Fragility in the data

Hannes Mueller

LSE | BLAVATNIK SCHOOL OF GOVERNMENT | UNIVERSITY OF OXFORD | IGC International Growth Centre

Commission on
State Fragility, Growth and Development

Hannes Mueller is a tenured scientist at the Institute for Economic Analysis (IAE-CSIC) and a Barcelona Graduate School of Economics Affiliated Professor.

About the commission

Contents

Background and definitions

The LSE-Oxford Commission on Growth in Fragile States aims to bring frontier research and policy knowledge together to support the development and implementation of effective growth policies in fragile states. The purpose of this note is to generate an overarching background piece that lays out the economic facts on fragile states and serve as a basis for discussion in the initial work of the Commission. The goal is to provide a broad view of the facts that emerge from existing data sets and approaches to measuring fragility, how these dimensions of state fragility are correlated with each other, and whether there is a relationship between political stability/security and economic development.
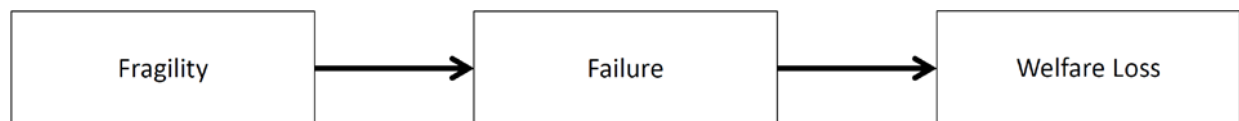
```
┌─────────────┐        ┌─────────────┐        ┌─────────────┐
│  Fragility  │───────▶│   Failure   │───────▶│ Welfare Loss│
└─────────────┘        └─────────────┘        └─────────────┘
```

Figure 1: Conceptual distinction

Before we approach the data, we will make a crucial conceptual distinction, which we also illustrate in **Figure 1**. In this note we will distinguish between fragility, failure, and the loss of welfare implied by it.

Fragility

Fragility means that *a state is at risk of failure*. As such, fragility is a concept that indicates future risks. This has many implications, which we discuss below. Fragility will be treated here as the likelihood of future failure. This interpretation leads us to adopt a forecasting framework in the second part of the note. We will argue that this framework caters perfectly to the task of a policy maker who is allocating effort in developing countries.

---

[1] I thank Gaurav Mehta for his excellent research assistance on this project.

Failure

Failure is when *a state does not manage to prevent a crisis, economic or political, which has the potential to harm the welfare of its population*. The distinction between failure and fragility also implies that without a notion of what failure is, it is impossible to define what fragility is. To define failure, we will use data on e*conomic decline, institutional changes, political turmoil* and *armed violence*. We then show that, based on the data of past failures, one can develop statistical models that evaluate the risk of failure. We then test the validity of these models in the data.

Welfare loss

Three facts emerge from the data on past failures and risk of failure:

1. Past failure is strongly correlated with economic development and well-being today

2. Existing indicators of fragility are related to past failures but less to risk, i.e. fragility

3. Political institutions play a key role in determining fragility on several dimensions

Past fragility explains present welfare

In this section, we will establish the link between fragility and welfare, i.e. the second relationship in **Figure 1**. We do this from a macro view looking at correlations in the cross- country data. We begin by *analysing the growth histories of poor and rich countries to show that past failures play an important role in explaining today's welfare*.

Fragility and economic growth

There is a simple way to motivate the focus on fragility in the economic growth data. *Economic growth has two dimensions:*

1. **Consistency**: Does the country grow consistently, or does it flip-flop back and forth between boom and bust?

2. **Intensity**: Does the country have large booms with high growth rates when it grows?

It turns out that *the first dimension is a lot more important when explaining gross domestic product (GDP) per capita today*. This is visualised in **Figure 2**. Here we plot the share of years with negative growth in the history of a country and compare it to GDP per capita in 2014. There is a clear negative correlation: *countries with more years of negative growth in their past are poorer today*.
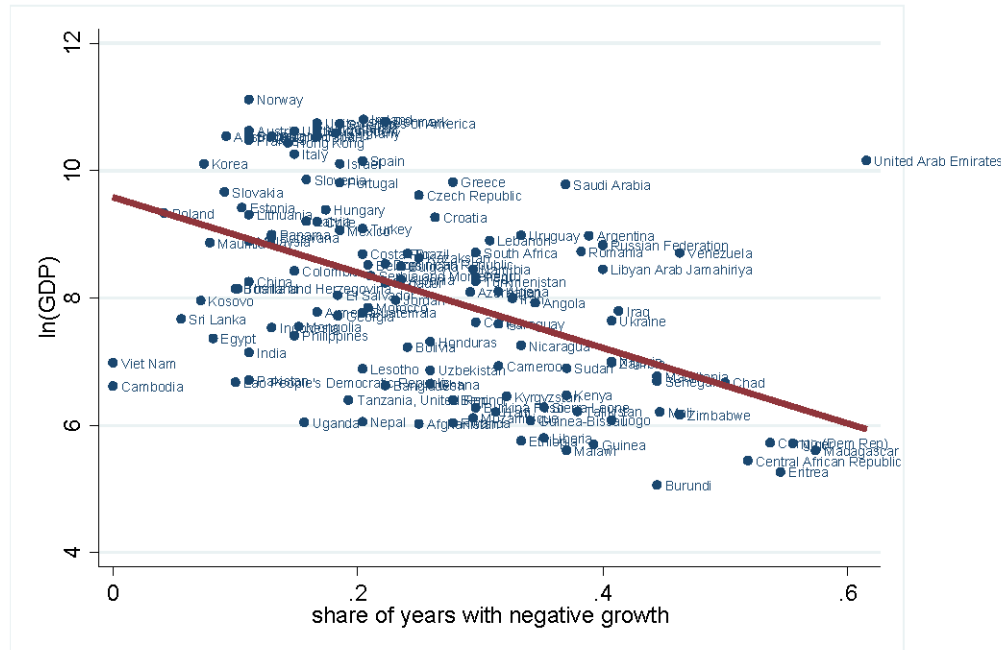
Figure 2: GDP per capita in 2014 and past declines

The correlation is quite strong. If we interpret it as causal (which is obviously problematic), shifting a country from a past with 60% negative growth years to 20%, would entail an increase in its GDP per capita of 2.5 log points. This is a 12-fold increase!

GDP per capita and past growth

What is important here is that the relationship between GDP today and average growth, conditional on growing, is considerably less strong. **Figure 3** shows that *the correlation between GDP per capita today and the past average growth rate is positive but is much less strong than one would expect*. In fact, the large variance of GDP per capita at all levels of average growth indicates that something else explains the variation in the data. *Consistent economic growth matters much more than intense economic growth*.
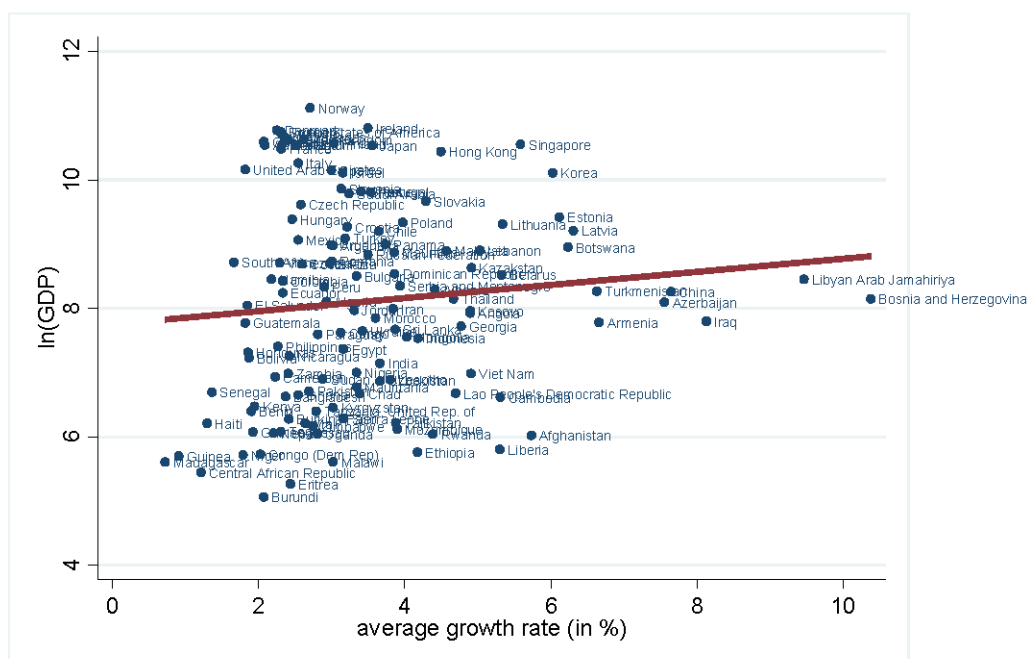
Figure 3: GDP per capita in 2014 and average (conditional) growth in the past

The example of Mexico illustrates this point. Mexico, one of the richer countries in our sample, grew by 2.5% in times of growth. Chad, a much poorer country, grew by 3.4% in times of growth. This means that in the years when both countries grew, Chad would catch up with Mexico. However, Mexico grew for over 80% of all years tracked, while Chad grew for exactly half, i.e. 50% of all the years tracked. *It is not growth per se that makes a country rich but growing consistently.* Therefore, a focus on increasing growth might be the wrong focus for policy. A better focus is ensuring stability, i.e. minimising failure.

Defining failure

However, negative growth is too common to be understood as failure. We therefore focus our statistical analysis on rare events.

We choose three main "failure" events:

1.   A *decline in GDP per capita* of more than 5% in a single year

2.   A *change in political institution*, indicated by a change in the Polity2 score

3.	An *outbreak of high intensity violence* in the country (civil war)

We also analyse other failures, like revolutions, purges, armed conflicts and refugee crises. The general notion of a correlation between past failures and today's outcomes always holds. However, we show in the final section of this note that *predicting these other failures is more challenging with standard data*, i.e. it is harder to understand fragility along different dimensions.

Illustrating the relationship between failure and GDP per capita:

- In **Figure 4** we present the relationship between GDP per capita and our first measure of failure - *large economic decline*. As with negative growth there is a significant negative relationship.

- This relationship is a little less strong in *countries that rely on a high share of natural resource rents in GDP*. To illustrate this, we marked these countries in red in **Figure 5**. These countries, of which the United Arab Emirates is the most extreme example, were very fragile in the past according to our first definition but are rich today.

- In **Figures 6** and **7** we report the correlation between GDP per capita in 2014 and our other two measures of failure: *civil war* and *institutional changes*. Once again, there is a clear relationship between fragility in the past and GDP per capita in 2014.

It is important to note, that *while our different measures of failure are correlated with each other, this correlation is not very strong*. In other words, there are countries with multiple events of economic decline, while other countries feature more civil wars or more unstable political institutions.

