent autocracies are also fairly stable, with 96.35% of the country-years remaining in the same status. The two inconsistent states are considerably less stable, with transition probabilities around 89.5%.

Another interesting thing to note is what the various regimes change into when they change. The majority of consistent autocracies transit into inconsistent autocracies when they change, and only one out of seven change into consistent democracy. Among the inconsistent autocracies, the majority of those that transit change into CA, and very few into CD. Among the inconsistent democracies, more than 40% change into consistent democracy. The few consistent democracies that transit predominantly change into inconsistent democracies. These transition patterns indicate that a large group of countries oscillate between consistent and inconsistent autocracy, and another body between consistent and inconsistent democracy. A smaller body makes the transitions from the autocratic pair to the democratic pair, or vice versa. It is possible to calculate the frequency of various trajectories countries follow in their democratization process and the expected long-run distribution of types. But since the stability of the types and the transition patterns depend on other variables, we will turn to a set of covariates before addressing this issue.

2.3 'Markov Regression': The multinomial logit model

One may estimate the transition matrix using a multinomial logit model with the institutional type at t as the outcome variable, and the type at $t-1$ as a set of dummy variables. The multinomial model (see Greene, 1997: 914–917; StataCorp 2005: 210–211) for the four outcomes ($j = 0 :$ CA, $j = 1 :$ IA, $j = 2 :$ ID, $j = 3 :$ CD) is then

$$p\left(Y_i = j\right) = \frac{e^{\boldsymbol{\beta}'_j \mathbf{x}_i}}{\sum_{k=1}^{4} e^{\boldsymbol{\beta}'_k \mathbf{x}_i}}$$

To identify the model, we set CA as the base outcome. The probabilities of the four outcomes are given by:

$$\begin{aligned}
p\left(Y_i = 0\right) &= \frac{1}{1 + e^{\boldsymbol{\beta}'_1 \mathbf{x}_i} + e^{\boldsymbol{\beta}'_2 \mathbf{x}_i} + e^{\boldsymbol{\beta}'_3 \mathbf{x}_i}} \\
p\left(Y_i = 1\right) &= \frac{e^{\boldsymbol{\beta}'_1 \mathbf{x}_i}}{1 + e^{\boldsymbol{\beta}'_1 \mathbf{x}_i} + e^{\boldsymbol{\beta}'_2 \mathbf{x}_i} + e^{\boldsymbol{\beta}'_3 \mathbf{x}_i}} \\
p\left(Y_i = 2\right) &= \frac{e^{\boldsymbol{\beta}'_2 \mathbf{x}_i}}{1 + e^{\boldsymbol{\beta}'_1 \mathbf{x}_i} + e^{\boldsymbol{\beta}'_2 \mathbf{x}_i} + e^{\boldsymbol{\beta}'_3 \mathbf{x}_i}} \\
p\left(Y_i = 3\right) &= \frac{e^{\boldsymbol{\beta}'_3 \mathbf{x}_i}}{1 + e^{\boldsymbol{\beta}'_1 \mathbf{x}_i} + e^{\boldsymbol{\beta}'_2 \mathbf{x}_i} + e^{\boldsymbol{\beta}'_3 \mathbf{x}_i}} \qquad (1)
\end{aligned}$$

The β estimates also has a direct interpretation in terms of relative probabilities:

$$\frac{p\left(Y = 1\right)}{p\left(Y = 0\right)} = e^{\boldsymbol{\beta}'_1 \mathbf{x}_i}$$

and:

$$\frac{p\left(Y = 3\right)}{p\left(Y = 0\right)} = e^{\boldsymbol{\beta}'_3 \mathbf{x}_i}$$

The estimates β_1 reported below, then, are interpreted as the impact of the explanatory variable on the probability of being inconsistent autocracy relative to consistent autocracy. The β_3 estimates approximate the probability of consistent democracy relative to consistent autocracy.

If we enter only the state at $t-1$ as explanatory variable(s), the predicted probabilities from estimating this model are identical to those reported in Table 2. The purpose of formulating this as a multinomial logit model, however, is to be able to account for a set of explanatory variables, listed in the next section.

2.4 Control Variables in the Transition Model

Lagged institutional type The institutional type at $t-1$ was entered in the form of three dichotomous indicator variables: IA_{t-1}, ID_{t-1}, and CD_{t-1}. These variables have the value 1 if the country was of this type the year before, and 0 otherwise. CA is the reference category. We will sometimes refer to the set of indicator variables by means of the name 'lagged regime type'.to

Time since previous regime change It is conceivable that political systems become more stable if they succeed in getting past the first few years. In particular, several scholars note that the first real election is an important step-stone for democracies. We coded a variable which has the value 1 if there was a change between the four categories from $t-2$ to $t-1$, and decreases at a constant rate as long as there are no new changes. The formula for the decay function is $2^{-time\ \sin ce\ change/12}$, and we call the variable 'proximity to regime change'. The denominator 12 implies that the effect of a change on the probability of future change is halved every 12 years. In the decay function formulation, we call this variable 'proximity of regime change'.

Time since independence Some political systems may also be more unstable just after a country's independence. We coded a variable which has the value 1 if the country became independent at $t-2$, and decreases at a constant rate afterwards. The formula for the decay function is $2^{-time\ \sin ce\ independence/8}$. In the decay function formulation, we call this variable 'proximity of independence'.

Dominant regime type in neighborhood The model includes a set of four dummy variables that denotes whether the four types CA, IA, ID, or CD are the dominant type in the neighborhood, coded along the lines of Gleditsch & Ward (2006). A regime type is considered dominant if a majority of the countries have regimes of this type.[1] All four dummy variables are included since they may all be 0 simultaneously.

Distribution of regime types in the world The model also takes into account the distribution of regime types in the world. This set of variables is coded as the share of the world's countries that are of type IA, the share of ID, and the share of CD. The share of type CA is the reference category.

Lagged GDP per capita This variable is the base-2 logarithm of constant-dollar GDP per capita.[2] The variable is lagged by one year.[3]

Lagged growth in GDP per capita This variable is the difference between log GDP per capita at $t-1$ and log GDP per capita at $t-2$.

2.5 Model selection

Which variables to include in a statistical model is routinely determined by the significance of the corresponding parameters. If a variable is significant, it is included in the model. If not, it is often removed. Since the aim of our model is to predict, we rather use out-of-sample predictive ability as the selection criterion. We split the country-years for which we have data into an estimation sample and a prediction sample. We estimated a set of candidate model on the estimation sample, and ran 100 simulations to predict the regime type for the prediction sample – ten independent simulations of the system for each of ten independent realization of parameters (more information on this below). We used two different methods to split the sample. In the first – the diacronic sample – we selected half of the countries by means of a random draw. All country-years for this sample constituted the estimation sample,

[1] We also tried to weight neighboring countries by population size, and to use continues variables representing the share of neighboring countries associated with each type. The reported formulation performed best in terms of effect on log likelihood.

[2] GDP per capita data were drawn from two sources. We use World Bank data for the period 1960 to 2000 (World Bank, 2000a; 2000b) and Penn World Tables, v5.6 (Summers and Heston, 1991) for 1950 to 1959. The two datasets refer to different baseline years for calculating constant dollar figures, and are based on different methods of measurement. To counter these differences, we calculate the average ratio in the three first overlapping years per country for the overlap, and use this ratio to adjust the numbers. To reduce endogeneity bias, we lag the variable. We use the average ln(GDP per capita) for the five years preceding the end-date of each time segment. That is, for the 1990–1994 segment of a right-censored polity observation, we use the average for the years from 1989 to 1993. For a non-censored observation (e.g., an observation ending with a polity change in July 1992) we use the average for the years 1987–1991.

[3] One problem with GDP per capita as an indicator of socio-economic development is that it fails to distinguish between oil-rich countries and other high- and upper middle-income countries. This is problematic, since oil-rich countries are systematically less prone to democratize than countries with comparable average income derived from manufacturing and services (Ross, 2001).

and all country-years for the remaining countries the prediction sample. In the second – the syncronic sample – all country-years up to 1985 constituted the estimation sample, and all country-years from 1986 onwards the prediction sample.

Model 2 – the core model – included the following variables: three indicator variables for regime type at $t-1$, Proximity of previous regime change, Proximity of independence, Three indicator variables for dominant regime in the neighborhood, three indicator variables for the distribution of regime types in the world, and GDP per capita.[4]

In model 1, we constrained six parameters in the core model that were not significant. In model 3, we added growth without constraining any parameters. In model 4, we added interaction terms to capture that the propensity for change from a regime type is dependent on GDP per capita (cf. Przeworski et al., 2000). The interaction term is the product of the indicator variables for regime type at $t-1$ and log GDP per capita at $t-1$. In model 5, we added growth to model 4. In model 6, we removed the variables representing the global distribution of regime types from model 4. In model 7, we constrained three insignificant parameters in model 6 to be zero.

Table 3 shows a set of statistics characterizing the prediction ability of the seven models. The two left-most number columns reflect the performance for the diachronic sample, the two right-most columns that of the synchronic sample. For each sample, we calculated the number of correct out-of-sample predictions of regime type as a percentage of total number of predictions. These figures are in columns labeled '% correct predictions'. The higher the percentage, the better is the model. We also calculated the average absolute value of the distance between predicted and observed regime type, treating the regime type variable as ordinal for the moment, setting $CA = 0$ and $CD = 3$. In other words, in cases where we predict CA but observe CD, the distance is three. The lower the average distance, the better is the model.

Unfortunately, different rankings of models emerge from the four indicators of predictive performance. For this version of the paper, we have chosen the model that perform best in terms of percentage correct predictions in the diachronic out-sample.[5] Model 6 performs best according to this criterion, with 48.47% correct predictions.

[4] The core model is numbered #2 for technical reasons.

[5] This choice is quite arbitrary. In the next version of the paper, we will extend the number of models to compare as well as update and correct the data set. We are aware of errors in the growth variable which may explain why this variable did not improve the predictive performance of the model.

		Diachronic 1950 − 2000		Synchronic 1985 − 2000	
Model	[Core model] plus	% corr. pred.	Dist. from observed	% corr. pred.	Dist. from observed
2	Core model	41.78	1.08	62.59	0.64
1	Constraints	46.48	1.01	63.77	0.60
3	Growth	46.89	0.89	62.53	0.65
4	GDP/Regime type int.	42.18	0.97	60.77	0.65
5	GDP/Regime type int. and growth	41.76	1.04	60.49	0.66
6	GDP/Regime type int., w/o world share	48.47	0.99	60.67	0.67
7	GDP/Regime type int. and constraints	42.26	1.05	61.81	0.64

Table 3: Out-of-sample predictive performance of different specifications of the linear component of the model

2.6 Estimation results

Table 4 presents the results of estimating the model. The three columns in the Table refer to the three multinomial logit equations (expression 1). CA is the reference outcome or equation. Within each of the columns, the three upper lines present the estimates for the dummy variables representing the state at $t-1$. CA is also the reference category in these variable sets.

The following patterns may be discerned from the estimates: As evident in the matrix of transition probabilities (Table 2), the probability of observing a regime type is considerably higher if the country had the same regime type the previous year. This is reflected in the positive estimates for IA_{t-1} in the IA equation and for CD_{t-1} in the CD equation, and the negative estimate for IA_{t-1} in the ID equation. This pattern of stability is reinforced when GDP per capita is higher – the estimates for the $GDP * IA_{t-1}$ and $GDP * ID_{t-1}$ terms are positive and significant in the IA and ID equations, respectively. We also see clear neighborhood effects. In the ID equation, the 'CA dominant' variable is negative, implying that the relative probability of observing ID is lower when there are more inconsistent autocracies in the neighborhood. Inconsistent democracy, on the other hand, is more likely when the neighborhood is dominated by either other inconsistent democracies or inconsistent autocracies.

Finally, the history variables have a clear impact on the stability of regime types. The 'Proximity of regime change' variable is positive and significant – the shorter the time is since a regime change, the higher is the probability of observing IA, ID, or CD relative to observing CA. These estimates reflect that IA, ID, and CD are relatively more frequently observed immediately after regime changes – over the 1950–2000 period, changes were more often in the direction of consistent and inconsistent democracy than in the direction of inconsistent autocracy ($3.77 > 2.29$) and even less frequent in the direction of consistent autocracy ($2.29 > 0$). The 'proximity of independence' estimates have the opposite signs, reflecting the same tendency. The closer the observation is to the independence year for the country, the less frequent do we observe consistent or inconsistent democracy.

The proximity of change variable is positive and signicantly larger than 0 in all three equations – if there has been a recent change, the largest fraction of countries have ID and the lowest fraction has CA. The proximity of independence variable is negative and significant in all three equations –

consolidation variable, for instance, was constrained to 0 in the ID and CD equations. In the IA equation, the consolidation variable is significant and negative: After the five years, the probability of remaining in the same state (i.e., IA) is reduced by 46%.[6]

The effect of the neighborhood on the relative risk varies considerably over the four states. The 'share of same type in neighborhood' is not significant in the IA and ID equations and was excluded. It is positive, large and clearly significant in the CD equation. The CD/CA relative risk is $\exp(1.44) = 4.2$ times higher when all neighbors are CD as compared to when none are CD. If a country has four neighbors and one of them changes into the same type as the country, the CD/CA relative risk increases by $\exp(1.44/4) = 1.4$.

The 'difference from neighborhood' variable is positive in all three equations, but clearly strongest for the two democratic outcomes. If the difference to the neighborhood is 1 (e.g. if the country is IA and all neighbors are ID), the CD/CA relative risk is $\exp(.47) = 1.6$ times higher than if the difference is 0.

Finally, the GDP per capita variable is clearly significant only for the CD outcome. The CD/CA relative risk increases by a factor of $\exp(0.36) = 1.4$ when GDP per capita is doubled.

2.7 Aggregate effects?

The aggregate effects of these estimates are difficult to read out of the estimates alone. Using the mathematical theory for Markov chains, we may calculate the long-run behavior of the transition matrix for a given, fixed set of values for the explanatory variables. But this requires us to assume that the transition probabilities are constant. This is impossible when neighborhood variables play an important role – a transition in one country affects the transition probabilities in the neighboring countries, and may lead to domino effects beyond the

[6] The probability of remaining in the same state is $1 - \exp(-.61) = 1 - 0.54 = 0.46$.

Variable	*Equation*		
	IA: Inconsistent Autocracy	*ID: Inconsistent Democracy*	*CD: Consistent Democracy*
	$\widehat{\beta}$ (*s.e.*)	$\widehat{\beta}$ (*s.e.*)	$\widehat{\beta}$ (*s.e.*)
CA: Consistent Autocracy at $t-1$	— *Reference category* —		
IA: Inconsistent	3.77**	−4.63**	1.39
Autocracy at $t-1$	(1.42)	(2.10)	(3.16)
ID: Inconsistent	0.226	−0.12	0.028
Democracy at $t-1$	(2.61)	(1.98)	(2.48)
CD: Consistent	3.04	1.76	6.21***
Democracy at $t-1$	(3.88)	(2.03)	(1.88)
Proximity of	2.29***	3.77***	3.78***
regime change$_{t-2}$	(0.37)	(0.38)	(0.46)
Proximity of	−0.95	−2.01**	−2.88**
independence	(0.79)	(0.82)	(1.03)
CA dominant	−0.33	−1.12***	−0.83**
in neighborhood$_{t-1}$	(0.23)	(0.25)	(0.30)
IA dominant	−1.29	−0.53	−1.69*
in neighborhood$_{t-1}$	(0.72)	(0.68)	(0.85)
CD dominant	−0.13	0.58	0.76*
in neighborhood$_{t-1}$	(0.43)	(0.35)	(0.38)
$\log_2$ (*GDP* per capita)	0.24*	0.78***	0.11
*IA$_{t-1}$	(0.14)	(0.21)	(0.30)
$\log_2$ (*GDP* per capita)	0.30	0.74***	0.47
*ID$_{t-1}$	(0.27)	(0.21)	(0.25)
$\log_2$ (*GDP* per capita)	−0.12	0.29	0.29
*CD$_{t-1}$	(0.40)	(0.20)	(0.18)
$\log_2$ (*GDP* per capita)	0.027 (0.095)	−0.26* (0.11)	0.19 (0.13)
Constant	−4.16*** (0.98)	−1.92 (1.08)	−6.93*** (1.39)
LL model	−1107.64		
LL null model	−6107.01		
No. of observations	4866		
Pseudo R^2	0.819		
***: $p < 0.001$; **: $p < 0.01$;*: $p < 0.05$			

Table 4: Multinomial logit regression estimates for probability distribution for institutional types, final model, 1950–2000

neighborhood.