Past failures and present welfare

In **Figures 8 to 11** we repeat the same exercise for other measures of well-being. In each case, we show the share of past years with failures on the x-axis, and measures of welfare today on the y-axis. *The correlation in the reported cases is quite striking*. As before, when using large declines in GDP as our measure of fragility, most outliers are countries with large resource rents that tend to be fragile but have lower poverty rates, low child mortality, and a relatively happy population.
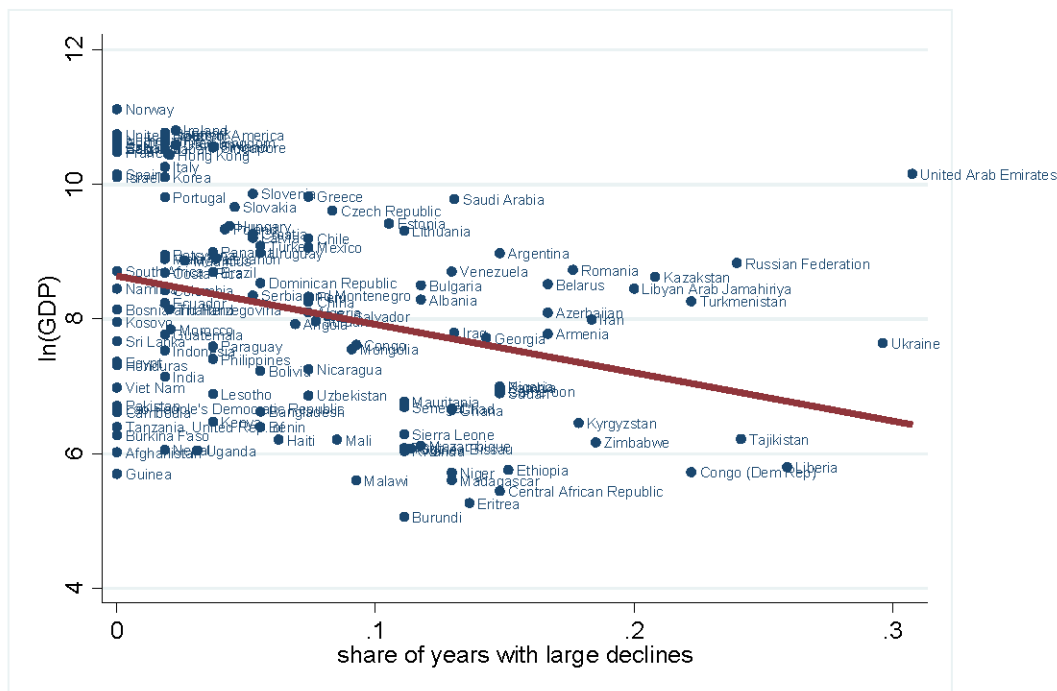
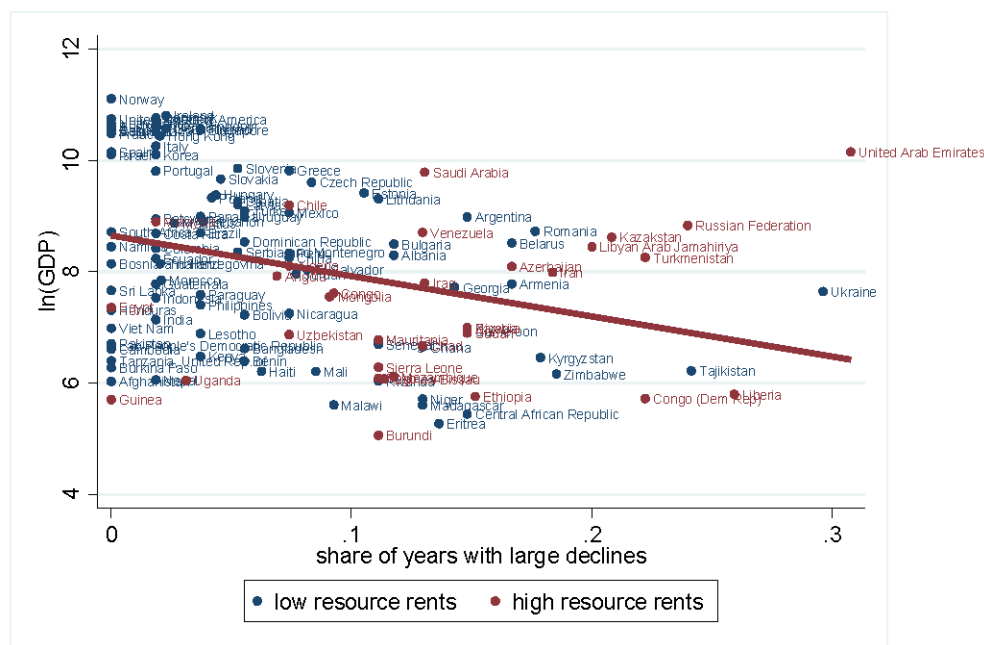**Figure 4:** GDP per capita in 2014 and large declines in the past

Figure 5: GDP per capita in 2014 and large economic declines in the past
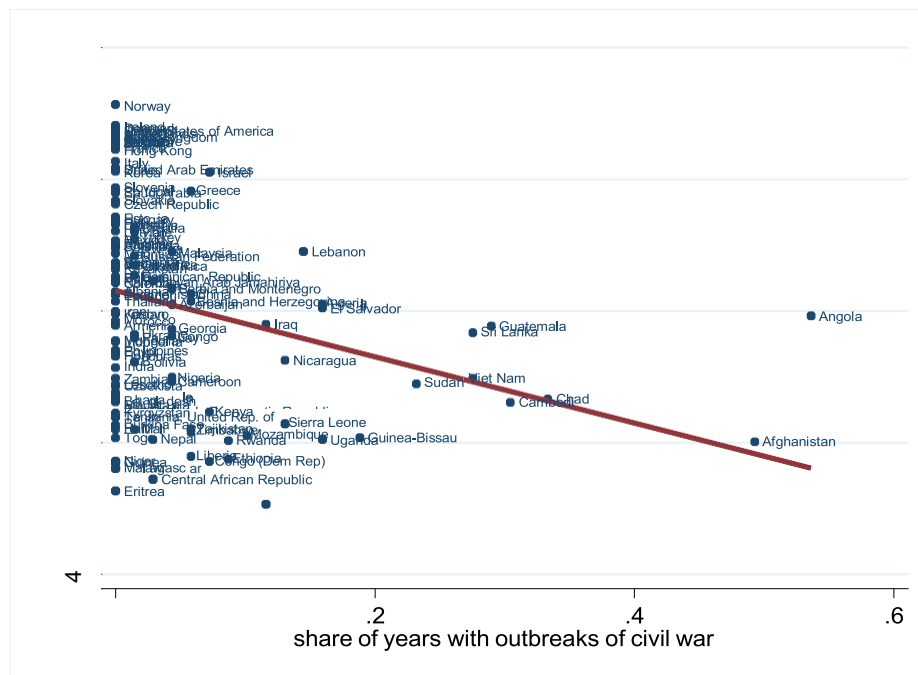
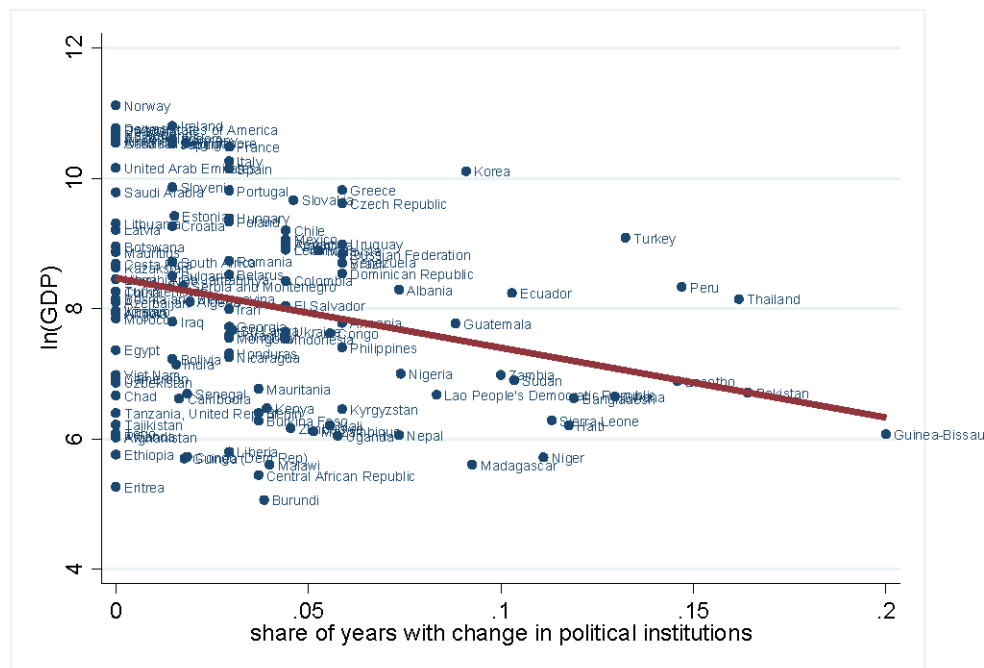**Figure 6:** GDP per capita in 2014 and past civil wars

Figure 7: GDP per capita in 2014 and past changes in institutions
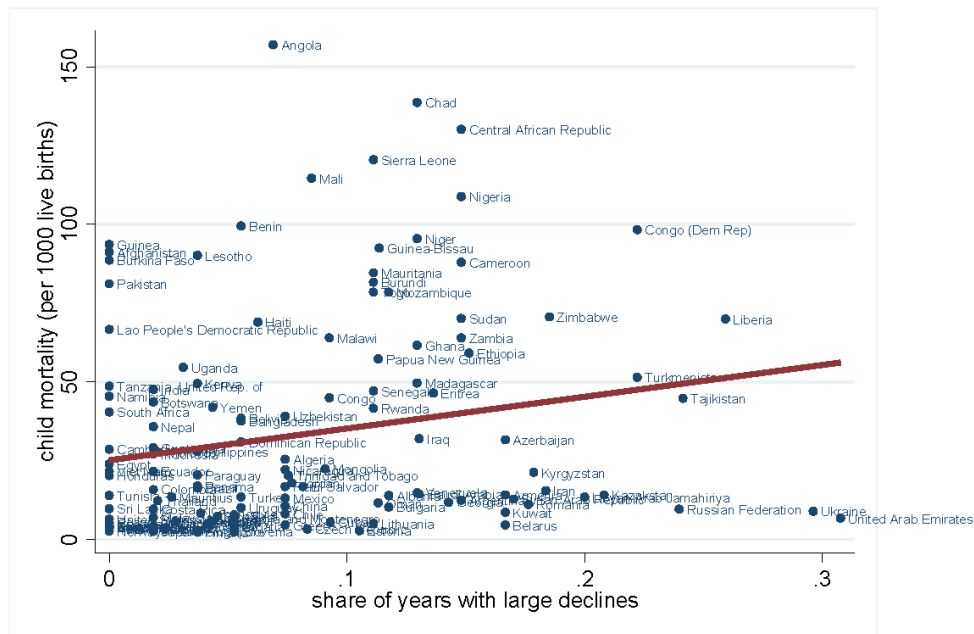
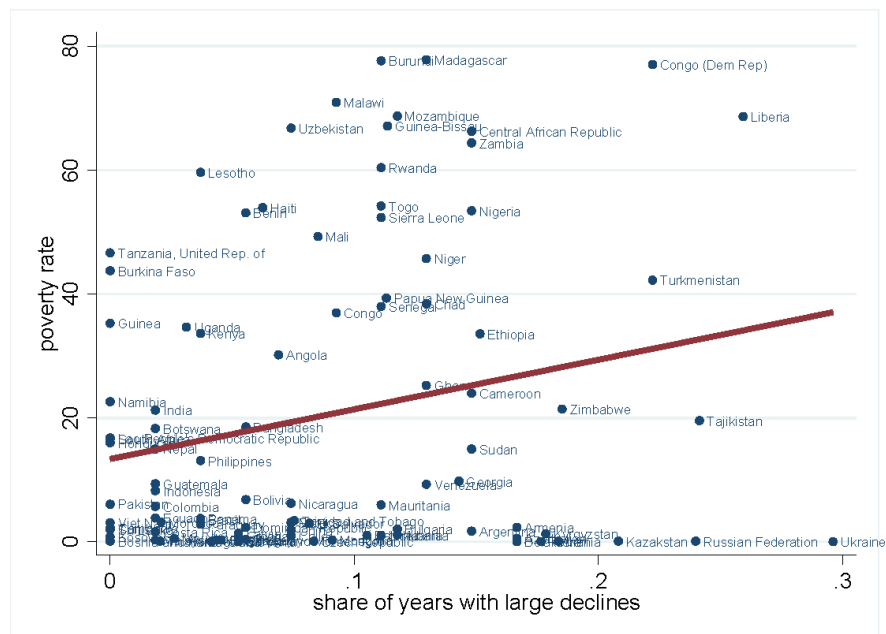**Figure 8:** Child mortality in 2015 and large declines in GDP

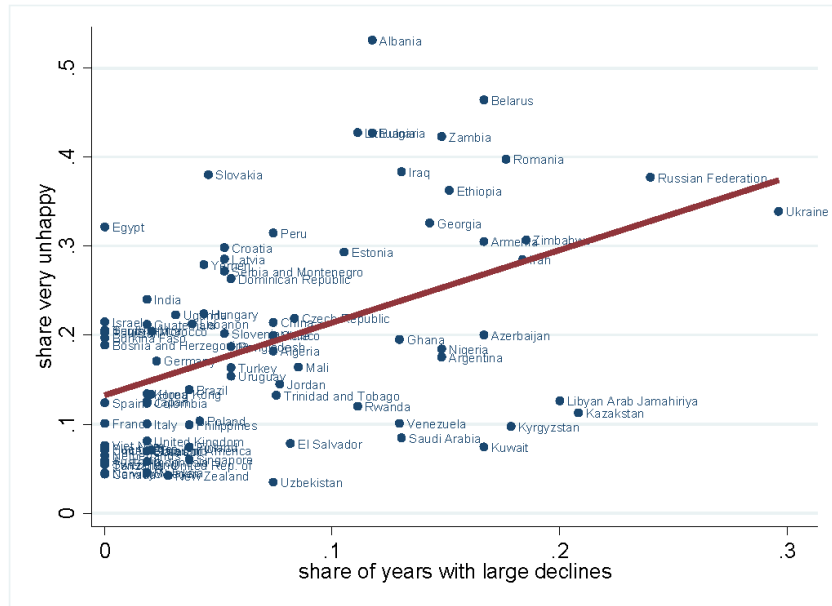**Figure 9:** Poverty rate in 2014 and large declines in GDP
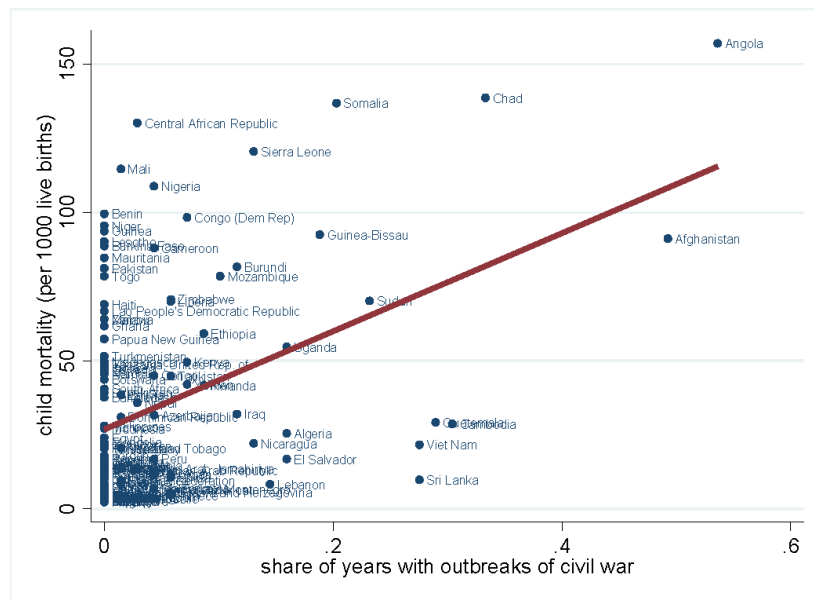
Figure 10: Happiness and large declines in GDP

Figure 11: Child mortality in 2015 and past civil wars

Implications

VARIABLES	(1) ln(GDP per capita) in 2014	(2) poverty rate in 2014	(3) child mortality rate in 2015	(4) share of respondants unhappy	(5) share of respondants unsatisfied with life	(6) change in GDP per capita between 1970 and 2014
share of years with large declines	-6.896***	83.31**	101.0***	60.54***	64.44***	-5.930***
	(1.910)	(37.62)	(37.88)	(15.96)	(23.73)	(1.035)
share of years with civil wars	-5.124***	31.86*	154.6***	32.61	48.44	-1.681**
	(1.220)	(18.39)	(38.79)	(22.76)	(33.32)	(0.653)
share of years with changes in political institutions	-10.68***	84.79	205.1***	26.06	0.712	-3.184**
	(2.529)	(54.36)	(61.63)	(22.79)	(37.59)	(1.336)
Observations	130	101	139	132	132	107
R-squared	0.264	0.088	0.264	0.184	0.085	0.321

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. GDP per capita is real GDP per capita at constant 2005 USD. The poverty rate is the headcount ration at $1.90 in PPP in percent of the population. Child mortality rate is the mortality rate of under 5 year olds per 1000 population. Data on happyness and life satisfaction is from the World Values Survey, various years after 2000. Happyness is measures as the share of respondants reporting to be "not very happy" and "not at all happy". Life satisfaction measure is the share of respondants is the share of respondants that report a life satisfaction below 5 on a scale from 1 to 10.

Table 1: Correlation of past fragility and present welfare

In Table 1 we summarise the findings on the correlation between past fragility and present welfare. In column 1 we show that *there is a strong negative correlation between GDP per capita in 2014, and our three measures of past failures.*

What do these coefficients imply?

Column 1: *GDP per capita in 2014*:

- An increase of fragility in terms of large declines by 10 percentage points coincides with a decrease in GDP per capita by 0.6896 log points, which is almost exactly 50%. In other words, *an economy that has experienced economic decline in 10% more years will have half the GDP per capita in 2014.* This is a very strong association.

- The other coefficients in column 1 are of similar size, which suggests that *fragility in the past might be key to understanding GDP per capita today.*

- We also show in column 6 that this relationship holds when looking at the changes in GDP per capita, i.e. *it is not (entirely) driven by poor countries being more fragile.* From this perspective, *preventing fragility or ensuring stability could have had dramatic effects on GDP per capita in the future.*

Column 2: *Poverty rate in 2014*:

- In column 2 of Table 1 we look at the poverty rate in 2014. Again, *the association between poverty and fragility is strong*, as measured by years experiencing economic decline.

- According to our estimates*, a reduction in global poverty by 8 percentage points would come about with a decrease in fragility of 10 percentage points*. The coefficients on the other measures of fragility are again similar but are estimated less precisely.

Column 3: *Child mortality rate in 2015*:

- The most striking results, perhaps, are in column 3 where we explain the variation in child mortality in 2015 with past fragility. The mean and variance of this variable is around 35 deaths in a population of 1000.

- The estimated correlation with the different fragility measures indicates that a *10 percentage point reduction in fragility would coincide with a reduction in child mortality of between 10 and 20 deaths in 1000*.

Columns 4 and 5: *Share of respondents unhappy or unsatisfied with life*:

- In columns 4 and 5 we check whether this decline in objective measures also coincides with declines in subjective well-being. Cross-country differences are a little more challenging to interpret, as a country's culture could play an important part in responses. Even so, using data from the World Values Surveys, we find the kind of correlations we anticipated from the previous columns.

- Column 4 suggests that *the share of respondents who report being unhappy or very unhappy falls by 6 percentage points when the share of economic declines in the past falls by 10 percentage points*. This is about half a standard deviation.

- As the effect of failure on happiness is, perhaps, less well studied we also look at changes across time in this data. We find that *in years with large economic decline, life satisfaction and happiness fall by more than one standard deviation compared to the average in the country in other years*. This suggests that there is indeed a link between experiencing a crisis and welfare.

- Needless to say, all these correlations are just that, correlations. However, there are a lot of better identified micro studies that underline the importance of, for example, political

uncertainty or violence for investment. *What the correlations in the macro data tell us is that these micro effects could indeed have severe macro consequences*.

Indicators of fragility and failure

Given the important role that fragility plays in explaining welfare, it is important to understand fragility and its indicators. This fact is now broadly accepted and, indeed, several organisations offer analysis of fragility in reports and capture their analysis in fragility indicators. In this section we first provide a brief overview over of this available data and then discuss two indexes of fragility.

Our argument is that, *despite an explicit discussion of risk, existing methods do not distinguish between fragility and failure. Instead, fragility is defined as something that has failed*. This is like evaluating the fragility of an object by observing whether it broke in the past. Of course, this can be useful if risk of failure is correlated across time. However, this approach to defining risk cannot determine future risks if they are new. We show that this is indeed the case for the only index with enough data to test whether it is able to determine future risks.

Existing methodologies and data

The Organisation for Economic Co-operation and Development (OECD) Report, *States of Fragility 2016, Understanding Violence***:** The newest methodology comes from the OECD. The OECD defines fragility *as a combination of exposure to risk and insufficient coping capacity of the state, systems and /or communities to manage, absorb or mitigate those risks*. Fragility can lead to negative outcomes including violence, the breakdown of institutions, displacement, humanitarian crises or other emergencies.

The framework in the 2016 report considers five dimensions of fragility based on a classical contextual risk typology: Economic, Environmental, Political, Security, and Societal. For all these dimensions of fragility, the OECD framework selects certain indicators that indicate risk and copying capacity.

For example, in the economic dimension of fragility, *resource rent dependence*, the *GDP growth rate* and the *unemployment rate* are used as some of the risk indicators. Copying indicators include *men and women in the labour force,* among others. Additionally, in the political dimension, *regime persistence* features as one of the risk indicators, and *decentralised elections* serves as

one of the copying indicators.

Overall, the report makes use of 43 indicators covering all dimensions of fragility. Then principle component analysis is used to construct an index on each dimension and an overall index. However, from the data going into the analysis it should be clear that there is no explicit distinction between welfare outcomes, failure and fragility.

In 2016, the OECD fragility framework ranked 56 countries. These were then divided into 'extremely fragile' and 'fragile'. Extreme fragility means high fragility in all five dimensions, as well as widespread armed conflict or very significant levels of collective and social violence. The remaining fragile contexts also include fragility in all dimensions; however, the level of violence is generally lower than in the extremely fragile category. Again, failure is explicitly used to define fragility for extremely fragile countries.

Fragile States Index: The most easily accessible and most commonly used data on fragility is from the Fund for Peace. The Fund offers a *Fragile States Index*, which captures the fragility of a country and is explained in detail in the appendix. While the details here are different from the OECD index, the basic idea of combining a multitude of variables on several dimensions and calculating an index from this is shared. Again, a potential problem is that welfare outcomes, failures and fragility are not explicitly distinguished in the data. The index is publicly available and, importantly, is provided for the years 2005-2016. This allow it to be used for panel analysis to test whether it picks up failures before they happen.

The World Bank: The World Bank defines "fragile or conflict-affected states". This classification is based on the World Bank's Country Policy and Institutional Assessment (CPIA) rating below a threshold of 3.2, and/or the presence of a United Nations (UN) and/or regional peacekeeping or peacebuilding operation. The CPIA is based on 16 criteria which try to capture a broad variety of aspects, from policies for social inclusion to institutions. Interestingly, there is ongoing work inside the World Bank, which tries to develop a forecasting framework for conflict to improve upon this definition (See Celiku and Kraay, 2017). The result is quite a dramatic difference in which countries are marked as fragile

Other indicators: According to the United States Agency for International Development (USAID), the key difference between fragile states and states that are more capable, is the nature of the relationship between the government and the governed. In fragile states, this relationship is poor.

The government acts in ways that create conditions (or outcomes) that are broadly seen as ineffective, illegitimate, or both.