In the simulation presented below, the initial growth rate of all countries is the average of their growth rates for the 1994–2000 period, and that all countries thereafter slowly converge to the average global growth rate. But what would happen if the world grew at a quicker pace? To explore the importance of the income variable, we ran an alternative set of simulations where we assume that countries converge to a growth rate that is 1% higher. The average growth rate in this high-growth scenario, then, is about 3%.

Clearly, these questions require that we both investigate how all the coefficients jointly affect the transition probabilities, and take into account the long-run impact of these changes in transition probabilities. To do this, it is necessary to run simulations. Below, we present a simulation routine that allows answering the questions above. The simulation routine starts at the year of intervention, draws a realization of the model parameters, calculates the transition probabilities, draws an outcome based on these probabilities, updates the explanatory variables, recalculates the transition probabilities for the next year, and repeats for several years into the future. To reduce the impact of individual draws, this procedure is repeated a large number of times.

3 Simulation setup

3.1 Realization of parameters and independent simulations

There are two sources of uncertainty for the simulated results. The first is due to the uncertainty regarding the magnitude of the parameter estimates, as reflected in the standard errors and the variance-covariance matrix. The second is due to the probabilistic nature of our dependent variable. To even out the effects of these two uncertainties, we ran $D * S$ simulations of a world where the probabilities of regime change are governed by the results presented in Table 4.

The first step in the simulation is to draw a realization of the parameter estimates summarized in Table 4. We drew 30 independent draws from the vector of parameter estimates and the associated variance-covariance matrix. Although we have observations for most countries of the world for the 1950–2000 period, it is convenient to regard the estimates as pertaining to one random draw from the universe of possible regime histories. The estimate vector in combination with the variance-covariance matrix gives us a model of the distribution of estimates if we obtained repeated independent samples from the universe. The estimate for the CA_{t-1}

term in the IA equation, for instance, is 3.77 *on average* according to our results. But not only is the variability estimated by the standard error (1.42 in this case), but the covariance matrix specifies how the estimate covary with other estimates in the model.

We drew $D = 30$ independent draws or 'realization of parameters' from the distribution specified by the set of estimates to account for the uncertainty of our parameter estimates. For each draw, we ran $S = 30$ simulations based. In each of these, we loop through the years starting in 2001, and draw regime types based on the transition probability matrix specified by the realized parameters and the current values for the exogenous and endogenous variables.

3.2 Initial States and Growth Scenarios

The initial state for the simulation is the world as of the last year of our dataset, currently the year 2000. We will refer to this year as t_0. All countries with data for all variables for this year were included in the simulation. For the first year of the simulation, we calculated the values for the variables included in Table 4. Except for GDP per capita, these variables endogenous – they are all functions of regime type at t_0, the neighborhood, and time.

The only exogenous variables are GDP per capita and growth in GDP per capita. For the first year of simulation (2001), we assume that the growth rate g_{it} is equal to the average rate for the years 1994–2000. This growth rate, however, may be atypically high or low. It may be hard to defend that China will continue to grow at 10% per year over the next 25 years, or that the economy of Sierra Leone will continue to shrink at almost the same rate. To avoid that what may be an extraordinarily high or low growth rate disproportionally affect the simulation results, we choose a middle ground: We set the growth rate g_{it} for country i at time as

$$g_{it} = \overline{g}_{t-1} + \delta\left(g_{i,t-1} - \overline{g}_{t-1}\right), 0 < \delta < 1 \tag{2}$$

where $\overline{g}_{t-1}$ is the world average growth rate the previous year, δ is a discount parameter, and $g_{i,t-1} - \overline{g}_{t-1}$ is the difference between the country's growth rate and the global average growth rate. The model assumes that all countries steadily converge to a joint growth rate. The δ parameter specifies the speed at which they converge. We set $\delta = 0.95$.[7]

[7] We currently have no procedure for validating this choice. Wt will later estimate it from the data.

3.3 Simulating regime transition

The regime transitions were simulated for a period of 25 years. For each year for each country, we calculated the transition probabilities as a function of the state at t and the other covariates using the estimates in Table 4 and the formula given in Equation (1). The program then draws the realization of these probabilities, using a random number generator. Assume for a moment that the estimated transition probabilities at t for a country are similar to those given in Table 2. If the country is a CA, the program determines that the regime type at $t+1$ is CA in 96.35% of the draws, IA in 1.99%, ID in 1.11%, and CD in 0.56%.

After having drawn the regime type at $t+1$, the program updates the variables depending on this for the focal country and for the other countries in the system: 'Share of same type in neighborhood', 'Regime type dominant in neighborhood', and 'Proximity of regime change'. Income was updated by calculating equation (2) for every country.

We implemented this by programming the simulation procedure in a combination of Stata and C++, a language that is object-oriented, 'simulation friendly' and very efficient.

3.4 Interventions

We ran two sets of simulations. In the first, we assume that all countries grow at rates that converge on the global average for the 1994–2000 period, which was $\overline{g} = 0.0207$. In the second, we assume that the global growth rate is one percentage point higher, or $\overline{g}' = 0.0307$

4 Simulation Results

Figure 1 shows the simulated global distribution of regime types for the first simulation for the first realization of parameters.[8] The x-axis represents time. The simulation was run for 25 years. The black shaded area corresponds to the average proportion of consistent autocracies (CA) for each year. The lightest grey shade corresponds to the proportion of consistent democracies (CD), and the other two shades to IA and ID. The distribution in 2000 is similar to the observed distribution for the 143 countries with data. In that year, 12.6% of the countries were CA, another 14.7% IA, 25.9% ID, and the remaining 46.9% CD. The distribution changes as simulated time goes by. Over the 25 years, the proportion of CD increases to about 55%. Each 'jump' in the shaded area marks one or more net transition between categories.

[8] This is a simulation where the growth rate converges to 2%.

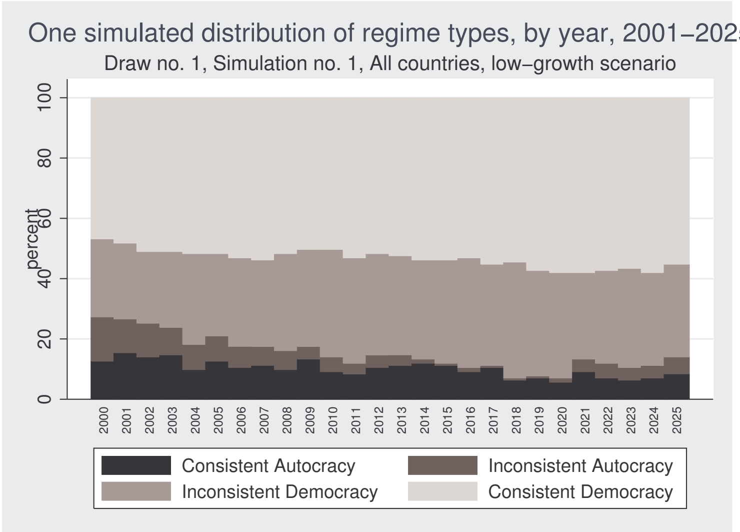

Figure 1: Simulated distribution of regime types, by year, draw no. 1, simulation no. 1