In the G7+ framework, a state of fragility can be understood as a period of time during nationhood when sustainable socio-economic development requires greater emphasis on complementary peace building and state building activities, such as building inclusive political settlements, security, justice, jobs, good management of resources, and accountable and fair service delivery. Fragility is categorised in five dimensions: Inclusive politics, Security, Justice, Economic Foundations, and Revenue and Service. A pilot study on the framework has been conducted in five countries where some indicators have been identified, however this framework is largely a work in progress.

Fragility index data and welfare

We now turn towards an analysis of the Fund for Peace *State Fragility Index and OECD fragility index data to see how they compare to our data on welfare*:

- From the OECD index, we recorded the ranking of countries according to the newly developed OECD fragility measure, which only looks at a sub-section of 56 countries. We rank countries such that a higher ranking implies higher fragility.

- From the Fund for Peace State Fragility Index, we look at the years 2005-2016 for panel analysis to test whether it picks up failures before they happen. Again, a higher index means higher fragility.

Figures 12 and **13** show that both *the fragility index from the Fund for Peace and from the OECD are negatively associated with the GDP per capita in 2014*. **Figure 12** shows that the Fund for Peace index is particularly closely associated with today's GDP. What is particularly interesting is how fragility differs from GDP per capita. What we see in Figure 12 is that resource-rich countries like Saudi Arabia, Libya, Qatar and Iraq are coded as relatively fragile for their level of GDP, i.e. above the fit line. This pattern also appears when we define fragility explicitly through failures. *Resource rich countries offer relatively high levels of welfare for their levels of fragility* – an intuitive result.

As shown in **Figure 13**, The correlation with the OECD measure is less close and, indeed, there is no association with our past fragility measures. However, it is important to note that the OECD only looks at a subset of countries – it does not aim to develop a globally valid measure of fragility.

It can also be shown that *our measures of fragility (the share of years with failures) are closely associated with the fragility index from the Fund for Peace*. This is comforting as it indicates that our measures, based on an explicit indicator of failure, are associated with a commonly used indicator of failure.
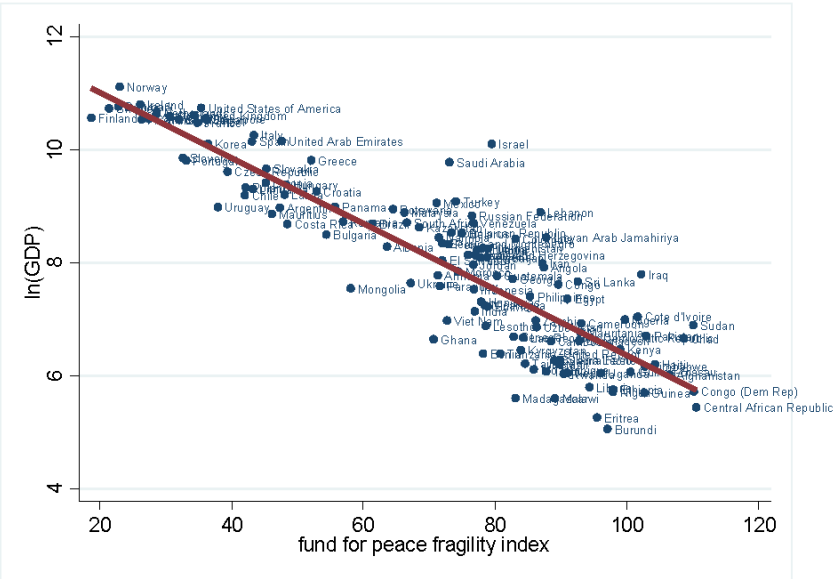

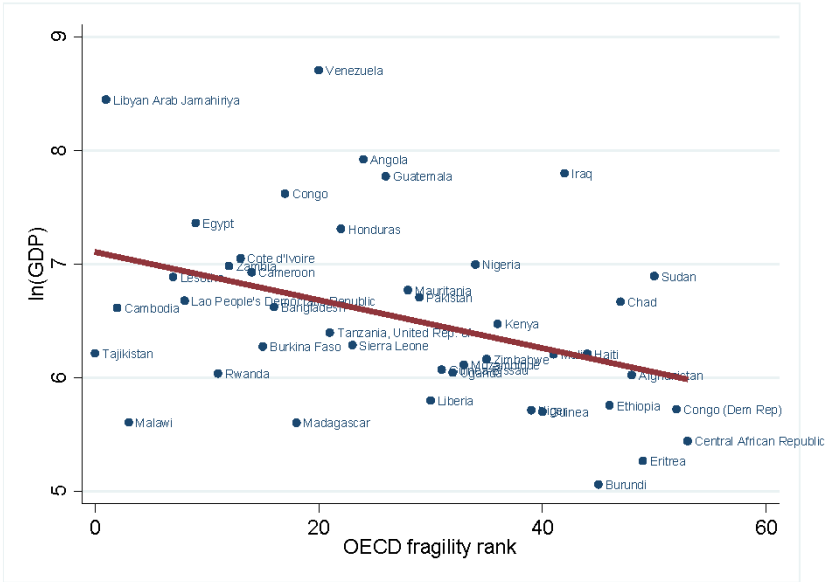**Figure 12:** GDP per capita in 2014 and the Fund for Peace Fragility Index

Figure 13: GDP per capita in 2014 and the OECD fragility ranking

Predictive capacity of fragility index data on future failure

Welfare and past fragility are closely related in the most commonly used fragility indexes. What about future failures? Can the fragility index predict future failure? There are two ways to interpret this question:

1. The first requires *countries with past failures to be ranked as more fragile than countries without past failures*. For example, Afghanistan should be ranked as more fragile than, say, the United Kingdom (UK). However, this is clearly a relatively low bar.

2. Much more interesting is *whether the fragility index increases before the failure occurs*, i.e. whether the timing of fragility can be predicted.

Ideally, one would want an index of fragility that fulfils both conditions. If the first requirement is fulfilled but the second is not, then new failures cannot be predicted. Essentially, the definition of future fragility then relies on past failures. This is akin to categorising objects as fragile because they broke.

Fund for Peace State Fragility Index predictions

The available data from the Fund for Peace Index allows us to run tests to determine its predictive capacity. In **Figure 14**, we show the example of a recent 'new' failure. It shows data for four Arab Spring countries, which experienced failures in several dimensions in the year 2011 (marked by a red line).

If the index does indeed capture future failures, we would expect a movement up in the fragility index before the failure, i.e. in the years 2009 to 2011. However, *there was no general anticipation of instability by the index*. In three out of four countries, the index even fell from 2009 to 2010, and again in 2011. In Tunisia, instability had already started in 2010, meaning the increase in 2011 is not surprising.

The Fund for Peace Index reflected dramatic movements only after the instability started, i.e. in the years following 2011. This is a clear indication that the index picks up failures after they happen, not before they happen. As failures typically happen where they happened before, this is a reasonable approach if one only wants to fulfil the first condition and spot renewed failures. However, it fails to predict or reflect new instabilities. The reason is undoubtedly that *the method by which the index is constructed does not explicitly distinguish between fragility and failure*.

Developing a model of risk of future failure

Using the *Fund for Peace State Fragility Index*, we develop a model to base our definition of fragility. This is explicitly understood as the risk of failure in the future, i.e. *we will base our measure of risk on the explicit aim of predicting our three definitions failures in the future*.

Methodology

To do this, we regress the failure dummy on the lagged index and include country fixed effects. We then ask whether the '***within model***', using the State Fragility Index, can be used to predict failure in the next year (out of the sample). This is as opposed to the *'**overall model**'*, which is used to ensure that countries who fail are ranked higher than countries that do not fail. The 'overall model' relies on changes in the State Fragility Index, predicting fragility based on whether it happened before.

The 'within model' is the approach a policy-maker would employ if they were interested in targeting fragile countries. They would first estimate a model, the relationship between the fragility index and failures, and then try to predict the future with this model. Clearly, such forecasting with the 'within model' is *only possible if the fragility index increases before failures and can therefore be used to predict them*.
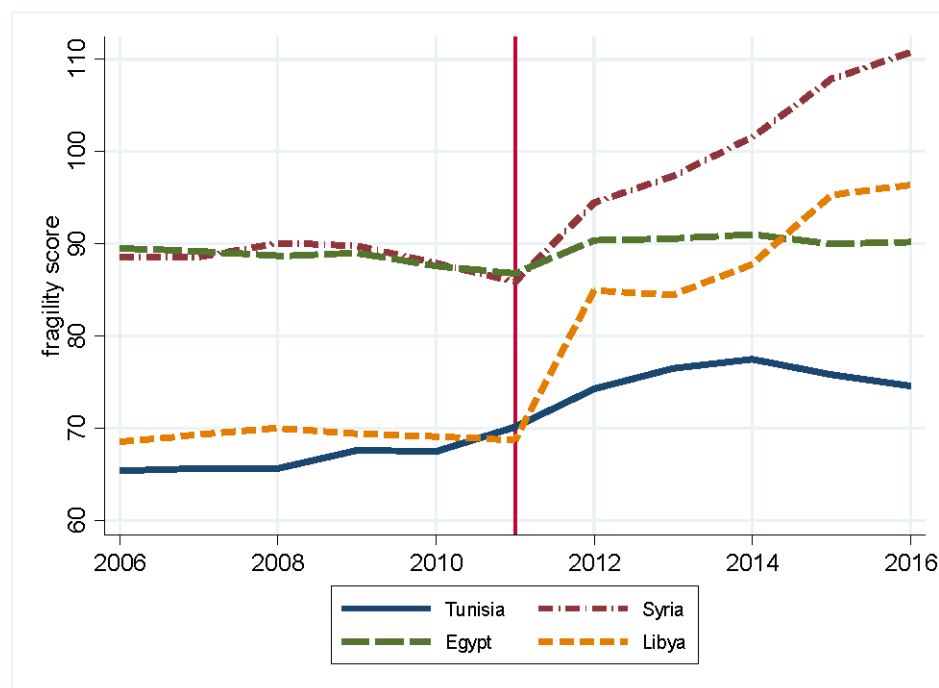

Figure 14: Fragility index and the Arab Spring

The results of the analysis of the forecasting are reported in so called receiver-operating curves (ROC), which show the forecasting power of a model on two dimensions:

1. The **y-axis** records the **true positive rate** – *the share of failures correctly anticipated*. We want this to be as high as possible.

2. However, a model that warns of failure will make many warnings in circumstances that do not turn out to be failures. This sort of mistake is captured by the **false positive rate** – the *share of non-failures that were thought to be failures by* the model. The false positive rate is reported on the **x-axis** in ROC. We want this rate to be a low as possible.

The perfect ROC curve therefore shoots up to a true positive rate of 1 at a false positive rate of 0. *The closer the curve is to the 45-degree line, the less correct the forecast is*. Random forecasts of failures would generate a ROC curve on the 45-degree line.

In **Figure 15** we show an attempt at *predicting large economic declines using the fragility index*:

- The **blue curve** reports *the predictive power of the 'overall model'*, i.e. the model one would use if one only wanted to make sure that countries who fail are ranked higher than countries that do not fail in the next year.

- The **red curve** reports the '*within model*' prediction, which one to use to make sure that movements in time are meaningful, i.e. *that countries are more likely to fail after their fragility increased*.

- The difference between the curves tells us how much the 'overall model' relies on changes in the index, as well as how much it relies on predicting fragility when it already happened before.