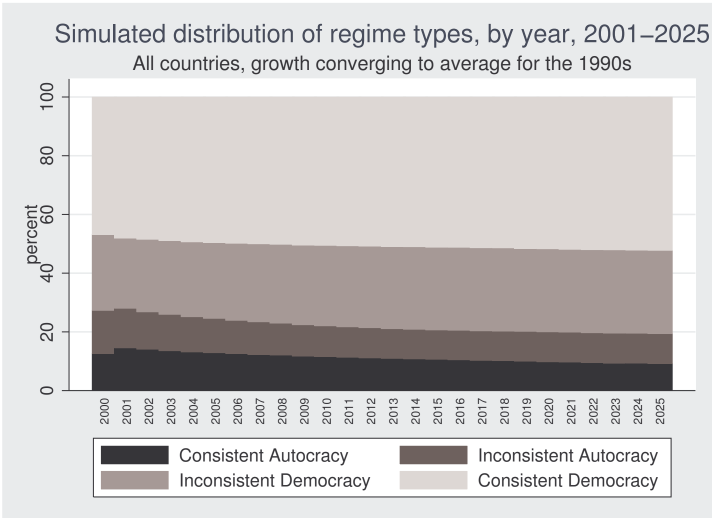

Figure 2:

Individual simulations may differ, however. If neighborhood effects are sufficiently strong, a few changes may tip the scale for several others, and a snowballing of democratization or autocratization may follow. Moreover, individual realizations of parameters are likely to yield somewhat different dynamics.

Figure 2 shows the simulated global distribution averaged over $D = 30$ times $S = 30$ simulations where annual growth rates converge to 2%, the average of the 1990s. Here, the idiosyncracies of individual simulations are evened out, and the predicted distribution changes gradually. The results suggest that the proportion of countries that are CA and IA will diminish by 2025. By then,9.1% of countries are CA and 10.2% IA. The proportion that are ID will increase to 28.4%, whereas that of CD will increase to 52.2% in 2025.

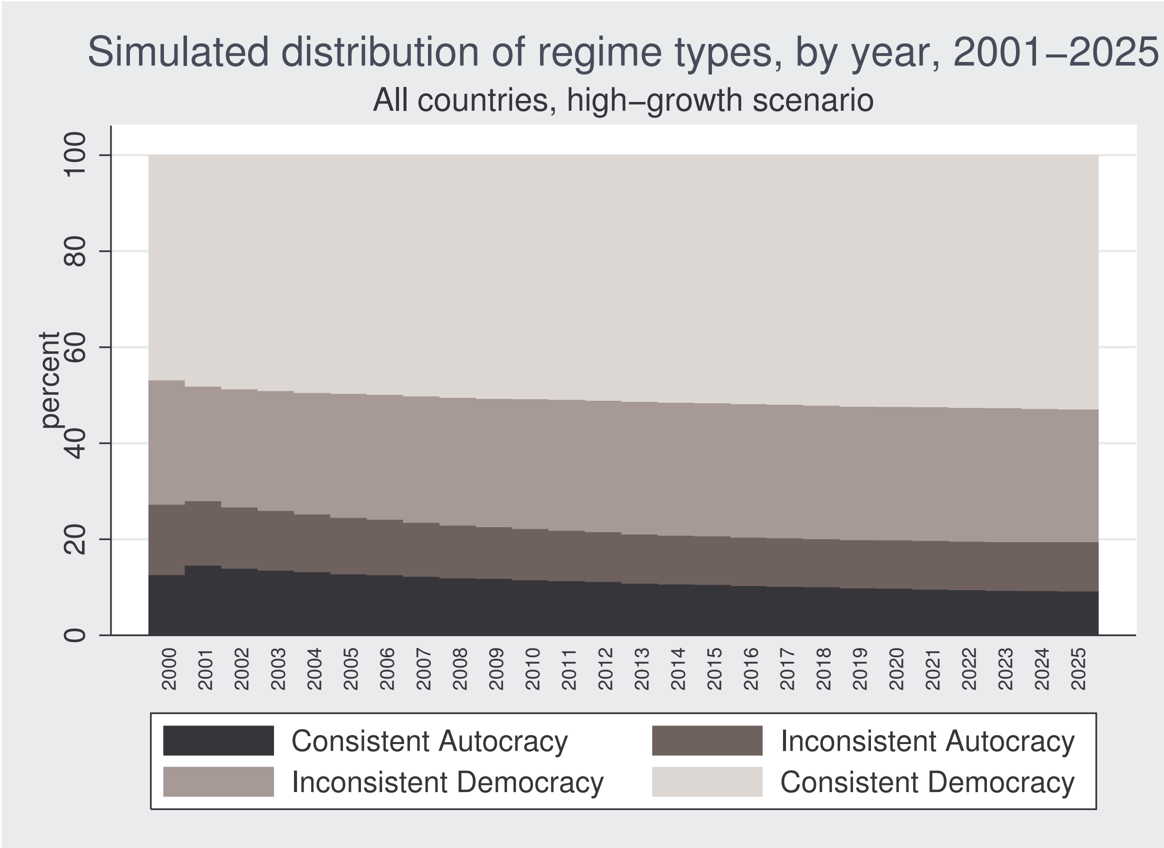

Figure 3: Prediction of global distribution of regime types

The proportion changes most quickly in the first 10–15 years. This can be attributed to two factors. First, the assumption that growth levels gradually converge to the global average plays a role. The impact of high growth in non-democratic countries then diminishes after the convergence point. Second, some of the change is due to a convergence to a steady-state equilibrium. In 2000, several countries have regime types that are outliers relative to the estimates in Table 4. A large number of simulations will predict regime changes in these countries so that their regime setup is more in line with what we would expect from the estimated results. As soon as this has happened, further changes are due to changes in globale income levels only.

What would the future global distribution of regime types look like if the world developed according to the alternative growth scenario, with average growth rates one percentage point higher? This is shown in Figure 3. The differences in the distribution are relatively marginal. In 2025, 52.9% of the countries are predicted to be democratic, 0.7 percentage points more than in the neutral growth scenario.

Figure 4 shows the predicted global shares of consistent democracies (CD) and of consistent

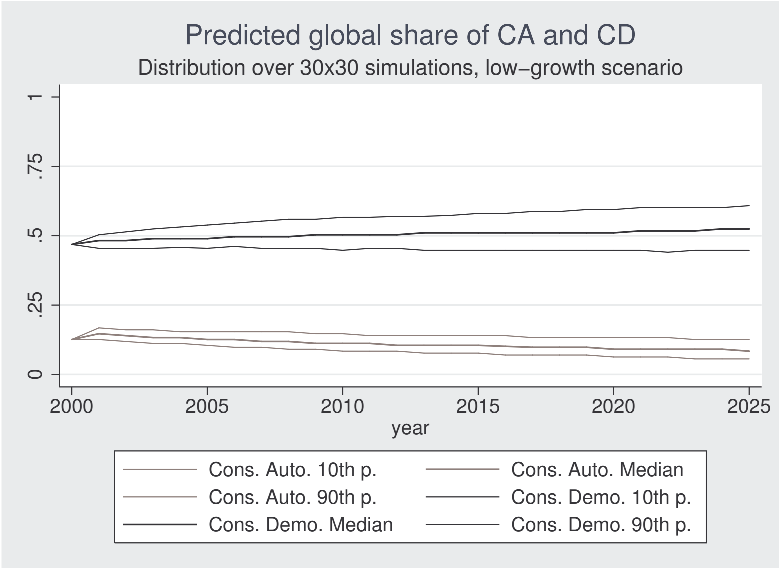

Figure 4: Prediction of share of countries that are consistent democracy or consistent autocracy, with 90% confidence bands

autocracy (CA) with 90% confidence bands. The thicker black and gray lines reflect the median share of CD and CA over all simulations. The thinner lines represent the 10th and 90th percentiles in the predicted shares of the two regime types. In other words, in 80% of the simulations, the predicted share of countries that are CD in 2025 is between 44.8 and 60.8%. The corresponding figures for CA are 5.6 and 12.6%.

4.1 Distribution of regime type by income group

The literature on development and democracy shows that global variation in regime type correlates strongly with income. Our estimates indicate this will be the case also in the future. In this section, we will group countries into income groups, labeled from 'Group 7' to 'Group 15'. Countries with average income between 128 and 256 dollars fall in 'Group 7', since $\log_2$ of the lower bound (128) is 7. 'Group 8' has income between 256 and 512 dollars, etc. USD dollars are given in constant dollars with 2000 as the reference year.

Figure 5 shows the simulated proportion of democracies in the lowest-income group – the group of countries with GDP per capita between 128 and 256 USD. In 2000, the group includes countries such as Burundi, Niger, and Nepal. Note that since we assume that all countries have an average annual growth of about 2%, a number of countries grow out of the income group. Among the remaining countries, the predicted proportion of consistent democracies remains very stable at around 15%. This means there is no discernible effect of other factors in the simulations. If neighborhood effects were sufficiently strong, for instance, we might

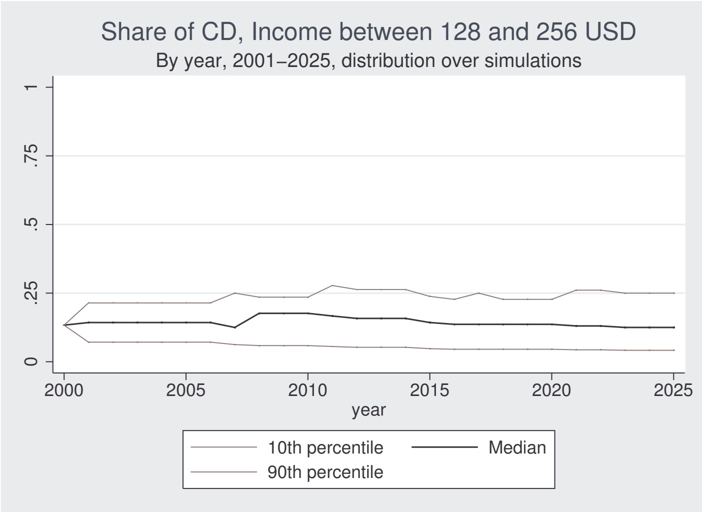

Figure 5:

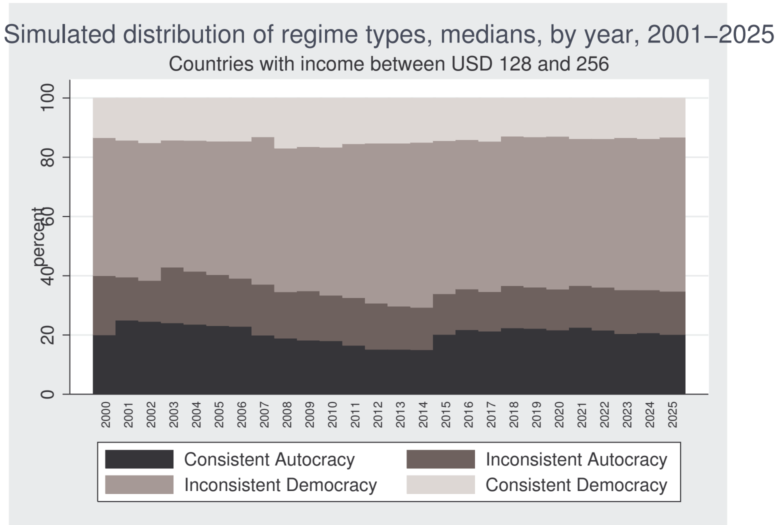

Figure 6:

have seen a spillover from predicted democratizations in neighboring countries.

The thin lines in Figure 5 indicate the 90% confidence band for the simulations. The band ranges from proportions of about 5 to around 25%. Most simulations yield fairly similar results for the share of democracies for the poorest countries in the world.

In Figure 6, the predicted distribution over all four regime types are reported for the low-income countries. The simulations indicate some expansion in the inconsistent democracy group among the low-income countires. This expansion may reflect the increase in the number of inconsistent democracies after the Cold war (e.g. in Sierra Leone and Mozambique).

Figure 7 aggregates information for the group of middle-income countries with average incomes between 1024 and 2048 US dollars (such as Guatemala, Iran, and the Philippines).

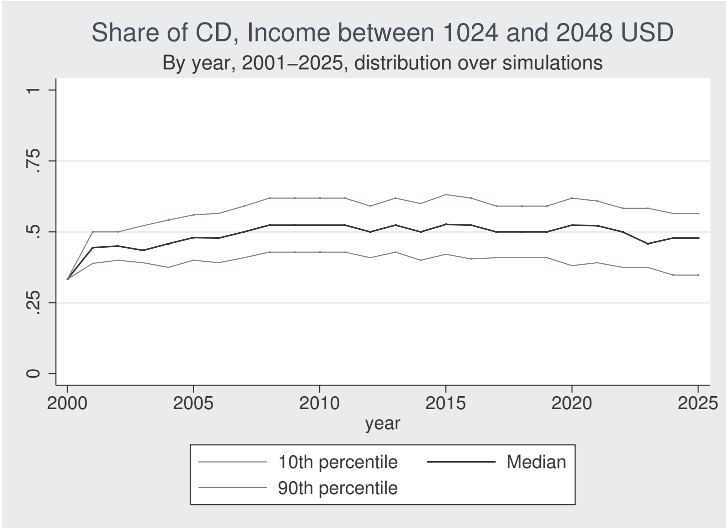

Figure 7:

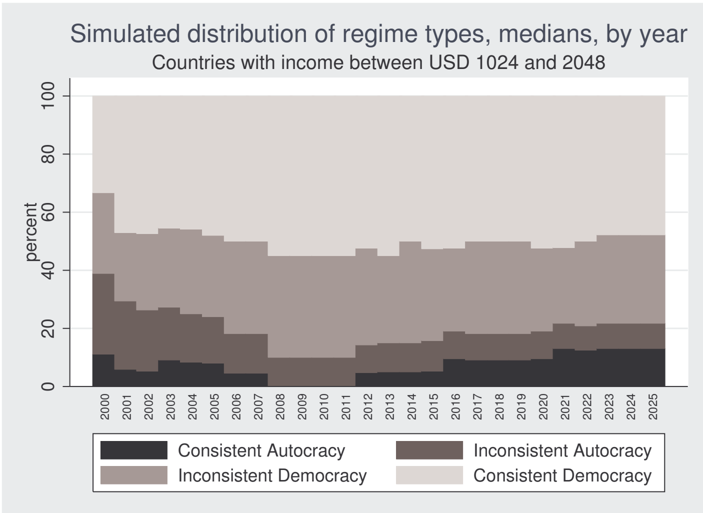

Figure 8:

The share of democracies in this group is predicted to increase considerably from 30 to 50% over the first 7–9 years, and stabilize at this share. Figure 8 shows the distribution over all four types for this income group.

Finally, Figure 9 shows the simulated distribution for countries between 8 and 16 thousand dollars per-capita income (Portugal, Slovenia, Greece, Cyprus, Kuwait, and South Korea constituted this group in 2000). The simulation results are unstable for this group given the low number of countries it comprises. The simulated prediction is a decrease in the share of the number of consistent democracies for this group.[9]

[9] Most of the remaining 10-25% are predicted to be inconsistent democracies.