Findings

As can be seen clearly, there is a gap between the curves, with the blue curve above the red curve for most values. This indicates that, indeed, *the fragility index more often does not anticipate fragilities through movements in time, instead ranking those countries that failed in the past as fragile*. This is like attributing fragility to an object after it broke.
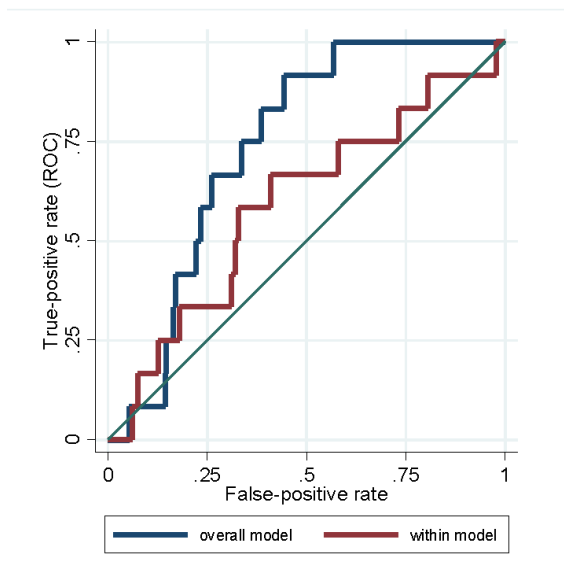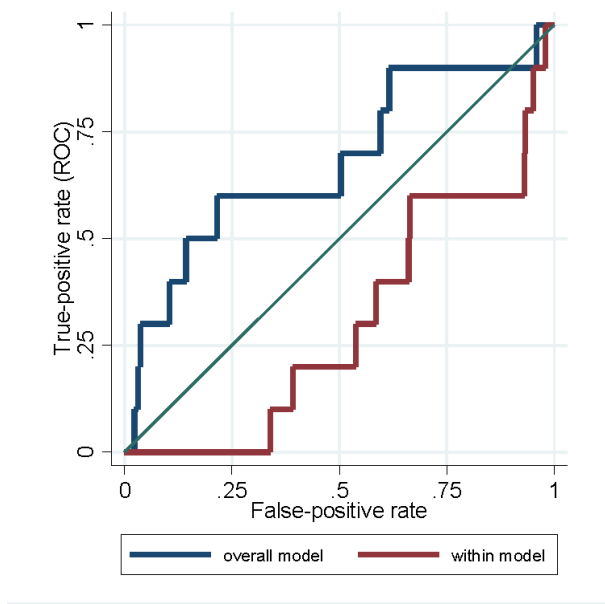
Figure 15: Fragility index forecasting large declines

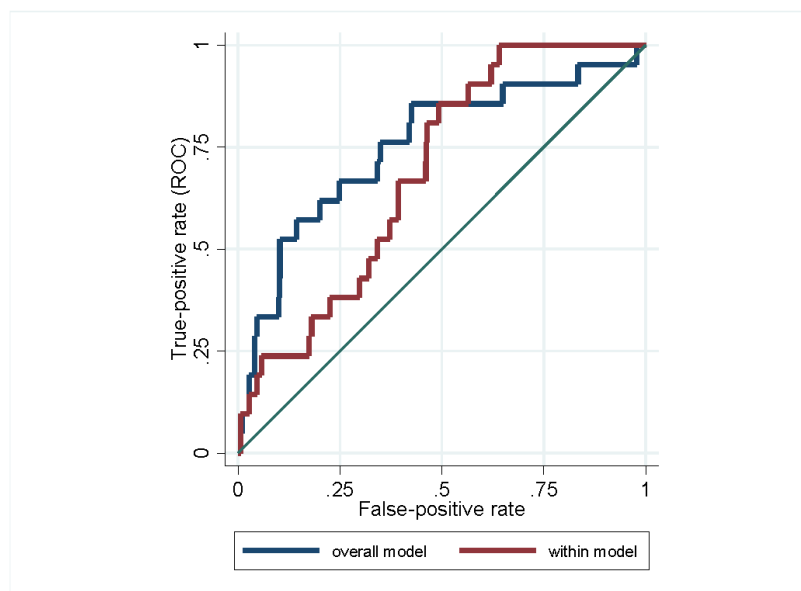**Figure 16:** Fragility index forecasting civil wars

Figure 17: Fragility Index Forecasting Changes in Institutions

In **Figure 16** we show the analogous ROC for the task of *forecasting civil war outbreaks*. It demonstrates how *the fragility index provides the incorrect forecast on the time dimension*. This is in line with the experience in the Arab Spring, in which countries were rated as relatively stable in the year before they descended into violence.

Figure 17 shows that the fragility index does a better job of *predicting institutional changes*. While the 'overall model' still performs better for most values of a false positive rate, *the 'within model' ROC has a relatively high predictive power for false positive rates above 0.5*.

Implications

- The important take-away is that *the Fund for Peace Fragility Index produces relatively little useful time variation to predict failures*.

- This implies that the index attributes fragility according to failures that happened previously. This is perceived as an effective predictor because failures often follow previous failures.

- However, *this does not mean that the variables and method used to construct the index capture future risks*. In fact, we show they do not.

The OECD fragility index cannot be evaluated in a similar manner but from the description in the previous section we already know that it was not developed using forecasting as an explicit goal. While risk is mentioned a lot in the methodology, it cannot be tested whether the risk of future failure is indeed captured in the OECD fragility index.

A model of fragility as risk

Variables of failure

- We conducted a large data-gathering and data-mining exercise in which we used a very simple machine learning algorithm: *the least absolute shrinkage and selection operator (**lasso**) in country fixed effects regressions to identify factors that predict failures in the data*.
- We fed 20 variables, such as foreign investment, GDP per capita, life expectancy, wars in neighbouring countries, political institutions, natural resource rents etc. into the lasso.
- It then provided the *automatically selected variables that best predict failures*.

Defining fragile political institutions

As a proxy for political institutions we use:

- a classification based on the Polity IV where we first follow the literature and define two *thresholds for democracy* (polity 2 score higher than 0 and higher than 5)

- and three dummies for the maximum *score on openness, competitiveness and constraints on the executive*.

- The latter approach to coding political institutions is proposed by Besley and Mueller (2017), which also reviews some of the literature on the link between institutions and stability. Together these variables allow us to track subtle changes in political institutions, while at the same time holding some connection to the literature.

Defining failure

To select the variables, we run the lasso repeatedly in different years. To get a better overview, we also add more failures in this exercise. *We use our three main measures of failure* (large economic declines, civil wars and changes in political institutions) and *add revolutions, purges, armed conflicts, large refugee streams and famines*. Note, that the lasso regressions are a 'within-sample' method, which means we do not explicitly optimise 'out-of-sample' forecasting.

Findings

- No failure leads to the same selection of variables.

- A similar set of variables is relevant when predicting the same failures in different periods (results of the lasso are available from the author upon request).

- This suggests that *the answer to what causes fragility will depend on how it is defined exactly*, i.e. which failure risk one tries to predict.

- However, for all failures there are both economic and political variables that play a role.

In **Appendix Figures A1 to A3** we show for our three main measures of failure how a set of four economic and political variables would predict future failures. Both have some forecasting power. As one would expect, the *political variables are more powerful when predicting civil war and changes in political institutions*.

Modelling risk across the three main indicators of failure

We build on the analysis using seven variables to run 'out-of-sample' forecasting, as we did

previously with the fragility index.

1. Political institutions

To keep the *focus on political institutions*, we only use one economic variable (economic growth) and six political variables, five of which we take from the Polity IV dataset. The aim here is to provide a model that can be used to forecast risk across our three main indicators of failure. We report the 'within-sample' performance of the model in **Table A1** in the Appendix. It shows that:

- *Economic growth* is negatively correlated with future failures.
- The share of the *population that is discriminated against* is positively associated with political turmoil.
- More *democratic institutions* are associated with less fragility. However, this is not a linear relationship, and the details here are different depending on the failure one tries to predict.
- Only *high openness* is negatively correlated with outbreaks of civil war and changes in political institutions. However, given that we control for the democracy dummies, this is not a general take-away, as different parts of the variation in the Polity IV dataset take on the forecasting power. Indeed, our analysis suggests that there could be a trade-off between good risk evaluation and interpretable results in the Polity IV data.

How does this model fare when predicting failures 'out-of-sample'? To assess this we first predict failures for the *years 1989 to 2014*, and always use the model estimated the year before, just like a policy maker would. We then summarise the ability to forecast in **Figures 18 to 21**.

2. Large economic declines

In **Figures 18** and **19** we first report on the task of forecasting *large economic declines* in the data. Strikingly, the 'overall' and 'within' model share similarly strong predictions. This implies that we are predicting with the time variation of the model. The forecast is also much better than what is achieved with the fragility index:

- The false positive rate is 25%.
- The true positive rate is close to 60%.
- This means that 60% of large economic declines are correctly anticipated.
- Mistakes in predictions are only made in 25% of the cases.
- Even better precision can be reached if only half the failures need to be anticipated. At a true positive rate of 50%, the false positive rate is only about 10%.
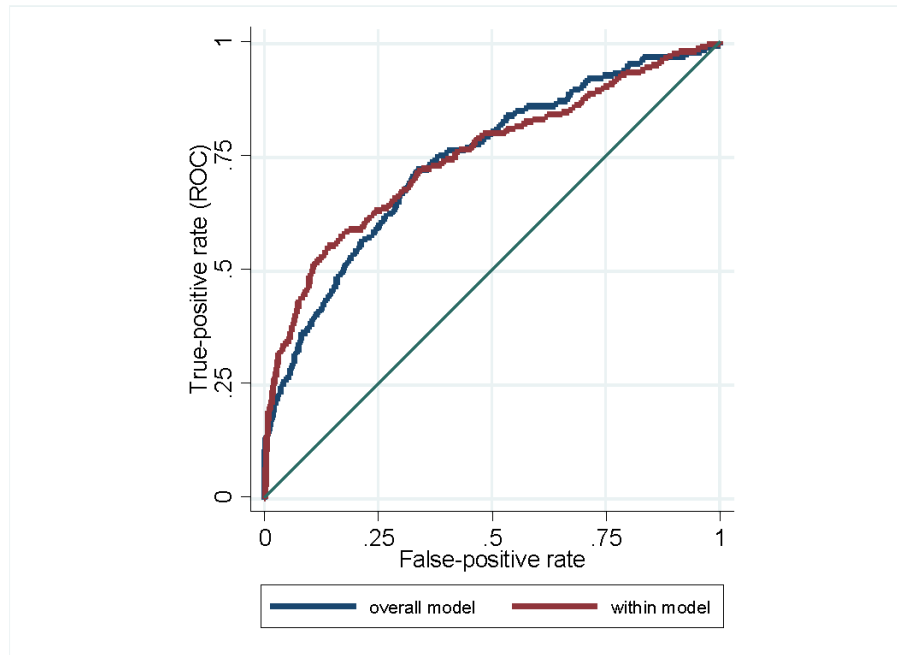
Figure 18: Forecasting large economic declines

In **Figure 19** we show the raw material going into the ROC. Here the **red bars** *represent actual large economic declines*, while the **black line** reports on (the 'within' part of) *the probability of failure* as suggested by the forecasting model. The over *3,000 observations* we have between *1989 and 2014* are ranked by this probability. Therefore, there should be increasingly more red bars as we move to the right of the graph. Indeed, this seems to be the case. The red bars become more and more common as the probability increases.

To make the connection between the two curves in **Figure 18 and 19** clear: the part of Figure 19 at the far right generates the steep part of the ROC in Figure 18. Here, very high fragility rankings coincided with an almost 100% of actual failures. This leads to a false positive rate of almost zero, at a true positive rate of 100%.

This illustrates, perhaps, the largest advantage of this approach to fragility: *It can be used for actual cost-benefit analysis*:

- Do policy-makers want to spread their money on many different fragile countries or concentrate on a few countries?
- How many failures can be prevented?
- How many countries will policy-makers invest in that are not predicted to fail?

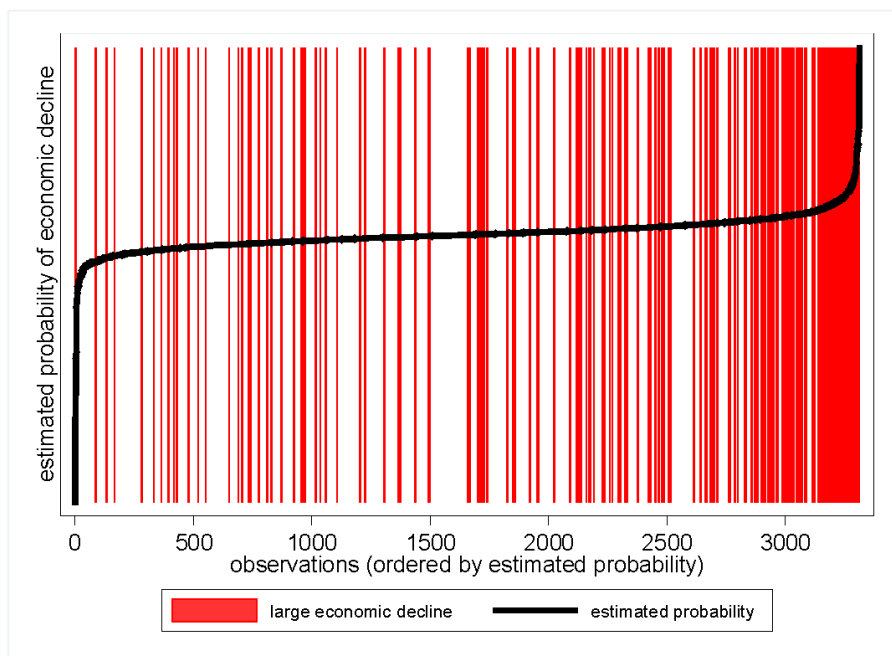*These questions can be answered with a model of fragility as the risk of failure.*

Figure 19: Forecasting large economic ceclines

3. Civil war and changes in political institutions

In **Figures 20 and 21** we show how our model fares when forecasting civil wars and changes in political institutions. *The results are very similar to the ones for large economic declines.* Again, the 'within' variation provides most of the forecasting power, meaning a policy-maker would be able to rely on the time variation in political institutions to spot arising fragilities.

In the Appendix we show the result for various other failures and the same model. Unfortunately, the forecasting power is relatively low in several of *these*. When thinking about fragility, one size does not seem to fit all. This has been acknowledged by existing measures of fragility, such as the OECD measure. However, the definition of fragility as the risk of future failure sharpens this message. A striking finding is, for example, that *the model has little explanatory power when predicting the start of an armed conflict or refugee numbers.*
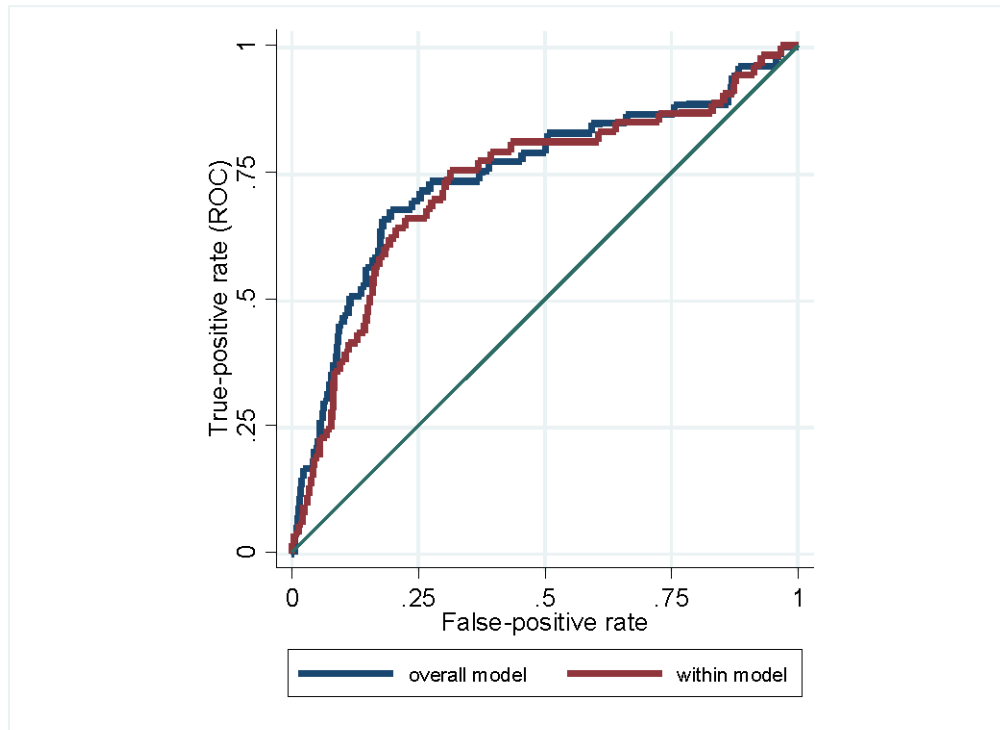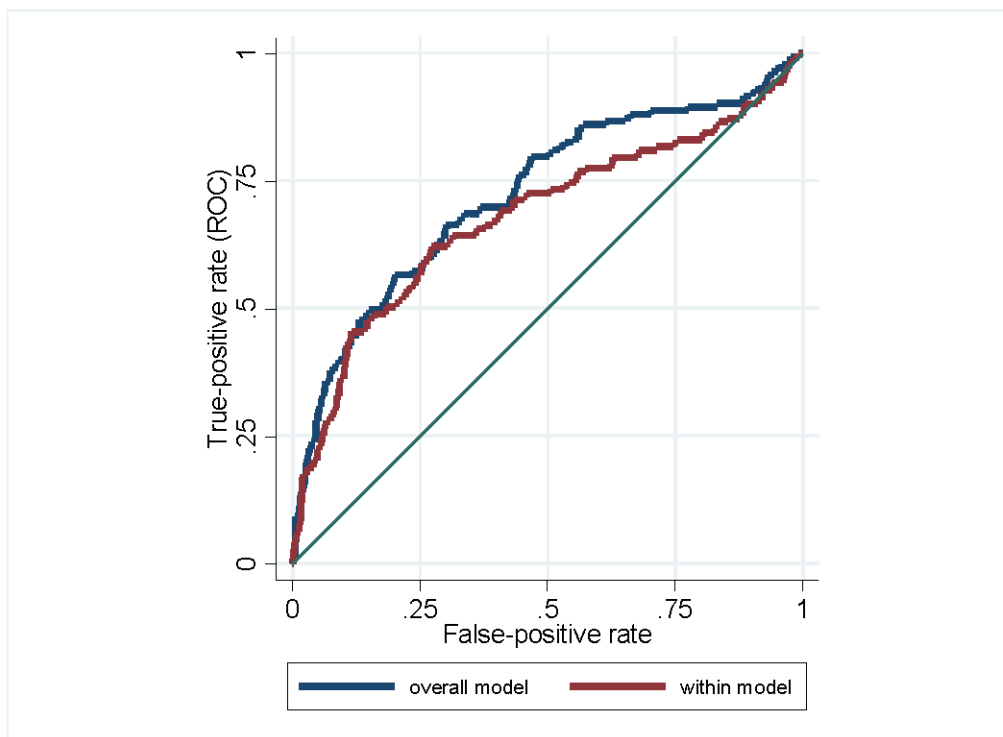
Figure 20: Forecasting civil war starts

Figure 21: Forecasting changes in political institutions

Conclusion

This note has departed from the premise that defining fragility is a crucial first step to analysing it in the data. We have opted to define fragility as the risk of future failure by the state, i.e. *fragility is the risk of experiencing a failure, not failure itself*. This implies that fragility precedes failure and can therefore only be defined once failure is defined. Here we have used data available in large indexes to define failure.

We have also argued that fragility and failures in turn should be distinguished from the consequences of failures, even if identifying causal relationships in the cross-country data is impossible. We have shown that a past failure, i.e. *past fragility, is strongly associated with low welfare today*. This is the motivation for looking at fragility elsewhere.

It is important to keep in mind that fragility is not the same as low growth or poverty. The current state of the art in measuring fragility does not make this distinction explicitly. Instead, it uses a kitchen sink approach to distinguish fragile from less fragile countries. The risk of such an approach is that *countries are categorised as fragile only after they have shown some sort of failure*, not before. This is not unlike an approach where objects are categorised as fragile after they break. Clearly, there is a need to understand what makes countries high or low risk before they destabilise (fail).

We have used a forecasting model to define fragility. The advantage of such an approach is that it can be used for actual *cost-benefit analysis*. Even so, our preliminary forecasting analysis indicates that *different models are needed to understand different fragilities*. Fragility can then only be defined in terms of specific instability risks. In the example of some simple models, we illustrated that *economic growth data, politics and political institutions can be used relatively effectively to identify some fragility risks*.

References

Besley, Timothy and Mueller, Hannes (2017). "Institutions, Volatility and Investment", *Journal of the European Economic Association*, jvx030. Accessible: https://doi.org/10.1093/jeea/jvx030

Celiku, Bledi and Kraay, Aart (2017). *Predicting Conflict*, Policy Research Working Paper; No. 8075. World Bank, Washington, DC. Accessible: https://openknowledge.worldbank.org/handle/10986/

Federal Reserve Bank of St. Louis (2017). *Penn World Table 7.1: GDP per capita*. Accessible: https://fred.stlouisfed.org/categories/33110

Fund for Peace (2005-2016). *Fragile States Index*. Accessible: http://fundforpeace.org/fsi/

G7+ (2018). *New Deal Implementation*. Accessible: http://www.g7plus.org/en/our-work/new-deal-implementation

Goldstone, Jack A, Bates, Robert H, Epstein, David, L, Gurr, Ted Robert, Lustik, Michael B, Marshall, Monty G, Ulfelder, Jay and Woodward, Mark (2010). "A Global Model for Forecasting Political Instability", *American Journal of Political Science*, 54(1): 190{208. Accessible: https://sites.duke.edu/niou/files/2011/06/goldstone-bates-etal.pdf

Marshall, Monty G. and Gurr, Ted Robert (2014). *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2013*, University of Maryland (Emeritus). Accessible: http://www.systemicpeace.org/polity/polity4.htm

North, Douglass, Wallis, John and Weingast, Barry (2009). *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge University Press.

Organisation for Economic Co-operation and Development (2016). *States of Fragility 2016, Understanding Violence*. Accessible: http://www.oecd.org/dac/states-of-fragility-2016-9789264267213-en.htm

The Peace Research Institute Oslo (PRIO). *Battle Deaths Data*. Accessible: https://www.prio.org/Data/Armed-Conflict/Battle-Deaths/

United States Agency for International Development (USAID) (2005). *Fragile States Strategy*. Accessible: https://www.fragilestates.org/about/articles-and-publications/topics/why-matter/fragile-states-strategy/

World Bank (2017). *Country Policy and Institutional Assessment (CPIA)*, the World Bank Group. Accessible: http://databank.worldbank.org/data/reports.aspx?source=Country-Policy-and-Institutional-Assessment

World Bank (2018). *Fragility, Conflict & Violence: Overview*, World Bank. Accessible: http://www.worldbank.org/en/topic/fragilityconflictviolence/overview

Data

The Fragile States Index (FSI): Produced by The Fund for Peace, it is a critical tool in highlighting the normal pressures that all states experience and identifying when those pressures are pushing a state towards the brink of failure. The Fund for Peace collects thousands of reports and information from around the world, detailing the existing social, economic and political pressures faced by each of the 178 countries analysed.