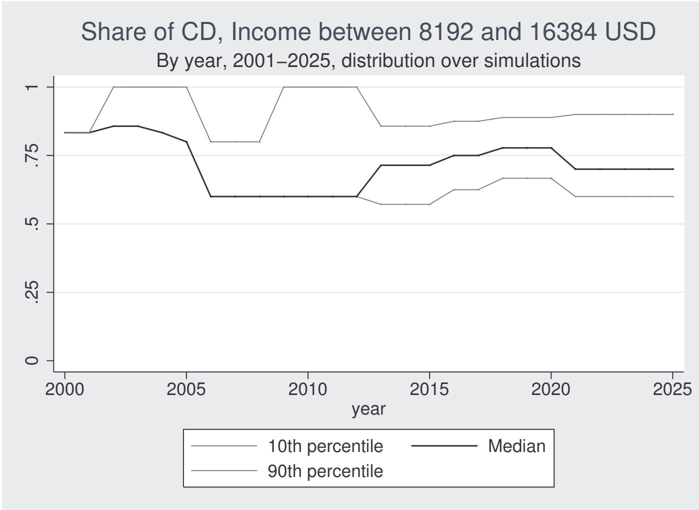

Figure 9:

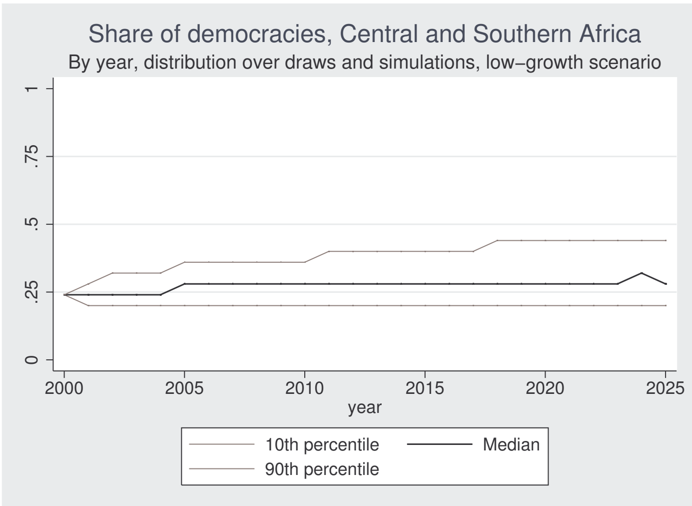

Figure 10:

4.2 Distribution of regime type by region

The following figures show the predictions by region for the share of countries that are consistent democracy by region. Figures 10–**??** show the predicted share of consistent democracy in Central/Southern and Western Africa separately for the two growth scenarios.

In Central and Southern Africa, a quarter of the countries were coded as consistent democracy in 2000 (these were Botswana, Mauritius, Madagascar, Malawi, Namibia, and South Africa). Across all simulations, the median predicted share of CD in this region is stable at just over 25%. The higher growth rate in the scenario underlying the simulations reported in Figure 11 does nothing to change the median share, although the 90th percentile then is somewhat higher (approaching 50%). Changes in average income has a fairly marginal im-

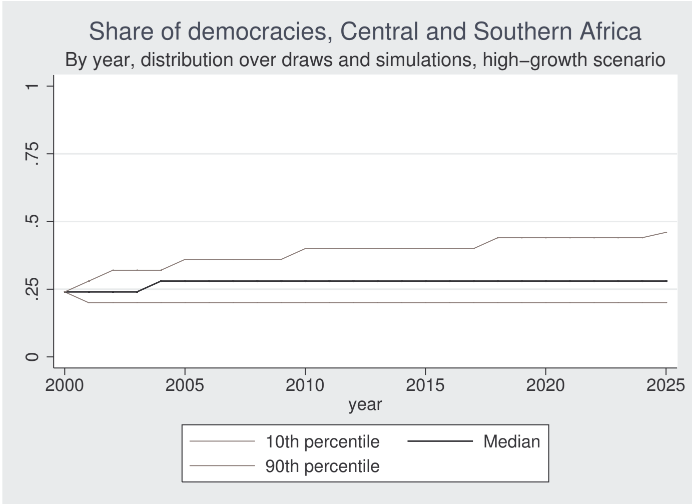

Figure 11:

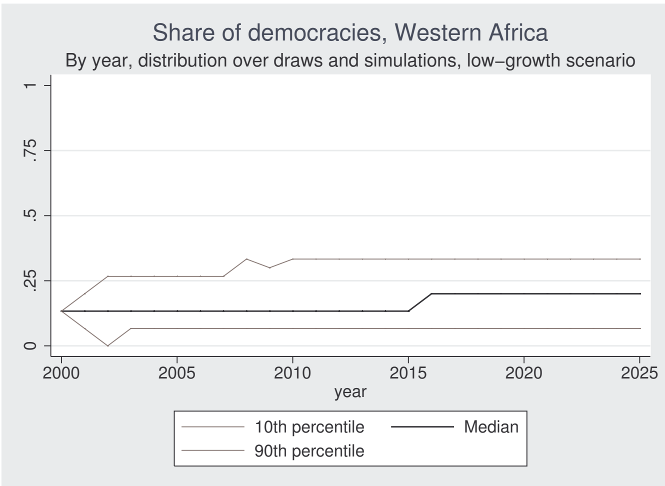

Figure 12:

pact. This probably reflects the huge cross-national income differences in the world and the widespread poverty in all of Africa South of Sahara. It will take considerable more than 25 years of 3% annual growth to change from the level of Mozambique to the level of Iran where democratizations are predicted to start to happen (see Figure 7).

Both the initial and predicted share of consistent democracies in Western Africa is lower than in Central and Southern Africa (the only two democracies in 2000 were Senegal and Benin). As in the other African region, a marginally higher growth rate does little to change the simulated median democratic share of countries. In both sets of simulations, however, the median increases somewhat around 2015.

Figures 14–16 show the predicted share of countries that are consistent democracies in three Asian regions. Most Asian countries are considerably better off than African countries.

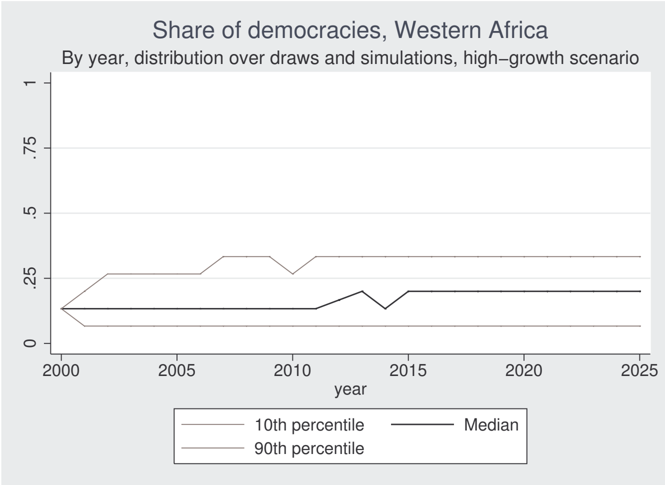

Figure 13:

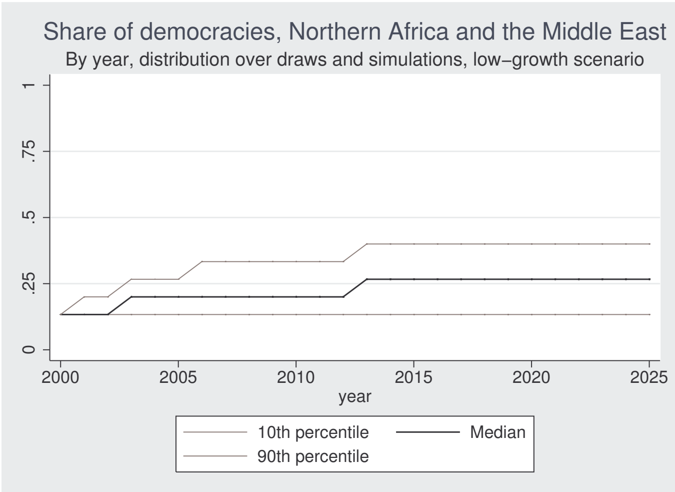

Figure 14:

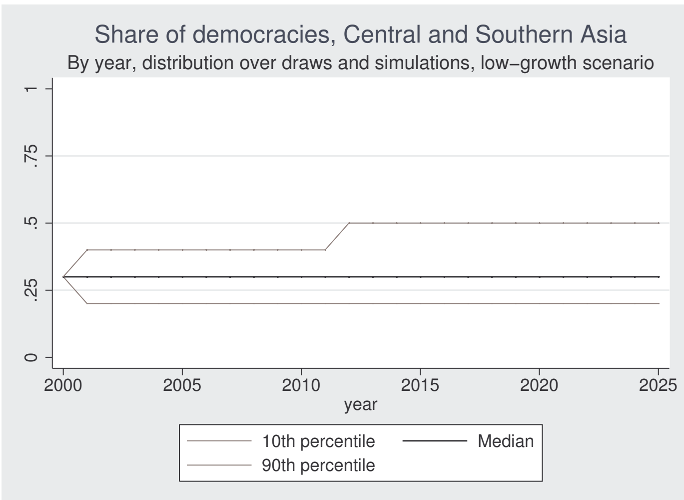

Figure 15:

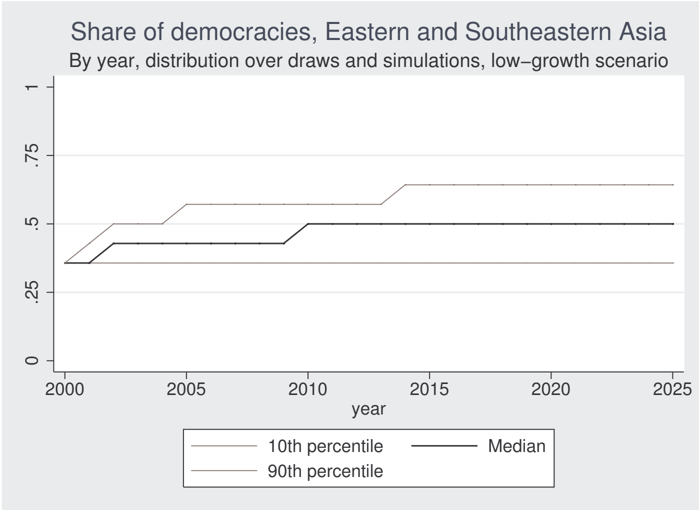

Figure 16:

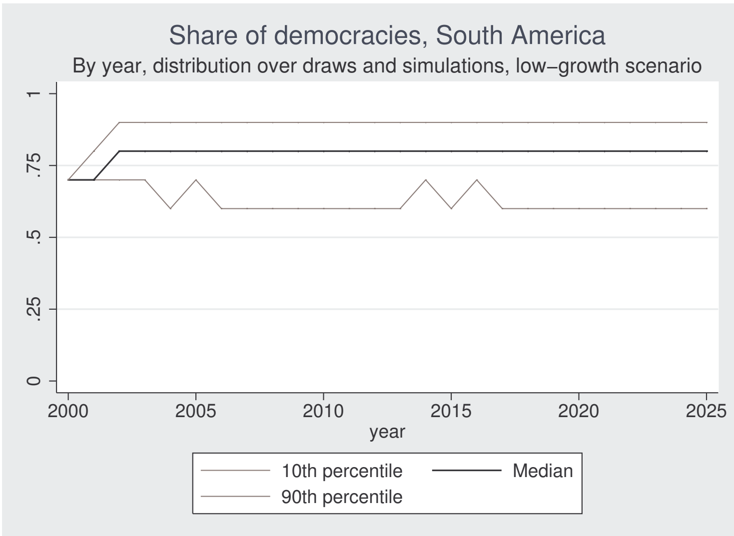

Figure 17:

In the Middle East/North Africa region, the predicted share of CDs increase from about 12% in 2000 to over 25% in the low-growth scenario. In Central and South Asia, the median predicted share is constant at a little above around 25%. In Eastern and South Eastern Asia, the simulations predict a real change over the coming 25 years. The predicted median predicted share increase from under 35% to 50%.

Figures 17 and 18 show the simulation results for two American regions. In both regions, the median prediction is stable just over 75%. In the most optimistic simulations, the share of democracies approach 90% in South America. In the most pessimistic, only 60% remain consistent democracies.

Finally, Figures 19 and 20 show the central tendency of predictions for Eastern and West-

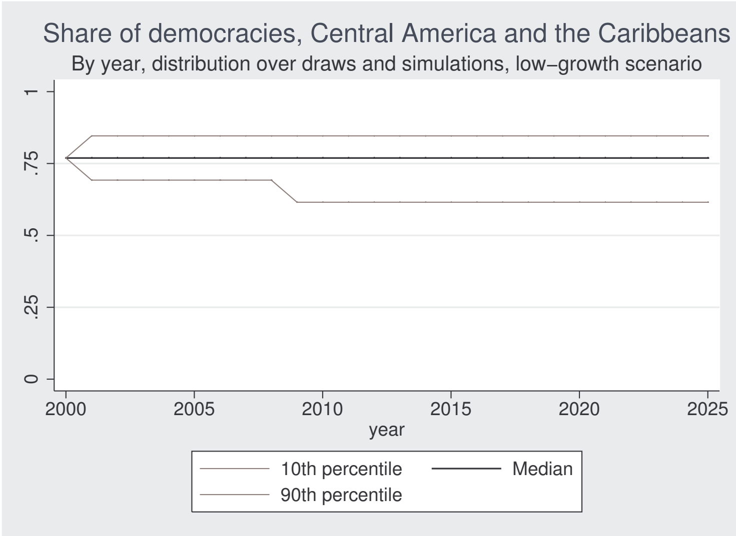

Figure 18:

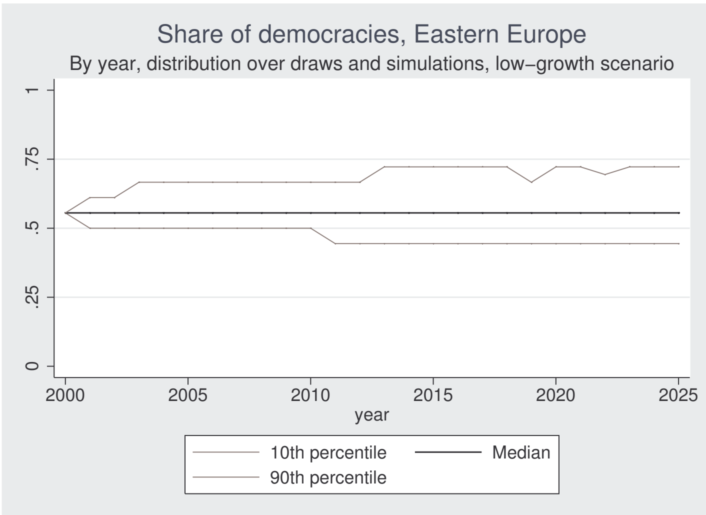

Figure 19:

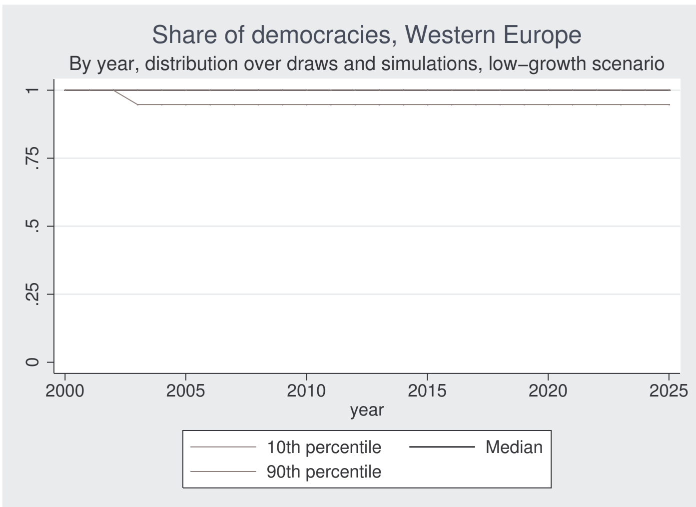

Figure 20:

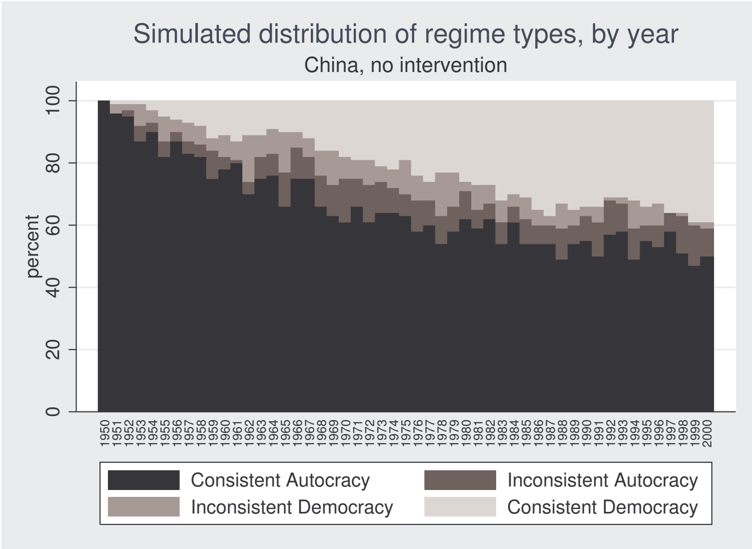

Figure 21:

ern Europe, respectively. Neither of the European regions are predicted to change much in terms of the share of consistent democracies. Eastern Europe is predicted to remain half democratic, whereas Western Europe will remain fully democratic except for in a minority of simulations.