The FSI is based on The Fund for Peace's proprietary Conflict Assessment System Tool (CAST) analytical platform. Using various algorithms, this analysis is then converted into a score representing the significance of each of the various pressures for a given country. Scores are derived for every country based on 12 key political, social and economic indicators (which in turn include over 100 sub-indicators).

These include: Demographic pressures, Refugees and IDPs, Group Grievance, Human Flight, Uneven Development, Poverty and Economic Decline, Legitimacy of the State, Public Services, Human Rights, Security Apparatus, Factionalised Elites and External Intervention. Further, the ranking of the fragile state is arrived at by comparing the sum of all pressures. The country with the largest total score is the most vulnerable, which in 2016 was Somalia.

The Organisation for Economic Co-operation and Development (OECD) Report, *States of Fragility 2016, Understanding Violence***:** The newest methodology comes from the OECD. The OECD defines fragility *as a combination of exposure to risk and insufficient coping capacity of the state, systems and /or communities to manage, absorb or mitigate those risks*. Fragility can lead to negative outcomes including violence, the breakdown of institutions, displacement, humanitarian crises or other emergencies.

Other data: We have used World Bank data wherever possible. However, the findings hold with GDP data from the Penn World Tables (version 7.1) as well. Poverty rates are defined by the poverty line of 1.9 USD per day.

We collected data on battle related deaths from UCDP/PRIO and have used type 3 and type 4 deaths to identify years with more than 25 battle-related deaths to define armed conflict. We have used a per capita measure to define civil wars, where a death-toll above 0.08 per 1000 population

is defined as a civil war.

The Polity2 index from the Polity IV dataset is a standard measure of democracies. It categorises the institutions of a country on a scale from -10 (autocracy) to 10 (democracy). For details, see the main text.

From the World Values Survey, we used the answers to the questions regarding the "Feeling of happiness" where responses are 1) very happy, 2) quite happy, 3) not very happy, 4) not happy at all. We have used the share of people responding 3) or 4) to code unhappiness. For life satisfaction we used the answers to the question "Satisfaction with your life", which is coded from 1 (dissatisfied) to 10 (satisfied). We defined dissatisfaction as a score below 5.

Appendix tables and figures

VARIABLES	(1) large decline next year	(2) start of civil war next year	(3) change in political institutions next yer
growth rate	-0.00817***	-0.000906***	-0.00122***
	(0.00158)	(0.000327)	(0.000439)
democracy score over 0	-0.0490***	0.00265	0.165***
	(0.0173)	(0.00963)	(0.0278)
democracy score over 5	-0.0297*	-0.0151	-0.129***
	(0.0172)	(0.0156)	(0.0229)
strong executive constraints	0.0287	-0.0120	-0.0250
	(0.0194)	(0.00845)	(0.0194)
high competitiveness of executive	0.00390	0.0212*	-0.0181
	(0.0212)	(0.0113)	(0.0197)
high openness of executive recruitment	0.00650	-0.0175**	-0.0958***
	(0.0191)	(0.00838)	(0.0211)
share of population discriminated	0.00104	0.0361	-0.0230
	(0.0477)	(0.0361)	(0.0343)
Observations	5,817	5,674	5,827
R-squared	0.046	0.007	0.057
Number of countryid	139	139	139

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A1: A forecasting model of three failure

We report ROC curves from the 'within' models in the following figures:
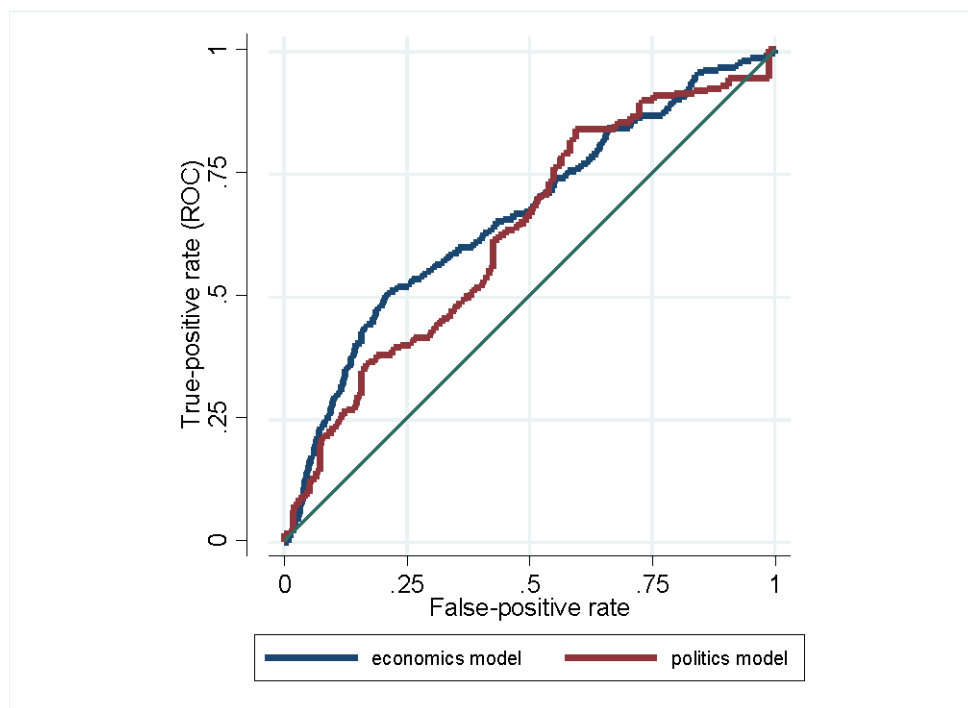
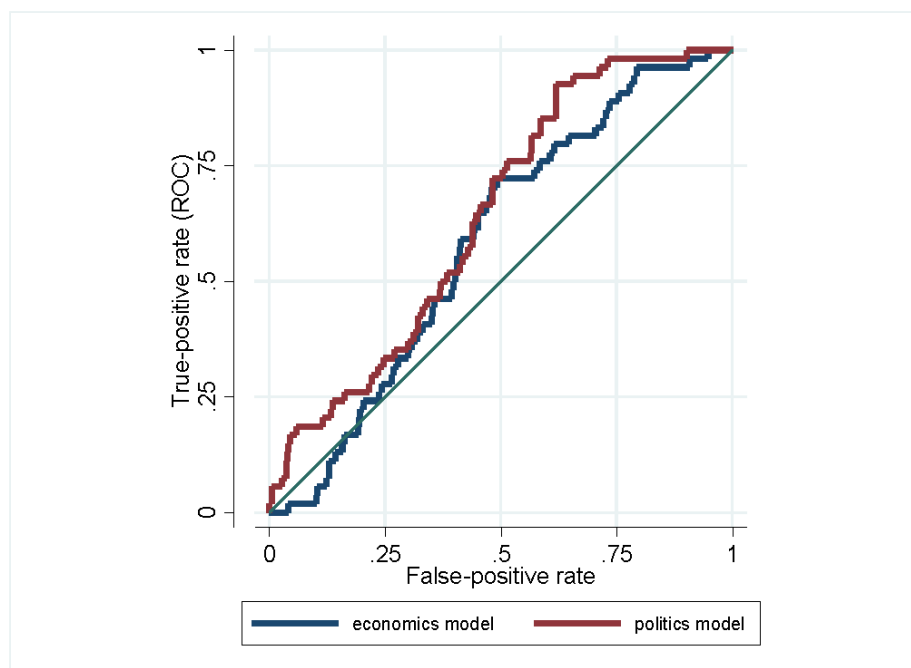**Figure A1**: Politics vs. economics in predicting large economic declines

Figure A2: Politics vs. economics in predicting civil war outbreaks
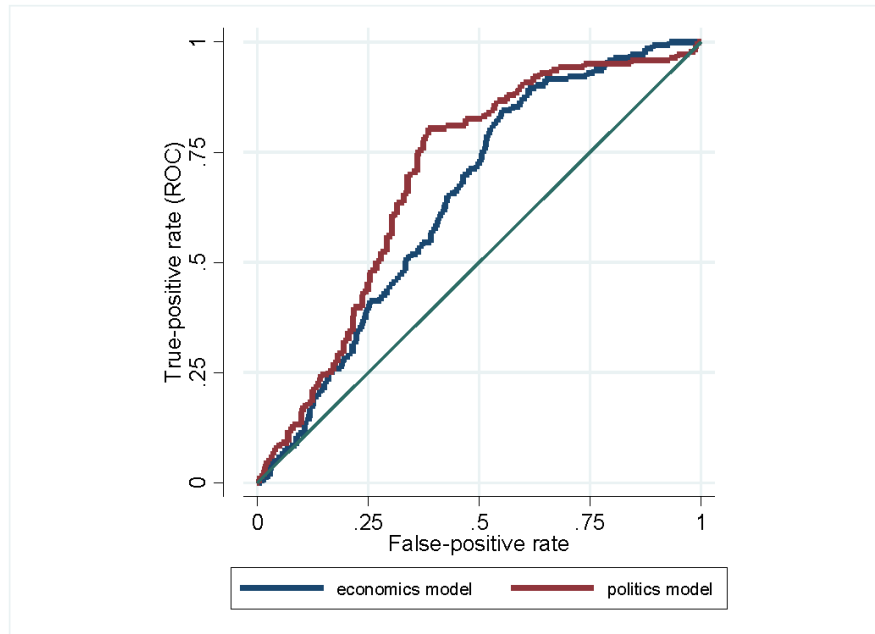
Figure A3: Politics vs. economics in predicting changes in institutions.

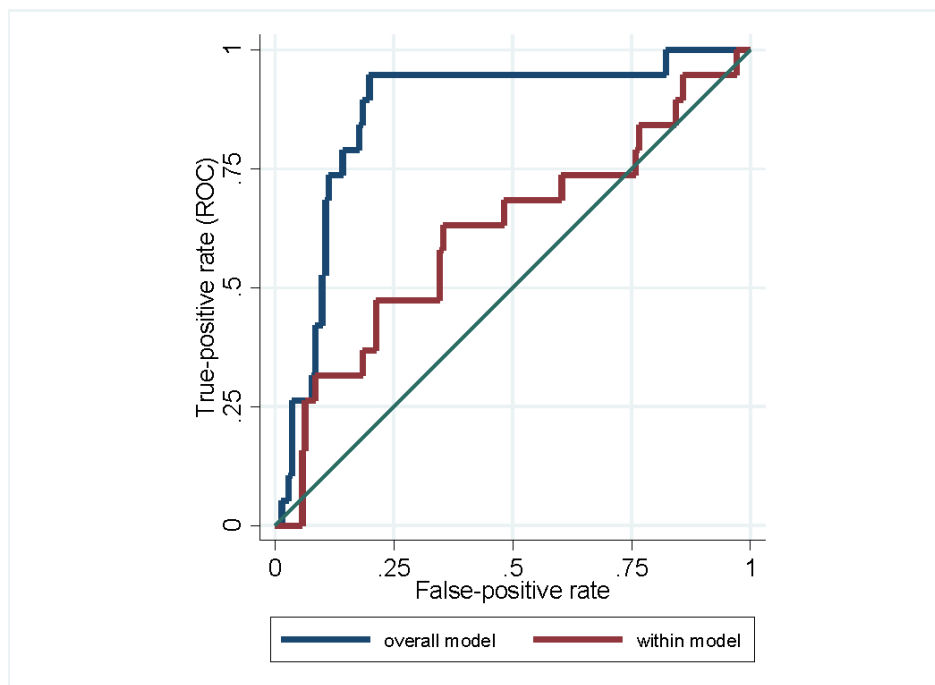The following figures report 'within' and 'overall' models:

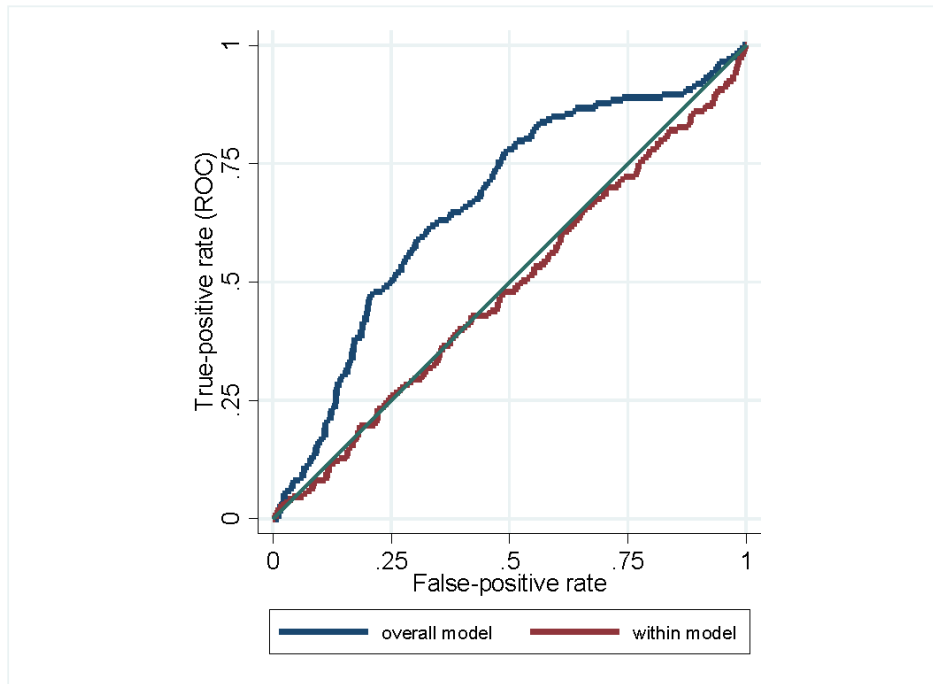**Figure A4**: Predicting failures to vaccinate with our model
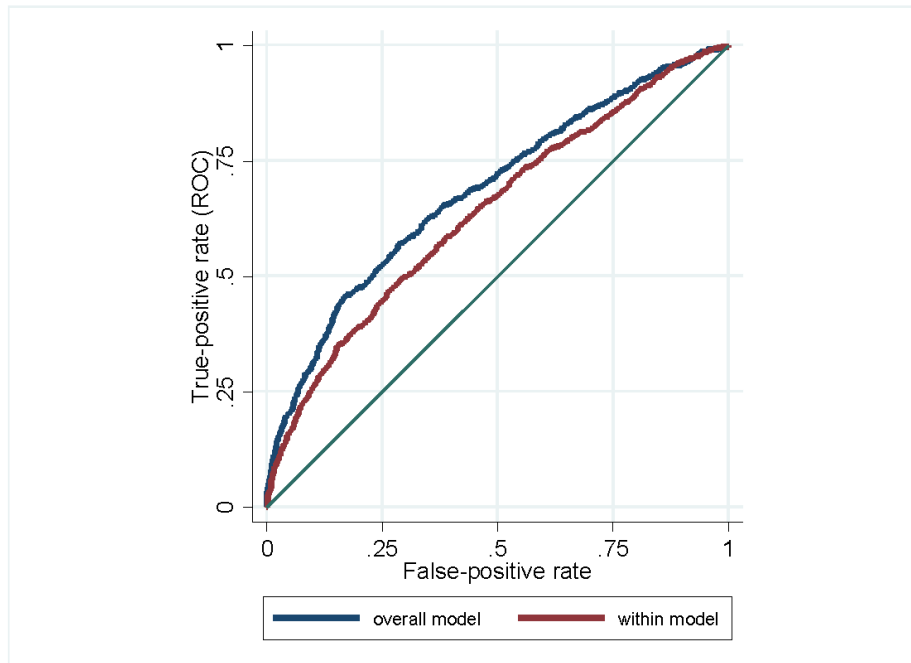
Figure A5: Predicting revolutions with our model

Figure A6: Predicting negative growth with our model
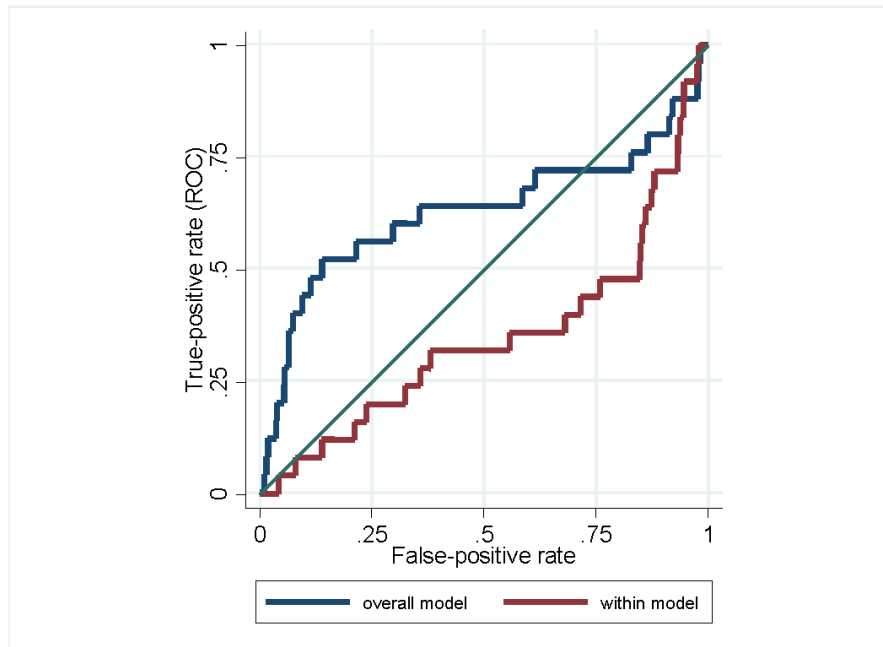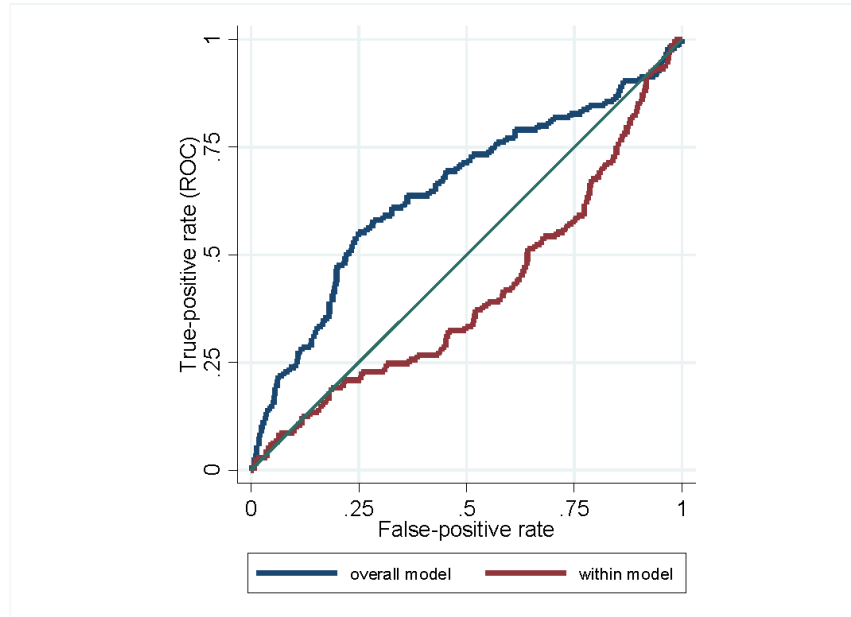
Figure A7: Predicting famines with our model

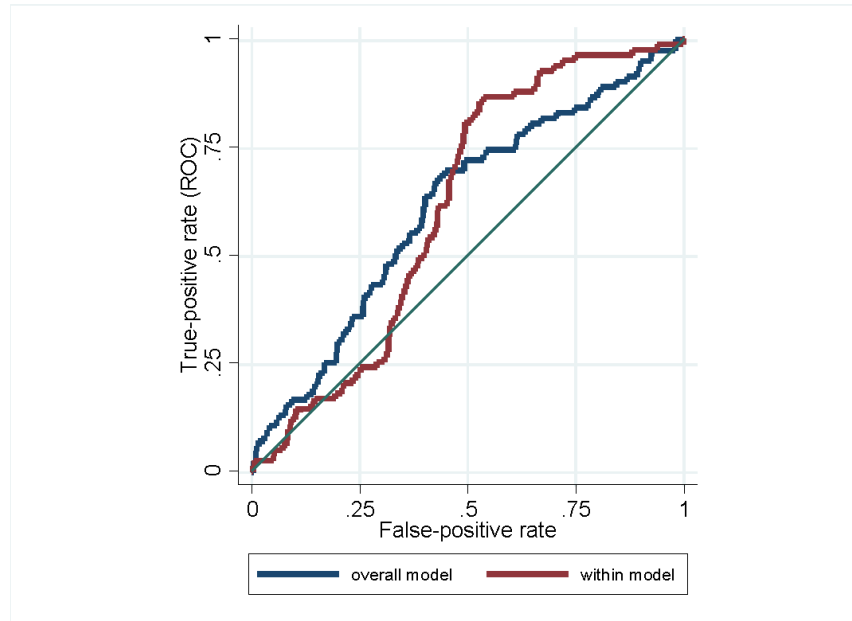**Figure A8**: Predicting armed conflict with our model

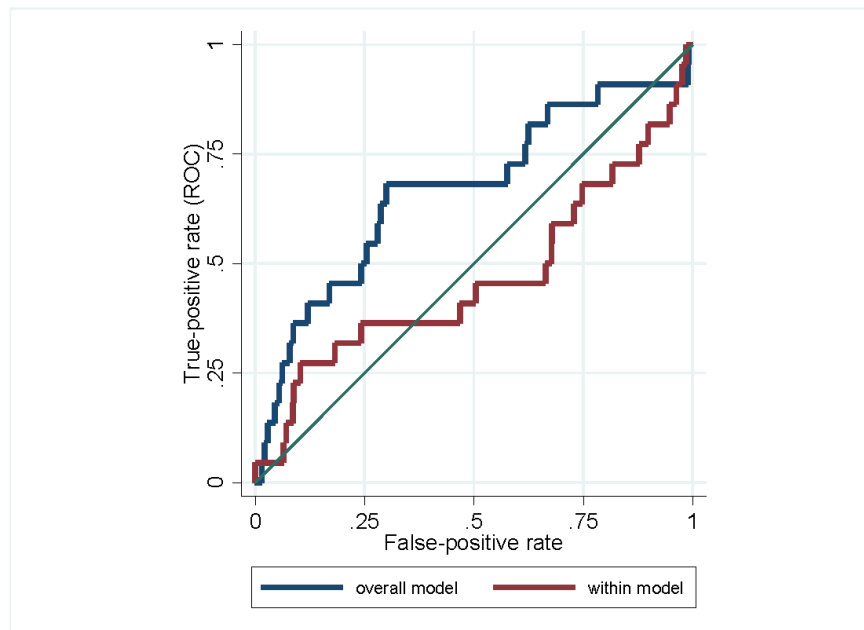**Figure A9**: Predicting purges with our model

Figure A10: Predicting refugees with our model

The LSE-Oxford Commission on State Fragility, Growth and Development was launched in March 2017 to guide policy to address state fragility.

The Commission, established under the auspices of the International Growth Centre (IGC), is sponsored by the London School of Economics and Political Science (LSE) and the University of Oxford's Blavatnik School of Government. It is funded from the LSE Knowledge Exchange for Innovation (KEI) Fund and the British Academy's Sustainable Development Programme through the Global Challenges Research Fund.

www.theigc.org/fragilitycommission