4.3 Individual-country simulation: The case of China

Our final Figure (Figure 21) shows the distribution of predicted regime type for a single country – China. China starts out as a low-income autocratic country with an annual growth rate of about 10%. In 2025, China remains autocratic in only half of our simulations. In more than 40% of the simulations, it has transitioned into a consistent democracy.

5 Conclusion

We have presented simulated predictions of changes in the global and regional distributions of regime types. The simulations were made based on a statistical model that borrows features from the most recent quantitative research on the determinants of democratization. The final specification of the model was guided by comparing the out-of-sample predictive performance of the various candidate specifications. Although projections over the next 25 years are inherently very uncertain, over prediction success for the 1985–2000 period exceeded 60%, based on estimates for the 1950–1984 period. We have some confidence in the aggregate predictions although the results for individual countries (e.g. that presented for China) should be treated with caution.

We find as previous studies (e.g. Pzeworski et al., 2000; Epstein et al., 2006; Gates et al., 2006) that economic development is important both for the transition into and stability of democracy. We find, however, that the impact of income is marginal in the sense that the poorest countries in the world have to grow for a much longer period than treated here for income to really make a difference.

The statistical model also identifies a statistically significant effects of geographic diffusion, confirming results in Gleditsch (2002) and Gleditsch & Ward (2006). Although the diffusion effects are statistically significant, they are not of substantively great importance: changes to full democracy are found to have fairly long-lasting effects in the country that changes, but the effect on the neighborhood beyond that is very marginal.

The pseudo-R^2 (the log likelihood ratio index) is as high as 0.819 in the model underlying the simulations. The model seemingly does a good job in explaining the regime type of countries. Still, the simulation shows that the predictions emerging from the model are quite uncertain. The out-of-sample predicitive ability is 48.5% when estimating the model on half of the world's countries for the $1950 - 2000$ period and predicting on the other half. The prediction success is somewhat higher (60.7%) when estimating the model on all countries for the $1950 - 1984$ period and predicting for the remaining years.

The relatively poor performance may be due to the sparsity of exogenous variables in the current version of the model. Most of the explanatory power lies in the variables representing the previous state of the country. Our model only performs marginally better than saying that 'a country's regime type next year is the same as it has this year'. Most of what determines *change* is not accounted for by the model. Correspondingly, the simulations may underestimate the forces that alter global democracy levels. Figure 2 still predicts a fairly consistent pattern of democratization – although many countries only change into inconsistent democracies, the simulations indicate that the autocratic regime types will largely disappear by 2025.

The analysis reported here needs development along several lines. First, we will exploit the opportunity given by our simulation tools to validate and improve the model using split-sample validation. Second, we will extend the set of exogenous explanatory variables in the model. The extent to which GDP per capita is based on oil extraction, for instance, is a major ommission here (Ross, 2001). We will also look into variables characterizing countries' history such as colonial history. Third, we will investigate whether variables have differential effects on the likelihood of change rather than on level as is currently assumed. Przeworski et al. (2000), for instance, argue that high income levels stabilize democracies but does not alter the probability of transition into democracy. This aspect is currently disregarded here.

We consider several extensions to the study reported here. First, it is necessary to extend the underlying dataset up to 2004 for which all variables are readily available, and to explore a larger number of explanatory variables (e.g., colonial history). Second, the methodology developed here is easily extended to other dependent variables of interest such as the onset and termination of internal armed conflict.

6 References

Beck, Nathaniel; Jonathan N. Katz & Richard Tucker, 1998. 'Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable'. *American Journal of Political Science* 42(4):1260–88.

Gates, Scott; Håvard Hegre, Mark P. Jones & Håvard Strand, 2006. 'Institutional Inconsistency and Political Instability: Polity Duration, 1800-2000', *American Journal of Political Science* 50(4): 893–908.

Epstein, David L.; Robert Bates, Jack Goldstone, Ida Kristensen, & Sharyn O'Halloran, 2006. 'Democratic Transitions', *American Journal of Political Science* 50(3): 551–569.

Gates, Scott; Håvard Hegre, Mark Jones & Håvard Strand, 2007. 'Democratic Waves? War, New States, and Global Patterns of Democracy, 1800–2000'. Paper presented to the 6th Pan-European Conference on International Relations, Turin, 12–15 September 2007.

Gleditsch, Kristian S., 2002. *All International Politics is Local. The Diffusion of Conflict, Integration, and Democratization .* Ann Arbor, MI: Michigan University Press.

Gleditsch, Kristian & Michael D. Ward, 2006. 'Diffusion and the International Context of Democratization' *International Organization* 60(4): 911–933.

Huntington, Samuel P., 1991. *The Third Wave: Democratization in the Late Twentieth Century.* Norman, OK & London: University of Oklahoma Press.

Jaggers, Keith & Ted Robert Gurr, 1995. 'Tracking Democracy´s Third Wave with the Polity III Data', *Journal of Peace Research* 32(4): 469–482.

McLaughlin, Sara, Scott Gates, Håvard Hegre, Ranveig Gissinger & Nils Petter Gleditsch,

1998. ‘Timing the Changes in Political Structures: A New Polity Database’, *Journal of Conflict Resolution* 42(2): 231-242.

Maddison, Angus, 1995. *Monitoring the World Economy 1820-1992.* Paris: Development Centre of the Organization for Economic Co-Operation and Development.Mooney, Christopher Z., 1997. *Monte Carlo Simulation.* QASS no. 116. Thousand Oaks, CA, etc.: Sage.

Przeworski, Adam; Michael E. Alvarez, José Antonio Cheibub, and Fernando Limongi., 2000. *Democracy and Development. Political Institutions and Well-Being in the World, 1950-1990.* Cambridge: Cambridge University Press.

Raknerud, Arvid & Håvard Hegre, 1997. ‘The Hazard of War. Reassessing the Evidence for the Democratic Peace’. *Journal of Peace Research* 34(4):385–404.

Ross, Michael, 2001. ‘Does Oil Hinder Democracy?’, *World Politics* 53(April):325–61.

Sanhueza, Ricardo, 1999. ‘The Hazard Rate of Political Regimes’, *Public Choice* 98: 337–367.

Summers, Robert & Alan Heston, 1991. ‘The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988’, *Quarterly Journal of Economics* 106(2): 327–368. [Version 5.6 of the data, released January 1995, available at http://datacentre.chass.utoronto.ca/pwt/.].

Ward, Michael D.; John O’Loughlin, Jordin S. Cohen, Kristian S. Gleditsch, David S. Brown, David A. Reilly, Corey L. Lofdahl, & Michael E. Shin, 1998. "The Diffusion of Democracy, 1946-1994," *Annals of the Association of American Geographers* 88: 545-574. Figure downloaded from http://weber.ucsd.edu/~kgledits/vita.html.

World Bank, 2000a. *World Development Indicators 2000.* Washington, DC: Development Data Center, International Bank for Reconstruction and Development.

World Bank, 2000b. *World Development Indicators on CD-ROM.* Washington, DC: Development Data Center, International Bank for Reconstruction and Development.